FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Evaluation of Depth Map Post-processing Techniques for Novel View Generation

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medieninformatik

eingereicht von

Matej Nezveda
Matrikelnummer 0402530

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: ao.Univ.Prof. Dipl.-Ing. Mag. Dr. Margrit Gelautz
Mitwirkung: Dipl.-Ing. Mag. Nicole Brosch

Wien, 15.03.2014

(Unterschrift Verfasser) (Unterschrift Betreuung)

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

Evaluation of Depth Map Post-processing Techniques for Novel View Generation

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Media Informatics

by

Matej Nezveda
Registration Number 0402530

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: ao.Univ.Prof. Dipl.-Ing. Mag. Dr. Margrit Gelautz
Assistance: Dipl.-Ing. Mag. Nicole Brosch

Vienna, 15.03.2014

(Signature of Author) (Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Matej Nezveda
Wagmeisterstraße 7, 3300 Amstetten

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum) (Unterschrift Verfasser)

Acknowledgements

I would like to thank my supervisor Margrit Gelautz and her project assistance Nicole Brosch for their generous support throughout the entire development process of this thesis.

I would also like to thank emotion3D GmbH, especially Florian Seitner, Tom Wilson and Georg Braun, for providing their software *Stereoscopic Suite X3* and for help and support regarding the use of this software.

A big thank you for all my friends, relatives and other people who participated in the subjective studies. I know that such studies are demanding and exhaustive.

I would like to thank the federal aid for students in Austria for their financial support of this thesis. Major parts of the work were carried out in close collaboration with the project *Paint3D*, a cooperation between TU Vienna and emotion3d GmbH supported by the Technology Agency of the City of Vienna (ZIT). During the final stage of the thesis, financial support was obtained through the project *Hyperion3D* under the program *IKT der Zukunft* of the Austrian Research Promotion Agency (FFG).

Finally, I would like to thank my parents Anton and Zuzana Nezveda who made it possible for me to study.

Above all, I cannot express my full gratitude to my partner, Dania, who constantly supported me through this entire thesis.

Abstract

Depth-Image-Based Rendering (DIBR) is a key technology for the processing and distribution of three-dimensional (3D) content. Given a two-dimensional image from a scene that was taken from a specific viewpoint and a corresponding depth or stereo-derived disparity map, DIBR enables the generation of images (i.e., novel views) that capture synthesized viewpoints of the scene. The ability of DIBR to synthesize novel views enables the generation of enhanced 3D content (i.e., additional views) for stereoscopic and multi-view displays and gives control over the 3D depth impression (e.g., adjusting the depth range). In DIBR the quality of an underlying depth map contributes to the quality of the novel views generated from it. For example, mismatches, misalignments of depth and color edges or over-smoothed depth edges can lead to visible artifacts in the novel views. We conduct a user study to investigate the effects of depth map post-processing on the perceived quality of 3D content that contains a novel view. A test environment for subjective quality assessment of the visual quality is introduced. In our study we find that filters based on local smoothing, i.e., the bilateral filter and the guided image filter, achieve significantly higher quality scores than filters based on local statistics or the unprocessed counterparts. In addition, our results indicate that the depth range within a scene has a strong impact on the visual quality of DIBR-based novel view generation and the effectiveness of depth post-processing. Furthermore, the obtained subjective results are compared against ten objective quality metrics. We observe only a weak correlation between subjective and objective quality results, which confirms the necessity of user studies in this field.

Kurzfassung

Depth-Image-Based Rendering ist eine Schlüsseltechnologie zur Verarbeitung und Verbreitung von dreidimensionalen (3D) Inhalten. Dabei können ausgehend von zweidimensionalen Bildern und dazugehörigen Tiefenkarten neue virtuelle Ansichten berechnet werden, welche in den Originalaufnahmen so nicht existiert haben. Diese neuen Ansichten können dazu verwendet werden, um einerseits 3D Inhalte für stereoskopische Bildschirme, einschließlich Multi-View-Displays, zu erstellen, und andererseits die Tiefenwirkung an die jeweiligen Ausgabegeräte anzupassen. Dabei hängt die Qualität der neuen Ansichten von der Qualität der zugrundeliegenden Tiefenkarten ab. So können Fehler in der Tiefenkarte zu sichtbaren Artefakten in den neuen Ansichten führen. Im Rahmen dieser Diplomarbeit werden unterschiedliche Nachbearbeitungsfilter untersucht, welche die Qualität der Tiefenkarten und in weiterer Folge die Qualität der neuen Ansichten verbessern sollen. Die Evaluierung wird mittels einer subjektiven Benutzerstudie durchgeführt. Dabei verbessert die Nachbearbeitung der Tiefenkarten mittels Filtern, die auf einer lokalen Glättung basieren wie beispielsweise der Bilaterale Filter und der Guided Image Filter, die subjektiven Ergebnisse der Novel Views signifikant. Die subjektiven Ergebnisse lassen weiter darauf schließen, dass die Tiefe innerhalb einer Szene sowohl die generelle visuelle Qualität der Novel Views als auch die Effektivität der Nachbearbeitungsfilter beinflusst. Ein Vergleich der subjektiven Ergebnisse mit den Ergebnissen von zehn objektiven Qualitätsmetriken zeigt, dass hier nur ein geringer Zusammenhang erzielt wird, wodurch die Notwendigkeit von subjektiven Untersuchungen in diesem Bereich bestätigt wird.

Contents

List of Acronyms

2D	Two-Dimensional
3D	Three-Dimensional
3D-TV	Three-Dimensional TeleVision
ACR-HR	Absolute Category Rating with Hidden Reference removal
BF	Bilateral Filter
D	Dilation
DCR	Degradation Category Rating
DIBR	Depth-Image-Based Rendering
DSCQS	Double Stimulus Continuous Quality Scale
DSIS	Double Stimulus Impairment Scale
FPF	Foreground Protecting Filter
GF	Guided image Filter
GF+W	Guided image Filter plus Weighting
GT	Ground Truth
IFC	Information Fidelity Criterion
JBMF	Joint Bilateral weighted Median Filter
MOS	Mean Opinion Score
MSSIM	Mean Structural SIMilarity index
NP	No Post-processing
NQM	Noise Quality Measure

PC	Pair Comparison
PSNR	Peak-Signal-to-Noise Ratio
PLCC	Pearson Linear Correlation Coefficient
SSX3	Stereoscopic Suite X3
SDSCE	Simultaneous Double Stimulus for Continuous Evaluation
SSCQE	Single Stimulus Continuous Evaluation
SSIM	Structural SIMilarity index
SSNCS	Single Stimulus Numerical Categorical Scale
UQI	Universal Quality Index
VIF	Visual Information Fidelity
VIFP	Visual Information Fidelity Pixel-based
VQM	Video Quality Metric
VSNR	Visual Signal-to-Noise Ratio
WMF	Weighted Mode Filter
WSNR	Weighted Signal-to-Noise Ratio

CHAPTER 1

Introduction

Three-dimensional television (3D-TV) was first demonstrated by John Logie Baird in 1928 and has gained increasing attention throughout the years. The key developments for a successful establishment of 3D-TV include backward-compatibility to traditional two-dimensional (2D) TV, simple and efficient production of three-dimensional (3D) content, support of a wide range of 3D-TV display technologies, low overhead regarding storage and transmission, and precise control over the 3D depth impression. [25].

Five different stages can be identified for an advanced three-dimensional television system [25]: 3D content creation, coding, transmission, view synthesis and presentation of the 3D content (see Figure 1.1). A key technology within this process is Depth-Image-Based Rendering (DIBR). This depth-based 3D-TV approach enhances the 2D video data with additional depth maps. A depth map provides information about the distance of each pixel in the scene to the

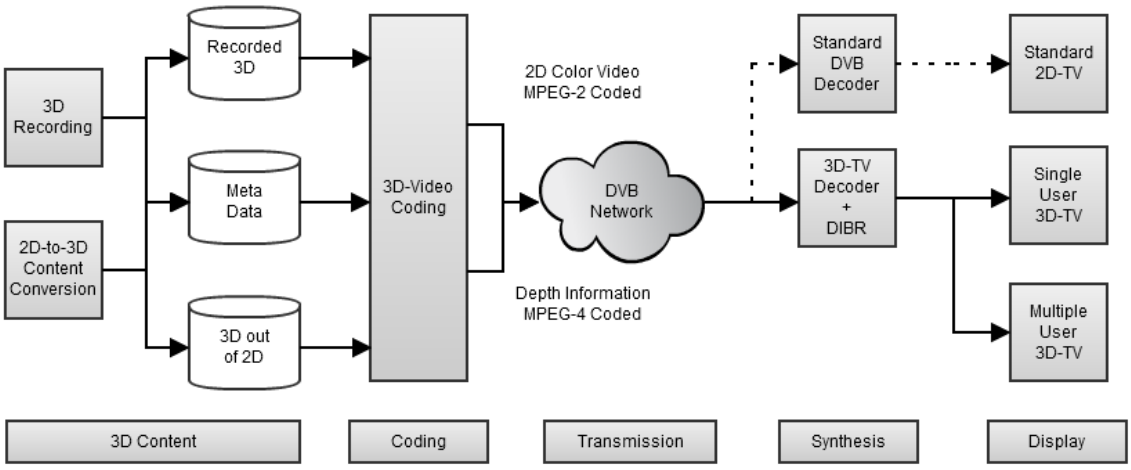


Figure 1.1: 3D signal processing chain (Figure reproduced from [25]).

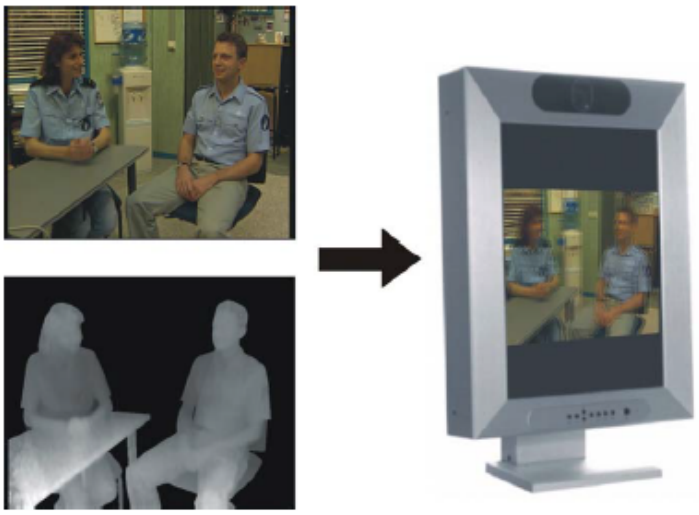
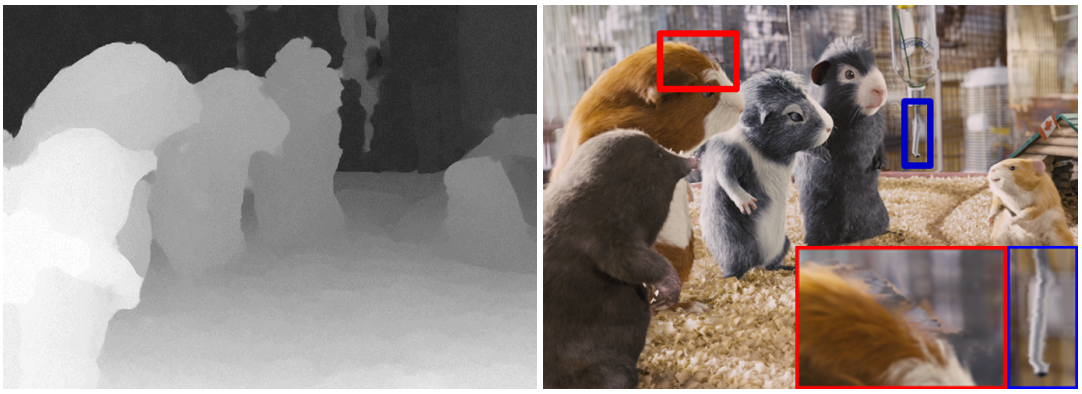


Figure 1.2: Illustration of video-plus-depth concept. A 2D color image with the corresponding depth map is used to create a stereoscopic image. A higher intensity value in the depth map means that the object is closer to the camera (Figure taken from [69]).

camera viewpoint. In consequence, the structure and geometry relationships of the scene are known. Depth maps can be derived from various methods such as special sensors (e.g., [39]), 2D-to-3D conversion (e.g., [9]) or stereo-matching algorithms (e.g., [36]). It should be noted that stereo-matching algorithms do not output depth data but disparity information. Disparity is defined as the displacement of a scene point shown on a stereoscopic image pair and is inversely proportional to the scene depth. A disparity map can be converted into a depth map and vice versa if the camera parameters (e.g., focal length, baseline) are known. For simplicity, in the course of this thesis no distinction will be made between depth and disparity maps [54, 91].

Compared to conventional stereoscopic 3D-TV acquisition using stereo cameras, a depth-based rendering approach has the following benefits: The ability of DIBR to synthesize an abitrary number of novel views (i.e., two and more) from color and depth images enables content generation for single- and multi-view stereoscopic displays and gives precise control over the 3D depth impression. In addition, the efficient representation of the 3D scene geometry using 2D color plus depth images (known as *video-plus-depth* format, see Figure 1.2) enables utilization of existing 2D distribution infrastructures.

In DIBR the quality of an underlying depth map affects the quality of the novel views generated from it. For example, mismatches, misalignments of depth and color edges or over-smoothed depth edges can lead to visible artifacts in novel views (see Figure 1.3). To improve the quality of a given depth map and, more importantly of the novel views generated from it, (depth map) post-processing techniques have been proposed [19, 20, 27, 46, 48]. However, most existing evaluations of these post-processing techniques focus on comparisons of the processed depth maps to depth ground truth, or comparisons of novel views to original views with 2D quality metrics. Alternatively to such objective evaluations, subjective studies can be performed



(a) (b)

Figure 1.3: Illustration of visual errors in a novel view caused by mismatches in the corresponding disparity map. (a) Given disparity map. For visualization, the disparity map is scaled to the intensity range of [0,255]. (b) Novel view. Note the visual artifacts in the novel view, especially in the transition area of foreground and background (red zoom-in) and around thin vertical structures (blue zoom-in).

to measure the quality of 3D content. In particular, it was shown that objective quality metrics are not able to sufficiently determine the subjective visual quality of stereoscopic images that contain a novel view [7]. Thus, the question of the subjective effectiveness of existing depth map post-processing techniques in the context of DIBR/3D-content arises.

1.1 Objective and Contributions

The aim of this thesis is to investigate the effects of depth map post-processing on the quality of stereoscopic image content that contains a novel view. Given an initial depth map created with the stereo-matcher from [23], mismatches and misalignments of depth and color edges can cause visual artifacts in novel views. Therefore, the primary research question is which depth map post-processing techniques can be used to improve the quality of 3D content that contains novel views. In particular, the following methods are considered: (1) depth map post-processing with edge-preserving filters that perform local smoothing [32, 75], (2) depth map post-processing with edge-preserving filters that perform local statistics [38, 53] and (3) depth map post-processing that especially focuses on depth and color edge alignment [31, 48].

This involves a further research question as a suitable method for the evaluation of the quality improvements must be determined. On the one hand this method should especially address the artifacts introduced by the depth based rendering approach and on the other hand permit to draw conclusions on the post-processing methods used. In this context, we perform both subjective and objective evaluations. Concerning the former, we adopt the pair comparison method [10] to fit our special needs. In our study, the two stereoscopic images that form a pair are not presented

in a sequential manner but the subjects can freely switch between the two stereoscopic images. This modification eases the assessment task. Concerning the assessment of the post-processing methods, we compare the subjective results to the objective results obtained from ten objective quality metrics. These metrics perform an automatic evaluation of the (stereoscopic) content.

Our results will show that post-processing the depth maps can significantly enhance the subjective quality of stereoscopic images that contain a novel view. In particular, edge-preserving filters that operate on local smoothing (e.g., bilateral filter) achieve the overall best results. We further find that the investigated objective quality metrics are not suitable to predict the perceived quality of DIBR/3D-content.

1.2 Outline of the Thesis

The current chapter has described the purpose and motivation of the thesis. The rest of the thesis is structured as follows:

Chapter 2 describes different visual quality assessment techniques of stereoscopic content. First, the principles of human (stereoscopic) vision are explained. Second, different types of distortions that can appear in stereoscopic content are discussed. These distortions are grouped into stereoscopic artifacts (e.g., artifacts caused by camera configuration, compression or display characteristics) and DIBR artifacts (e.g., artifacts introduced by the rendering approach itself). Third, different subjective quality assessment methodologies for the evaluation of stereoscopic content are addressed. Fourth, objective quality metrics for an automatic quality evaluation are discussed. A special focus is put on the suitability of these objective quality metrics for the evaluation of stereoscopic content that contains novel views.

Chapter 3 provides an overview of existing methods for depth map post-processing. The presented methods are grouped into three classes. For each class, the presented methods are explained and its advantages and disadvantages are discussed. In particular, the first group is based on Gaussian filtering, the second group relies on bilateral filtering and the third group comprises additional techniques.

Chapter 4 discusses the evaluation methodology applied in the preliminary and the main study. First, the dataset used for both studies is introduced. Next, the chosen subjective quality assessment methodology is described. This description includes details regarding the test environment and the subjective data processing (i.e., quality score computation). Finally, the chosen objective quality metrics are presented. Hereby, the correlation computation of subjective and objective scores is outlined.

Chapter 5 presents the results of the preliminary study. First, the evaluated approaches are briefly described. Then, the results of the subjective and objective quality evaluation are presented. Finally, the impacts on the main study are discussed.

Chapter 6 presents the results of the main study. First, the differences between the evaluated approaches regarding the preliminary study are outlined. Next, the obtained subjective

and objective quality scores are described. Afterwards, the gained insights from the main study are discussed in detail.

Chapter 7 compactly summarizes all covered topics, discusses the overall conclusion that can be drawn and covers possible future work.

CHAPTER 2

3D Content and Evaluation

This chapter focuses on subjective and objective evaluation methods for the quality assessment of 3D content. In this context, we especially concentrate on 3D content that contains a novel view. As an introduction to the topic, Section 2.1 addresses the principles of human 3D perception and outlines the requirements for 3D consumer applications. Section 2.2 points out stereoscopic and DIBR related artifacts which need to be taken into consideration by 3D assessment methods. Section 2.3 describes subjective evaluation approaches, Section 2.4 objective ones. The study design chosen in this thesis incorporates insights gained from the discussed methods and is addressed in Section 4.4.

2.1 Principles of 3D Vision

Stereoscopic systems trick the human visual perception into seeing a 3D scene from a planar image. In order to design a study concerning 3D image quality, we need to understand the human 3D perception and its impacts on 3D technologies.

2.1.1 Human Perception of Depth

The human visual perception is the ability of humans to process and interpret information contained in visible light. The eyes, which are responsible for visual perception, consist of an optical system (i.e., the lens) and a neural system (i.e., the retina). Figure 2.1 illustrates the anatomy of the human eye. The cornea is a transparent membrane and covers the pupil. The pupil controls the amount of light passed through the optical system. It is located at the center of the iris and dilates in the dark and contracts in the light. The lens refracts the light and allows to focus on objects at various distances. The lens can change its shape and control the focal length of the eye. In the retina, the light is converted to an electrical signal and transmitted through the optical nerve to the visual cortex, where the further processing towards visual sensations takes place. The fovea is a central part of the retina. It is exposed directly to the incoming light and is responsible for sharp central vision [30].

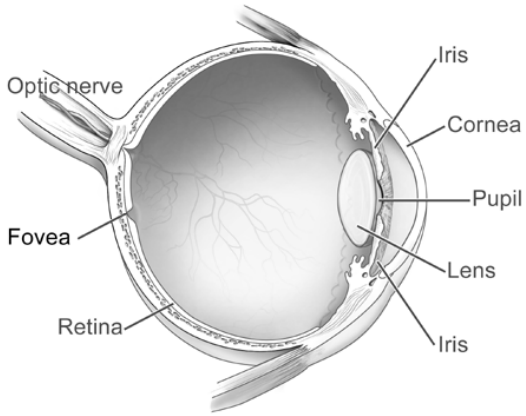


Figure 2.1: Illustration of the human's eye anatomy (Figure taken from [26]).

Depth perception relies on various cues for estimating depth information. These cues can be grouped in oculomotor cues and visual cues (see Figure 2.2). Oculomotor cues involve movement of the eyes, visual cues are based on images projected on the retina:

- **Oculomotor cues** are accommodation and convergence. Accommodation is the process of focusing on objects at different distances from the eye by alterations of the lens. Convergence is the simultaneous movement of the eyes to locate the area of interest. Accommodation and convergence are coupled mechanisms and are associated with adoptions in pupil diameter. The pupil narrows with near accommodation-convergence and widens with far accommodation-convergence [30, 50, 72].

- **Visual cues** can be monocular or binocular. Monocular cues depend either on the image content or on motion parallax. Occlusion, linear parallax and size constancy are three examples for monocular cues. However, many other monocular cues are existent which include texture gradient, aerial perspective, lightning and shading, and defocus blur. Motion parallax corresponds to displacement differences at different lateral positions of the head. As we move, objects that are closer to us show greater displacement than objects that are in the distance [30, 50, 72].

  Although monocular cues provide an indication of depth, binocular parallax provides the strongest depth impression [72]. Due to the lateral displacement of the human eyes, the same visual scene is perceived from two slightly different viewpoints. The resulting differences in the images of the left and right eye are referred to as binocular disparity. The brain fuses the two different images and uses the binocular disparity to obtain the depth information of the visual scene. The fovea is used as reference point for the binocular system and indicates a disparity of zero. When the eyes converge on a specific object, corresponding points in both eyes are stimulated. These points lie on a line called *horopter*. All points that fall within an area close to the horopter called *Panum's fusional range* [56], will fuse and appear in single. All points outside this area will not fuse and appear double (see Figure 2.3(a)) [30, 50, 72].

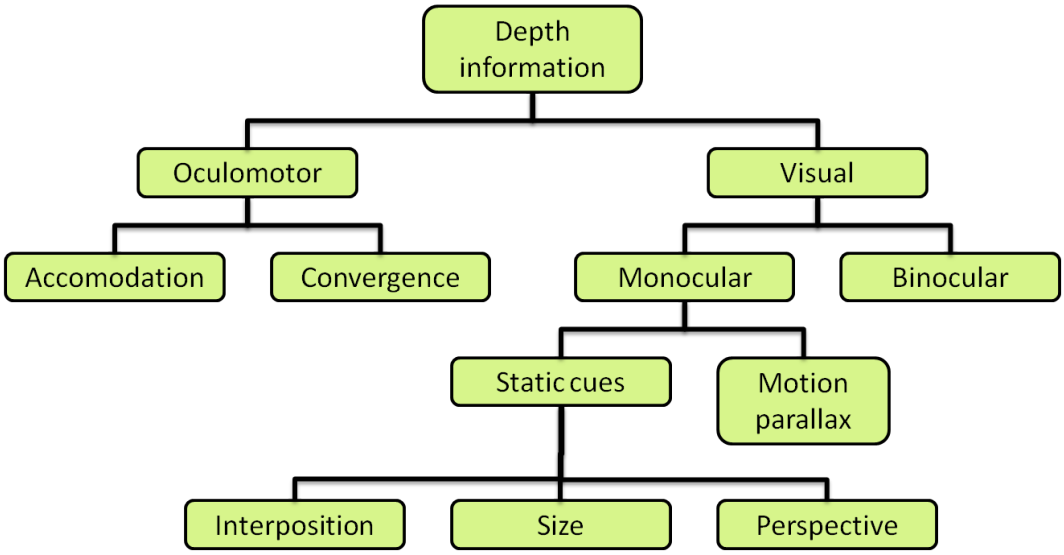


Figure 2.2: Oculomotor and visual depth cues (Figure reproduced from [72]).

