SPATIAL AUGMENTED REALITY IN OUTDOOR AREAS USING LONG-RANGE VISION SYSTEMS

BEHNAM MANESHGAR

A THESIS
IN
THE DEPARTMENT
OF
COMPUTER SCIENCE AND SOFTWARE ENGINEERING

PRESENTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF COMPUTER SCIENCE
CONCORDIA UNIVERSITY
MONTRÉAL, QUÉBEC, CANADA

JUNE 2017

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: **Behnam Maneshgar**
Entitled: **Spatial Augmented Reality in Outdoor Areas using Long-Range Vision Systems**

and submitted in partial fulfillment of the requirements for the degree of

Master of Computer Science

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

Dr. Yuhong Yan _____ Chair

Dr. Thomas Fevens _____ Examiner

Dr. Marta Kersten-Oertel _____ Examiner

Dr. Sudhir P. Mudur _____ Co-supervisor

Dr. Charalambos Poullis _____ Co-supervisor

Approved _____
Chair of Department or Graduate Program Director

_____ 20 _____ _____

Amir Asif, Ph.D..P.Eng., Dean
Faculty of Engineering and Computer Science

Abstract

Spatial Augmented Reality in Outdoor Areas using Long-Range Vision Systems

Behnam Maneshgar

Projection mapping [also known as Spatial Augmented Reality] is a well known augmented reality technique which changes the appearance of the physical, real-world environment by projecting media on them. Projection Mapping is currently used in a variety of applications and it is increasingly becoming more popular. In this research, we investigate the use of computer graphics and computer vision techniques that can enable us to create an immersive augmented reality environment in an outdoor setting with large number of participants. Specifically, we address long range projection of stereoscopic media in anaglyph format. This is inexpensive and less intrusive among user immersion technologies, since participants have only to wear low cost anaglyph glasses. Since the projection surface is at a long distance, and is a real world object(s), and there is always ambient lighting, there are many challenges. These include calibration of the cameras and projectors, modeling the projection surface(s) geometry and light behavior and transforming the projection media to compensate for the surface properties. We address these problems and propose an innovative new framework which uses image-based techniques for automatic calibration, geometry and appearance properties capture, and a simulation based technique for creating compensated media. We have implemented our framework and demonstrate its effectiveness through multiple user studies in real world settings.

Acknowledgments

I would like to thank my supervisors Prof. Sudhir Mudur and Prof. Charalambos Poullis and my external advisor Prof. Leila Sujir. Back to the day, I decided to immigrate and continue my education abroad, I had a couple of reasons which were very important to me. Today, I am so glad and grateful that I successfully reached my goals after two years and it was not possible without their endless support.

I would also like to thank my parents, my brother and his family especially my cutie niece who made my life more joyful with her birth.

Finally, the last but not the least, thanks to all my friends for their passionate support.

Behnam Maneshgar

Contents

List of Figures **vii**

1 Introduction **1**

2 Background **4**

2.1 Review of Related Work 4
2.2 Calibration 5
2.2.1 Camera Calibration 6
2.2.2 Projector Calibration 10
2.3 Geometry and Light Response Acquisition 10
2.3.1 Structured Light Scanning 11
2.4 Reflectance Properties 13
2.5 Illumination model 14
2.6 Specular detection 16

3 Proposed Approaches **18**

3.1 Fixed Camera Method 19
3.1.1 Calibration 19
3.1.2 Geometry Acquisition 20
3.1.3 Reflectance Properties 20
3.1.4 Color Compensation 22
3.2 Roving Camera Method 24
3.2.1 Bidirectional Reflectance Distribution Function 26
3.2.2 Compensation 28
3.3 Implementation 29
3.3.1 Implementation of Fixed Camera Method 29
3.3.2 Implementation of Roving Camera Method 31

4 Experiments **33**

4.1 Fixed Camera Method 33
4.2 Roving Camera Method 34
4.2.1 Indoor Testing 35

4.2.2 Public Place Experiment #1 35
4.2.3 Public Place Experiment #2 39

5 Conclusion and Future Work **47**

Bibliography **48**

List of Figures

1 Projecting stereoscopic content onto Roman Bath building. The projection is done by our partner "elasticspaces group" (leading by Dr. Leila Sujir). (a) shows the entrance of roman bath (b) shows same place after projecting stereoscopic 3D content onto it. 2
2 Pinhole camera model [Matb] 6
3 Camera radial lens distortion [Mata] 7
4 Captured image from a checker board for calibration. 7
5 Vertical (a) and Horizontal (b) Phase-shifting fringe patterns to be displayed by a monitor or TV. 8
6 This figure shows the encoding process of the feature points. 9
7 Two rays $\vec{X_L}$ and $\vec{X_R}$ are generated from the devices and their intersection in 3D coordinate system is V_p. $\alpha = \beta = 90$ [@tr] 11
8 Vertical and horizontal binary patterns [GHP14] 12
9 In Gray codes, the codes of two consecutive numbers differ just in one bit, however in binary code, it may vary in more than one bit [@GR] 13
10 Algorithm used by SLS to Re-mesh the point cloud stored in image format 13
11 First two figures demonstrate light reflection from surface of a specular and a diffuse object respectively. [@re17] 14
12 (a) Light reflection geometry model used by Cook and Torrance [CT82] and Shafer [Sha85]. [ABC11] (b) Diffuse and specular reflection lobe of a light source [JK10] 14
13 Phong illumination model components [Wikb] 15
14 Smart Projection Mapping System Pipeline 18
15 One of our setups while testing the system in long range. Three cameras have been calibrated corresponding to each other using phase-shifting method while they are focused at infinity. . 19
16 Fixed camera method: this technique requires a minimum of 3 fixed cameras and a long-range video projector. 19
17 Phase-Shifting camera calibration set-up 20
18 A rendering of the reconstructed geometry of the Roman Bath. 44 images captured by three cameras. 21
19 (a) The object/projection surface. (b) The estimation of the specular parameters without the smoothness term and (c) the estimation with the smoothness term. 22

20 (a) Three synthetic test cases: perfectly specular, perfectly diffuse, and diffuse/specular. (b) Energy minimization for shininess. (c) Energy minimization for diffuse coefficients. (d) Energy minimization for specular coefficients. 23
21 System Overview 24
22 Roving camera method: this technique requires a roving camera which captures N samples from the scene by sweeping in front of it; A minimum of two fixed cameras for geometry reconstruction, and a long-range video projector. 25
23 Long-range camera calibration. Traditional calibration techniques fail for long-range vision systems. In the proposed approach we use SfM for estimating the camera poses and intrinsic parameters, thereby eliminating the need for special calibration boards and procedures [BXZ16, ABL$^+$16]. Setup shown for experiment #1. 26
24 Experiment #1: (a) The XYZ map of the scene's structure generated with SLS. (b) The Normal map of the scene's structure generated with SLS. (c,d) Render from two novel viewpoints of the reconstructed geometry. 27
25 diffuse map(left), specular map(right). The inset picture in the specular map (b) is a close-up of the region indicated with red and demonstrates the level of detail captured by this approach. 28
26 (a) A stereoscopic projection without compensation. Color distortions due to the reflectance properties of the projection surface negatively affect the depth perception of the viewer. (b) The stereoscopic projection after compensation using the proposed technique. Color distortions are minimized by taking into account the effect of the reflectance properties of the projection surface. (c) The expected projection. This cannot always be achieved due to the limitations of additive color mixing and the hardware.(d) The original stereoscopic image taken from [ste] 34
27 Our test in lab environment. (a) Projection surface, (b) rendered scene. (c) one frame of our projection media before compensation, (d) same frame after compensation. (e) Projected scene with a frame without compensation, (f) adjusted same frame after compensation, (g) XYZ map and (h) diffuse map of the scene. 36
28 An example image used in Experiment #1. First row: (left) original image, (right) adjusted image. Second row: (left) projection of original image on scene, (right) projection of adjusted image on scene. Third row: (left) adjusted images after one iteration, (right) adjusted image upon convergence of the algorithm. (g) Capturing the geometry by SLS and (h) is the update map (Absolute difference between (a) and (b). 38
29 Experiment #1: Statistical significance of participants' questionnaire responses between before (vertical) and after (horizontal) the application of the proposed approach. The question for both projections was 'Rate your color and depth perception for this projected image/video'. A likert scale was used ranging from [1,10]. 95% confidence level. Sample size: 34. 39
30 The distributions of the responses to the question 'Rate your color and depth perception for this projected image/video' with the original image (top) and adjusted image (bottom) 40

31 In this figure (a) shows our projection setup, (b) shows the distance and direction of our viewers and projection (c) shows the XYZ map and (d) is the normal map of the scene . . . 42
32 Experiment #2 [Butterfly video]: Participants' questionnaire responses between before (horizontal) and after (vertical) the application of the proposed approach. A likert scale was used ranging from [0,5]. 95% confidence level. Sample size: 37. 43
33 An example frame (from the video) used in Experiment #2. (a) original frame, (b) adjusted frame, (c) projection of original frame on scene, (d) projection of adjusted frame on scene, (e) the projection onto the building's facade; note the excessive ambient lighting present in the scene; distance over 20m, (f) diffuse map of the surface. 44

Chapter 1

Introduction

Spatial Augmented Reality, or its more commonly known name Projection Mapping (PM), is a projection technique which transforms a real-life object or scene into a surface for video projection [RWF98]. Although this technique has been pioneered and used by Disney since the seventies, it is only in recent years that it has gained significant popularity due to the availability of specialized software which simplifies the otherwise cumbersome calibration process [RWC+98]. Currently, PM is being widely used in advertising, marketing, cultural events, live performances, theater, etc as a way of enhancing an object/scene by superimposing visual content [RRL+14].

Stereoscopic content is one step further in PM. More detail can be exposed by adding one more dimension to the content. Perhaps the most popular proposed method to generate stereoscopic content is the anaglyph 3D. The anaglyph 3D demonstrates the disparity using two different colours, typically red and cyan. The anaglyph 3D media is created by blending two differently filtered coloured views of the scene one for each eye. Viewers perceive depth by wearing anaglyph 3D glasses which filters the view which is not intended for each eye [Wika]. In contrast with passive and active vision technologies, anaglyph 3D does not impose any additional requirements such as the number of projectors needed [as in the case of polarized stereo or active stereo which require two], expensive active vision/shutter glasses which also have a limited usage time-span due to recharging requirements, or expensive passive stereo glasses with polarized optics.

Hence, in comparison with the other available techniques, it is no surprise that the anaglyph 3D is still the most popular technique for creating stereoscopic content, primarily due to its simplicity and cost-effectiveness. The creation of an anaglyph 3D involves the encoding of two images, one for each eye, with different colors. Historically, the colors which have been used are red and cyan. The viewer can then perceive 3D with the use of anaglyph glasses: glasses which have one red and one cyan lens. Each colored-lens allows only the image which is encoded with the same color to pass through therefore ensuring that only one image is seen by each eye. The brain, and in particular the visual cortex, fuses the two images seen by the eyes into the perception of a three-dimensional scene.

Anaglyph stereoscopic content has been successfully used in demonstrating scientific results [mar], visualizing maps [BR09] health [RGVP+14], and also used as an easy way of presenting 3D content online [Wik17]. More recently, researchers from several fields including artists, designers and computer scientists among others, have begun exploring the potential of stereoscopic technologies with artistic practices.

Figure 1: Projecting stereoscopic content onto Roman Bath building. The projection is done by our partner "elasticspaces group" (leading by Dr. Leila Sujir). (a) shows the entrance of roman bath (b) shows same place after projecting stereoscopic 3D content onto it.

Projecting onto outdoor surfaces such as building facades, poses several challenges. Firstly, the projection surface may be coloured (not necessarily white) and may also contain parts with complex reflectance properties or low reflectivity which may interfere with the projected content. Secondly, to be able to account for the occurrence of these types of object properties, one has to capture the geometry and light behavior properties of the projection surface, which typically requires the use of a calibrated camera-projector system. The calibration of the camera-projector system becomes a non-trivial task because of the long-range. Standard procedures such as those described in [Zha00, Tsa87] require that all optical systems are focused on the calibration board/object. However, in the case of long-range outdoor projection, the camera-projector system is focused on the projection surface which is located at a long distance away e.g. $> 15m$. Capturing images of the calibration board from such a distance leads to poor coverage within the image which in turn results in improper calibrations. On the other hand, capturing images by placing the calibration board at a shorter distance leads to blurry images which again results in poor calibrations.

An early example of one of our outdoor projection mapping experiments is shown in Figure 1. In this: (a) The projected media was stereoscopic. (b) The projector alignment was done manually. (c) The media was not compensated for the projection surface properties. (d) The viewers were wearing anaglyph 3D glasses in order to perceive depth. Objects with complex reflectance properties such as the windows and columns could not be handled during the projection and resulted in color distortions to the red-cyan stereo content. This caused loss of depth perception and noticeable visual artifacts. Yet the general audience response was quite positive.

In this work, we address many of the problems of long range projection mapping of stereoscopic content in outdoor areas and propose a complete framework which automates the following processes: (a) system calibration, (b) structure and appearance information acquisition, (c) approximating model of projection surface's reflectance properties, and (d) color compensation to the extent possible with the given projection surface. The result is compensated image/video content such that its projection onto the particular surface will produce an image/video which, when viewed, will seem as close to the original as possible. This is, of

course, limited by the projection surface properties, as it may not always be possible to completely compensate for the surface reflectance behavior.

We have proposed two different approaches to improve the quality of long-range stereoscopic projection mapping in outdoor areas. In the first approach, which we term as the Fixed Camera approach, we present a solution which uses a minimum of 3 fixed video cameras and a fixed long range projector. Using this setup, we present a complete framework for calibration, geometry acquisition and reconstruction, estimation of reflectance properties, and finally color compensation; all within the context of outdoor long-range PM of stereoscopic content. Using the proposed technique, the observed projections are as close as possible [constrained by hardware limitations] to the actual content being projected; therefore ensuring the perception of depth and immersion when viewed with stereo glasses. The 3 standard fixed video cameras are used to reconstruct the geometry and to capture reflection samples from the scene.

Computing reflectance properties of the scene is a model fitting problem which means the more samples, the more accurate result. However, in the Fixed Camera approach utilizing a large number of cameras to have different samples from the scene is time consuming and expensive while the risk of confusion due to multiple samples from different cameras is higher as well. Therefore, we proposed the second approach, termed the Roving Camera approach, which overcomes the sampling problem. This system is composed of 2 standard fixed video cameras, a long range fixed projector, and a roving video camera for multi-view capture. While most of the time, we have used two fixed cameras, one on each side of the projector, in cases when we have too many holes in 3D reconstruction caused by self-occlusions, we have experimented with a third camera on top, and it has worked very well.

In both of the above methods, the overall computational framework comprises of four modules: (a) calibration of a long-range vision system without the need for calibration boards, (b) dense 3D reconstruction of projection surface geometry from the multiple calibrated camera images, (c) modeling of the reflectance properties of the projection surface from roving camera images and, (d) the adjustment of the stereoscopic content in an iterative manner. In addition to cleverly adapting established computer vision techniques, our principal contribution is the system design which is distinct from previous work. The proposed methods have been implemented and tested in real-world applications. The methods have been experimentally evaluated in real and/or simulated environments. Specifically, the Roving Camera method has been evaluated via three non-trivial user experience studies in public places uses stereo projection onto building facades after dark. The results are reported and show considerable improvement in the quality of perception of the projected media and immersion in the augmented reality space.

Chapter 2

Background

In this chapter we present related work in the following order. First, the existing methods and techniques for projection mapping and for controlling the appearance of objects have been reviewed. Then, the basic techniques required for projection mapping including camera calibration, geometry acquisition, and reflectance properties are explained.

2.1 Review of Related Work

Many different methods have already been proposed for controlling and adjusting the appearance[1] of projections onto various surfaces. The majority of which seem to work quite well for controlled environments. Below we provide a brief review of earlier work which is close to our research objective.

Grossberg et al. [GPNB04] presented a method to control the appearance of small objects by using one camera and one projector. In their work, the focus is primarily on the spectral responses, spatially varying fall-offs, and non-linear responses in the projector-camera system which creates a high dependency between the camera view and camera response. The computed radiometric model of the system is then used to calculate the compensated image.

In a similar approach, Aliaga et al. [AYL+12] addressed appearance editing of an object using a projector. Multiple projections are used to improve the resolution and compensate the images by reformulating the problem as a constrained optimization. An elliptical Gaussian is used to model projector pixels and their interaction between projectors.

Perhaps the closest work to that proposed here is in [BWEN05]. Bimber et al. introduced a view-dependent stereoscopic projection for compensating distortions caused by the scene's structure. They present an elaborate process which involves computing of inverse light transport to create compensated images and demonstrate the success in controlled environments.

A recent work was also done by Ahmed et al. [ALLL16]. They used multiple projectors to reproduce the appearance of an object. Their system can achieve better black levels and less contrast compression.

Several works [BGJC14, BGJ+15, SCT+15] have addressed dynamic and moving scenes. One or more projectors have been used to continuously compensate the projected image while the present objects are

[1] The appearance of objects is given by the object color, emitted light and object's light reflection and transmission

moving. Y. Zhou et al. [ZXT+16] recently introduced a method for dynamic objects. They use a Microsoft Kinect and a projector for their setup. In the first stage, the system needs to be calibrated. Later they require the 3D geometric model of the object either by scanning it or using a 3D printed object. Afterward, The scene is scanned in real time using a kinect and gives the dense point cloud of the scene. By meshing and registering the low-density version of the scanned scene, the authors find the position and rotation of the object in the 3D coordinate system. In the next step, the texture of the virtual model is projected onto the real object. Since white colored objects are used, no compensation is needed. Law et al. [LAS+11] have attempted a perceptually-based object appearance modification. They partitioned the projection surface into patches based on the target appearance colors to make it appear as similar as possible to the target.

