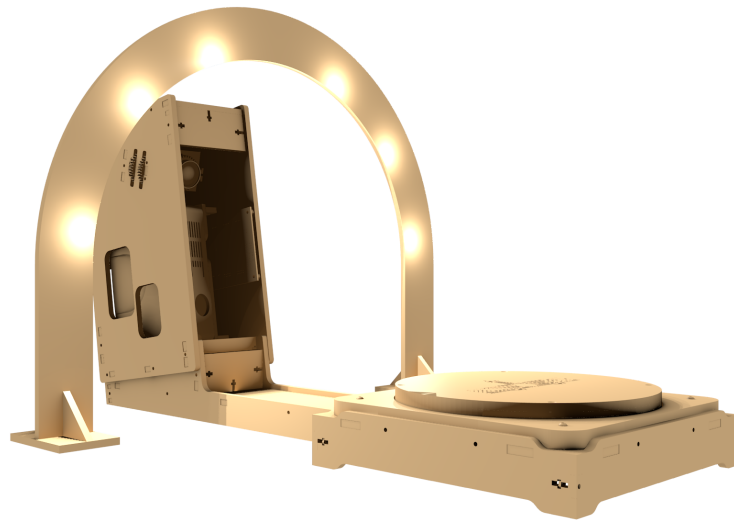# Combining Positions and Photometric Stereo Normals in Linear Time with Belief Propagation

Ruben Dulek

16th July 2015

Abstract

To increase the accuracy of 3D scanners, photometric stereo scanning is combined with structured light with a new method that scales well with high-resolution cameras. This new method defines a Markov random field initially with the low-frequency data of the structured light scan and applies the data of the photometric scan locally. In an experiment, the resulting scans are found to be more accurate than the original scan. Also, a new method is proposed to calibrate the positions of lights for photometric scanning by using a rotating platform. The technique finds the reflections of the lights in reflective spheres and traces back the positions of the lights. It then rotates the platform and repeats the experiment. While the rotation is found to improve the result, the sequence in total is found to perform badly.

1 Introduction

The rise of 3D printing has provided some especially useful tools for those who produce prototypes, fabricate plastic items and artistic modellers. However, not everyone is endowed with the skills to use complex 3D design software. The field of 3D scanning has provided some relief, but 3D scans are often too inaccurate to capture the precise dimensions of an item or to capture the surface structure of something the designer would like to incorporate into his design. Yet, 3D scanning has a wide range of applications, such as reverse engineering [VMC97], modelling and prototyping, preserving culture by scanning statues, tablets and paintings [KKOF04], special effects in films and quality control [Bos10]. This thesis is concerned with increasing the accuracy of 3D scanners.

Photometric stereo is a technique that analyses the reflection of light onto a surface to reconstruct the 3D geometry of that surface. It provides more surface detail where other techniques are lacking [Td91], but is vulnerable to the offset-error problem. This causes low-frequency noise in the scan. Nehab et al. popularised the idea to combine the high-frequency data of photometric stereo, which is often quite accurate, with the low-frequency data of another range imaging technique [NRDR05]. The combination of the two has often been shown to be more accurate than both scans individually [ATD$^+$08]. This thesis introduces a new technique to combine the two range imaging techniques.

1.1 Problem Statement

The research in this thesis is twofold. Firstly, a new technique is described to calibrate for photometric stereo using a rotating platform. The rotating platform allows for multiple measurements to be made which are more or less independent of each other. This should reduce the noise in the system. Secondly, a new technique is introduced to combine photometric stereo with structured light. This technique involves a grid of random variables, all dependent on each other. The Maximum A-Posteriori (MAP)-estimate of this Markov random field is found using a series of Kalman filters.

Both techniques are subject to some restrictions in order to make it usable and to prepare it for future advances in hardware. The result is a working 3D scanner that is intended to be usable by persons that are not technically-inclined. This poses the following restrictions:

- The scanning and calibration process must be automatic, involving as few interventions of the user as possible.

- The work flow of the scanner must remain simple. This makes the scanner accessible to the greater public, which in turn allows research to advance faster.

- The scanning and calibration algorithms must scale well with new hardware. Since the product is built as a kit, a technical user may choose to replace parts of the 3D scanner if he so pleases, such as the camera or the projector. The expectation is that the resolution of cameras and the quality of lenses will greatly increase in the future. The scanning and calibration algorithms must therefore scale linearly in their time complexity with the number of pixels and be easily paralellisable to process the data with multiple processors.

- The result of the fusion must be more accurate than the original scans.

The first three requirements are applied to both the fusion and the calibration techniques. The last is only applicable to the fusion technique.

1.2   Structure of This Thesis

The first section after this introduction is on related academic work. It will introduce some related topics of research upon which this thesis is based. It will start by introducing various range imaging techniques and will gradually converge on photometric stereo and how it is fused with other scans.

After that, the scanning process is described in three sections. Section 3 will describe the calibration process of the scanner. It will extensively detail the hardware components that need to be calibrated and give a short description of the calibration process for each, as well as the assumptions that the calibration does not account for. The last hardware component is the arc of lights. The calibration sequence for the lights is a new method, so it will be explained in much greater detail in Section 3.5.

In Section 4, the photometric stereo method is described. Though some of its initial estimates are obtained from the structured light scan, this scanning technique is not new. Still it is required to have a complete understanding of how the 3D scanner works and to show it satisfies the linearity constraint and the parallelisability constraint. Therefore, the algorithm is explained in some detail.

Section 5 describes the algorithm that is used to merge the photometric stereo data with the structured light data. The Markov random field is described and how its random variables are updated. Then lastly a belief propagation algorithm is described that converges to the final result.

Having described the process of the 3D scanner, Section 6 introduces three experiments to validate the new techniques described in Sections 3 and 5. Two of these experiments will test the accuracy of the calibration technique from Section 3.5. The last will test the accuracy of the scanning and the fusion technique described in Section 5. The results of these tests are then explored in Section 7.

