

High-Detail 3D Reconstruction of Scenes with Complex Materials

Liam Nortje

*Master of Science in Computer Vision, Robotics and Machine
Learning*

from the

University of Surrey



Department of Electronic Engineering

Faculty of Engineering and Physical Sciences

University of Surrey

Guildford, Surrey, GU2 7XH, UK

May 2021

Supervised by: Dr Jean-Yves Guillemaut

©Liam Nortje 2021

DECLARATION OF ORIGINALITY

I confirm that the project dissertation I am submitting is entirely my own work and that any material used from other sources has been clearly identified and properly acknowledged and referenced. In submitting this final version of my report to the JISC anti-plagiarism software resource, I confirm that my work does not contravene the university regulations on plagiarism as described in the Student Handbook. In so doing I also acknowledge that I may be held to account for any particular instances of uncited work detected by the JISC anti-plagiarism software, or as may be found by the project examiner or project organiser. I also understand that if an allegation of plagiarism is upheld via an Academic Misconduct Hearing, then I may forfeit any credit for this module or a more severe penalty may be agreed.

High-Detail 3D Reconstruction Of Scenes With Complex Materials

Liam Nortje

Author Signature  Date: 08/09/2021

Supervisor's name: Dr Jean-Yves Guillemaut

WORD COUNT

Number of Pages: 56

Number of Words: 14338

ABSTRACT

Depth estimation and 3D reconstruction from images is a core technology used in various applications such as, film, computer graphics, earth observation, and medical imaging. There are two main categories of 3D reconstruction: active methods, which interact physically with the object, and passive methods, that measure the radiance or irradiance of the object with a sensor.

There is an abundance of 3D reconstruction techniques using passive methods. However, the goal of achieving high-quality 3D reconstruction using these techniques has posed a considerable Computer Vision challenge. The main reason is because they assume the bidirectional reflectance distribution function (BRDF) of materials are Lambertian (diffuse) leading to discontinuities in their reconstructions. However, the approach Helmholtz Stereopsis does not assume the BRDF of an object, yet this approach alone produces lower quality depth maps compared with other passive approaches such as epipolar plane image volumes.

From a review of literature on this subject, it is evident that a gap exists with current methods and neither Helmholtz Stereopsis nor epipolar planes offer a complete solution. Therefore, the project detailed in this paper focuses on improving passive 3D reconstruction methods by combining Helmholtz Stereopsis with epipolar plane images to form a novel approach to achieve high-quality 3D reconstruction of complex materials.

As part of the project a novel image acquisition technique was created to produce a dataset of 1200 images within Blender. Using these images disparity maps were calculated using epipolar plane image volumes eliminating unknown depth variables in Helmholtz reciprocity equations, thereby enabling the calculation of object surface normals from only two linear path pairs. The results obtained from four objects showed an average mean difference of normal angles from their true value of 45.09 degrees with an average standard deviation of 9.22. This best standard deviation achieved was 6.8.

From an overall project evaluation perspective, the project successfully achieved the four objectives it set, notwithstanding that the results achieved by the project were not perfect. The results were affected by the presence of occlusions, reflections, and the object point matching algorithm used. The results furthermore showed that the approach used suffered from a poor foundation of disparity maps that affected the outcome. However, it is evident that the combination of Helmholtz Stereopsis and epipolar plane images can form a novel approach to extract surface normals. Thereby, contributing to filling the identified gap in the literature and current passive 3D reconstruction approaches.

This paper further proposes another novel method for future work that could fully integrate Helmholtz Stereopsis into epipolar plane image volumes. This would allow for reflective materials to be modelled and their normals calculated.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Jean-Yves Guillemaut for his support, valuable advice, and contributions towards this project. Furthermore, I would like to thank my family for their support throughout this project.

TABLE OF CONTENTS

Declaration of originality.....	ii
Word Count.....	iii
ABSTRACT.....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
GLOSSARY.....	viii
LIST OF FIGURES	ix
INTRODUCTION.....	2
1.1 Background and Context.....	2
1.2 Scope and Objectives	3
1.3 Achievements.....	4
1.4 Overview of Dissertation.....	5
2 BACKGROUND THEORY AND LITERATURE REVIEW.....	6
2.1 3D Reconstruction.....	6
2.2 Multiview Stereo.....	7
2.3 Epipolar Plane Images.....	7
2.4 Photometric Stereo.....	9
2.5 Helmholtz Stereopsis.....	10
2.6 Summary.....	12
3 APPROACH OVERVIEW	13
3.1 Introduction.....	13
3.2 First Approach.....	13
3.3 Second Approach.....	14
3.4 Prerequisites.....	15
4 METHOD	15
4.1 EPI depth estimation software.....	15
4.1.1 Use of the software.....	15
4.1.2 Issues	16
4.1.3 Solutions.....	16
4.1.4 Results.....	17
4.2 Synthetic data acquisition.....	18
4.2.1 Method.....	19
4.2.2 Results.....	24
5 USING HELMHOLTZ AND EPI APPROACH.....	30
5.1 Introduction.....	30

5.2	Approach taken.....	30
5.3	Issues and Solutions.....	33
5.4	Results and Evaluation	34
6	CONCLUSION	39
6.1	Evaluation.....	40
6.2	Future Work.....	41
6.3	Self-Reflections	42
7	REFERENCES.....	43

GLOSSARY

BRDF – Bidirectional reflectance distribution function

EPI – epipolar plane image

EPI volume – a 3D structure of epipolar plane images

SRHARL – Scene Reconstruction from High-Spatio Angular Resolution Light Fields

SVD – Singular Value Decomposition

2D – Two Dimensional

3D – Three Dimensional

LIST OF FIGURES

Figure 1 – Geometry of epipolar plane images.....	8
Figure 2 – EPI volume [23].....	8
Figure 3 – Helmholtz reciprocity camera and light setup [7].....	10
Figure 4 – First approach image acquisition.....	13
Figure 5 – Second approach EPI setup	14
Figure 6 – Bike dataset with corresponding disparity map	17
Figure 7 – Mansion dataset with corresponding disparity map	17
Figure 8 – Mansion depth map from Kim et al.'s paper [4].....	18
Figure 9 – Bike depth map from Kim et al.'s paper [4].....	18
Figure 10 – Disparity map with ambient illumination.....	19
Figure 11 – Disparity map with active illumination.....	19
Figure 12 – Helmholtz setup showing camera and light pairs positions.....	20
Figure 13 – Blender representation EPI setup	22
Figure 14 – Blender representation of Helmholtz setup, showing a camera (black pyramid) and spotlight (black rings).....	22
Figure 15 – Blender representation of opposite Helmholtz pair to figure 14.....	23
Figure 16 – EPI Blender image render.....	24
Figure 17 – A) Helmholtz Blender render with camera in up position - B) Helmholtz Blender render with camera in down position.....	25
Figure 18 – A) Blue Monkey disparity map - B) Blue Monkey true depth map.....	25
Figure 19 – A) Drum set EPI Blender render - B) Drum set Helmholtz Blender render (camera in up position) - C) Drum set Helmholtz Blender render (camera in down position)	26
Figure 20 – A) Drum set EPI disparity map - B) Drum set true depth map.....	27
Figure 21 – A) Textured Monkey EPI Blender render - B) Textured Monkey Helmholtz Blender render (camera in down position) - C) Monkey Helmholtz Blender render (camera in up position)	27
Figure 22 - A) Textured Monkey EPI disparity map - B) Textured Monkey true depth map.....	28
Figure 23 - A) Chair EPI Blender render - B) Chair Helmholtz Blender render (camera in down position) - C) Chair Helmholtz Blender Render (camera in up position).....	28
Figure 24 - A) Chair EPI disparity map - B) Chair true depth map.....	29
Figure 25 - Drum set high resolution EPI disparity map	30
Figure 26 – Inverted Monkey normal map.....	34
Figure 27 - Double Chair object normals.....	34
Figure 28 - A) Chair calculated normal map - B) True Chair normal map – C) Chair heat map.....	35
Figure 29 – A) A copy of figure 28A and B, highlighting point discrepancies.....	35

Figure 30 – A) Textured Monkey calculated normal map – B) True Monkey normal map – C) Textured Monkey heat map.....	36
Figure 31 – Poly Haven [45] normal map	36
Figure 32 – A) Drum set calculated normal map – B) True Drum set normal map – C) Drum set heat map.....	37
Figure 33 - A) Blue Monkey normal map calculated - B) True Blue Monkey normal map – C) Blue Monkey heat map.....	37

INTRODUCTION

Depth estimation and 3-dimensional (3D) reconstruction from images is a fundamental challenge in Computer Vision and has benefited significantly from recent improvements in computer and camera hardware, algorithm techniques and advancements in neural networks. 3D reconstruction is concerned with modelling the 3D world and is used in a variety of applications: such as film; computer aided geometric design; computer graphics; earth observation; computer vision; medical imaging; and city planning. This facilitates the modelling of objects for the placement in a virtual environment, enabling easier viewing, use and experimentation.

There are two main types of 3D reconstruction methods: active and passive. Active methods interact physically with the object they are trying to reconstruct, for example using a laser scanner. With a laser scanner the time delay of the reflection can be used to determine its distance and generate a depth map for reconstruction. Conversely, passive methods do not interfere with the object and typically use two-dimensional (2D) images to create depth maps.

Active methods are normally more accurate than passive methods. This is because active methods can directly measure the object. These methods, however, typically require niche and expensive hardware like Lidar sensors, which make the methods less accessible. Nowadays through advancements in technology cameras have improved and reduced in cost, making them widely accessible to most people. This has made bridging the quality gap between active and passive methods more important as it allows people and companies access to technology that was less accessible.

This paper therefore focuses on the use of passive Computer Vision techniques to achieve high quality 3D reconstruction and describes a novel approach to fill a gap in current and previous methods of software.

1.1 Background and Context

A 3D reconstruction can be made from combining multiple depth maps of an object. Depth estimation from a pair of images is calculated based on the disparity of a pixel in one image to the same pixel in the second image or in mathematical terms $d = x_{left} - x_{right}$ where d is the disparity and x is a pixel's position given by $x = [u, v]$. This disparity is used to calculate the depth from motion parallax: pixels closer to the camera will move more than pixels far from the camera [1]. From this it is apparent that disparity is inversely proportional to depth. The matching of pixels is known as the correspondence problem and is affected by several issues, namely noise and occlusion [2].

There are several techniques to find matching pixels. The approaches range in accuracy, ease of use and output quality. Some use a dataset estimation approach [3] while others use epipolar plane

images (EPIs) [4]; but commonly assume that the surfaces of objects are Lambertian (i.e., uniformly reflective). This leads to limitations in many approaches available and the models they generate as most objects in the real world are not completely Lambertian.

One approach that does include the generation of models for non-Lambertian materials is Helmholtz Stereopsis [5]. This approach exploits reciprocity to find points on a surface of an object using 2D images. While this method works with non-Lambertian materials, it produces low quality reconstructions compared to other methods such as High Spatio-Angular Light Fields that uses epipolar plane images [4].

This document details a project that aims to improve upon existing approaches by incorporating Helmholtz Stereopsis in Scene Reconstruction from High Spatio-Angular Light Fields (SRHARL). This would improve the quality of the depth map and therefore the 3D reconstruction previously seen by Helmholtz Stereopsis. This would also allow for additional normal information to be gathered that SRHARL lacks.

1.2 Scope and Objectives

Overall Goal

The goal of this paper is to form a novel approach by using a combination of High Spatio-Angular Resolution Light Fields and Helmholtz Stereopsis to improve upon high quality 3D reconstructions of models with complex materials.

Primary Objectives

This paper seeks to achieve the following objectives which will contribute to fulfilling the overall goal of the paper explained above:

1. Ensure the code for epipolar images can be used to model 3D objects.
2. Combine the Helmholtz Stereopsis approach and epipolar plane images approaches.
3. Produce at least two 3D models of Lambertian objects.
4. Produce at least two 3D models of non-Lambertian objects.

The scope of the project is detailed below in table 1.

Table 1 – Scope of project

In Scope	Out of Scope
3D reconstruction from 2D images, of small	Other forms of 3D reconstruction

objects.	
Modelling of small objects only, for example a drum set.	Modelling of whole environments
Combination of the Helmholtz Stereopsis and the EPI approach.	Combination of any other approaches not discussed
3D modelling from high resolution images.	Image resolutions larger than 1000x1000 pixels.

The success of achieving the objectives and goal of the project will be detailed in Evaluation section 6.1 to determine if the project was successful or not. The overall success of the project is dependent on the overall goal. Primary objectives lead towards the overall goal and will also be assessed in section 6.1.

1.3 Achievements

This paper provides a novel approach to 3D reconstruction enabling normal information to be extracted from 2D images. This was achieved through combining two previously known methods, Helmholtz Stereopsis and epipolar plane image volumes, in Computer Vision leveraging the advantages into a new approach.

As the approach is new there were no available datasets to achieve results. This paper therefore generated a dataset of 1200 images using a novel image acquisition technique in Blender [5]. This technique reduces the need for complex camera and light setups used by Helmholtz Stereopsis.

Using Blender, models were acquired, and two setups were created with lighting and a stage to move a camera on. Three hundred models for each setup were then rendered in high-resolution to achieve objective two.

The novel approach was then used on the datasets to produce disparity maps and normal maps showing that surface normal extraction is possible.

Furthermore, this paper proposed another novel method based on the one used in this project which could extract surface normals from non-Lambertian objects while achieving better quality 3D reconstructions compared to current methods.

1.4 Overview of Dissertation

Section 1:

This section contains an introduction to the project giving the project background and context. This section also contains the goals, objectives and summarises the project's achievements.

Section 2:

Background theory and a literature review is presented showing a brief history of relevant works, state of the art Multiview stereo approaches, their issues, solutions including photometric stereo, and Helmholtz Stereopsis and a novel approach that will combine solutions.

Section 3:

An overview of two novel approaches is then detailed, explaining how they achieve 3D reconstruction. This section also gives details one why one approach was selected over the other.

Section 4:

This section provides technical detail on the EPI software used, the issues and problems encountered are described together with a section showing how relevant synthetic image datasets were created to achieve the project's main goal.

Section 5:

Section 5 provides an explanation of the code used to combine Helmholtz Stereopsis and SRHARL. This section includes issues encountered and their solutions along with results obtained.

Section 6:

A conclusion then follows summarising the whole thesis with an evaluation on the success of the project based on the objectives achieved. This is followed by a recommendation on future uses and improvements. This last section of the paper details a short self-reflection of what I have learnt, and what could have been improved.

Section 7:

This section contains all references used throughout the paper.