All the aforementioned techniques have been shown to perform reasonably satisfactorily in cases involving projections of non-stereoscopic content from short-ranges i.e. $\leq 3m$. The primary reason is that, the calibration procedure involved which for longer distances in uncontrolled environments produces poor results; both in terms of geometric and radiometric calibration.

In general, the following steps are needed to compensate for projection on the objects/scenes. System calibration is a fundamental step in most of the machine vision systems. Hence, in our case, the first step of the pipeline is calibration of the system. Usually, one or more camera(s) and one projector are used in the system, and they need to be geometrically calibrated with respect to each other. Next, the geometry of the scene is acquired which allows for the per-scene-point normal computation. As it is known, the normal is required for computing how a surface point reflects light. In addition to the normal, the scene's reflectance properties e.g. diffuse, specular, transparent, translucent, etc are also required for determining the reflected light. Lastly, the media should be compensated based on the scene's properties, both in terms of geometry and surface reflectance properties, which have to be modeled. In the following sections, the background theory of each stage of the pipeline is explained in detail.

2.2 Calibration

Cameras are an intrinsic component of any vision system. Since the late 20th century, the pinhole camera has become very popular because of its simplicity and low cost of manufacturing. The pinhole camera model is a mathematical model which describes the projection of 3D points into the image plane of a camera. Figure 2 shows the projection of a 3D object (tree) onto a 2D plane using an ideal pinhole camera model. For simplicity, this model describes a camera without the lens, and therefore it does not simulate lens distortions.

A pinhole camera model can be described with two matrices; one describes **Intrinsic** parameters of the camera like focal length and optical center expressed in pixel coordinates and the other one describes **Extrinsic** parameters which consist of rotation and translation of the camera. The process of determining these two matrices is called camera resectioning or commonly known as geometric camera calibration. In another word, camera calibration is the estimation of location, orientation, lens parameters and sensor parameters of a camera. Camera calibration has a significant role in the field of computer vision and robotics. Moreover, accurate calibration is of crucial importance when dealing with long-range vision-projection systems, such as in the case of long-range projection mapping. A small error in image space i.e. few pixels, can lead to large displacements in the projected space. In this section, we describe our system calibration requirement, which

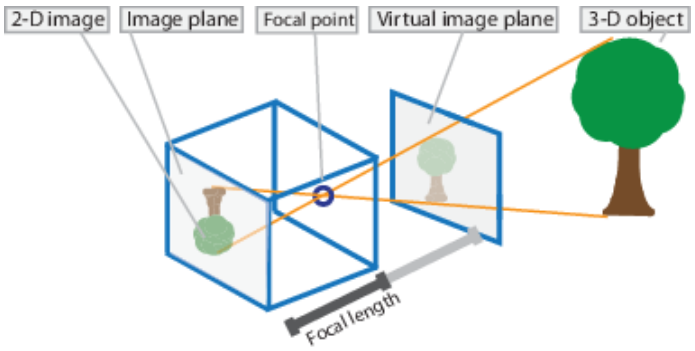


Figure 2: Pinhole camera model [Matb]

involves:

- the calibration of the cameras
- the pose recovery with respect to the cameras and intrinsic parameter calibration of the projector

2.2.1 Camera Calibration

The pinhole camera model is used to describe these parameters which are specified by the camera matrix C in the Equation 1,

$$C = \underbrace{\begin{bmatrix} \alpha & -\alpha cot(\theta) & u_0 \\ 0 & \frac{\beta}{sin(\theta)} & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{intrinsic} \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}}_{extrinsic} \tag{1}$$

where $\alpha = kf_x$, $\beta = kf_y$, (f_x, f_y) is the focal length on the x and y-axis respectively, θ is the skew angle, u_0, v_0 is the principal point on the x and y-axis respectively and r_{1-3}, t_{x-z} determine the camera's rotation and translation relative to the world. An ideal pinhole camera does not account for cameras with lenses, however, real cameras are using complex lenses. Two different distortions have been proposed for camera lenses,

- **Radial Distortion** which appears when light rays bending amount is different from the center of the camera sensor to the edges. Barrel distortion results when bending near the edges is more than at the center, and Pincushion distortion happens if bending is more at the center of the camera. [Figure 3]
- **Tangential Distortion** occurs once the lens and camera image plane are not parallel to each other.

Lens distortion can be modeled by following parameters:

$$D = \begin{bmatrix} k_1 & k_3 & P_1 & P_2 & k_3 \end{bmatrix}^\top \tag{2}$$

Perhaps the most popular technique for calibrating a camera is the one proposed in Tsai et al. [Tsa87] and extended in Zhang et al. [Zha00]. Given a set of points in world space and their corresponding image points, one can recover both the intrinsic and extrinsic parameters of the camera. In the method proposed by Zhang

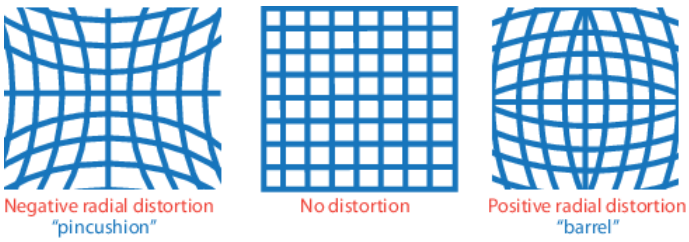


Figure 3: Camera radial lens distortion [Mata]

et al. [Zha00], multiple images from an in-focus checkerboard in different poses of a known geometric object are captured. Since a checker board is a collection of planar feature points, we can suppose that Z is the same for all points on a checkerboard which could be equal to a constant $Z = c = 0$. Also, the top left corner of the checkerboard is considered as the $(0,0,0)$. Therefore the position of all the other corners of the checkerboard can be calculated by the size of each square of the checkerboard. For example, if the size of each square is 5cm, we will have feature points like $(0,0,0)$, $(0,5,0)$, $(10,5,0)$ in each pose of the checkerboard. Next, as using 2D and 3D correspondences, camera parameters are recovered.

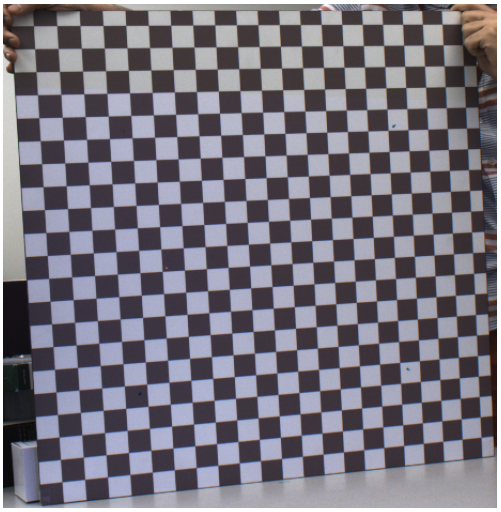


Figure 4: Captured image from a checker board for calibration.

Out Of Focus Calibration method

This method encodes feature points into phase shifting patterns being displayed on a monitor visible to the cameras. The phase shifting patterns are a sequence of periodic intensity patterns. The patterns are shifted equally in the sequence until it covers the entire period [PMS10].The feature points can then be accurately decoded even when the captured images are blurred, because this does not affect the phase of the pattern sequence. To encode the feature points, one vertical and one horizontal phase map are required in which each vertical/horizontal line has a unique phase value [Figure 6]. Thus, each pixel appearing on the monitor has a single pair of (Φ_v, Φ_h) to identify the feature. These phase maps are carried by the phase shifting fringe patterns. Equation 3 is used to generate N equally phase-shifted vertical and horizontal fringe patterns (Figure

5) and is given by,

$$\begin{aligned} I_v^i(u,v) &= 0.5\Big[1+\cos\left(\Phi_v + 2i\pi/N\right)\Big], \\ I_h^i(u,v) &= 0.5\Big[1+\cos\left(\Phi_h + 2i\pi/N\right)\Big] \end{aligned} \tag{3}$$

where i is the index of the fringe pattern, Φ_v and Φ_h represent the vertical and horizontal phase maps, respectively. Equation 4 generates the vertical and horizontal phase map (Φ_v, Φ_h) in order, for point/ monitor pixel (u_0^d, v_0^d).

$$\begin{aligned} \Phi_v(u_0^d, v_0^d) &= 2\pi u_0^d / P_v, \\ \Phi_v(u_0^d, v_0^d) &= 2\pi u_0^d / P_h. \end{aligned} \tag{4}$$

After projecting and capturing the patterns, the phase value of each pixel of the captured image sequence extracts as follows,

$$\phi(x,y) = \tan^{-1}\left[\frac{\sum_{i=1}^{N} I^i \sin\left(2i\pi/N\right)}{\sum_{i=1}^{N} I^i \cos\left(2i\pi/N\right)}\right] \tag{5}$$

Where I^i is the intensity of a particular pixel in the i^{th} captured image. This equation generates a non-continuous wrapped phase map with values in the range of $[-\pi, \pi]$. Next, the phase maps are unwrapped to produce a unique phase value for each of their columns/rows in the pattern. Adding an offset to each section of the phase map generates an unwrapped phase map which has a unique phase value at each horizontal/vertical pixel line as shown in the following equation,

$$\Phi(x,y) = \phi(x,y) + k \times 2\pi \tag{6}$$

After decoding of the correspondences, the camera can be calibrated using the traditional technique. The result is the intrinsic and extrinsic parameters for each camera. The extrinsic parameters are given with respect to the first (top-left) encoded feature point in the monitor.

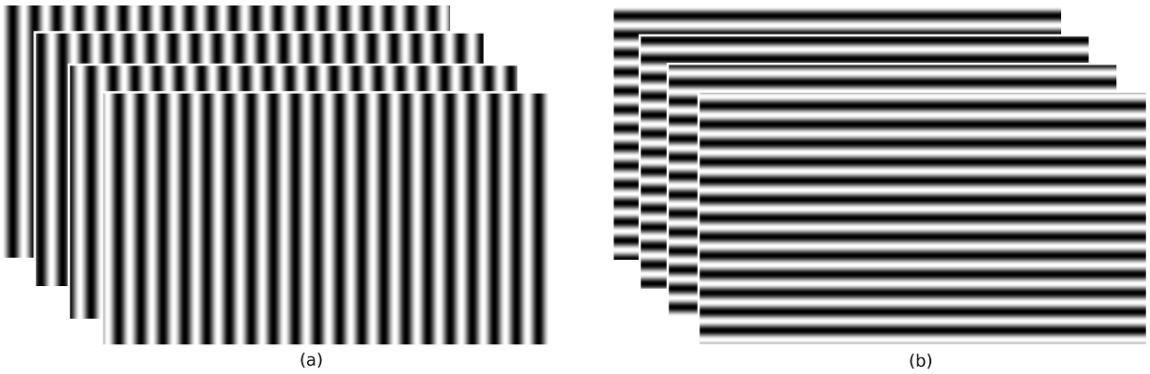


Figure 5: Vertical (a) and Horizontal (b) Phase-shifting fringe patterns to be displayed by a monitor or TV.

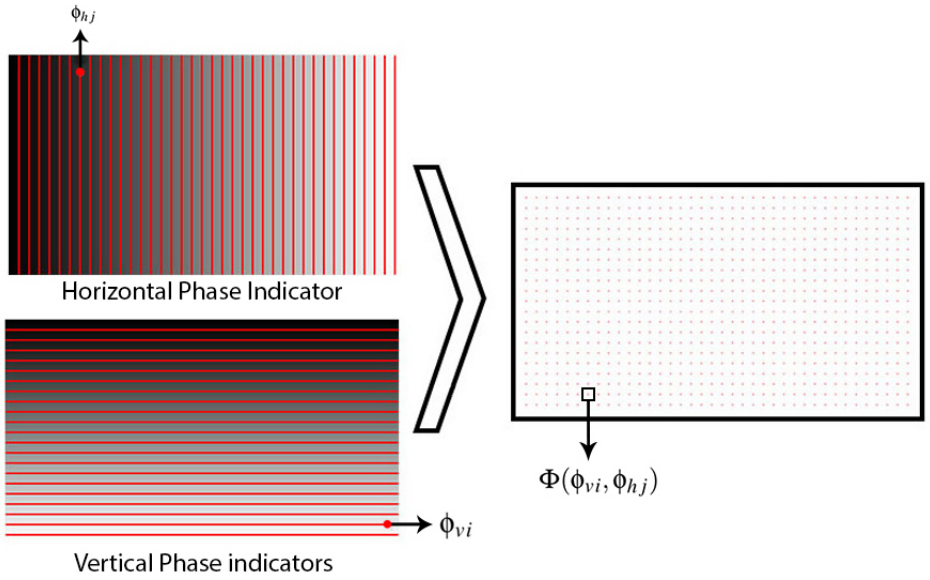


Figure 6: This figure shows the encoding process of the feature points.

Algorithm 1: Camera calibration and pose estimation using phase shifting.

1 generate vertical and horizontal fringe patterns
2 display and capture images from monitor at different poses
3 compute wrapped phase-map
4 calculate k in Equation 6
5 calculate unwrapped phase map
6 match the encoded feature points with the decoded phase map
7 calibrate camera using the traditional technique

Self Calibration

Self-Calibration is another technique for calibrating multiple cameras set-up. In this method, multiple cameras observe a common scene, and by discovering 2D correspondences between the cameras, the camera parameters of the system can be recovered. Svoboda et al. [SMP05] proposed a fully multi-camera self-calibration method for virtual environments. They have a detectable bright spot in the working volume and capture the volume with at least three synchronized cameras. These points are considered as correspondences and are validated through pairwise epipolar constraints. The system is calibrated using these correspondences.

Structure from Motion (SfM) is a simple, powerful and robust vision based technique for estimating the 3D geometry of a 3D object. It receives a sequence of 2D images taken from the object by the camera and recovers its 3D points. Briefly, Structure from motion first uses SIFT to find and match the feature points of the image sequence. Therefore it can track ”P” points in ”F” frames. By having the correspondences between the frames, SfM calibrates all the cameras/frames using a self-calibration method. Afterward, it triangulates, and reconstructs a sparse geometry of the scene. Lastly, the bundle adjustment technique is applied to minimize the error. The bundle adjustment technique optimizes the 3D coordinates and calibration parameters to refine the 3D structure.

2.2.2 Projector Calibration

The projector is usually treated as an inverted camera and is calibrated using the traditional method introduced by Zhang et al [Zha00]. Given a set of 3D world points and their corresponding 2D image positions, the projector's intrinsic and extrinsic parameters are recovered. Several works have been done to calibrate a multi-camera-projector system which means calibration of all camera(s) and projector(s) of the system with respect to one coordinate system. As an example, R. R. Garcia and A. Zakhor [GZ13] take advantage of binary code patterns and encode each pixel of the projector. First, they project the patterns on a screen and capture the image sequence by the cameras. Then they decode the binary codes and generate dense correspondences between the cameras and projector. Lastly, they perform bundle adjustment to calibrate the system. In a similar approach, An et al. [ABL+16] also calibrate a multi-camera-projector system, however instead of using binary code patterns, they take the benefit of phase-shifting and fringe patterns. They encode each pixel of the projector by a pair of vertical and horizontal phase amount. Cameras can capture projected fringe patterns and decode the phase value. Using the phase shifting technique makes the system robust in the case of an out-of-focus set-up.

2.3 Geometry and Light Response Acquisition

The geometry of the projection surface directly impacts the appearance of the surface as well as the projected content. Identifying surface points with complex reflectance properties requires that the geometry (surface points and normals), and reflectance properties (response to light) of the projection surface be known.

3D reconstruction is a well-studied area in computer graphics. Many techniques have been proposed [TS67], [Pho75], [ON95], [LFTG97a] and a lot of systems exist. As an example, Y. Furukawa and J. Ponce [FP10] proposed an algorithm to reconstruct an object/scene by using stereo cameras. The algorithm detects feature points of each image, finds the matches between each pose and outputs a dense set of patches covering the surface of the object/scene. Many commercial products are also available on the market. Microsoft Kinect is one of the most popular devices and comes with game consoles. It scans the scene/object in high resolution with high accuracy in short range. The existing methods mostly have been particularly successful in cases where the acquisition is performed under controlled lab conditions without the presence of any dynamic elements. Although PM can also be used indoors in a similar fashion, the majority of its applications involve large-scale and/or outdoor objects/scenes. Perhaps the only work reported in the literature to address the capture of complex geometry and the estimation of reflectance properties of outdoor objects is by Debevec et al. [DTG+04].

Finding corresponding pixels between calibrated devices like projectors and cameras is the primary requirement for 3D reconstruction. As an easy and straightforward but time-consuming solution, a single red dot could traverse among all the pixels of the projector. Therefore at each moment, just one single pixel of the projector is illuminating the scene. Simultaneously all cameras capture the scene. The single dot projected on the building and its projection onto the camera's image plane provides the correspondences between all the devices. Next, by triangulation, the 3D point can be computed.