2.1.2 Parallax and 3D Displays

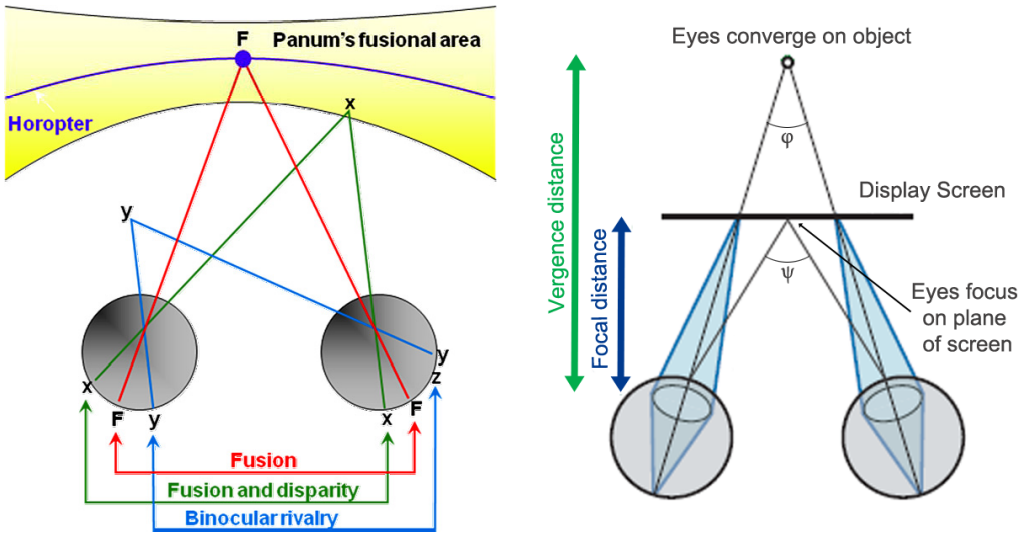
Binocular parallax is the most dominant visual depth cue. Stereoscopic systems exploit this fact and present two slightly different views to the viewer's eyes. The resulting difference in relative positions for a scene point in the two views is referred to as parallax and can be grouped as follows [52]:

- **Zero parallax** occurs when the point is at the same position for the left and the right eye. The eyes focus and converge at the same distance, e.g., the monitor screen.
- **Positive parallax** occurs when the point is shifted to the left for the left eye and vice versa for the right eye. The eyes focus on the screen but converge on a point behind the screen.
- **Negative parallax** occurs when the image is shifted to the right for the left eye and vice versa for the right eye. The eyes focus on the screen but converge on a point in front of the screen.

These three different kinds of parallax allow to create a feeling of depth in the content. Zero parallax is used as reference point and produces no depth impression, whereas positive parallax results in points seen behind the screen and negative parallax results in points seen in front of the screen. However, an exhaustive use of the parallax techniques can lead to unpleasant viewing situations [44].

2.1.3 Visual Discomfort

Visual discomfort is experienced subjectively and may result in eye strain, headache or tension in the neck and shoulder area. To avoid visual discomfort in consumer applications (e.g., stereo-



(a) Panums's fusional range

(b) accommodation-convergence conflict

Figure 2.3: Performance limits of the binocular system: (a) Object F is fixated by both eyes, passes through the horopter and will fuse to a single image. Object X is located within Panum's fusional area and thus will fuse as well. Object Y is located outside of Panum's fusional area, provokes binocular rivalry and will not fuse (Figure taken from [58]). (b) Eyes focus on the screen but are fixated on an apparent object. If the difference in angles between convergence (φ) and accommodation (ψ) exceeds an acceptable maximum level, visual discomfort occurs (Figure inspired by [34, 72]).

scopic television), several factors of binocular vision have to be considered. Excessive screen disparities may not fall within Panum's fusional range and are a potential cause of visual discomfort (see Figure 2.3(a)). Another cause is the mismatch between accommodation and convergence when viewing stereoscopic content on a display. The accommodation stimulus will focus on the screen, the convergence stimulus will vary depending on the depth of the scene. If this conflict exceeds an acceptable maximum level, it can result in a blurred image caused by loss of accommodation, double vision caused by loss of fusion, or both. This acceptable maximum level should not exceed a specific angular difference between accommodation and convergence (see Figure 2.3(b)) [44, 58].

Therefore, limits for a comfortable zone of viewing can be defined. Within this limits, visual discomfort caused by excessive screen disparity or accommodation-convergence conflict can be prevented. This limit is defined as one arc min of the screen disparity. Arc min is a measure of binocular disparities in human visual perception, where a 1 cm wide object 57 cm away from the eye subtends approximately one arc min. Table 2.1 shows limits of comfortable viewing, according to the one arc min rule [44].

However, three factors have been determined that can cause visual discomfort even within

	Limits for comfortable viewing	
View distance (mm)	**Near (mm)**	**Far (mm)**
500	440	580
1000	780	1400
2000	1300	4800
3000	1600	23000

Table 2.1: Limits of comfortable viewing according to 1 arc min of the screen disparity. The limits are obtained in respect to the viewing distance from the observer (Table from [44]).

this limit. The first factor concerns fast motion in spatial or depth direction, which is demanding on accommodation and convergence. The second factor concerns uncertain and unnatural depth perception caused by unnatural blur. The last factor concerns 3D artifacts which can cause spatial and temporal inconsistencies. The following section discusses these potentially occurring 3D artifacts in more detail [44].

2.2 Types of Distortions

The quality of 3D content can be affected by various distortions. In our case, we identify two types of distortions than can lead to a degradation of 3D content that contains a novel view. First, general stereoscopic artifacts caused by camera configuration, compression or characteristics of the display technology used [51]. Second, DIBR artifacts induced by the depth based rendering approach itself [5].

2.2.1 Stereoscopic Artifacts

Stereoscopic artifacts can be introduced in various stages of 3D content delivery and also affect different layers of human 3D vision [3]. The stages are capture, representation, coding, transmission and visualization whereas the layers are structure, color, motion and binocular. In this context, the layers are defined as follows: Structure denotes the perception of contours and texture within images and corresponds to spatial and color-less vision. Color indicates color vision and motion corresponds to motion vision. Binocular means the perception of the environment with two eyes. Figure 2.4 visualizes the classification and dependencies of artifacts and layers of human vision. Note that the artifacts can be introduced in different stages of the 3D signal processing chain and also affect different layers. In the following, six selected distortion types that can appear in stereoscopic conditions are explained in more detail:

- **Depth plane curvature** is caused by a toed-in camera setup where the two cameras are positioned with an angle to each other. Objects at the corner of the image appear further away from the observer than objects in the middle of the image. This can lead to a misleading perception of relative object distances and annoying image motions during panning [51, 83].

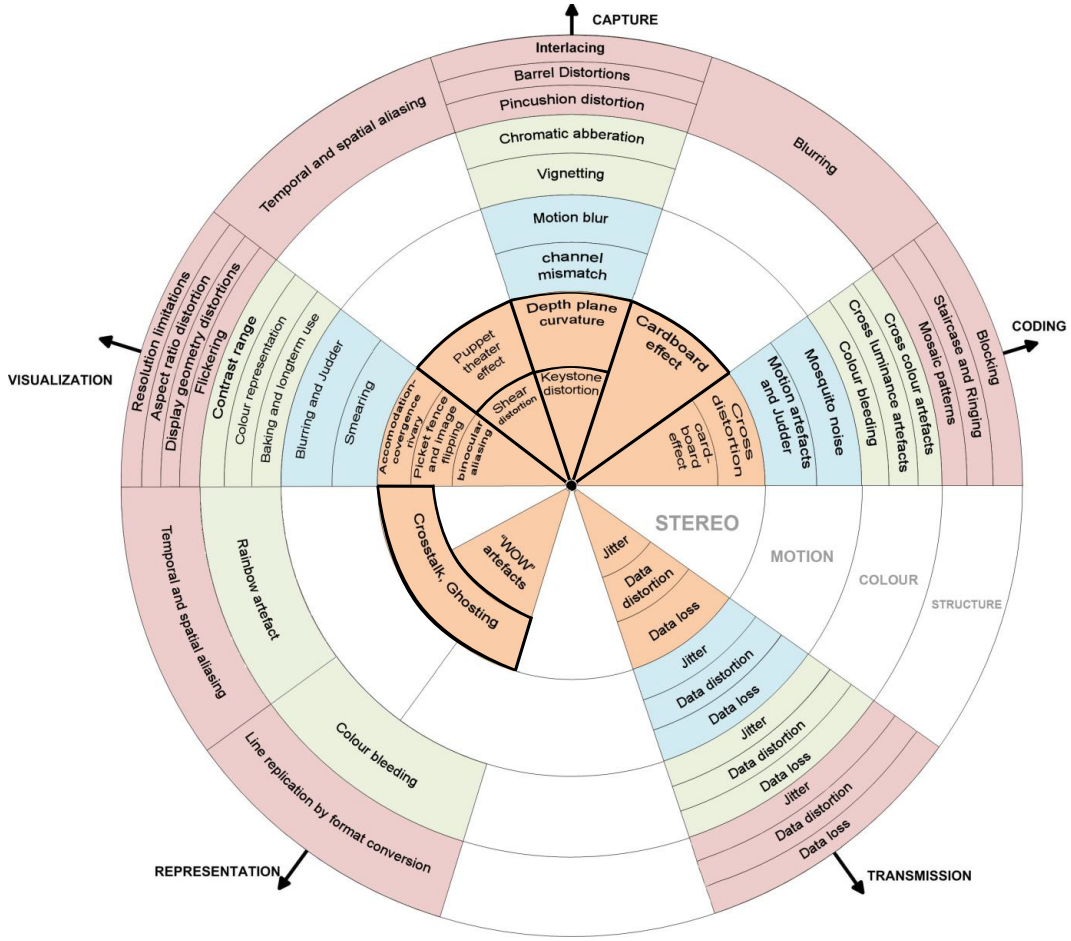


Figure 2.4: Classification of stereoscopic artifacts. Artifacts highlighted in bold are explained in the text, a description of the other artifacts can be found in [2] (Figure taken from [3]).

- **Keystone distortion** is also introduced by a toed-in camera setup and is related to depth plane curvature. However, a shift-sensor camera system does not exclude the keystone distortion. A vertical difference between corresponding points is introduced and as a result, the image looks like a trapezoid. The distortion is greatest in the corner [51, 83].

- **Crosstalk** can be caused by an incorrect separation between left and right view or an incorrect head position. It can be expressed in distortions like ghosting, shadowing or double contours. In addition, it depends on the display system used (active, passive, with/without glasses) and can occur in stereoscopic as well as auto-stereoscopic displays [51].

- **Puppet-theatre effect** is a visual size distortion in which objects appear unnaturally small. The perception of this effect depends on prior knowledge about the appearance of the misaligned object. More familiar objects like humans are more affected [51, 85].

- **Cardboard effect** leads to wrong depth perception of objects. The objects appear flattened as if they would lie on a cardboard in the 3D scene. The causes of this effect are acquisition parameters (e.g., focal length, camera baseline, convergence distance) and coding parameters (e.g., coarse quantization of the depth map) [51].
- **Shear distortions** appear in viewing conditions where the viewing position for the intended depth perception is fixed. A change in the viewing position suggests the impression that the stereoscopic image follows the viewing position of the observer. Employment of head or eye tracking can be used for a correction of the viewpoint in order to avoid this effect [51, 83].

2.2.2 DIBR Artifacts

DIBR utilizes depth maps to generate novel views. Therefore, the quality of the novel views depends on the underlying depth map. Inaccurate depth values can cause visual artifacts in the novel views because the mapping of the pixels is related to the provided depth information [5]. Another issue in DIBR consists in the handling of exposed areas (i.e., disocclusions in novel views). Hole-filling and depth map post-processing are two possible approaches to tackle these exposures [24]. Nevertheless, both approaches might lead to artifacts in the novel views. Among others, the following artifacts can be observed in novel views:

- **Rubber sheet artifacts** occur at object boundaries and stretch the foreground objects to the background objects behind them. These artifacts look like as if the foreground was blurred in the background. They are introduced by hole-filling based on linear interpolation of foreground- and background image color [24, 29].
- **Strip-like impairments** are related to rubber sheet artifacts. They also appear at object boundaries and are caused by a misalignment between foreground and background pixels during hole-filling. Contrary to rubber sheet artifacts which look blurry, these impairments introduce strips at the object boundaries [24].
- **Geometric distortions** result from depth map post-processing that over-smoothes depth edges in horizontal direction. The novel view is created out of the original view and the associated depth map. Therefore, changes in the depth map affect the appearance of the rendered objects which can lead to geometrically misrepresented objects [5].

2.3 Subjective Quality Assessment

Subjective quality assessment is used to measure the quality of images or video sequences. The subjective evaluation methods used are standardized by the International Telecommunication Union (ITU). In 2012, ITU has released the recommendation ITU-R BT.2021 [12] that is concerned with the assessment of stereoscopic content. In particular, the methods Absolute Categorical Rating (ACR)[1], Double Stimulus Continuous Quality Scale (DSCQS), Pair Comparison (PC) and Single Stimulus Continuous Quality Evaluation (SSCQE) are recommended

[1] This methodology is also known as single stimulus method.

Reference	Abbr.	Method	M/S	Presentation	Voting
[12, 55]	ACR	Absolute Category Rating	M/S	Test	5Q/5I
[55]	ACR-HR	Absolute Category Rating with Hidden Reference	M	Test	5Q/5I
[12, 14]	DSCQS	Double Stimulus Continuous Quality Scale	M/S	Test&Ref	C
[14]	DSIS	Double Stimulus Impairment Scale	M	Ref/Test	C
[12, 55]	PC	Pair Comparison	M/S	Test/Test	P
[12, 14]	SSCQE	Single Stimulus Continuous Quality Evaluation	M/S	Test	CC

Table 2.2: Overview of subjective assessment methods mentioned in the course of this thesis. First and second column denote reference and name of subjective methodology, respectively. Third column denotes whether subjective methodology is recommended by ITU for monoscopic (M) and/or stereoscopic (S) viewing conditions. Fourth column denotes display order: only test sequence (Test); First reference, then test sequence (Ref/Test); Both sequences simultaneously (Ref+Test); Test and reference in random order (Test&Ref); Two test sequences compared to each other (Test/Test). Fifth column denotes voting type: 5-grade quality scale (5Q); 5-grade impairment scale (5I); Continuous scale, single voting (C); Continuous scale, continuous voting (CC); Preference (P) (Table inspired by [1]).

to evaluate image/video quality, depth quality and visual comfort of stereoscopic content. Since the recommendation ITU-R BT.2021 does not deal with 3D content which contains a novel view, additional subjective methodologies originally proposed by ITU for 2D content have been used for assessing the quality of DIBR content. For example, the MPEG call for proposals on 3D video coding technology [42] evaluates the proposed technologies with Double Stimulus Impairment Scale (DSIS). The call addresses the questions of efficient compression and high quality view reconstruction. Bosc et al. [7] investigate Absolute Category Rating with Hidden Reference Removal (ACR-HR) for quality assessment of 3D content generated through DIBR.

In the following, the six mentioned subjective methodologies are described in more detail, whereas Table 2.2 briefly summarizes their most important factors. It should be noted that trial structure, grading scale and opinion score calculation differ with regard to the methodology used. The grading scale can be discrete, where the quality is rated by selecting fixed rating points or categories, or continuous, where intermediate values between the rating points or categories are possible. If a methodology supports both grading scales, one grading scale can be selected. All methods except for SSCQE can be used for still images and video content. SSCQE is designed to address the impact of quality fluctuations over time and thus is intended for video content. In the course of the following description of the subjective methodologies, the terminus *reference stimulus* is used to describe the ground truth visual content, whereas the terminus *test stimulus* is used to describe a degraded version of the reference stimulus:

- **Absolute Categorical Rating** (ACR) [12,55]: Each test stimulus is individually presented to the subjects. The subjects then assess the test stimulus in terms of perceived quality on a five grade scale. This scale can either be discrete or continuous. The test stimuli should be shown in random order. The final quality scores are expressed as Mean Opinion Score (MOS). For one test stimulus it is defined as the mean of all individual scores.

- **Absolute Categorical Rating with Hidden Reference** (ACR-HR) [55]: Compared to ACR, this methodology differs in the quality score computation. The reference stimulus must be included in the evaluation and is rated like any other test stimulus. The quality score of a test stimulus is determined by the difference between its MOS and the associated hidden reference stimulus.

- **Double Stimulus Continuous Quality Scale** (DSCQS) [12, 14]: Each test stimulus is shown in pairs with the reference stimulus. Reference and test stimulus of each pair are presented in random order and the subjects are not informed about their assignment. Depending on the visual content, the pairs are shown several times and the assessment task is performed in the last representation. The subjects are asked to rate the quality of both stimuli on a continuous quality scale. The mean difference between the reference and the test stimulus for all subjects represents the quality score of one test stimulus. DSCQS is especially practical when it is not possible to provide test conditions that cover the full range of quality conditions.

- **Double Stimulus Impairment Scale** (DSIS) [14]: The reference and test stimulus are presented in pairs. Contrary to DSCQS, the reference stimulus is always shown first, followed by the test stimulus. The subjects rate the quality of the test stimulus on a five grade discrete scale keeping in mind the reference stimulus.

- **Pair Comparison** (PC) [12,55]: The entire set of test stimuli is grouped into pairs. Hence, for a set of N stimuli, $\frac{N(N-1)}{2}$ pairs are generated. The pairs are shown to the subjects, who decide which of the two stimuli is better. Depending on the study design, a PC can also involve ties, where both stimuli are rated equally. This method allows to compare stimuli which only differ slightly in quality. However, the quality results are relative in terms of preferences. Furthermore, the assessment task is time consuming.

- **Single Stimulus Continuous Quality Evaluation** (SSCQE) [12,14]: This method specifically addresses artifacts that appear over time, such as flickering, and hence uses time-varying stimuli, such as video content. The quality of a single test stimulus is evaluated continuously by operating a score recording device (e.g., a slider). The content that belongs to different genres (e.g., drama, sport or news) is shown in random order. Each genre should take at least five minutes. The overall evaluation can be split into several evaluation sessions. In this case, each session should cover all genres and quality parameters. However, per session not all genre and parameter combinations have to be considered. The duration of one evaluation session should be between 30 and 60 minutes.

In addition to the evaluation method, the design of the subjective evaluation including general viewing condition, test material (i.e., stimuli) selection, number of subjects and session

Item	Value
Room illumination	low
Chromaticity of background	D_{65}
Peak luminance	70-250 cd/m^2
Monitor contrast ratio	≤ 0.02
Ratio of luminance of background behind picture monitor to peak luminance of picture	≈ 0.15

Table 2.3: General viewing conditions for subjective quality assessment of stereoscopic content in a laboratory environment (information taken from [13]).

duration has to be considered as well. As we concentrate on stereoscopic content in our evaluation, the recommendations of ITU-R BT.2021 are described in more detail [12]:

- **General viewing conditions**: Viewing conditions like screen luminance, contrast, background illumination and viewing distance should be consistent with viewing conditions used for 2D content. This is motivated by practical considerations. On the one hand, the users will use the same display to watch 2D and 3D content, on the other hand the progress in performance of 3D-TV can be compared to standard high definition television. Table 2.3 summarizes the principal conditions. Note that these settings are indented for a laboratory environment, thus different settings have to be adapted to a home environment.

  Concerning viewing distance, the so called *design viewing distance* has to be used for the evaluation of stereoscopic content. It is defined as the distance at which two adjacent pixels subtend at one arc min. It can also be expressed in multiples of the picture height. It is 4.8 and 3.1 times the picture height for image resolutions of 1280$\times$720 and 1920$\times$1080.

- **Test material**: The test material should be selected according to the addressed research question. If possible, the mean, standard deviation and minimum/maximum of the parallax should be provided. The parallax should lie within the visual comfort limits. For example, for a 1920$\times$1080 image resolution watched from a distance of 3.1 times the picture height, the visual comfort limits are approximately $\pm 2\%$ and $\pm 3\%$ of the screen parallax.

- **Subjects**: A minimal amount of 30 subjects is recommended. All subjects should pass a visual acuity test using Snellen charts, a color vision test using Ishihara plates and a stereo vision test. Appendix B gives detailed information on these visual tests and shows examples.

- **Session design**: The session duration should be in the range of 20 to 40 minutes. Furthermore, the subjects should be informed about the goal of the study and possible occurring side effects like visual discomfort.

2.4 Objective Quality Assessment

In order to evaluate (stereoscopic) content automatically and in a more efficient manner, objective quality assessment is used. The goal of such objective quality metrics is to predict the perceived quality of (stereoscopic) image/video content. In particular,

- **2D intended quality metrics** are originally developed for 2D content. However, they are used by various authors for the evaluation of 3D content, which is generated either by stereoscopic acquisition or DIBR [19, 43, 68].
- **3D specific quality metrics** are specially designed for stereoscopic content. These metrics are either based on 2D-like methods and perform the quality computation only on the images or they also take the additional depth information into account [5].

This section gives an overview of different quality metrics in these two categories. In addition, their suitability and reliability for video-plus-depth content is addressed. Concerning the latter, the Pearson linear correlation coefficient (PLCC) is a commonly used measure [76]. It quantifies the linear correlation between obtained objective and subjective scores and returns a value in the interval $[-1; +1]$. A value of $+1$ corresponds to a positive correlation, a value of -1 corresponds to a negative correlation and a value of 0 corresponds to zero correlation.

2.4.1 2D Intended Quality Metrics

Mean Squared Error (MSE) [78] measures the squared error between an $m \times n$ image I and its noisy approximation $\bar{I}$. It is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - \bar{I}(i,j)]^2, \tag{2.1}$$

where i and j are pixels within the image.

Signal-to-Noise Ratio (SNR) [17] measures the ratio between a signal and the background noise. For an $m \times n$ image I and its noisy approximation $\bar{I}$ it is defined as:

$$SNR = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j)]^2}{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - \bar{I}(i,j)]^2}. \tag{2.2}$$

Peak Signal-to-Noise Ratio (PSNR) [78] is a logarithmic representation of MSE and is defined as:

$$PSNR = 10 log_{10} \frac{MAX_I^2}{MSE}, \tag{2.3}$$

where MAX_I is the maximum possible value of the image I (i.e., 255 for 8-bit images). The PSNR score is a single number and is expressed in decibels (dB). Compared to MSE, PSNR contains no new information but is useful if the images being compared have different dynamic

ranges.

