View rendering for 3DTV

Suryanaryana M. Muddala

Department of Information and Communication Systems
Faculty of Science, Technology and Media
Mid Sweden University

To My Family and Friends

Abstract

Advancements in three dimensional (3D) technologies are rapidly increasing. Three Dimensional Television (3DTV) aims at creating 3D experience for the home user. Moreover, multiview autostereoscopic displays provide a depth impression without the requirement for any special glasses and can be viewed from multiple locations. One of the key issues in the 3DTV processing chain is the content generation from the available input data format video plus depth and multiview video plus depth. This data allows for the possibility of producing virtual views using depth-image-based rendering. Although depth-image-based rendering is an efficient method, it is known for appearance of artifacts such as cracks, corona and empty regions in rendered images. While several approaches have tackled the problem, reducing the artifacts in rendered images is still an active field of research.

Two problems are addressed in this thesis in order to achieve a better 3D video quality in the context of view rendering: firstly, how to improve the quality of rendered views using a direct approach (i.e. without applying specific processing steps for each artifact), and secondly, how to fill the large missing areas in a visually plausible manner using neighbouring details from around the missing regions. This thesis introduces a new depth-image-based rendering and depth-based texture inpainting in order to address these two problems. The first problem is solved by an edge-aided rendering method that relies on the principles of forward warping and one dimensional interpolation. The other problem is addressed by using the depth-included curvature inpainting method that uses appropriate depth level texture details around disocclusions.

The proposed edge-aided rendering method and depth-included curvature inpainting methods are evaluated and compared with the state-of-the-art methods. The results show an increase in the objective quality and the visual gain over reference methods. The quality gain is encouraging as the edge-aided rendering method omits the specific processing steps to remove the rendering artifacts. Moreover, the results show that large disocclusions can be effectively filled using the depth-included curvature inpainting approach. Overall, the proposed approaches improve the content generation for 3DTV and additionally, for free view point television.

Keywords: 3DTV, view rendering, depth-image-based rendering, disocclusion filling, inpainting.

Acknowledgements

Contents

List of Papers

This thesis is mainly based on the following papers, herein referred by their Roman numerals:

   I Suryanarayana. M. Muddala and M. Sjöström and R. Olsson. Edge-Preserving Depth-Image-Based Rendering Method. In *International Conference on 3D Imaging, Liége, Belgium*, 2012.

  II Suryanarayana. M. Muddala and M. Sjöström and R. Olsson and S. Tourancheau. Edge-aided virtual view rendering for multiview video plus depth. In *3D Image Processing (3DIP) and Applications, IS&T/SPIE, Burlingame, CA, USA*, 2013.

III Suryanarayana. M. Muddala and R. Olsson and M. Sjöström. Disocclusion Handling Using Depth-Based Inpainting. In *The Fifth International Conferences on Advances in Multimedia, Venice, Italy*, 2013.

List of Figures

List of Tables

Terminology

Abbreviations and Acronyms

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
3DTV	Three-Dimensional Television
3DV	Three-Dimensional Video
ATTEST	Advanced Three-Dimensional Television System Technologies (EU project)
CDD	Curvature Driven Diffusion
DIBR	Depth-Image-Based Rendering
DM	Disocclusion Map
FTV	Freeview point Television
HDTV	High Definition Television
HVS	Human Visual System
IBR	Image Based Rendering
IST	Information Society Technologies
ITU	International Telecommunication Union
MPEG	Motion Picture Expert Group
MRF	Markov Random Field
MSSIM	Mean Structural Similarity Index Metric
MVD	Multiview Video plus Depth
MVC	Multiview Video Coding
MVP	MPEG-2 Multiview Profile
PC	Pair Comparison
PDE	Partial Differential Equations
PSNR	Peak Signal-to-Noise Ratio
RGB	Red, Green, Blue
SSD	Sum of Squared Differences
VSRS	View Synthesis Reference Software
V+D	Video plus Depth
YCbCr	Color space with one luma component and two chrominance com-

ponents, blue and red

Mathematical Notation

α	Scalar value
β	Depth enhancement parameter
λ	Interpolation parameter
γ	Depth variance threshold
ϵ	Scalar constant
$\nabla\cdot$	Divergence operator
$\mathbf{p}$	Pixel in an image
B	Baseline
C	Confidence term
D	Data term
H	Convolution kernel
J	3D Structure tensor
L	Depth regularity term
P	Priority term
$\mathbf{I}$	Identity matrix
$\mathbf{K}$	Intrinsic parameters matrix of a camera
$\mathbf{P}$	Projection matrix of a camera
$\mathbf{R}$	Rotation matrix
$\mathbf{P}_O$	Projection matrix for original camera
$\mathbf{P}_V$	Projection matrix for virtual camera
$\mathbf{t}$	Translational vector
$\mathbf{t}_L$	Translational vector for Left view
$\mathbf{t}_R$	Translational vector for Right view
$\mathbf{t}_V$	Translational vector for Virtual view
Z_{near}	Minimum depth value
Z_{far}	Maximum depth value
x	X-coordinate
d	disparity
f	focal length
g	Control function
k	Scalar curvature
$\mathbf{m}$	Camera pixels coordinate
$\mathbf{m}_o$	Original camera pixels coordinate
$\mathbf{m}_v$	Virtual camera pixels coordinate
z_o	Original camera depth value
z_v	Virtual camera depth value
I	Image or video frame
$I_{\mathbf{p}}$	Image intensity value at pixel $\mathbf{p}$

M	3D World Coordinate
Ω	Empty region or Hole region
Φ	Source region
Φ_b	Background source region
Φ_f	Foreground source region
$\delta\Omega$	Boundary region
$\delta\Omega_1$	One sided boundary region
$\delta\Omega_1^{'}$	Depth-guided boundary region
$\Psi_{\mathbf{p}}$	Patch centered at pixel P
$\Psi_{\hat{\mathbf{q}}}$	Estimated patch centered at $\hat{\mathbf{q}}$ from patch matching
Y	Pixel value in depth map
$Z_{\mathbf{p}}$	Depth patch centered at pixel $\mathbf{p}$
Z_c	Depth threshold

Chapter 1

Introduction

Over the last century, Television(TV) has greatly evolved from black and white television to High definition television (HDTV) and is known as one of the most visual entertainment systems. In recent years, however, there has been a significant increase in the development of video over internet and handheld devices, although TV still remains commonplace in terms of home entertainment.

3D video became increasingly popular with the technical advancements of communication, display technology and related fields and attracting the people's interest more and more since it provides natural viewing experience of the real world [Onu11]. From the recent developments in display technology, it is possible to envision the 3D TV being the next generation of home entertainment system by providing the viewer with immersive feeling of visual content. Multiview (multiple perspectives of the same scene) and free view (a scene from any perspective) point televisions are the most promising applications of 3DTV, as they create depth feeling without the need for 3D glasses. To provide such a 3D experience, these applications require many videos of the same scene captured from different perspectives. However, the problem can be reduced by providing only a few views and reproducing the remaining views on display. The quality of virtual views is affected by the rendering methods and the available data since it is not possible to capture and transmit the video of every view point. The important task in this area is to produce high quality virtual views using the available 3D video data representations.

This thesis work investigates the problems that appear in the rendered images and presents a novel depth-image-based rendering approach by utilizing actual projected information from the warping. Additionally, it investigates the inpainting methods in the context of filling missing areas in the rendered images and presents a novel inpainting method.

1.1 Background

This thesis work addresses the Depth-Image-Based Rendering (DIBR) for 3DTV. This section provides the motivation behind the DIBR and an overview of the thesis.

1.1.1 Motivation

Today the commercial success of 3D movies in the cinema and advancements in 3D technologies has increased the interest in developing the 3D TV and free view point television (FTV). These technologies aim at creating the immersive 3D experience and navigating around the scene in living room environments. In general, 3D video is obtained from two cameras capturing the same scene point from different perspectives and 3D displays project the captured scene onto each eye. Typically, a user uses special glasses to perceive the 3D experience. In theatre environments wearing these glasses is acceptable but in living room environments, wearing glasses is considered as an obstacle since people do move around.

The conventional 3D displays create 3D experience by providing a stereo pair of the same scene. Today, auto stereoscopic displays are available in the market associated with the recent developments in display technology [MVi, Phi]. These displays enable a multi user sensation in living room environments without any additional glasses. The displays require a number of video captures from different perspectives to provide of immersive 3D experience but, it is inefficient to capture and transmit every view point of a video due to capturing setup and transmission cost [MSD$^+$08]. To solve these issues, specific data representations are considered for 3DTV i.e. scene geometry information (i.e. depth in the scene) together with a specified number of videos. Utilizing the available information, additional views are rendered, which are required to be displayed on 3D displays. The concept of producing multiple views using scene geometry and reduced number of views is shown in Fig. 1.1. In the context of generating new view images, the DIBR method is used [SKS05]. Though it is a simpler and efficient method in generating virtual views, it has several limitations that are due to its inherent artifacts and the available data. The number of artifacts can be reduced by increasing the input data information. However limitations still remain in relation to producing the high quality virtual views using rendering methods from the captured scene in order to create a multiscopic 3D-presentation.

1.2 Overall Aim

The aim of this thesis work is to improve the 3D video quality experience by surmounting the limitations in the current 3D video distribution framework, particularly with regards to rendering. Rendering is a necessary task to produce a sufficient number of virtual views from the set of available views in order to create a multiscopic 3D-presentation.
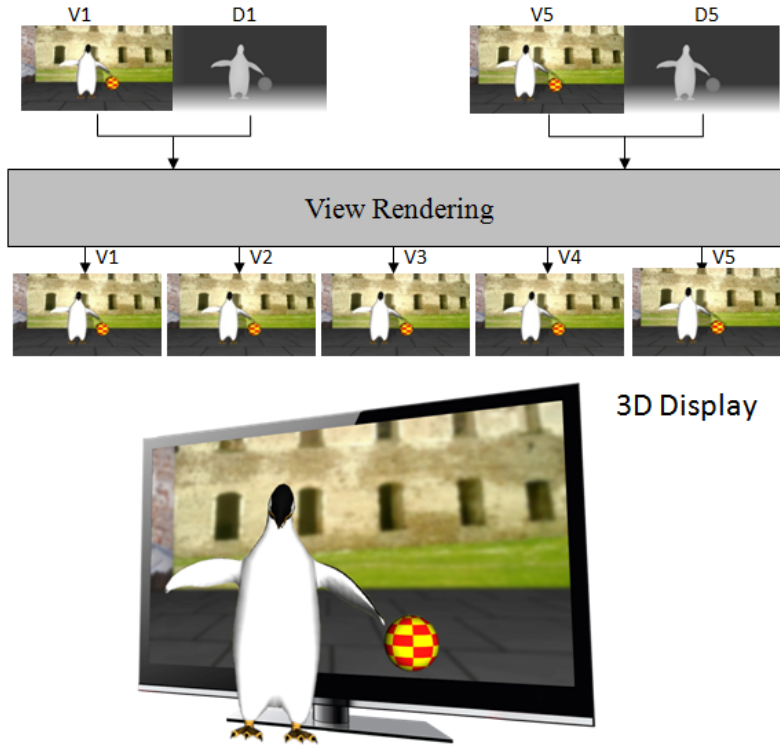
Figure 1.1: View rendering for multi view display using set of view plus depth information (Texture images are represent with V and Depth images represent with D and the number indicates the view point).

1.3 Scope

This thesis work will address the DIBR artifacts in the view rendering context (See Fig. 1.2). The rendering process is limited to the horizontal camera arrangement. Both horizontal and rotated camera setups are considered in the rendering process, where inpainting is used to fill the missing regions.

The pre-existing available uncompressed videos and depths data are used in the view rendering evaluation [DGK+09, SS02]. The perceived quality of the rendered views is difficult to measure without employing subjective tests, however, these tests do require time, proper planning and execution. Thus, subjective test is conducted for one case. An alternative to the subjective tests objective metrics considered due to the limitations of time and other resources.

Currently this thesis is limited to visual quality so the complexity has not been considered in evaluating the rendering methods.
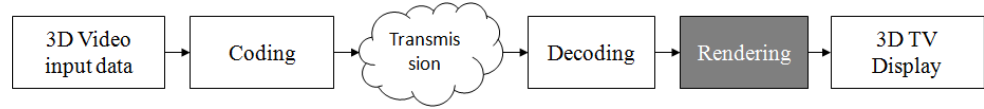
Figure 1.2: The 3DTV distribution chain.

1.4   Concrete and Verifiable Goals

In order to achieve a better 3D video experience on stereo and multiview autostereo-scopic displays, high-quality view rendering algorithms are required to produce good virtual views. The possible view rendering method should, effectively, use the information from both depth and texture video and process this information in such a way so as to produce better virtual views. The following goals are defined.