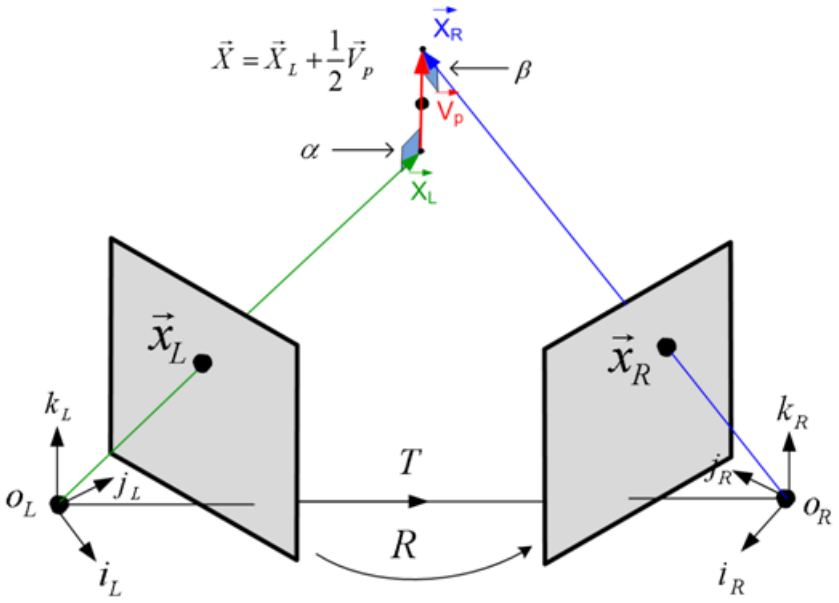


Figure 7: Two rays $\vec{X_L}$ and $\vec{X_R}$ are generated from the devices and their intersection in 3D coordinate system is V_p. $\alpha = \beta = 90$ [@tr]

Triangulation

By having a calibrated device like a camera, we can cast a ray for each pixel. The ray starts from the position of the camera, and it passes through the pixel of the camera which is located on its image plane. Camera pose and focal length are needed in order to generate a ray. The pose can be recovered by the rotation and translation of the camera, and its focal length is used to retrieve the position of the image plane. Therefore if we send rays from the corresponding pixels of each device, their intersection is the 3D point of the scene. Figure 7 shows triangulation for two devices O_L and O_R. R and T are the rotation and translation of the cameras according to each other, $\vec{X_L}$ and $\vec{X_R}$ are rays cast from left camera and right camera through the X_L and X_R corresponding pixels respectively. As it is shown in the Figure 7, because of the possible errors during the calibration of each device and in finding corresponding pixels, in 3D coordinates, rays rarely intersect at a single point. Therefore, the nearest point to all of the rays i.e. $(\bar{X})$, can be considered as their intersection. By repeating this process for all corresponding pixels in the devices, we can reconstruct the scene/object.

2.3.1 Structured Light Scanning

Structured Light Scanning (SLS) [GHP14] is a fast and accurate technique for scanning objects in high resolution. A sequence of patterns is projected on the scene and camera(s) capture the projected patterns on the scene. By processing the captured images, the geometry of the real object is reconstructed.

As was explained before, the red dot algorithm illuminates the scene with one single pixel for a moment. Thus the system can find the corresponding pixels between devices. However, in this method, we want to speed up the process and capture the correspondences by projecting a small number of patterns. To achieve this goal, SLS assigns a unique number/ID to each pixel of the projector, and by projecting this ID on the

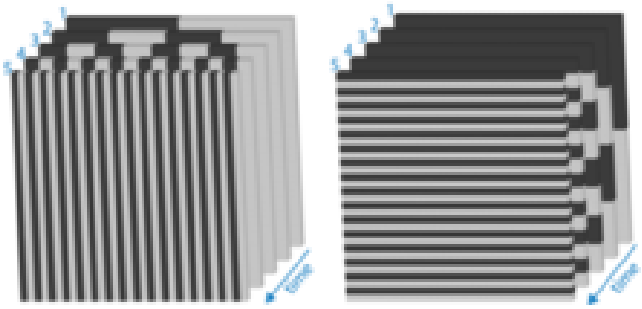


Figure 8: Vertical and horizontal binary patterns [GHP14]

scene, cameras can distinguish pixels. However projecting the number itself is not feasible and instead a sequence of patterns are used. Binary patterns, shown in Figure 8 are typically used for this. This technique presents the unique ID of the pixel in binary representation mode. Each pattern indicates one bit of the number. If the bit is 0, the pixel is off (black), if the bit is 1, it projects white color. Therefore cameras can read the ID of each pixel doing the inverse steps. First, it captures black and whites and then it can assign zero and ones based on the color of the pixel. Although binary codes can be used to encode the pixels, Gray codes have been used in SLS. Gray codes work in a similar approach to the binary codes. However, using gray codes ensures that consecutive codes differ just in one bit. Figure 9 shows the difference between binary codes and Gray codes. It is evident that in Gray codes each column differs from its adjacent column just in one bit. However, in binary systems all the bits are changing in the middle columns.

Finally, the correspondences between the cameras and the projector can be defined. By triangulation of the calibrated devices, the 3D model can be reconstructed. Despite the possibility of using one calibrated projector and one calibrated camera, usually, two calibrated cameras are used in SLS because of the simplicity of camera calibration compared to projector calibration.

The resolution of the projector is the most important factor in the quality of the scanned model. However, cameras also should have enough resolution to be able to capture all the projected pixels of the projector. Since SLS finds the 3D point related to each pixel of the projector, it also saves the 3D data into a three channel image. Each pixel of the image has three data R, G and B that maps to X, Y and Z respectively.

Hole Filling

In most of the cases, because of the self-occlusions, which is not able to reconstruct all the points (projector pixels) of the scene. Adding more cameras to the system, calibrating the projector and/or another scan of the object can obtain more information about the geometry of the scene. Nevertheless, a hole filling technique also can estimate the position of missed points/pixels during the scanning process. Since SLS saves data in image format, pixels with no depth/position information are known, a simple method of averaging the position of its neighbors can recover the geometry information.

Filtering

Errors in practical experiments are inevitable. Bilateral filtering [TM98] is a noise cancellation method which reduces the error/noise of the image. The advantage of bilateral filtering compared to other methods is that

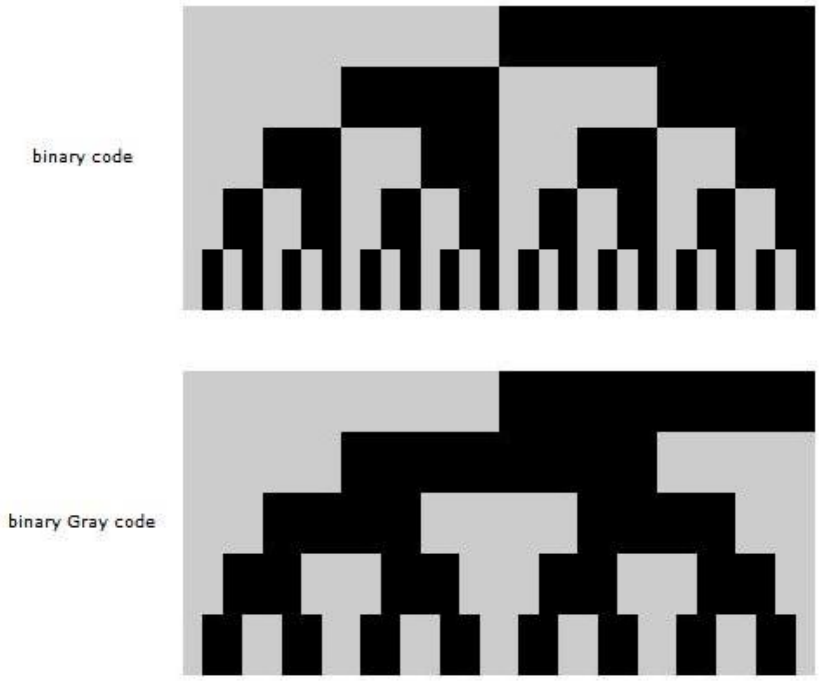


Figure 9: In Gray codes, the codes of two consecutive numbers differ just in one bit, however in binary code, it may vary in more than one bit [@GR]

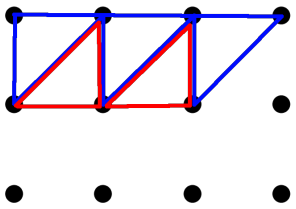


Figure 10: Algorithm used by SLS to Re-mesh the point cloud stored in image format

edge information is preserved.

Re-meshing

SLS results a point cloud, whereas the 3D model and mesh of the scene is needed. Thus a re-meshing algorithm is required. SLS uses a simple re-meshing algorithm. Since data is stored in image format, the neighbors of each point are known and by connecting neighbors to each other, we can triangulate the point cloud and generate the 3D model. Figure 10 shows how SLS connects neighbors in re-meshing process.

2.4 Reflectance Properties

There are thousands of objects with different materials in the world. Each one responds to the received light in a specific way. Some absorb all the received light; some reflect part of the light, some of them are transparent

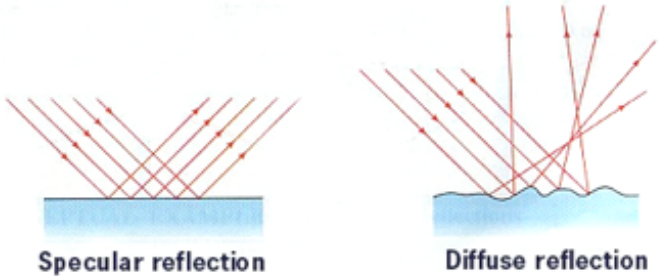
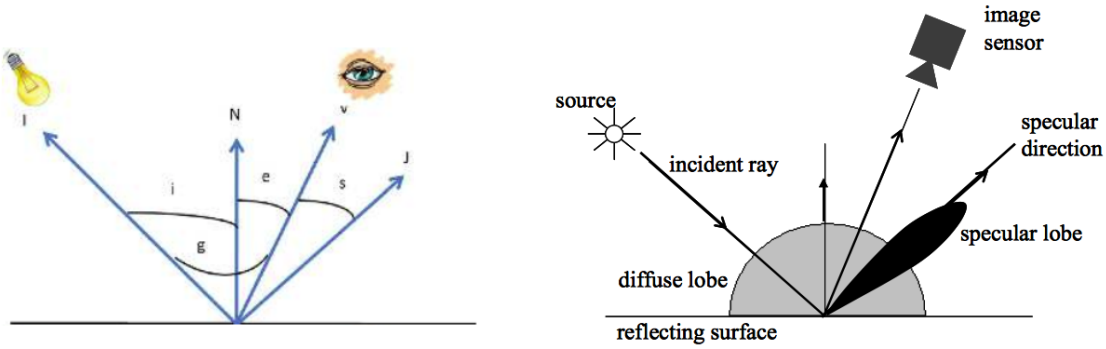


Figure 11: First two figures demonstrate light reflection from surface of a specular and a diffuse object respectively. [@re17]



(a) (b)

Figure 12: (a) Light reflection geometry model used by Cook and Torrance [CT82] and Shafer [Sha85]. [ABC11] (b) Diffuse and specular reflection lobe of a light source [JK10]

and light passes entirely or partially through. The Bidirectional Scattering Distribution Function (BSDF) is a model which describes light behavior of an object which is illuminated by a light source directly or indirectly. BSDF is a superset of Bidirectional Reflectance Distribution Function (BRDF) and Bidirectional Transmittance Distribution Function (BTDF). BRDF models the light reflection from the object and BTDF models the light transmitted through the object. In the context of projection mapping, Bidirectional Reflectance Distribution Function (BRDF) of an object demonstrates how an observer can perceive the reflection of the light illuminated by the projector. Diffuse and Specular are two major elements of the reflection. Diffuse objects reflect light in all directions, whereas in pure specular reflection, the light reflects mostly in a particular angle related to the incident light. Figure 11 shows the comparison between specular and diffuse surfaces.

2.5 Illumination model

Many BRDFs have been introduced. Each models one or more aspects of the light reflection and has its strength. The Lambertian model is the simplest reflection model, which simulates only diffuse reflection.

Phong illumination model

Phong illumination model [Pho75] is a simple BRDF. It introduces three different light reflection behaviours. Ambient light, diffuse light, and specular light reflection. Ambient is the environmental light in which the intensity of light is invariant for all the points. As it is also apparent in Figure 13, the ambient light illuminated all the points with the same intensity and the object has no shading. The ambient light can be modeled by:

$$I = L_a * K_a \tag{7}$$

Where I is the intensity of the light reflected from the surface of the object and received by the viewer, L_a is the ambient light received by the surface and K_a is the ambient reflection constant. Diffuse reflection has been modeled as follows:

$$I = L_d * K_d * \cos(\theta) \tag{8}$$

where L_d is the diffuse light received by the surface, K_d is the diffuse reflection constant and θ is the angle between incident light and the normal of the surface which is equivalent to angle i in Figure 12a. Thus the viewer position and the view direction does not affect I neither in ambient reflection nor in diffuse reflection. Whereas specular reflection depends on view direction. (vector "v" in Figure 12a). This means that each observer receives unique amount of light based on view direction.

$$I = L_s * K_s * \cos(\phi)^{\alpha} \tag{9}$$

The above equation simulates specular reflection, where ϕ is the angle between the reflected light from the surface and view direction (marked as s in Figure 12a), α is the shininess constant of the object, the larger alpha the smaller the highlight. A small α outcome is a wide and smooth specular highlight. The Phong illumination Equation 10 is the sum of these three components. Moreover, in real life, usually there are more than one light source in the environment. The Phong model computes the reflection of each light separately and the sum of all components will be the reflected light (Equation 10).

$$I = L_a * K_a + \sum_{i \in Lights}^{\infty} (L_{i,d} * K_d * \cos(\theta) + L_{i,s} * K_s * \cos(\phi)^{\alpha}) \tag{10}$$

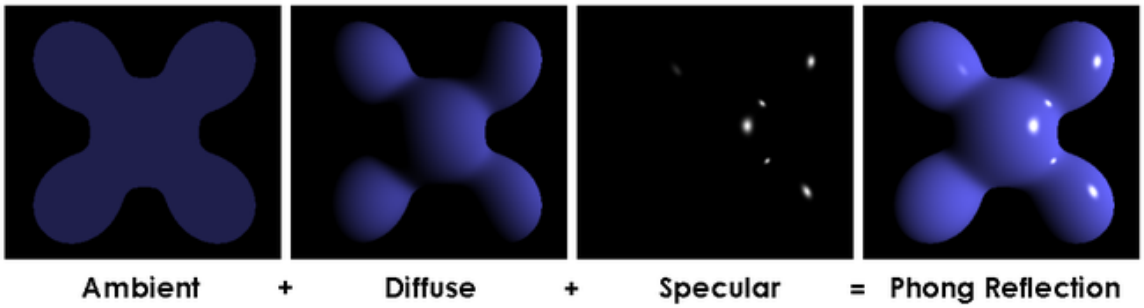


Figure 13: Phong illumination model components [Wikb]

The Phong illumination model follows the RGB color model and computes each color separately. It means that I, L and K consist of three RGB values for all ambient, diffuse and specular reflections.

$$I = \begin{pmatrix} I_r \\ I_g \\ I_b \end{pmatrix} \qquad L = \begin{pmatrix} L_r \\ L_g \\ L_b \end{pmatrix} \qquad K = \begin{pmatrix} K_r \\ K_g \\ K_b \end{pmatrix} \tag{11}$$

After Phong, Blinn [Bli77] introduced a BRDF known as the Blinn-Phong reflection model. Despite the simplicity of this model, it became very popular. It is the standard lighting model used in both the DirectX and OpenGL rendering pipelines [MSUA12]. By using $(n \cdot h)$ instead of $(r.v)$ Blinn simplified the model and decreased computations since there is no need to compute the reflection of the light r anymore.

$$D(h) = (n \cdot h)^n = cos(\phi)^n \tag{12}$$

where h is the halfway vector and defined as follow

$$h = \frac{(L+V)}{\left\|(L+V)\right\|} \tag{13}$$

where L is light vector and V is view vector.

This reflection model is very fast and straightforward. However, it cannot simulate complex materials. The Lafortune [LFTG97b] introduced a more multi-functional BRDF which supports more complex objects. This model handles multiple specular lobes which help in simulating complex objects.

$$f_r(u,v) = \sum_{i=1}^{N} (C_{x,i}u_x v_x + C_{y,i}u_y v_y + C_{z,i}u_z v_z)^{n_i} \tag{14}$$

Equation 14 represents this model, where u is the incident direction, and v is the extent direction, which is not in the same plane, and C is the coefficient of each XYZ direction.

Many samples are needed with different viewpoints to acquire and model the material of an object. The sampling process is usually performed under lab conditions. MERL [MPBM03] is a BRDF database which contains the light reflection of an object in different conditions. After measuring the reflections and collecting the data, they must be fitted into a BRDF model. Lafortune model is a favorite candidate. Eventually, computed materials can be used to render realistic scenes.

2.6 Specular detection

Detection and separation of the diffuse component and specular component of the objects is a vital issue in computer graphics, computer vision, and general vision systems including PM. Typically, techniques first try to detect the specularity of the scene and then cancel out or reject them as an outlier [ABC11]. A plethora of methods with different approaches have been proposed to tackle this problem, Single image techniques, and multiple image techniques have been suggested to extract diffuse map and specular map of the scene. Lin et al. [LLK$^+$02] proposed a color-based method to identify and separate the specular component from an input image sequence by using a multi-baseline stereo system. Fries et al. [FRTT04] proposed a multi-flash method to achieve this separation. They used a fixed camera and a portable flashlight to capture images of

the scene with different light source positions. Seitz et al. [SMK05] proposed a method for the cancellation of n-bounced inter-reflection light. In this, they first proved the existence of a set of linear operations by applying the inverse light transportation theory. Then, by probing the scene using a very narrow beam of the light, they can compute the light transport operators.

Specular objects impose problems in projection mapping. In specular reflection the light received by the observer depends on the position of the observer and its view direction. In projection mapping, participants are watching the scene from different angles, and therefore each person receives a distinct amount of light. A mechanism is needed to minimize its effect.

In the following chapter, we describe in detail, the two methods, Fixed Camera and the Roving Camera and the implementation of our complete framework for projection mapping in outdoor situations.

Chapter 3

Proposed Approaches

The previously techniques described have been applied to short-range capture and projection. The work reported in this thesis is mainly concerned with long range and stereoscopic projection. We present our findings in extending these existing techniques, as well as the development of new methods due to new problems that arise during long-range capture, calibration, and projection.