Weighted Signal-to-Noise Ratio (WSNR) [17] uses an additional weighting function to adjust the results of SNR to the impact of the human visual system.

Noise Quality Measure (NQM) [17] is based on SNR and further uses a degradation model to measure the noise injection.

Universal Quality Index (UQI) [77] combines the mean, variance and covariance of pixels. It models any distortion as a combination of the three factors loss of correlation, luminance distortion and contrast distortion. For two images I and $\bar{I}$ the metric output UQI is defined as follows:

$$UQI = \frac{4\sigma_{I\bar{I}}\mu_I\mu_{\bar{I}}}{(\sigma_I^2 + \sigma_{\bar{I}}^2)(\mu_I^2 + \mu_{\bar{I}}^2)}, \tag{2.4}$$

where μ_I and $\mu_{\bar{I}}$ are the mean values of I and $\bar{I}$, σ_I and $\sigma_{\bar{I}}$ are the variances of I and $\bar{I}$, and $\sigma_{I\bar{I}}$ is the covariance of I and $\bar{I}$. The UQI index is in the range $[-1; +1]$, where a score of 1 can only be reached if both images are identical, so $I = \bar{I}$.

Structural SIMilarity index (SSIM) [79] can be considered as an extension of UQI and also combines the mean, variance and covariance of pixels. It operates on local windows of size $N \times N$. The score of two windows x and y is defined as follows:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{2.5}$$

where μ_x and μ_y are the mean values of x and y, σ_x and σ_y are the variances of x and y, and σ_{xy} is the covariance of x and y. $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$, where L is the dynamic range of the pixel values, K_1 = 0.01 and K_2 = 0.03.

In practice, the quality of an image is expressed by only one single value. Therefore, the Mean SSIM (MSSIM) is used as a quality indicator [79]:

$$MSSIM(I, \bar{I}) = \frac{1}{M} \sum_{j=1}^{M} SSIM(x_j, y_j), \tag{2.6}$$

where $I, \bar{I}$ are the reference and distorted image, M is the number of samples in the quality map, and x_j, y_j are the image contents of the local window j. The resulting MSSIM index is in the range of 0 and 1, where 1 corresponds to the best quality score. In the following course of the thesis, the term SSIM corresponds to MSSIM.

Multi-Scale Structural SIMilarity index (MS-SSIM) [80] is an extension of SSIM. In SSIM, the quality score computation is performed by means of a luminance, contrast and structure comparison. However, variations of viewing conditions are not considered. Such variations may include the image resolution or the viewing distance and affect the perceptibility of the image details. Therefore, MS-SSIM performs a low-pass filtering and down-sampling (factor: two) of

the input reference and distorted images. The contrast and structure comparisons are performed on all scale factors, the luminance comparison is performed only at the highest scale factor. The final MS-SSIM is defined as the combination of the all contrast and structure comparisons and the luminance comparison.

Visual Signal-to-Noise Ratio (VSNR) [15] measures the visual fidelity of an image. Thus, it is sensitive to geometric distortions, e.g., spatial shifting, rotations and transformations. The quality score computation consists of two stages. In the first stage, a wavelet-based approach is used to quantify the presence of perceptible distortions. The metric terminates, if no visible distortions are detected. Otherwise, the second stage measures the degradation of the image structure.

Information Fidelity Criterion (IFC) [67] uses a source and a distortion model in order to measure the information shared between the source and the distorted image. Both models are described in the wavelet domain. The quality score is calculated by quantifying the amount of statistical information shared between source and distortion model.

Visual Information Fidelity (VIF) [66] is an extension of IFC, where in addition to the information shared between the source and the distorted image, the information content of the source scene is also considered. Thus, the information fidelity between source and distorted image can be quantified relative to the information content of the source scene.

Visual Information Fidelity Pixel-based (VIFP) [66] is a computational simpler pixel-based version of VIF.

Video Quality Metric (VQM) [61] measures the perceived video quality. It involves calibration techniques, perception-based feature extraction, video quality parameter calculation, and combination of those parameters to a single metric. The calibration steps include spatial alignment, valid region estimation, gain and level offset calculation, and temporal alignment. The video quality parameter calculation considers perceptual video impairments like blurring, jerky/unnatural motion, global noise, block distortion and color distortion. Due to its high correlation results with subjective quality ratings, VQM is adapted as American national standard [73] and ITU recommendation [11, 40].

2.4.2 3D Specific Quality Metrics

Peak Signal-to-perceptible-noise ratio [90] is oriented on PSNR but unlike PSNR it measures the perceptible temporal noise and not the spatial noise. The temporal noise is defined as the inter-frame change in the processed and reference sequence. A just noticeable distortion threshold is used for the detection of perceptible changes. However, the authors mention two limitations of this metric: First, the metric is not suitable for moving camera scenarios because temporal noise within moving objects is not considered. Second, a combination of temporal noise and spatial noise is more realistic according to characteristics of the human visual system.

Perceptual quality metric [41] operates on the luminance channel of the reference and distorted color view and measures the contrast distortion and the luminance difference. In a subjective evaluation, four video-plus-depth videos synthesized with a DIBR algorithm are evaluated by 30 subjects on an auto-stereoscopic display. The proposed metric obtains a higher correlation with MOS than VQM. It achieves an average PLCC of 0.978 whereas VQM has an average PLCC of 0.795.

Color quality metric [71] is also a perceptual based quality metric for DIBR generated content. Similar to [41], only the luminance channel is considered. The basic operation unit of this metric is a 8×8 pixel block. Two distortions are quantified, namely (1) distortion of the block content considering luminance and contrast features and (2) distortion for block boundary detection. The metric is evaluated with twelve test sequences (dataset one: six videos, resolution: 560×420; dataset two: six videos, resolution: 640×1024) and rated by 28 test persons on a stereoscopic display. Correlation coefficients indicate that the proposed metric provides a better alignment with MOS compared to [41]. The proposed metric obtains an average PLCC of 0.963 (dataset one) and 0.943 (dataset one), [41] obtains an average PLCC of 0.607 (dataset one) and 0.557 (dataset two).

Edge-based structural distortion indicator [4] compares the contours of the original view with those of the virtual view. According to a displacement vector acquired from the contours, three quality indicators are extracted: (1) mean ratio of inconsistent displacement vectors per contour pixel, (2) ratio of inconsistent vectors and (3) ratio of new contours. The authors evaluate their proposed indicator against the metrics used in [6] and compare the results to MOS values from subjective evaluations. In all of the cases the indicator is close or the closest to the MOS. However, the proposed method does not take color consistency into account and thus is just an indicator for structural distortions in DIBR views.

Depth based perceptual quality metric [22] assigns more importance to regions where errors caused by DIBR may occur. This is especially in the vicinity of front objects and areas with motion inconsistency between original and synthesized view. It uses a weighting function based on the depth range and also a temporal consistency function to consider motion activity. The inputs of this method are the reference view, the reference depth map and the synthesized view. The output is a per-pixel weighting coefficient map which can be multiplied with the error maps from PSNR or SSIM. The results show that the correlation of PSNR and SSIM can be enhanced with the proposed method. The increase is 0.012 and 0.088 for PSNR and 0.225 and 0.099 for SSIM for the two tested sequences, respectively.

Depth enhanced video quality metric [87] adapts VQM in such a way that the 2D color image quality is combined with depth image quality. The color image quality is based on the VQM scores whereas the depth quality is obtained by separating the depth into a number of depth planes and combining three features out of them. These features are (1) distortion of the relative distance within each depth plane, (2) distortion in the consistency of each depth plane and (3) structural error of the depth. Three color plus depth videos are used for the evaluation of

the metric. The proposed metric correlates better with subjective scores than VQM and obtains PLCC of 0.8369 compared to PLCC of 0.8008 for VQM.

3VQM [70] is a quality metric for DIBR content. It combines three quality measures to get a final quality score, namely (1) temporal outliers, (2) temporal inconsistencies and (3) spatial outliers. Therefore, the quality of the depth map is compared with a so called ideal depth map. In this context the ideal depth map is defined as the depth map that would generate a distortion free virtual view given the same reference image and DIBR parameters. The performance of 3VQM is evaluated with subjective experiments. The test material includes DIBR synthesized videos from depth map and colored video compression, stereo matching and 2D to 3D conversion. The 3VQM metric obtains PLCC of 0.8942. Moreover, all of the results are in the range of two time difference MOS standard deviation and 80 percent of the results are in the range of one times difference MOS standard deviation. Therefore, the authors claim that 3VQM is a significantly consistent and accurate quality measure. However, no comparative study to other objective quality metrics is performed.

2.4.3 Suitability for Video-plus-Depth

Thirteen 2D intended quality metrics and seven 3D specific quality metrics were described in the course of this section. Concerning 3D specific quality metrics, all methods are evaluated in respect to subjective evaluations and all methods except of 3VQM are compared to 2D dedicated quality metrics. In addition, color quality metric is compared to perceptual quality metric and depth based perceptual quality metric is compared to peak signal-to-perceptible-noise ratio. However, it is difficult to select an appropriate method. On the one hand, an overall comparison of all methods is missing. On the other hand, the evaluation approaches used differ regarding display types (stereoscopic, auto-stereoscopic, 2D) and the characteristics of the visual content.

Investigations on the reliability of the 2D methods for video-plus-depth content are twofold. On the one hand, Yasakethu et al. [86] observe that PSNR and SSIM can be used for depth perception prediction, whereas VQM can be used for image quality prediction. On the other hand, Bosc et al. [6, 7] show in two subjective experiments that 2D metrics do not meet the requirements for quality prediction in the field of DIBR. The authors use all objective metrics available in the MeTriX MuX visual quality assessment package [35] and VQM. In a first study, synthesized views observed in monoscopic conditions are considered. Three different multi-view video-plus-depth sequences are selected and 84 synthesized views are generated with seven different DIBR algorithms. The evaluation includes whole video sequences as well as still images, which are key frames out of these sequences. Subjective quality scores are determined from 43 non-expert subjects with ACR-HR and PC. In a second study, synthesized image quality is evaluated in stereoscopic conditions. Similar to the preceding experiment, the same set of 84 synthesized views is used. The stereoscopic images are created from the original view and the synthesized view. The quality is evaluated by 25 non-expert subjects with ACR-HR. The analysis of the subjective tests reveals that both testing methods, ACR-HR and PC, are consistent in monoscopic conditions. Due to the mode of assessment, the PC method requires less subjects than ACR-HR to reach a statistical difference because the impairments among the tested images are small. The authors conclude that it is more difficult to rate the visual quality of the

synthesized images via ACR-HR. In order to check the consistency between the subjective and objective results, PLCC are calculated. In total, 12 metrics are compared in monscopic and stereoscopic conditions. Figure 2.5 depicts the observed differences between PLCC in monoscopic and stereoscopic viewing conditions (see Table 2.4 regarding abbreviations used). In all of the cases except for PSNR, VSNR and NQM, the objective metrics are slightly closer to human perception in stereoscopic conditions. The correlation coefficients are in both cases under 50 percent except for MSSIM in both conditions, NQM in both conditions and WSNR in stereoscopic conditions. The authors conclude that the objective metrics under investigation detect and penalize non-annoying artifacts. Thus, these metrics do not meet the requirements for quality prediction in the field of DIBR. Furthermore, the authors observe that the rankings from monoscopic conditions are completely different than the rankings from stereoscopic conditions. This leads to the assumption that the perception of artifacts works differently in the context of 2D and 3D.

2.5 Summary

This chapter summarized different visual quality assessment techniques of stereoscopic content. A special focus was put on 3D content that contains a novel view. In the beginning, it was pointed out that the parallax of the generated stereoscopic content must be within certain limits, as a violation of these limits can lead to visual discomfort. Furthermore, two types of artifacts were addressed that can degrade stereoscopic content even within these limits. While general stereoscopic artifacts can be neglected in the study design of the planned evaluation, DIBR artifacts are of main interest and must be tackled by the quality assessment techniques. In this context, two basic approaches were addressed, namely subjective studies and objective quality metrics. Six potential subjective evaluation methodologies were discussed, which address different kinds of content characteristics. Regarding objective quality metrics, 2D intended quality metrics can be one choice to evaluate stereoscopic content. For the case of video-plus-depth content the results of two case studies have been discussed, where a poor correlation between these metrics and subjective perception was observed. Promising developments towards a dedicated 3D quality metric have been treated as well. However, none of these approaches have been standardized so far. For the planned study this particularly means that the main emphasis is put on a subjective evaluation because objective quality metrics do not tackle the issues of the addressed research question in a satisfying way.

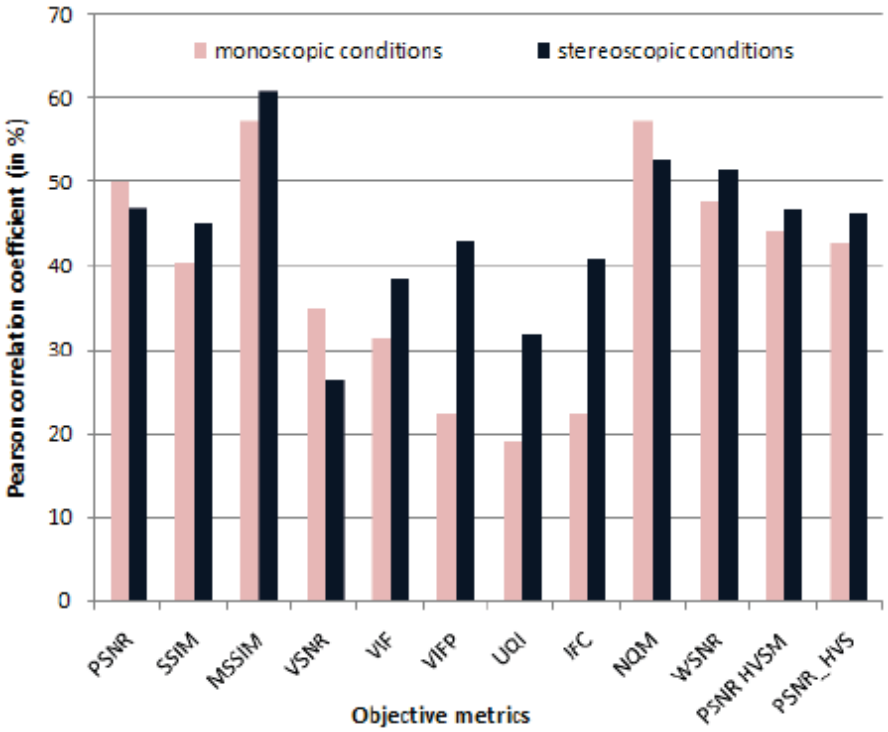


Figure 2.5: Results of comparison study performed by Bosc et al. [7]. PLCC of twelve objective metrics are compared in monoscopic and stereoscopic viewing conditions (Figure taken from [7]).

Abbreviation	Full Name	Reference
PSNR	Peak-Signal-to-Noise-Ratio	[78]
SSIM	Structural SIMilarity index	[79]
MSSIM	Mean Structural SIMilarity index	[80]
VSNR	Visual Signal-to-Noise Ratio	[15]
VIF	Visual Information Fidelity	[66]
VIFP	Visual Information Fidelity Pixel-based	[66]
UQI	Universal Quality Index	[77]
IFC	Information Fidelity Criterion	[67]
NQM	Noise Quality Measure	[67]
WSNR	Weighted Signal-to-Noise Ratio	[17]
PSNR-HVSM	-	[62]
PSNR-HVS	-	[21]

Table 2.4: Overview of metrics used in comparison study performed by Bosc et al. [7].

CHAPTER 3

Depth Map Post-Processing

This chapter gives an overview of state-of-the-art methods for depth map post-processing. The basic ideas of each approach are pointed out and its advantages and disadvantages are discussed. The techniques presented in this chapter are grouped into three classes:

- Gaussian filter techniques (Section 3.1): Broadly speaking, a Gaussian filter denoises data and results in smooth depth maps. We discuss four methods that are based on the principles of Gaussian convolution [20, 46, 74, 89]. These methods differ either in the general design of the Gaussian filter or especially consider the characteristics of depth maps.

- Bilateral filter techniques (Section 3.2): The bilateral filter is related to the Gaussian filter. It smoothes depth maps but in addition preserves edges. We describe four methods that utilize bilateral filtering [19,27,53,68]. One of them transfers the developed concepts from the Gaussian techniques to the bilateral filter, the remaining three investigate adaptions of the general bilateral filter or are closely related to these adaptions.

- Other techniques (Section 3.3): In addition to Gaussian and bilateral filter techniques, the use of other approaches is proposed in literature. We highlight three of those approaches [32, 48, 63].

3.1 Gaussian Filter Techniques

This section exemplarily discusses depth map post-processing methods that are based on Gaussian filtering. These methods differ either in the general design of the Gaussian filter or especially consider the characteristics of depth maps. The description of these methods is structured hierarchically, i.e., the later methods are based on the preceding ones.

Generally, a Gaussian filter computes the weighted sum of the pixels in a window centered at a pixel. This window can be symmetric (i.e., horizontal diameter d_h and vertical diameter d_v

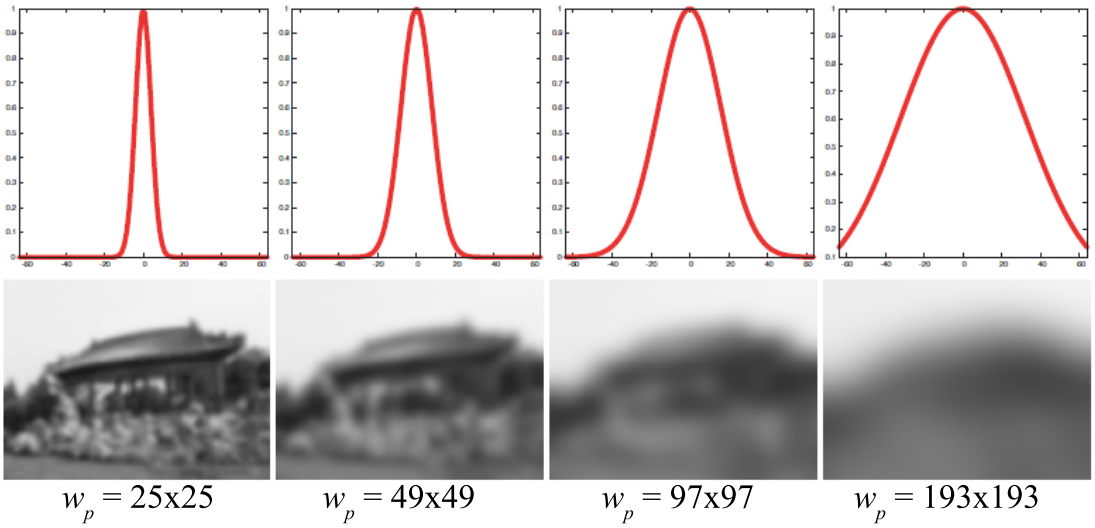


Figure 3.1: Comparison of Gaussian filter with different settings for window w_p. Top row shows the profile of a 1D Gaussian kernel and bottom row shows the obtained results after applying the corresponding 2D Gaussian convolution. Note that with increasing w_p edges are lost because Gaussian filtering is performed over a larger area (Figure taken from [57]).

of the window are identical) or asymmetric (i.e., horizontal diameter d_h and vertical diameter d_v of the window can differ). In particular, the Gaussian convolution is defined as follows:

$$GC[I]_p = \sum_{q \in w_p} G_\sigma(\| \text{ p} - \text{q } \|)\, I_q. \tag{3.1}$$

$GC[I]$ is the output of the Gaussian convolution GC applied on the image I. p and q are two-dimensional pixel coordinates. I_q is the intensity value of pixel q. $w_p = d_h \times d_v$ and is the window centered at pixel p. The two-dimensional Gaussian kernel $G_\sigma(x)$ is given by:

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} \exp\,(-\frac{x^2}{2\sigma^2}). \tag{3.2}$$

The weight for pixel q is defined as $G_\sigma(\| \text{ p} - \text{q } \|)$ and decreases with the spatial distance to center pixel p, where σ defines the standard deviation of the Gaussian convolution. Note that the spatial extent of the filter window w_p can be further restricted by σ. It is worth noting that the Gaussian filter only considers the spatial distance of the pixels to compute the weights in a filter kernel. Figure 3.1 gives an example for this case and shows an image that is processed with different parameters.

Gaussian Symmetric Filter

Tam et al. [74] investigate the impact of depth map post-processing with Gaussian filters with symmetric windows on the perceived image and depth quality. The objective of the Gaussian

Smoothing	Symmetric	Asymmetric
None	H = $\{0, 0 \times 0\}$ V = $\{0, 0 \times 0\}$	H = $\{0, 0 \times 0\}$ V = $\{0, 0 \times 0\}$
Mild	H = $\{4, 13 \times 13\}$ V = $\{4, 13 \times 13\}$	H = $\{4, 13 \times 13\}$ V = $\{12, 41 \times 41\}$
Strong	H = $\{20, 61 \times 61\}$ V = $\{20, 61 \times 61\}$	H = $\{20, 61 \times 61\}$ V = $\{60, 193 \times 193\}$

Table 3.1: Parameters used for comparison study between Gaussian symmetric and Gaussian asymmetric filtering performed by Zhang et al. [89]. Settings on standard deviation σ and window size w_p are expressed as $\{\sigma$,$w_p\}$. H corresponds to horizontal direction, V corresponds to vertical direction (Table from [89]).

filtering is to tackle inaccuracies in the depth maps and to reduce the size of newly exposed areas in the novel views. Thus, the overall quality of the novel views shall be improved.

In particular, five different settings on standard deviation σ and window size w_p are considered[1]: $\{\sigma, w_p\} = [\{000, 00{\times}00\}, \{010, 10{\times}10\}, \{020, 20{\times}20\}, \{060, 30{\times}30\}, \{120, 80 \times 80\}]$. Larger values indicate a greater level and extent of smoothness. The depth maps are post-processed with these settings and used to create novel views. Stereoscopic images are formed from either the original left view and the novel right view or from novel images for both views. Twenty-three subjects rate the quality of the resulting stereoscopic images with DSCQS methodology.

In the context of the user study performed by Tam et al. [74], two important findings for depth based 3D systems have been made. First, a higher amount of smoothing often leads to a better perceived image as well as depth quality. Second, smoothing may reduce the quality of depth maps but improves the subjective quality of novel views. This occurs because on the one hand noise in the depth maps is reduced and on the other hand smooth transitions between object boundaries of foreground and background objects in the depth maps are produced. However, Gaussian symmetric filtering can cause geometric distortions in the novel views [89].

Gaussian Asymmetric Filter

The authors of [89] propose to use Gaussian filters with asymmetric windows for depth map post-processing to address geometric distortions in novel views that are a result of over-smoothing of horizontal depth borders. To avoid these distortions, the depth maps are smoothed stronger in vertical than in horizontal direction, which leads to better visual results in the novel views (see Figure 3.2).

The result of Gaussian asymmetric filtering is evaluated in a subjective study and compared to the quality scores of Gaussian symmetric filtering. Ten subjects rate the image quality of stereoscopic images using DSCQS methodology. The stereoscopic images contain one original view and one novel view. The depth map of the novel view is post-processed either with

[1] Tam et al. [74] do not explicitly mention whether their window size w_p is a radius or a diameter.