Finally, Section 8 will conclude this thesis with a summary of the scanning process and the experimental results. This section will also describe research that could be conducted to expand upon this thesis.

2   Related Work

This section will lay out some of the research upon which this work is based. The canonical research on photometric stereo is displayed first, after which this section will gradually scope towards the directly relevant works preceding this thesis.

2.1   Photometric Stereo

There are many strategies to measure depth, collected under the denominator of range imaging. The most commonly used techniques for a 3D scanner are:

- *Stereo triangulation*, matching parts of images from different viewpoints and measuring their disparity [MP79].

- *Time-of-flight*, measuring how long light or sound takes to reflect back from the surface [LJ77].

- *Structured light*, projecting patterns of light on the surface and measuring how these patterns are deformed [AAT81].

- *Depth from motion*, measuring flow in a video to estimate the parallax in the scene [Ull79].

- *Shape from shading*, measuring the surface orientation by measuring how much light is reflected by it.

This work concerns the last technique, *shape from shading*, or more specifically, *photometric stereo*. Shape from shading aims to measure depth by looking at how light reflects on the object. In photometric stereo, lights are cast upon the surface from different angles. The direction of the surface is then estimated by measuring how much of that light is reflected into the camera.

Photometric stereo was first shown to be viable by Woodham in 1980 [Woo80]. Woodham showed how to obtain the surface normal vectors of a completely differentiable surface if the surface was perfectly diffuse, illuminated by a directional light source with a known direction and had uniform albedo. He then showed how these surface normals can be integrated to obtain the three-dimensional geometry of the surface. Research since then has split into two major strands: addressing the assumptions on the surface and lights and improving the process to transform surface normals into surface geometry.

The research concerning Woodham's assumptions concentrated mostly on generalising photometric stereo to other types of surfaces. As such, researchers have been able to scan specular surfaces [Ike81], surfaces with patches of different reflectance [GCHS10] and the more general case of textured surfaces [BP03]. This has progressed to increasingly complex reflection models, including the Torrance-Sparrow model [HB88] and the Cook-Torrance model [YSL10], even up to anisotropic surfaces [HLHZ08]. Some progress is made to account for

interreflections as well [NIK91] [LHY11], but this is still restricted to many assumptions. Photometric stereo in the presence of translucent or light-emitting materials or materials allowing subsurface scattering is not yet pursued at the time of this writing. This branch of research is out of the scope of this thesis; a simple Lambertian model is assumed, though allowing variations in diffuse albedo. Deviations from this model are filtered out up to a reasonable extent, but the reflection model itself is kept simple.

The other major branch of research concerns itself with using the obtained normal vectors to create 3D geometry. The first work in this topic appeared as a technique for shape from shading and worked by integrating pixels in predetermined straight lines over the surface [Rin66] [Hor70]. This old technique has been improved upon since in several ways [ACR05] [FH06], mostly to improve reliability in cases where the surface is not integrable. However, this technique has been shown to be relatively sensitive to noisy normal vectors and depends on the choice of integration paths [SSL+11]. A more popular technique is to take a fast Fourier transform of the normal vectors, integrate the data in the frequency domain and transform them back [FC88] [Kov05]. This technique is very robust to noise but unsuitable to handle non-integrable surfaces. Even more popular in research literature are Poisson-like methods such that solve a sparse linear system of constraints using least squares [Hor90] or n^{th}-order polynomials [ARC06]. Poisson-like solvers commonly produce very good results, but are expensive to compute, having exponential time complexity in the number of pixels, and they are difficult to implement. The method proposed in this document integrates the normal vectors using information from another scan. The technique used for this belongs to yet another class, belief propagation, which will be explained in more detail in Section 5.

2.2 Combining Photometric Stereo with Other Range Imaging

Combining the orientations from photometric stereo with the positions of other range imaging techniques has known a paradigm shift over time. Initially, photometric stereo was used to fill holes of information in the other scan, because of its dense grid of normal vectors. Here they assume that the other scan provides a sparse set of positions on the surface. They interpolate them using the photometric stereo data, for instance by solving a set of nodal equations to minimise an error function [Ter88] or least squares to minimise the quadratic error [BZK85]. This approach was limited due to the noise inherent in both techniques. Newer approaches therefore seek to combine the best properties of photometric stereo with the best properties of the other data. Most prominent in this paradigm shift is Nehab et al. [NRDR05], who explicitly combine the low frequency data of range imaging with the high frequency data of photometric stereo.

By far the most common approach to combine surface gradients with surface geometry is to minimise some error function, that combines the deviation from some model, with a Poisson solver [FL94] [ZYY+12] [HCG14]. This approach defines a cost function based on the depth estimates from the range images and the differential depth estimates from the photometric scan. The influence of each of these can be weighted using just pre-set constants [WLDW11] or adaptively based on other data, such as the steepness of the surface [ZYY+12].

To minimise this cost function, the cost is reformulated as a Poisson equation and then solved using a least squares fit. Like with ordinary normal vector integration, these methods are expensive to compute but produce good results. This has made them the main focus of research in this field. Using adaptive weighting of the influence of the normal vectors can bring out the strengths of both the range scan and the photometric stereo scan. This can also be done with preliminary filtering [NRDR05].

Competing techniques to the Poisson-like method are few. Perhaps the simplest is to linearly interpolate between the two surfaces, such as in [CKC03]. Here they assign weights to a surface obtained by a photometric scan based on the angle of the surface and assign complementary weights to a surface obtained from a range scan. The final surface is a weighted average between the two. This method is fast to compute, but comparatively inaccurate.