There are two important factors in this work. First, the focus is long-range and we are projecting outdoors. Second, light behavior properties of the projection surface are computed at pixel level resolution. An overview of our system is shown in Figure 14. In the first stage, the system is calibrated; which includes calibration of the individual cameras, calibration of each camera with respect to other cameras, and the calibration of the projector with respect to the cameras. Next, the geometry of the projection surface is captured using a structured-light scanning technique. Using the images and geometry of the surface, the reflectance properties at each point are estimated. Surface points with complex reflectance properties are identified. Finally, the original stereoscopic content is compensated to account for the reflectance properties of the projection surface prior to projecting it on the surface.

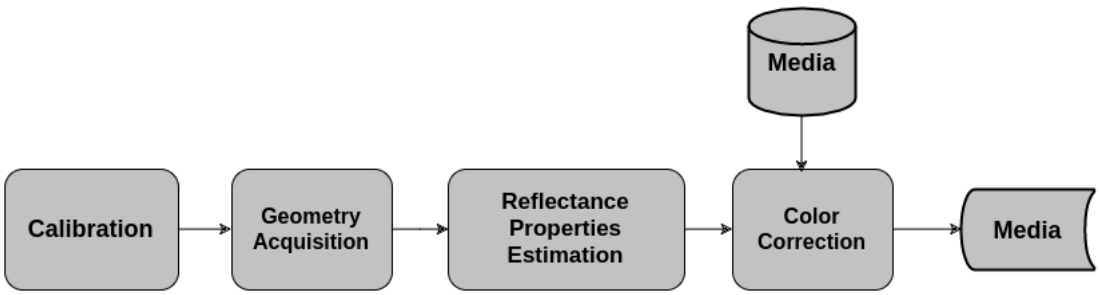


Figure 14: Smart Projection Mapping System Pipeline

As mentioned earlier, we have proposed two methods for long-range projection of stereoscopic content in outdoor areas, the Fixed Camera method and the Roving Camera method. In the first, we encountered the problems caused by limited number of samples of the scene (images from the scene are captured by fixed camera(s)). Therefore in the second, we propose a more robust method which can obtain and handle a very large number of samples from the scene. These methods are discussed further in Sections 3.1 and 3.2

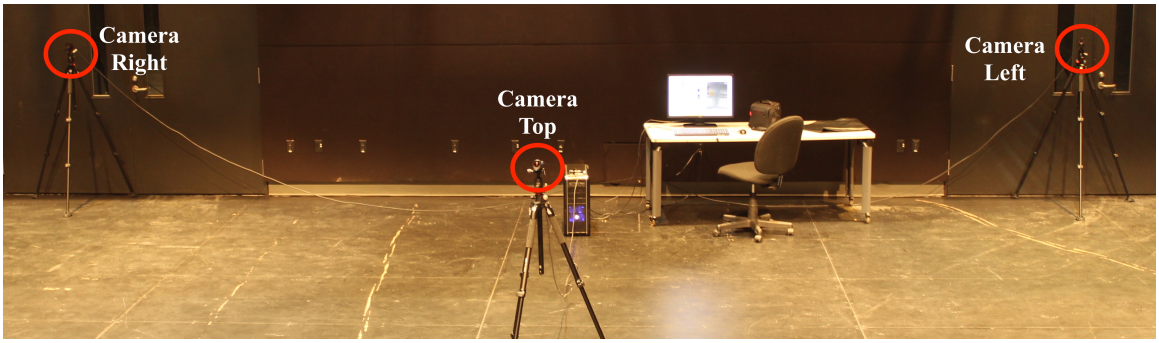


Figure 15: One of our setups while testing the system in long range. Three cameras have been calibrated corresponding to each other using phase-shifting method while they are focused at infinity.

respectively.

3.1 Fixed Camera Method

In this method, we use a minimum of three cameras and a projector, all fixed in their location and orientation in the outdoor area. In the following, their use in each step of the pipeline (Figure 14) is explained.

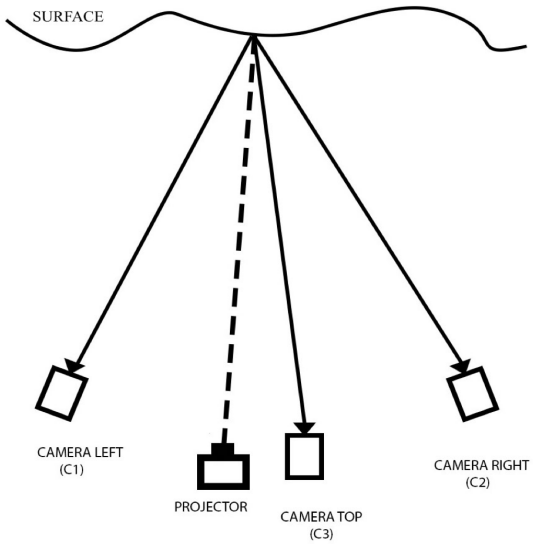


Figure 16: Fixed camera method: this technique requires a minimum of 3 fixed cameras and a long-range video projector.

3.1.1 Calibration

In the first stage, the system needs to be calibrated. Out-Of-Focus camera calibration 2.2.1 is used to calibrate the cameras. Figure 17 shows a camera capturing the fringe patterns. This technique calibrates each camera separately, and the extrinsic parameters of each camera are computed based on the position of the checkerboard on the first image. Therefore we need to either calibrate all the cameras together in one shot, which

is not feasible if we have a larger number of cameras, or calibrate cameras two by two in such a way that it permits linking all of them to each other. To calibrate two or more cameras, we need to capture the fringe patterns simultaneously. Next, each camera extracts its wrapped phase map. Synchronization is necessary to use the same offset unwrapping of their phase map. Therefore, an image with a red dot in the middle of the screen is added to the beginning of the phase-shifting fringe pattern sequence which synchronizes the unwrapping offset among all the cameras. Afterward, by using the traditional technique, we can extract the rotation and translation between two cameras.

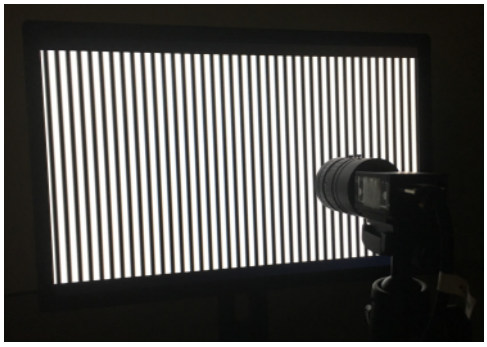


Figure 17: Phase-Shifting camera calibration set-up

The projector is treated as an inverted camera and is calibrated with the traditional method using a set of 2D to 3D correspondences resulting from the geometry acquisition. The explanation on the extraction of the 2D-3D correspondences is deferred to Section 3.1.2. Given a set of 3D world points and their corresponding 2D image locations, the projector's intrinsic and extrinsic parameters are recovered.

3.1.2 Geometry Acquisition

Structured-Light-Scanning (SLS) [HP14] is used to capture the geometry of the projection surface. Encoded patterns are projected onto the surface in sequence. The cameras capture one image per pattern. These images are decoded to produce a dense correspondence between the projector's pixels and the camera's pixels. A rendering of the reconstructed geometry of the Roman Baths (Figure 1) is shown in Figure 18. This was generated from 44 images captured by three cameras.

3.1.3 Reflectance Properties

As part of the geometry, acquisition images are captured from each camera from different viewpoints. Each image records the brightness at every visible surface point from that particular viewpoint. Given three such measurements per surface point, we approximate the local reflectance properties.

We use the Phong illumination model, which is a local illumination model that is easy to compute and use, to describe the local interactions between the material and the light and is given by:

$$I = L\kappa_d \cos\theta + L\kappa_s \cos\phi^{\alpha} \tag{15}$$

where I is the brightness of the reflected light as recorded by the camera, L is the incident radiance emitted

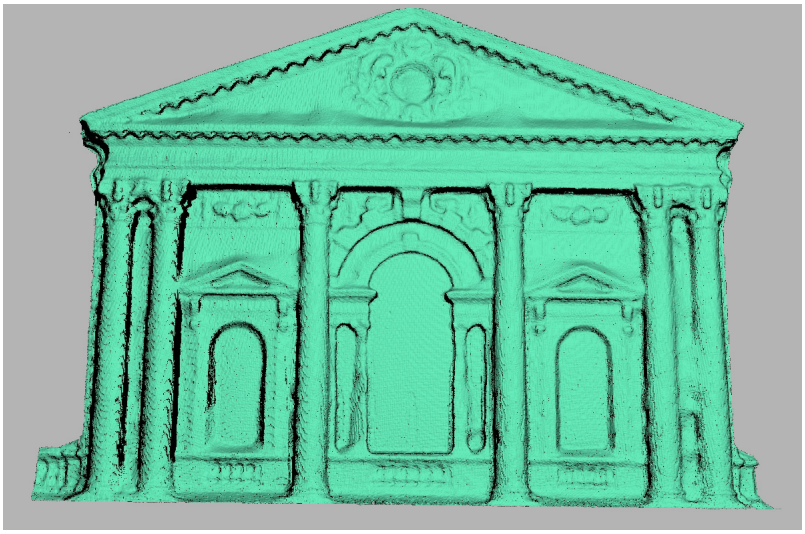


Figure 18: A rendering of the reconstructed geometry of the Roman Bath. 44 images captured by three cameras.

from the projector, K_d is the 3-vector diffuse reflection coefficient, K_s is the 3-vector specular reflection coefficient and α is the shininess coefficient of the material.

A non-linear optimization [Lev44] is used to compute the optimal values for the material parameters such that the energy function $E(f)$ is minimized,

$$E(f) = E_{data}(f) + E_{smooth}(f) \tag{16}$$

where $E_{data}(f)$, the energy data term and $E_{smooth}(f)$, the energy smoothness terms are as defined below.

Energy data term $E_{data}(f)$: This term is a measure of how appropriate the optimized material parameters are, given the observed data. It is defined as,

$$E_{data}(f) = \sum_{i=0}^{n} |I_r^i - I_m^i|^2 \tag{17}$$

where n is the number of cameras, I_r^i is the rendered image as viewed from camera i using the acquired geometry and the material parameters being optimized, and I_m^i is the observed image captured by camera i.

Energy smoothness term $E_{smooth}(f)$: This term is a measure of the smoothness between brightness values in neighboring pixels of the rendered image and is given by,

$$E_{smooth}(f) = \sum_{i=0}^{n} \left[\sum_{j=0}^{w \times h} \left[\sum_{m}^{8} |B_j - B_m|^2 \right] \right] \tag{18}$$

where n is the number of cameras, j is the number of pixels within the rendered image from camera i, m is the 8-neighbourhood around the pixel j, B_j is the brightness at pixel j, and B_m is the brightness at pixel m. This term ensures that optimal values will provide smooth results. This is illustrated in Figure 19 which shows a comparison between the smoothed/non-smoothed procedures for computing material coefficients. For the statue in Figure 19a the specular map without smoothing can be seen in Figure 19b and with smoothing in Figure 19c. As evidenced, without the smoothness term there is noise between neighbouring pixels which is removed when the smoothness term is introduced.

(a) (b) (c)

Figure 19: (a) The object/projection surface. (b) The estimation of the specular parameters without the smoothness term and (c) the estimation with the smoothness term.

Figure 20 shows the results of the energy minimizations for three synthetic test/validation cases shown in Figure 20(a): a perfectly specular, a diffuse/specular and a perfectly diffuse sphere, respectively from top to bottom. Figures 20(b), 20(c), 20(d) show the progress of error minimization corresponding to the shininess, diffuse and specular coefficient respectively.

3.1.4 Color Compensation

Changes in the appearance of the projected content caused by the reflectance properties of the projection surface need to be compensated. We use the estimated geometry and reflectance properties at each surface point to compensate the content prior to its projection. For each frame of the stereoscopic content-to-be-projected, we calculate the compensated projector's brightness L_p for each pixel p as follows,

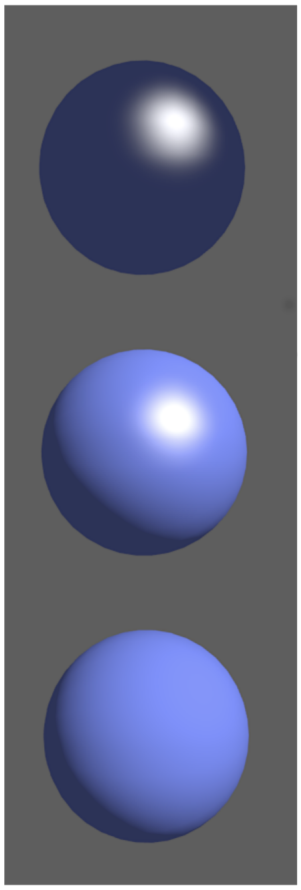
$$\begin{aligned} L_o &= L_p \left[\kappa_d cos(\theta) + \kappa_s cos(\phi)^{\alpha} \right] \\ L_p &= \frac{L_o}{\left[\kappa_d cos(\theta) + \kappa_s cos(\phi)^{\alpha} \right]} \end{aligned} \tag{19}$$

where L_o is the original image's brightness value at pixel p, L_p is the compensated image which, when projected on the surface, ideally yields L_o.

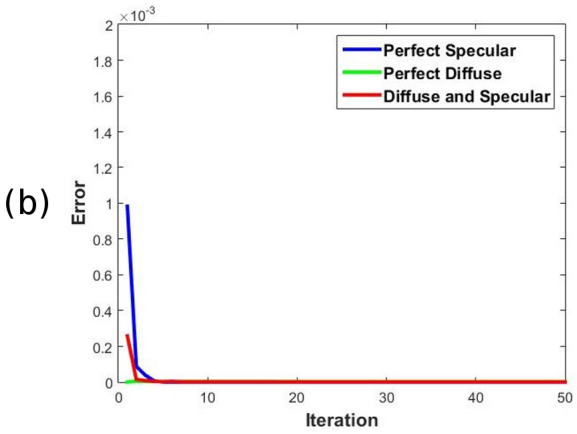
Due to the limited range of brightness values the projector can produce, the colors for which compensation will work are restricted.

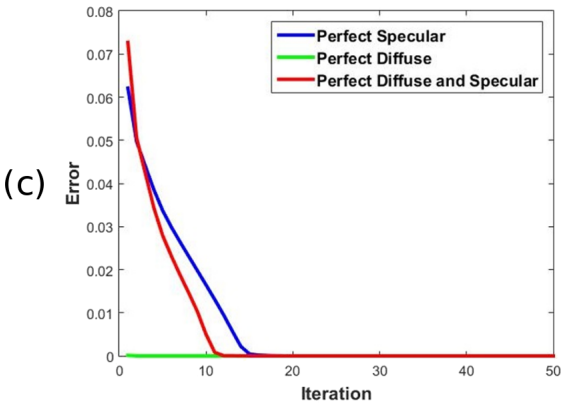
Furthermore, in this work, the content is stereoscopic. This means the range of values is already limited because of the anaglyph processing. We have found that this works well for stereoscopic content being projected on outdoor surfaces, such as building facades, where the surface is primarily diffuse with sharp specular components in the presence of windows, light fixtures and other such objects.

As previously mentioned in section 2.3.1, the reflected color from the object depends on the material's reflectance properties and the emitted light. Knowing the reflectance properties of each surface point allows us to compensate. First, using additive color mixing, we calculate an image which when projected cancels out [if needed] the colors on the projection surface and makes the surface appear grayish. Next, we calculate the image which when projected on top of the 'grayish' surface will be as close to the original as possible.



(a)





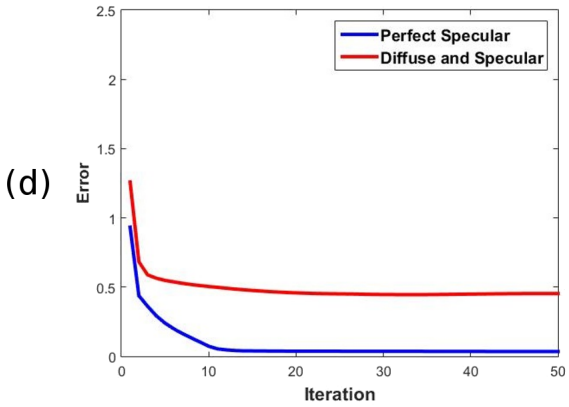


Figure 20: (a) Three synthetic test cases: perfectly specular, perfectly diffuse, and diffuse/specular. (b) Energy minimization for shininess. (c) Energy minimization for diffuse coefficients. (d) Energy minimization for specular coefficients.

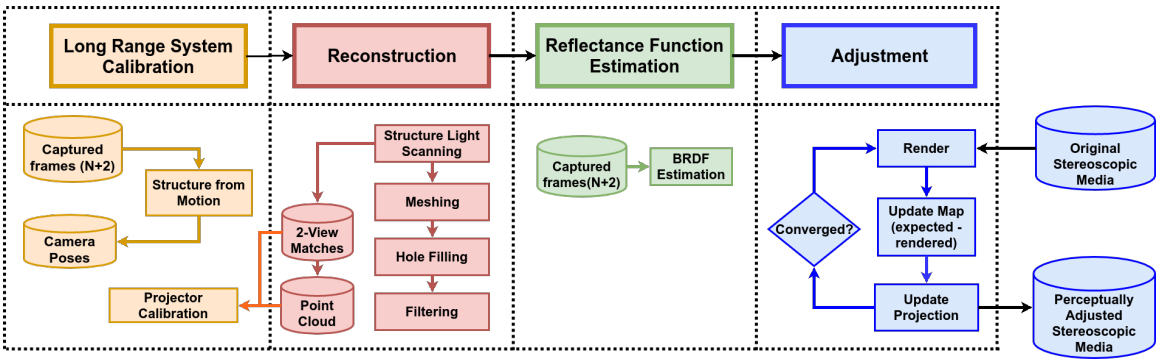


Figure 21: System Overview

The result is the original image with increased brightness and reduced contrast depending on how bright the grayish image is.