(a) Gaussian symmetric filter

(b) Gaussian asymmetric filter

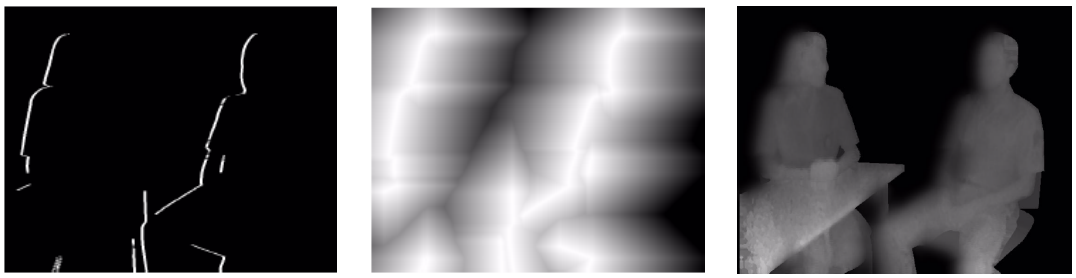
Figure 3.2: Comparison between (a) Gaussian symmetric and (b) Gaussian asymmetric filtering. The left image shows the processed depth map, the center image shows the obtained novel view and the right image shows an enlarged segment. Note that geometric distortions (pink arrow) caused by Gaussian symmetric filtering are reduced by Gaussian asymmetric filtering. A higher intensity value in the depth map means that the object is closer to the camera (Figures taken from [89]).

symmetric or asymmetric Gaussian filtering. Three levels of smoothing (none, mild, strong) are applied on the depth maps (see Table 3.1). For asymmetric smoothing, the level of vertical smoothing is three times larger than in horizontal direction.

The resulting mean ratings for asymmetric smoothing are higher in all three scenarios meaning that less smoothing in horizontal direction in the depth maps leads to subjectively better results in the novel views. In particular, asymmetric smoothing achieves mean ratings of 48.6 (symmetric smoothing: 44.8), 58.0 (symmetric smoothing: 52.9) and 68.4 (symmetric smoothing: 62.1) for the three scenarios none, mild and strong smoothing, respectively. However, Gaussian asymmetric filtering strongly degrades the depth maps, which can affect the depth perception [19].

Gaussian Distance Dependent Filter

In order to take advantage of Gaussian filtering on the perceived image quality and to counteract the issue of depth map degradation, Daribo et al. [20] propose a weighted Gaussian filter based



(a) binary map (b) distance map (c) depth map

Figure 3.3: Illustration of the processing steps from the Gaussian distance dependent depth filter proposed by Daribo et al. [20]. (a) Binary map of depth map edges (white: edge pixel, black: no edge pixel). (b) Distance map with distance information to the previously detected depth map edges. The intensity value in the distance map indicates the shortest distance from a pixel to the detected depth map edges, where brightness decreases with the distance. (c) Smoothed depth map. A higher intensity value in the depth map means that the object is closer to the camera (Figures taken from [18]).

on the distance to the depth edges. The authors mainly address the reduction of exposed areas (i.e., disocclusions) in novel views and apply stronger smoothing in the respective areas of the depth maps. These exposed areas are located at object edges and correspond to sharp depth discontinuities in the depth maps. In addition, the direction of the warping determines the location of the exposures (i.e., novel right views have exposures on the right side of object edges, novel left views on the left side).

The approach consists of three steps (see Figure 3.3). First, the depth map is used to create a binary map which reveals where the displacement between adjacent pixels is larger than a predefined threshold. These displacements correspond to edges in the depth map that cause exposures in the novel views. Second, the city-block distance is used to compute a distance map that stores the shortest distance of each pixel to the detected edges in the depth map. For example, a zero value in the distance map indicates that a pixel belongs to the edge. Finally, the distance information is used to weight the output of the Gaussian filter. The closer a pixel is located to an edge, the stronger it is affected by the Gaussian filter. In contrast, a further distance results in less Gaussian filtering. This results in stronger smoothing near an edge and lower smoothing further away from an edge. It is worth noting that the method by Daribo et al. [20] can be applied to asymmetric windows as well in order to reduce the amount of geometric distortions in the novel views.

The evaluation is performed on one stereoscopic video sequence. The resulting depth map and novel view quality are compared to Gaussian symmetric smoothing with PSNR. The parameters of the Gaussian kernel are $\sigma = 20$ and $w_p = 61 \times 61$. Concerning depth map quality, Gaussian distance dependent filtering obtains PSNR values between 26 and 28. Gaussian symmetric smoothing achieves PSNR scores between 20 and 22. Concerning synthesized view quality, Gaussian distance dependent filtering achieves PSNR scores in the range of 29 to 31.

Gaussian symmetric smoothing performs weaker and obtains PSNR scores between 24 and 26. Although Gaussian distance dependent filtering can reduce the depth map degradation by limiting the filtered areas, the particular depth is not considered in the filtering step itself [46].

Gaussian Depth Discontinuity Filter

Lee et al. [46] extend the Gaussian distance dependent filtering [20] and propose a method where the weighting of the Gaussian filter considers not only the distance information to the depth edges but also the strength of depth discontinuities. These depth discontinuities correspond to edges in the depth map, whereas the analysis of adjacent pixels that form them provides information on their depth relation. The authors notice that the length of exposed areas (i.e., disocclusions) in novel views can be directly predicted by the strength of depth discontinuities in the depth maps. Therefore, the additional analysis of the depth discontinuities during smoothing results in a higher novel view quality and less degraded depth maps (see Figure 3.4).

In particular, the approach is similar to [20] and consists of three steps. First, edges in the depth map are detected that cause exposures in the novel view. Second, the depth discontinuity of these edges is calculated. Third, the weighting of the Gaussian filter for each pixel is determined by the strength of depth discontinuity and the city-block distance of this pixel to the beforehand detected depth edges. According to the obtained weighting, the original depth map and the filtered depth map are blended.

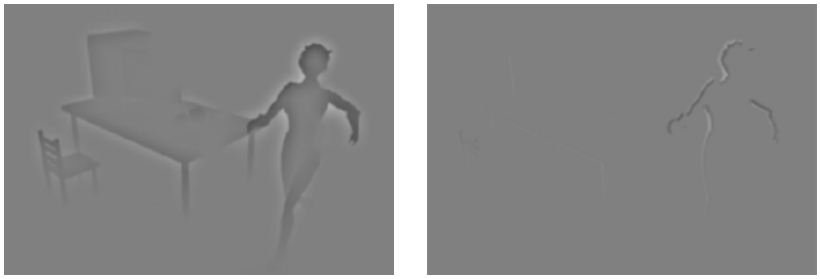
The evaluation is performed with PSNR on one stereoscopic video sequence concerning depth map quality and novel view quality. The parameters used for evaluation are not listed. The proposed method by Lee et al. [46] outperforms Gaussian asymmetric filtering and Gaussian distance dependent filtering regarding these two quality parameters. Gaussian depth discontinuity filtering achieves a PSNR between 31 and 32 concerning depth map quality and a PSNR between 28 and 28.4 concerning novel view quality. Gaussian asymmetric filtering obtains PSNR scores which vary between 22 and 22.1 for depth map quality and between 27 and 27.4 for novel view quality. Gaussian distance dependent filtering achieves a mean PSNR score of approximately 31 for depth map quality and PSNR scores in the range of 27.7 to 28 for novel view quality.

3.2 Bilateral Filter Techniques

This section addresses bilateral filter based methods for depth map post-processing. The first method is based on the concepts from the Gaussian based filter techniques and transfers them to bilateral filtering. The following two methods investigate the application of the bilateral filter derivatives. The last method performs local statistics rather than local averaging.

The bilateral filter [75] is an edge-preserving filter. Additionally to the spatial distance of the pixels it also considers the difference in intensity values (see Figure 3.5). As a result, dissimilar pixels that e.g., occur at edges of the processed image are maintained. In particular, the bilateral filter is defined as follows:

(a) (b)



(c) (d)

Figure 3.4: Illustration of depth map degradation by Gaussian asymmetric filtering, Gaussian distance dependent filtering and Gaussian depth discontinuity filtering. (a) Original depth map, where a higher intensity value in the depth map means that the object is closer to the camera. (b)-(d) Difference images between original depth map and depth maps that were post-processed with (b) Gaussian asymmetric filter, (c) Gaussian distance dependent filter and (d) Gaussian depth discontinuity filter. The difference is encoded in gray values where a mid gray value corresponds to no differences. Note that the depth map processed by Gaussian depth discontinuity filtering results in the smallest difference compared to the original depth map (Figures taken from [46]).

$$
\begin{aligned}
BF[I]_p &= \frac{1}{W_p} \sum_{q \epsilon w_p} G_{\sigma_s}(\| \mathrm{p} - \mathrm{q} \|)\, G_{\sigma_r}(|I_p - I_q|)\, I_q, \\
W_p &= \sum_{q \epsilon w_p} G_{\sigma_s}(\| \mathrm{p} - \mathrm{q} \|)\, G_{\sigma_r}\, (|I_p - I_q|).
\end{aligned}
\tag{3.3}
$$

$BF[I]$ denotes the output of bilater filter BF applied on the image I. p and q are two-dimensional pixel coordinates. I_p and I_q are the intensity values of pixels p and q. w_p is a window centered at pixel p. W_p is a normalization term and ensures pixel weights in the window sum to 1.0. $G_{\sigma_s}(\| \mathrm{p} - \mathrm{q} \|)$ is a spatial Gaussian filter which decreases the influence of distant pixels, $G_{\sigma_r}(|I_p - I_q|)$ is a range Gaussian filter which decreases the influence of dissimilar pixels (in color or intensity). Parameters σ_s and σ_r are the standard deviation of the spatial filter G_{σ_s} and the range filter G_{σ_r}.

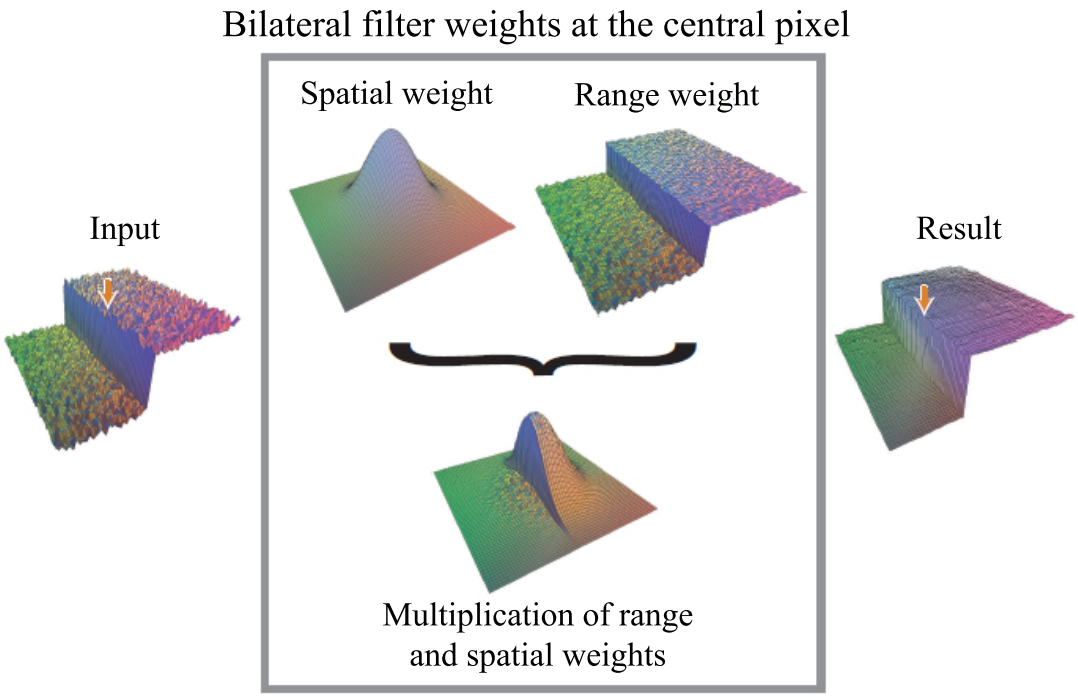


Figure 3.5: Illustration of weight calculation for a pixel located at an edge (under the arrow). Spatial and intensity closeness are combined. This ensures that only nearby similar pixels are considered in the smoothing process (Figure taken from [57]).

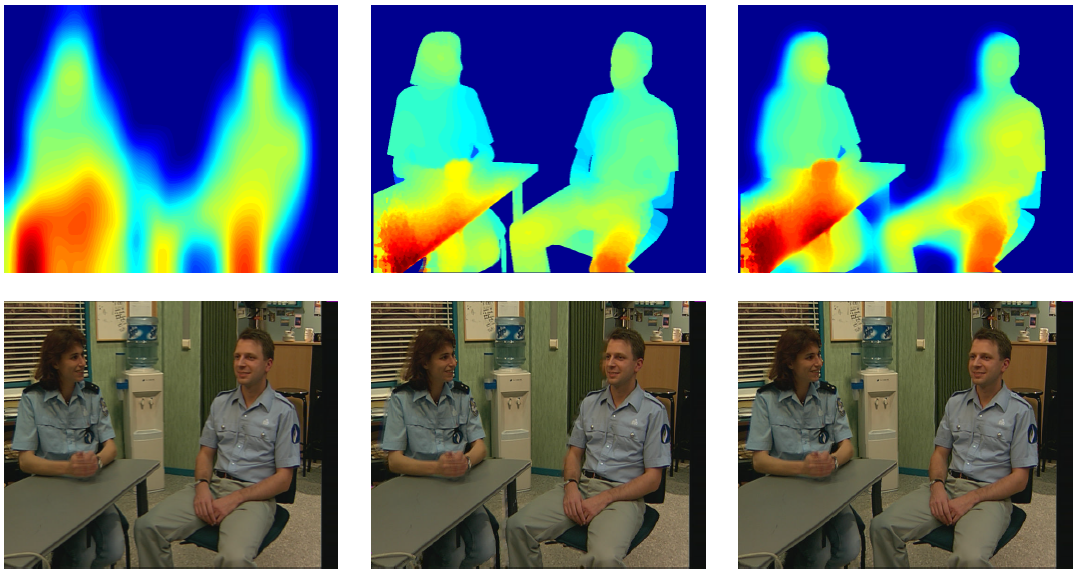
Two important variations of the bilateral filter have been proposed in literature. These modifications can also be applied in depth map post-processing, as addressed in the scope of this section:

- **Joint bilateral filter** [59]: The bilateral filter can be generalized by performing the weight computations in a guidance image that differs from the input image. Thus, details from the guidance image can be extracted and transferred to the input image. In the context of depth map post-processing, misalignments of depth and color edges can be reduced.

- **Adaptive bilateral filter** [88]: This version of the bilateral filter determines the range filter parameters automatically. In particular, the center of the range filter and the smoothing parameter at each pixel position adapt to the image content. These modifications make the adaptive bilateral filter capable of sharpness enhancement and noise removal.

Bilateral Depth Discontinuity Filter

Daribo et al. [19] combine the bilateral filter and the discontinuity analysis from the Gaussian depth discontinuity filter [46]. Traditionally, the bilateral filter smoothes images but preserves



(a) (b) (c)

Figure 3.6: Comparison of (a) Gaussian asymmetric filter, (b) bilateral filter, and (c) bilateral depth discontinuity filter for depth map post-processing. Top row shows the post-processed depth maps, bottom row shows the obtained novel views. Warm colors in the depth map indicate that the object is closer to the camera (Figures taken from [19]).

edges. However, the quality of novel views can be increased, if depth edges that result in dis-occlusions in novel views are smoothed (see Section 3.1). Therefore, this approach applies bilateral filtering according to the strength of depth discontinuities and takes advantage of both approaches. The smoothing of the required depth discontinuities improves the visual result of these areas in the novel views, while the other depth discontinuities remain preserved. In addition, homogeneous areas are being de-noised. Thus, the described approach is a trade-off between depth map degradation and novel view quality.

The result of bilateral depth discontinuity filtering is compared against Gaussian asymmetric filtering and traditional bilateral filtering. Figure 3.6 shows qualitative results of these three approaches. For a quantitative evaluation, the depth maps of one video-plus-depth sequence are post-processed with these three filters and the PSNR values of the depth maps and the novel views are compared. The parameters σ_s and σ_r are set to 20 and 11, respectively. The window size w_p is not explicitly given, but determined from σ_s. Concerning depth map quality, the bilateral filter achieves the highest PSNR values with an average of approximately 43. The PSNR of the proposed filter by Daribo et al. [19] varies between 31 and 34, whereas the PSNR of the Gaussian asymmetric filter is less than 20. On the contrary, the Gaussian asymmetric filter achieves the highest PSNR values concerning novel view quality. The PSNR varies between

(a)

(b)

(c)

(d)

(e)

Figure 3.7: Comparison between joint bilateral filter and bilateral scaling based filter on the quality of the post-processed depth maps. (a) Guidance image. (b) Ground truth depth map. (c) Input depth map. (d) Joint bilateral filter post-processed depth map. (e) Bilateral scaling based filter post-processed depth map. Note that the visibility of the guidance image object's texture is higher in (d) than in (e). This visibility can even be reduced with a smaller size for the kernel window. A higher intensity value in the depth map means that the object is closer to the camera (Figure (a) taken from [65], Figures (b)-(e) taken from [27]).

23 and 23.5 for the Gaussian asymmetric filter, between 22.7 and 23.4 for the bilateral depth discontinuity filter and between 22.1 and 22.7 for the bilateral filter.

Bilateral Scaling Based Filter

Gangwal et al. [27] use the joint bilateral filter for depth map post-processing. Their approach consists of three steps. First, the initial depth map and the guidance image are down-scaled using a 2D box filter. This step reduces the visibility of the guidance image object's texture within the post-processed depth maps and removes local outliers. Second, the joint bilateral filter is applied to the down-scaled depth map in order to align depth and color edges. Last, the filtered depth map is up-scaled using a multi-step implementation of joint bilateral up-sampling [64] to the full image resolution.

The described approach is compared to traditional joint bilateral filtering which does not consider the additional down- and up-scaling. The evaluation is performed with PSNR. Parameters σ_s and σ_r are set to 0.5 and 0.1, respectively. Five different window sizes w_p are considered: $w_p = \{25 \times 25, 41 \times 41, 73 \times 73, 105 \times 105, 137 \times 137\}$. The proposed method by Gangwal et al. [27] achieves higher PSNR scores for all settings. The increase is up to 1.8 dB depending on the size of the filter window. Moreover, the computation cost can be reduced by a factor of 16 to 64.

Bilateral Compression Recovery Filter

De Silva et al. [68] address the related topic of compressed depth map recovery. The authors propose an adaptive bilateral filtering technique to reduce artifacts in novel views caused by depth map compression. In contrast to [88], in [68] the range filter parameters are determined by a local histogram analysis. To this end, the depth map is segmented into 64×64 pixel blocks and for each block a histogram is obtained. The histogram is smoothed using a one-dimensional average filter. Afterwards, dominant peaks and their enclosing minima of the smoothed histogram are identified. The peaks and their enclosing minima are used to aid the adaptive bilateral filtering. For each pixel, the offset of the range filter is given by the distance to the nearest peak and the size of the range filter is based on the distance to the enclosing minima.

For evaluation, depth maps of two video sequences are compressed with the H.264/AVC reference encoder at five different quantization parameter settings. The compressed depth maps are post-processed with the approach proposed by De Silva et al. [68] and compared to their unprocessed counterparts regarding PSNR. The parameters are set to $\sigma_s = 12$ and $w_p = 7 \times 7$, wherein σ_r is obtained through the described histogram analysis. The authors claim that the post-processed depth maps increase the quality up to 1.9 dB depending on the quantization parameter used for compression.

Weighted Mode Filter

Min et al. [53] propose a weighted mode filter for the enhancement of depth videos. It is related to the joint bilateral filter because it uses a guidance image and considers identical color and a spatial similarity measure. However, it differs concerning two points. First, an additional parameter is introduced that models errors that may exist in the depth data. Second, the joint bilateral filter provides a mean value through adaptive summation in a window, whereas the weighted mode filter selects the largest value within a window. This step significantly reduces the smoothing effect in depth maps that can cause object deformations in novel views.

3.3 Other Techniques

This section discusses three methods that rely on other techniques than Gaussian or bilateral filtering. The first method is a general filter concept which is ubiquitous in computational photography applications and thus is also of interest for depth map post-processing. The last two methods have been designed specifically for the quality improvement of novel views, whereas

one of them is based on color image and depth map registration and the other one is based on a two-step rendering approach.

Guided Image Filter

The guided image filter [32] is an edge preserving filter. Similar to the joint bilateral filter, depth values are locally averaged based on a color and spatial similarity and thus misalignments of depth and color edges can be reduced. However, the computation time of the guided image filter is independent of the window size. In particular, the guided image filter is defined as follows:

$$\begin{aligned} GF[I]_p &= \sum_{q \in w_k} W_{GF_{pq}}(G) I_q, \\ W_{GF_{pq}} &= \frac{1}{|w|^2} \sum_{k:(p,q) \in w_k} \left(1 + \frac{(G_p - \mu_k)(G_q - \mu_k)}{\sigma_k^2 + \epsilon}\right). \end{aligned} \tag{3.4}$$

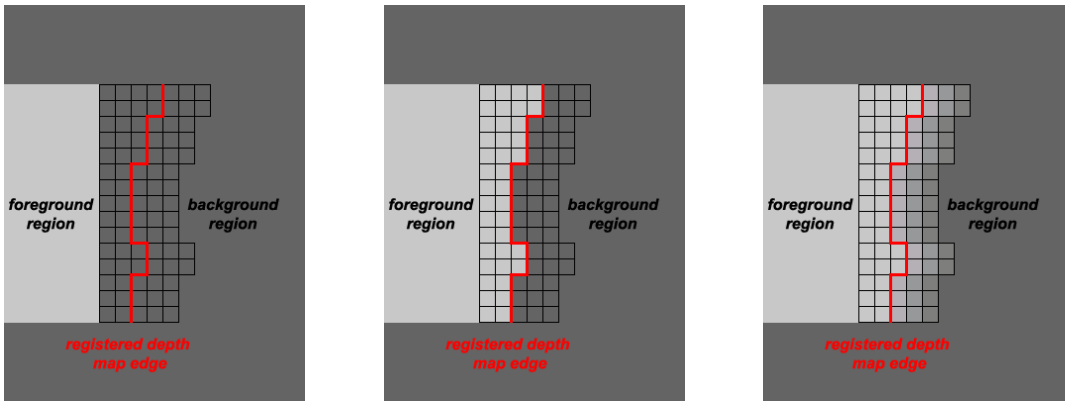
$GF[I]$ is the output of the guided image filter GF applied on the input image I. G is the guidance image (e.g., color image of a scene). I_p and I_q are the intensity values at pixel coordinates p and q of the input image I (e.g., depth map of a scene). G_p and G_q are the intensity values at pixels p and q of the guidance image G. w_k is the window centered at pixel k and $|w|$ represents the number of pixels in this window. W_{GF} is the kernel weights function. μ_k and σ_k^2 denote the mean and variance of the guidance image G in the local window w_k.