A bit more advanced are the belief propagation methods. Here the information is combined in a Markov random field, of which the MAP estimate is computed [ATD$^+$08]. This gives a most likely explanation of the observed measurements (assuming a good model). In fact, it can be used to combine any type of measurements, not just normal vectors [ZWYD08]. The largest problem with these solutions is that finding the MAP-estimate can be difficult for such large Markov random fields. That is why this thesis makes the assumption that the noise in all measurements is normally distributed, as can be estimated with a Gaussian function. *Gaussian Markov random fields* have come under much attention in image processing recently due to their applications on spacial models such as images, so they are well understood. They possess mathematical qualities that make it feasible to compute the MAP-estimate in polynomial time using their *precision matrix* [RH05]. This thesis approximates the MAP-estimate with one-dimensional Kalman filters, which is even less time-complex.

3 Calibration

In this section, the calibration sequence of the scanner is explained. The scanner has four hardware components, which need to be calibrated in a specific order. This section maintains that order. First, an overview of the set-up is given. Then, in order, the camera calibration is explained, the projector calibration is explained, the platform calibration is explained and finally the light calibration is explained. The calibration of the lights uses a novel technique that makes use of the rotating platform to obtain a more accurate calibration. That technique is one of the scientific contributions of this thesis. It is described in Section 3.5.

3.1 Hardware Overview

In order to get an idea of how the various are calibrated in relation to each other, this section gives an overview of the hardware components and their relative position and orientation. The 3D scanner used for the experiments of this thesis consists of a camera, a projector, a rotating platform and six LEDs. The set-up of the scanner hardware is depicted in schematic form in Figure 1 and a photo is shown in Figure 2.

The camera and the projector are positioned in a wooden housing and aimed at the platform. Six LEDs are positioned on a wide arc in front of the camera.
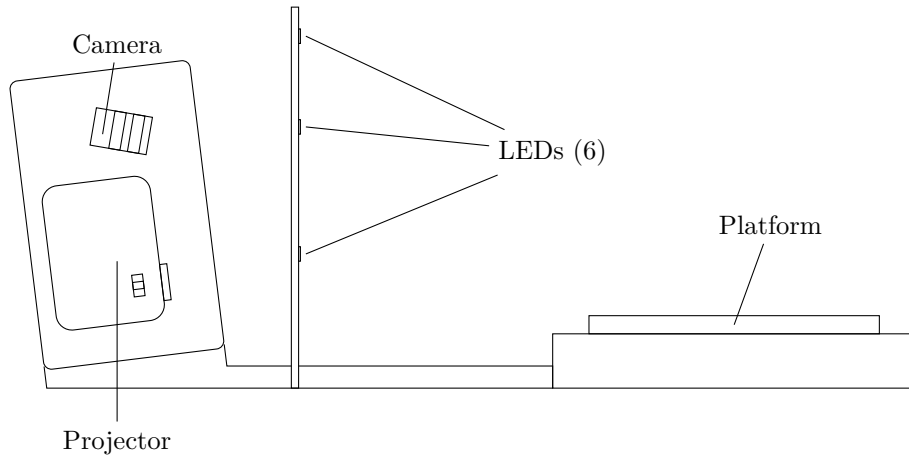
Figure 1: The set-up of the 3D scanner.

The arc is wide enough to be out of the field of view of both the camera and the projector. Alongside the scanner, two calibration items are provided. The first is a *calibration board*, which is a flat wooden board, to calibrate the camera, projector and platform. The second is a *calibration statue*, which is a plastic structure with three spheres on top, to calibrate the lights. The calibration board, calibration statue and all items to be scanned will be placed on the platform, where they are in view of the camera and projector and where they can be lit by the LEDs.

To make the software easy to use for the layman, the user is guided through all steps of the calibration in sequence with clear instructions on what to do. The interaction with the user is kept to a minimum, but the user must place the calibration board and statue onto the platform when the software asks for it. If needed, the user is also required to adjust the focus of the camera and the projector and the aperture of the camera.

3.2 Camera

To capture any data for both structured light and photometric stereo, a camera is required.

3.2.1 Hardware

The camera used in these experiments is an *E-Con See3CAM_10CUG* monochrome camera by *E-Con Systems*. It has a resolution of 1280 by 960 pixels, but supports a lower frame rate at that resolution. The camera produces its maximum frame rate of 60 frames per second at a resolution of 1280 by 720 pixels, so that is the resolution that was used in these experiments. It is equipped with a Fujinon DF6HA-1B lens, with a diameter of 12.7mm and an angle of view of 56°09' horizontally and 43°36' vertically.

The camera was chosen due to several constraints. Most importantly, the camera needs to be cheap, but still give a sharp image with little noise. A global shutter is desired to prevent tearing, which is disastrous for the structured light
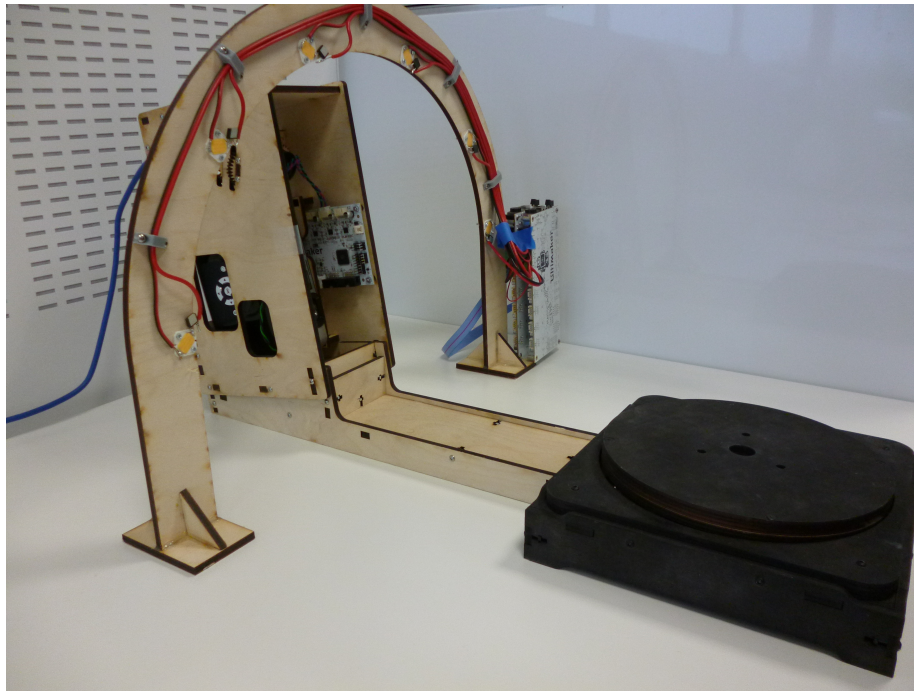
Figure 2: A photo of the 3D scanner.