3.2 Roving Camera Method

Recovering the reflectance properties of an object/scene requires sampling of the scene with an arbitrary incident light ray direction and view direction of the sampler. Afterward, a suitable BRDF model is needed to fit sampled data. Since it is a model fitting problem, the more samples we have, the more accurate the result will be. In this approach, we mainly focus on covering a wide range of samples from different directions. In this way, the result will be more accurate and, in addition, we are able to use more complex BRDF models.

The first step, as before, is calibration of the system. This involves the estimation of $N+2$ camera poses corresponding to the N frames captured by the roving camera and the 2 frames captured by the fixed cameras. Uniquely, our proposed approach leverages the sub-pixel accuracy of the dense reconstructions resulting from structured light scanning (SLS) techniques with the robustness and accuracy of the camera pose estimations resulting from Structure-from-Motion techniques (SfM). On the one hand, SLS techniques inherently involve a complex calibration procedure which becomes even more challenging as the number of cameras increases or the distance from the projection surface increases; this is due to the fact that the reconstruction can only be performed on the area visible to all the cameras i.e. the intersection area of all views. However, once the calibration is complete, an accurate and dense reconstruction can be generated. On the other hand, SfM techniques make an assumption on the scene's rigidity, and estimate the camera poses and intrinsic parameters, which also can be used to create a sparse reconstruction of the scene, if needed.

In this method, we overcome the difficulties of calibration and subsequent 3D reconstruction imposed by the long-range focus of multiple cameras. We do this by first computing the camera poses and intrinsic parameters using SfM on the set of $N+2$ images captured by the roving video camera and the two fixed cameras. In a second step, the two fixed cameras and a long-range projector are used for SLS which yields a dense reconstruction of the scene. This involves (i) projecting encoded Grey-coded patterns, (ii) capturing the patterns with the two cameras, (iii) decoding the patterns, and (iv) identifying the per-pixel matches. The per-pixel matches are then triangulated to produce a dense point cloud which is further processed to generate a dense mesh representing the projection surface.

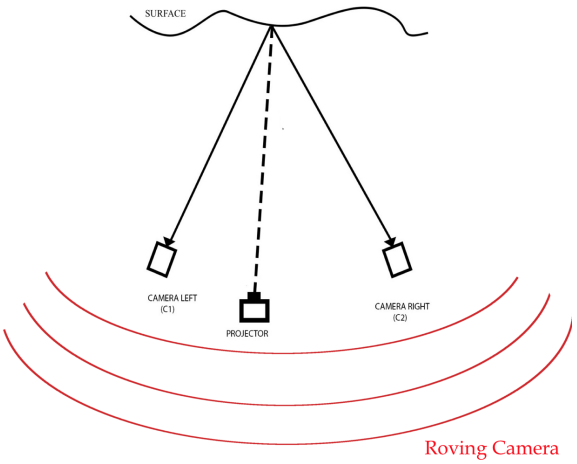


Figure 22: Roving camera method: this technique requires a roving camera which captures N samples from the scene by sweeping in front of it; A minimum of two fixed cameras for geometry reconstruction, and a long-range video projector.

With a fully calibrated system, the per-pixel reflectance function is estimated using non-linear optimization on the $N+2$ observations. This results in a reflectance map of the projection surface which is subsequently used for the adjustment of the images/videos to be projected. Figure 21 shows a diagram of the proposed framework summarizing the four processing modules: long-range system calibration, reconstruction, reflectance function estimation, and adjustment. Each of these processing stages is described in greater detail in the following sections.

We employ SfM for the calibration of the cameras. SfM involves capturing a large number of images from different viewpoints and performing feature extraction and matching using Scale-Invariant Feature Transform (SIFT). The roving video camera is swept randomly in front of the projection surface, which captures the surface from different viewpoints and view angles. At the same time, the two fixed cameras are focused on the projection surface. This process yields $N+2$ images. Bundle adjustment is then performed to simultaneously refine the parameters of the camera motion $[R|t]_{3x4}$, intrinsic camera parameters K_{3x3}, and the 3D points P in the scene while minimizing the re-projection error in all the images.

Figure 23 shows an example of the results obtained from SfM. Each camera is represented by its position and its oriented image plane. SfM generates a sparse reconstruction of the scene's structure which is also shown in the figure. However as we need a dense point cloud to accurately reconstruct the projection surface geometry, we do not use this sparse point cloud in further processing. Instead, with these calibrated fixed cameras, we resort to the use of the SLS technique to obtain a dense 3D reconstruction.

Once the cameras are calibrated, we proceed with the estimation of the projector's pose and intrinsic parameters with respect to the calibrated cameras. Traditional techniques for camera-projector calibration such as [BXZ16, ABL$^+$16, MT12, FHMF09] suffer from similar problems when applied to long-range systems. These techniques require projecting a pattern [usually a checkerboard] onto [or next to] the calibration board and detecting the corners in order to estimate the projector parameters. In other words, the projector is treated as an inverted camera. However, projective geometry dictates that the farther away you move from the projector, the larger the projected pattern. Hence, it becomes almost impossible to employ these techniques in

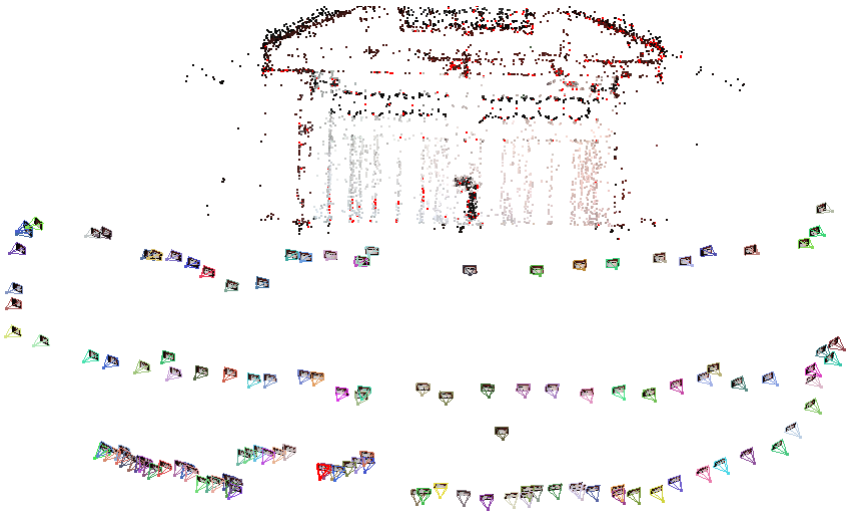


Figure 23: Long-range camera calibration. Traditional calibration techniques fail for long-range vision systems. In the proposed approach we use SfM for estimating the camera poses and intrinsic parameters, thereby eliminating the need for special calibration boards and procedures [BXZ16, ABL+16]. Setup shown for experiment #1.

long-range scenarios.

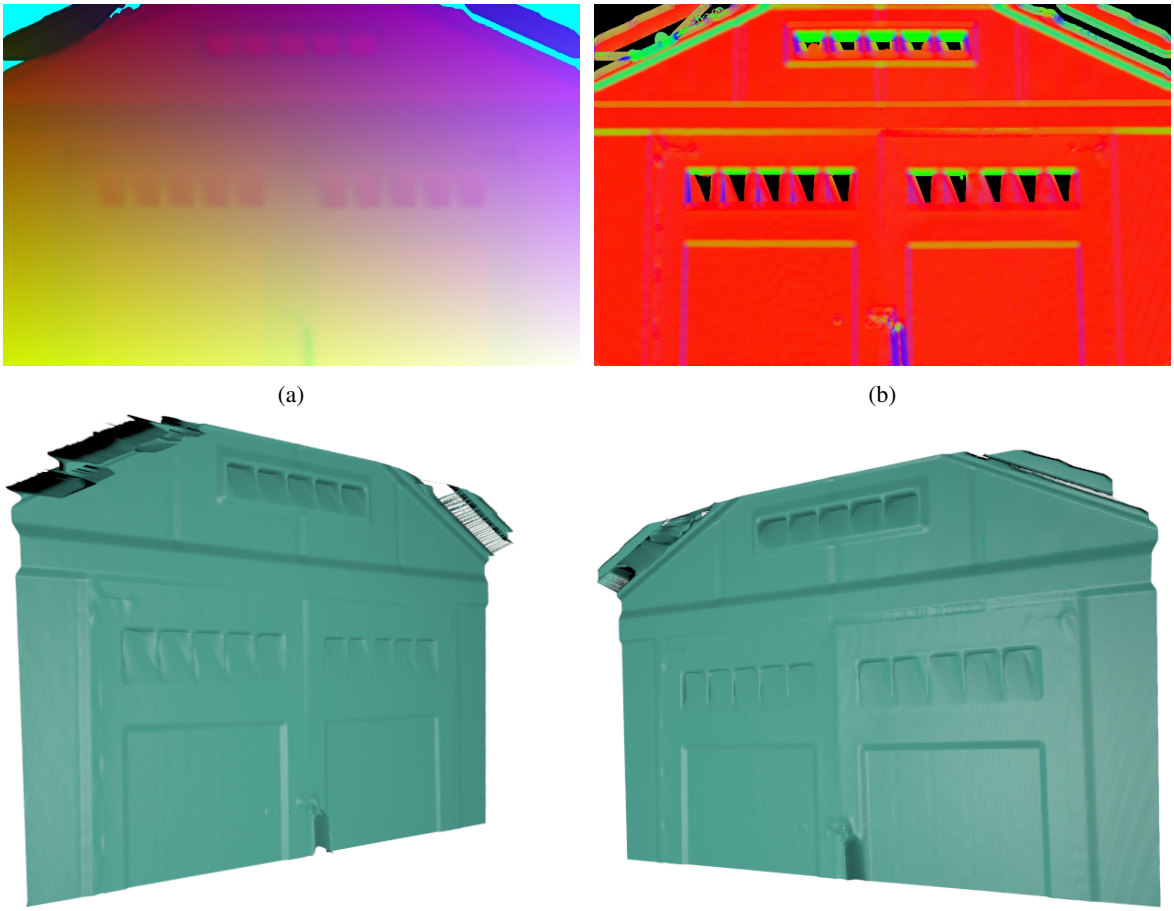
To overcome such problems, we employ SLS [GHP14] for simultaneously (a) calibrating the projector with respect to the other cameras and (b) capturing the scene's geometry accurately. This process involves projecting encoded [with each pixel's location] patterns which are captured by the two fixed cameras and are then decoded to identify each pixel's location. Given the dense correspondences between the two images, a 3D position can be computed by triangulation. The result is a dense reconstruction of the scene's structure which in combination with the correspondences in the 2D projected image can be used to estimate the projector's pose and intrinsic parameters as in [Zha00].

As previously stated, SLS yields a dense point cloud representation of the projection surface, which is depicted as an XYZ map as shown in Figure 24a. Holes can result due to occlusions. These are filled with neighborhood information. Bilateral filtering is used to remove noise if present in the mesh while at the same time preserving the edges. Finally, a mesh is created by triangulating the nearest neighbors in the XYZ map; an example is shown in Figure 24.

The quality of the reconstructions using the above method was quantitatively evaluated by reconstructing known objects and comparing them with the ground truth. For objects located at a distance of $6m$ the estimated surface fitting error [RMSE] between the ground truth and the reconstructed object was of the order $0.1cm^2$ which is comparable to the errors reported in [GHP14].

3.2.1 Bidirectional Reflectance Distribution Function

The Bidirectional Reflectance Distribution Function (BRDF) represents the light behavior (reflection) properties of the projection surfaces. The BRDF is estimated based on the per-pixel samples captured by the roving camera and the two-fixed cameras. Given the fact that during the projection mapping the audience is typically spread over a limited area in front or next to the projector's direction, we only consider the BRDF



(a) (b)

(c) (d)

Figure 24: Experiment #1: (a) The XYZ map of the scene's structure generated with SLS. (b) The Normal map of the scene's structure generated with SLS. (c,d) Render from two novel viewpoints of the reconstructed geometry.

over a range of incident light and viewing directions ranging from $0 \leq \theta \leq \pi$ on the horizontal and $0 \leq \phi \leq \frac{\pi}{4}$ on the vertical.

The Lafortune BRDF [LFTG97b] is another alternative for data fitting. This analytical model leverages the simplicity of the Phong model while capturing realistic BRDFs from measured data. It should be noted here that our method of fitting the measured data to a BRDF analytical model is not limited to Phong and Lafortune models.

We use the $N+2$ per-pixel samples captured by the cameras to recover the BRDF parameters via non-linear optimization. To sample the incident light from the limited area mentioned earlier, we perform a random walk using the roving video camera in front of the building (projection surface) so as to cover the surface and obtain an adequate number of samples. During the estimation, the per-pixel normals computed from the XYZ map are also incorporated. The result is the per-pixel BRDF; a diffuse map is shown in Figure 25a and the corresponding specular map is shown in Figure 25b. The inset picture in Figure 25b is a close-up of the region indicated with red and demonstrates the high level of detail we are able to capture using this

(a) (b)

Figure 25: diffuse map(left), specular map(right). The inset picture in the specular map (b) is a close-up of the region indicated with red and demonstrates the level of detail captured by this approach.

approach.

3.2.2 Compensation

The colors of the original image are adjusted such that any interference or colour distortions due to the surface's geometry and reflectance properties are mitigated to the fullest extent possible. Using the recovered scene geometry, the surface's per-pixel BRDFs, and camera poses, an image is rendered from the projector's viewpoint and the colour difference is iteratively diminished. The steps are summarized in the algorithm below,

$i \leftarrow 0$ $I^i_{rendered} \leftarrow Raytrace(K[R|t], I_{specular}, I_{diffuse}, I_{original})$ $||I^i_{rendered} - I_{original}||^2 \geq \tau$ $I^i_{update} \leftarrow I^i_{rendered} - I_original$ $i \leftarrow i+1$ $I_{adjusted} \leftarrow I^i_{rendered}$

where $Raytrace(.)$ is a physics-based renderer [PJH16] which given the [inverted] camera parameters $K[R|t]$ i.e. projector, the two maps containing the specular and diffuse coefficients for each surface point $I_{specular}, I_{diffuse}$, and the original images $I_{original}$ produces a render from a virtual camera [with identical pose and parameters to the projector] of how the original image will appear if projected onto the surface. During this simulation, the projector is modeled as an array of $N \times M$ point light sources each emitting light only to their corresponding surface point, where $N \times M$ is the size of the original image $I_{original}$.

The second row in Figure 28 shows an example of the adjusted image after a single iteration (28c) and upon convergence of the algorithm (28d). The iterative optimization is performed offline, since for an image with a resolution of 1600×1200 it typically takes 4 minutes to converge due to the rendering required at each iteration. Given this was a proof of concept, the computations have not been optimized, though clearly there is ample scope for it.

3.3 Implementation

Our framework is implemented in the C++ language, and the OpenCV library is used for most of our computations. In the following section, we describe the implementation detail for each method separately.

3.3.1 Implementation of Fixed Camera Method

The implementation follows the same pipeline shown in Figure 14. First the system is calibrated using phase shifting patterns, then the geometry is acquired using the structured light scanning method/reflectance properties recovery, and finally the compensation of the media is applied.

Phase Shifting Camera Calibration

An application is needed to display phase shifting patterns. In this application, we generate vertical and horizontal phase maps using the Equation 4. Then fringe patterns are produced using Equation 3. In this step, we need to decide and determine the feature points' location on the monitor. Since we have a grid, we just need to keep the first row and the first column of the feature points (Figure 6). Next, we need to display and capture patterns. For a faster and better result, we first show a black and a white image, then we can generate the mask image by subtracting them. In the monitors that have a reflection on its frame, we added a black border to the white image which makes our mask image more accurate. During the calibration process, we need the exact metric distance between two feature points, while so far we only know the distance in pixels. With this in mind, we generate a checkerboard based on our feature points. By displaying it, we can measure the exact distance between the feature points (corners of the checkerboard). For testing the program after decoding, we can check if the extracted feature points match the corners of checkerboard.

After capturing the patterns 10-15 times with different monitor orientations, each set of patterns should be decoded. Each pose is independent of the others, which means that multi-threading is beneficial. We assign each sequence a thread to decode the patterns. Although we did not use the faster GPU implementation in this case, each pixel in a sequence is also independent and could potentially have its own thread. Equation 5 is used to extract the phase value of a pixel. However, this will produce a wrapped phase map. The value of K in Equation 6 is known in the middle of the screen. We call it $K_{start} = (K_{min} + K_{max})/2$. Then by applying edge detection, we can identify where exactly K is changing. We start with K_{start} from the middle of the screen (already detected via the "Red Dot Pattern") and traverse all the other pixels. Every time we reach an edge, we either add or subtract one to K ($K++$ or $K--$). By storing all the K values in a matrix, we can unwrap the phase by

```
1   Unwrapped = Wrapped + (K * PI);
```

The next step is matching the computed phase value of each pixel with our predetermined values. A point is a feature point if both the vertical and horizontal phase amount of the pixel matches our values. In some cases, more than one pixel (neighbors) matches with an encoded feature point. We have chosen the average of the pixels as a feature point. Now we are able to calibrate the camera using OpenCV's "calibrateCamera" function. After calibrating all the devices separately the following equation will return the rotation and

translation of a device which corresponds with another one.

$$R = R_1^t * R_2 T = R_1^t * (T_2 - T_1) \tag{20}$$

Where R_1 and R_2 are the rotation matrices and T_1 and T_2 are translation matrices of the first and second device respectively. As in the traditional method, during the calibration process both devices should capture one pose of the monitor simultaneously to have a common origin for their calibration.