A large weight is assigned to pixel q if G_p and G_q are located on the same side of an edge. On the other hand, pixel q will have a small weight if G_p and G_q are on different sides of an edge. The amount of smoothing is controlled by the parameter ϵ. If the guidance image is identical to the input image, similar smoothing results compared to the bilateral filter can be obtained, by setting $\epsilon = \sigma_r^2$ for intensity values in the range of $[0, 1]$. It is worth noting that cross-based local multipoint filtering [47] and adaptive guided image filtering [60] are further developments of the guided image filter and can reduce the undesired smoothing effect of the GF at edges. However, both approaches lead to an increase in complexity and runtime.

Foreground Protecting Filter

The authors of [48, 84] address the problem of edge misalignments between depth and video data. These inaccuracies can lead to artifacts in novel views because wrong color information is used for the handling of exposures. Furthermore, the authors observe that smoothing based approaches for depth map post-processing result in edge distortions of the foreground objects in the novel views. Thus, the authors propose to process only depth values of background pixels and leave the foreground pixels unchanged.

In the first step, the edges in the color image and the corresponding depth map are detected. Concerning the former, Xu et al. [84] use a horizontal Prewitt operator, Lu et al. [48] use a horizontal Laplacian operator instead. Concerning the latter, neighbouring pixels with a depth discontinuity larger than a predefined threshold are detected. Next, the edges in the color image and the depth map are aligned. Therefore, the foreground depth values are shifted towards the corresponding color edge. Note that this step only changes depth values of background pixels.



(a) (b) (c)

Figure 3.8: Comparison of foreground protecting filter approaches proposed by Xu et al. [84] and Lu et al. [48]. (a) Depth map before post-processing. (b) Depth map after post-processing according to Xu et al. [84]. (c) Depth map after post-processing according to Lu et al. [48] (Figure inspired by [48]).

In order to produce smooth edge transitions, Lu et al. [48] perform an additional piecewise smoothing after depth map registration. [84] and [48] are similar and differ only in the edge detector used and in the additional smoothing step in [48] (see Figure 3.8). The foreground objects in depth maps get aligned with foreground objects in the color image. As a result, the foreground objects remain preserved in the novel views.

Xu et al. [84] carry out the evaluation on two video sequences and compare their method with no post-processing and two different post-processing methods which include Gaussian asymmetric filter. The parameters used for the comparative analysis are not specified. Five different inpainting techniques are also used in the evaluation. Their method achieves the highest PSNR scores in all scenarios. Compared with Gaussian asymmetric filter, the average PSNR improvements are 0.31 dB and 1.75 dB for the two video sequences under investigation. Lu et al. [48] accomplish their evaluation on three images and compare their method also with no post-processing and two different post-processing methods, which also include Gaussian asymmetric filter. The parameters used for evaluation are not listed. Compared to asymmetric Gaussian filter, the average PSNR increase is 5.65 dB for the three images.

Two-Step Rendering Approach

Riechert et al. [63] propose a two-step rendering approach consisting of disparity forwards mapping and image backwards mapping. Their approach does not focus explicitly on disparity map post-processing. However, the disparity maps are used to enhance the quality of novel views. To this end, for every novel view the corresponding disparity map is obtained. Thus, advanced interpolation filters can be used in the novel views because pixels in the novel view can be projected to the original view (see Figure 3.9).

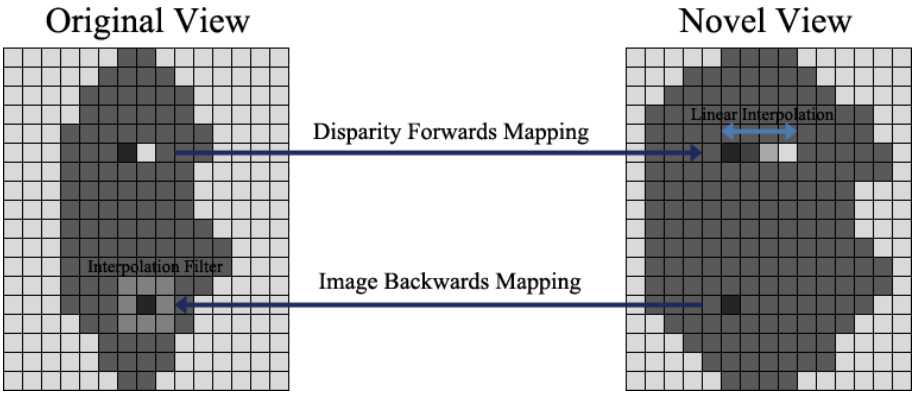


Figure 3.9: Illustration of the two-step rendering approach of Riechert et al. [63]. Disparity forwards mapping is used to create a disparity map for a novel view. This disparity map enables an image backward mapping of the novel view's pixels to the original view (Figure reproduced from [63]).

The disparity forwards mapping is defined as follows: Two adjacent disparity pixels are warped forward. As the target positions of those pixels must not be located next to each other, all disparity values in between are interpolated linearly. However, the linear interpolation does not take place in disoccluded areas which are detected in advance. In the image backwards mapping, pixels of the novel view can be projected to the source view according to the beforehand obtained disparity map. Therefore, any kind of interpolation filters can be used. The authors show qualitative results for three images where the performance of three different interpolation filters are compared. However, the authors give no quantitative evaluation of their approach.

3.4 Summary

In this chapter, methods for depth map post-processing have been reviewed. First, Gaussian based techniques were described. For the simple case where the whole image is processed with a Gaussian filter, a stronger smoothing in vertical than in horizontal direction results in a better novel view quality because the incidence of geometric distortions in the novel views is reduced. However, processing the whole depth map with Gaussian filters can strongly degrade the quality of depth maps which can have a negative impact on the depth perception. Therefore, two approaches were addressed where the Gaussian filtering of depth maps is weighted according to a distance information to the edges in those depth maps. One of these two methods additionally considers the strength of depth discontinuities to further improve the novel view quality.

Then, approaches based on bilateral filtering were discussed. The bilateral filter locally averages similar depths but preserves edges. The amount of smoothing is controlled by a depth similarity and a spatial similarity. In order to improve the alignment of depth edges with color edges in the corresponding view, the bilateral filter weights can be computed according to color similarities in the corresponding view. Instead of locally averaging depths, an approach based

on local statistics that considers color similarity, spatial similarity and similarity of depths was addressed. This approach significantly reduces the smoothing effect in depth maps that can cause object deformations in novel views.

Last, three approaches were addressed where depth map post-processing does not rely on Gaussian or bilateral filtering. Guided image filtering is related to bilateral filtering because it locally averages depths with similar colors in the corresponding view. This approach can be implemented very efficiently because its runtime is independent of the window size. Foreground protecting filtering aligns those depth edges with color edges that result in disocclusions in novel views by performing edge detection in the depth map and the corresponding view. The two-step rendering approach uses depth maps to enhance the quality of novel views but performs no explicit depth map post-processing.

CHAPTER 4

Experimental Set-up and Evaluation

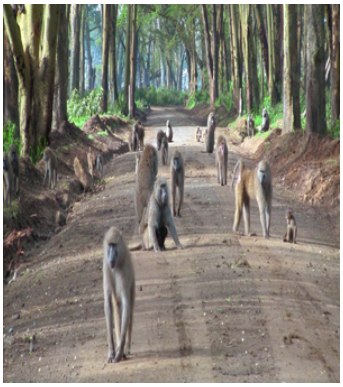
This chapter describes the experimental set-up and evaluation applied in this thesis. The aim of our evaluation is to investigate the effects of different depth map post-processing methods on the quality of stereoscopic images that contain a novel view. We perform a preliminary and a main study. The results of the preliminary study are used to optimally define the main study. Both studies are divided into a subjective and an objective evaluation in order to assess the correlation between subjective and objective scores.

This chapter is structured as follows. Section 4.1 describes the generation of the stereoscopic images that are used in our studies. Section 4.2 addresses the subjective test methodology, Section 4.3 the objective one. Section 4.4 summarizes the evaluation process.

4.1 Dataset

In the subjective and objective evaluation, six stereoscopic image pairs are used (see Figure 4.1). For all of the image pairs the original left and original right views are available. These image pairs contain challenging scenes for disparity map generation (e.g., fuzzy object borders in Figure 4.1c or thin vertical structures in Figure 4.1d) that cause errors in the corresponding disparity maps (see Figure 4.2). The original images have different resolutions. Since our user study is conducted on a monitor with a native resolution of 1680×1050 pixels, all images are down-sampled to match the resolution of 1680×1050 either in width or in height (see Table 4.1).

From the down-sampled left and right image pairs, the corresponding disparity maps are created using Stereoscopic Suite X3 (SSX3) [23]. Different post-processing filters are applied on the generated disparity maps. Both the unprocessed and the post-processed disparity maps are used to generate novel views. Occurring disoccluded areas in the novel views are filled using the built-in inpainting algorithm of SSX3. The original left views and the generated novel right views form the 42 stereoscopic images that are used in our evaluations.



(a) monkeys000

(b) tiger000

(c) gforce033

(d) gforce105

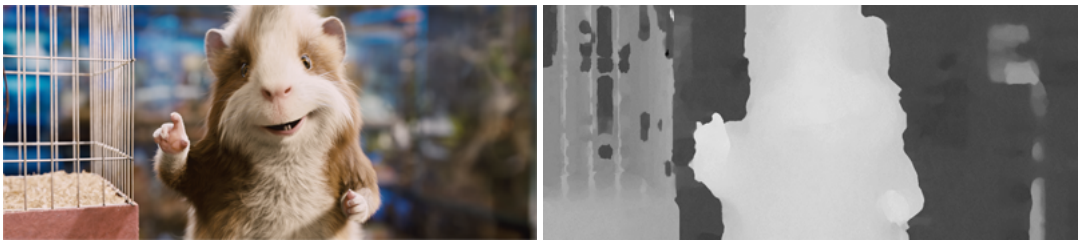
(e) musketiers264

(f) musketiers333

Figure 4.1: Original left views from the stereoscopic images that are used in this study.

		Disparity Statistics			
Title	**Resolution**	**Min**	**Max**	**Mean**	**Std**
gforce033	1680×743	-34	-2	-16.23	1.56
gforce105	1680×739	-28	9	-8.54	2.93
monkeys000	933×1050	-18	12	-4.19	1.27
musketiers264	1680×749	-55	-4	-14.16	4.47
musketiers333	1680×749	-13	15	-2.89	1.92
tiger000	1680×945	-150	19	-63.03	23.04

Table 4.1: This table lists the title, resolution and disparity statistics in pixels of the stereoscopic images that are used in this study.



(a) (b)

Figure 4.2: Example of (a) one original left view of the dataset and (b) the corresponding disparity map. For visualization, the disparity map is scaled to the intensity range of [0,255]. Note the mismatches of the disparity map in the area of the cage (i.e., thin vertical structures).

4.2 Subjective Quality Assessment

The Pair Comparison (PC) methodology was used for the subjective quality assessment in this study [10]. In PC, a pair of stimuli (i.e., stereoscopic images) is displayed to the subjects and the quality of the stimuli is assigned in terms of preferences. When comparing the stimuli that are based on post-processed depth maps on a 3D display, the differences were often subtle. In this context, PC enables the comparison of stimuli which differ only slightly.

4.2.1 Environment

To perform the subjective study, we set up a lab (see Figure 4.3). In particular, a table was placed in front of a wall. Both table and wall were covered with a black cardboard in order to prevent the subjects from being distracted by the environment. We additionally covered all windows so that no daylight could enter the room. The light in the room was turned off during the test. The stimuli were displayed on a 22 inch stereoscopic display (i.e., Samsung Syncmaster 2233RZ) with a native resolution of 1680×1050 pixels, with NVIDIA 3D vision controller. The subjects were seated approximatively two meters away from the display and in line with the center of the display.

4.2.2 Experiment Design

The assessment task was designed as described in the following. A pair of stimuli that captured the same scene was displayed successively to the subjects. We generated a comparison set for each scene that consisted of different stimuli according to the investigated post-processing methods. Each subject was asked to rate whether stimulus ‘A is better’, stimulus ‘B is better’, or stimuli A and B were the ‘same’. Therefore, the subjects could freely switch between the two stimuli that formed a pair by using the arrow keys of a keyboard (see Figure 4.4). Only one subject per session was performing the assessment task. In order to prevent a bias in the studies, the image pairs were presented in random order.

(a) (b)

Figure 4.3: Test environment and 22 inch stereoscopic display (i.e., Samsung Syncmaster 2233RZ) used in this study.

Before each test session, the subjects were given written instructions and a brief explanation of the experiment design (see Appendix A). Afterwards, all subjects were screened for visual acuity, color vision and stereo vision according to [10] (see Appendix B). Before the actual test session, a trial run was performed, in which the test methodology was introduced to the subjects by using three example pairs out of the test stimuli. These example pairs had been selected by an expert viewer and matched the three options 'A is better', 'B is better' and 'same'. In the middle of the test, each subject was given the opportunity to take a short break. After the quality assessment task was finished, the subjects were asked to fill in a questionnaire about their impressions on the perceived quality and the test methodology itself (see Appendix C).

4.2.3 Subjective Data Processing

We analyse the study results by applying an outlier detection algorithm [45] and computing quality scores. The outlier detection algorithm detects subjects whose preferences are contradictory. In particular, the algorithm quantifies the number of circular triads among three stimuli i, j and k. For example, a circular triad occurs when stimulus i is preferred over stimulus j, stimulus j is preferred over stimulus k, but stimulus k is preferred over stimulus i (see Figure 4.5). For pair comparison data that involves ties such as ours, a circular triad is formed in the following four cases [45]:

$$\begin{aligned} &i > j \cap j > k \cap k > i, \\ &i > j \cap j > k \cap k = i, \\ &i > j \cap j = k \cap k > i, \\ &i = j \cap j > k \cap k > i, \end{aligned} \tag{4.1}$$

where $i > j$ means that stimulus i is preferred over stimulus j and $i = j$ means a tie between stimuli i and j. When the ratio of non-circular triads compared to all possible circular triads

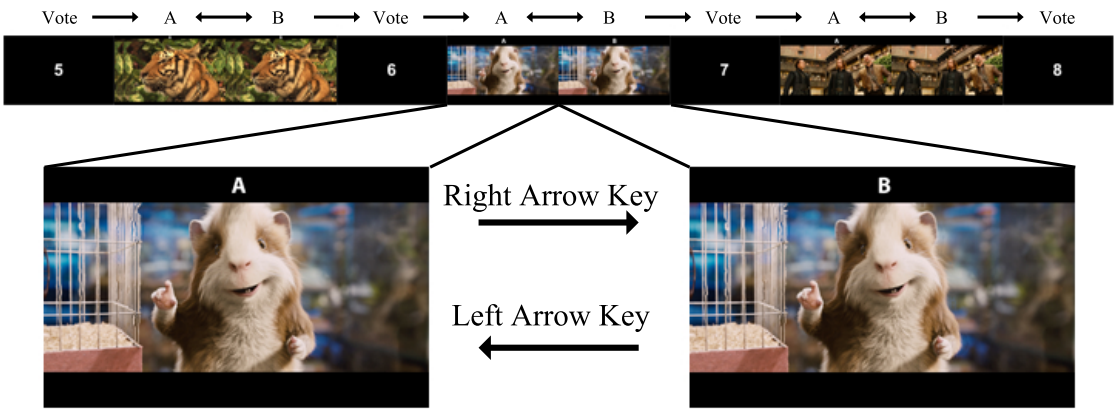


Figure 4.4: General scheme of the quality assessment task. Each image pair was initiated by its ID, where also the preference vote for the preceding image pair had to be given. The arrows indicate the navigation through the content by using the arrow keys, i.e., left arrow means only the left arrow key is allowed, right arrow means only the right arrow key is allowed, and left-right arrow means that the left and the right arrow keys are allowed.

(a) (b)

Figure 4.5: Illustration of a circular triad. (a) No circular triad is present: stimulus i is preferred over stimulus j, stimulus j is preferred over stimulus k, and stimulus i is preferred over stimulus k. (b) One circular triad is present: stimulus i is preferred over stimulus j, stimulus j is preferred over stimulus k, but stimulus k is preferred over stimulus i (Figure inspired by [33]).

(herein referred to as *transitivity satisfaction rate*) is relatively low, the subject can be considered as an outlier and its ratings are discarded from further analysis.

Next, we compute quality scores for each comparison set individually and for all comparison sets together. These quality scores measure the subjective quality of the stimuli according to the preferences of the subjects. Following, we give an example of the quality score computation for a single comparison set. A comparison set consists of n stimuli, T_1, ..., T_n. This results in $\binom{n}{2}$ stimuli pairs. The number of comparisons for a pair (T_i, T_j) is given by n_{ij}. The results of each comparison set are summarized by a matrix of choice frequencies $\{c_{ij}\}$. Table 4.2 shows an example of a matrix for four stimuli. Each entry c_{ij} consists of the number of preferences w_{ij},

	T_1	T_2	T_3	T_4
T_1	-	c_{12}	c_{13}	c_{14}
T_2	c_{21}	-	c_{23}	c_{24}
T_3	c_{31}	c_{32}	-	c_{34}
T_4	c_{41}	c_{42}	c_{43}	-

Table 4.2: Example of a matrix of choice frequencies with four stimuli.

where T_i is preferred over T_j, and the number of ties t_{ij}, where no preference between T_i and T_j is present. Ties are treated as half way decision, thus c_{ij} is obtained as follows [28]:

$$c_{ij} = 2 \times w_{ij} + t_{ij}. \tag{4.2}$$

Note that $c_{ij} + c_{ji} = 2 \times n_{ij}$, which is two times the number of comparisons. The Bradley-Terry-Luce model [8, 49] is used to convert the pair-comparison data to a continuous quality score. In this model, the probability p_{ij} of choosing T_i against T_j is expressed as:

$$p_{ij} = \frac{\pi(T_i)}{\pi(T_i) + \pi(T_j)}, \tag{4.3}$$

where $\pi(T_i)$ is the quality score of T_i, $\pi(T_i) \geq 0$ and $\sum_i \pi(T_i) = 1$. The parameters for $\pi(T_i)$ are estimated by maximizing a log-likelihood function and the confidence intervals are obtained from the Hessian matrix of the log-likelihood function [81]. The obtained parameters for $\pi(T_i)$ are referred to as (hypothetical) MOS in the course of this thesis.

4.3 Objective Quality Assessment

In this study, the performance of ten objective quality metrics is assessed (see Table 4.3). All quality metrics are computed using the MeTriX MuX Visual Quality Assessment Package [35]. The objective quality scores are computed from the novel views of the stereoscopic images that correspond to the right views and the original right views. A description of each metric can be found in Section 2.4.

In order to estimate the accuracy of the objective quality metrics used, we compute the Pearson linear correlation coefficient (PLCC) between our subjective and objective scores. The results from the subjective quality assessment are used as reference solution. The first step of the computation of the correlation coefficient aligns the range of the objective quality scores to the range of the subjective quality scores by using a linear least squares regression:

$$MOS_p = a \times score + b. \tag{4.4}$$

$score$ is the obtained score from the objective quality metric. a and b are the obtained parameters from the linear regression. MOS_p is the predicted MOS from the objective quality score. Next,

Abbreviation	Name
PSNR	Peak signal-to-noise ratio [78]
SSIM	Structural similarity index [79]
MSSIM	Multi-scale structural similarity index [80]
VSNR	Visual signal-to-noise ratio [15]
VIF	Visual information fidelity [66]
VIFP	Visual information fidelity pixel-based [66]
UQI	Universal quality index [77]
IFC	Information fidelity criterion [67]
NQM	Noise quality measure [17]
WSNR	Weighted signal-to-noise ratio [17]

Table 4.3: Objective quality metrics evaluated in this study.

$PLCC$s are obtained between MOS (corresponds to the obtained subjective quality scores as defined in Section 4.2) and MOS_p for each comparison set:

$$PLCC = \frac{\sum_{i=1}^{N}(MOS_i - \overline{MOS})(MOS_{p_i} - \overline{MOS_p})}{\sqrt{\sum_{i=1}^{N}(MOS_i - \overline{MOS})^2}\sqrt{\sum_{i=1}^{N}(MOS_{p_i} - \overline{MOS_p})^2}}, \tag{4.5}$$

where $\overline{MOS_p}$ and $\overline{MOS}$ are the average scores of MOS_p and MOS over the N stimuli of the corresponding comparison set. Finally, PLCC is averaged across the different comparison sets.

4.4 Summary

This chapter introduced the dataset, the general scheme and the evaluation process of the preliminary and the main study. The dataset consists of six stereoscopic images with different depth range and content. The evaluation process of both studies is divided into two parts, subjective quality assessment and objective quality assessment. We describe the methodology that we use in our subjective quality assessments, i.e., the paired comparison method. The subjective performances can be evaluated by analysing the given preferences of the subjects. Concerning the objective quality assessment, ten objective quality metrics are selected and the computation of the correlation between objective and subjective scores is explained.

CHAPTER 5

Results - Preliminary Study

This chapter discusses the results of the preliminary study. In this preliminary study, the following questions are addressed:

1. How do the subjects rate the design of the study?
2. What is the overall performance of the selected post-processing approaches?
3. Do the objective evaluation results correlate with the subjective evaluation results?

In order to answer these questions and to obtain a first idea of the selected post-processing approaches, the subjective and objective quality scores are obtained. Moreover, the correlation between the subjective and objective scores is computed. The results of the preliminary study are used to design the main study.

This chapter has the following structure. Section 5.1 briefly discusses the post-processing approaches used in the preliminary study. The results of the subjective quality assessment are given in Section 5.2, the results of the objective quality assessment in Section 5.3. Section 5.4 discusses the results of the preliminary study and their impact on the main study.

5.1 Evaluated Approaches

In the preliminary study, six different post-processing approaches were used to improve the disparity maps before the novel right views were generated. In addition to these generated novel views, the novel views that were generated from the unprocessed disparity map and the original views were considered as well. Table 5.1 lists the approaches and the corresponding parameters used in the preliminary study. The selection of the parameters was based on a visual judgement of the resulting novel views. Particular attention was paid to the reduction of visible artifacts in the novel views through the selected parameters of each post-processing approach. All post-processing approaches are developed in MATLAB. In total, 48 stereoscopic images

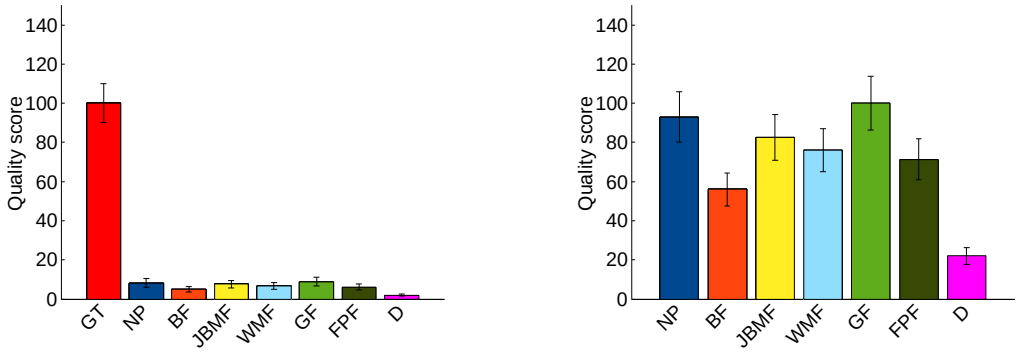
Abbr.	Name	Description
BF	Bilateral filter [75]	$r = 7, \sigma_s = 1.0, \sigma_d = 0.022$
JBMF	Joint bilateral weighted median filter [38]	$r = 7, \sigma_s = 1.0, \sigma_r = 0.022$
GF	Guided image filter [32]	$r = 7, \epsilon = 0.022^2$
WMF	Weighted mode filter [53]	$r = 7, \sigma_s = 1.0, \sigma_r = 0.022, \sigma_d = 25.5$
FPF	Foreground protecting filter [48]	$r = 5, s = 5$
D	Dilation [31]	$r = 7$
NP	No post-processing	unprocessed depth map
GT	Ground truth	original left and original right view

Table 5.1: This table lists the evaluated approaches of the preliminary study. For every approach, its abbreviation, name and parameter description is given.