scan and calibration. The camera should not have a Bayer filter, since that effectively increases the pixel size on the sensor for the light intensity data. Colour can be measured by projecting light of different colours from the projector but is not used in the surface reconstruction. No housing is required either, since the camera will be placed in a wooden box that already provides sufficient protection. Meeting these requirements, this camera strikes a good balance between price and quality, though it is a bit on the cheap side at approximately €100.

The camera features two rings on its lens. One ring adjusts the camera's focus distance, which should be centred on the platform. The other adjusts the camera's aperture. This should be adjusted such that the images captured by the camera use the maximum possible range of intensity values without clipping the whites. The user is provided with live camera footage and asked to adjust the focus and aperture on the device. The focus and aperture rings can be accessed from the outside of the wooden case via a few gears.

3.2.2 Calibration

To calibrate the camera, the intrinsic parameters of the camera have to be found. To this end, the user must place a calibration board on top of the platform. The calibration board is a flat wooden board. It is held upright by a plastic piece, which also features pins on the underside that fix it to the platform. The board is printed with a matte reflective paint with 44 dark circles in a specific pattern. This pattern is designed to be asymmetrical with respect to point symmetry, to eliminate any ambiguous orientations of the platform. The board is depicted in Figure 3.

Figure 3: The calibration board used for the calibration of the camera and projector and a euro for scale.

The circles are recognised by the software. The centre of each circle is found in the camera image. The software knows the actual relative positions between the circles. It can then project the measured positions onto its digital model, resulting in a transformation matrix. This transformation matrix contains all intrinsic parameters for the camera. The implementation of this technique is provided by *OpenCV*, a library of computer vision algorithms. The camera is assumed to be positioned at the coordinate origin and facing in direction of positive z. Therefore, no extrinsic parameters need to be acquired.

3.2.3 Assumptions

The calibration of the camera relies on a number of assumptions that are made to simplify the process. Many of these assumptions could be detrimental to the photometric stereo method [KKS98] as well as to structured light. As future research, the calibration sequence could be extended to prevent having to make some of these assumptions.

- There is no pre-knee circuit in the camera. In reality, the images in modern cameras (including the E-Con See3CAM_10CUG) are processed by a so-called pre-knee circuit which compresses the higher end of the signal. This allows for a greater range of intensities to be detected, but makes the signal no longer linear.

- No gamma correction is made for the camera. In reality, both the camera itself and the video driver of the computer may apply some gamma

correction for display purposes, which uncouples the image intensity from the linear relation with the actual intensity. This requires gamma re-correction to undo, but some data is inevitably lost.

- The black level of the camera is zero. This means that the camera would give a signal of 0 if the scene is completely dark. This is often not the case.

- The white of the camera is balanced. This means that a perfectly un-saturated surface should be represented by an equal signal in each colour channel. For a monochrome camera such as the one used in these experiments, this means that an equally intense light should give an equal signal regardless of the hue or saturation of the light. This is almost never the case in reality and some cameras explicitly respond better to green light to reflect the workings of the human eye.

- There is no blooming in the camera. Blooming is caused when a very bright light overloads the sensors in the camera such that they bleed to neighbouring sensors (for neighbouring pixels). Cameras with anti-blooming drains are expensive and not used in these experiments.

- The camera is assumed to fit the Brown-Conrady model with 6 radial distortion coefficients and 2 tangential distortion coefficients. More complex lens distortions are normally very small, but they will cause distortion in the resulting scan.

- Any noise in the camera image is normally distributed. Since the noise has many causes, this is not a safe assumption. However, the noise in a digital light sensor is mostly caused by the random distribution of photons and the variation in the amount of electrons each photon generates. This noise can safely be assumed to be normally distributed.

3.3 Projector

To project the structured patterns of light for the structured light technique, a projector is required. The projector effectively functions as a point light too, which could be useful for the photometric stereo technique. However this breaks the assumption that all lights have equal intensity. In the future, the projector (having a known position) could be used as a point light to increase the accuracy, but this requires that the intensity of lights is calibrated for too.

3.3.1 Hardware

The projector used in these experiments is a *Movying MOV298A* LED projector with 600 Lumens of power, a contrast ratio of 1000:1 and an effective resolution of 1280 by 720 pixels. It theoretically supports a frame rate of 60 frames per second, but this frame rate is difficult to implement in the firmware for the scanner.

The constraints on the projector are primarily that it must provide enough light for the camera to differentiate between the different brightness values in the projected image. The Gray pattern requires the camera to differentiate only between light and dark, but the phase pattern requires the camera to

differentiate accurately which brightness value is being projected on a piece of surface by the projector. Other constraints are that the projections of the projector's pixels should be reasonably dense-packed on the item's surface, since that is a limitation of the scan's resolution in the Gray pattern. Lastly, the projector must be cheap to reduce the cost of the final product. The bulk rate for this projector comes down to €285 per projector. This makes it good value.

The projector features one slider to adjust its focus distance. This focus distance must be centred around the platform's centre, like the camera.

3.3.2 Calibration

To calibrate the projector, both the intrinsic and the extrinsic parameters of the projector have to be found. The technique used for this is similar to the scanning itself. The projector is made to project a series of patterns onto the calibration board. The patterns are binary, consisting only of completely black and completely white pixels. Therefore each pattern subdivides the camera image in two. After 20 patterns, the scanner can uniquely identify each projector pixel in the camera image. This way, it can find the albedo of the surface under each pixel of the projector. The scanner reconstructs what the camera would see if it were a camera.