2 BACKGROUND THEORY AND LITERATURE REVIEW

Depth estimation and 3D reconstruction from images is a fundamental challenge in Computer Vision that has benefited significantly from improvements in computer and camera hardware, algorithm techniques and neural networks advancements. There are several areas within depth estimation and 3D reconstruction, including single image depth estimation using labelling [5], deep convolution and neural network depth estimation [6], depth estimation using stereo techniques [7], and multiple image depth estimation and 3D reconstruction [8]. This literature review will not include reviews of single image depth estimation or the use of deep neural networks unless it provides insight into the current context.

The following section will review and discuss the literature, models and frameworks used in 3D reconstruction research. The structure of the literature review is as follows: firstly, a brief introduction to 3D reconstruction will be presented to ensure alignment and knowledge before going in-depth with theoretical concepts of depth estimation and 3D reconstruction. This will be followed by the three main categories for 3D reconstruction research regarding 2D images. The concept of Multiview stereo and a review of epipolar plane images will briefly be explored to understand the background of research regarding multiple image dense 3D reconstruction of Lambertian materials. Then a review of Photometric stereo will be made to understand how objects with priori knowledge of the bidirectional reflectance distribution function (BRDF) can be reconstructed. Finally, Helmholtz Stereopsis will be presented in order to show how objects can be reconstructed with arbitrary or unknown BRDFs. All areas will cover advantages and disadvantages to identify gaps in the literature of depth estimation and 3D reconstruction to motivate a novel approach.

2.1 3D Reconstruction

To create a 3D reconstruction depth estimation of pixels is required. This normally involves taking images of an object from different viewpoints and calculating the disparity/motion parallax or in other words the distance a pixel has moved from one image to another. The disparity calculated is inversely proportional to the depth of that pixel: pixels that move more from one image to another are respectively closer than those that move less [1]. This is relatively simple, however matching the exact object point/pixel from one image to another is very challenging. This matching is called the correspondence problem [2] and is made more difficult due to noise and occlusions. There are several approaches to solving the correspondence problem and the most relevant to this paper are explored below.

2.2 Multiview Stereo

Multiview stereo is the use of multiple 2D images for 3D reconstruction. Images are taken as the camera moves to produce motion parallax and estimate disparity.

One of the first methods to find corresponding changes in images was the use of image correlation [9] whereby windows are used to find correspondence between pixel intensities. A window is a small square or rectangle which is parsed over the images in a grid. The windows for the first image are compared to the windows in the next image and so on. The goal is to find two windows that contain the same image segment. The comparison of windows can be approached with two main classes of algorithms. These classes are correlation-based algorithms [10] which produce a dense set of correspondences or feature-based algorithms [11] which produce a sparse set of correspondences. Sparse features for example are edges and corners whereas dense features relate to all pixels and may include estimation based on intensity of colour. Sparse algorithms are typically faster whereas dense allows for estimations on all pixels. Using windows, however, can come with downsides as a too small a window may be too sensitive to noise and too large a window will be less sensitive to differing image intensities, causing discontinuities in the final disparity map [12]. To solve this issue better window approaches can be used including adaptive [10], multiple [13], [14], predictable [15], and shiftable [16] windows all of which aim to mitigate these issues.

Discontinuities often occur at occlusions as a pixel in one image should match with at least one pixel in another image. Methods that perform well with occlusions involve energy minimisation [17] and are classed as a global approach. Energy minimisations use a source (foreground) and sink (background) normally involving graph cuts [18] to segment the image based on the minimum energy cut. Minimum energy cuts are found due to certain properties, specifically that a pixel is dependent only on its neighbouring pixels [18]. Weights are given between pixels to show their relationship towards one another and their relationship to the source and the sink. The line through the pixels where the weights are smallest is the minimum energy cut. This allows them to set a pixel with no pair pixel to infinite energy disregarding it [17]. To find the minimum of the energy function a form of global optimisation is needed. Global optimisation is a slow process as there are many calculations to compute. This computation time increases significantly for higher resolution images. Despite the use of global minimisation showing promising results for 3D reconstruction [17] this method will not be considered as a viable option for the theoretical foundation of this paper because it is argued to be highly time consuming [19] and therefore out of scope for this paper.

2.3 Epipolar Plane Images

Previously explored approaches can be used on any image pair as they are not reliant on the geometric positions of the camera. However, if the images can be captured in a certain way geometric properties

of the scene become available to use in the disparity calculation [20]. One example is epipolar plane images (EPI) [21]. EPI are images that are taken in the same plane. To achieve this, pictures must be taken by moving a camera uniformly on a linear stage. This enables advantage to be taken of some interesting characteristics. This is because of the specific geometric properties of epipolar plane images [9]. One of these characteristics is the creation of epipolar lines [22]. Epipolar lines are scanlines as a pixel's disparity between one image and another will only change within one axis. This is shown in figure 1 below where the camera has moved to the right from camera position 1 to camera position 2. Note that the pixel has only changed its x direction staying within the scanline.

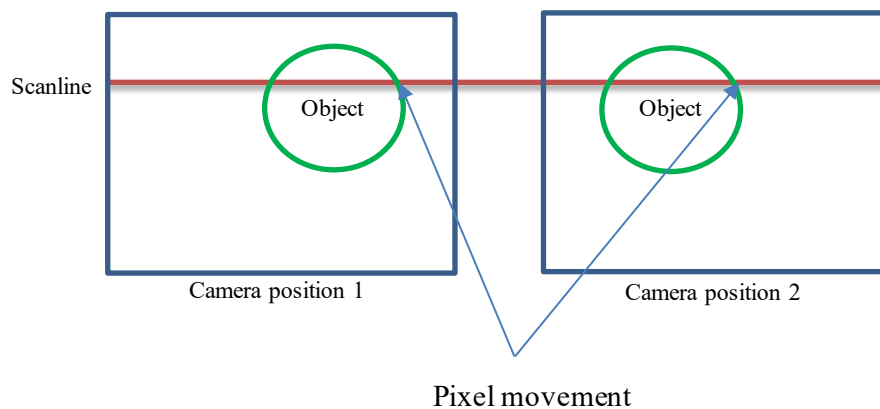


Figure 1 – Geometry of epipolar plane images

This characteristic reduces the time needed to search for pixel pairs as only scanlines need to be searched. This offers an advantage over approaches like window correlation mentioned above in section 2.2 as there is less area to search through. If multiple epipolar images are collected and stacked, more geometric advantages may be used. Criminisi et al. [23] show that an epipolar volume can be created giving rise to EPI strips. These strips may be represented as a slice horizontally through this 3DEPI volume. An image representation of these EPI strips looks like straight lines of different colours or intensities as seen in figure 2 [23].



Figure 2 – EPI volume [23]

What makes them useful is that they each have their own slant [19]. From this slant a gradient can be calculated which is proportional to disparity and therefore their depth within the image [23].

There are many benefits to EPI geometry including their flexibility to combine with other approaches, this makes them especially useful. Consequently, they have been used in several papers [24], [25], [26].

One paper that uses EPIs is by Kim et al. [4] which uses these EPI strips with high resolution images allowing for higher quality 3D reconstruction. This is achieved through dense feature sets focusing on pixel level calculation. This mitigates the discontinuity problems highlighted earlier in section 2.2. The algorithm is faster as there is no global regularisation as they use a fine-to-coarse method [4]. This contrasts with the normal coarse-to-fine approach used in energy minimisation [27]. Global regularisation is needed in the coarse-to-fine approach because of grey level constancy assumption and so the algorithms are normally overly sensitive to noise [27]. Kim et al.'s [4] fine-to-coarse approach operates by going from high resolution to low resolution. A sparse edge confidence measure is used to calculate the most promising EPI's to determine depth estimates. This is iterated through to down sampled lower resolution images and continued until all EPI pixel depths have been calculated [4]. Kim et al. [4] found that their method was faster and more detailed than other existing methods including graph cuts [17] and semi global matching [28]. None of the other methods were able to run the 21MP image resolutions. However, Kim et al.'s approach [4] suffered from spatially varying reflectance. The problem of reflectance is a major issue within 3D reconstruction methods as the approaches work on the assumption that the materials, they are measuring depth for are Lambertian (diffuse). The reflection issue is common as seen in [26] and [17]. A paper by Wanner et al. [29] showed that EPI geometry can be used to reconstruct the Lambertian and or reflection present in the image using higher order tensors. They also take advantage of the vertical slices of the EPI volume. As the paper is still based on a Lambertian model errors still occur. The reconstruction also does not give rise to the normal of the reflected material and so reconstructs the reflection as a non-Lambertian model. Future work on this paper [29] may yield promising results. The sections below cover approaches which can reconstruct objects with priori knowledge, arbitrary or unknown BRDFs.

2.4 Photometric Stereo

Photometric stereo differs from the other approaches explored in section 2.2 and 2.3. Instead of changing the position of the camera to take pictures, the camera is kept stationary, and a light is moved instead, to illuminate the object from different angles. The first introduction of photometric stereo was from Woodham [30]. By keeping the object stationary Woodham [30] shows that the previous correspondence problems can be solved because pixels do not move from one image to another. He explains that by making assumptions about the bidirectional reflectance distribution functions (BRDFs) and creating orthographic projection the viewing angle to the light source (phase) remains constant for all points on the surface of the object. Orthographic projection creates a certain

property whereby image points are directly related to object points, (x,y) where $f(x,y)$ is the BRDF. The surface normal can then be written as a vector of gradients of the function of these object points. These gradients can then be represented in a gradient space [30] which can be looked through to find the meeting points of conic curves [31]. Two different illuminations would provide two meeting points where one is the right point, therefore at least three different illuminations would be needed. The problem with photometric stereo is that the objects cannot be self-occluding. This gives rise to several limitations in the method as most real-world objects contain occlusions or are textured. Li et al. [32] argue that photometric stereo is prevented from being practical in use because of the error of depth reconstruction and surface reflectance. Furthermore, the priori BRDF must also be known which is difficult to obtain in reality [33].

2.5 Helmholtz Stereopsis

Helmholtz Stereopsis [7] is another approach to depth estimation where materials are not considered to have Lambertian reflectance. Magda et al. [34] were the first to introduce the use of Helmholtz reciprocity for 3D reconstruction. This is based on the Helmholtz principle of how our eyes view in 3D [35] and was used to reconstruct surfaces with arbitrary BRDFs. Zickler et al. [7] further developed the principle and proposed Helmholtz Stereopsis, a method for reconstruction of surfaces with unknown and arbitrary surface reflectance.

Zickler et al. [7] presented a method that exploits the symmetry of an object's reflectance. To do this, pairs of images are collected where the camera and light source swap exact positions for each image. This creates symmetric outgoing radiance and incident irradiance for each surface point on the object. This symmetry is called Helmholtz reciprocity. The setup for this is shown in figure 3 [7] below where o_l and o_r are camera and light positions, \hat{v}_l and \hat{v}_r are vectors of irradiance and incident light rays, p is the point on the object and \hat{n} is the normal vector of point p .

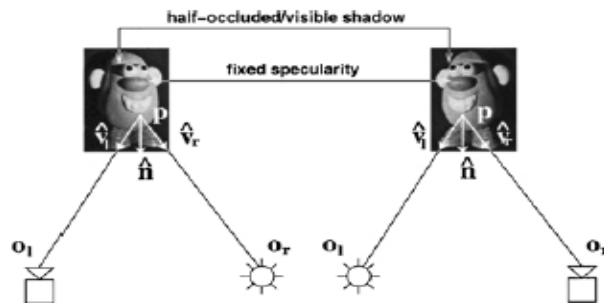


Figure 3 – Helmholtz reciprocity camera and light setup [7]

This reciprocity is not obtained from approaches that have fixed light sources unless the objects are assumed to be Lambertian. In using Helmholtz reciprocity Zickler et al. [7] show that the BRDF

will cancel out. Equations 1 and 2 [7] below show that the intensity of a pixel for the camera in the left position and the camera in the right position i_l and i_r respectively is based on the BRDF (f_r and f_l). Because of the specific setup used in Helmholtz Stereopsis [7] these two equations can be equated. This is shown in equation 3 [7].

$$i_l = f_r(\hat{\mathbf{v}}_r, \hat{\mathbf{v}}_l) \frac{\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_r}{|\mathbf{o}_r - \mathbf{p}|^2} \quad \text{Equation 1}$$

$$i_r = f_r(\hat{\mathbf{v}}_l, \hat{\mathbf{v}}_r) \frac{\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_l}{|\mathbf{o}_l - \mathbf{p}|^2} \quad \text{Equation 2}$$

$$f_r(\hat{\mathbf{v}}_r, \hat{\mathbf{v}}_l) = f_r(\hat{\mathbf{v}}_l, \hat{\mathbf{v}}_r) \quad \text{Equation 3}$$

The BRDFs therefore cancel each other and equation 4 [7] below is derived.

$$\left(i_l \frac{\hat{\mathbf{v}}_l}{|\mathbf{o}_l - \mathbf{p}|^2} - i_r \frac{\hat{\mathbf{v}}_r}{|\mathbf{o}_r - \mathbf{p}|^2} \right) \cdot \hat{\mathbf{n}} = \mathbf{w}(d) \cdot \hat{\mathbf{n}} = 0 \quad \text{Equation 4}$$

$$\text{Where } v_l = \frac{1}{\|\mathbf{o}_l - \mathbf{p}\|} (\mathbf{o}_l - \mathbf{p}) \text{ and } v_r = \frac{1}{\|\mathbf{o}_r - \mathbf{p}\|} (\mathbf{o}_r - \mathbf{p}).$$

If the position of the camera and light source are known and the angles of the incident and irradiance can be calculated a vector w can be created for that point. The pixel pairs in the Helmholtz images are then iterated over creating w vectors which can be stacked to create a W matrix as shown in equation 5.

$$W \begin{bmatrix} w \\ w \\ w \end{bmatrix} \cdot \begin{bmatrix} \hat{n} \\ \hat{n} \\ \hat{n} \end{bmatrix} = 0 \quad \text{Equation 5}$$

This matrix lies in an epipolar plane defined by the object point, irradiance and incident, the only remaining unknowns are the surface normal and the depth. In a special case where at least three pairs

of images are taken a constraint can be made where if the hypothesised depth has a surface point the normal will be determined as a unit vector in the W matrix 1-D null space [7]. The depth values are estimated, and W matrices are calculated. The depth values that lie on a real point on the object will be based on the ratio of rank two and three singular decompositions of the matrix [7]. The normal is then found from the right singular vector which is related to the smallest singular value from the singular value decomposition.

While Helmholtz Stereopsis is shown in practise to be effective in finding surface normals and depths for non-Lambertian materials, the quality of depth estimation struggles with discontinuities [7] and therefore the quality of the final depth map is lower. As Helmholtz reciprocity requires certain light sources, capturing real images with ambient light can become an issue. By measuring the ambient light and subtracting it from the captured images Tu et al. [36] solved this issue, making the setup more accessible for depth estimation. The setup of multiple cameras and light pairs to satisfy Helmholtz reciprocity is more complex than some other approaches explored in this literature review [24], [26], [4] but less complex than photometric stereo [30].