(6 images $\times$ 8 approaches) were evaluated. For one comparison set, this resulted in $\frac{8\times7}{2} = 28$ pair comparisons, for all comparison sets in $28 \times 6 = 168$ pair comparisons.

The following description summarizes the most important aspects of each applied post-processing method (a detailed description is given in Chapter 3). The bilateral filter [75] (BF) is applied on the depth maps to locally average similar depths. The amount of smoothing is controlled by the parameters σ_d and σ_s for the depth range and the spatial range, respectively. In particular, larger σ_d and σ_s allow the averaging of less similar pixels in the local neighborhood. The joint bilateral weighted median filter [38] (JBMF[1]) changes depths according to local statistics of color similarity and spatial similarity. To preserve depths at object boundaries in the corresponding color image, JBMF computes filter weights according to color similarities in the corresponding color image. Similarly to BF, the parameters σ_r and σ_s control the color range and spatial range, respectively. Instead of locally averaging depths, JBMF chooses a neighboring depth according to the median filter weight. In some similarity to JBMF, the weighted mode filter [53] (WMF) is based on local statistics that consider color similarity, spatial similarity and similarity of depths that are adjusted by the parameters σ_r, σ_s and σ_d, respectively. WMF changes depths by seeking the mode of these statistics. Guided image filtering [32] (GF) averages depths with similar colors in the corresponding view. Thus, this edge-preserving filter is able to improve the alignment of depth edges with color edges in the corresponding view. The parameter ϵ adjusts the smoothing effect, i.e., large ϵ generate smoother results than small ϵ. The foreground protecting filter [48] (FPF) aligns depth edges with color edges by performing edge detection in the depth map and the corresponding color view followed by piecewise smoothing. The parameters r and s control the maximum expansion of a depth edge and the number of smoothing steps, respectively. Dilation (D) uses a structuring element (i.e., a squared object is used in our evaluations) to expand the shapes of objects within an image. The parameter r controls the size of the structuring element.

[1]Contrary to Hosni et al. [38], we apply JBMF to all pixels of a disparity map.



(a) (b)

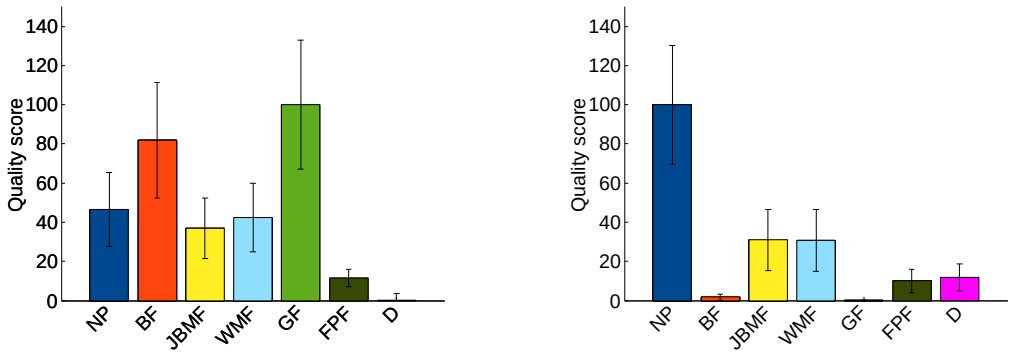
Figure 5.1: Comparison between subjective quality scores of the preliminary study where (a) the GT is included for score computation and (b) the GT is excluded from score computation. Note that GT denotes the reference solution where the stereoscopic images consist of original left and original right views. When the GT is not considered in the subjective score computation, the performance of the investigated post-processing methods can be better compared. The quality scores are normalized to obtain a maximum quality score of 100 per diagram for better visibility.

5.2 Results of Subjective Quality Assessment

Eight subjects (four female, four male) participated in the preliminary study (see Appendix, Table D.1 for a detailed information about the subjects). Seven of them were expert viewers with advanced experience in image processing and 3D content, one of them was a non-expert. The age ranged from 20 to 32 with an average of 28. All of the subjects were screened for visual acuity, color vision and stereo vision according to [10]. It should be noted that the number of subjects is relatively low. However, the main intention of the preliminary study was to determine whether the general study design was appropriate to evaluate the performance of depth map post-processing. To this end, the presented subjective scores are intended to give a first impression about the performance of the selected approaches.

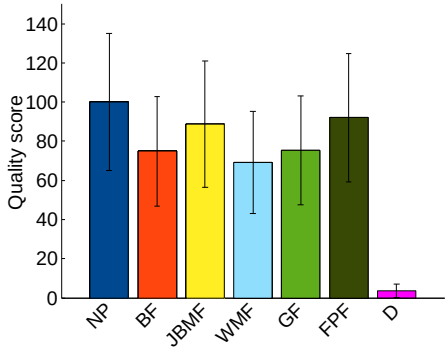
Based on the comments of the subjects regarding the experiment design, the following conclusions can be drawn. First, it was often difficult to judge the quality changes between the two stimuli. In this context, the opportunity of a no-preference decision turned out to be especially important. Without having this opportunity of a no-preference decision, the subjects would have to select a preferred stimulus randomly in inconclusive cases. This in turn could potentially have given a false result. Next, the opportunity to switch between the two stimuli helped the subjects to identify the visually distorted areas and, in consequence, to make a preference decision. Finally, the duration of the subjective experiment was perceived as too long. The subjects mentioned that especially towards the end their attention declined.

Figure 5.1(a) shows the results of the subjective study for all methods as described in Sec-

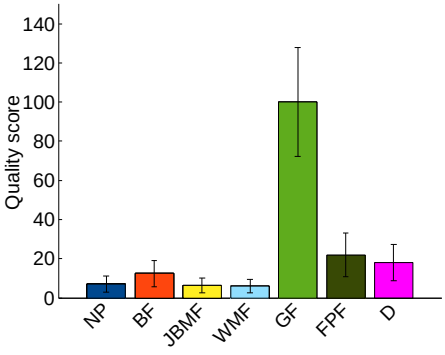


(a) monkeys000

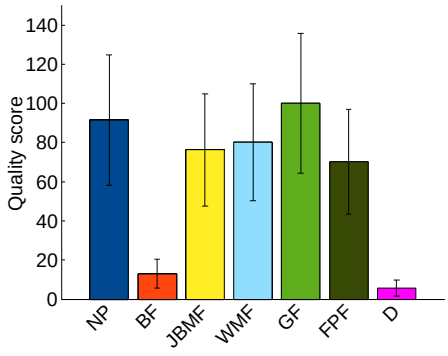
(b) tiger000



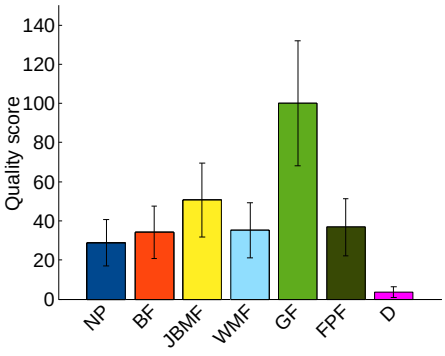
(c) gforce033



(d) gforce105



(e) musketiers264



(f) musketiers333

Figure 5.2: Quality scores of the preliminary study for each scene. The quality scores are normalized to obtain a maximum quality score of 100 per diagram for better visibility. Note that GT is not considered in these representations (see the text for a detailed explanation).

	NP	BF	JBMF	WMF	GF	FPF	D
PSNR	2	1	6	7	3	4	5
SSIM	2	1	6	7	3	4	5
MSSIM	2	1	6	7	3	4	5
VSNR	2	1	6	7	3	4	5
VIF	2	1	7	6	3	4	5
VIFP	2	1	7	6	3	4	5
UQI	2	1	6	7	3	4	5
IFC	1	2	7	6	4	3	5
NQM	3	1	6	7	2	4	5
WSNR	2	1	6	7	3	4	5
PC	2	6	3	4	1	5	7

Table 5.2: Rankings according to the measurements of the preliminary study.

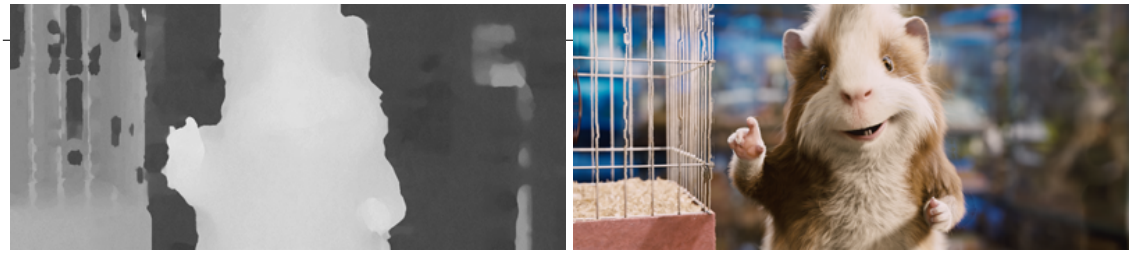
tion 5.1. It can be seen that the reference solution GT obtains the highest quality score of 100. In comparison to GT, the quality scores of the remaining methods are below 10 percent. This result was expected because GT consists of original left and original right views. Thus, no DIBR related artifacts are present. We exclude GT from the further analysis in order to better visualize the performance of the different post-processing methods on the subjective quality. To this end, Figure 5.1(b) and 6.3 show the subjective results where GT is not considered. It can be observed that filters that consider a guidance image (e.g., GF, JBMF, WMF) perform subjectively better than filters that only consider depths (e.g., BF, D). Concerning the former, filters that perform local smoothing (e.g., GF) rather than local statistics (e.g., JBMF, WMF) achieve higher subjective scores.

Nevertheless, all post-processing methods except for GF achieve overall lower quality scores than the unprocessed counterpart NP (as can be seen in Figure 5.1(b)). The parameters of GF applied the strongest smoothing on the disparity maps (see Figure 5.3). Thus, we decided to adjust the parameters of methods in the main study in order to ensure a fair comparison.

5.3 Results of Objective Quality Assessment

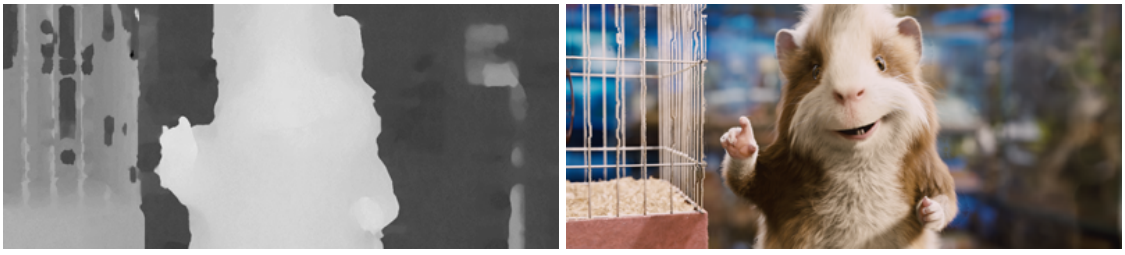
Table 5.2 shows the rankings of the different methods according to the objective (all rows except of the last one) and subjective (last row) quality scores. It can be seen that the rankings between the subjective and objective quality scores differ. For example, all objective quality metrics except for IFC rate BF as the best performing filter. However, according to the subjective scores, BF is only ranked on the sixth place. Another observation is that the rankings of the eleven metrics are similar.

Table 5.3 shows the obtained correlation between subjective and objective scores expressed in percent. All tested metrics have a correlation below 50 percent. SSIM obtains the highest correlation with 42.55 percent, VSNR the lowest one with 34.30 percent.

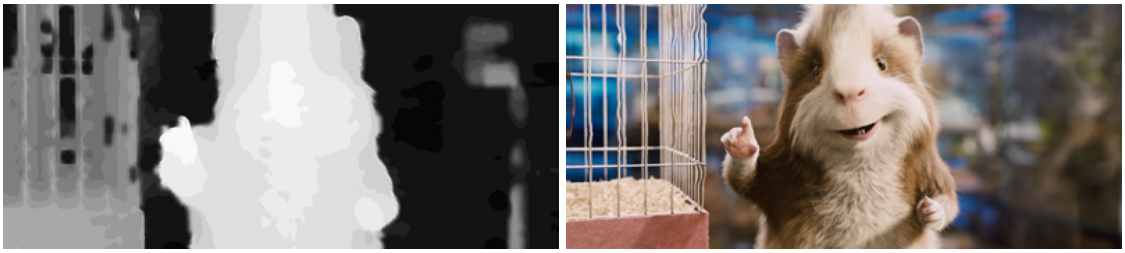


(a)

(b)



(c)



(d)

Figure 5.3: Comparison of four post-processed disparity maps and their impacts on the novel view quality. Left column shows the disparity maps and right column the obtained novel views. For visualization, the disparity maps are scaled to the intensity range of [0,255]. (a) No post-processing of disparity map. (b) Disparity map is post-processed with BF ($r = 7$, $\sigma_s = 1.0$, $\sigma_d = 0.022$). (c) Disparity map is post-processed with JBMF ($r = 7$, $\sigma_s = 1.0$, $\sigma_r = 0.022$). (d) Disparity map is post-processed with GF ($r = 7$, $\epsilon = 0.022^2$).

	PSNR	SSIM	MSSIM	VSNR	VIF
PLCC$_{PC}$	38.65	42.55	40.62	34.30	39.72
	VIFP	**UQI**	**IFC**	**NQM**	**WSNR**
PLCC$_{PC}$	40.32	40.08	36.39	39.29	38.60

Table 5.3: Correlation between subjective and objective scores in percent of the preliminary study. It can be seen that all objective quality metrics have a correlation below 50 percent.

5.4 Discussion

In the preliminary study, the disparity maps were post-processed with six selected post-processing approaches. Stereoscopic images were created out of the original left views and the novel right views. In addition, one stereoscopic image was created with the unprocessed disparity map and a second stereoscopic image was created from the original left and original right view. Eight subjects rated the perceived quality of the stereoscopic images with PC methodology. The obtained subjective results were evaluated in respect to the resulting quality scores. An objective quality assessment was performed between the novel right views and the original right views. The correlation between subjective and objective results was computed. From the obtained results, we can conclude the following:

- GT (Ground Truth) obtains the highest subjective scores and outperforms the remaining methods significantly. This result was expected because for GT the stereoscopic images consist of the original left and original right view. From the analysis of the results after the exclusion of GT, the performance of the remaining methods could be compared better. Therefore, GT will not be considered in the main study.
- GF (Guided Filter) was the best performing method. As the parameters of GF applied the strongest smoothing on the disparity maps, the parameters of the other methods will be adjusted in the main study. That should ensure a fair comparison of the post-processing methods.
- D (Dilation) was clearly the worst performing method and will not be considered in the main study.
- The survey showed that the selected study design is appropriate to subjectively evaluate the quality of stereoscopic images that contain a novel view. The subjects especially liked the opportunity to switch between the two stimuli.
- The duration of the experiment was perceived as too long. Thus, the number of evaluated approaches for the main study will be reduced.
- The correlation between objective and subjective results is weak. Thus, the focus of the main study will be on a subjective evaluation.

CHAPTER 6

Results - Main Study

This chapter presents the results of the main study. In the main study, the following questions are addressed:

1. Which depth map post-processing approaches achieve the best results in terms of novel view quality?

2. Do the characteristics of a scene (e.g., depth range) have an impact on the quality improvements of novel views through depth map post-processing?

3. Can the observation of the preliminary study be confirmed, that the objective quality metrics under investigation are not suited for the quality prediction of novel views?

This chapter is organized as follows. Section 6.1 briefly discusses the post-processing approaches used in the main study. Section 6.2 presents the results of the subjective quality assessment, Section 6.3 presents the results of the objective quality assessment. Section 6.4 discusses the results from the subjective and the objective evaluation and compares the results to each other.

6.1 Evaluated Approaches

In total, seven different methods consisting of six post-processing approaches and their unprocessed counterparts were used in the main study. All post-processing approaches were developed in MATLAB. This resulted in 42 stereoscopic images (6 images $\times$ 7 approaches), $\frac{7\times6}{2} = 21$ pair comparisons for one comparison set, and $21 \times 6 = 126$ pair comparisons for all comparison sets. Note that the number of total comparisons in the main study is lower compared to the preliminary study (i.e., 168 comparisons in the preliminary study). This reduces the duration of the subjective quality assessment because subjects complained about the duration in the preliminary study.

Abbr.	Name	Description
BF	Bilateral filter [75]	$r = 7, \sigma_s = 3.0, \sigma_d = 0.1$
JBMF	Joint bilateral weighted median filter [38]	$r = 7, \sigma_s = 3.0, \sigma_r = 0.1$
GF	Guided image filter [32]	$r = 3, \epsilon = 0.1^2$
GF+W	Guided image filter plus weighting	$r = 3, \epsilon = 0.1^2, T = 0.09$
WMF	Weighted mode filter [53]	$r = 7, \sigma_s = 3.0, \sigma_r = 0.1, \sigma_d = 25.5$
FPF	Foreground protecting filter [48]	$r = 5, s = 5$
NP	No post-processing	unprocessed depth map

Table 6.1: This table shows the evaluated approaches of the main study. For every approach its abbreviation, name and parameter settings are given.

Table 6.1 shows the post-processing approaches and their corresponding parameters used in the main study. The main study considers all approaches of the preliminary study except GT and D. GT was not considered because the use of this method allowed no conclusions with respect to the other methods in the preliminary study. Due to its worse performance in the preliminary study, D was replaced with a different post-processing filter GF+W. GF+W performs an additional weighting of GF's filter kernel that considers color differences in between views (e.g., down-weight differences that exceed a fixed threshold T) to further reduce the influence of erroneous depth assignments. A brief description of the remaining post-processing approaches and the influence of their parameters can be found in Section 5.1. A detailed description of these approaches is given in Chapter 3.

6.2 Results of Subjective Quality Assessment

Eighteen subjects (eight female, ten male) participated in the main study (see Appendix, Table D.2 for a detailed information about the subjects). Nine of them were expert viewers with advanced experience in image processing and 3D content, the remaining nine of them were non-experts. The age ranged from 22 to 38 with an average of 27. All subjects were screened for visual acuity, color vision and stereo vision according to [10]. Two subjects did not pass the vision screening and their ratings were discarded from the further analysis. Thus, the further subjective data processing is based on the ratings of 16 subjects.

The outlier detection method as described in Section 4.2.3 was used to detect outliers among the remaining 16 subjects. In particular, Figure 6.1 shows the transitivity satisfaction rate per subject for the performed main study. The scores of all 16 subjects are balanced. Thus, none of the subjects was rejected as an outlier.

Figure 6.2 and 6.3 show the results of the subjective study, i.e., the quality scores that are based on the preferences given by the subjects. For a better visibility, in the shown diagrams the quality scores are normalized to obtain a maximum quality score of 100 per diagram. As shown in Figure 6.2 in normalized representation, BF and GF on average achieve the highest quality scores of 100 percent and 58 percent, respectively. When examining the subjective quality scores for each category (i.e, scene) individually, we observe that BF has relatively large quality scores

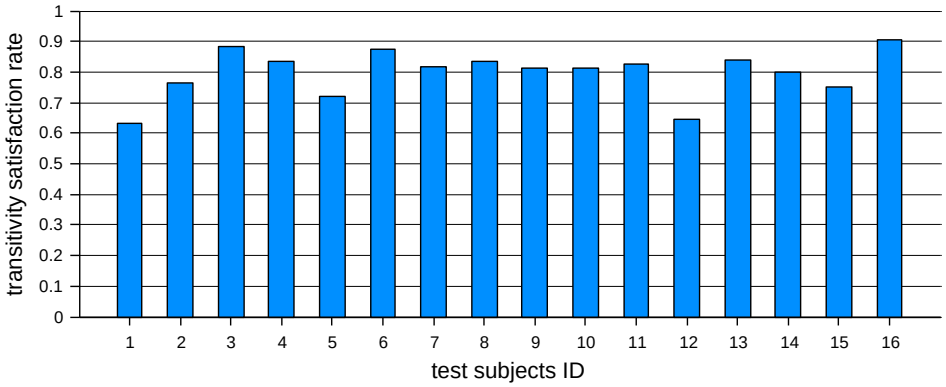


Figure 6.1: Transitivity satisfaction rate of the main study for each subject.

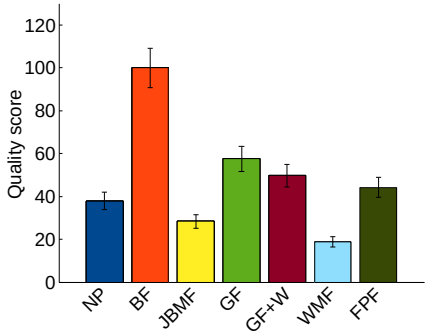
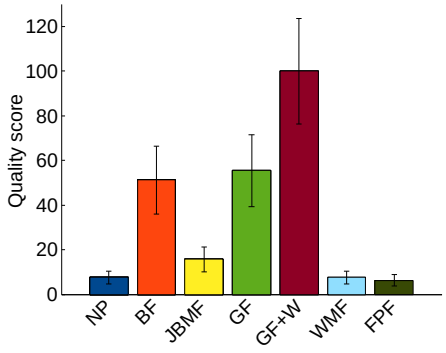


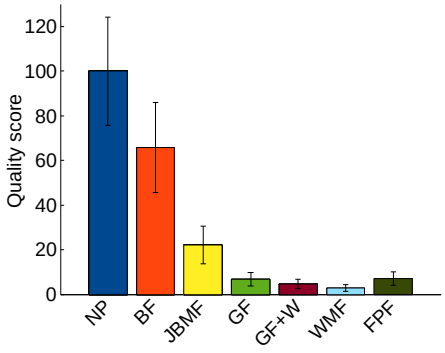
Figure 6.2: Overall subjective quality scores of the main study. The quality scores are normalized to obtain a maximum quality score of 100 for better visibility.

for scenes that contain large foreground objects in front of a distant background. This owes to the fact that BF tends to preserve depth edges with large depth differences (e.g., foreground object versus distant background), but averages similar depths (e.g., in the background). In comparison to BF, GF, GF+W and FPF do not consider local depth similarities when processing the depth maps. These filters can improve the alignment of depth edges with color edges, and also smooth depth values near depth edges with large depth differences. The performed smoothing might introduce smoothed depth edges that can lead to visible object deformations in novel views (which are especially visible at edges with large depth differences). The investigated post-processing methods that are based on local statistics (contrary to local smoothing) tend to show relatively small or no improvements compared to their unprocessed counterparts.