Treating the projector as if it were a camera as well, the intrinsics and extrinsics of the projector are calibrated. This functions in the same way as the calibration of the camera: The centre of each circle is determined. These centres are then fit to the known model and the projection matrix contains all required parameters.

3.3.3 Assumptions

The calibration of the projector relies on the similar assumptions as the calibration of the camera. For a more detailed description, please see Section 3.2.3. The calibration sequence could be extended to prevent having to make some of these assumptions.

- The output light intensity of the projector is linear to the input signal. This projector is a DLP projector, causing it to achieve excellent linearity in greyscale.

- The black level of the projector is zero. This means that the projector does not project any light when given a signal of 0. This is rarely true. In fact, this is found to be the most significant source of noise in the calibration algorithm for the lights, when the projector is made to project black in an effort to turn it off.

- The white of the projector is balanced. If this is not the case, then surfaces of different hue would reflect different amounts of light, resulting in noise for textured items.

- The projector is assumed to fit the Brown-Conrady model with 6 radial distortion coefficients and 2 tangential distortion coefficients, like the camera.

3.4 Platform

A rotating platform offers numerous advantages for 3D scanners, many of which have not yet been explored in scientific literature. Most importantly, it offers the ability to view the item from different sides without any need for human interference or multiple cameras and projectors. It can be calibrated to obtain the axis of rotation, to which items can be fixed. Lastly, it allows for making multiple measurements which are more or less independent of each other, increasing the accuracy of these measurements. This last advantage is explored in the calibration technique in Section 3.5.

3.4.1 Hardware

The platform a wooden disc, made by a laser cutter out of plywood. The disc is encased by a wooden box, which rests stable on any flat surface. The disc has a diameter of 0.20m. This diameter was chosen to provide users with a guide of what items can be replicated with an Ultimaker 3D printer, after scanning. The volume that can be scanned using this set-up approaches the maximum print volume of an Ultimaker 2. The centre of the platform is set at a distance of 0.5m from the camera.

The platform is powered by a stepper motor. The motor is a Nema 14HR08-0654S with a maximum radial force of 12N·cm. The force is transported via reducer gears to the platform. It has a step angle of 0.900°, which is reduced to 0.00398°by the gears.

The platform is painted black. This minimises the amount of light reflected from its surface. The surface of the platform is slightly lower than the lens of the projector. This allows the bottom of the item to be properly scanned. However, the projector's light that does hit the surface of the platform will get stretched such that the platform is barely illuminated by the projector. This prevents the platform from showing up in the resulting scan.

3.4.2 Calibration

To calibrate the platform, the rotational axis of the platform has to be found. This axis is again found by looking at the calibration board with the camera. With a calibrated camera and known dimensions of the calibration board, the precise 3-dimensional location of the calibration board can be deduced. Through the positions of the circles on the platform, a plane is fit. This plane coincides with the surface of the platform.

The platform is rotated a number of times. This produces a set of planes. If the calibration board was rotated around a line on its surface, these planes should all intersect in one line. This line is the axis of rotation. However, due to noise inherent in measuring with a camera and using a wooden platform, the planes rarely intersect in one line. To still obtain an axis, each pair of consecutive planes is intersected, producing an axis. The axis is expressed as an origin and a direction. For every consecutive pair, the origins and directions are averaged to produce a single axis, which is the result of the calibration.

3.4.3 Assumptions

The platform is also subject to a few assumptions in order for the scanner to function. The calibration sequence could be made to prevent having to make these assumptions in future research.

- The platform's rotational velocity is *approximately* known and can be controlled. All techniques used by the scanner described here allow for some error in the rotational velocity, with varying accuracy constraints, but the platform's rotational velocity and resulting rotation angle is assumed to be reasonably precise. Because the motor is a stepper motor, this assumption is reasonable to make. The use of markers on the side of the platform could further increase accuracy using computer vision.

- The calibration board is fixed onto the platform and made to rotate around its visible surface. The calibration board is fixed with magnetic pegs, but inevitably vibrates slightly with the platform.

3.5 Lights

The lights are required for the photometric stereo method. They illuminate the scene from different angles, after which the camera can measure the amount of light that gets reflected. For the calibration of the lights a novel technique was used. To calibrate the lights means to find the 3-dimensional positions of the lights. This subsection details the process of finding these light positions. In brief, the position of each light is found by positioning three reflective spheres in front of the camera on the rotation platform. Each light is turned on in sequence and the three reflections are seen in the camera. A ray is then modelled from the camera through the centre of each reflection and reflected off the spheres. The light position is then the intersection of these three rays.

3.5.1 Hardware

The lights in this set-up are model *Sharp GW5BTF15L0B*. These are high-power LED emitters, consisting of 48 individual LEDs each, mounted in 4 rows to form a rectangle of 10.00mm by 12.96mm. The LED emitters are capable of a total radiance of 360 Lumens. Their light has a colour temperature of 3045K, making them slightly yellow to the human eye. There are 6 of these lights in the set-up, which is twice of the minimum requirement for the photometric stereo method that is used here.

The lights are suspended on a wooden arc. This arc is sliced from a plywood board with a laser cutter. The arc is placed on small footings that are cut from the same board to keep it upright. Since the Lambertian model for photometric stereo requires at least 3 light sources and there are 6 available, it is theoretically possible to place the lights such that the Lambertian model is always invertible (see Section 4) [Td91]. In practice though this is neither achievable nor required. The chance of a model occurring that is not invertible is negligible and if it occurs, a pseudo-inversion should find a nearby model that is invertible. The lights are therefore more or less equally spaced around the arc, to maximise the difference between lighting conditions and thereby increasing the signal-to-noise ratio of the normal vectors.