Helmholtz Stereopsis is a powerful technique as it is the only BRDF independent technique [37]. Because of this there have been several recent papers using Helmholtz Stereopsis [38], [39], [40]. These papers all expand upon Zickler et al.'s [7] paper using the reciprocity constraint to improve the approach across different areas. Guillemaut et al. [37] shows that the constraint when used on single object points failed when using an object with a rough texture because of inter-reflections. This was solved by taking radiance measurements from surface point neighbourhoods. Another paper [41] shows the use of wavelength multiplexing combining photometric calibration and a technique for surface calibration using Helmholtz Stereopsis.

2.6 Summary

There are multiple methods to solving the correspondence problem. Some methods struggle with discontinuities [10], [11] and others suffer with occlusions [27]. However, all current approaches have problems with reflectance. Two methods including Photometric Stereopsis [30] and Helmholtz Stereopsis [7] solve this issue but produce low quality 3D reconstructions. This literature review identified a gap in the approaches and academic literature of high-quality 3D reconstruction. This literature review therefore finds that the combination of Kim et al.'s [4] approach on Scene Reconstruction from High Spatio-Angular Resolution Light Fields and Helmholtz Stereopsis [7] could contribute to filling the gap needed in 3D reconstruction.

3 APPROACH OVERVIEW

This section describes two different approaches to combine Helmholtz Stereopsis [7] and Scene Reconstruction from High-Spatio Angular Resolution Light Fields (SRHARL) [4] into a novel system that would improve upon previous methods. These two approaches are detailed below.

3.1 Introduction

As highlighted by the literature review in section 2.6 a gap exists for the creation of a novel approach to 3D reconstruction. This would entail using two methods discussed in sections 2.3 and 2.5, namely SRHARL [4] and Helmholtz Stereopsis [7]. Helmholtz Stereopsis uses Helmholtz reciprocity of light to find the normals of object points from images and SRHARL uses EPI volumes to find depths of pixels in images. From the research carried out by this project these two methods have never been used in conjunction. This paper therefore proposes two different approaches to leverage their complementarity to improve upon existing methods and fill the gap discussed in the literature review.

3.2 First Approach

The first approach was to add the W matrices to the EPI pixel gradients found in the EPI volume. As shown before in section 2.3 EPI volumes are created from several images taken at regular intervals along a linear stage. These images are then stacked on top of each other creating slanting lines in the horizontal slice of the volume. Using a program that looks for similar pixel intensities these lines are found from a 2D slice, and their gradient is calculated. As the gradient is proportional to depth, the depth of that pixel or groups of similar intensity pixels can be calculated. EPI methods work well when these lines are uninterrupted: when there are no occlusions or reflections. In contrast, Helmholtz Stereopsis works well with reflections as the BRDF can be arbitrary or unknown. The idea to fill the gap was to replace these slanted lines with the Helmholtz approach. In theory to do this you would use a modified version of the Helmholtz image capture setup. This setup is shown below in figure 4 where a camera and a light are moved at regular intervals along a linear stage. Then their positions are switched, and the same process is repeated.

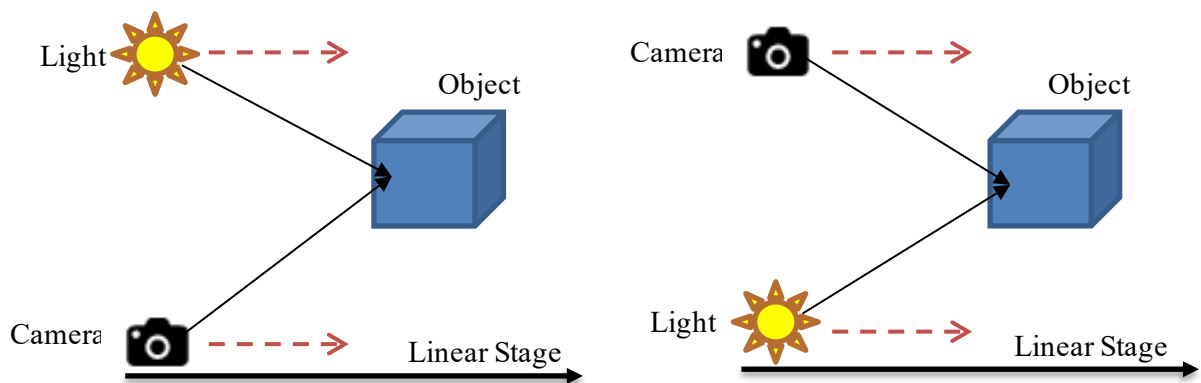


Figure 4 – First approach image acquisition

Position one of the camera and light, and position two of the camera and light pair would then form an EPI volume each. It is possible then to iterate through each pixel for each volume and find its w matrix and form a 4D EPI volume, but instead of pixels you would have w matrices. The idea then is to find some correlation between the w matrices to form a W matrix for each point on the object. This approach would then allow for depth and accurate surface normal estimation of non-Lambertian objects, while increasing the quality of standard Helmholtz Stereopsis. The idea seemed simple, however due to adding another camera the EPI slice would increase in dimensions from 2D to 3D. In practise finding a correlation between w matrices in a higher dimensional 3D EPI slice would be very time consuming and unfortunately out of scope for this master's dissertation. Therefore, the second approach described below in section 3.3 was used.

3.3 Second Approach

The second approach is a simplified version of the first and stems from ideas of the first approach. In Helmholtz Stereopsis depths are estimated and ranked based on their likelihood of being a real scene point on the object through singular value decomposition (SVD). If the SVD of the W matrix is rank 2 then the point is said to be a real point on the object. The computation and accuracy of Helmholtz Stereopsis is reliant on the depth estimations. This estimation process can be eliminated by first using the EPI method to get point depths and then using these depths in the W matrix. The Helmholtz Stereopsis constraint shown in equation 4 in section 2.5 also has two undefined variables, the depth and the normal.

This means that three pairs of images are needed to get a reliable normal. This method would eliminate the unknown depth d and only require pairs of Helmholtz reciprocity images along a linear path to calculate the normals. This reduces the tedious camera and lighting setups needed for Helmholtz Stereopsis. Instead, a normal EPI camera setup is used which can be seen in figure 5 with a linear stage to capture images and calculate depths.

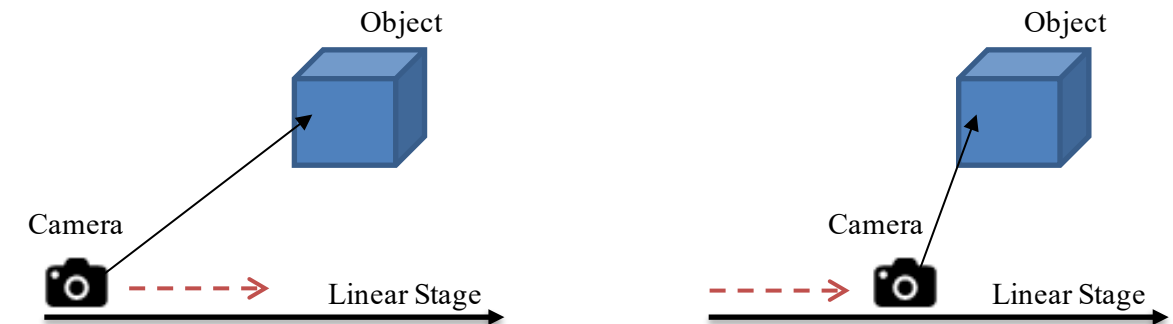


Figure 5 – Second approach EPI setup

Then using the same linear stage, a camera is used, but with a light above it, taking pictures at the same positions as the camera in the EPI setup. The light and camera are then swapped, and images are again captured at the same intervals along the linear stage. This setup can be seen in figure 4. This creates a Helmholtz reciprocity pair for every EPI image. The depth recovered for each pixel in one EPI image will be substituted into a w matrix so that every pixel for every EPI image has a W matrix that can be singularly decomposed giving the normal of that point as the 3rd singular vector of the right matrix. This should yield more information than just the EPI reconstruction and reduce the complexity of the Helmholtz Stereopsis setup. However, as the EPI method is being used to calculate depths, this method will still struggle with non-Lambertian objects. The method of how this was setup is shown below with section 5 providing more detail about how Helmholtz was then applied.

3.4 Prerequisites

Some prerequisites are needed before Helmholtz Stereopsis is applied. These are detailed below in section 4 and includes the synthetic data creation used to test the approach and the EPI depth estimation software; the software used to calculate depths of points. The issues encountered and respective solutions are also presented in each section.

4 METHOD

4.1 EPI depth estimation software

Third party software was acquired from a GitHub repository [42] relating to Kim et al.'s [4] paper Scene Reconstruction from High Spatio-Angular Resolution Light Fields. The program is based on the algorithms used in the paper and contains command line execution within Python [43]. Python was installed as well as all dependencies required. The environment was also configured for use.

4.1.1 Use of the software

The program is designed for command line execution and contains five command line utilities. The most important two of these five are called 'disparity' and 'imgs2lf'. The 'imgs2lf' runs executables to convert image files into lightfield.hdf5 files. HDF5 files are hierarchical files which allow for large data compression however, to read or write to them requires additional coding which in this case was done in Matlab and Python as there are useful libraries and data representations. The most important

command is ‘disparity’ which takes one positional argument, and several optional arguments. The positional argument is the lightfield file that was created from the previous function ‘imgs2lf’ above. The ‘disparity’ function takes the lightfield file and calculates all disparities for every image in the file, plotting graphs of disparity maps as an output. The other important optional arguments for ‘disparity’ are parameter changes including the minimal disparity to sample for, the max to sample for, the disparity step size, edge confidence threshold, depth confidence threshold, bilateral median filter threshold, disabling radiance update and debug mode. The default parameter settings should give reasonable results. Defaults are selected automatically if the arguments are not inputted to the command line.

4.1.2 Issues

Several issues were encountered during the initial setup of the software. The first issue was with ‘imgs2lf’. The program ran through an image folder collecting all the files via a string format. However, since this code was last updated in 2015 some of the code was outdated. This led to errors where the function could not read the file names. Another issue faced was with the ‘disparity’ function. Before the program calculates the disparities, the images are down sampled into a temporary hdf5 file where EPI calculation will take place. When this down sampling happens a copy of the lightfield file is held in a temporary file. However, there is a float conversion from the data type chosen using ‘imgs2lf’. This runs into issues as the temporary file is also the data type that was chosen and therefore the values of the lightfield would all be zero after the EPI calculation. The last issue encountered was that the paper [4] says GPU computation takes place, however in the publicly released program only the CPU is used for computation. This significantly affects computation time, taking four days to finish the disparity calculation on one hundred images using an i9 9900k CPU and an Nvidia 2080 super GPU. This hardware should run the program faster than the speeds mentioned within the paper [4] stating that 50 images should take 50 minutes.

4.1.3 Solutions

To remedy the first issue of strings a simple update to the code was sufficient to enable lightfield creation. The second issue regarding data type conversions can be fixed in two ways. Either when the lightfield.hdf5 file is created using ‘img2lf’, float should be chosen as the data type or the temporary file changed to be created with the float type. The latter method was selected in order to limit changes to the code which should work as is. To achieve faster computation time, smaller resolutions were used. However, this did affect the quality of depth estimations.

4.1.4 Results

After resolving the above issues, the EPI software was used on data sourced from Kim et al.'s paper [4]. The datasets varied in depth and detail. The datasets chosen were the “bikes” images and the “mansion” images. These can be seen below in figure 6 and 7 with their resulting EPI disparity map calculation.

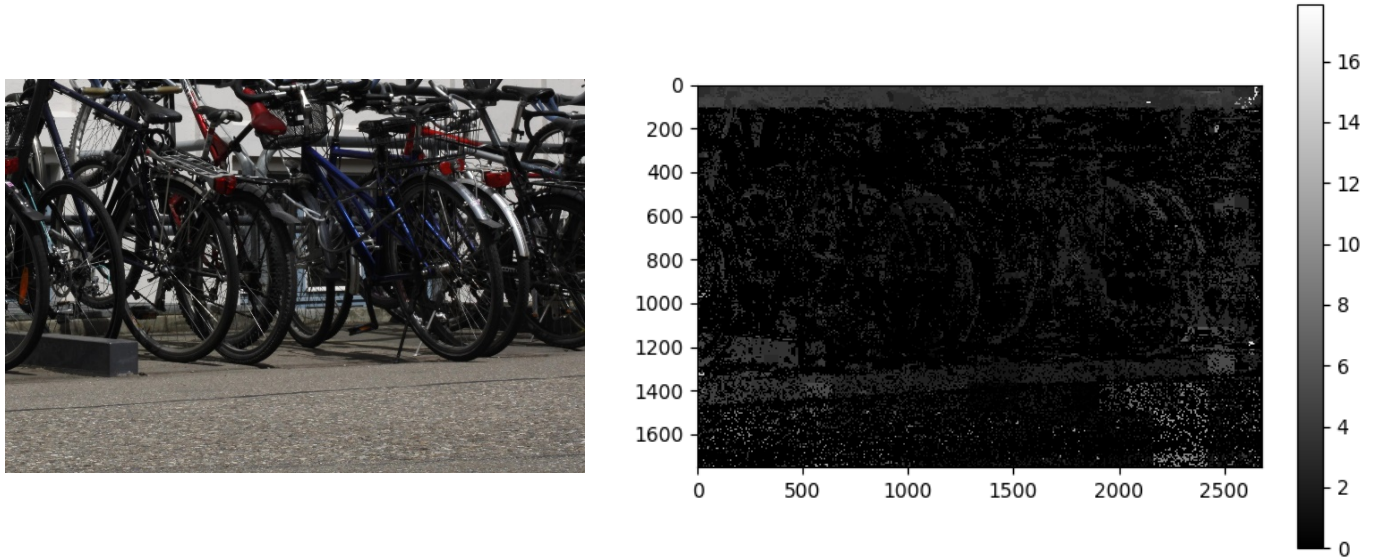


Figure 6 – Bike dataset with corresponding disparity map

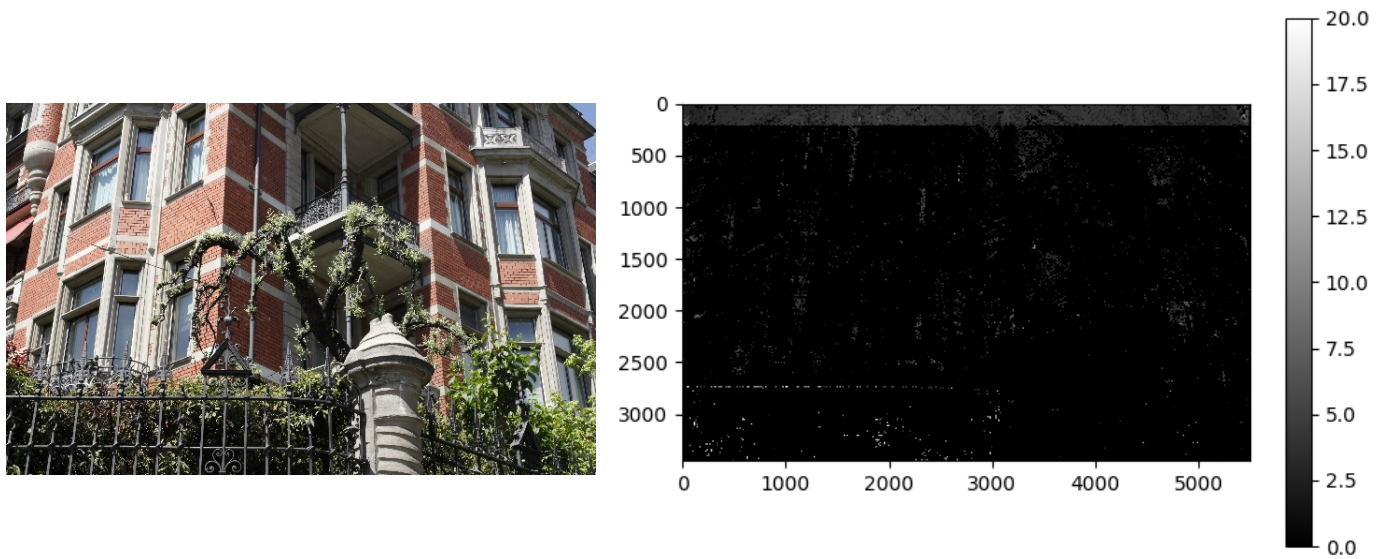


Figure 7 – Mansion dataset with corresponding disparity map

As can be seen even after issues were solved the disparity images did not turn out as expected. These are in contrast, drastically different from the results Kim et al. [4] managed to produce. The results from this paper can be seen below in figure 8 and figure 9.