Goal I: Examine the problems of DIBR method.

Goal II: Propose an alternative DIBR method that reduces the rendering artifacts utilizing the available information for texture and depth data.

Goal III: Investigate the visual quality experienced by users.

Goal IV: Examine the methods to fill large disocclusions.

Goal V: Propose an inpainting method that addresses the disocclusion artifacts in the rendered views by utilizing the available information from the render-ing process.

1.5   Outline

The remainder of the thesis is organized as follows: In chapter 2, an overview of 3D video, human depth perception and 3D displays are presented. In chapter 3, an overview of DIBR, the related work and main contribution of this thesis are pre-sented. A more detailed description of hole filling methods inpainting, related work and the main contribution of this thesis are presented in chapter 4. The author's contributions are presented in chapter 5. A summary of the conclusions and future works are presented in chapter 6.

1.6   Contributions

The author's contributions of this thesis are presented in three papers. The co-authors have contributed to the formulation of evaluation criteria and the analysis. The contributions of this thesis are:

I An alternative formulation of DIBR method is proposed by introducing the edge-pixels idea.

The author contributed to the implementation of the edge-pixel idea, selection of evaluation criteria for rendering methods and presentation at conference.

II Extension of paper I for multiview video plus depth data and subjective evaluations.

The author contributed to the merging idea development and implementation. The author was involved in the selection of the objective evaluation, subjective test and wrote the paper with suggestions from the co-authors. The final presentation was done by the author.

III A method to fill the disocclusions in the virtual views by using depth-based texture synthesis.

The author is responsible for the depth-included curvature inpainting idea development and implementation. The author was involved in the selection of evaluation criteria for rendering methods that relied on inpainting and wrote the paper with suggestions from the co-authors. The final presentation was done by the author.

Each contribution is described in chapter 5.

Chapter 2

Overview of 3D video

In the previous chapter, the motivation, overall aim and goals of this thesis have been introduced. This chapter presents an overview of the 3D video i.e. history of 3D, distribution chain and 3D video formats considered in this thesis. Moreover, a brief description regarding the depth perception and 3D displays is provided.

2.1   History of 3D

The stereoscopic imaging or three dimensional imaging was initiated in 1838 when Charles Wheatstone introduced a mirror device that creates the illusion of depth by means of the principle of binocular vision [Whe38]. The principle is about fusing two slightly different views of the same scene into a single stereoscopic image. Sir David Brewster further developed the stereoscope in 1844. The Lumière brothers, at a later stage showed the first short 3D motion picture to the public in 1903. In the 1920s there was a boost in relation to stereoscopic cinema. Even though stereo films had been previously presented, the first 3D feature film was released in 1922. The first 3DTV experiments were started in 1928 by using the principle of the stereoscope. Despite these successful demonstrations it was not until 1950 that Hollywood tuned into 3D. Later success of 3D movies was affected because of the lack of stereographic experience, inadequate quality control and poor projection systems.

The first non experimental stereoscopic broadcast was aired in 1980, even though the first stereoscopic broadcast experiments started in 1953, due to difficulties in transmission and displays. In the early 1990s, 3DTV research efforts were raised due to the gradual transition from analog to digital service. Inspired by the interest in 3DTV broadcast services; the Motion Picture Expert Group (MPEG) started working on a compression technology for stereoscopic video sequence that resulted in a Multi View Profile (MVP) as part of the MPEG-2 standard [IL02]. It provides backward compatibility for conventional 2DTV.

After some attempts involving the integration of stereoscopic TV and HDTV, stereoscopic broadcast started with the Winter Olympics in Nagano in 1998 [JO02] and the first stereoscopic channel was started in 2010. On the other hand, 3D cinema had obtained a good reputation by deeper understanding of depth perception and image analysis. The improvements in 3D video technologies raised more interest in 3DTV and FTV [MFW08]. While 3DTV provides depth impression without wearing glasses and FTV allows the user to choose the view point of a scene in any direction. The 3D viewing experience was greatly improved by the support of advanced digital cameras and more understanding of depth perception. Moreover, the development of digital technology and 3D displays were significantly improved, which creates more interest in multiview video (MVV) applications.

2.2   3DTV approach

The Three-Dimensional Television (3DTV) framework was presented by the European Information Society Technologies (IST) project "Advanced Three-Dimensional Television System Technologies" (ATTEST) [RMOdBaI$^+$02]. In the succession, 3D4YOU, Multimedia Scalable 3D for Europe (MUSCADE) and more advanced projects aimed at supporting the wide range of 3D displays and applications [3DY], [3DI]. Before these projects, 3DTV relied on the concept of stereoscopic video, i.e. capturing, transmission and display of two separate videos for each eye. The basic 3D video transmission chain is shown in Fig. 1.2. In recent years, the display technology has been

(a) Texture image   (b) Depth map
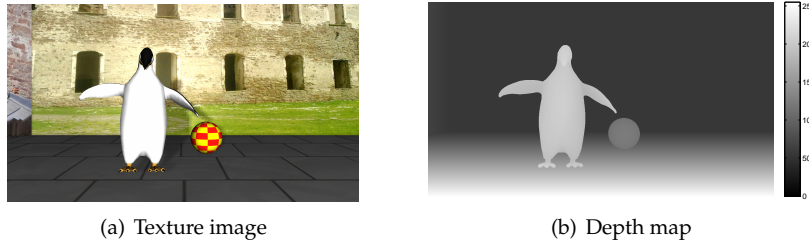
Figure 2.1: View plus depth format, texture image (a) and corresponding depth map (b).

developed immensely. Multiview 3D displays require a large number of input views to create immersive depth perception of the scene. The problem is now in relation to capturing and in the transmission of all the required views due to capturing and bandwidth limitations. To solve this issue, more flexible data representations are introduced that distribute the videos and associated per-pixel depth information. From this data representation, a number of required virtual views can be produced at the receiver side using the view rendering method, DIBR.

2.3 3D Video data representations

The new Three-Dimensional Video (3DV) data representations have different advantages and disadvantages over conventional stereoscopic video data formats. The advantages include bandwidth reduction, depth adjustment at the receiver and backward compatibility with 2D TV. Despite the advantages, there are several disadvantages, which must be considered in order to produce better quality 3D. The quality of the rendered view depends on the depth values of the original image and filling in the missing information in the stereo views. The new 3DV representations are video plus depth and multiview video plus depth [SMM$^+$09].

2.3.1 Video-plus-depth

Video-plus-Depth (V+D) is one of the well known data format for 3D video, consisting of a regular 2D video and corresponding depth per pixel information see Fig. 2.1. The 2D video represents colour, texture and structure information of a scene, whereas depth data represents the distance of per pixel data in a 3D scene from the camera centre. The depth is a gray scale image, i.e. the darkest value represents the farthest point and the brightest value represents the nearest point.

2.3.2 Multiview video-plus-depth

With the recent development in display technology, multiview autostereoscopic displays are now available within market. These displays required a greater number

of input videos in order to create multi user sensation without wearing glasses. Acquisition and transmission of a high number of views is inefficient, and the V+D data supports only to provide a limited number of views. Rendering artifacts can be added by increasing the number of virtual views to be rendered from single V+D. To solve this problem, new advanced 3DV format Multiview-Video-plus-Depth (MVD) is introduced. MVD consists of two or more number of V+D data and this data representation supports in producing the high quality virtual views.

2.4   Rendering

At the receiver side of the 3DTV distribution chain (See Fig. 1.2), the transmitted data is decoded in to colour video and depth data. With the decompressed data, stereo virtual views for the left and right eye are generated using the DIBR algorithm. These generated views are used to present on stereoscopic or multiview auto stereoscopic displays. As this thesis work is related to the rendering part, more details are presented in chapter 3.

2.5   Depth perception

The idea behind 3D technology is to create depth impression as are seen through naked eye. It is important to understand the basics of human depth perception in order to create the depth impression using display and 3D technology. The human visual system utilizes a variety of available visual information or cues to perceive the depth. The so called cues can be categorized into monocular and binocular cues. Monocular cues can be perceived by means of a single eye.

2.5.1   Monocular cues

- Accommodation: Change in the focus provides information about the relative depth or distance.

- Occlusion: Objects appear closer when one is in front of another; this is one of the strongest monocular cues.

- Perspective: Gives the relative depth when the objects are known to be the same size but the absolute depths are unknown.

- Motion parallax: when an object moves relative to stationary objects.

2.5.2   Binocular cues

- Stereopsis (binocular disparity): The projections of 3D space points onto the left and right eye at different positions. The difference between the projections is binocular disparity; closer objects have a larger disparity.

- Convergence: The two eyes are focused on same point on the objects, and so the angle between the eyes provides the depth cue.

Monocular cues are used to create depth on flat image surfaces. All these cues should be as consistent as possible in order to obtain a better depth perception [CV95]. Interested readers can refer to the detailed discussion of the effect of depth cues [ISM05, CV95].

2.6   Displays

Human beings see different perspectives of a scene through the left and right eye. Human Visual System (HVS) uses the difference in the perspectives such as disparity visual information and creates the sensation of depth. The majority of the stereoscopic 3D displays have been implemented by imitating the HVS. The 3D displays require at least two camera views captured from different perspectives to create depth impression by distributing each view to the left and right eye.

With recent developments, the 3D display technology can be categorized into stereoscopic displays and auto stereoscopic displays. Stereoscopic displays that rely on some kind of additional glasses are projecting the left and right views onto appropriate eye while auto stereoscopic displays do not requires any additional eye wear and the required optical elements are integrated in the display itself. These displays are also called free view displays.

2.6.1   Stereoscopic displays

The stereoscopic displays provide two slightly different perspective views and uses different multiplex methods to create the required signal for the appropriate eye. These displays can be categorized into colour multiplexed displays, polarization multiplex and time multiplex displays. In colour multiplexed or anaglyph displays, the two views of the left and right images are combined using the near complimentary colours (red and green, red and cyan or green and magenta). The user wears a pair of respective anaglyph glasses so that the colours are separated and thus receives the corresponding images for the left and right eye. The disadvantages associated with these displays are in relation to the loss of colour information and crosstalk. The shortcomings associated with anaglyph displays can be reduced by means of polarization multiplexed displays, in which two views with different perspectives are projected onto a display using orthogonally polarizing filter sheets (linear or circular polarization). Time-multiplexed displays exploit the memory effect of HVS. These displays require battery powered shuttered glasses.
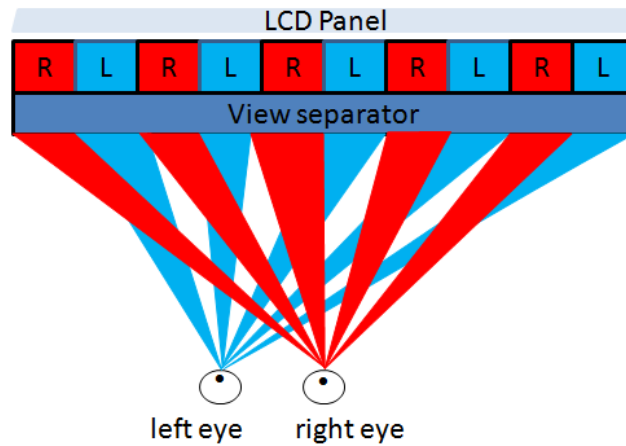
Figure 2.2: Illustration of auto stereoscopic display view generation.

2.6.2 Autostereoscopic displays

Autostereoscopic displays produce two or more of views and create a depth impression without additional eye wear and also create motion parallax. The autostereoscopic displays provide different views for the left and right eye using a view separator such as a parallax barrier or lenticular lenses see Fig. 2.2.

Two view displays produce a single stereo pair; this can be repeated to create motion parallax. These displays can be divided into parallax barrier displays and lenticular displays. In parallax barrier displays, parallax barriers are placed in front of or behind the LCD displays, which hide the parts of the image for one eye and show it to the other eye. The disadvantages associated with these displays are in terms of brightness loss and spatial resolution. With slanted barrier system, the spatial resolution problem can be avoided to some extent. The lenticular displays use cylindrical lenslets with flat panel displays in order to direct the light from a pixel to being visible from particular viewing angle. Lenticular sheet alignment with the displays is one of the difficult issues, which causes distortion for increased displayed resolution. In addition, intensity variations are another problem. These problems can be reduced by using a slanted lenticular array.

In *multi view displays*, large numbers of stereo pairs are presented. The stereo pairs can be viewed from different locations and multiple viewers can watch them simultaneously. More stereo pairs are required in order to get the smooth motion parallax, when viewer is moving around. The most common multiview auto stereoscopic displays are lenticular displays. The problem of reduction in horizontal resolution is reduced by means of slanted lenslets [LR06].