The wooden arc is also equipped with plastic clips modelled on a computer and fabricated with a 3D printer to hold the cables. Attached to the side of the arc are two microcontroller boards to provide sufficient power to all the lights and allow control of these lights via a USB cable.

3.5.2 The Calibration Statue

Figure 4: The calibration statue with three spheres and a euro for scale.

To facilitate the correct placement of the three spheres, a *calibration statue* is fabricated. It is modelled on a computer and fabricated by a 3D printer. The finished product can be seen in Figure 4. The dimensions of this calibration statue needs to be known by the software, so it is critical that they are accurate. To verify that the printer was accurate, the dimensions of the printed model were measured with a vernier caliper. They are found to be within 0.04mm of the digital model, of which some error can be attributed to the use of the caliper.

Placed upon the calibration statue are three reflective spheres. The spheres used in these experiments are Arasmith billiard balls with a diameter of 52.4mm. The spheres are black and glossy, to reduce the influence of environmental light but maximise the influence of the scanner's light sources in the camera images. Billiard balls are used to make certain that the balls are as close to being spherical as possible while keeping the cost low.

3.5.3 Capturing

After the calibration statue is placed on the platform, the calibration of the lights is automatic. The first step in this process is to capture data by activating the

lights and taking pictures of the calibration statue with the camera.

The lights are activated in sequence while the camera takes one picture of the scene with each light illuminating it. All lights are then deactivated and the scene is captured with the lights turned off. The image of the scene without lights is then subtracted from the image of the scene with lights. This effectively measures the influence of the lights on the scene. This image will be noisy, due to temporal interference patterns with fluorescent lamps in the room, clouds passing in front of the sun, people moving in the room or other changes in the environmental light. Therefore this process is repeated a number of times. The results are averaged and samples which deviate more than one standard deviation from the mean are discarded to remove such noise.

3.5.4 Approximating Sphere Positions

Calibrating the light direction with one reflective sphere is common in photometric stereo [SW13]. However, this makes the assumption that the light is a directional light rather than a point light and the assumption that the position of the sphere is known. Since the lights are fairly close to the scene in this set-up, a directional light is an inaccurate model. To find the position of a point light though, multiple spheres are required and their position needs to be known. This is the second step in the calibration process.

After the capturing is completed, the three brightest reflections are found in each light's image, to match the expected reflections in each of the spheres. This produces 3 groups of 6 two-dimensional coordinates each. To find an initial estimate of the sphere's position, the smallest enclosing circle is taken for each of these clusters, as depicted in Figure 5. The centres of these circles are taken as initial guess for the centres of the spheres in the images. The smallest enclosing circle is chosen because it minimises the chance for a ray to miss the sphere if it is cast through each of the light positions. When a ray misses the sphere, no reflection can be constructed and the calibration fails. Constructing the smallest enclosing circle maximises the room for error around all of the light reflections.

The rotational axis of the platform is known from calibrating the platform. The spheres are assumed to be positioned on the calibration statue which is positioned in the centre of the platform. Since the shape of the calibration statue is known, the positions of the spheres are constrained. Let the (known) distance of the sphere from the rotational axis be R. The two possible positions for each sphere are determined by the intersection of a line and a cylinder. The line goes through the camera and through the initial guess of the sphere's position. It represents all possible positions of the sphere according to the measurements. The cylinder is a hollow cylinder with the same axis as the rotating platform, infinite height and radius R. It represents the space of possible positions due to the shape of the calibration statue.

The line and the cylinder intersect at most in two positions. To find these, first project the line and the cylinder onto the plane orthogonal to the rotational axis. This produces a two-dimensional line and circle. In the formula below, $\vec{r}$ is a point on the rotational axis and $\vec{s}$ the normalised direction of the axis. $\vec{l}$ is a point on the line and $\vec{m}$ the normalised direction of the line. Name $\vec{q}$ the centre of the projected cylinder, $\vec{o}$ a point on the projected line and $\vec{p}$ the direction of the projected line. See Figure 6b for an illustration. Then $\vec{q}$, $\vec{o}$ and $\vec{p}$ can be
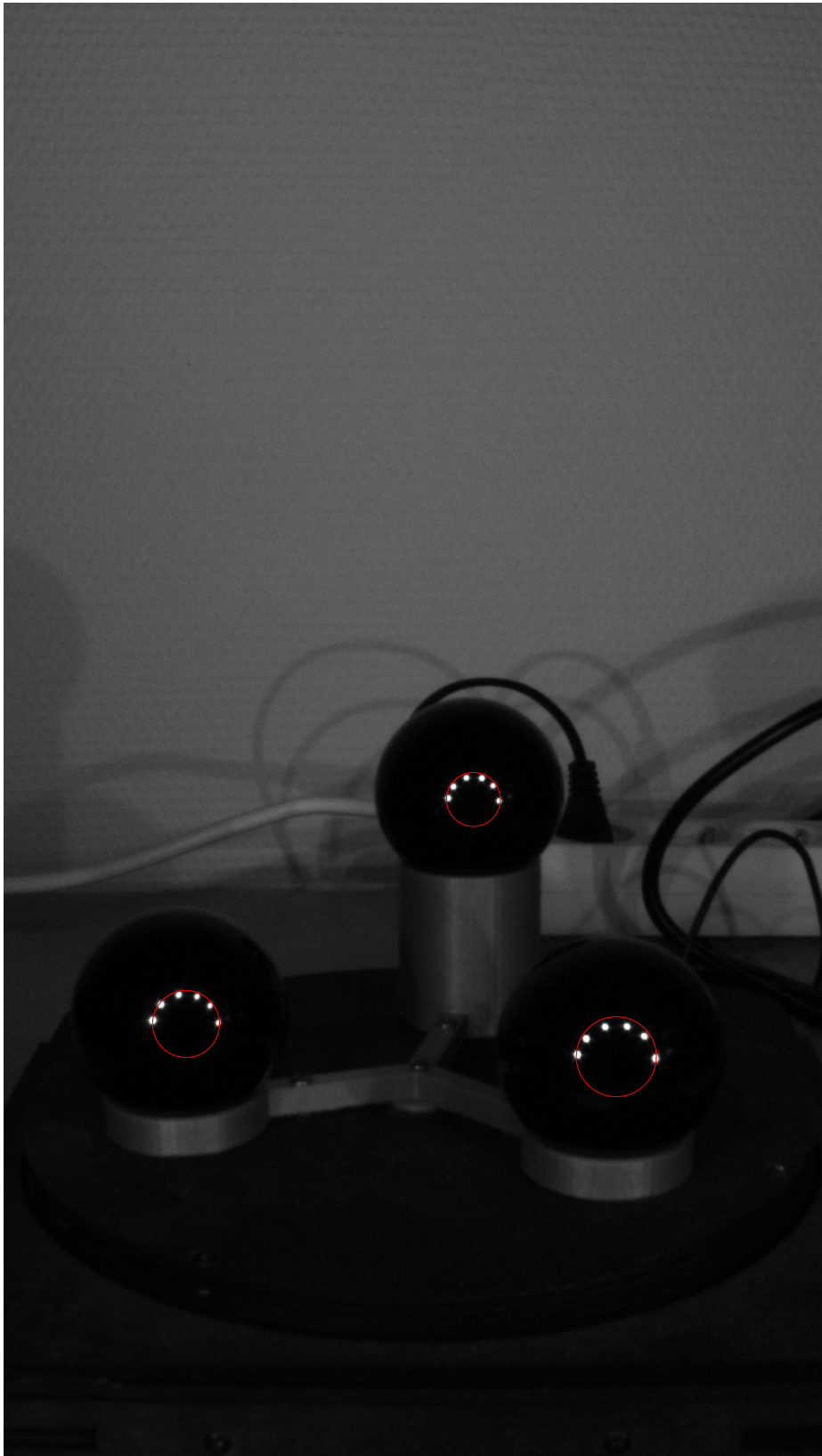
Figure 5: The captured reflections and their three smallest enclosing circles