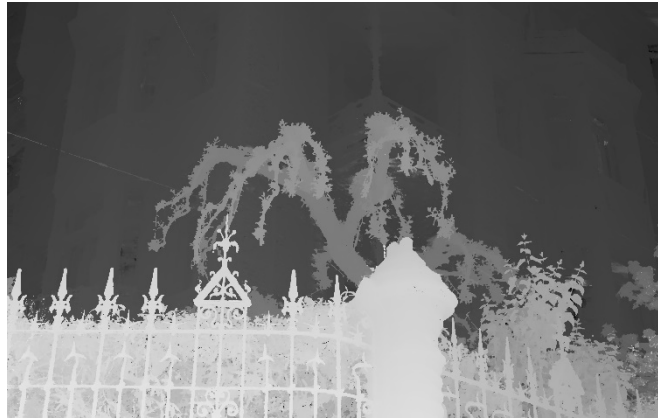


Figure 8 – Mansion depth map from Kim et al.'s paper [4]



Figure 9 – Bike depth map from Kim et al.'s paper [4]

These images shown above in figures 8 and 9 are depth maps and not disparity maps. However, in appearance these should look very similar. The cause of the discrepancy is unknown and would need to be explored in more detail. It could be the case that the public software [42] used for figures 6 and 7 is not the same as that used in the paper for figure 8 and 9.

4.2 Synthetic data acquisition

To test the EPI code and achieve objectives by combining Helmholtz Stereopsis, images are needed. For this project images were created synthetically, this was due to the current pandemic and to save time. Synthetic images have advantages over real life images including but not limited to lighting and material control, faster setup, and less noise. To create the synthetic data Blender [44] was used in an animation style render to create the images needed.

4.2.1 Method

Both Helmholtz Stereopsis and High Spatio-Angular approaches require specific image acquisition to work properly. The easier of the two setups is used for the High Spatio-Angular approach [4] (EPI approach). This approach uses a linear stage, moving the camera at regular intervals to capture images. Depending on how far the camera is from the object and the spacings between camera motions will affect the different disparities of different pixels. The goal is to achieve noticeable differences in disparity between objects close to the camera and objects further from the camera. To compensate for distance the object can be enlarged to still fill the focal length of the camera (to fill the frame). The EPI setup also requires ambient illumination to keep intensities of the same pixels constant. In contrast the Helmholtz setup requires Helmholtz Stereopsis' setup which requires active illumination to meet constraints in equation 4. The setup for the EPI approach can be seen in section 3.3 figure 5. If active illumination is used in the EPI setup, the disparity calculation is worst. This can be seen below in figure 10 showing better disparity measurements (in white) compared to figure 11.

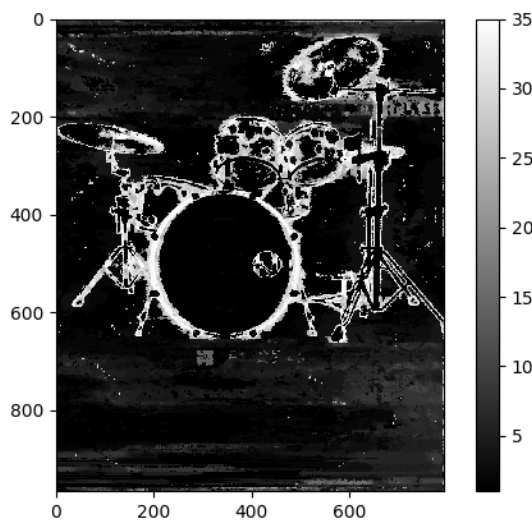


Figure 10 – Disparity map with ambient illumination

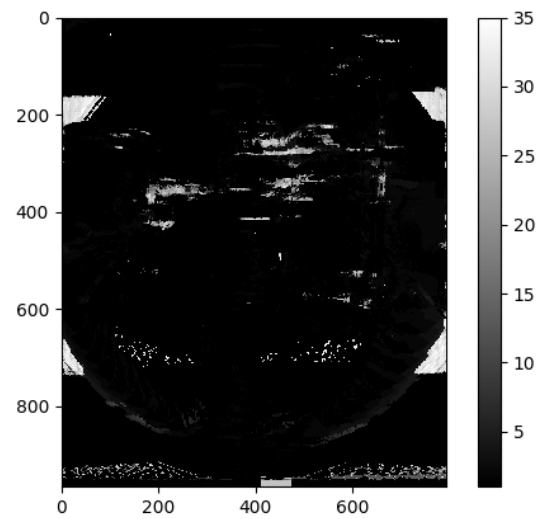


Figure 11 – Disparity map with active illumination

To take advantage of Helmholtz reciprocity a camera and light pair are needed. These still need to move along a linear stage but also need to switch positions creating pairs of images for each position on the linear stage. The position they occupy is also very important as the plane in which the camera and light occupy needs to be different from one another [37]. This means that the camera and light cannot be in line with the direction they are moving. To compensate for this the light can be positioned above the camera and then switched for pair capture. This ensures that they do not occupy the same plane.

Currently there are no datasets available online that fit these parameters. To apply and test the objectives, new datasets were created using Blender [44]. Two datasets were needed. One for the EPI

calculation, this gives the depth estimation of each point on the object, and a second for Helmholtz reciprocity collecting double the images of the EPI setup as the camera and light pairs switch position. This is shown below in figure 12.

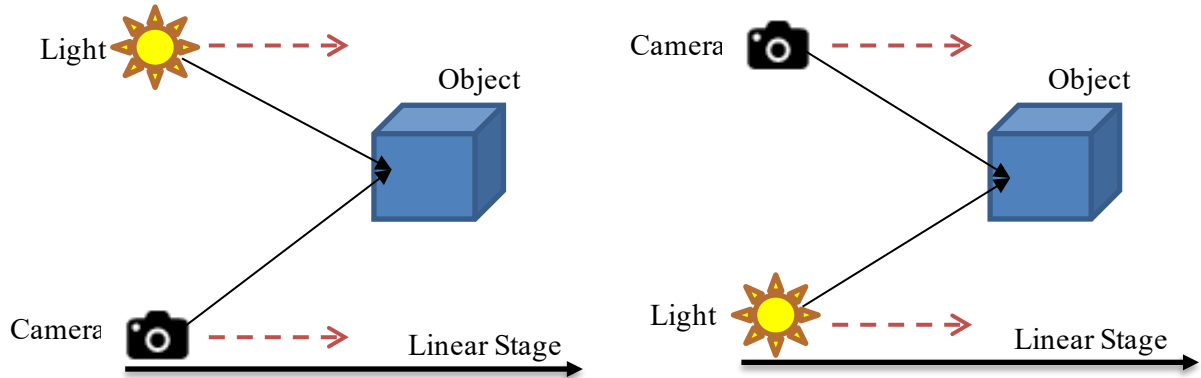


Figure 12 – Helmholtz setup showing camera and light pairs positions

To achieve these two setups, specific requirements for each setup were used, as shown below in table 2.

Table 2 – Requirements for synthetic data setups

Requirement	EPI setup one	Helmholtz setup two
100 images	✓	✓
Camera and Light	✗	✓
Evenly spaced camera movement	✓	✓
Linear stage	✓	✓
Ambient Illumination	✓	✗
Active Illumination	✗	✓
Reflective material	✗	✓

Blender [44] was used for several reasons, these include it being, graphics card supported, free, highly supported and updated, and the main reason: real life light physics simulations which were specifically needed in this case. Using the EPI third party code from Kim et al.'s [4] paper a few objects were tested in Blender. Already pre-made objects can be found online [45] which can be imported into Blender. Simple textured objects are best to use to test the parameters as stated in section 3. "Simple", meaning without many detailed parts like a tree with several leaves.

The EPI setup shown below in figure 13 was created using the monkey object that Blender supplies. The white meshes behind and below the object were added to create a background for the images captured. This is not necessary, as the objects may be masked later, but can be added to potentially make a better contrast between pixel intensity values of the object. In the centre of the figure in black is the camera object. This object has normal camera properties like focal length and resolution. The resolution has been set to 793 x 967 as this is a high-resolution while being low enough to not increase EPI calculations significantly. The resolution should stay the same for all setups as there should be a point on the object that exists for every setup. This setup does not have a light setup as this would create occlusions on the object (some points on the object would be in shadow). This would not be ideal as those points need to exist in the other setups. Instead, ambient occlusion is enabled. This lights the whole "world" therefore lighting the object in its entirety. At the bottom of the figure, an animation panel can be seen. This is created to capture all 100 images needed to create a dense EPI volume. The camera object position is key framed at each frame. This means that the camera moves every frame and the distance it moves can be controlled precisely between each frame. 100 frames were allocated but this can be increased.

It should be noted that Blender does not create equal distances between camera movements in animation mode but speeds up and slows down movement to make the camera movement smoother for video playback purposes. For precise camera positions these need to be added manually or programmed as done in these image acquisitions. The object properties for initial tests were chosen to be diffuse, this was to provide a good starting point to test the current implementation. The object colour was changed to contrast against the background, any colour can be chosen, in this case it was blue. "World" properties in Blender decide the physics of the world. There are many parameters that can be changed to affect how light interacts with objects. For the first few tests, reflections were eliminated from the scenes with everything being set to diffuse. In subsequent tests shown later in this paper the world was set to model real life lighting physics including reflections, subsurface scattering etc.. Finally, the images were rendered using "cycles" instead of "eevee". The differentiation between the two is that cycles use ray tracing (real light path rendering) to achieve a more real life image whereas eevee does not use ray tracing and only tries to follow physically based principles [46].

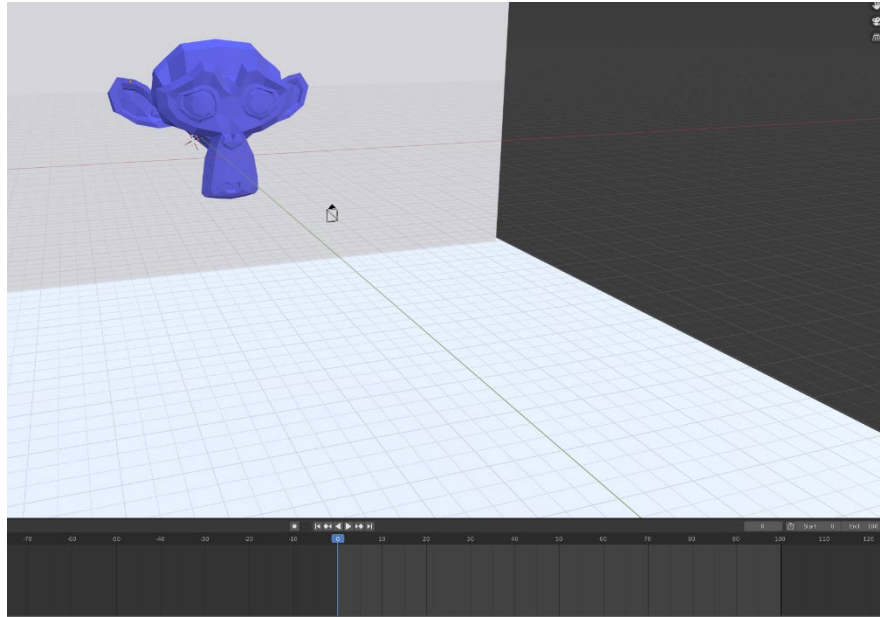


Figure 13 – Blender representation EPI setup

For the Helmholtz setup the blender representations are very similar, however the camera and light change positions. Below in figure 14 you can see the Blender representation of Helmholtz setup.