SLS

We have used the approach presented by [GHP14] for structured light scanning and bilateral filtering. Another minor modification that was also applied was the averaging of the 3D reconstruction of 2x2 cameras to support multiple cameras.

Estimation of Reflectance Properties

In this step, we estimate the color of the object. The Phong Model represents a simple and fast reflection model. To estimate K_d, K_s and α we optimize the Phong Model with our collected data.

$$I = LK_d \cos\theta + LK_s \cos\phi^{\alpha} \tag{21}$$

To compute $\cos\theta$ and $\cos\phi$, we need the position of the projector and cameras in the world coordinate system. The following equation will yield the position of the device based on its origin.

$$T = \begin{bmatrix} 1 & 0 & 0 & -T_x \\ 0 & 1 & 0 & -T_y \\ 0 & 0 & 1 & -T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{22}$$

$$P = R^t * T \tag{23}$$

$\cos\theta = N.L$, we already have the surface normals from SLS which allows the calculation of the light direction via

$$L = Normalize(P_{light} - P_{surface}) \tag{24}$$

Also $\cos\phi = R.V$. R and V can be computed by

$$V = Normalize(P_{camera} - P_{surface}) \tag{25}$$

$$R = L - ((2 * (N \cdot L) * N); \tag{26}$$

After computing $\cos\theta$ and $\cos\phi$, we need to minimize the Phong equation to find the unknowns. For each pixel, we should have at least three samples. We already have the correspondences between all the devices from SLS. Projection of a white image with maximum power ($L = [1.0, 1.0, 1.0]$) gives us a sample from the scene.

We use C/C++ cminpack [Dev07] Levenberg-Marquardt algorithm implementation for our optimization. The following pseudocode shows our energy function for optimization.

```
1 for (int camID = 0; camID < m; ++camID) {
2     cosPhi_Alpha = pow(cosPhi,alpha);
3     fvec[camID] = pow((((L * Kd * cosT + L * Ks * cosPhi_Alpha)) - (I)),2) * 1000;
4 }
```

As it is shown in the code, we use the square of the error and multiply by an arbitrarily large number to make the optimization more accurate. However, this code does not include the smoothing energy. The difference between the intensity at a pixel and the intensity of its neighbors is an additional term added to our energy function. For a more accurate result, we have weighted the neighbors based on their distance from the pixel. By estimating K_d, K_s and α we are able to simulate the surface behavior. Using Equation 19, we compute our projection media.

3.3.2 Implementation of Roving Camera Method

There are two main distinguishing differences between this implementation and the previous one. An entirely new method of camera calibration and a new algorithm for media compensation are introduced.

Camera Calibration and Reconstruction

We calibrate cameras by using the Structure from Motion technique. **VisualSfM** is a well established SfM application, and we use this application for our calibration process. We pass all the captured images which include the fixed and roving camera images. After running the SfM and the bundle adjustment, the "N-View Match" file will export the workspace which includes image path, 3D model, and camera information in text format. Below is the NVM file format for camera information.

```
1 <Camera> = <File name><focal length><quaternion WXYZ><camera center><radial distortion> 0
```

VisualSfM is an open source application. Thus we can use their implementation for loading the NVM format. It requires "LoadNVM" function, "DataInterface.h" and "interface.h" classes.

Note that, we need the calibration of all cameras to occur in the same coordinate system. To accomplish this, we run SfM for all the images together (roving and fixed cameras/images). We use calibrations for SLS. The same implementation of the SLS is used in this approach. However, in this version, since we do re-projection and ray tracing, hole filling will improve the resolution of our projection. We have followed the approach presented by Poullis [Pou13] for hole filling.

Reflectance Properties and Compensation

The next step is finding the correspondences between image pixels and 3D points. Each 3D point is re-projected onto the image plane by using OpenCV's "projectPoints" function. This method receives a vector of 3D points and the camera calibration information as input. The output is the projection of 3D points onto the camera image plane. It is important to re-project the 3D points onto the undistorted image. Nearest neighbour or linear interpolation can be used for getting color after re-projection onto the image. The re-projection is valid if it is inside the image and there is no self-occlusion in the scene.

Many samples have been captured and extracted for each 3D point. Due to the large number of the samples, we are able to fit our data into more complex BRDF models like Lafortune. However, it also works

perfectly with the Phong Model. Once we have recovered the reflectance properties of the scene, we should ray trace the scene to get the projection result. PBRT [PJH16] (a Physically Based Ray Tracing renderer) has been used to ray trace and render the scene. In our test cases we suppose that the projector and camera are in the same location, so we update our projection as follows:

```
1    Update_map = (Expected_Image - Rendered_Image);
2    Projection_Image += Update_map / step;
```

Division of the update_map by $step$ guarantees that our rendered_image will converge to the Expected_Image.

We update the Projection_Image 10 - 15 times. In each iteration we also check that the RGB values are between zero and one. Otherwise we clamp the value to 0 or 1 and change the others to keep the ratio of R, G and B the same. After clamping the values, we stop updating the pixel in the next iteration. The result will be the compensated version of our image according to the scene.

Using the above implementations we conducted a number of experiments of stereoscopic projection mapping in real world situations and user studies through random participants who were present/passing through the experimental sites. These experimental setups and user study results are described in the following chapter.

Chapter 4

Experiments

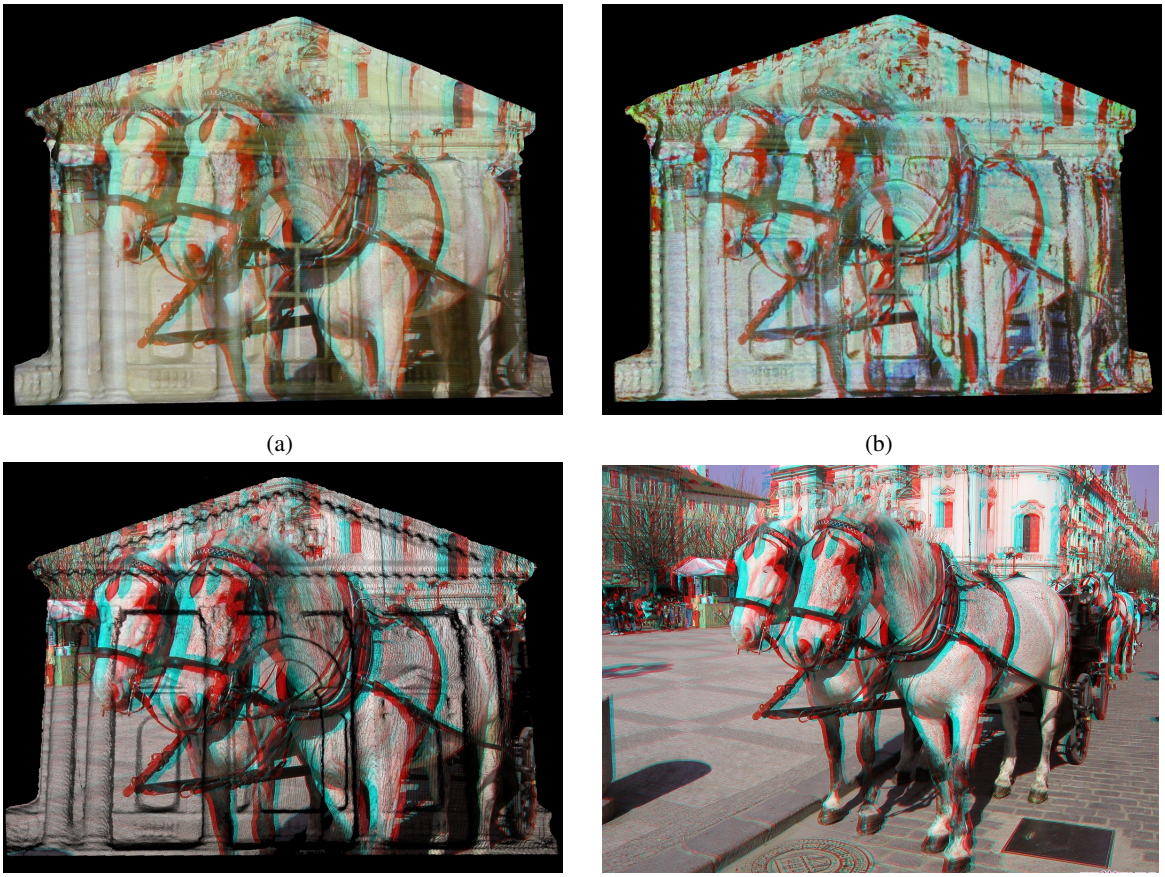
In this chapter we explain our projection mapping experiments for each method. The chapter is organized as follows: in section 4.1 we show the results of the Fixed Camera method and in section 4.2 the results of the Roving Camera method. These include one short range experiment and two long range experiments.

4.1 Fixed Camera Method

All reported results were generated using an implementation of our framework on an Intel-i7 PC with commodity hardware. The projector used was a Panasonic PT-VW435N projector with a native resolution of WXGA 1280×800. The statue of Alexander shown in the experiments has dimensions $25.5cm \times 19cm \times 14.5cm$ and was used for comparison purposes with [HP14].

A 3D print of the Roman Baths was used for experimentation due to access restrictions on the real site. These experiments were conducted in relative scale. Figure 1 shows a projection on the real Roman Bath's building. Figure 26a shows a stereoscopic projection [without compensation] being projected onto the Roman Baths. Color distortions occurring on the anaglyph image due to the reflectance properties of the projection surface have a negative impact on the depth perception of the viewer. Figure 26b shows the stereoscopic projection after compensation using the proposed technique. Color distortions are minimized by taking into account the effect of the reflectance properties of the projection surface. A render of projection is shown in Figure 26c which is the expected result.

The expected projection cannot always be achieved [as in the this case] because of the limitations of additive color mixing and the hardware. Intuitively, a projection surface with a white-ish color can reflect a larger percentage of the projected light by the projector, therefore, more colors can be compensated. For example, a bright red color projected onto a white surface will appear as red. On the other hand, a projection surface with a darker color will absorb the projected light; a bright red color projected onto a dark surface will appear as dark red.



(a) (b)

(c) (d)

Figure 26: (a) A stereoscopic projection without compensation. Color distortions due to the reflectance properties of the projection surface negatively affect the depth perception of the viewer. (b) The stereoscopic projection after compensation using the proposed technique. Color distortions are minimized by taking into account the effect of the reflectance properties of the projection surface. (c) The expected projection. This cannot always be achieved due to the limitations of additive color mixing and the hardware.(d) The original stereoscopic image taken from [ste]

4.2 Roving Camera Method

The proposed framework was tested first in the lab environment and was then evaluated by human participants during two separate experiments. In both experiments, we wanted to expose participants to the projection mapping of stereoscopic content, with the goal of assessing the effectiveness of the proposed approach with respect to color and depth perception, and the sense of immersion. This was done through questionnaires which the participants had to fill out before, during, and after the experiments. The research team conducting the experiments included three computer science researchers and a fine arts scholar. The experiments received approval by the University Research Ethics Committee and informed consent forms were obtained from each of the participants. Furthermore, the participants were asked to fill out a demographic and background information form. This established their prior experience, if any, with anaglyph 3D stereo. They were informed that they could withdraw from the experiment any time. The consent form and questionnaire are

attached to the end of this chapter on pages [45 and 46]

4.2.1 Indoor Testing

This test was conducted in one of the university department's reception rooms. The red color of the wall as well as three paintings of various colors [shown in Figure 27a] test the accuracy and performance of the system. In this test, the projection was from mid-range [i.e. around 5 meters] using a Panasonic PT-VW435N projector with a native resolution of WXGA 1280×800 and three Pointgrey Grasshopper 3 cameras [two fixed and one roving]. A 3D anaglyph video [You] was projected on the wall once without compensation and then again after compensation. As is evident in Figure 27 the projected frame after compensation shows far more detail. Colors closer to the original coloring of the frame [so-called "true colors"] are also apparent. As an example, in the top half of the middle painting, trees and their separation from the sky are more prominent and the viewer can enjoy a fuller and more vibrant scene. The specularity and shiny reflectance were also removed, which better emphasizes the projected pathway.

4.2.2 Public Place Experiment #1

The first experiment involved projection mapping of anaglyph 3D images on a shed. The shed was already installed as an exhibit in an indoor public place, the lobby of a building. The shed was chosen because of its red color which would interfere with the projection of the red/cyan stereoscopic content, therefore, causing problems with the color and depth perception of the viewers. The doors were white and provided a reference to minimum change.

In this experiment the projection was from a medium-range [i.e. a little over 3 meters] because we specifically wanted to address only the following questions without introducing possible bias due to the distance of the projection:

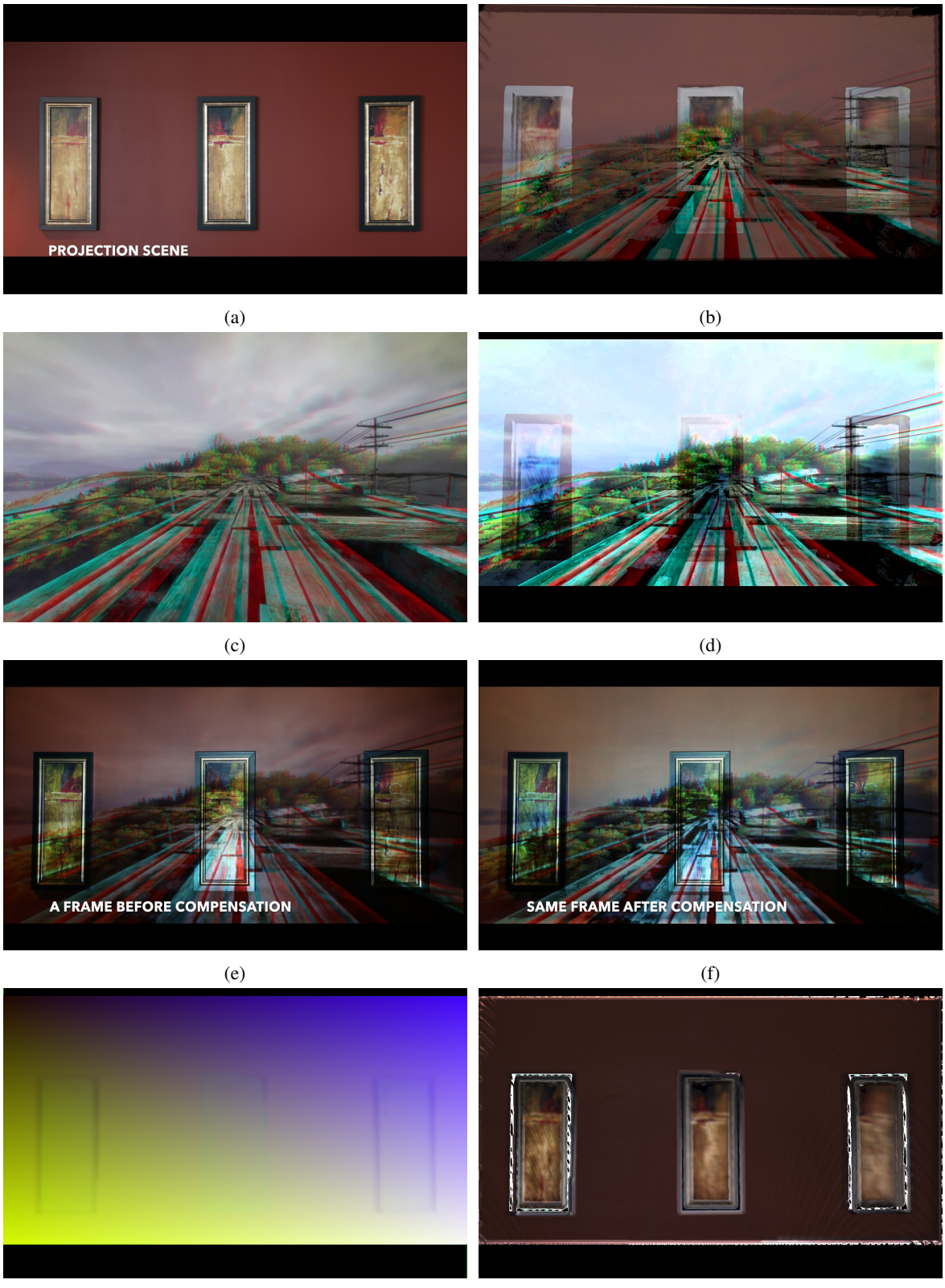
- Does projection mapping of anaglyph 3D content on surfaces containing red or cyan colors cause interference with the color and depth perception of the viewer?
- Does the proposed approach improve the color and depth perception of the viewer?

Participants

This experiment involved 34 participants [70.6% male, 29.4% female] which according to the demographic form ranged from 19-39 years old [35.2% within the age group 25-26]. Of the 34 participants 44.4% had graduated [or were in the process of] from an art field, and 55.6% had graduated [or were in the process of] from an engineering field. These participants were random people who were around in the venue during the time the stereo projection was set up. The majority [55.8%] of the participants indicated that they had prior experience with anaglyph 3D although 2.9% of them could not perceive 3D.

Setup and Equipment

As previously mentioned, this experiment involved projection mapping onto a red shed with white doors. The size of the projection surface was $1.35x2.40m$ and the distance of the projector $3.3m$. A projector



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Figure 27: Our test in lab environment. (a) Projection surface, (b) rendered scene. (c) one frame of our projection media before compensation, (d) same frame after compensation. (e) Projected scene with a frame without compensation, (f) adjusted same frame after compensation, (g) XYZ map and (h) diffuse map of the scene.

[SANYO PLC-ZM5000L] was used with a native resolution of 1920x1200. The two fixed cameras [Pointgrey Grasshopper 3] with a resolution of 1920x1200 were placed at a distance of 3.5. A third similar camera [Pointgrey Grasshopper 3] was used as the roving camera to capture over 200 images of the scene.