(a) The intersection of a line with a cylinder

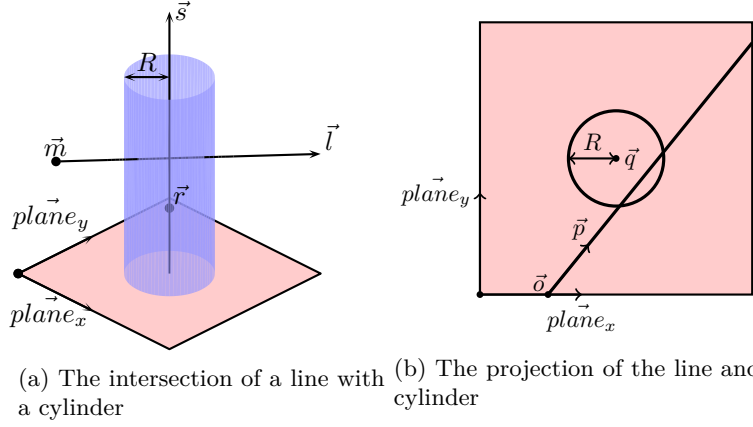(b) The projection of the line and cylinder

Figure 6: Computing the intersection of a line and a cylinder

computed as follows.

$$\vec{plane}_x = \vec{s} \times \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{1a}$$

$$\vec{plane}_y = \vec{s} \times \vec{plane}_x \tag{1b}$$

$$\vec{q} = \begin{pmatrix} \vec{r} \cdot \vec{plane}_x \\ \vec{r} \cdot \vec{plane}_y \end{pmatrix} \tag{1c}$$

$$\vec{o} = \begin{pmatrix} \vec{l} \cdot \vec{plane}_x \\ \vec{l} \cdot \vec{plane}_y \end{pmatrix} \tag{1d}$$

$$\vec{p} = \begin{pmatrix} \vec{m} \cdot \vec{plane}_x \\ \vec{m} \cdot \vec{plane}_y \end{pmatrix} \tag{1e}$$

In Formula 1a and 1b, two vectors are found that are orthogonal to the rotation axis, forming the orthogonal plane on which to project the cylinder and line. Now the problem is reduced to finding the intersection of a line and a circle in two dimensions and then finding the corresponding position on the 3-dimensional line back. To find the intersections between a line and a circle, set the standard form of the circle to be equal to the slope-intersect form of the line. For simplicity, let $\vec{q} = \vec{0}$, which can be achieved by subtracting $\vec{q}$ from $\vec{o}$. The standard form for a circle centred at position $\vec{0}$ with radius r is then as follows.

$$x^2 + y^2 - R^2 = 0 \tag{2}$$

The slope-intersect form for a line given the point $\vec{o}$ on the line and the direction

$\vec{p}$ can be found as follows.

$$y = t \cdot \vec{p}_y + \vec{o}_y$$
$$x = t \cdot \vec{p}_x + \vec{o}_x$$
$$t = \frac{x - \vec{o}_x}{\vec{p}_x}$$
$$y = \vec{o}_y + \vec{p}_y \frac{x - \vec{o}_x}{\vec{p}_x}$$
$$y = \vec{o}_y + \vec{p}_y \frac{x}{\vec{p}_x} - \vec{o}_x \frac{\vec{p}_y}{\vec{p}_x} \tag{3}$$

Equations 2 and 3 must both hold in the intersections of the line and the circle. The slope-intersect form of the line can substitute the y in the standard form of the circle.