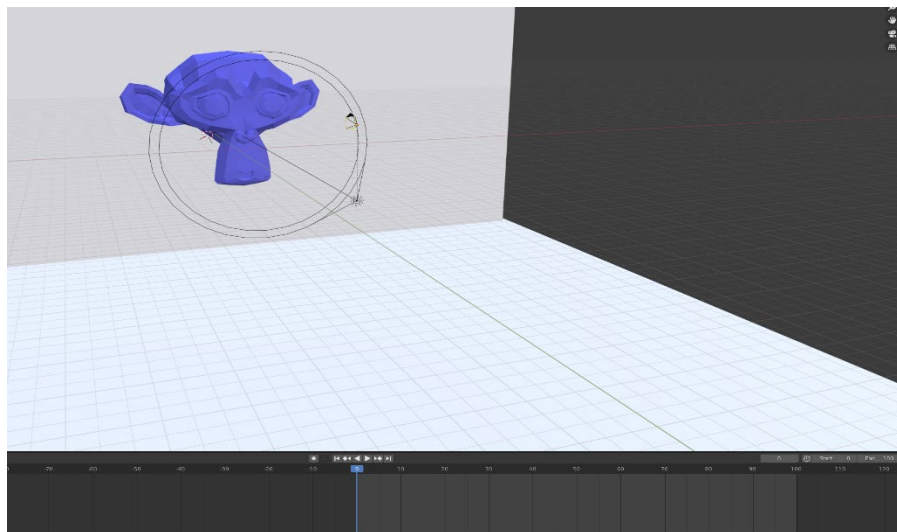


Figure 14 – Blender representation of Helmholtz setup, showing a camera (black pyramid) and spotlight (black rings)

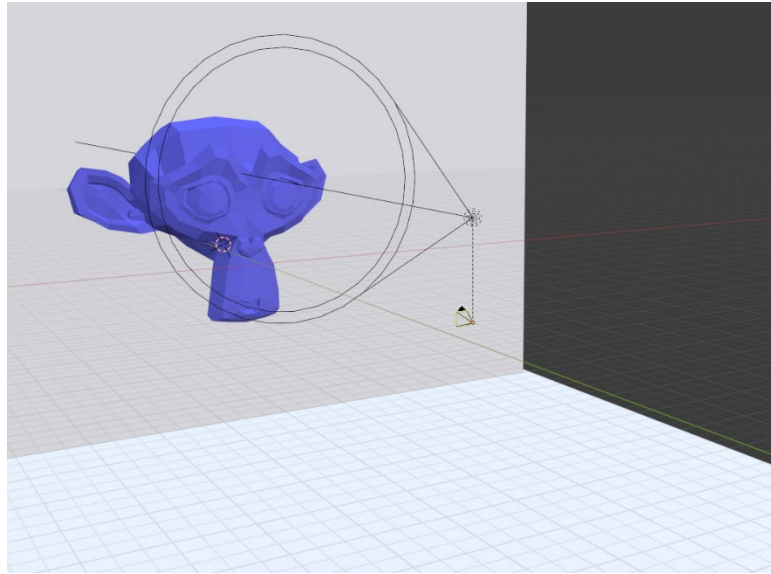


Figure 15 – Blender representation of opposite Helmholtz pair to figure 14

In figure 15 it can be noted that the position from figure 14 of the camera has been replaced by a light source. To achieve a pair of images for the Helmholtz setup shown in figure 14 and 15, the camera and light are swapped precisely using their coordinates as positions to achieve Helmholtz reciprocity outlined in section 2.5. The light is tied to the camera as a “child” object. This means that when the camera moves the light also moves. Therefore, when all images are rendered for the Helmholtz setup the images will be pairs of each other. The light object chosen was “spotlight” this was to ensure that the largest area of the object would be lit. This is to reduce occlusion issues found to limit 3D reconstruction as stated in the literature review section 2.5 specifically in Zickler et al.’s paper [7]. Ambient light was turned off for the Helmholtz setup as Helmholtz reciprocity requires only active illumination to achieve the constraints as explained at the beginning of this section. The starting positions for the cameras/lights in the Helmholtz setup and the EPI setup are in the same positions, to match the point positions on the object when the depth estimation is needed for Helmholtz reciprocity calculations. Only one pair of the Helmholtz setup needs to match the EPI setup as all positions of lights and cameras are known. This allows for the depth estimation to be calculated geometrically for the second Helmholtz pair.

After the images were created, they were used to test the EPI software detailed in section 4.1. These tests showed issues in the EPI software itself which are detailed in section 4.1 with their solutions. After errors were fixed within the code the images were tested again. Results from the EPI software were better, but still not as advertised in Kim et al.’s [4] paper. As no more errors could be found in the software the images were manipulated to check if better results could be obtained. Several different images were created and were tested with the EPI software.

These parameters were:

- distance of object to the camera
- reflection of the object
- texture of the object

The main variable that had the biggest effect on the result was the material of the object, and its texture. The results of these tests are detailed in section 4.2.2 below.

4.2.2 Results

From the two setups described above a total of 300 images were created: 100 images for the EPI setup and 200 images for the Helmholtz setup as they are pairs. The images show the differences in each setup, for example figure 16 shows the EPI setup as shown by its full illumination of the object.

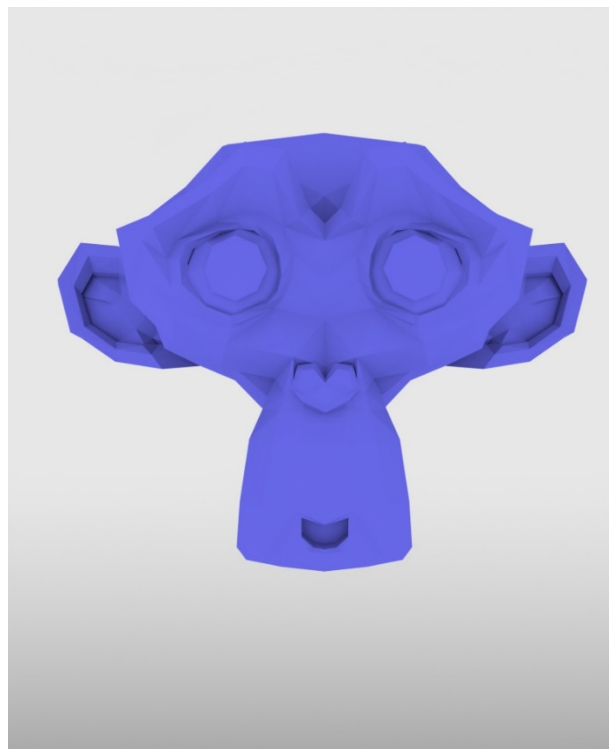


Figure 16 – EPI Blender image render

Figure 17 A and B show the Helmholtz setup pairs. These images were the first images rendered in each animation frame. This means that the camera/light positions for Helmholtz and EPI renders are in the same positions. This is shown in figure 17 where the object has been displaced in the image collected compared to figure 16 and 18.



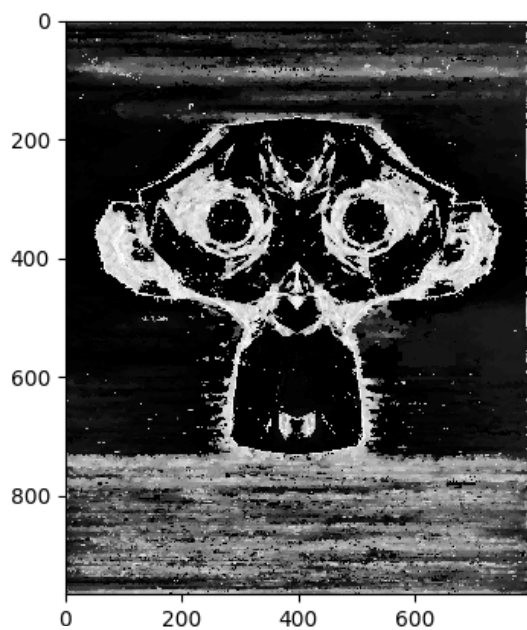
17 - A



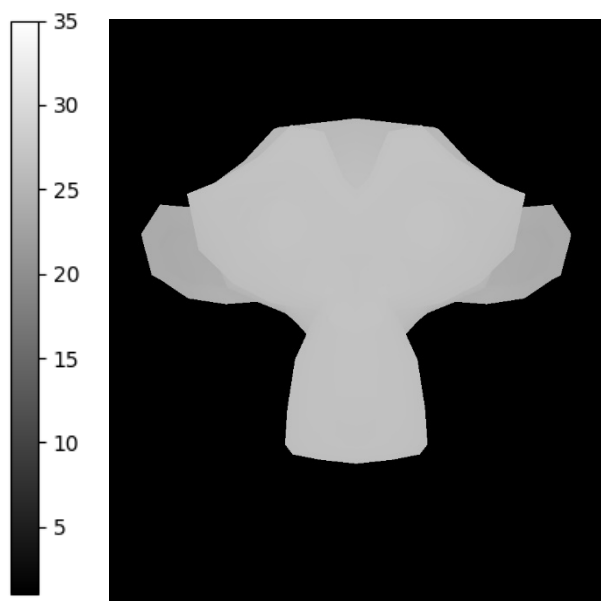
17 - B

Figure 17 – A) Helmholtz Blender render with camera in up position - B) Helmholtz Blender render with camera in down position

As noted in section 4.2.1 these images were tested using the EPI software. One hundred disparity maps were calculated. From these 100 disparity maps the 50th one is shown below in figure 18A. The 50th image is shown because normally the middle disparity image is the best out of the 100.



18 - A

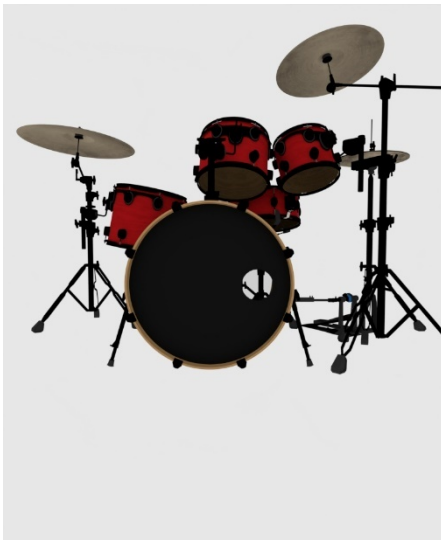


18 - B

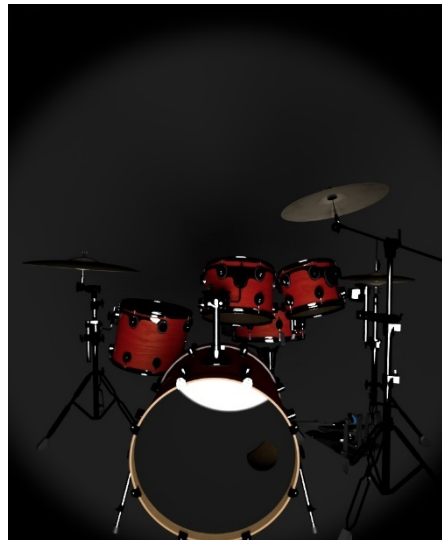
Figure 18 – A) Blue Monkey disparity map - B) Blue Monkey true depth map

From figure 18 above you can see that it is in black and white. Pixels that have a larger disparity (displacement) from one image to another have a whiter colour assigned to them and pixels with lower disparities have a darker colour assigned to them. The key (bar with numbers from 1 to 35) in figure 19 on the right shows the colour of a pixel regarding its disparity change. As stated in section 4.1 the resulting disparity map is not quite accurate as the monkey has mostly been outlined in white indicating that the black parts on the monkey's face are further away. An average pixel difference between the true map and the calculated map was calculated to compare all of the disparity maps calculated in this section. The average pixel difference for figure 18A compared to 18B was 51.61.

To test different objects and textures, more datasets were created. A total of 1200 images including the monkey images shown above were created. Three different objects were used, namely a reflective drum set, a chair with reflections and the monkey with and without a texture. These can be seen below with their respective EPI and Helmholtz image acquisition and their resulting disparity map calculation.



19 - A



19 - B



19 - C

Figure 19 – A) Drum set EPI Blender render - B) Drum set Helmholtz Blender render (camera in up position) - C) Drum set Helmholtz Blender render (camera in down position)

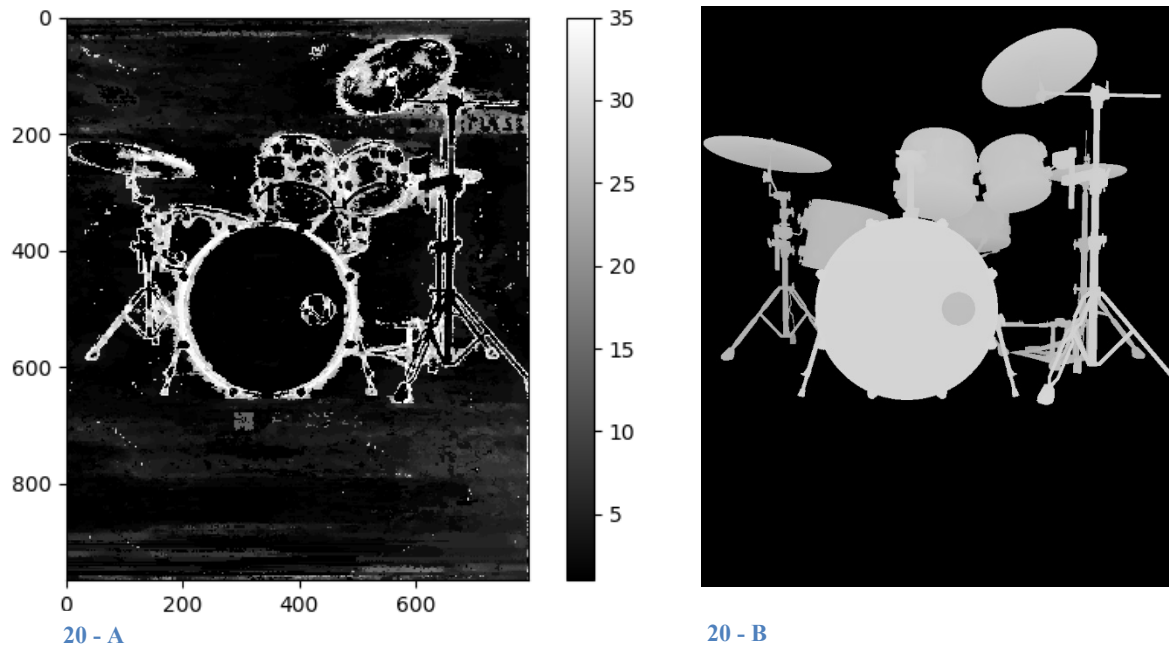


Figure 20 – A) Drum set EPI disparity map - B) Drum set true depth map

The drum set disparity map shown in figure 20A worked well with some inner sections of the drums being white, showing that the disparity has been calculated for those sections. The big bottom drum (foot drum) of the drum set is reflective as shown in figures 19A and 19B. This may have affected the EPI disparity calculation as stated in the literature review in section 2.3. The mean pixel difference between the true depth map in figure 20B and the calculated depth map in figure 20A is 47.31.