Modern 3D displays Holographic, Light filed, and Volumetric displays can solve the accommodation convergence problem. Interested readers can follow these articles [Pas05, UCES11].

Chapter 3

Depth-image-based rendering

The concept of 3DTV and the distribution of 3D content using V+D and MVD formats has been presented in section. 2.3. View rendering is a necessary task in order to produce a sufficient number of views for the distribution of 3D content. View rendering and view synthesis are terms used in the computer vision field for producing novel views from the available information. The available data consists of regular 2D videos and associated depth information, which is a gray scale representation of a scene. The common algorithm used for generating the novel views is DIBR, also known as 3D image warping. Although DIBR is an efficient method in producing novel views, it suffers from rendering artifacts within the virtual views. These artifacts limit the quality of the rendered view and consequently affect the perception of depth and visual quality.

This chapter starts by classifying the rendering methods, and then introduces background information regarding pinhole camera model, warping and depth characteristics in order to understand the DIBR approaches. Afterwards, the related work in the DIBR is presented. In particular, the section involving the proposed edge-aided rendering approach addresses the DIBR artifacts for the application of 3DTV.
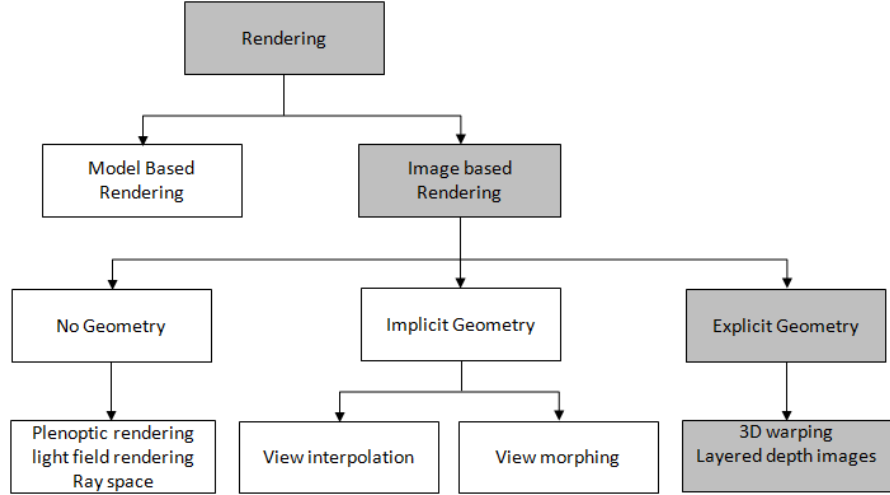
Figure 3.1: Rendering methods classification.

3.1   Overview of rendering methods

Rendering methods are categorized into Model Based Rendering methods (MBR) and Image Based Rendering methods (IBR) depending on the amount of 3D information being used. Model based methods, require information about scene models to generate new views, whereas IBR methods use available images and geometric data. The IBR methods are further classified, based on the availability of geometric information (see Fig. 3.1). DIBR methods use the explicit geometry information depth maps in producing additional views. This thesis work is related to rendering methods using explicit geometry.

3.2   Pinhole camera model

Projective geometry is a well organized mathematical framework that can be used to describe a pinhole camera model. This geometric principle is used to model image formation, view generation and 3D scene reconstruction from multiple images.

Pinhole camera is the simplest imaging device and its model is described by optical centre (also known as camera projection centre) and the image plane [SKS05]. The distance between the optical centre and the image plane is the focal length f and the line perpendicular to the image plane is the principal axis. The intersection of the principle axis with the image plane is called the principal point and the plane parallel to the image plane is the focal plane or principal plane (see Fig. 3.2(a)).

The consideration is that the camera centre is located at the origin of Euclidian coordinate system and that the optical axis is collinear with the z-axis. Suppose,

(a) Projections illustration   (b) Geometric relations
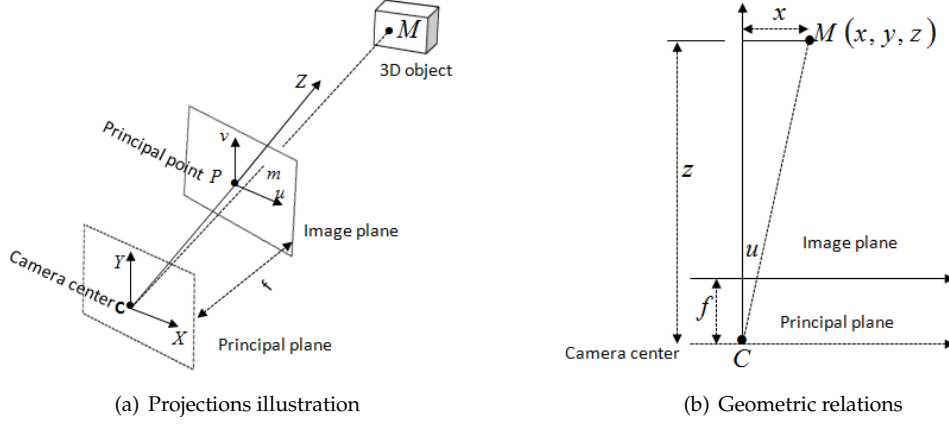
Figure 3.2: Pinhole camera model.

3D world point $(x, y, z)^T$ is projected onto the image plane position at $(u, v)^T$. The mapping of 3D world point of a scene to the 2D image plane is called the perspective projection and using homogenous coordinates it can be reformulated into:

$$z \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \tag{3.1}$$

where $u = \frac{xf}{z}$, $v = \frac{yf}{z}$ are pixel positions on the image. The matrix in the Eq. (3.1) is called the projection matrix, which defines the linear mapping of a 3D point on a 2D image plane and it can be written as:

$$z\mathbf{m} = \mathbf{PM}, \tag{3.2}$$

where $\mathbf{M} = (x, y, z, 1)^T$ and $\mathbf{m} = (u, v, 1)^T$ are homogenous coordinates of the 3D world point and its projection coordinate on the image plane. $\mathbf{P}$ is the projection matrix of the camera and contains only information concerning the focal length. In general the projection matrix is a 3x4 full rank matrix with 11 degree of freedom and it can be decomposed into

$$\mathbf{P} = \mathbf{K}\left[\mathbf{R}|\mathbf{t}\right], \tag{3.3}$$

where $\mathbf{K}$ is intrinsic parameters matrix, $[\mathbf{R}|\mathbf{t}]$ is extrinsic parameters matrix. $\mathbf{K}$ is the camera calibration matrix with 5 degree of freedom that describes the focal distance f, skew parameter s, image centre o_x, o_y and camera pixels size s_x, s_y and is defined as

$$\mathbf{K} = \begin{bmatrix} f/s_x & s & o_x \\ 0 & f/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.4}$$
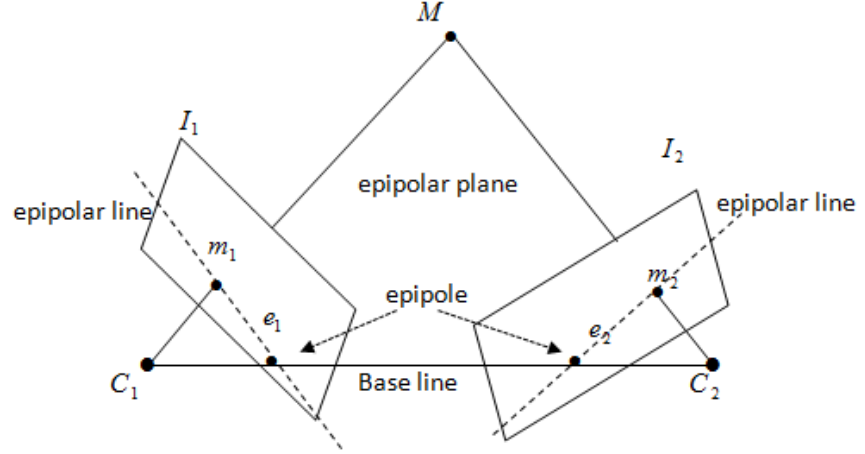
Figure 3.3: Two view geometry.

The extrinsic parameters, the rotation matrix $\mathbf{R}$ with 3 degree of freedom and the translation vector $\mathbf{t}$ with 3 degree of freedom describe the camera location and orientation with respect to the world coordinate. Moreover, the projection is not defined for the camera projection centre.

3.2.1   Two-view geometry

In the previous part, the geometry of a single camera is presented; now the geometry of two camera views (stereo camera) is briefly described. Suppose two cameras capturing the same scene from different perspectives and the principle of acquiring the 3D scene point $\mathbf{M}$ is shown in Fig. 3.3. Usually, the 3D scene information is available in both views apart from the occluded parts. Let $\mathbf{M}$ be an unoccluded 3D point that is projected onto two images through their camera centres C_1, C_2 at positions $\mathbf{m}_1$, $\mathbf{m}_2$ respectively. The two points are projections of the same 3D point so they are called corresponding points. On the other way, the 3D scene point $\mathbf{M}$ can be estimated from the corresponding image points $\mathbf{m}_1$ and $\mathbf{m}_2$ using triangulation. In general, the two cameras C_1 and C_2 are related by their position, orientation and internal geometry. The concept that describes the relations between the two cameras from the corresponding points is called epipolar geometry [HZ04]. By using this geometry the object points in one camera view can be extracted from the same object in the other camera's view.

3.3   Depth map characteristics

Depth map (aka depth image) is crucial information for producing the new viewpoints from the existing view images. In general, a depth map provides information

about the distances of objects in 3D space. The depth map can be obtained by means of computer graphics, range sensors and stereo analysis.

Computer generated depth data provides the true depth information and the common problems in depth from stereo matching such as colour differences and occlusions are avoided but, the drawback is that it is not real data.

Range sensors depth data extracted from range sensors, which are triangulation based sensors or time-of-flight based sensors. The main principle of the triangulation-based methods involves the triangle formed by a distinct point on an object using a laser pointer and a camera. Another approach, measuring the distance, depends on the signal travel time between the sender and detector. The cameras relying on this principle are named as Time-of-Flight (ToF) cameras. One of the main problems associated with range sensors is a resolution mismatch between the colour image and the acquired depth data [LS01].

Stereo depth is obtained from the extracted disparity information. The difference between the corresponding points in the two views is defined as disparity, where depth is expressed as the distance of a 3D object from the camera position. Let f and B be the camera's focal length and baseline. The relation between disparity d and depth z are related as:

$$z = \frac{Bf}{d}.$$
(3.5)

Intially, the disparity is estimated from the pixel correspondence between the two views i.e. for each pixel in one image, its corresponding pixel can be obtained by searching along the epipolar line. This process is called as disparity estimation. Once the disparity for each pixel is obtained then depth value Y can be determined using maximum and minimum disparity as follows:

$$Y = 255 \cdot \frac{d - d_{min}}{d_{max} - d_{min}}.$$
(3.6)

The depth map is an 8 bit gray scale values, i.e. gray value 0 represents the farthest point and value 255 represents the nearest point. In order to generate virtual views, the depth value should translate into real metric values with reference to a near and far depth clipping plane. Usually, depth obtained from stereo suffers from occlusions, low texture areas, and repetitive textures in the stereo images. These shortcomings can be reduced by post processing. However, the errors in the depth map lead to artifacts in the rendered views.

3.4   3D warping

The depth map and camera parameters for each view are available; a new view can be rendered using the principle of 3D warping. It is a two step process, in which the original image points $\mathbf{m}(u, v)$ are firstly reprojected into 3D space points or world

coordinates utilizing depth information. In the next step, the 3D points are projected onto a new camera image plane at the required viewing position. The process of converting 2D-to-3D and 3D-to-2D also called forward warping.