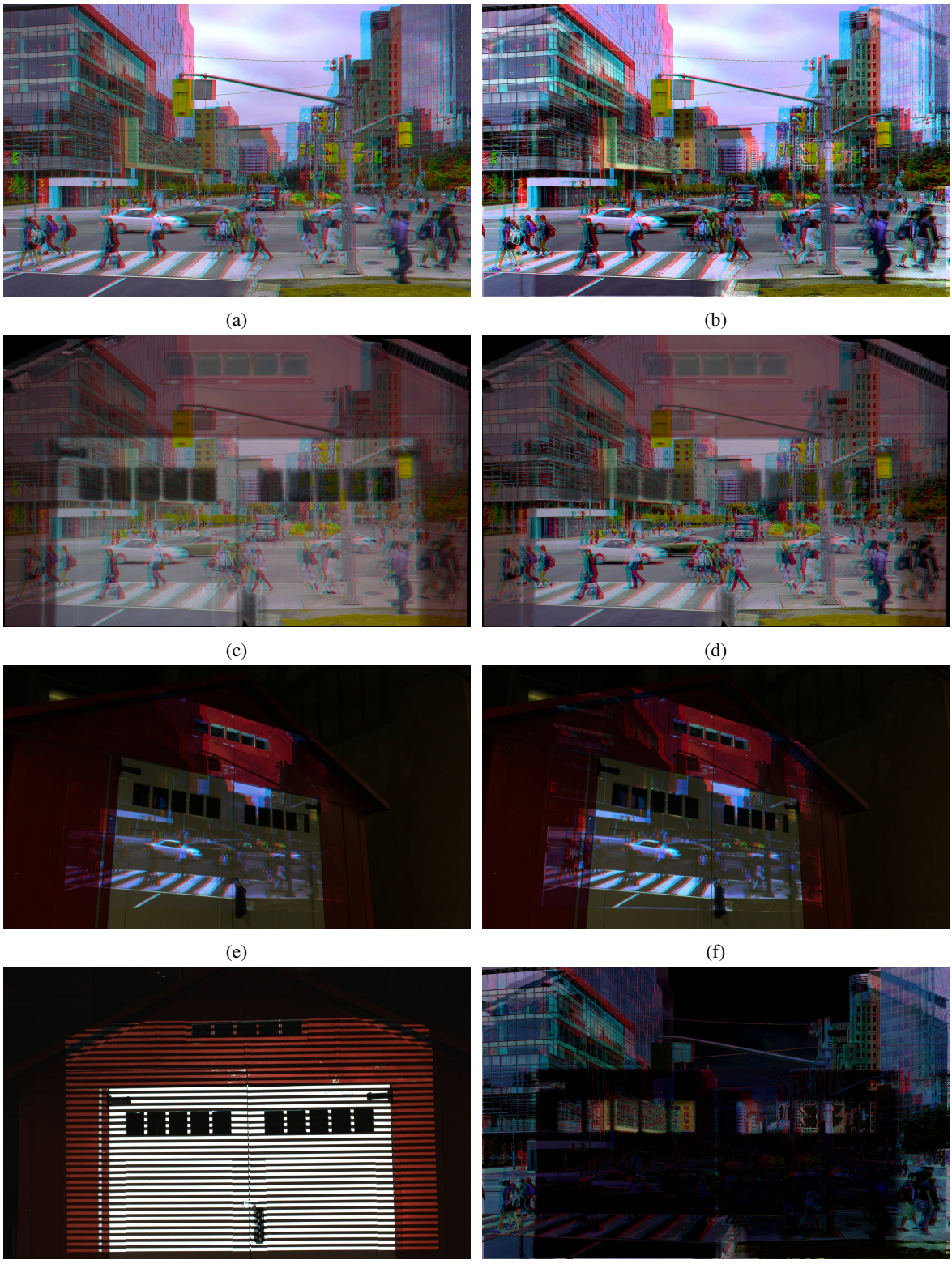
Procedure

The participants were provided with anaglyph 3D glasses so that they could observe the stereoscopic projection. Initially, the original non-compensated stereoscopic image was projected and after 10 seconds the adjusted stereoscopic image was projected. The participants were allowed to move freely in the general area of the projection during the experiment and observe the projection from different viewpoints.

Results

We present the qualitative and quantitative assessment of the results from this first experiment. Figure 28 shows the stereoscopic content before and after the application of the proposed approach. The first row shows the original non-compensated image (left) and the adjusted image (right). The middle row shows the projection of the original image (left), and the projection of the adjusted image (right) onto the scene. In the images in the first row, the change is almost not noticeable, however as it is evident from the images in the second row there is a significant difference in the projections of the two. For example, the content projected onto the red areas of the scene is almost not distinguishable in the original image which results in very poor or no depth perception. On the other hand, the projection of the adjusted image appears brighter and perceptually correct in these areas. Finally, the third row shows the resulting image after a single iteration of the adjustment process (left) and the final adjusted image upon convergence of the algorithm.

Figure 29 shows a table of the statistical significance between the participants' questionnaire responses about any perceptual change in the projection before (vertical) and after (horizontal) the application of the proposed approach. The question for both projections was 'Rate your color and depth perception for this projected image/video'. A Likert scale was used ranging from [1,10]. The table shows a high statistical significance for the improvement in the perception of the participants with the adjusted images. In particular, 50% of the participants who had rated the perception of the projection of the original image with a '3' had rated their perception of the adjusted image with a '5' which is higher than the average of '3'. Similarly, 45% with an initial rating of '6' had increased their rating to '8', 67% with an initial rating of '7' had increased their rating to '10', and 33% with an initial rating of '9' had increased their rating to '10'. In Figure 30 we show the distributions of the participants' responses for these questions.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 28: An example image used in Experiment #1. First row: (left) original image, (right) adjusted image. Second row: (left) projection of original image on scene, (right) projection of adjusted image on scene. Third row: (left) adjusted images after one iteration, (right) adjusted image upon convergence of the algorithm. (g) Capturing the geometry by SLS and (h) is the update map (Absolute difference between (a) and (b).

Column %	5	6	7	8	9	10	NET
3	50% ♦	0%	0%	0%	8%	0%	6%
4	50%	0%	40%	18%	17%	0%	21%
5	0%	100%	20%	18%	8%	0%	15%
6	0%	0%	20%	45% ♦	17%	0%	24%
7	0%	0%	0%	9%	33%	67% ♦	21%
8	0%	0%	0%	9%	17%	0%	9%
9	0%	0%	20%	0%	0%	33% ♦	6%
NET	100%	100%	100%	100%	100%	100%	100%

Figure 29: Experiment #1: Statistical significance of participants' questionnaire responses between before (vertical) and after (horizontal) the application of the proposed approach. The question for both projections was 'Rate your color and depth perception for this projected image/video'. A likert scale was used ranging from [1,10]. 95% confidence level. Sample size: 34.

4.2.3 Public Place Experiment #2

The second experiment took place in an outdoor public space at night and involved a long-range projection onto a building's facade containing glass windows, complex decorative sculptures, carved columns, etc. with no control over the illumination conditions.

In this experiment, the projection was from a distance of over 20 meters. Participants viewed the projection standing on the pavement of a downtown street. The experiment was conducted over two nights, as during the first night it was drizzling. A good number of people did watch the projection in spite of the rain. The experiment was designed to focus on the following:

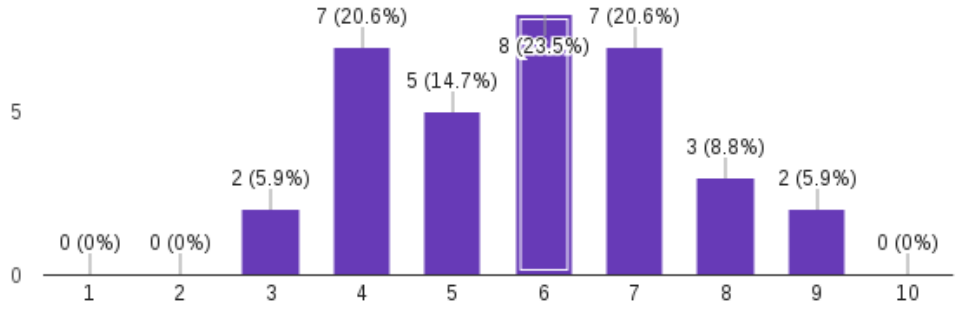
- Is the proposed approachable to enhance the color and depth perception of the viewer for long-range projections in outdoor areas?

Participants

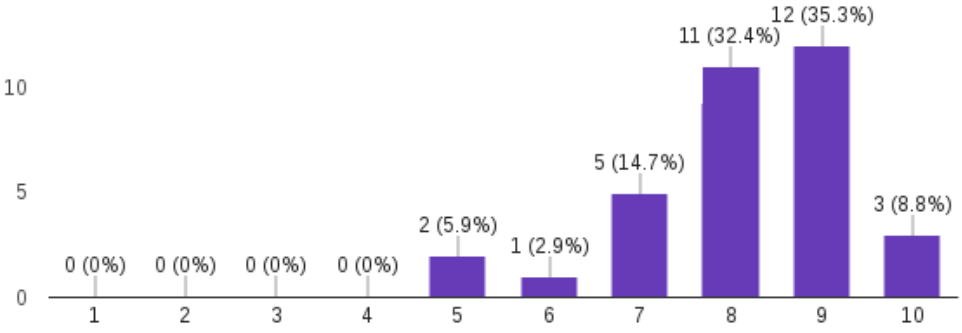
The second experiment involved 37 participants [64.9% male, 35.1%] over two days chosen at random from the street, which according to the demographic range from 18-68 years of age [62.1% within the age group of 22-29]. Of the 37 participants, 73% had an engineering, 16.2% arts, 8.1% social sciences background. One participant reported he/she had suffered or was prone to epilepsy or seizure and did not take part in the experiment. 78.4% reported that they were able to perceive depth when using anaglyph 3D glasses in the past.

Setup and Equipment

The experiment involved projecting stereoscopic content from the third-floor window of a building onto the facade of another building across a busy street at a distance of over 20 meters. The third column in Figure 33 (top) shows the setup used for the second experiment. The projector and cameras remained the same as in experiment #1 and were used in a similar fashion for all the steps in the pipeline.



(a)



(b)

Figure 30: The distributions of the responses to the question 'Rate your color and depth perception for this projected image/video' with the original image (top) and adjusted image (bottom)

Procedure

Similar to experiment #1 the participants were provided with anaglyph 3D glasses. Two stereoscopic videos [(a) a butterfly flying, (b) a roller coaster] were projected first without, and later with the adjustment. The participants were again allowed to move freely in the general vicinity in order to observe from different viewpoints.

Results

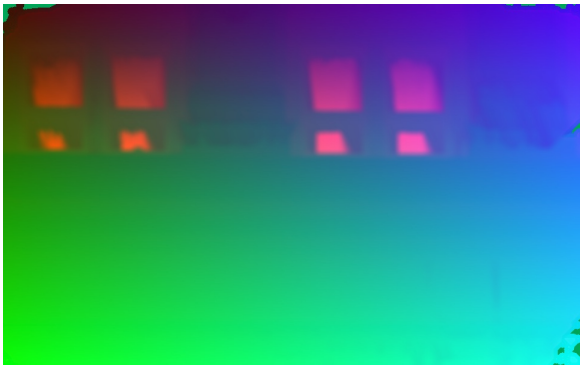
The analysis of the responses of the second experiment indicates a significant improvement in the color and depth perception of the viewers. The first and second columns in Figure 33 show a sample frame from the butterfly video, original and adjusted, and projected onto the building facade. Figure 32 shows the statistical significance between the ratings of the participant's perception before and after the experiment for the butterfly video. 83% of the participants who initially rated their perception of 3D with a '3' had increased their rating to '4' when viewing the adjusted video which is higher than the average for '4'. Similar reported improvements can be noted for those who initially rated their perception with '1'. Almost identical results were reported for the second video with the roller coaster. There was an 8% decrease between the two categories for the rating of '3' which we believe was due to the responses of the participants of the first night's experiment. Because of the drizzle, the facade had an even darker color than the one captured and used to calculate

the adjusted video. In the categories of 'Rate your perception of color and depth', 'Have you experienced nausea, dizziness, or eye strain?', and 'Were you able to perceive 3D?', there were no significant differences from the earlier experiment.

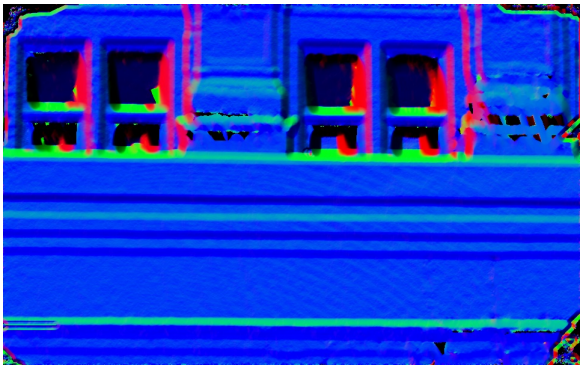
The above experiments have shown that the proposed methods do indeed help improve the quality of colour and depth perception, and correspondingly the sense of immersion in long-range out door projection mapping.

(a)

(b)



(c)



(d)

Figure 31: In this figure (a) shows our projection setup, (b) shows the distance and direction of our viewers and projection (c) shows the XYZ map and (d) is the normal map of the scene

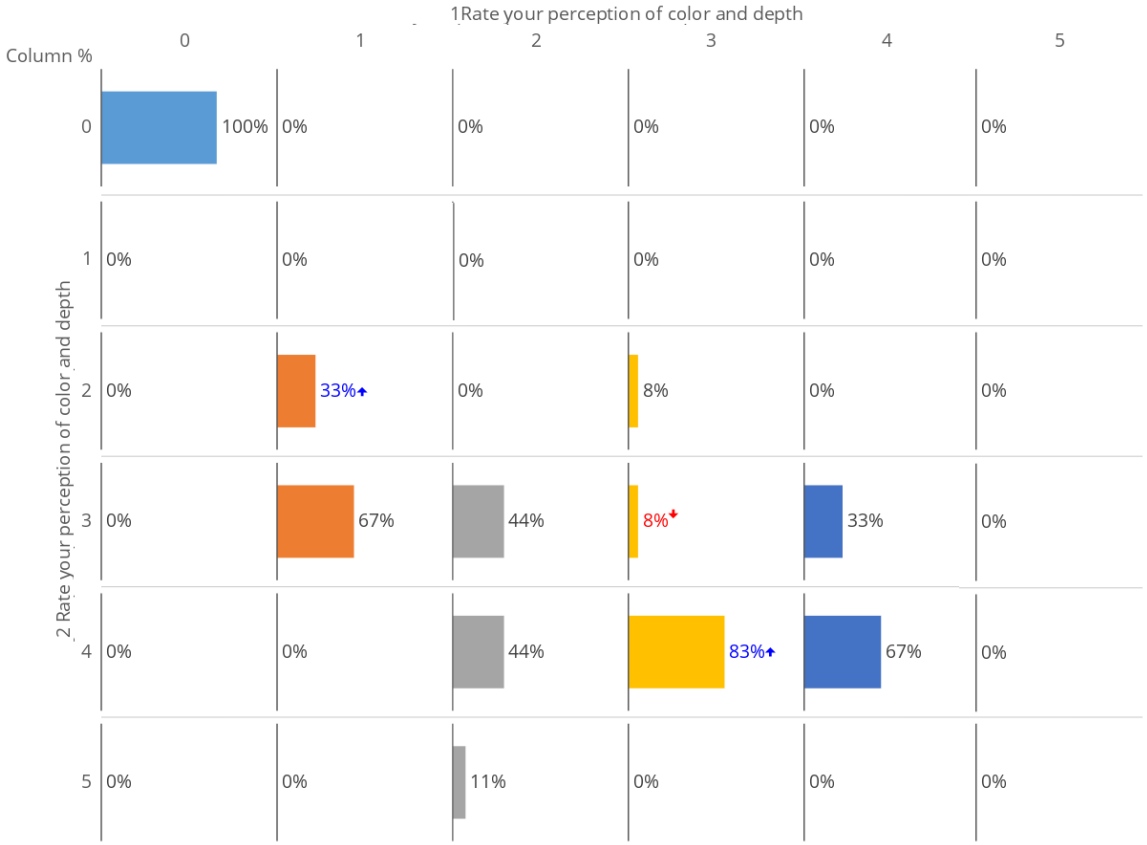
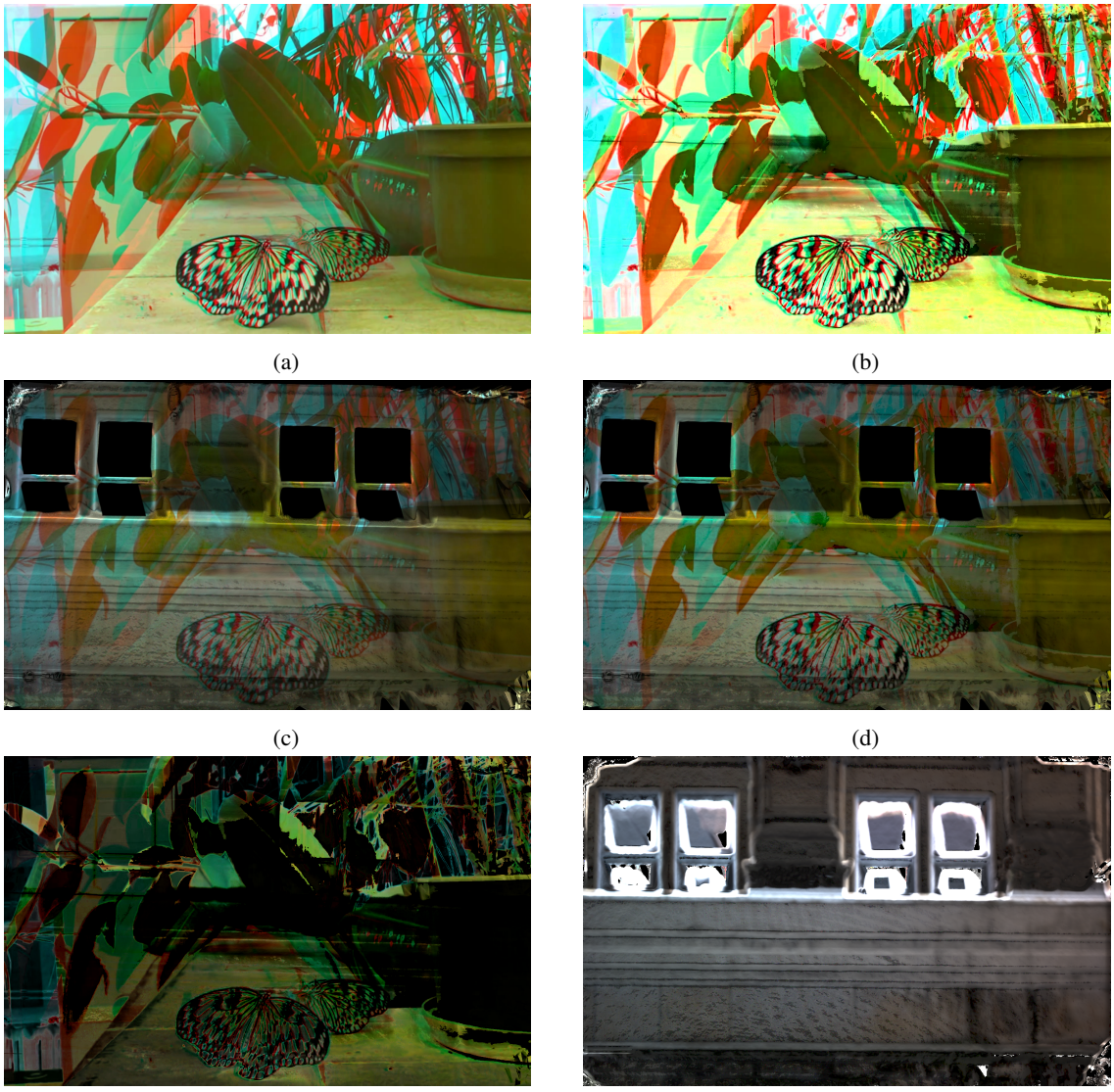


Figure 32: Experiment #2 [Butterfly video]: Participants' questionnaire responses between before (horizontal) and after (vertical) the application of the proposed approach. A likert scale was used ranging from [0,5]. 95% confidence level. Sample size: 37.