21 - A



21 - B



21 - C

Figure 21 – A) Textured Monkey EPI Blender render - B) Textured Monkey Helmholtz Blender render (camera in down position) - C) Monkey Helmholtz Blender render (camera in up position)

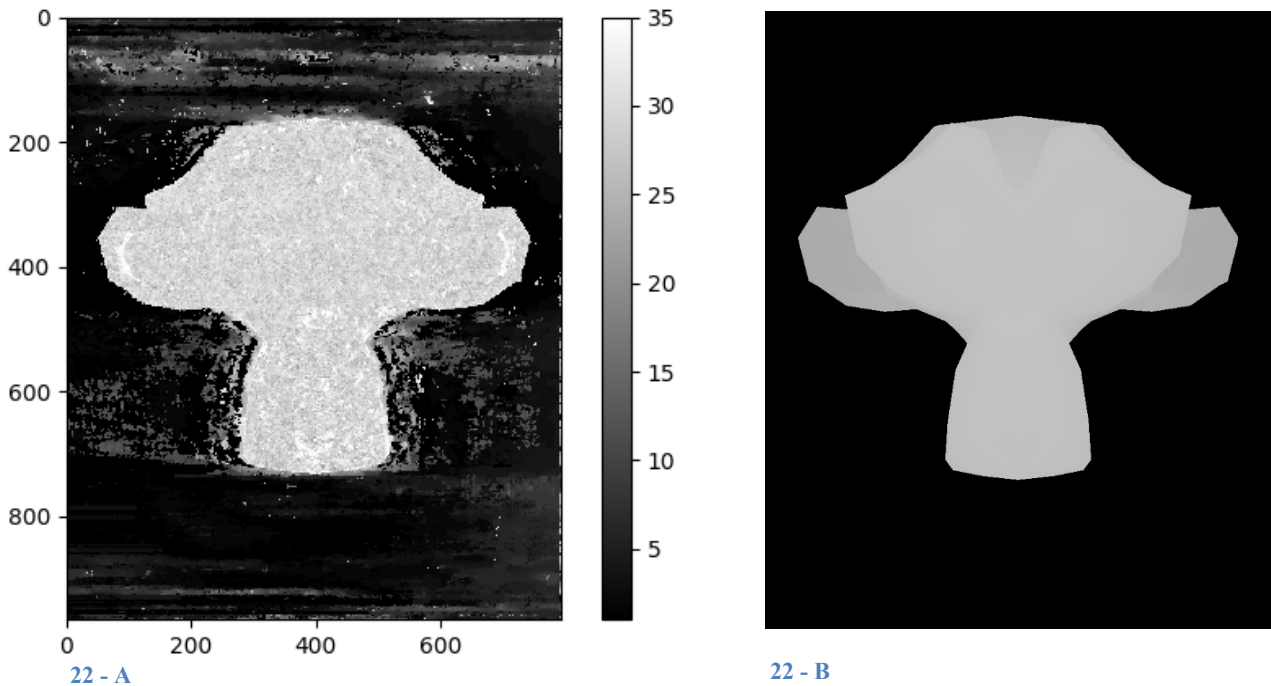


Figure 22 - A) Textured Monkey EPI disparity map - B) Textured Monkey true depth map

The textured monkey disparity map in figure 22A was the best so far with all of the monkey having disparity calculations attributed to it shown in white. This contrasts with the blue monkey shown in figure 18A where sections are clearly missing, shown in black. Texture helps in the EPI volume as the slanted lines mentioned in section 2.3 are more distinguishable from one another. As there are more lines there are therefore more gradients to be calculated and therefore more points on the object that have disparity values. The mean pixel difference between the true depth map in figure 22B and the calculated depth map in figure 22A is 44.01.



Figure 23 - A) Chair EPI Blender render - B) Chair Helmoltz Blender render (camera in down position) - C) Chair Helmoltz Blender Render (camera in up position)

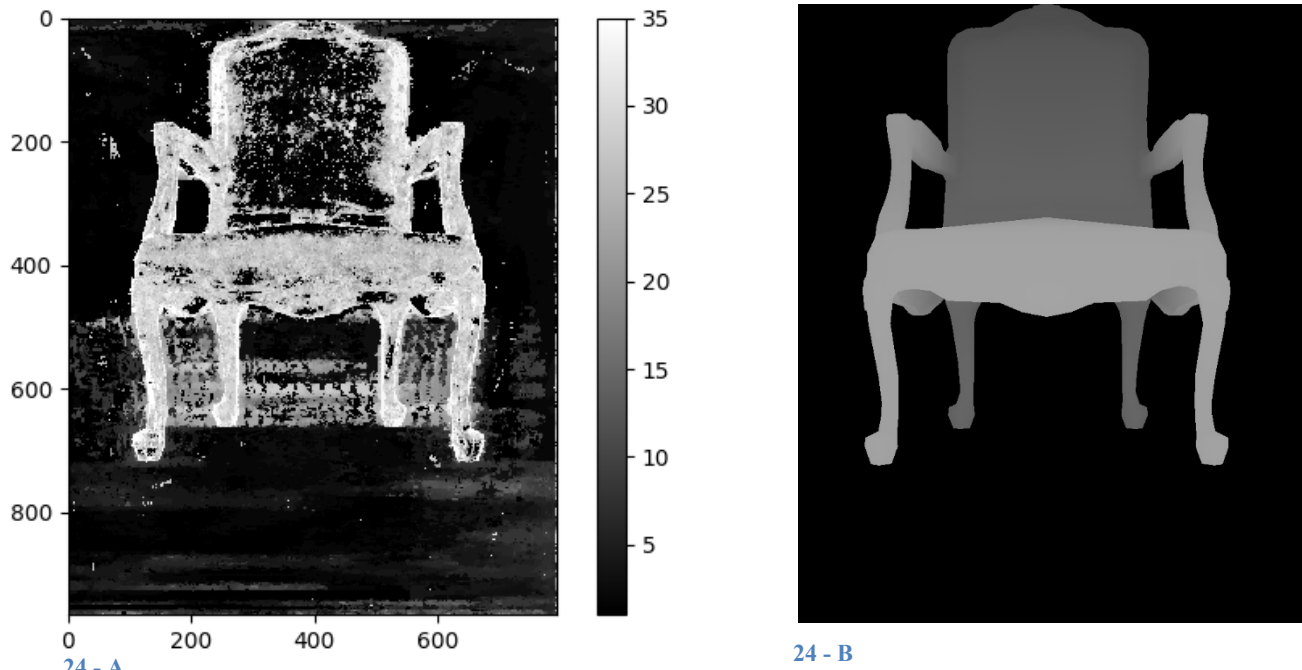


Figure 24 - A) Chair EPI disparity map - B) Chair true depth map

The chair disparity map, figure 24A was also better than the blue monkey disparity map and the drum set disparity map shown in figure 18A and figure 20A respectively. This is shown by the increased points on the object that have disparity values. The mean difference of the pixel values was also calculated with a difference of 31.96. This means that each pixel intensity value in figure 24A is on average 31.96 less than the true value.

In all disparity maps mentioned the background was ignored as the core focus was the object in question and not the reconstruction of the background.

Table 3 below summarises the average difference in pixel value for the disparity maps compared to their true depth map with the best disparity map

Table 3 – Average pixel value differences from calculated disparity map to true disparity map

Disparity Map	Average Pixel Difference from True Value
Blue Monkey	51.61
Drum Set	47.31
Textured Monkey	44.01
Chair	31.96

Initially higher resolution images of 1277x1269 were used as shown in figure 25. However, this required collecting multiple datasets and including testing, was extremely time consuming. The time taken to render these images was 60 minutes and the time to calculate the EPI disparity from 100 views took 129 minutes. Although these datasets showed improvements a decision was taken to use lower resolution images in order to manage the project timeline.

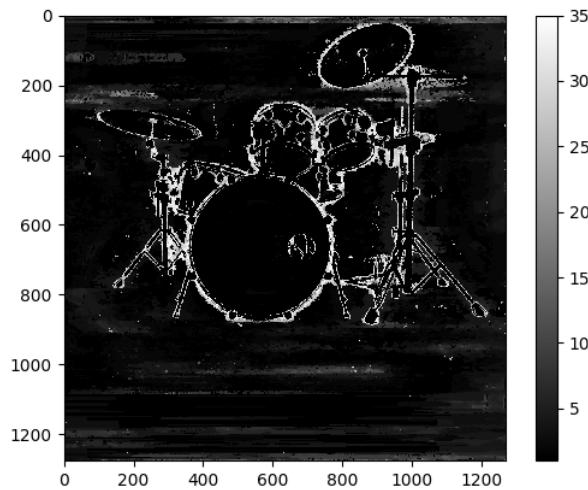


Figure 25 - Drum set high resolution EPI disparity map

5 USING HELMHOLTZ AND EPI APPROACH

5.1 Introduction

As stated in section 3, the second approach outlined in section 3.3 was used in this paper. This was chosen over the first approach in section 3.2. The first approach was declared to be out of scope as it would be too time consuming because details of the theory behind the approach would need to be studied further.

Therefore, in section 5.2 an outline is given about how the approach was taken using Python [43]. This is followed by issues encountered in section 5.3 and their solutions. Results of the approach are then shown in section 5.4 with their evaluation.

5.2 Approach taken

Firstly, synthetic images were created using Blender [44] this is shown in detail in section 4.2. The EPI images were used in the third-party EPI software from Kim et al's paper [4] detailed in section 4.1. This gave a total of 100 disparity maps for every 100 EPI images in that dataset. The dataset that was first used was the blue monkey shown in figure 16 and 17 in section 4.2.2. The disparity map values from the EPI software in Kim et al.'s paper [4] stores all the disparity values for every image in a high-performance data management and storage suite or hdf5 file in short. An hdf5 file stores

data in a hierarchical manner which is very similar to paths of files, using backslashes “\” to separate groups. Using one of the command line executions in Kim et al.’s software [4], namely “imgs2lf” (images to lightfield) mentioned in section 4.1 the Helmholtz pair images of the blue monkey shown in figure 17 and 18 were stored into hdf5 files to be imported into Python [43].

Python was chosen as the programming language because Kim et al.’s software [4] was in Python. This meant that if anything was to be integrated, as thought of for the first approach shown in section 3.2, then it would be a better approach to use the same language. For an integrated development environment (IDE), PyCharm [47] was chosen because it is free, easy to use, updated regularly and can run command line execution within the IDE.

The coordinates (x,y,z) of the camera positions were also needed as stated in section 2.5 for Helmholtz reciprocity. These were written into Excel [48]. The coordinates from Blender [44] are in inches and not millimetres. To satisfy the constraint in equation 4 in section 2.5 the coordinates and measurements need to be in the same system, either metric or imperial. They were therefore converted to millimetres before being copied into a text file which would later be inputted into the code.

Three scripts were made in the approach. These were main.py, CalculateNormal.py, and ShowNormalMap.py. The main functions of these scripts are shown in table 3 below.

Table 4 – Script input, calculation, and outputs

Script Name	Input/s	Calculation/s	Output/s
main.py	hdf5 lighfield files for Helmholtz images, the hdf5 file for the EPI disparity, and coordinates of cameras and lights in the Helmholtz images.	It then takes this information calculates the x,y,z coordinate of every pixel and uses all the information above to create W matrices mentioned in section 3.3	A matrix in the shape of (image width, image height, W Matrix, w vector).
CalculateNormal.py	Takes the Matrix containing all the W matrices calculated in main.py as an input.	Uses SVD to find the 3 rd right vector mentioned in section 2.5 of each W Matrix.	Outputs a matrix of shape (image width, image height, 3) which contains the normal vector of each pixel.

ShowNormalMap.py	Takes the normal matrix calculated in CalculateNormal.py as input.	Encodes the normal vectors into RGB values.	Outputs a Matrix of size (image width, image height, 3) which is a normal map image that is then displayed.
------------------	--	---	---

As detailed in table 3 above, main.py takes the required files and image data and calculates the x, y, and z coordinate of each pixel. To do this the disparity values from the EPI calculation are converted to depth values using equation 6 below.

$$depth = (baseline * focal\ length) / disparity \quad \text{Equation 6}$$

With baseline being the distance between the camera movements for each subsequent image in the EPI image dataset. To achieve the depth in millimetres the focal length needs to have a unit of pixels in order to return a result of $mm = (mm * pixels) / pixels$ cancelling the pixel unit leaving millimetres. Blender [44] gives focal length in millimetres so a conversion was made using equation 7 below.

$$focal\ length\ (pixels) = (focal\ length(mm) / sensor\ width\ (mm)) * image\ width(pixels) \quad \text{Equation 7}$$

From this the depth of every pixel can be found and stored into a matrix array. By using the field of view (FOV) of the camera, the centre position of the camera, the shape of the images taken and the depth values an x, y, and z coordinate can be found for each pixel through similar triangles. This can be done because the camera is modelled as a pinhole camera [49]. A function to do this was found in the Open3d library [50] that transforms a depth map into a 3D point cloud.

The script then iterates through every pixel in the four dimensional matrix of shape (image width, image height, image dataset, depth (x, y, z)) and uses equation 4 shown in section 2.5 to create a w vector for each pixel in the dataset and stacks them to form a new matrix of shape (image width, image height, W matrix, w vector). This matrix is saved for use in the CalculateNormal.py script.

The matrix is read into the CalculateNormal.py script which iterates over every pixel and calculates the SVD of the W matrix. The third matrix calculated is then transposed so that the 3rd column vector is now on the 3rd row, this enables it to be copied into a new matrix of shape (image width,

image height, normal vector). The right 3rd singular vector is taken as this corresponds to the normal constraint shown in equation 4 in section 2.5. The matrix is then saved to be used in the ShowNormalMap.py script.

A normal map is inputted into the ShowNormalMap.py script that it reads in. It then iterates through the normals of the map and encodes them into RGB values. RGB values range from 0 to 255 while the normal vectors can range from -1 to 1. To encode them the equations 8, 9, and 10 were used shown below. The x coordinate of the normal is encoded to the red channel, the y coordinate is encoded to the green channel and the z coordinate is encoded to the blue channel.

$$R = \left(\frac{(x+1)}{2} \right) * 255 \quad \text{Equation 8}$$

$$G = \left(\frac{(y+1)}{2} \right) * 255 \quad \text{Equation 9}$$

$$B = \left(\frac{(z+1)}{2} \right) * 255 \quad \text{Equation 10}$$

This RGB image was then saved as a .png file to view.

5.3 Issues and Solutions

Two issues were encountered while programming these scripts. These issues are detailed below.

The first issue encountered was that normal maps were inverted so that points on the object that should protrude were concave and vice versa. This was a simple issue and was solved by flipping the z coordinate using a negative sign in front of it. This means that if the vector was facing away as shown in figure 26 it will now face towards the viewer as shown in 30A and vice versa.

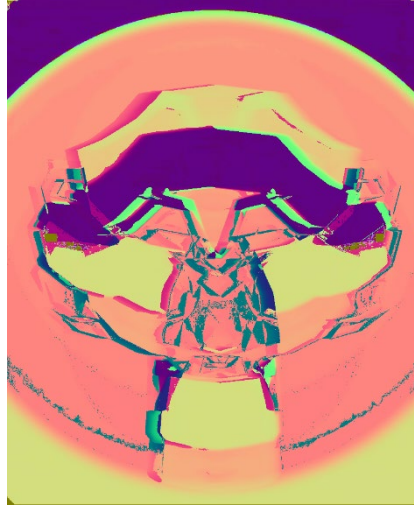


Figure 26 – Inverted Monkey normal map

The second issue encountered was that the script was not matching object points between the Helmholtz pairs. This meant that two normal object vectors would be shown. This can be seen in figure 27 as a purple and a yellow chair are visible. To mitigate this issue the images were cropped on the top and bottom so that the Helmholtz images would align. This can be done because of vertical epipolar plane scanlines detailed in section 2.3 figure 1. Retrospectively a better solution would be to use images without a background, allowing for the objects to be masked.