Consider two camera views and an arbitrary 3D scene point at homogenous coordinates $\mathbf{M} = (x, y, z, 1)^T$ (See Fig. 3.4). The projection of the 3D point onto the reference and virtual image planes at pixels positions $\mathbf{m}_L = (u_L/z_L, v_L/z_L, 1)^T$, $\mathbf{m}_V = (u_V/z_V, v_V/z_V, 1)^T$ respectively. The perspective projection of the 3D point onto the two views is given by

$$z_L \mathbf{m}_L = \mathbf{P}_L \mathbf{M}, \tag{3.7}$$

$$z_V \mathbf{m}_V = \mathbf{P}_V \mathbf{M}. \tag{3.8}$$

The 3D point $\mathbf{M}$ can be expressed by rearranging Eq. (3.7):

$$\mathbf{M} = z_L \mathbf{P}_L^{-1} \mathbf{m}_L. \tag{3.9}$$

By substituting Eq. (3.9) into Eq. (3.8), the relation between the virtual view pixel and reference view pixel is give by

$$z_V \mathbf{m}_V = z_L \mathbf{P}_V \mathbf{P}_L^{-1} \mathbf{m}_L. \tag{3.10}$$

Eq. (3.10) is applied to general warping case meaning the rendering of virtual views when the input cameras are involved in some rotation. Camera misalignment leads to the keystone distortion in the stereo visualization that can be avoided by capturing the images with the 1D parallel camera arrangement or by the rectification of the images after capture. Once the cameras are aligned, there is no rotation change in the two cameras, only a change in the translational vector i.e. along the x-direction. Hence, Eq. (3.10) is simplified into

$$u_v = u_o + \frac{f \cdot (t_{x,v} - t_{x,o})}{z_o} + (o_{x,v} - o_{x,o}), \tag{3.11}$$

where $t_{x,v}$, $t_{x,o}$ are the horizontal components of the translation vector $\mathbf{t}$, for the virtual and original views; z_o is translated real depth value from the depth map, $o_{x,v}$, $o_{x,o}$ are the principal component offsets for the virtual and original views, respectively.

Using Eq. (3.10) or Eq. (3.11), the synthesis of the virtual view I_V is obtained from the reference view I_L and the corresponding depth information Z_L. Normally, the input pixel locations are not mapped to the integer positions of the virtual view, so that projected pixels are rounded to the nearest integer. Following this the colour and depth values of input pixels m_L are copied to m_V. Because of this rounding to the nearest value, a few undefined pixels will appear in the virtual view. In some situations, more than one pixel from the original view are projected onto the virtual view at the same location; usually this case occurs when the pixels are occluded due to less distance between the pixels.
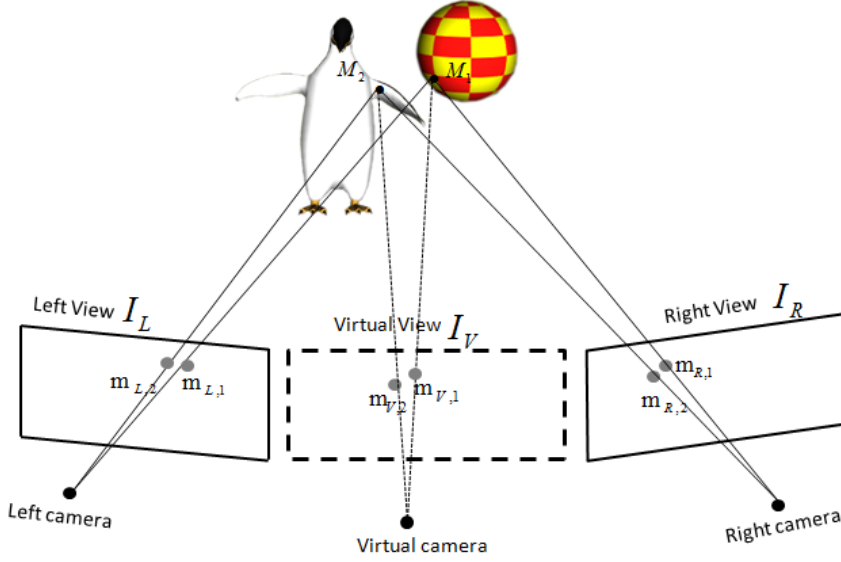
Figure 3.4: Forward warping of a pixel from left or right view on to virtual view.

3.4.1 Merging

So far, virtual view generation from a single V+D has been considered. When MVD information is available, then the rendering of intermediate view is as follows: two reference views are firstly warped to a desired view position. In the next step, the two warped views are combined into one view. This step minimizes the number of artifacts that result from a single view. The possible means to combine the two views information involves prioritizing the view information, which either depends on the virtual camera distance or density of the warped pixels [CKS06]. Another solution involves blending the two views pixel by using bilinear interpolation based on distance to rendered view as follows [MFY+09]:

$$I_V(u_V, v_V) = wI_L(u_L, v_L) + (1 - w)I_L(u_R, v_R),$$ (3.12)

$$w = \frac{|\mathbf{t}_V - \mathbf{t}_R|}{|\mathbf{t}_V - \mathbf{t}_L| + |\mathbf{t}_V - \mathbf{t}_R|},$$ (3.13)

where $\mathbf{t}$ is the translational vector of a camera and the subscripts L,R,V represents the left, right and virtual cameras respectively and w is the linear weight, that depends on the distance between the reference and virtual view translation vector.

3.4.2 Rendering artifacts

In general, virtual views are produced by applying warping and merging. The virtual views suffer from several rendering artifacts that result in a reduced visual qual-

(a) Empty cracks   (b) Translucent cracks

(c) Empty regions   (d) Corona
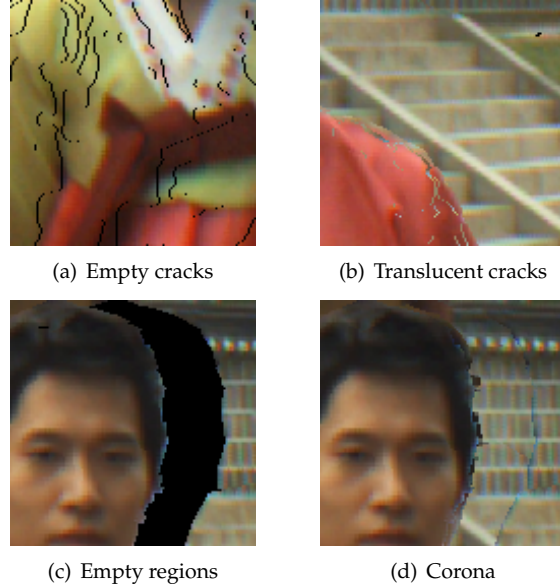
Figure 3.5: Rendering artifacts.

ity. These artifacts become less severe when the rendered views are closer to the reference view or views rendered from MVD. The common rendering artifacts are now listed see Fig. 3.5.

The undefined pixels in the rendered images are called *empty cracks*. These are the consequence of assigning each warped pixel to the nearest integer coordinates in the virtual view (see Fig. 3.5(a)). In the majority of these DIBR methods, these cracks are filled by applying a median filter.

Another similar type of artifact involves *translucent cracks*, which occur for the same reason as the *empty cracks* but contain background colour information at that position due to occlusions. (see Fig. 3.5(b)).

The next kind of artifacts involves *empty regions* or *holes*. *Empty regions* are disoccluded areas as a consequence of the warping i.e. scene parts, which are occluded in the original view become visible in the virtual view. The disocclusion occurs at the object boundaries, where the depth discontinuities are larger (See Fig. 3.5(c)). These artifacts still appear in the virtual views rendered from MVD information because information is not available in both reference views. Reducing the disocclusions is also called, hole filling or disocclusion filling. Two different filling strategies, namely interpolation and inpainting, are described in this thesis, one in the next section and the other in the next chapter.

Smearing of edges is a consequence of smooth texture edges or depth-texture misalignments. The edges, positions in the texture image are not aligned with the depth information due to not good depth data and as a result, foreground edge pixels are warped onto background at disocclusion areas. They appear as a shadow of the
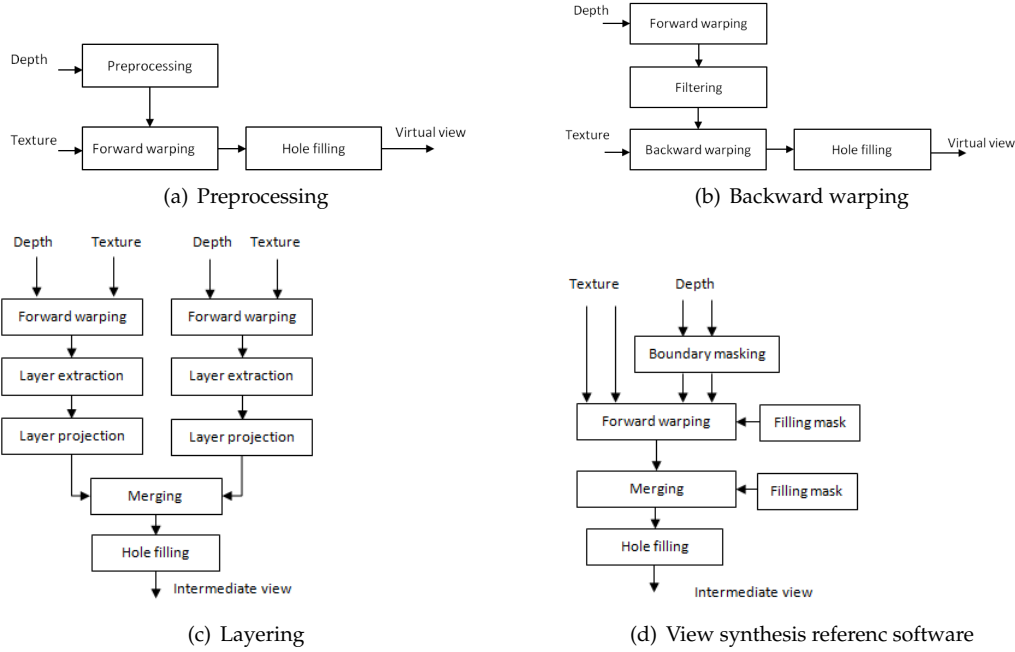
Figure 3.6: Existing rendering approaches.

foreground with few pixels wide and are called coronal-like effects (see Fig. 3.5(d)).

Unnatural contours are the consequence of sharp edges between the foreground and background regions where the colours of the pixels at the edge have not been blended.

3.5 Existing approaches

This section provides a brief overview of existing solutions that reduces the depth image based rendering artifacts.

3.5.1 Preprocessing of depth map

The rendering artifacts can be reduced by preprocessing of the depth map see Fig. 3.6(a). Tam et al. proposed a depth preprocessing approach to smooth the depth images using well known smoothing filters, namely the Gaussian filter and median filters. The Gaussian filter is usually used as a smoothing filter. The low pass filter reduces the sharp depth continuity in the depth images, which implies reduced artifacts at the object boundaries [Feh04, TAZ$^+$04].

Symmetrical smoothing filters result in distorted boundaries in the rendered views

such as vertical lines becoming curves, so Zhang et al. introduced asymmetrical filter smoothing. This smoothing applies only to the vertical direction [ZT05]. The previous methods apply smoothing to entire images, thus increasing the blur and to reduce this effect edge dependent filtering is introduced in [ZwPSxZy07, DTPP07]. Park et al. further proposed an adaptive filter that consists of both discontinuity-preserving smoothing and gradient direction based smoothing [PJO+09].

3.5.2   Backward warping

Backward warping is another approach to reducing the artifacts. In this approach, first the depth map is firstly warped to the virtual view position using the geometrical 3D warping and this is followed by filtering the depth map to fill the missing regions in the depth map. In the next step, the virtual view pixels are obtained by finding the corresponding pixels in the original view, using the warped depth map. Finally, a hole filling step fills the missing regions in the rendered views, see the block diagram for the backward warping Fig. 3.6(b). This process avoids the crack like rendering artifacts but it adds additional complexity [MSD+08].

3.5.3   Layering approach

Zitnick et al. introduced the layering approach to obtain the rendered views by dividing the input data into layers based on reliability. Firstly, the layers from each view are warped separately and thereafter merged into one view. Finally, post processing step is applied in order to reduce the rendering artifacts [ZKU+04]. The block diagram of this approach is shown in Fig. 3.6(c). This approach was extended by dividing the layers into a main layer, background layer and a boundary layer [MSD+08]. This approach was further improved using constraints on the layers, selections [SHKO11]. However artifacts are still present in the rendered views.

3.5.4   View synthesis reference software

MPEG has developed View Synthesis Reference Software (VSRS) to generate virtual views using MVD data. The software is comprised of many strategies to reduce the artifacts. It takes the two reference views and two depth maps to produce the virtual views at the required position. There are two modes in the rendering process, namely the general and 1D mode. In the general mode, the reference view pixels are mapped onto target using pixel to pixel mapping by using backward warping and in addition inpainting, is used for hole filling. In 1D mode, the input views are up-sampled so as to obtain fractional precision accuracy. Two input views are firstly warped to virtual viewpoints and then the blending the two views with different choices. Finally, the remaining holes are filled with background propagation, a block diagram of VSRS 1D mode is shown in Fig. 3.6(d). The detailed description of VSRS refers to [vsr10]. Further, additional strategies including reliability map creation and similar enhancements are used in the VSRS 1D fast mode, a detailed description of
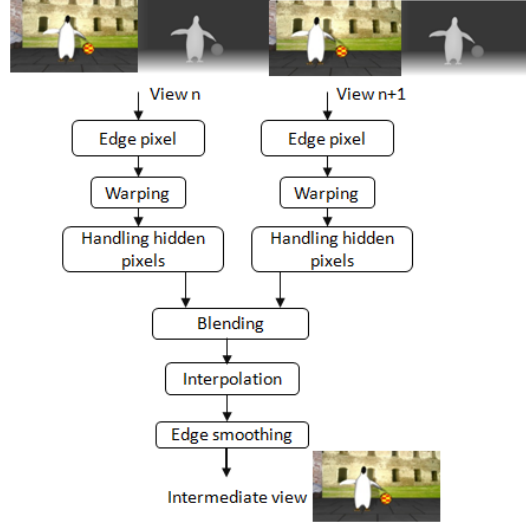
Figure 3.7: Proposed edge-aided rendering method flow.

which can be found in [Hhi12]. Although, these methods show significant improvements, enhancements can still be achieved by means of direct approaches without specific processing steps in order to reduce artifacts.