(a) (b) (c) (d) (e) (f)

Figure 33: An example frame (from the video) used in Experiment #2. (a) original frame, (b) adjusted frame, (c) projection of original frame on scene, (d) projection of adjusted frame on scene, (e) the projection onto the building's facade; note the excessive ambient lighting present in the scene; distance over 20m, (f) diffuse map of the surface.

SPM EXPERIMENT- Consent Form

Experiment Purpose & Procedure

Dear Participant,

We are conducting a study observing how people perceive anaglyph 3D content projected onto outdoor areas and we would like to invite you to participate.

The purpose of this experiment is to study your perception of color and depth in projection mapping of stereoscopic content.

The experiment consists of 2 parts during which you will be asked to observe various projections of stereoscopic content being projected onto an outdoor scene. After the experiment, you will be asked to complete a questionnaire.

Please note that none of the tasks is a test of your personal intelligence or ability. The objective is to test the effectiveness of our research systems.

If you have suffered or are prone to any form of epilepsy or seizure, we regret that we are unable to accept your participation in the virtual reality part of the study.

Confidentiality

The following data will be recorded: demographic data, responses to questionnaires

All data will be coded so that your anonymity will be protected in any research papers and presentations that result from this work.

Finding out about result

If interested, you can find out the result of the study by contacting the researcher Mr. Behnam Maneshgar, after 01 July 2017. He can be contacted in room EV9.113. His phone number is x7163 and his email address is b_manesh@encs.concordia.ca.

Record of Consent

Your signature below indicates that you have understood the information about the SPM experiment and consent to your participation. The participation is voluntary and you may refuse to answer certain questions on the questionnaire and withdraw from the study at any time with no penalty. This does not waive your legal rights. You should have received a copy of the consent form for your own record. If you have further questions related to this research, please contact the researcher.

Participant ---------- Date ----------

Researcher ---------- Date ----------

Participant ID : _____

Demographic Information

- Gender : Male Female Other: _____
- Age : _____
- Educational background:

  Social Science Engineering Arts Other: _____
- Have you suffered or are prone to any form of epilepsy or seizure? If yes, please talk with one of the members of the SPM team. YES NO

Experience Questions:

- Have you ever used anaglyph 3D glasses? YES NO
- If yes, were you able to perceive 3D? YES NO

==========

Experiment Questions - Butterfly Projection #A

Were you able to perceive 3D? YES NO
Rate your perception of color and depth. ____/ 5
Did you experience nausea, dizziness or eye strain? ____/ 5

Experiment Questions - Butterfly Projection #B

Were you able to perceive 3D? YES NO
Rate your perception of color and depth. ____/ 5
Did you experience nausea, dizziness or eye strain? ____/ 5

==========

Experiment Questions - Roller Coaster Projection #A

Were you able to perceive 3D? YES NO
Rate your perception of color and depth. ____/ 5
Did you experience nausea, dizziness or eye strain? ____/ 5

Experiment Questions - Roller Coaster Projection #B

Were you able to perceive 3D? YES NO
Rate your perception of color and depth. ____/ 5
Did you experience nausea, dizziness or eye strain? ____/ 5

Chapter 5

Conclusion and Future Work

Providing low cost minimally intrusive immersive 3D experiences in outdoor public places will open up interesting opportunities for creative artworks by media artists, commercial advertisements, messaging for masses and public education. Anaglyph 3D is clearly a contending technology for this, provided it enables high-quality 3D experience. Our work reported in this thesis is one step towards improving the quality of 3D stereoscopic content projected onto long range outdoor surfaces such as building facades for the public experience. We presented a complete framework to transform stereoscopic anaglyph 3D content so that the quality of depth and color perception is maintained even after projection onto relatively less accommodating surfaces, i.e., surfaces with complex geometry, texture, color and material properties. We introduced two comprehensive methods for our smart projection mapping. Firstly, we developed a framework which compensated the media with minimum number of samples available. Three fixed cameras and one long range projector have been used. By using the phase shifting calibration method, all the cameras have been calibrated corresponded to each other. Then through structured light scanning techniques we simultaneously recover shape and reflectance properties of the surface using non-linear optimization. By applying one shot compensation using the phong model, we are able to generate compensated media for projection. In the second method, we use two fixed long range cameras, one long range projector, and one roving video camera to capture the surface images from different viewpoints. With this captured data, we use structure from motion to overcome the previously persistent problem of long range calibration. By using the same technique as previous methods we recover shape and reflectance properties of the surface. In a final step, we iteratively adjust the projection content to mitigate depth and color perception problems due to the projection surface. We have successfully tested our framework on a number of long range indoor and outdoor scenarios. For validation of the improvement in the quality of 3D experience provided by the use of our framework, we specifically set up two user study experiments with participation from random people present at the experiment venues.

Building facades often include glass windows, which scatter light in different directions, letting some of the incident light to pass through. Our present system places cameras in front of the building and hence can only capture reflected light. Correspondingly our framework can only model reflectance properties of the projection surface. We would like to extend our system to model more general light scattering behavior. We would like to make various computations in the framework more efficient. One thought which we have is that since the physical nature of the projection surface implies material property coherence, we could consider

clustering of the captured pixels/point cloud into regions, and then use this coherence of material property with a cluster to improve computational efficiency. Lastly. we would also want to consider the use of GPU computing for improving computational speeds.

Bibliography

[ABC11] Alessandro Artusi, Francesco Banterle, and Dmitry Chetverikov. A survey of specularity removal methods. In *Computer Graphics Forum*, volume 30, pages 2208–2230. Wiley Online Library, 2011.

[ABL+16] Yatong An, Tyler Bell, Beiwen Li, Jing Xu, and Song Zhang. Method for large-range structured light system calibration. *Applied Optics*, 55(33):9563–9572, 2016.

[ALLL16] Bilal Ahmed, Jong Hun Lee, Yong Yi Lee, and Kwan H Lee. Mimicking an object using multiple projectors. In *Mixed and Augmented Reality (ISMAR-Adjunct), 2016 IEEE International Symposium on*, pages 61–63. IEEE, 2016.

[AYL+12] Daniel G Aliaga, Yu Hong Yeung, Alvin Law, Behzad Sajadi, and Aditi Majumder. Fast high-resolution appearance editing using superimposed projections. *ACM Transactions on Graphics (TOG)*, 31(2):13, 2012.

[BGJ+15] Panagiotis-Alexandros Bokaris, Michèle Gouiffès, Christian Jacquemin, Jean-Marc Chomaz, and Alain Trémeau. One-frame delay for dynamic photometric compensation in a projector-camera system. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2675–2679. IEEE, 2015.

[BGJC14] Panagiotis-Alexandros Bokaris, Michèle Gouiffès, Christian Jacquemin, and Jean-Marc Chomaz. Photometric compensation to dynamic surfaces in a projector-camera system. In *European Conference on Computer Vision*, pages 283–296. Springer, 2014.

[Bli77] James F Blinn. Models of light reflection for computer synthesized pictures. In *ACM SIGGRAPH Computer Graphics*, volume 11, pages 192–198. ACM, 1977.

[BR09] Maged N. Kamel Boulos and Larry R. Robinson. Web gis in practice vii: stereoscopic 3-d solutions for online maps and virtual globes. *International Journal of Health Geographics*, 8(1):59, 2009.

[BWEN05] Oliver Bimber, Gordon Wetzstein, Andreas Emmerling, and Christian Nitschke. Enabling view-dependent stereoscopic projection in real environments. In *Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 14–23. IEEE Computer Society, 2005.

[BXZ16] Tyler Bell, Jing Xu, and Song Zhang. Method for out-of-focus camera calibration. *Applied optics*, 55(9):2346–2352, 2016.

[CT82] Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (TOG)*, 1(1):7–24, 1982.

[Dev07] Frédéric Devernay. C/c++ minpack. `http://devernay.free.fr/hacks/cminpack/`, 2007.

[DTG+04] Paul Debevec, Chris Tchou, Andrew Gardner, Tim Hawkins, Charis Poullis, Jessi Stumpfel, Andrew Jones, Nathaniel Yun, Per Einarsson, Therese Lundgren, et al. Estimating surface reflectance properties of a complex scene under captured natural illumination. *Conditionally Accepted to ACM Transactions on Graphics*, 19, 2004.

[FHMF09] G Falcao, Natalia Hurtos, J Massich, and D Fofi. Projector-camera calibration toolbox. *Erasumus Mundus Masters in Vision and Robotics*, 2009.

[FP10] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8):1362–1376, 2010.

[FRTT04] Rogerio Feris, Ramesh Raskar, Kar-Han Tan, and Matthew Turk. Specular reflection reduction with multi-flash imaging. In *Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium on*, pages 316–321. IEEE, 2004.

[GHP14] Qing Gu, Kyriakos Herakleous, and Charalambos Poullis. 3dunderworld-sls: An open-source structured-light scanning system for rapid geometry acquisition. *arXiv preprint arXiv:1406.6595*, 2014.

[GPNB04] Michael D Grossberg, Harish Peri, Shree K Nayar, and Peter N Belhumeur. Making one object look like another: Controlling appearance using a projector-camera system. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2004.

[@GR] Binary gray code. URL: `http://demonstrations.wolfram.com/BinaryGrayCode`.

[GZ13] Ricardo R Garcia and Avideh Zakhor. Geometric calibration for a multi-camera-projector system. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 467–474. IEEE, 2013.

[HP14] Kyriakos Herakleous and Charalambos Poullis. 3dunderworld-sls: An open-source structured-light scanning system for rapid geometry acquisition. *arXiv preprint arXiv:1406.6595*, 2014.

[JK10] Joongki Jeong and Min Young Kim. Adaptive imaging system with spatial light modulator for robust shape measurement of partially specular objects. *Optics express*, 18(26):27787–27801, 2010.

[LAS+11] Alvin J Law, Daniel G Aliaga, Behzad Sajadi, Aditi Majumder, and Zygmunt Pizlo. Perceptually based appearance modification for compliant appearance editing. In *Computer Graphics Forum*, volume 30, pages 2288–2300. Wiley Online Library, 2011.

[Lev44] Kenneth Levenberg. A method for the solution of certain non–linear problems in least squares. 1944.

[LFTG97a] Eric P. F. Lafortune, Sing-Choong Foo, Kenneth E. Torrance, and Donald P. Greenberg. Non-linear approximation of reflectance functions. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 117–126, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co. URL: http://dx.doi.org/10.1145/258734.258801, doi:10.1145/258734.258801.

[LFTG97b] Eric PF Lafortune, Sing-Choong Foo, Kenneth E Torrance, and Donald P Greenberg. Non-linear approximation of reflectance functions. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 117–126. ACM Press/Addison-Wesley Publishing Co., 1997.

[LLK+02] Stephen Lin, Yuanzhen Li, Sing Bing Kang, Xin Tong, and Heung-Yeung Shum. Diffuse-specular separation and depth recovery from image sequences. In *European conference on computer vision*, pages 210–224. Springer, 2002.

[mar] mars.nasa.gov. Mars 3d Images. URL: http://mars.jpl.nasa.gov/mars3d/.

[Mata] Mathwork. Distortion in camera calibration. Accessed: 2017-04-23. URL: https://www.mathworks.com/help/vision/ug/camera-calibration.html.

[Matb] Mathwork. Pinhole camear model. URL: https://www.mathworks.com/help/vision/ug/camera-calibration.html.

[MPBM03] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3):759–769, July 2003.

[MSUA12] Rosana Montes Soldado and Carlos Ureña Almagro. An overview of brdf models. 2012.

[MT12] Daniel Moreno and Gabriel Taubin. Simple, accurate, and robust projector-camera calibration. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 464–471. IEEE, 2012.

[ON95] Michael Oren and Shree K Nayar. Generalization of the lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, 1995.

[Pho75] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[PJH16] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016.

[PMS10] Tomislav Pribanić, Saša Mrvoš, and Joaquim Salvi. Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning. *Image and Vision Computing*, 28(8):1255–1266, 2010.

[Pou13] Charalambos Poullis. A framework for automatic modeling from point cloud data. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2563–2575, 2013.

[@re17] Nature of reflection, 2017. Accessed: 2017-05-10. URL: `http://sciencevault.net/11hscphys/82worldcommunicates/824%20Reflection.htm`.

[RGVP+14] Gonzalo M. Rojas, Marcelo Gálvez, Natan Vega Potler, R. Cameron Craddock, Daniel S. Margulies, F. Xavier Castellanos, and Michael P. Milham. Stereoscopic three-dimensional visualization applied to multimodal brain images: clinical applications and a functional connectivity atlas. *Frontiers in Neuroscience*, 8:328, 2014.

[RRL+14] Brett Ridel, Patrick Reuter, Jérémy Laviole, Nicolas Mellado, Nadine Couture, and Xavier Granier. The revealing flashlight: Interactive spatial augmented reality for detail exploration of cultural heritage artifacts. *Journal on Computing and Cultural Heritage (JOCCH)*, 7(2):6, 2014.

[RWC+98] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM, 1998.

[RWF98] Ramesh Raskar, Greg Welch, and Henry Fuchs. Spatially augmented reality. In *First IEEE Workshop on Augmented Reality (IWAR'98)*, pages 11–20, 1998.

[SCT+15] Christian Siegl, Matteo Colaianni, Lucas Thies, Justus Thies, Michael Zollhöfer, Shahram Izadi, Marc Stamminger, and Frank Bauer. Real-time pixel luminance optimization for dynamic multi-projection mapping. *ACM Transactions on Graphics (TOG)*, 34(6):237, 2015.

[Sha85] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

[SMK05] Steven M Seitz, Yasuyuki Matsushita, and Kiriakos N Kutulakos. A theory of inverse light transport. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1440–1447. IEEE, 2005.

[SMP05] Tomáš Svoboda, Daniel Martinec, and Tomáš Pajdla. A convenient multicamera self-calibration for virtual environments. *Presence*, 14(4):407–422, 2005.

[ste] Anaglyph image. `https://goo.gl/9vSgkP`. Accessed: 2016-11-25.

[TM98] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.

[@tr] Head motion tracking in 3d space for drivers. URL: `http://archimede.bibl.ulaval.ca/archimede/fichiers/25229/ch06.html`.

[TS67] Kenneth E Torrance and Ephraim M Sparrow. Theory for off-specular reflection from roughened surfaces. *JOSA*, 57(9):1105–1112, 1967.

[Tsa87] Roger Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987.

[Wika] Wikipeida. Anaglyph 3d. Accessed: 2017-05-17. URL: https://en.wikipedia.org/wiki/Anaglyph_3D.

[Wikb] Wikipeida. Phong reflection model. Accessed: 2017-05-10. URL: https://en.wikipedia.org/wiki/Phong_reflection_model.

[Wik17] Wikipedia. Anaglyph_3d, 2017. URL: https://en.wikipedia.org/wiki/Anaglyph_3D.

[You] Youtube. 3d jumpscare on bridge 3d anaglyph 3d video hd 1080p. Accessed: 2017-05-01. URL: https://youtu.be/rM0FMxzFO4Y.

[Zha00] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[ZXT+16] Yi Zhou, Shuangjiu Xiao, Ning Tang, Zhiyong Wei, and Xu Chen. Pmomo: Projection mapping on movable 3d object. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 781–790. ACM, 2016.