Figure 27 - Double Chair object normals

5.4 Results and Evaluation

The following section shows results obtained from the code on all the datasets detailed in section 4.2 followed by an evaluation of those results.

From the chair dataset highlighted in section 4.2.2 figure 23B and 23C the following RGB normal map was created in figure 28A and is shown next to its true normal map in figure 28B. The true normal map was calculated in Blender [5]. The heat map is also shown in figure 28C and was

calculated by taking the angle between the normal vectors of the calculated map compared to the true map. The whiter the pixel colour the smaller the angle between the normal vectors. This means that the whiter the pixel the more accurate the normal calculation.

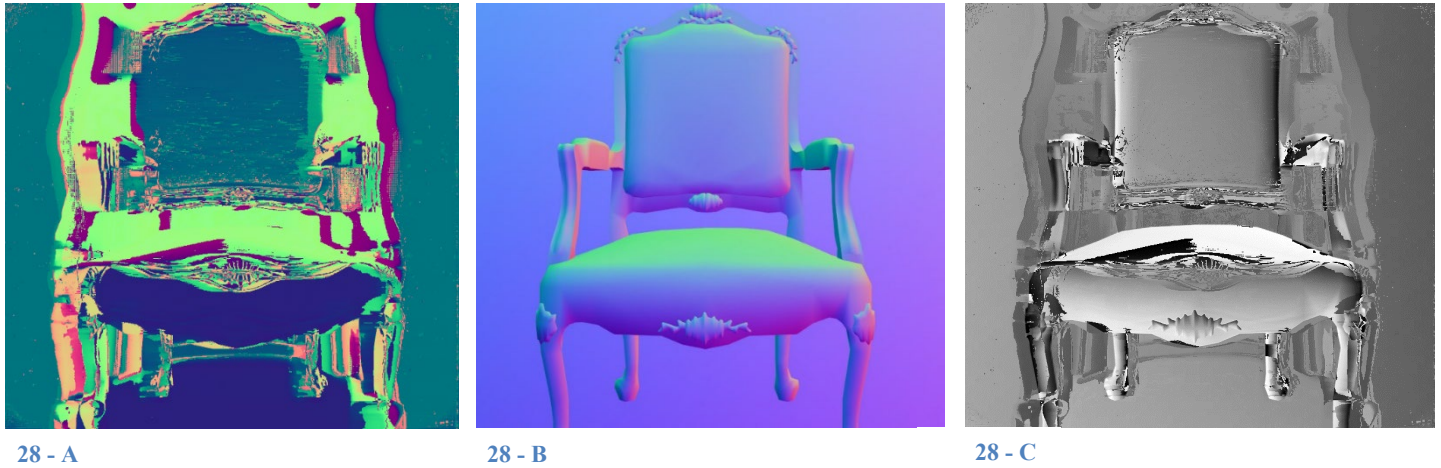


Figure 28 - A) Chair calculated normal map - B) True Chair normal map - C) Chair heat map

As shown in figure 28A the calculated normal map is not perfect. This discrepancy can be explained by two issues. Firstly, the bottom of the back and seat of the chair were occluded in one of the Helmholtz Stereopsis paths as shown in figure 23B in section 4.2.2. As stated by Zickler et al. [7] Helmholtz Stereopsis struggles when occlusions are present. Therefore, the normals have only been calculated for one of the Helmholtz camera paths. Secondly, the matching of points on the object is not exact. This is shown below in figure 29A and B. This could either again be because of occlusions, or that the matching of object points within the programming is incorrect. The heat map in figure 28C showed a mean difference in angles of normals to their true values of 57.96 degrees with a standard deviation of 6.80. This was the lowest standard deviation of all the images. This shows that overall normal difference values had the least variance of 46.27 as shown in table 5 below.

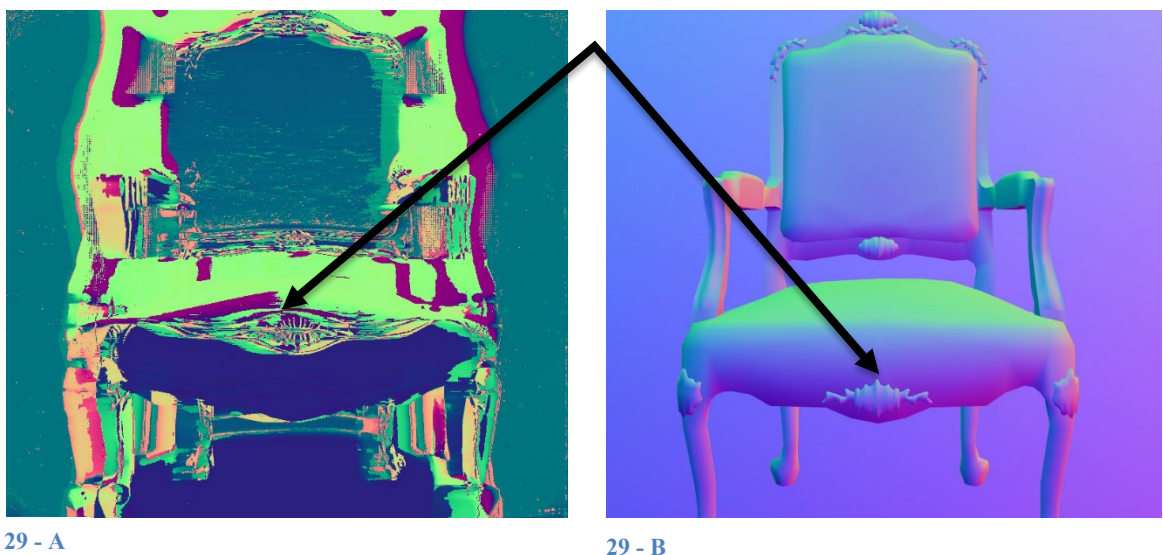


Figure 29 – A) A copy of figure 28A and B, highlighting point discrepancies

From the textured monkey dataset highlighted in section 4.2.2 figure 21B and 21C the following RGB normal map can be seen in figure 30A, with the true normal map shown in 30B and the heat map in 30C.



Figure 30 – A) Textured Monkey calculated normal map – B) True Monkey normal map – C) Textured Monkey heat map

The normal map shown in figure 30B is not the true normal map for the textured monkey. This is because an already made texture was used from Poly Haven [45]. This includes a normal map which was applied to the monkey to give it texture. The normal map for this texture can be seen below in figure 31. This could be the reason why there is a textured appearance of the normals. Nevertheless, a heat map was calculated from 30B with a standard deviation of 9.98 and a mean angle difference of 44.41 as shown in table 5 below.

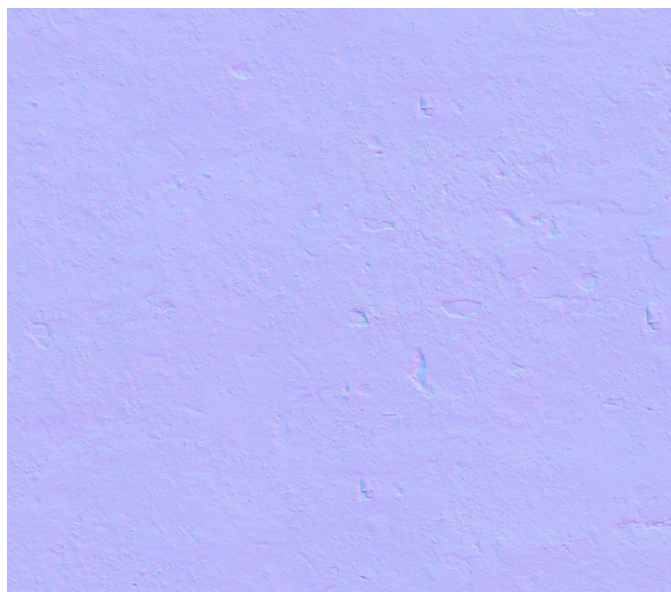


Figure 31 – Poly Haven [45] normal map

From the drum set dataset highlighted in section 4.2.2, figure 19B and 19C, the following RGB normal map and true normal map can be seen below in figure 32A and 32B respectively. The heat map is also shown in figure 32C.



32 - A

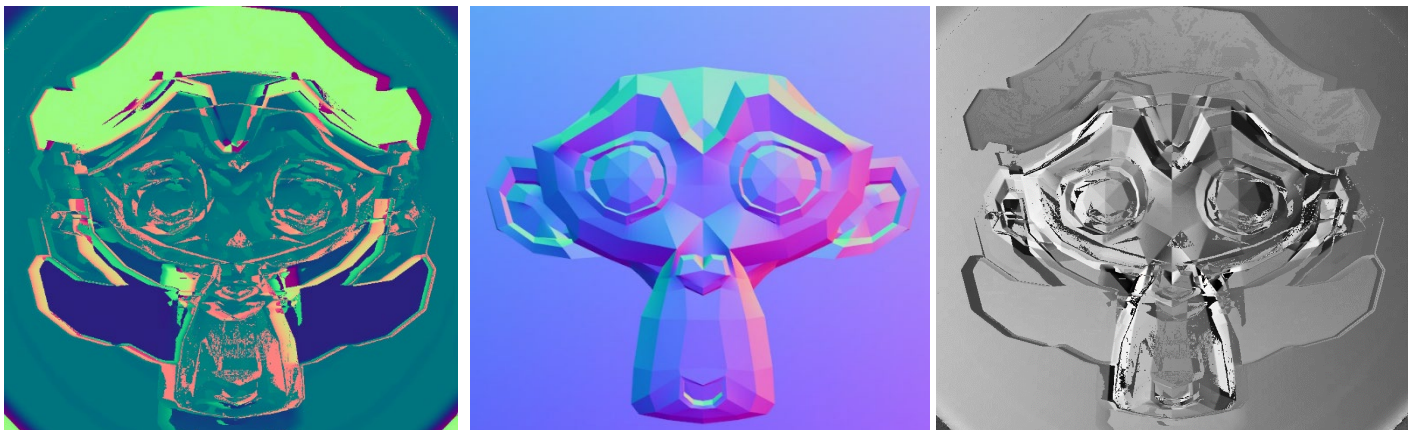
32 - B

32 - C

Figure 32 – A) Drum set calculated normal map – B) True Drum set normal map – C) Drum set heat map

The drum set normals calculated again suffered from occlusions and point matching mentioned previously. As can be seen in the bottom middle of figure 32A, the foot drum face is reflective. As the approach taken in this project uses EPI depth calculations before using Helmholtz Stereopsis, the method still suffers from reflections as detailed in section 2.3. The heat map shown in 32C had a standard deviation of 10.58 and a mean angle difference of 47.37 as shown in table 5 below.

From the blue monkey dataset highlighted in section 4.2.2, figure 17A and 17B, the following RGB normal map can be seen in figure 33A with its true normal map shown in figure 33B. The heat map is also shown in figure 33C.



33 - A

33 - B

33 - C

Figure 33 - A) Blue Monkey normal map calculated - B) True Blue Monkey normal map – C) Blue Monkey heat map

The Blue Monkey heat map had the best mean angle difference of 42.49 and had the second-best standard deviation of 9.55 as shown below in table 5.

Table 5 – Normal angle differences from true values

Object	Mean (degrees)	Median (degrees)	Standard Deviation	Variance
Chair	46.10	45.94	6.80	46.27
Textured Monkey	44.41	43.47	9.98	99.52
Drum set	47.37	47.89	10.58	111.95
Blue Monkey	42.49	41.59	9.55	91.25
Average over all objects.	45.09	44.72	9.22	87.25

Table 6 – Disparity quality vs standard deviation

Object	Disparity Quality (lower = higher quality)	Standard Deviation of Normal Differences to True Value
Chair	31.96	6.80
Textured Monkey	44.01	9.98
Drum set	47.31	10.58
Blue Monkey	51.61	9.55

From table 6, it should be noted that the Textured Monkey values cannot be compared to the other objects. As stated above a normal map was placed as a texture on top of the object. This caused the resulting normals to be calculated from that map. However, the calculation of the standard deviation is from a normal map of the standard monkey object in Blender.

As shown in table 6 the Chair object had the best disparity map calculation and standard deviation of normal differences. This could indicate that the quality of the disparity map is correlated to the quality of normals calculated. Disparity quality of the Drum set and Blue Monkey were very similar and therefore, so were their standard deviations. The Drum set standard deviation was slightly higher than that of the Blue Monkey. This could have resulted from the Drum set having more occlusions

and reflections.

All of the normal maps shown in figures 28A, 30A, 32A, and 33A seemed to suffer from the same two issues: occlusions and point matching. Time permitting these issues would have been better mitigated. Firstly, a better algorithm approach could have been explored to match points between the images, taking advantage of epipolar planes. Secondly, the distance between the camera and the light pairs in the Helmholtz setup described in 4.2.2 could have been reduced. This would decrease the size of occlusions present when the camera and light swap positions.

The results achieved were not perfect, however this was expected as the depth maps used in the normal calculations were inaccurate, as discussed in section 4.1 and 4.2. It is evident that if better depth maps were used, as well as the improvements stated above, this novel approach could have provided significant advantages over current methods.

It must also be noted that as the background was taken into account when calculating the heat maps, the standard deviations, means, medians and variances will be higher than their real values.

6 CONCLUSION

At the start of this project a review of available literature identified a gap in the current approaches to passive 3D reconstruction. The present state-of-the-art passive 3D reconstruction methods assume the BRDF of the objects they are constructing as Lambertian. This leads to errors in the final reconstruction as most objects are non-Lambertian. One approach that does not assume the BRDF is Helmholtz Stereopsis which takes advantage of Helmholtz reciprocity to calculate the normals of objects whose BRDF is unknown or arbitrary. As a result of this research, the project detailed in this paper was undertaken with the goal of improving upon current passive methods to reconstruct complex materials.

This project, therefore proposed two novel approaches to achieve this goal. Both concentrated on combining Helmholtz Stereopsis and epipolar plane image volumes to develop a new approach. The first approach detailed a method of creating 4D epipolar plane image volumes from Helmholtz Stereopsis. However, this approach was not followed as it was evident that further research was needed to find similarities in the 3D epipolar plane slice to make it functional as a method, which would have been outside of the time scale for the project and so was put out of scope. The second approach therefore formed the focus of the project which used the epipolar plane images to find depth points for use inside of the Helmholtz Stereopsis equations instead. This was useful as the Helmholtz reciprocity equation has two unknown variables, depth and normal. By eliminating the unknown depth, only one pair of Helmholtz Stereopsis images was needed to meet the constraint of this equation and

calculate the normals of the object.