3.6   Proposed edge-aided rendering approach

In the previous sections, basic concepts of 3D warping, view merging, rendering artifacts and existing approaches were presented. VSRS is the best performing method among the existing approaches due to integrating specific solutions for each artifact. However, the rendered views still exhibit artifacts.

In this section, the problems of rendering artifacts are addressed using a straight-forward approach edge-aided rendering method for 3DV data. The edge-aided rendering method relies on the fundamental principle of 3D warping, that is the observation of each pixel position in reference and virtual views. Post-processing is one of the tools used to fill the empty regions in the virtual views by the interpolation and propagation of background information. Instead of applying a post-processing step to recover the missing information, the edge-aided rendering method reconstructs the missing and required pixel information through interpolation of the actual projected samples.

The method consists of the following steps: Firstly, specific edge-pixels are introduced at depth discontinuities for the sake of filling empty regions. Then the original view pixels are warped into virtual view positions and the actual values (floating positions) are retained without assigning nearest integer pixel positions as other DIBR methods. Next, the method applies the merging step to two warped pixels followed
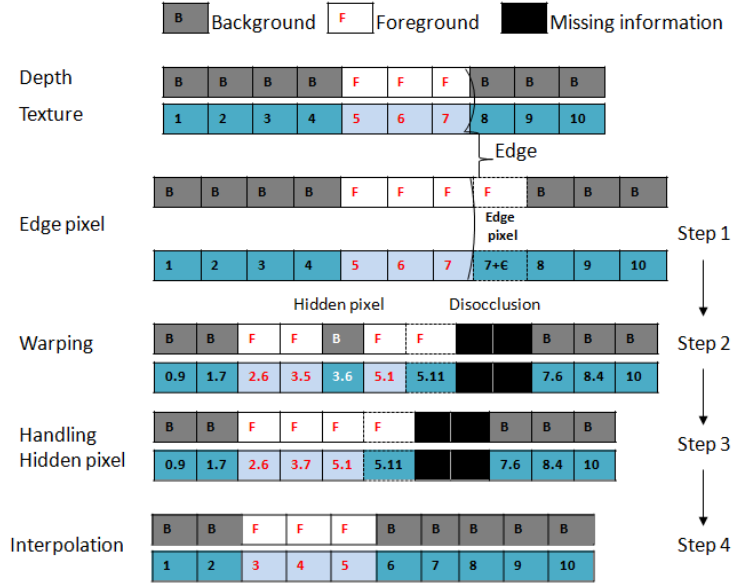
Figure 3.8: Edge-aided rendering method description (The numbers in the texture box indicates assumed pixel positions and the pixels positions changes with depth after warping).

by linear interpolation in order to obtain the pixel grid of the virtual view. Finally, edge smoothing is applied to cancel out the pixilation at boundaries.

Fig. 3.7 shows the block diagram of the edge-aided rendering method. The virtual view pixel has displacements in the horizontal and vertical directions in the general warping case. Because of the assumption of horizontal displacement (1D parallel camera arrangement) of the camera, the virtual view pixel is displaced along the horizontal direction. The proposed method thereby turns into a line-by-line algorithm with the following steps as depicted in Fig. 3.8.

3.6.1 Edge-pixels

The so called *edge-pixel* has position and colour value, similar to a general pixel, but the position is a rational number i.e. the position of the edge is horizontally shifted by a small amount. The *edge-pixel*, which is introduced at the depth discontinuities, depends on the warping direction, for example: a right warped view has disocclusions at the right side of the foreground objects, so the edge-pixel $(7 + \epsilon)$ is introduced at the right side of the foreground (see step1 in Fig. 3.8 at pixel position 7). The *edge-pixel* contains foreground depth and background colour, so that it follows the foreground (see step2 in Fig. 3.8 *edge-pixel* position after warping at position 5.11). In general, images consist of mixed colours of both foreground and background at object boundaries over transition regions. The transition range is about 1 or 2 pixels wide due to the averaging of colours in the camera pixel sensor. Hence, background

(a) Merging1   (b) Merging2
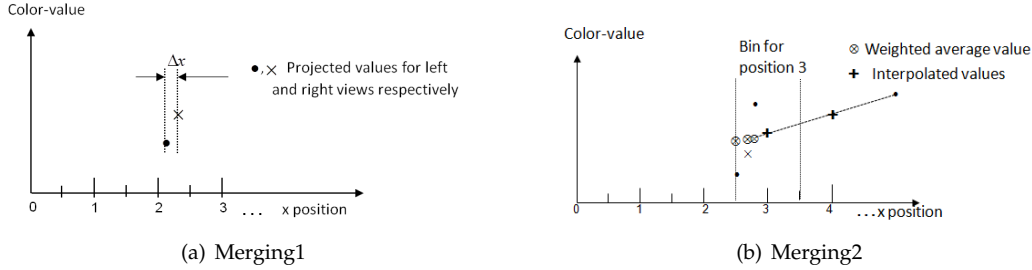
Figure 3.9: Different steps in merging.

colour values are selected outside the transition area to avoid smearing colours over a larger area.

3.6.2 Handling Hidden pixels

Once the original view pixels are warped to the new view point, there are a few pixels that become occluded and appear through the cracks in the virtual views. They are identified after warping by taking the depth differences between the neighbouring pixels. After identifying the *hidden pixel* (see the step2 in Fig. 3.8 at pixel position 3.6) sample, they are simply ignored in further processing steps, see step3 in Fig. 3.8.

3.6.3 Blending two views

When the two views are warped from the reference views into the virtual view point, the next step is to combine the views information, the so called merging or blending of views. In the edge-aided rendering approach, the *merging* step (see Fig. 3.7) combines the projected pixels from the two views by applying a weighted averaging of subsets: Firstly, sort the two projected view coordinates in ascending order. If either of the two projected pixel coordinates are close enough (less than a threshold), then select the pixel, which is closer to the camera by observing the depth information. In this step, the horizontal differences $\Delta x = x_i - x_{i-1}$ between two of the projected coordinates are calculated and compared with the threshold d_0. This step is applied irrespective of the pixels origin information see Fig. 3.9(a). After the previous step, the total image width is divided into one-pixel-wide bins, i.e. the virtual view pixel position $x \pm 0.5$. A weighted average $\otimes$ (See Fig. 3.9(b)) is computed over all pixels projected in each of these bins, where the weight is based on the distance to each pixel's original image. All pixels within each bin are assigned this averaged value $\otimes$ before interpolation. After interpolation the colour value of the virtual view grid becomes $+$ (see Fig. 3.9(b)).

The causes in relation to the rendering artifacts and the solutions of the proposed approach and VSRS are given in Table 3.1.

Table 3.1: Comparison of rendering artifacts solutions

Artifacts	Cause	Remedy	
		VSRS	Proposed Method
Empty cracks	Integer round offs of projected coordinates	Median filtering	Interpolation
Translucent cracks	Background pixels seep through cracks	Constraints on pixel mapping order	Removal of hidden pixel
Empty regions	Disocclusions	Background propagation	Introduction of edge pixels, interpolation
Smearing of edges	Smooth texture edges or depth texture misalignment	Removal of unreliable pixels before warping	Selection of "pure" background and foreground colours for edge pixels
Unnatural contours	Pixilation at edges due to new border between foreground and background	Splatting along edges	Upscale filtering along edges

Chapter 4

Disocclusion filling using inpainting

Disocclusions are one of the severe problems in virtual views when reconstructed from DIBR. In the previous chapter, they are addressed using background colour interpolation. The background interpolation offers acceptable results for small baseline (i.e. distance between cameras) setups and objects at different distances from the cameras. While in other cases, for example, FTV, involves large baseline setup, which results in bigger disocclusion areas. Then nearest colour interpolation does not yield better results.

The disocclusions occur at object boundaries by the properties of warping. The disocclusion problem can be considered as missing information in the texture and is recovered by using inpainting techniques [TLD07]. Inpainting is a process of filling the missing information using its available neighbourhood information. In this chapter, we briefly describe the classification of inpainting methods and existing approaches for disocclusion (also known as hole) filling. In the end, the proposed depth based inpainting is introduced in order to handle disocclusion areas.

4.1   Overview of inpainting methods

In the past, inpainting referred to simply painting defects such as cracks or missing areas manually, but in the texture synthesis context, inpainting is a process of recovering missing information from the available information around the holes. Inpainting methods are broadly classified into texture inpainting, structure inpainting and hybrid inpainting.

Texture inpainting methods fill the missing regions by replicating the repetitive patterns in an image. These methods use small amounts of textures to model Markov Random Field (MRF) and generate the new textures by sampling and copying textures from neighbouring regions [EL99]. Texture inpainting methods achieve better results for homogeneous texture images but encounter difficulties in filling real world scenes. The real images consists of composed textures at objects boundaries due to mixing of different textures and linear structures.

To solve these problems, structural inpainting methods propagate linear structures (aka constant colour values) into missing regions through a diffusion process [BSCB00]. Structural inpainting methods are often used to solve the partial differential equations (PDE) iteratively in order to fill the missing information. The disadvantage associated with these methods is that it is not able to recover the large textured regions due to diffusion process and is not aware of edges, information, the result of which is a blurring in the inpainted image. Thereafter, in total the variation approach, anisotropic diffusion is applied, which depends on the isotropic strength [CS00]. These worked well for smaller missing regions but is unable to deal with the broken edges and texture patterns. Thus total variation model is extended to the curvature driven diffusion (CDD) model, which improved the inpainting process by using the curvature of the isophote together with its strength [CS01]. However, blurring artifacts still remain in larger missing regions although this works well for smaller missing regions.

Another kind of inpainting method includes exemplar-based methods, which combine the advantages of structural and textural inpainting [CPT04]. It uses the priorities of linear structures within a specified area and fills the regions with similar data from the neighbourhood. It can be called hybrid inpainting since it combines the benefits of both textural and structural inpainting.

4.2   Exemplar-based inpainting

Exemplar-based methods have proved as efficient inpainting algorithms for filling large missing regions since they use the advantages of structural, textural inpainting and filling order i.e. in which missing pixels are filled first [CPT04]. Actually the first effort of an exemplar method was introduced in [Har01]. In this approach, filling order of a pixel in a missing region is extracted from the amount of textured detail in its neighbouring region. The order in the Criminisi et al. method is derived from linear structures in its neighbourhood regions. The block diagram of the Criminisi et al.
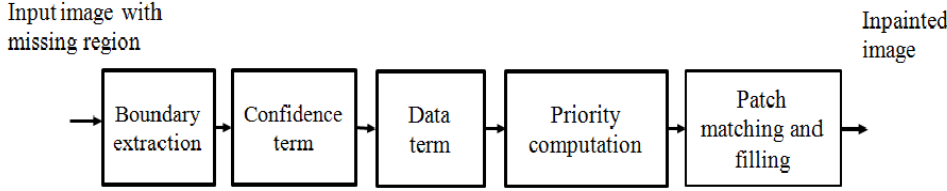
Figure 4.1: Exemplar-based inpainting method.

method is shown in Fig. 4.1. The Criminisi et al. method consists of following steps: the filling order is firstly computed for missing pixels using a priority term, which is related by the isophote values and the amount of non missing data in the neighbourhood. Next, the selected missing region is filled by assigning the best match texture from neighbourhood. The output quality of the inpainted image depends on the order in which the filling is performed. The basic steps of this method are presented in the following steps.