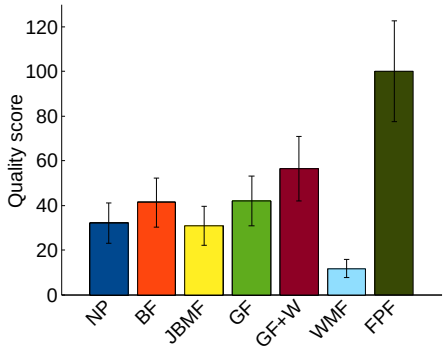
The relative ranks of the subjective quality scores of the post-processing methods vary across the investigated scenes. As can be seen in Figure 6.3(b), on *tiger000* the investigated post-processing methods do not achieve a visual improvement over the unprocessed counterpart (i.e., NP). In comparison to the remaining scenes, *tiger000* has a significantly larger depth range (see



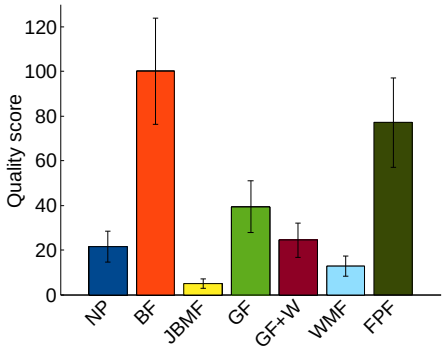
(a) monkeys000



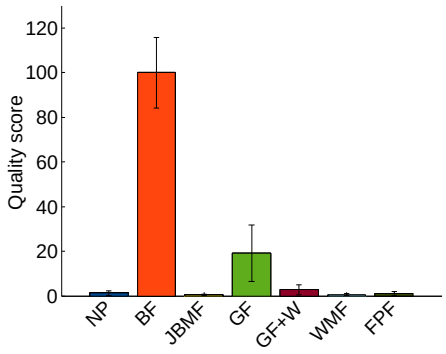
(b) tiger000



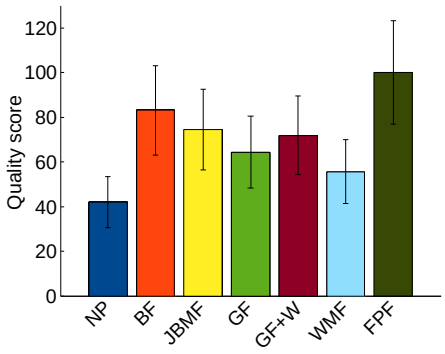
(c) gforce033



(d) gforce105

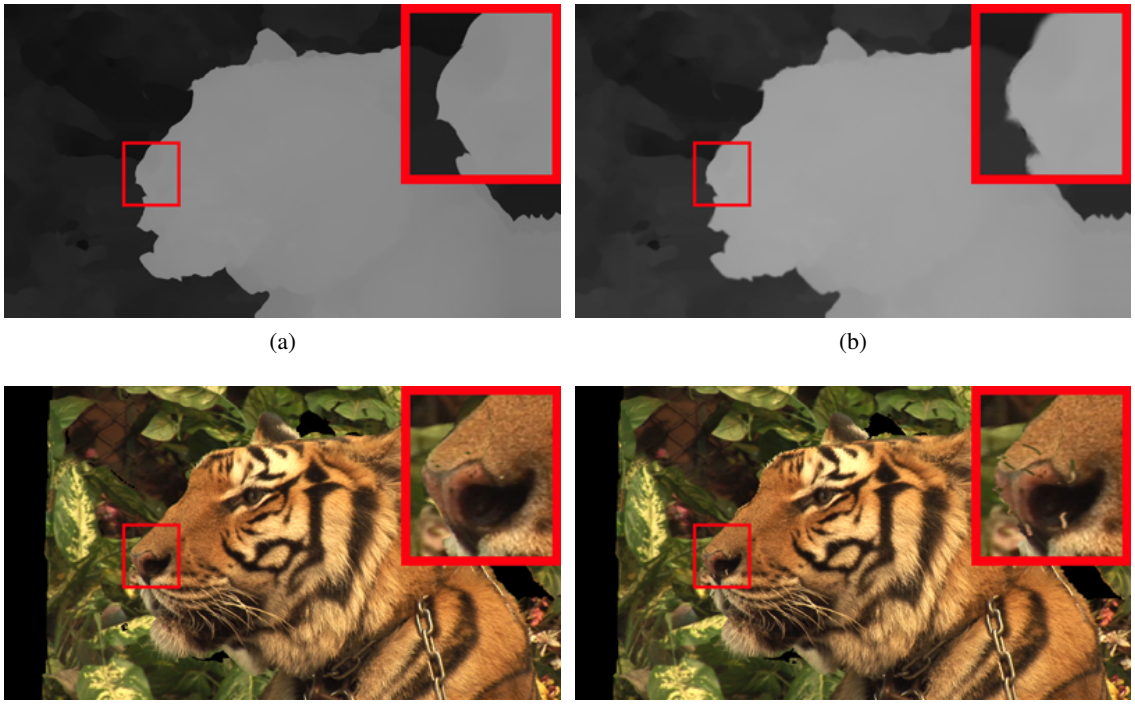


(e) musketiers264



(f) musketiers333

Figure 6.3: Quality scores of the main study for each scene. The quality scores are normalized to obtain a maximum quality score of 100 per diagram for better visibility.



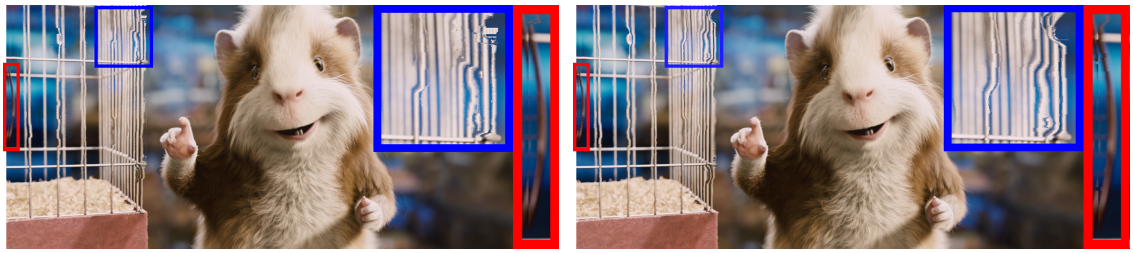
(a) (b)

(c) (d)

Figure 6.4: Comparison of GF post-processed disparity map to its unprocessed counterpart NP and the results on the novel view for the scene *tiger000*. (a) Unprocessed disparity map of *tiger000*. (b) Disparity map of *tiger000* post-processing with GF. (c) Novel view of *tiger000* obtained with unprocessed disparity map. (d) Novel view of *tiger000* obtained with disparity map post-processed with GF. The parameters of GF are $r = 3$ and $\epsilon = 0.1^2$. Note that in this example GF over-smoothes depths at object borders in the disparity map which leads to visual artifacts in the novel views.

Table 4.1), which increases the visibility of disparity errors in novel views and causes larger disocclusions near object boundaries. In this context, especially misalignments of depth and color edges (e.g., due to fuzzy object borders), and smoothed depths at object borders introduce large object deformations and visual artifacts at disocclusion boundaries (see Figure 6.4). In contrast to *tiger000*, for scenes with smaller depth ranges, e.g., *musketiers333*, the subjective preferences indicate improvements resulting from the investigated post-processing filters, when compared to their unprocessed counterparts.

A survey taken among the subjects regarding their personal impressions of the stereoscopic image quality gleaned the following details: First, the subjects perceived the inpainted disocclusions on the image borders as disturbing, especially for the scene *tiger000*. These disocclusions are caused by non-overlapping camera views and cannot be prevented by depth map post-processing. However, the quality of a depth map can influence the quality of the performed



(a) NP (b) BF

Figure 6.5: Comparison of (a) novel view generated with unprocessed disparity map and (b) novel view generated with GF post-processed disparity map. Blue zoom-in: artifacts at thin vertical structures caused by mismatches in the corresponding disparity maps. Post-processing the disparity map with GF improves the visual quality. Red zoom-in: artifacts at image borders caused by non-overlapping camera views. Post-processing has also an impact on these artifacts because pixels that are used for the inpainting of these areas are affected by depth map post-processing. In this example, the parameters of GF are $r = 7$, $\sigma_s = 3.0$ and $\sigma_r = 0.1$.

inpainting. Second, visual distortions at thin vertical structures (e.g., area around the cage for scene *gforce105*) were also perceived as disturbing. Although the selected post-processing approaches (e.g., GF) could improve the overall quality of these areas, artifacts in the novel views were still visible. Figure 6.5 shows examples of the two above mentioned observations.

6.3 Results of Objective Quality Assessment

Table 6.2 compares the rankings of the subjective and objective quality metrics. The objective evaluation was carried out using the ten quality metrics listed in Table 4.3. As can be seen from the main part of Table 6.2 (all rows except for the last one), the rankings achieved by the different objective metrics are relatively stable across the different metrics. For example, most of the tested objective metrics rate GF as best performing depth post-processing filter and in all cases BF was ranked last. For comparison, the last row of Table 6.2 shows the ranking results of the subjective study (from Figure 6.2). In particular, we observe a change in BF's rank from the last place in the objective evaluation to the first place in the subjective ranking. An important observation of our study is that there is only a weak correlation between the subjective and the objective results. More detailed correlation results are shown in Table 6.3. It can be seen that in all cases the correlation is below 50 percent.

6.4 Discussion

The main study was designed according to the results obtained from the preliminary study (i.e., selection of the post-processing methods, adaptation of the experiment duration). The disparity maps were post-processed with six different approaches. The unprocessed disparity map was considered as well. These disparity maps were used to create novel views. Stereoscopic

	NP	BF	JBMF	GF	GF+W	WMF	FPF
PSNR	2	7	4	1	6	3	5
SSIM	3	7	2	1	6	4	5
MSSIM	4	7	2	1	6	3	5
VSNR	3	7	4	1	6	2	5
VIF	3	7	1	2	6	4	5
VIFP	4	7	2	1	6	3	5
UQI	4	7	2	1	6	3	5
IFC	3	7	1	2	6	4	5
NQM	3	7	2	1	5	4	6
WSNR	2	7	4	1	6	3	5
PC	5	1	6	2	3	7	4

Table 6.2: Rankings according to the measurements of the main study.

	PSNR	SSIM	MSSIM	VSNR	VIF
PLCC$_{PC}$	40.23	44.11	43.39	36.48	44.45
	VIFP	**UQI**	**IFC**	**NQM**	**WSNR**
PLCC$_{PC}$	44.98	44.92	44.38	34.45	34.96

Table 6.3: Correlation between subjective and objective scores in percent of the main study. It can be seen that all objective quality metrics have a correlation below 50 percent.

images were formed that consisted of the original left views and the resulting novel views. The perceived quality of the stereoscopic images was evaluated with the PC methodology. Sixteen valid subjects were considered for the subjective score computation. In addition, an objective quality assessment with ten objective quality metrics was performed and the correlation between objective and subjective scores was computed.

The following conclusions can be drawn from the subjective quality assessment of the main study. Filters that perform local smoothing, i.e., the BF and the GF, achieve significantly higher quality scores than filters based on local statistics, i.e., the WMF and the JBMF, or their unprocessed counterparts. Both types of filters have edge-preserving properties. However, the filters based on local smoothing tend to result in smoother transitions at depth edges. Contrary to filters that perform local smoothing, the filters based on local statistics tend to preserve the sharpness of depth edges. This occurs because no averaging is performed. We believe that selective smoothing in depth maps can improve the visual quality of novel views and thus can be an explanation for the subjective performance of these local smoothing based filters. This observation would be consistent with those of various authors [19, 20, 46, 74] where a smoothing at object boundaries in the depth maps results in a higher novel view quality. In addition, the results of the main study indicate that the depth range within a scene affects the quality improvements through depth map post-processing. In particular, scenes with a low depth range gained higher quality improve-

ment than scenes with a large depth range. Especially the scene *tiger000* showed that large disocclusions at object boundaries and image borders can lead to visual artifacts in novel views.

The evaluation of the objective quality assessment leads to the following results: (1) subjective and objective methods obtained different rankings, (2) objective methods obtained similar rankings among each other and (3) the correlation between subjective and objective scores is weak. These three findings are compliant with [6], where the authors observe a different ranking between subjective and objective scores, a high correlation between objective scores, and a poor correlation between objective and subjective scores for DIBR generated content. Therefore, our results confirm the conclusions of [6]. The investigated objective quality metrics show a weak correlation compared to subjective scores. As the objective metrics used are originally proposed for 2D content, they detect and penalize artifacts that may not be disturbing in stereoscopic viewing conditions. In addition, none of the ten objective quality metrics has been designed for the special use case of novel views. Therefore, the objective quality metrics do not reflect the perceived quality and are not suitable to access the visual quality of stereoscopic images that contain novel views.

CHAPTER 7

Conclusion

In this thesis, we have analysed the impact of depth map post-processing on the quality of novel views. In particular, we have evaluated stereoscopic content that is formed from one original view and one novel view. We have focused on a subjective study. To this end, we have determined a suitable subjective evaluation methodology, which enables subjects to compare the visual quality of depth based stereoscopic content easily and that permits to draw conclusions on the post-processing methods used. In addition, we have explored the reliability of ten objective quality metrics for an automatic evaluation of DIBR/3D-content.

The main lessons learned from our study were that depth map post-processing can significantly enhance the quality of stereoscopic content that contains a novel view. In particular, the bilateral filter and the guided image filter achieved the overall best results for the scenes investigated in our main study. Both smoothing filters have edge-preserving properties and perform local averaging. However, the guided filter uses a guidance image to average depth values with similar colors in the corresponding views. Contrary to filters that perform local smoothing, non-averaging filters that are based on local statistics (e.g., joint bilateral median filter, weighted mode filter) showed no improvement compared to the unprocessed counterpart. An important observation was that the depth range within a scene had a strong impact on the quality of DIBR-based novel view generation and the effectiveness of depth post-processing. For scenes with a low depth range, post-processing of depth images resulted in more noticeable quality improvements than for scenes with large depth ranges. An additional finding was that the objective metrics under investigation were not successful in predicting the quality of stereo pairs that contain a novel view.

Future work could include the evaluation of depth map post-processing on stereoscopic video content, as opposed to still image content. In this context, particular attention should be paid to ensure a consistent filtering of the depth maps over time in order to avoid flickering artifacts in the novel views. Therefore, the filters must be adopted to consider not only the spatial domain but also the temporal domain. For instance, the work of [37] presents a spatio-temporal version of the guided filter. This temporal extension could be used in order to transfer the results of our work to stereoscopic video content.

Bibliography

[1] M. Barkowsky. *Subjective and objective video quality measurement in low-bitrate multimedia scenarios*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2009.

[2] A. Boev, D. Hollosi, and A. Gotchev. Classification of stereoscopic artefacts. Technical report, Department of Signal Processing, Tampere University of Technology, 2008.

[3] A. Boev, D. Hollosi, A. Gotchev, and K. Egiazarian. Classification and simulation of stereoscopic artifacts in mobile 3DTV content. In *Stereoscopic Displays and Applications XX*, pages 72371F–72371F–12, 2009.

[4] E. Bosc, P.L. Callet, L. Morin, and M. Pressigout. An edge-based structural distortion indicator for the quality assessment of 3D synthesized views. In *Picture Coding Symposium*, pages 249–252, 2012.

[5] E. Bosc, P.L. Callet, L. Morin, and M. Pressigout. Visual quality assessment of synthesized views in the context of 3D-TV. In *3D-TV System with Depth-Image-Based Rendering: Architectures, Techniques and Challenges*. Springer, 2013.

[6] E. Bosc, R. Pepion, P.L. Callet, M. Koeppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin. Towards a new quality metric for 3-D synthesized view assessment. *IEEE Journal of Selected Topics in Signal Processing*, 5(7):1332–1343, 2011.

[7] E. Bosc, R. Pepion, P.L. Callet, M. Pressigout, and L. Morin. Reliability of 2D quality assessment methods for synthesized views evaluation in stereoscopic viewing conditions. In *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pages 1–4, 2012.

[8] R.A. Bradley and M.E. Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3):324–345, 1952.

[9] N. Brosch, C. Rhemann, and M. Gelautz. Segmentation-based depth propagation in videos. In *Proceedings of the ÖAGM/AAPR Workshop*, pages 1–8, 2011.

[10] ITU-R Recommendation BT.1438. Subjective assessment of stereoscopic television pictures, 2000.

[11] ITU-R Recommendation BT.1683. Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference, 2004.

[12] ITU-R Recommendation BT.2021. Subjective methods for the assessment of stereoscopic 3DTV systems, 2012.

[13] ITU-R Recommendation BT.2022. General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays, 2012.

[14] ITU-R Recommendation BT.500. Methodology for the subjective assessment of the quality of television pictures, 2012.

[15] D.M. Chandler and S.S. Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 16(9):2284–2298, 2007.

[16] Colorvisiontesting. Available: `http://colorvisiontesting.com/ishihara.htm`. Accessed: 26 February 2014.

[17] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, and A.C. Bovik. Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, 9(4):636–650, 2002.

[18] I. Daribo. *Coding and rendering of 3D video sequences; and applications to three-dimensional television (3DTV) and free viewpoint television (FTV)*. PhD thesis, Graduate College of Telecom ParisTech, 2009.

[19] I. Daribo and H. Saito. Bilateral depth-discontinuity filter for novel view synthesis. In *IEEE International Workshop on Multimedia Signal Processing*, pages 145–149, 2010.

[20] I. Daribo, C. Tillier, and B. Pesquet-Popescu. Distance dependent depth filtering in 3D warping for 3DTV. In *IEEE International Workshop on Multimedia Signal Processing*, pages 312 –315, 2007.

[21] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli. A new full-reference quality metrics based on HVS. In *International Workshop on Video Processing and Quality Metrics*, pages 1–4, 2006.

[22] E. Ekmekcioglu, S. Worrall, D. De Silva, A. Fernando, and A.M. Kondoz. Depth based perceptual quality assessment for synthesized camera viewpoints. In *International Conference on User Centric Media*, pages 76–83, 2010.

[23] Emotion3D. Available: `http://www.emotion3d.tv/`. Accessed: 26 February 2014.

[24] C. Fehn. Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV. In *Stereoscopic Displays and Virtual Reality Systems XI*, pages 93–104, 2004.

[25] C. Fehn, K. Hopf, and B. Quante. Key technologies for an advanced 3D-TV system. In *Three-Dimensional TV, Video, and Display III*, pages 66–80, 2004.

[26] Food and Drug Assistance. Available: `http://www.fda.com/fdamd/claucoma.htm`. Accessed: 26 February 2014.

[27] O.P. Gangwal and R.-P. Berretty. Depth map post-processing for 3D-TV. In *International Conference on Consumer Electronics*, pages 1–2, 2009.

[28] M.E. Glickman. Parameter estimation in large dynamic paired comparison experiments. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 24(6):510–5523, 1999.

[29] M. Gong, J.M. Selzer, C. Lei, and Y.-H. Yang. Real-time backward disparity-based rendering for dynamic scenes using programmable graphics hardware. In *Proceedings of Graphics Interface*, pages 241–248, 2007.

[30] P.M. Grove. The psychophysics of binocular vision. In *3D-TV System with Depth-Image-Based Rendering: Architectures, Techniques and Challenges*. Springer, 2013.

[31] R.M. Haralick, S.R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):532–550, 1987.

[32] K. He, J. Sun, and X. Tang. Guided image filtering. In *European Conference on Computer Vision*, pages 1–14, 2010.

[33] M. Hecher. A comparative perceptual study of soft shadow algorithms. Master's thesis, Vienna University of Technology, 2012.

[34] D.M. Hoffman, A.R. Girshick, K. Akeley, and M.S. Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):1–30, 2008.

[35] MetriX MuX homepage. Available: `http://foulard.ece.cornell.edu/gaubatz/metrix_mux/`. Accessed: 26 February 2014.

[36] A. Hosni, M. Bleyer, C. Rhemann, M. Gelautz, and C. Rother. Real-time local stereo matching using guided image filtering. In *International Conference on Multimedia and Expo*, pages 1–6, 2011.

[37] A. Hosni, C. Rhemann, M. Bleyer, and M. Gelautz. Temporally consistent disparity and optical flow via efficient spatio-temporal filtering. In *Pacific Rim Conference on Advances in Image and Video Technology*, pages 165–177, 2012.

[38] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):504–511, 2013.

[39] G. J. Iddan and G. Yahav. 3D imaging in the studio and elsewhere ...,. In *Videometrics and Optical Methods for 3D Shape Measurements*, pages 48–55, 2001.

[40] ITU-T Recommendation J.144. Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference, 2004.

[41] P. Joveluro, H. Malekmohamadi, W.A.C Fernando, and A.M. Kondoz. Perceptual video quality metric for 3D video quality assessment. In *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pages 1–4, 2010.

[42] ISO/IEC JTC1/SC29/WG11. Call for proposals on 3D video coding technology. Doc. N12036, 2011.

[43] S.-W. Jung. Enhancement of image and depth map using adaptive joint trilateral filter. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(2):258–269, 2013.

[44] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Technology and Science*, 53(3):1–14, 2009.

[45] J.-S. Lee, L. Goldmann, and T. Ebrahimi. Paired comparison-based subjective quality assessment of stereoscopic images. *Multimedia Tools and Applications*, 67(1):31–48, 2012.

[46] S.-B. Lee and Y.-S. Ho. Discontinuity-adaptive depth map filtering for 3D view generation. In *Conference on Immersive Telecommunications*, pages 1–6, 2009.

[47] J. Lu, K. Shi, D. Min, L. Lin, and M.N. Do. Cross-based local multipoint filtering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 430–437, 2012.

[48] X.-H. Lu, F. Wei, and F.-M. Chen. Foreground-object-protected depth map smoothing for DIBR. In *IEEE International Conference on Multimedia & Expo*, pages 339–343, 2012.

[49] R.D. Luce. *Individual choice behavior: A theoretical analysis*. Wiley, 1959.

[50] G. Mather. *Foundations of perception*. Psychology Press, 2006.

[51] L.M.J. Meesters, W.A. IJsselsteijn, and P.J.H. Seuntiens. A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3):381–391, 2004.

[52] Bernard Mendiburu. *3D Movie Making*. Focal Press, 2009.

[53] D. Min, J. Lu, and M.N. Do. Depth video enhancement based on weighted mode filtering. *IEEE Transactions on Image Processing*, 21(3):1176–1190, 2012.