For the reasons stated above the second approach was chosen as the project method to achieve the goal of improving upon current passive 3D reconstruction methods. Given that a novel approach was used there were no available image datasets that met the specifications. Typically, Helmholtz images can be collected in any way so long as there are reciprocity pairs. However, once epipolar image plane volumes are combined, random pairs will not work as depths from the epipolar plane images will not match the points within the Helmholtz images. To overcome this challenge this project invented a novel image acquisition technique to collect Helmholtz reciprocity pairs along a linear path. By using this novel technique within Blender this project created a dataset of 1200 synthetic images.

Using a third-party epipolar plane image technique, depths were calculated, and their quality quantified by taking the difference in pixel value from their true value. The Chair object had an average pixel difference of 31.96 as the best, the Textured Monkey had a value of 44.01, followed closely by the Drum set at 47.31 and finally the Blue Monkey with 51.61. These depths calculated were then used within software, created by this project, to calculate the Helmholtz object surface normals. These were then converted to RGB and plotted. Heat maps were then calculated from the angle difference of the calculated normals to the true normal angle values. This gave a mean angle difference and a standard deviation. For the Chair object the mean was 46.10 with a standard deviation of 6.80. For the Textured Monkey a mean of 44.41 with a standard deviation of 9.98. For the Drum set a mean of 47.37 with a standard deviation of 10.58 and for the Blue Monkey a mean of 42.49 with a standard deviation of 9.55. However, it is noted that these standard deviations and means are higher than their true value as the background was considered when calculating the normal angle differences.

These results were not perfect due to occlusions, reflections, the disparity maps calculated, and the object point matching algorithm used. However, it is evident that the combination of Helmholtz Stereopsis and epipolar plane images can form a novel approach to extract surface normals of objects. Therefore, had more time been available, this approach could have been improved further to yield better results than existing methods. In addition, this paper also proposed another novel method that integrates Helmholtz Stereopsis into epipolar plane image volumes that has the potential to improve the results achieved in this paper by allowing surface normal and depth information extraction of arbitrary and unknown bidirectional reflectance distribution functions.

6.1 Evaluation

The overall goal's success is based on the primary objectives. The evaluation of the success of these

objectives is discussed below.

The first objective was to ensure that the epipolar plane image volume software, in this case taken from Kim et al.'s [4] paper, could be used to model 3D objects. As highlighted in section 4.1.1 and 4.2.2 the software was shown to be able to produce disparity maps and therefore objective one was achieved successfully.

The second objective was to combine two approaches namely, the Helmholtz Stereopsis approach and the epipolar plane image approaches. As highlighted in section 3.3 a novel approach was created from the combination of these two methods. This novel approach was tested using a novel dataset of 1200 images created within the project, from which surface normals were successfully extracted as evidenced in section 5.4. Therefore, objective two was achieved.

The third and fourth objectives were to create at least two models of Lambertian objects and at least two models of non-Lambertian models. These two were successfully completed as shown in section 5.4 with the two non-Lambertian objects being the Drum set, and Chair and the Lambertian objects being a Textured Monkey and a non-textured Blue Monkey.

The overall goal of the project was to form a novel approach by using a combination of High Spatio-Angular Resolution Light Fields and Helmholtz Stereopsis to improve upon high quality 3D reconstructions of models with complex materials. Considering that all primary objectives were achieved the overall goal was met despite the results not being perfect. The reasons for this has already been noted as being due to occlusions, reflections, the disparity maps calculated, and the matching point algorithm used. Notwithstanding, it is evident that the combination of Helmholtz Stereopsis and High Spatio-Angular Resolution Light Fields offers a novel approach to improve upon high quality 3D reconstructions of models with complex materials and if polished beyond this project has significant potential to improve upon existing passive 3D reconstruction methods.

6.2 Future Work

This paper provided a solid insight into leveraging the complementarity of Helmholtz Stereopsis and epipolar plane images by using the advantages of both methods to create a novel approach for 3D reconstruction. Some issues were faced with the resulting normal maps. Therefore, future research would be recommended to use a different EPI software to improve the quality of the depth maps used in the calculation of normals. Furthermore, it is recommended that the approach be further polished by using better point matching algorithms, which possibly could lead to higher quality surface normal extraction.

Furthermore, this paper focused on the second approach detailed in section 3.3. Future research could consider exploring the first approach in section 3.2. This would eliminate the need for EPI

software entirely allowing for depth estimation and normal calculation through Helmholtz Stereopsis alone.

As a relatively new approach, Helmholtz Stereopsis provides many advantages as it does not require a BRDF to be known. Therefore, it would be interesting for future research to see if it could be combined with other methods in Computer Vision.

6.3 Self-Reflections

As mentioned previously this research had difficulties regarding the depth estimation software from Kim et al's [4] paper. Retrospectively, if more time was available tests could be done to find a more accurate EPI volume software to improve the results. Likewise, a better point matching algorithm would be used.

At the beginning of this research issues such as bugs within the code, time taken to calculate disparity images, and time taken to render high-resolution datasets were not anticipated to be as prevalent as they were. Therefore, retrospectively the timeline should have been changed to be more forgiving.

7 REFERENCES

- [1] S. Steinman, B. Steinman and R. Garzia, *Foundations of Binocular Vision: A Clinical Perspective*, First Edition ed., New York: McGraw-Hill Education, 2000, pp. 2-5.
- [2] A. Ogale and Y. Aloimonos, "Shape and the Stereo Correspondence Problem," *International Journal of Computer Vision*, vol. 65, pp. 147-162, 2005.
- [3] T. Hassner and R. Basri, "Example Based 3D Reconstruction from Single 2D Images," in *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, New York, USA, 2006.
- [4] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung and M. Gross, "Scene Reconstruction from High Spatio-Angular Light Fields," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 1-12, 2013.
- [5] B. Liu, S. Gould and D. Koller, "Single image depth estimation from predicted semantic labels," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, 2010. pp. 1253-1260.
- [6] A. Geiger, M. Roser and R. Urtasun, "Efficient Large-Scale Stereo Matching," in *10th Asian Conference on Computer Vision - ACCV 2010*, Queenstown, New Zealand, 2010, pp. 25-38.
- [7] T. Zickler, P. Belhumeur and D. Kriegman, "Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction," *International Journal of Computer Vision*, vol. 49, no. 2-3, pp. 215-227, 01 September 2002.
- [8] R. Zhen and R. L. Stevenson, "Motion deblurring and depth estimation from multiple images," in *IEEE International Conference on Image Processing (ICIP)*, Phoenix, Arizona, 2016, pp. 2688-2692.
- [9] T. Keating, P. Wolf and F. Scarpace, "An Improved method of digital image correlation," in *XIII International Congress of Photogrammetry*, Helsinki, 1975, pp. 993-1002.
- [10] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: theory and experiment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 920-932, 1994.
- [11] W. E. L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, no. 1, pp. 17-34, 1985.
- [12] University of Nevada, "Reno, Stereo Correspondence Problem," University of Nevada, Reno, Computer Science & Engineering, [Online]. Available: <https://www.cse.unr.edu/~bebis/CS791E/Notes/StereoCorrespondenceProblem.pdf>. [Accessed July 2021].
- [13] A. Fusiello, V. Roberto and E. Trucco, "Efficient stereo with multiple windowing," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, USA, 1997, pp. 858-863.

- [14] B. L. A. Y. D. Geiger, "Occlusions and binocular stereo," in *Proceedings European Conference Computer Vision*, Santa Margherita Ligure, Italy, 1992, pp. 425–433.
- [15] J. Mulligan and K. Daniilidis, "Predicting disparity windows for real-time stereo 1842:," in *Computer Vision - ECCV 2002. Lecture Notes in Computer Science*, vol. 1842, Berlin, Springer, 2000, p. 220–235.
- [16] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *International Journal of Computer Vision*, vol. 33, no. 3, pp. 181–200, 1999.
- [17] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proceedings Eighth IEEE International Conference on Computer Vision (ICCV)*, Vancouver, Canada, 2001, pp. 508–515, vol.2..
- [18] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algo-rithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, September 2004.
- [19] G. Graber, T. Pock and H. Bischof, "Online 3D reconstruction using convex optimization," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, 2011, pp. 708–711.
- [20] D. Liu, Y. Huang, Q. Wu, R. Ma and P. An, "Multi-Angular Epipolar Geometry Based Light Field Angular Reconstruction Network," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1507–1522, 2020.
- [21] R. Bolles, H. Baker and D. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Internationa Journal of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.
- [22] L. Huijin, G. Kaiyu, Z. Yongbing and i. D. Qiongha, "Light field depth estimation ex-ploiting linear structure in EPI," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Turin, Italy, 2015, pp. 1–6.
- [23] A. Criminisi, S. B. Kang, R. Swaminathan, B. Szeliski and P. Anandan, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," *Computer Vision and Image Understanding (CVIU)*, vol. 97, no. 1, pp. 51–85, 2005.
- [24] Y. Zhang, H. Lv, Y. Liu, H. Wang, X. Wang, Q. Huang, X. Xiang and Q. Dai, "Light-Field Depth Estimation via Epipolar Plane Image Analysis and Locally Linear Embedding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 739–747, 2017.
- [25] L. Kunyuan, J. Zhang, R. Sun, X. Zhang and J. Gao, "arXiv.org- Computer Vision and Pattern Recognition: EPI-based Oriented Relation Networks for Light Field Depth Estimation," Cornell University, 26 August 2020. [Online]. Available: <https://arxiv.org/abs/2007.04538>. [Accessed July 2021].
- [26] Y. Fang, K. Wei, J. Hou, W. Wen and N. Imamoglu, "Light Filed Image Quality Assessment by Local and Global Features of Epipolar Plane Image," in *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, Xi'an, China, 2018, pp. 1–6.
- [27] T. Amiza, E. Lubetzky and N. Kiryati, "Coarse to over-fine optical flow estimation," *Pattern Recognition* ,

vol. 40, no. 9, pp. 2496-2503, 2007.

- [28] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 807-814, 2005.
- [29] S. Wanner and B. Goldluecke, "Reconstructing Reflective and Transparent Surfaces from Epipolar Plane Images," in *Pattern Recognition - 35th German Conference, GCPR 2013*, Saarbrücken, 2013.
- [30] R. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images," *SPIE seminar on Image Understanding Systems and Industrial Applications*, vol. 133, no. 1, pp. 139-144, 1980.
- [31] E. R. Davies, "Chapter 16: The Three-Dimensional World," in *Machine Vision: Theory, Algorithms, Practicalities. Signal Processing and its Applications*, 3 ed., Morgan Kaufmann, 2005, pp. 445-485.
- [32] M. Li, C. Diao, W. Xing and D. Lu, "A non-lambertian photometric stereo under perspective projection," *Frontiers of information Technology & Electronic Engineering*, vol. 219, pp. 1191-1205, 2020.
- [33] G. Addari, J.-Y and Guillemaut, "An MRF Optimisation Framework for Full 3D Helmholtz Stereopsis," in *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP*, Prague, Czech Republic, 2019, pp. 725-736.
- [34] S. Magda, D. Kriegman, T. Zickler and P. Belhumeur, "Beyond Lambert: reconstructing surfaces with arbitrary BRDFs," in *Proceedings Eighth IEEE International Conference on Computer Vision*, Vancouver, Canada, 2001.
- [35] H. V. Helmholtz, *Treatise on Physiological Optics*, New York: Dover Publications, 1925.
- [36] P. Tu, P. R. S. Mendonça, J. Ross and J. Miller, "Surface Registration with a Helmholtz Reciprocity Image Pair," *IEEE Workshop on Color and Photometric Methods in Computer Vision*, 2003.
- [37] J.-Y. Guillemaut, O. Drbohlav, R. Sara and J. Illingworth, "Helmholtz Stereopsis on rough and strongly textured surfaces," *Conference: 2nd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, pp. 10-17, 2004.
- [38] H. Mori, R. Köhle and M. Kamm, "Scene Depth Profiling Using Helmholtz Stereopsis," in *Computer Vision – ECCV 2016*, Cham, Springer International, 2016, pp. 462-476.
- [39] M. Weinmann, R. Ruiters, A. Osep, C. Schwartz and a. R. Klein, "Fusing structured light consistency and helmholtz normals for 3D reconstruction," *BMVC Press*, pp. 1-12, 2012.
- [40] A. Delaunoy, E. Prados and P. N. Belhumeur, "Towards full 3d helmholtz stereovision algorithms," in *In Computer Vision - ACCV 2010 - 10th Asian Conference on Computer Vision*, Queenstown, New Zealand, 2010, pp. 39-52.
- [41] N. Roubtsova and J.-Y. Guillemaut, "Colour Helmholtz Stereopsis for Reconstruction of Complex Dynamic Scenes," in *2014 2nd International Conference on 3D Vision*, Tokyo, 2014. pp. 251-258.
- [42] M. Tuschen, "manuSrep/DisneyDispPy," GitHub, 20 October 2016. [Online]. Available:

- <https://github.com/manuSrep/DisneyDispPy>. [Accessed 21 04 2021].
- [43] Python, “Python,” Python Software Foundation, 04 March 2021. [Online]. Available: <https://www.python.org/>. [Accessed 04 March 2021].
- [44] <https://www.blender.org/>, “Blender,” <https://www.blender.org/>, 2021. [Online]. Available: <https://www.blender.org/>.
- [45] P. Haven, “Models,” Poly Haven, 20 March 2021. [Online]. Available: <https://polyhaven.com/models>. [Accessed 20 March 2021].
- [46] J. Lampel, “Cycles vs. Eevee - 15 Limitations of Real Time Rendering in Blender 2.8,” CGCOOKIE, 07 March 2019. [Online]. Available: <https://cgcookie.com/articles/blender-cycles-vs-eevee-15-limitations-of-real-time-rendering>. [Accessed 04 August 2021].
- [47] Pycharm, “Pycharm,” Jet Brains, 10 March 2021. [Online]. Available: <https://www.jetbrains.com/pycharm/>. [Accessed 10 March 2021].
- [48] Office, “Office,” Microsoft, 25 February 2021. [Online]. Available: <https://www.office.com>. [Accessed 25 February 2021].
- [49] yodayoda, “From depth map to point cloud,” Medium, 24 September 2020. [Online]. Available: <https://medium.com/yodayoda/from-depth-map-to-point-cloud-7473721d3f>. [Accessed 19 August 2021].
- [50] Open3d, “open3d.geometry.create_point_cloud_from_depth_image,” Open3d, 03 March 2019. [Online]. Available: http://www.open3d.org/docs/0.7.0/python_api/open3d.geometry.create_point_cloud_from_depth_image.html. [Accessed 30 August 2021].