Suppose an input image I having a missing region Ω, also known as a "hole" to be inpainted. The remaining portion of the image, apart from the missing region, is defined as the source region $\Phi = I - \Omega$, which is used to find a similar texture for the missing region. The boundary of the missing region is defined as the hole boundary $\delta\Omega$ (see Fig. 4.2). In the first step, priorities are calculated along the boundary. The priority gives the information about, which patch (small portion of missing region), should be filled first. Let us consider a patch $\Psi_{\mathbf{p}}$ centered at a pixel $\mathbf{p}$ and priority term $P(\mathbf{p})$ is defined as:

$$P(\mathbf{p}) = C(\mathbf{p}) \cdot D(\mathbf{p}),　\quad (4.1)$$

where $C(\mathbf{p})$ is the confidence term, which gives the percentage of non-missing pixels in a patch and $D(\mathbf{p})$ is data term, which gives the information about the isophote in the neighbourhood and these are defined as:

$$C(\mathbf{p}) = \frac{1}{|\Psi_{\mathbf{p}}|} \sum_{\mathbf{q} \in \Psi_{\mathbf{p}} \cap \Omega} C(\mathbf{q}),　\quad (4.2)$$

$$D(\mathbf{p}) = \frac{\langle \nabla^{\perp} \cdot n_{\mathbf{p}} \rangle}{\alpha},　\quad (4.3)$$

where $|\Psi_{\mathbf{p}}|$ is the area of $\Psi_{\mathbf{p}}$ (in number of pixels), α is normalization factor. $n_{\mathbf{p}}$ is a unit vector orthogonal to $\delta\Omega$ at a point $\mathbf{p}$, and $\nabla^{\perp} = (-\partial_y, \partial_x)$ is the direction of the isophote. Actually $C(\mathbf{q})$ is set to the initialization $C(\mathbf{q}) = 0$ for hole pixel in Ω and $C(\mathbf{q}) = 1$ for source region pixels.

From the priorities along the hole boundary, the highest priority patch $\Psi_{\hat{\mathbf{p}}}$ centred at $\hat{\mathbf{p}}$ is selected to be filled first. Then patch matching is applied in order to fill the missing region in the selected patch. The patch matching equation is given as:

(a) Warped texture view   (b) Notation
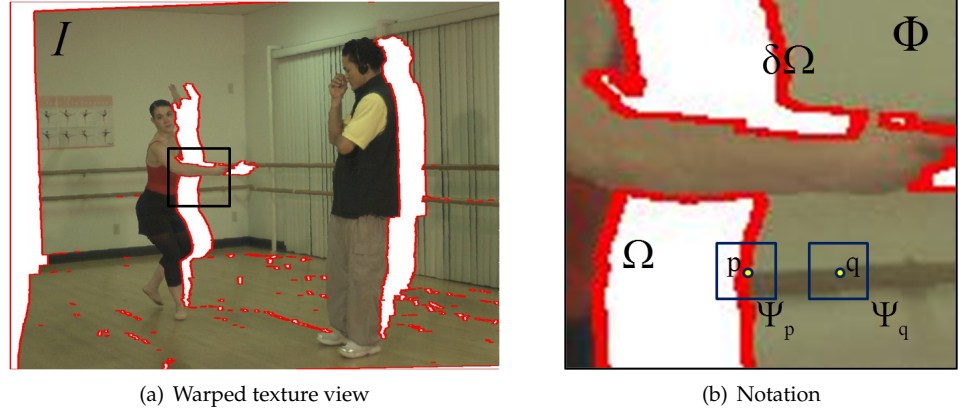
Figure 4.2: Exemplar based inpainting notation (white colour represents disoccluded areas).

$$\Psi_{\hat{\mathbf{q}}} = \arg \min_{\Psi_{\mathbf{q}} \in \Phi} \left\{ \text{SSD}(\Psi_{\hat{\mathbf{p}}}, \Psi_{\mathbf{q}}) \right\}, \tag{4.4}$$

where SSD is the sum of squared difference between the two patches. Once the best patch $\Psi_{\hat{\mathbf{q}}}$ is found, the missing pixels information in the patch is filled from the corresponding pixel inside $\Psi_{\hat{\mathbf{q}}}$. After filling the patch, the confidence term is updated according to:

$$C(\mathbf{q}) = C(\hat{\mathbf{p}}), \forall \mathbf{q} \in \Psi_{\hat{\mathbf{p}}} \cap \Omega. \tag{4.5}$$

The process continues until the empty regions are filled in. The exemplar method performs well for texture synthesis, however, for view rendering context it is unable to distinguish foreground and background parts around the disocclusions and as a consequence, the foreground parts might propagate into missing regions.

4.3   Existing approaches

There are several derivatives of the exemplar methods which do exist so far and the following approaches also come under exemplar methods. In the view rendering context the disocclusions are assumed as holes in textures synthesis.

4.3.1   Depth-aided inpainting

Daribo et al. addressed the disocclusion problem by adopting the Criminisi approach and added depth map as additional information in the filling process [DPP10]. They extended the Criminisi et al. approach by using depth regularity term in the priority term by multiplying with $C(\mathbf{p})$ and $D(\mathbf{p})$ terms, $P(\mathbf{p}) = C(\mathbf{p}) \cdot D(\mathbf{p}) \cdot L(\mathbf{p})$ and is given as:

$$L\left(\mathbf{p}\right) = \frac{|Z_{\mathbf{p}}|}{|Z_{\mathbf{p}}| + \sum_{\mathbf{q} \in, Z_{\mathbf{p}} \cap \Phi}\left(Z_{\mathbf{p}}(q) - \overline{Z_{\mathbf{p}}}\right)^2}, \tag{4.6}$$

where $L\left(\mathbf{p}\right)$ is depth regularity term and is approximated as the inverse variance of the depth patch $Z_{\mathbf{p}}$ centred at $\mathbf{p}$. They state that $L\left(\mathbf{p}\right)$ favours the inpainting process order from the background. Additionally, the patch matching step is modified by taking into account the depth together with texture with a parameter. The parameter gives the importance to the depth while searching for similar structures. Though added depth information improved the inpainting process in giving priority to the background the same problem exists from the exemplar approach, which is that foreground parts copied in to the missing regions.

4.3.2   Depth-based inpainting using structure tensor

Gautier et al. also followed the Criminisi et al. approach by extending the exemplar approach by using a depth based priority and a depth based search in the filling process [GMG11]. They used the structure tensor J as a data term instead of using structure propagation along the isophote directions.

$$J = \sum_{l=R,G,B,Z} \nabla I_l \nabla I_l^T, \tag{4.7}$$

$$D(\mathbf{p}) = a + (1 - a)\exp\left(\frac{-C_1}{(\lambda_1 - \lambda_2)^2}\right), \tag{4.8}$$

where ∇I_l is local spatial gradient over a 3x3 window. J is the 3D structure tensor and λ_1, λ_2 are the eigen values of J, which gives amount of structure variation, C_1 is a constant positive value and $a \in [0, 1]$. The tensor term aims at propagating the strongest structure into the missing region. In addition to previous methods, they used a one sided priority i.e. propagating the structure from the background. Further in the patch matching process, instead of using the best patch, they considered the average of a few best patches based on the idea of [WSI07].

4.4   Proposed depth-included curvature inpainting

The existing depth-based inpainting methods perform better than the conventional exemplar-based method based on their knowledge of both the background and foreground parts in the filling process. There are still problems associated with the existing methods. For [GMG11] there are holes in the image after inpainting and for [DPP10] there are holes filled with foreground objects. Moreover, these two methods relied on the original depth map at the rendered view position. In general, the original depth map at the rendered view position is not available.

Disocclusions occur at depth discontinuities and they are belongs to background information hence, filling from the background favours inpainting process. The

proposed depth-included curvature inpainting has the following steps: firstly, the background-side boundary is computed to start filling from the background direction. In the next step, depth-included curvature data term is introduced to favour the isophote curvature. Finally, the source region is divided into foreground and background regions to help the inpainting process.

4.4.1 Depth-guided background priority

By definition, filling starts from the background side in order to avoid leaking of foreground parts into disoccluded areas. Hence, the depth-guided background priority is given as follows: firstly, one sided background boundary is computed from the knowledge of warping i.e. a view warped to the right direction have holes on the right side of the background and vice versa. After this, the depth-guided boundary of a hole is obtained by applying a threshold in such a way background depth regions should fill first, because disocclusions belong to background objects in a scene at different depth levels (see Fig. 4.3). The depth-guided boundary is given as:

$$\delta\Omega_1 = DM * H, \tag{4.9}$$

$$\delta\Omega_1^{'} = \delta\Omega_1(\mathbf{q})|_{\mathbf{q}\in\delta\Omega_1\cap(Z(\mathbf{q})<M\cdot max(Z))}, \tag{4.10}$$

where DM is a disocclusion map, $*$ is a convolution operator, $\delta\Omega_1^{'}$ is the depth-guided boundary, Z is the depth map, $Z(\mathbf{q})$ is a pixel depth value at pixel $\mathbf{q}$ and H is the convolution kernel defined according to the warping direction as follows:

$$H = \begin{cases} \begin{bmatrix} 0 & 0 & 0 \\ 1 & -8 & 0 \\ 0 & 0 & 0 \end{bmatrix} & \text{if left warped view;} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & -8 & 1 \\ 0 & 0 & 0 \end{bmatrix} & \text{if right warped view.} \end{cases} \tag{4.11}$$

In some situations, some holes are unfilled due to threshold M applied to one depth level. Then the remaining holes are filled with one sided boundary extraction using (4.9). Also in few cases, the hole areas at boundary of images are irrespective of warping direction then total boundary is used to perform the remaining process.

4.4.2 Depth-included curvature data term

In addition to the CDD depth data term proposed in [LWXD12] the depth information is included to consider depth curvature in to account along with texture curvature. While the existing approaches favours the isophote directions and strong structure gradients in order to propagate the linear and strong structures, the depth-included curvature term aids the inpainting process by considering the curvature of

(a) Warped texture

(b) Disocclusion boundary

(c) One sided disocclusion boundary

(d) Depth-guided boundary
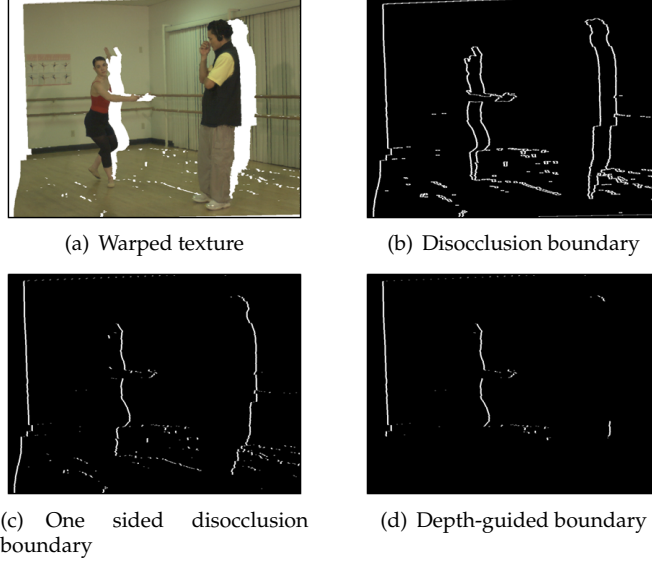
Figure 4.3: Illustration of disocclusion region boundary extraction.

the isophote. The motivation in relation to the CDD model is filling the discontinuity parts by using the strength and curvature of the isophote. Hence the data term is given by:

$$D\left(\mathbf{p}\right) = \left| \nabla \cdot \left(\frac{|k_{\mathbf{p}}|}{|\nabla I_{\mathbf{p}}|} \nabla I_{\mathbf{p}} \right) \right|, \tag{4.12}$$

$$k_{\mathbf{p}} = \nabla \cdot \left(\frac{\nabla I_{\mathbf{p}}}{|\nabla I_{\mathbf{p}}|} \right), \tag{4.13}$$

where $k_{\mathbf{p}}$ is isophote curvature passing through a pixel $\mathbf{p}$, $\nabla \cdot$ is the divergence at $\mathbf{p}$, Once the depth-guided boundary and depth-included curvature data term are obtained then the priority term is computed using (4.1).