[54] K. Müller, P. Merkle, and T. Wiegand. 3-D video representation using depth maps. *Proceedings of the IEEE*, 99(4):643–656, 2011.

[55] ITU-T Recommendation P.910. Subjective video quality assessment methods for multimedia applications, 1997.

[56] P.L. Panum. *Physiologische Untersuchung über das Sehen mit zwei Augen*. Schwer, 1858.

[57] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, 4(1):1–73, 2009.

[58] R. Patterson. Human factors of stereo displays: An update. *Journal of the Society for Information Display*, 17(12):987–996, 2009.

[59] G. Petschnigg, R. Szeliski, M. Agrawala, M.F. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. In *Special Interest Group on Graphics and Interactive Techniques*, pages 664–672, 2004.

[60] C.C. Pham, S.V.U. Ha, and J.W. Jeon. Adaptive guided image filtering for sharpness enhancement and noise reduction. In *5th Pacific Rim conference on Advances in Image and Video Technology*, pages 323–334, 2011.

[61] M.H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50(3):312–322, 2004.

[62] N. Ponomarenko, F. Silvestri, K. Egiazarian, J. Astola, M. Carli, and V. Lukin. On between-coefficient contrast masking of DCT basis functions. In *International Workshop on Video Processing and Quality Metrics*, pages 1–4, 2007.

[63] C. Riechert, F. Zilly, M. Müller, and P. Kauff. Advanced interpolation filters for depth image based rendering. In *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pages 1–4, 2012.

[64] A.K. Riemens, O.P. Gangwal, B. Barenbrug, and R.-P.M. Berretty. Multi-step joint bilateral depth upsampling. In *Visual Communications and Image Processing*, 2009.

[65] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002. Middlebury Stereo Dataset: Available: `http://vision.middlebury.edu/stereo/data/` Accessed: 26 February 2014.

[66] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[67] H.R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005.

[68] D.V.S.X. De Silva, W.A.C. Fernando, H. Kodikaraarachchi, S.T. Worrall, and A.M. Kondoz. Adaptive sharpening of depth maps for 3D-TV. *Electronics Letters*, 46:1546–1548, 2010.

[69] A. Smolic, K. Müller, 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 *Conference on Picture Coding Symposium*, pages 389–392, 2009.

[70] M. Solh, G. AlRegib, and J.M. Bauza. 3VQM: A vision-based quality measure for DIBR-based 3D videos. In *International Conference on Multimedia and Expo*, pages 1–6, 2011.

[71] C. Sun, X. Liu, and W. Yang. An efficient quality metric for DIBR-based 3D video. In *International Conference on High Performance Computing and Communication & International Conference on Embedded Software and Systems*, pages 1391–1394, 2012.

[72] P. Surman. Stereoscopic and autostereoscopic displays. In *3D-TV System with Depth-Image-Based Rendering: Architectures, Techniques and Challenges*. Springer, 2013.

[73] ANSI T1.801.03. American national standard for telecommunications - Digital transport of one-way video signals - Parameters for objective performance assessment. Available: `http://www.ansi.org/`, 2003. Accessed: 26 February 2014.

[74] W. J. Tam, G. Alain, L. Zhang, T. Martin, R. Renaud, and D. Wang. Smoothing depth maps for improved stereoscopic image quality. In *Three-Dimensional TV, Video, and Display III*, volume 5599, pages 162–172, 2004.

[75] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *International Conference on Computer Vision*, pages 839–846, 1998.

[76] Video Quality Experts Group (VQEG). Final report from the Video Quality Experts Group on the validation of objective models of multimedia quality assessment, phase I. Available: `http://www.its.bldrdoc.gov/vqeg/projects/multimedia-phase-i/multimedia-phase-i.aspx`, 2008. Accessed: 26 February 2014.

[77] Z. Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.

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

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

[80] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multi-scale structural similarity for image quality assessment. In *Asilomar Conference on Signals, Systems, and Computers*, pages 1398–1402, 2003.

[81] F. Wickelmaier and C. Schmid. A Matlab function to estimate choice model parameters from paired-comparison data. *Behavior Research Methods, Instruments, and Computers*, 36(1):29–40, 2004.

[82] Wikipedia. Snellen chart. Available: `http://en.wikipedia.org/wiki/Snellen_chart`. Accessed: 26 February 2014.

[83] A. Woods, T. Docherty, and R. Koch. Image distortions in stereoscopic video systems. In *Stereoscopic Displays and Applications*, pages 36–48, 1993.

[84] X. Xu, L.-M. Po, K.-W. Cheung, K.-H. Ng, K.-M. Wong, and C.-W. Ting. A foreground biased depth map refinement method for DIBR view synthesis. In *International Conference on Acoustics, Speech and Signal Processing*, 2012.

[85] H. Yamanoue, M. Okui, and F. Okano. Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(6):744–752, 2006.

[86] S.L.P. Yasakethu, C.T.E.R. Hewage, W.A.C. Fernando, and A.M. Kondoz. Quality analysis for 3D video using 2D video quality models. *IEEE Transactions on Consumer Electronics*, 54:1969–1976, 2008.

[87] S.L.P. Yasakethu, S.T. Worrall, D.V.S.X. De Silva, W.A.C. Fernando, and A.M. Kondoz. A compound depth and image quality metric for measuring the effects of packet loss on 3D video. In *International Conference on Digital Signal Processing*, pages 1–7, 2011.

[88] B. Zhang and J.P. Allebach. Adaptive bilateral filter for sharpness enhancement and noise removal. *IEEE Transactions on Image Processing*, 17(5):664–678, 2008.

[89] L. Zhang and W.J. Tam. Stereoscopic image generation based on depth images for 3D TV. *IEEE Transactions on Broadcasting*, 51(2):191–199, 2005.

[90] Y. Zhao and L. Yu. A perceptual metric for evaluating quality of synthesized sequences in 3DV system. In *Visual Communications and Image Processing*, 2010.

[91] Y. Zhao, C. Zhu, L. Yu, and M. Tanimoto. An overview of 3D-TV system using depth-image-based rendering. In *3D-TV System with Depth-Image-Based Rendering: Architectures, Techniques and Challenges*. Springer, 2013.

APPENDIX A

User Instructions

The following written instructions were given to each subject before the start of the subjective evaluation. Depending on the preferred language of the subject, the instructions were given in English or German.

Subjektives Experiment "3D POST"

Willkommen bei der Forschungsgruppe "Interactive Media Systems" der Technischen Universität Wien. Sie werden an einem Experiment teilnehmen, welches Teil meiner Diplomarbeit ist. In diesem Experiment versuchen wir, den Einfluss von verschiedenen Nachbearbeitungsschritten bei der Darstellung von 3D-Inhalten zu untersuchen.

Was muss ich tun?

Bitte lesen Sie diese kurze Einführung genau. Hier wird erklärt, wie der Test abläuft. Da die Resultate für uns sehr wichtig sind, bitten wir Sie um Ihre volle Aufmerksamkeit während der nächsten halben Stunde.

Wie läuft das Experiment ab?

Sie werden nun auf einem Computerbildschirm immer nacheinander zwei 3D-Bilder (Bild A, Bild B) sehen. Um den Tiefeneindruck der Bilder wahrnehmen zu können, müssen Sie sich eine Shutterbrille aufsetzen. Diese wird Ihnen vor dem Beginn des Experiments zur Verfügung gestellt. Bei diesem Test werden sechs unterschiedliche Bilder verwendet, wobei diese zu 126 Bildpaaren zusammengefasst sind. Die einzelnen Bilder eines Bildpaares sind grundsätzlich identisch, wurden aber unterschiedlich nachbearbeitet und unterscheiden sich somit in der Qualität. Sie haben die Möglichkeit, mit den Cursortasten der Tastatur zw. den Bildern hin- und herzuwechseln, um so die Qualität beurteilen zu können.

Nachdem Sie sich eine Meinung über die Qualität der Bilder gemacht haben, bitten wir Sie, diese zu beurteilen, indem Sie uns folgendes mitteilen:

1) A ist besser

2) B ist besser

3) Gleich

Zunächst sehen Sie drei Testsequenzen, deren Bewertung nicht aufgezeichnet wird. Diese Testphase dient dazu, dass Sie sich mit dem Ablauf des Experiments vertraut machen und eventuell auftauchende Fragen beantwortet werden können. Ist der Testdurchgang beendet, teilt Ihnen der Testverantwortliche dies mit. Danach sehen Sie die 3D-Bilder des tatsächlichen Tests.

Wenn Sie fertig sind, teilt Ihnen der Testverantwortliche mit, dass Sie das Experiment beendet haben. Sie werden dann gebeten, einen Fragebogen auszufüllen. Falls Sie Kommentare zu dem Experiment haben, teilen Sie uns diese bitte nach dem Experiment mit.

Subjective Experiment "3D POST"

Welcome to the research group "Interactive Media Systems" at the Technical University of Vienna. You will participate in an experiment that is part of my master thesis. In this experiment, we try to investigate the influence of different post-processing steps in the presentation of 3D content.

What do I have to do?

Please read this short introduction. It explains how the whole test procedure is working. Since the results are very important for us, we ask for your full attention during the next half hour.

How does the experiment look like?

You will see two successive 3D images (image A, image B) on a computer screen. In order to perceive the depth impression of the images, you have to put on shutter glasses. The glasses will be provided to you before the experiment starts. Six different images are used in this test. They are combined to form 126 image pairs. The individual images of an image pair are basically the same, but were post-processed differently and thus differ in quality. By using the cursor keys on the keyboard you can switch between the two images in order to assess the quality.

Once you have made an opinion about the quality of the images, we ask you to judge this by telling us the following:

1) A is better

2) B is better

3) Same

First, you will see three test sequences whose vote is not recorded. During this testing phase you can make yourself familiar with the course of the experiment and possibly emerging questions can be answered. You will be informed about the end of the testing phase and subsequently the actual experiment starts.

When you are done with the experiment, you will be asked to answer a questionnaire. If you have any comments about the experiment, please let us know after the test has finished.

APPENDIX B

User Screening

All of the subjects performed a visual acuity test using Snellen charts, a color vision test using Ishihara plates and a stereo vision test according to [10]. In the following, the three vision tests are explained in more detail.

The Snellen chart (see Figure B.1) was printed in A4 format. The subjects were placed away at a distance of 2.8 meters from the chart. Subjects with contact lenses or glasses were tested with these visual aids. Each eye was tested individually by covering the opposite eye without pressing on it. The test operator showed on a specific letter and the subjects had to read the letter. Subjects with normal vision can read the letters of the 8th line.

The Ishihara plates (see Figure B.2) were printed out in A4 format and were arranged side by side. The subjects could come as close to the plates as they liked to and had to read the numbers inside the plates. Subjects with normal vision can read all of the depicted numbers inside the circles.

The stereo vision test consisted of two stereoscopic images denoted as VT-04 and VT-07. The first stereoscopic image VT-04 (see Figure B.3) tests the ability to perceive fine depth. Nine patches are provided and each of them has four circles in which only one circle has a small parallax. Subjects with normal vision can perceive the circle with a small parallax in front of the display screen (see Table B.1 for the correct answers). The second stereoscopic image VT-07 (see Figure B.4) tests the ability to perceive depth in random dot images. Subjects with normal vision can perceive a rectangular shape.

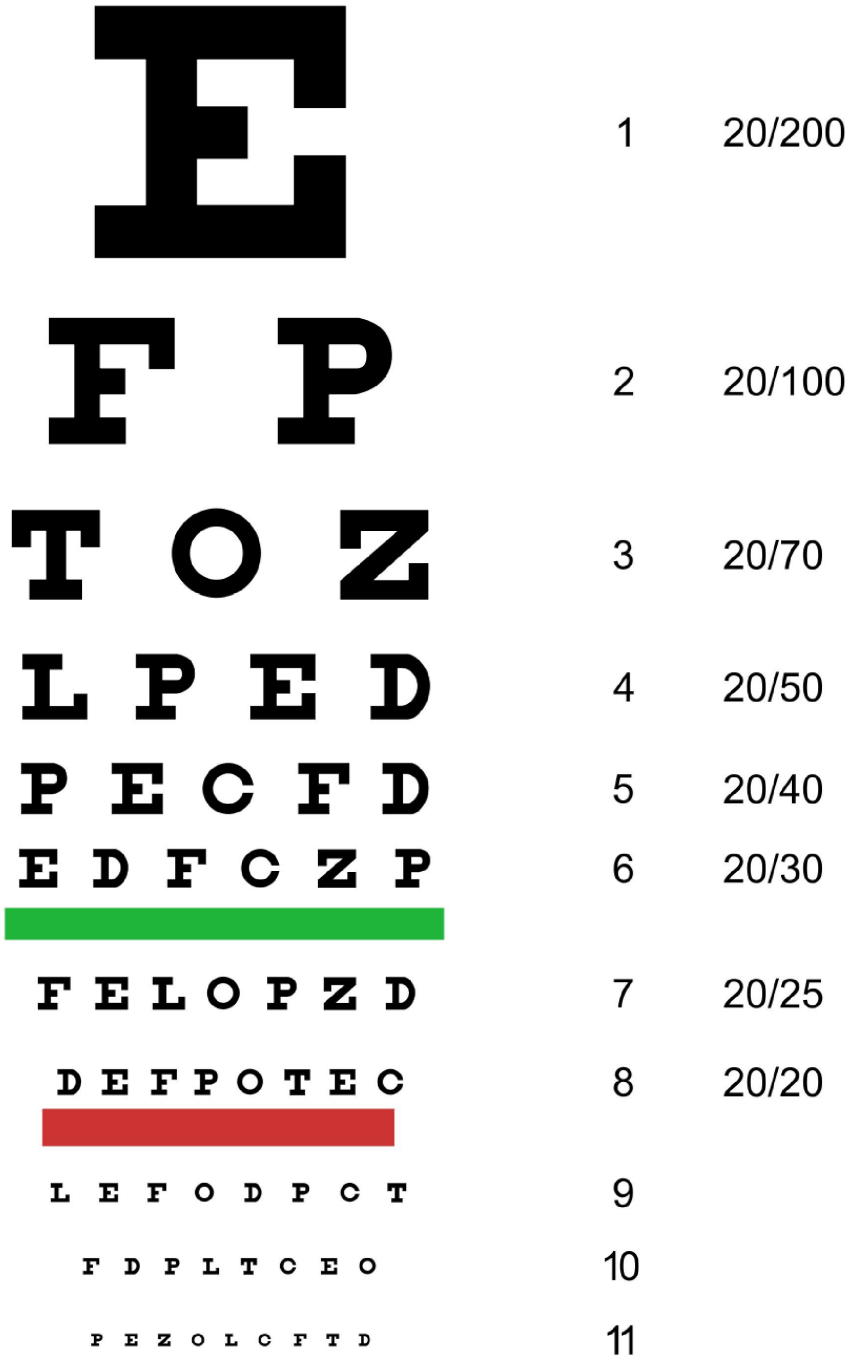


Figure B.1: Snellen chart used to determine the visual acuity of the subjects. Note that the chart is scaled to fit this page (Figure taken from [82]).

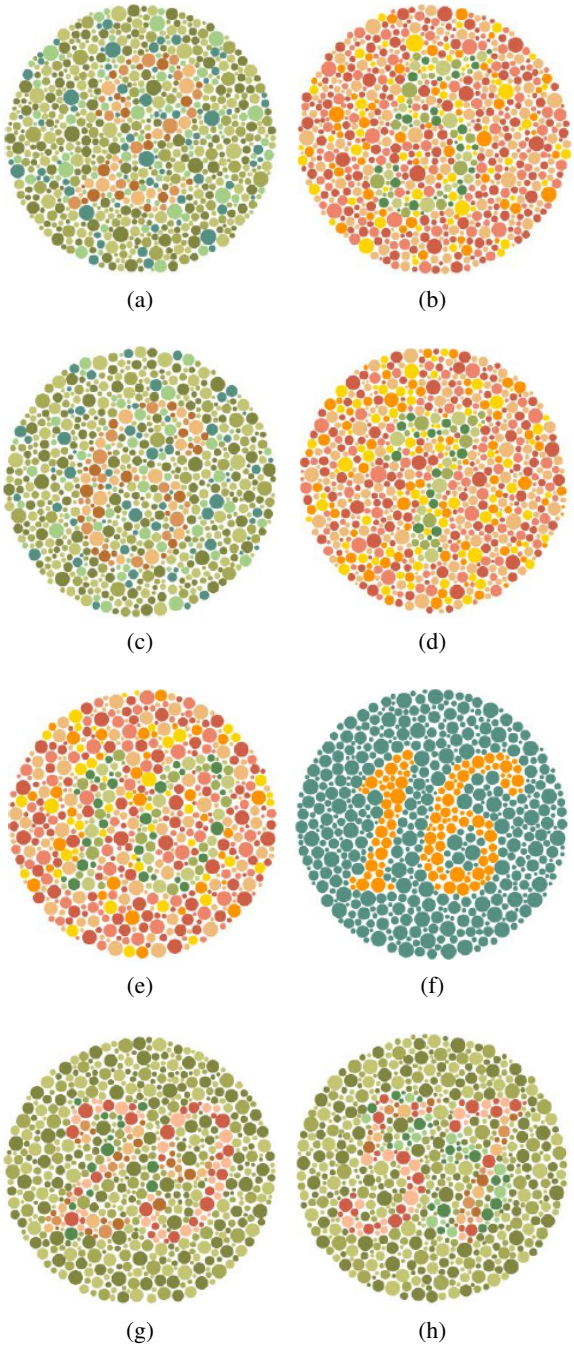


Figure B.2: Ishihara plates used to determine the color vision of the subjects. The following numbers are color coded in each plate: (a) 2, (b) 5, (c) 6, (d) 7, (e) 10, (f) 16, (g) 29, (h) 57. Note that the plates are scaled to fit this page (Figures taken from [16]).

(a) right view (b) left view

Figure B.3: Right and left view of stereoscopic test image VT-04 to determine the ability to perceive fine depth of the subjects (Figures taken from [10]).

Test number	Correct Answer	Angle of stereopsis at 3H
1	bottom	480
2	left	420
3	bottom	360
4	top	300
5	top	240
6	left	180
7	right	120
8	left	60
9	-	0

Table B.1: Correct answers and parallax of stereoscopic test image VT-04 (Table reproduced from [10] where more details can also be found).

(a) right view (b) left view

Figure B.4: Right and left view of stereoscopic test image VT-04 to determine the ability to perceive depth in random dot images of the subjects (Figures taken from [10]).

APPENDIX C

User Questionnaire

The subjects had to fill in the following questionnaire. The first page contains information about the subjects and had to be filled in before the start of the subjective evaluation. The second page contains questions about the test procedure and personal impressions and had to be filled in after the end of the subjective evaluation. The information at the top of the first page regarding test number, Snellen, Ishihara, Stereo VT04, Stereo VT07, test start, and test end had to be filled in by the test instructor and was used during the evaluation.

FAKULTÄT FÜR INFORMATIK
Faculty of Informatics

Test Number: _____

Snellen: 20 / __
Ishihara: __ / 8
Stereo VT04: __ / 9 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9
Stereo VT07: _____

Test Start: __:__
Test End: __:__

- **Age:** _____
- **Sex:**

  ☐ female ☐ male

- **Do you wear glasses or contact lenses?**

  ☐ yes ☐ no

- **Highest educational degree:**

  ☐ undergraduate ☐ graduate ☐ doctoral

- **Professional status:**

  ☐ employed ☐ unemployed ☐ student ☐ retired

- **Have you ever seen 3D movies?**

  ☐ yes ☐ no

  If "Yes", which display system was used?

  ☐ active shutter glasses ☐ passive shutter glasses

  ☐ without glasses ☐ I don't know

- **Do you have experience in image processing (Matlab, Photoshop, etc.)?**

  ☐ yes ☐ no

- **Did you suffer from any kind of headache or other discomfort when watching the 3D images?**

  ☐ yes ☐ no

  If "yes", what exactly?

  _____

  _____

- **Did any unpleasant picture noise or other unpleasant effects occur?**

  ☐ yes ☐ no

  If "yes", what exactly?

  _____

  _____

- **Would you like to try 3D at home?**

  ☐ yes ☐ no

  Why?

  _____

  _____

- **What do you think about the test setup?**

  _____

  _____

  _____

- **Do you have any additional comments?**

  _____

  _____

  _____

APPENDIX D

Detailed User Information

ID	Age	Gender	Education	Job	Optics	3D mov. exp.	Img. proc. exp.	Snellen	Ishihara	Stereo
1	20	w	Matura	Student	no	yes	no	20/25	8/8	8/9 (9)
2	28	m	Graduate	Student	no	yes	yes	20/20	8/8	8/9 (9)
3	26	m	Undergraduate	Student	no	yes	yes	20/50	8/8	5/9 (6,7,8,9)
4	28	w	Graduate	Student	yes	yes	yes	20/20	8/8	9/9
5	32	w	Doctoral	Employed	no	yes	yes	20/40	8/8	9/9
6	32	m	Doctoral	Employed	no	yes	yes	20/20	8/8	9/9
7	27	w	Graduate	Student	yes	yes	yes	20/25	8/8	7/9 (6,7)
8	30	m	Graduate	Student	no	yes	yes	20/20	8/8	8/9 (8)

Table D.1: Detailed information about the subjects of the preliminary study.

ID	Age	Gender	Education	Job	Optics	3D mov. exp.	Img. proc. exp.	Snellen	Ishihara	Stereo
1	24	m	Undergraduate	Student	yes	yes	no	20/20	8/8	8/9 (9)
2	29	f	Graduate	Student	yes	yes	yes	20/20	5/8	8/9 (9)
3	23	m	Graduate	Student	no	yes	yes	20/25	1/8	9/9
4	27	m	Matura	Employed	yes	yes	no	20/25	8/8	9/9
5	28	f	Graduate	Student	yes	yes	yes	20/20	8/8	8/9 (9)
6	30	m	Graduate	Employed	no	yes	yes	20/20	8/8	8/9 (9)
7	28	m	Graduate	Unemployed	no	yes	no	20/20	8/8	7/9 (8,9)
8	24	m	Undergraduate	Student	yes	yes	yes	20/25	8/8	4/8 (6,7,8,9)
9	31	f	Graduate	Employed	no	yes	no	20/25	8/8	8/9 (9)
10	28	m	Undergraduate	Student	no	no	no	20/20	8/8	7/9 (8,9)
11	28	m	Graduate	Student	yes	yes	yes	20/25	8/8	8/9 (9)
12	26	f	Undergraduate	Student	no	yes	no	20/20	8/8	9/9
13	24	f	Graduate	Student	yes	yes	no	20/20	8/8	8/9 (9)
14	26	f	Graduate	Employed	yes	yes	no	20/25	8/8	8/9 (9)
15	23	m	Matura	Student	yes	yes	yes	20/25	8/8	7/9 (8,9)
16	22	f	Undergraduate	Employed	yes	yes	no	20/20	8/8	8/9 (9)
17	28	m	Graduate	Employed	yes	yes	yes	20/20	8/8	8/9 (9)
18	38	f	Undergraduate	Employed	no	yes	yes	20/25	8/8	7/9 (8,9)

Table D.2: Detailed information about the subjects of the main study.