4.4.3 Depth-based patch matching

Patch matching is another important step in the exemplar-based inpainting method to find the similar texture from the source region. While the previous depth-based exemplar approaches search a similar patch in both texture and depth, the proposed approach applies the search in a specific source region by dividing the source region into foreground and background regions. The patch-matching step in the depth-included curvature inpainting method is an improvement to the method of [DPP10] and [GMG11]. The idea of segmenting the background region has been previously employed by [AK12] using patch averages. However, the depth-based patch matching classifies source region so as to enhance the patch-matching step. Let us con-

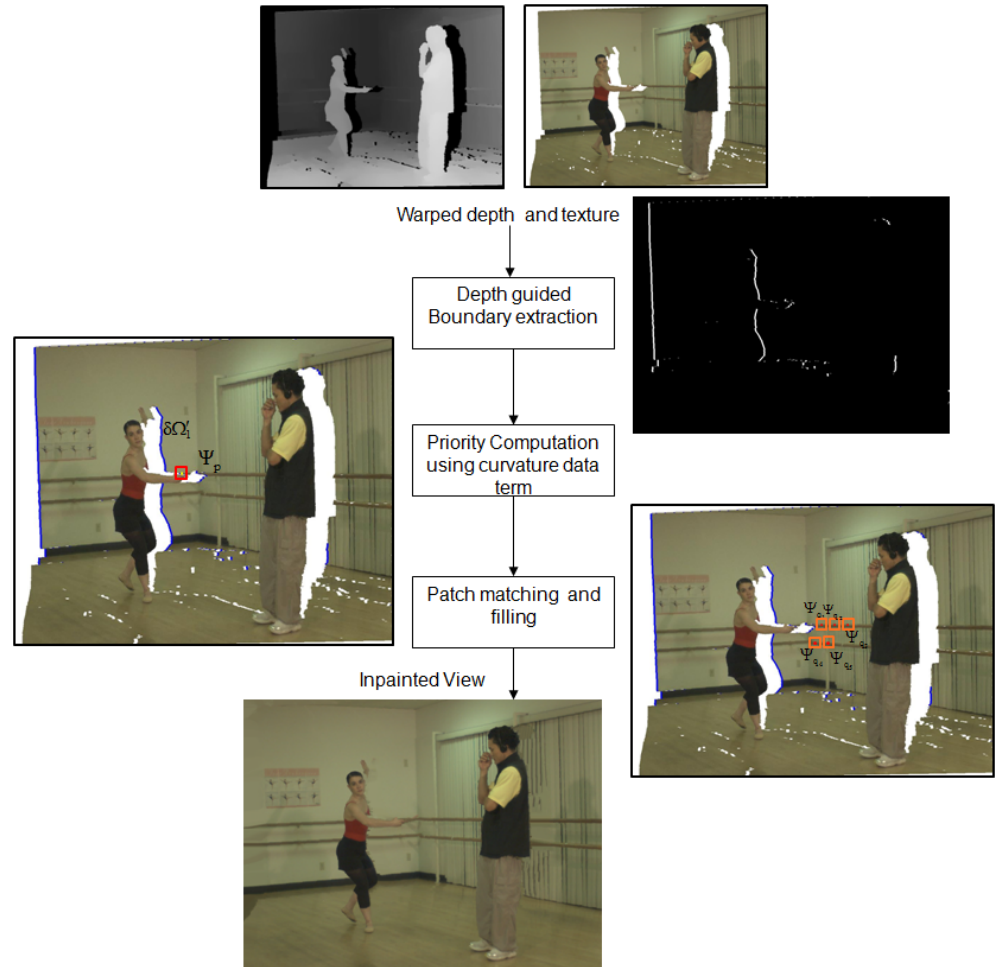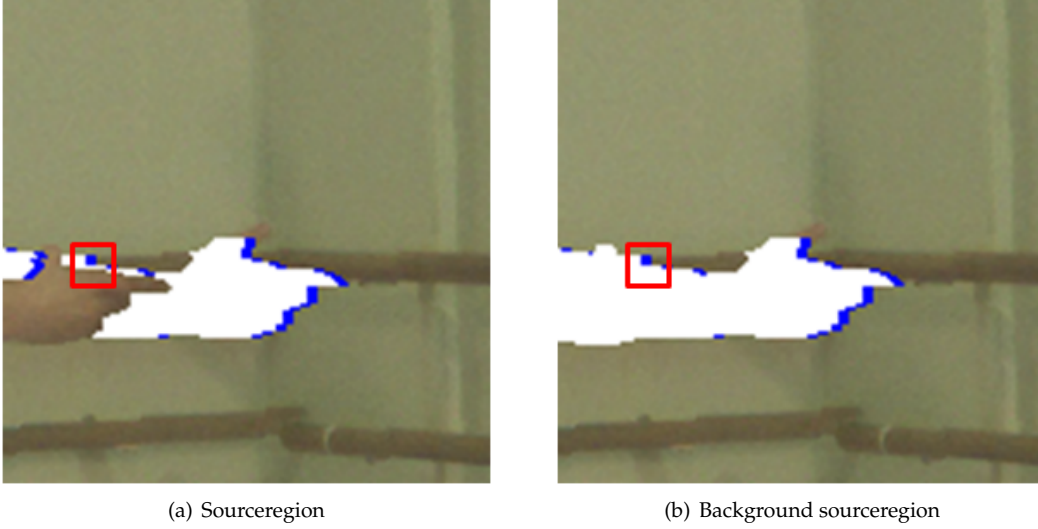Figure 4.4: Proposed depth-included curvature inpainting method.

(a) Sourceregion   (b) Background sourceregion

Figure 4.5: Source region selection for patch matching.

sider Φ is a known source region, which contains both foreground and background regions. The background source region Φ_b is obtained by subtracting the foreground source region Φ_f from the total source region Φ using a depth threshold Z_c:

$$\Phi_b = \Phi - \Phi_f. \tag{4.14}$$

The Z_c takes two different values depends on the variance of pixels in a depth patch. Suppose, the variance of non missing pixels in a depth patch is greater than the threshold γ average of the patch is used, because the patch might be placed on unwanted foreground parts (see Fig. 4.5(a)). In other case, the patch may lie on a homogenous region or gradual transition regions; the farthest value of the depth patch is used as the depth threshold to obtain similar textures according to the depth level. Hence the depth threshold Z_c is defined as follows:

$$Z_c = \begin{cases} \overline{Z_{\hat{\mathbf{p}}}} & \text{if } var(Z_{\hat{\mathbf{p}}}(\mathbf{q})|_{\mathbf{q} \in \Psi_{\hat{\mathbf{p}}} \cap \Phi}) > \gamma; \\ \max(Z_{\hat{\mathbf{p}}}) & \text{otherwise.} \end{cases} \tag{4.15}$$

where $\Psi_{\hat{\mathbf{p}}}$ is the highest priority patch, $Z_{\hat{\mathbf{p}}}$ is the depth patch centered at $\hat{\mathbf{p}}$ and $\overline{Z_{\hat{\mathbf{p}}}}$ is the average value of the depth patch. $Z_{\hat{\mathbf{p}}}(\mathbf{q})$ is the depth value at pixel $\mathbf{q}$ and γ is the depth variance threshold.

After the highest priority patch $\Psi_{\hat{\mathbf{p}}}$ and depth-based source region Φ_b are obtained (see Fig. 4.5(b)), the search is for the best N number of patches within the specified source region. Warped depth information is used to aid the filling process and to simultaneously fill the disocclusion in both the texture and the depth map. Similar patches obtained from the patch matching are not equally reliable [WSI07].

Therefore, a weighted average of N patches is adopted for filling the missing pixels in the disocclusion. The weighted average removes the noise from the patches.

The next chapter provides the separate contributions of this thesis.

Chapter 5

Contributions

The previous chapters have described the methodology of edge-aided DIBR and depth-included curvature inpainting. In this chapter, the novelty and evaluation criteria of the three papers of this thesis are explained. The contributions are:

- Paper I: Edge-preserving depth-image-based rendering method.

- Paper II: Edge-aided virtual view rendering for multiview video plus depth.

- Paper III: Disocclusion handling using depth-based inpainting.

5.1   Paper I: Edge-preserving depth-image-based rendering method.

The first paper introduced the edge-preserving rendering method for generating virtual views using the V+D data and compared it with the state-of-the-art method.

5.1.1   Novelty

The novelties of the method lies in the following ways: firstly, introducing the edge pixels at the depth discontinuities for the sake of background colour interpolation. Next, it applies arbitrary 1D interpolation methods to reconstruct a virtual view grid. The method further applies the edge smoothing by low pass filtering. The proposed edge-preserving rendering method omits the specific processing steps for handling the rendering artifacts.

5.1.2   Results

The first investigated the causes of rendering artifacts in virtual views and a comparison between the state-of-the-art and proposed edge-preserving rendering approach solutions were performed. Furthermore the edge-preserving rendering method was evaluated by producing the virtual views and comparing them with the original views captured at those positions. Two different kinds of input data, computer generated and photographic content were used for evaluation in order to test different types of depth data. A computer generated image consists of "Penguin", which has sharp depth discontinuities and structured background. The photographic sequence "Poznan Street" has complex depth distributions and contains many edges and the later sequence "Lovebird1" has large depth discontinuities and complex textures [DGK+09, UBH+08]. The virtual views were compared with the original view using the objective metrics Luminance Peak Signal-to-Noise-Ratio (YPSNR) and Mean Structural Similarity Index (MSSIM). Even if both metrics measures the view quality, MSSIM only matched to perceived visual quality [WBSS04]. Along with objective measurements, a visual inspection also conducted. The objective results demonstrated that the proposed method performs better than the reference method VSRS for photographic content and achieved similar quality results for synthetic data. Moreover, visual inspection clearly showed the improvements over the reference method.

5.2   Paper II: Edge-aided virtual view rendering for multiview video plus depth.

Paper I addressed DIBR artifacts for virtual views using single V+D data. This paper extends the V+D setup to MVD and introduced a rendering method to generate

intermediate views.

5.2.1 Novelty

The novelty of this paper lies in the merging of two warped views information before the interpolation process. The merging applies weighting to the actual projected positions with respect to the distances of the original cameras.

5.2.2 Results

In addition to the previous objective evaluation, subjective experiments were also conducted to assess the rendered view quality. Objective results provide differences in the rendered views whereas subjective tests reflects the end user preferences. Together with the previous photographic input data another sequence "Poznan hall" was used in the evaluation. The selection of subjective quality assessment method is still challenging for evaluating the rendering method. From the description of the subjective methods standards ITU-R Rec BT.500.11 and ITU-T Rec P.910, pair comparison subject quality assessment method was employed to evaluate the view rendering methods [ITU97, ITU93]. Pair comparison offers reliable quality ratings, finds the small differences and has no requirement of reference images to assess rendered images. The obtained subjects preferences from the test were converted to a quality score using the Bradley-Terry's model [Har01]. Again the proposed method showed improvements over the reference methods [vsr10, Hhi12] using objective metrics. Although subjective results provided different opinion scores for various input data, overall quality demonstrates that no significant differences exist between edge-aided rendering method and state-of-the-art methods. Nonetheless, the edge aided rendering method offers a straightforward approach to obtaining the intermediate views without addressing each individual artifact in the processing.

5.3 Paper III: Disocclusion handling using depth-based inpainting.

The motivation behind the last paper included in this thesis was based on the texture synthesis methods. This paper addressed the disocclusion regions in the warped views using depth- included curvature inpainting.

5.3.1 Novelty

The novelties in this paper: Firstly, the background boundaries of disocclusion are obtained by using depth in order to start filling from the background. Next, depth information is included in the curvature data term to give importance to the depth curvature together with the texture image. Finally, the source region is limited to be

approximately at the depth as the boundary of the disocclusion, in order to favour a background patch.

5.3.2 Results

In similar manner to Paper I, objective quality assessment and visual inspection were conducted with different kinds of test sequences. The proposed depth-included curvature inpainting was compared with three reference inpainting methods. The reference method [CPT04] does not use the depth information during the inpainting process, while the two other reference methods [DPP10, GMG11] use the original depth information. Three different MVD test sequences "Ballet", "Break dancers" and "Lovebird1" were used for evaluation [ZKU+04]. Each sequence has different depth and texture characteristics that make them suitable for evaluating inpainting methods. The "Ballet" sequence has large depth discontinuities at two different depth ranges, which results in large disocclusion areas at different depth levels. The "Break dancers" sequence has a large number of objects. And finally the "Lovebird1" sequence has a complex texture and more structured background, with larger depth discontinuities.

The objective results and visual inspection consistently demonstrated that the proposed approach performs better than the three reference methods. Although the proposed depth-included curvature inpainting used warped depth map, it favours the inpainting process over reference approaches.

Chapter 6

Conclusions

A concise summary of the contributions is presented in the previous chapter, which consists of the novelty of each paper, evaluation criteria and important results. This final chapter provides the overall aim and conclusions of the three contributions in respective of view rendering. Future work is then presented with possible extensions and applications of the current approaches.

6.1 Overview

The presented thesis work aimed at achieving a better 3D video quality experience. In this context, high quality view rendering methods are essential. A new algorithm edge-aided DIBR for view rendering and a hole filling strategy were proposed. The edge-aided DIBR offers a straightforward approach, while other state-of-the-art approaches apply specific tools for minimizing the rendering artifacts in virtual views. The edge-aided rendering method introduces the edge-pixels and interpolation to resample the virtual views grid and to interpolate the background colours in to disoccluded regions. Edge-aided rendering method performs better than the state-of-the art methods objectively and shows similar results subjectively.

Moreover the depth-included curvature inpainting a new hole filling strategy was proposed for bigger disocclusions. It effectively uses the warped depth and texture information to fill the disocclusion regions. The edge-aided rendering approach and disocclusion handling approach can be employed for rendering new perspectives views from different kinds of camera setups in order to generate high quality content for 3D displays. Thus the use of the proposed approaches satisfies the aim of the research with regards to providing a better 3D video experience on stereo and multiview auto stereoscopic displays.

6.2 Outcome

The following goals were defined in this thesis work in order to corroborate the outcome of the presented research. The goals were addressed in the contributions. A concise summary about the outcome with regards goals are presented here:

Goal I: Examine the problems of DIBR method.

The rendered virtual view images suffer from several artifacts, which are due to various reasons. The problems of DIBR were examined and their causes were listed and further compared with the remedies of the proposed and the existed state-of-the art approach.

Goal II: Propose an alternative DIBR method that reduces the rendering artifacts utilizing the available information for texture and depth data.

An alternative DIBR method was proposed to address the rendering artifacts. The method allows producing virtual views using texture and depth data. It utilizes the actual projected information to reconstruct the virtual views. Further, the proposed method applies a merging step to combine the rendered views. Specifically, it uses the introduced edge-pixels to interpolate background colours into missing regions. The proposed method fully avoids the post processing for reducing the artifacts. Objective evaluations of the proposed method showed increased quality over the state-of-the-art methods. Thus the quality virtual views can be achieved with

straightforward approaches. Therefore it achieved the Goal II by introducing the new rendering method, which addressed in Paper I and II.

Goal III: Investigate the visual quality experienced by users.

Subjective evaluations were conducted in order to investigate the visual quality in addition to the objective evaluations. The quality evaluations revealed that the proposed edge-aided rendering method achieves improved objective quality and similar subjective quality over state-of-the art methods.

Goal IV: Examine the methods to fill large disocclusions.

Different inpainting methods were examined in the context of disocclusion filling. Further, the problems with existing inpainting methods for disocclusion filling were investigated.

Goal V: Propose an inpainting method that addresses the disocclusion artifacts in the rendered views by utilizing the available information from the rendering process.

In Paper III, a depth-included curvature inpainting method was proposed to address the disocclusion problems in the rendered views. The proposed inpainting method utilizes the warped depth information and guides the inpainting process to propagate the true texture and structured data into missing regions. The warped depth information was utilized in determining the background boundary, curvature data term and background source region extraction in the inpainting process, whereas the reference depth-based inpainting methods relied on the true depth maps. Objective evaluations and visual inspection showed that the proposed method favours background propagation over the reference methods.

6.3   Impact

The edge-aided rendering method offers improved visual quality for stereoscopic and multiview autostereoscopic displays by producing virtual views. The research presented in this thesis allows for the rendering of high quality virtual views for 3D displays with a straightforward approach. The high quality virtual views from the proposed approach can have an impact on the content generation for 3D displays and thus can improve the commercial success of 3DTV. Moreover, the depth-included curvature inpainting method offers better handling of disocclusions. Further, the approaches are not only limited to 3DTV, but can also be utilized in FTV applications as well, since the proposed approach is able to handle large disocclusions.

6.4   Future work

A few possible future works are:

- Investigating temporal problems with the edge-aided rendering method, further defining a temporal filter to avoid temporal effects such as flickering.

- In general, inpainting methods require more processing time to fill a single frame. Thus, investing the steps in the depth-included curvature inpainting method and improving the method further to reduce the computational time.

- The current inpainting method fills the disocclusion on a frame wise basis, which leads to temporal artifacts when the filling is not consistent over time. Thus, defining a temporal inpainting by extending the proposed image inpainting approach to a temporal domain.

- More elaborate subjective tests on different 3D displays to evaluate rendering methods.

Bibliography

[3DI]   3D Immersive, Interactive Media Cluster.   [online] `www.3diim-cluster.eu`.

[3DY]   3D4YOU Content Generation and Delivary for 3D Television. [online] `www.3d4you.eu`.

[AK12]   I. Ahn and C. Kim. Depth-Based Disocclusion Filling for Virtual View Synthesis. In *2012 IEEE International Conference on Multimedia and Expo*, pages 109–114, 2012.

[BSCB00]   M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of ACM Conf. Comp. Graphics (SIGGRAPH)*, pages 417–424, 2000.

[CKS06]   E. Cooke, P. Kauff, and T. Sikora. A novel view creation approach for free viewpoint video. *Signal Processing; Image Communication*, 21:476–492, 2006.

[CPT04]   A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13:1200–1212, 2004.

[CS00]   T. Chan and J. Shen. Mathematical Models for Local Nontexture Inpaintings. CAM TR 00-11, March 2000.

[CS01]   T. F. Chan and J. Shen. Non-Texture Inpainting by Curvature-Driven Diffusions (CDD). *J. Visual Comm. Image Rep.*, 12:436–449, 2001.

[CV95]   J. E. Cutting and P. M. Vishton. Perceiving layout and knowing distances: the integration, relative potency and contextual use of different information about depth. *Handbook of perception and Cognition*, 5:69–117, 1995.

[DGK+09]   M. Domański, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner. Poznań multiview video test sequences and camera parameters. ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xianl, China, October 2009.

[DPP10]   I. Daribo and B. Pesquet-Popescu. Depth-aided image inpainting
             for novel view synthesis. In *Multimedia Signal Processing*, pages 167–
             170, 2010.

[DTPP07]   I. Daribo, C. Tillier, and B. Pesquet-Popescu. Distance Dependent
             Depth Filtering in 3D Warping for 3DTV. In *IEEE 9th Workshop on
             Multimedia Signal Processing*, pages 312–315, October 2007.

[EL99]   A. Efros and T. Leung. Texture Synthesis by Non-parametric Sam-
             pling. In *International Conference on Computer Vision*, pages 1033–
             1038, 1999.

[Feh04]   C. Fehn. Depth-image-based rendering (DIBR), compression, and
             transmission for a new approach on 3D-TV. *Proc. SPIE Stereoscopic
             Displays and Virtual Reality Systems XI*, pages 93–104, 2004.

[GMG11]   J. Gautier, O. L. Meur, and C. Guillemot. Depth-based image com-
             pletion for view synthesis. In *3DTV conference*, pages 1–4, 2011.

[Har01]   P. Harrison. A Non-Hierarchical Procedure for Re-Synthesis of
             Complex Textures. In *Winter School of Computer Graphics Conf. Proc.
             (WSCG)*, pages 190–197, 2001.

[Hhi12]   Test Model under Consideration for HEVC based 3D Video cod-
             ing. ISO/IEC JTC1/SC29/WG11 MPEG2011/N12559, February
             2012. San Jose, CA, USA.

[HZ04]   R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer
             Vision*. Cambridge Univeristy Press, 2nd edition, 2004.

[IL02]   H. Imaizumi and A. Luthra. *Three-Dimensional Television, Video, and
             Display Technologies.*, chapter Stereoscopic Video Compression Stan-
             dard "MPEG-2 Multi-view Profile", pages 167–181. In [JO02], 2002.

[ISM05]   W. A. IJsselsteijn, P. J. H. Seunties, and L. M. J. Meesters. *3D Video
             communication Algorithms, concepts and real-time systems in human cen-
             tered communication*, chapter Human Factors of 3D displays, pages
             219–233. In [SKS05], 2nd edition, 2005.

[ITU93]   ITU-R BT.500-11. Methodology for the subjective assessment of the
             quality of television pictures, November 1993.

[ITU97]   ITU-T p.910. Subjective video quality assessment methods for mul-
             timedia applications. ITU-T Study Group 12, 1997.

[JO02]   B. Javidi and F. Okano. *Three-Dimensional Television, Video, and Dis-
             play Technologies.* Springer Press Berlin, 2002.

[LR06]   Y. G. Lee and J. B. Ra. Image distortion correction for lenticula mis-
             alignment in three-dimensional lenticular displays. *Optical Engi-
             neering*, 45(1):017007–017007–9, 2006.

[LS01]   R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of Quantum Electronics*, 37:390–397, March 2001.

[LWXD12]   S. Li, R. Wang, J. Xie, and Y. Dong. Exemplar Image Inpainting by Means of Curvature-Driven Method. In *2012 International Conference on Computer Science and Electronics Engineering (ICCSEE)*, pages 326–329, 2012.

[MFW08]   Y. Morvan, D. Farin, and P. H. N. De With. System architecture for free-viewpoint video 3D-TV. *Consumer Electronics, IEEE Transactions*, 54(2):925–932, 2008.

[MFY$^+$09]   Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto. View generation with 3D warping using depth information for FTV. *Signal Processing; Image Communication*, 24(1-2):65–72, 2009.

[MSD$^+$08]   K. Muller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand. View synthesis for advanced 3D video systems. *EURASIP Journal on Image and Video processing*, 2008:1–11, 2008.

[MVi]   Multiview 3D displays. [online] `www.dimenco.eu`.

[Onu11]   L. Onural. 3D video technologies: An overview in research trends. In *SPIE*, 2011.

[Pas05]   S. Pastoor. *3D Video communication Algorithms, concepts and real-time systems in human centered communication*, chapter 3d displays, pages 235–260. In [SKS05], 2nd edition, 2005.

[Phi]   Autostereoscopic 3D. [online] `www.usa.philips.com/c/3d-autostereoscopic-series/303923/cat/en/professional`.

[PJO$^+$09]   Y. K. Park, K. Jung, Y. Oh, J. K. Kim, G. Lee, H. Lee, K. Yun, N. Hur, and J. Kim. Depth-image-based rendering for 3DTV service over T-DMB. *Signal Processing: Image Communication*, 24:122–139, 2009.

[RMOdBaI$^+$02] A. Redert, C. Fehn M. O. de Beec and, W. A. Ijsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, I Sexton, and P. Surman. Advanced three-dimenstional television systems technologies. . In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*, pages 313–319, 2002.

[SHKO11]   M. Sjöström, P. Härdling, Linda S. Karlsson, and R. Olsson. Improved depth-image-based rendering algorithm. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4, 2011.

[SKS05]   O. Schreer, P. Kauff, and T. Sikor. *3D Video communication Algorithms, concepts and real-time systems in human centered communication*. John Wiley & Sons, 2nd edition, 2005.

[SMM+09]   A. Smolic, K. Muller, P. Merkle, P. Kauff, and T. Wiegand. An overview of available and emerging 3d video formats and depth enhanced stereo as efficient generic solution. In *Picture Coding Symposium, 2009. PCS 2009*, pages 1–4, 2009.

[SS02]   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.

[TAZ+04]   W. J. Tam, G. Alain, L. Zhang, T. Martin, and R. Renaud. Smoothing depth maps for improved steroscopic image quality. *Three-Dimensional TV, Video, and Display III*, 5599:162–172, 2004.

[TLD07]   Z. Tauber, Z. N. Li, and M. S. Drew. Review and preview: Disocclusion by inpainting for image-based rendering. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, 37(4):527–540, 2007.

[UBH+08]   G. M. Um, G. Bang, N. Hur, J. Kim, and Y. S. Ho. 3d video test material of outdoor scene. ISO/IEC JTC1/SC29/WG11/M15371, April 2008.

[UCES11]   H. Urey, K. V. Chellappan, E. Erden, and P. Surman. State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE*, 99(4):540–555, April 2011.

[vsr10]   Report on experimental framework for 3D video coding. ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631, October 2010. Guangzhou, China.

[WBSS04]   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.

[Whe38]   C. Wheatstone. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions*, 120:371–394, 1838.

[WSI07]   Y. Wexler, E. Shechtman, and M. Irani. Space-Time Completion of Video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(3):463–476, 2007.

[ZKU+04]   C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szelisk. High-quality video view interpolation using a layered representation. *ACM Trans. Graph.*, 23(3):600–608, August 2004.

[ZT05]   L. Zhang and W. J. Tam. Stereoscopic image generation based on depth images for 3D TV. *IEEE Transactions on Broadcasting*, 51(2):191–199, June 2005.

[ZwPSxZy07] L. Zhan-wei, A. Ping, L. Su-xing, and Z. Zhao-yang. Arbitrary view generation based on DIBR. In *Intelligent Signal Processing and Communication Systems*, pages 168–171, December 2007.

Biography

Suryanarayana Murthy Muddala was born on the 14th of August 1984 in Malakapalli, Andhra Pradesh, India. He received Master of Science in Electrical Engineering with emphasis of Signal Processing from Blekinge Institute of Technology, Sweden in 2009. In 2011 he started his PhD studies at the department of Information Technology and Media at Mid Sweden University in Sundsvall, Sweden. His main research interest are rendering and capturing 3D content.