$$x^2 + (\vec{o}_y + \vec{p}_y \frac{x}{\vec{p}_x} - \vec{o}_x \frac{\vec{p}_y}{\vec{p}_x})^2 - R^2 = 0$$

$$x^2 + \vec{o}_y^2 + \frac{2x\vec{o}_y\vec{p}_y}{\vec{p}_x} - \frac{2\vec{o}_x\vec{o}_y\vec{p}_y}{\vec{p}_x} + \frac{x^2\vec{p}_y^2}{\vec{p}_x^2} - \frac{2x\vec{o}_x\vec{p}_y^2}{\vec{p}_x^2} + \frac{\vec{o}_x^2\vec{p}_y^2}{\vec{p}_x^2} - R^2 = 0$$

$$\boldsymbol{x}^2 + \frac{\boldsymbol{x}^2\vec{p}_y^2}{\vec{p}_x^2} + \frac{2\boldsymbol{x}\vec{o}_y\vec{p}_y}{\vec{p}_x} - \frac{2\boldsymbol{x}\vec{o}_x\vec{p}_y^2}{\vec{p}_x^2} + \vec{o}_y^2 - \frac{2\vec{o}_x\vec{o}_y\vec{p}_y}{\vec{p}_x} + \frac{\vec{o}_x^2\vec{p}_y^2}{\vec{p}_x^2} - R^2 = 0 \tag{4}$$

In Equation 4, the variable x is marked bold and the quadratic formula can be recognised. This has the form $ax^2 + bx + c = 0$. Let a, b and c be computed as follows.

$$a = 1 + \frac{\vec{p}_y^2}{\vec{o}_x^2} \tag{5a}$$

$$b = \frac{2\vec{o}_y\vec{p}_y}{\vec{p}_x} - \frac{2\vec{o}_x\vec{p}_y^2}{\vec{p}_x^2} \tag{5b}$$

$$c = \vec{o}_y^2 - \frac{2\vec{o}_x\vec{o}_y\vec{p}_y}{\vec{p}_x} + \frac{\vec{o}_x^2\vec{p}_y^2}{\vec{p}_x^2} - R^2 \tag{5c}$$

Solving this equation for x is trivial with the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-\frac{2\vec{o}_y\vec{p}_y}{\vec{p}_x} + \frac{2\vec{o}_x\vec{p}_y^2}{\vec{p}_x^2} \pm \sqrt{(\frac{2\vec{o}_y\vec{p}_y}{\vec{p}_x} - \frac{2\vec{o}_x\vec{p}_y^2}{\vec{p}_x^2})^2 - 4(1 + \frac{\vec{p}_y^2}{\vec{o}_x^2})(\vec{o}_y^2 - \frac{2\vec{o}_x\vec{o}_y\vec{p}_y}{\vec{p}_x} + \frac{\vec{o}_x^2\vec{p}_y^2}{\vec{p}_x^2} - R^2)}}{2(1 + \frac{\vec{p}_y^2}{\vec{o}_x^2})}$$

$$= \frac{-\frac{\vec{o}_y\vec{p}_y}{\vec{p}_x} \pm \sqrt{\frac{\vec{o}_y^2\vec{p}_y^2}{\vec{p}_x^2} - \frac{\vec{o}_x\vec{o}_y\vec{p}_y^3}{\vec{p}_x^3} + \frac{\vec{o}_x^2\vec{p}_y^4}{\vec{p}_x^4} - (1 + \frac{\vec{p}_y^2}{\vec{o}_x^2})(\vec{o}_y^2 - \frac{2\vec{o}_x\vec{o}_y\vec{p}_y}{\vec{p}_x} + \frac{\vec{o}_x^2\vec{p}_y^2}{\vec{p}_x^2} - R^2)}}{1 + \frac{\vec{p}_y^2}{\vec{o}_x^2}}$$

$$\tag{6}$$

This will find at most two points on the line where the line intersects with the circle. Here, x is expressed in multiples of $\vec{p}$, the projected direction vector of the line. Therefore to obtain the 3-dimensional position of the intersection between

the line and the cylinder, the same x can be filled in the original formula of the line.

$$\vec{i} = x\vec{m} + \vec{l} \tag{7}$$

If the determinant $b^2 - 4ac$ is equal to 0, then the two outcomes will be equal, but this does not influence the outcome of the calibration. If the determinant is negative, then $\sqrt{b^2 - 4ac}$ is undefined and no solution can be found. The estimated initial position misses the cylinder of possible solutions. In that case, the point on the cylinder that is closest to the line is taken as initial estimate of the sphere's position. The intersections are visualised in Figure 7. The positions of the spheres are visualised in yellow. The cylinder is visualised in blue. The rays of the initial guesses are visualised as black lines.
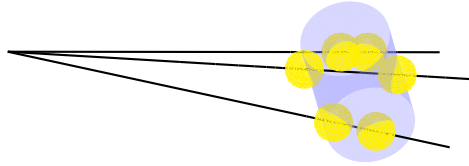

Figure 7: Finding the positions of the 3 spheres via line-cylinder intersection

For every sphere there will now be two possible locations. To choose from among these, all possibilities are simply enumerated. Since the three spheres are known to be equally spaced, the combination of spheres that best approximates an equilateral triangle is chosen, with the extra requirement that the highest sphere must be in the back. This entails a good approximation of the locations of the spheres using only the centroid of the reflections, but it can be further refined.

3.5.5 Refining Sphere Positions and Computing Light Positions

The size of the spheres is also known. Furthermore, the assumption was made that the spheres are perfectly specular and that a reflection originates from a single point in 3-dimensional space. This information can be used to refine the positions of the spheres. A hill-climbing technique is employed to minimise the error. The error measure used to this end is based on the latter assumption, that the light originates from a single point in 3-dimensional space.

To measure the error, the light rays are traced to find their source. A ray is cast through the three pixels where the reflections are seen for a single light source. The rays are bounced off their respective spheres. The error measurement indicates how well these three rays converge into one point.

The reflection of the rays requires finding the point of intersection between a line and a sphere. This can be found with a similar approach as the line-cylinder intersection, but the formulas are simpler. The standard form of spheres is set equal to the slope-intersect form of lines. Let $t\vec{d} + \vec{e} = \vec{x}$ be the slope-intersect form for a line with slope-intersect e and direction d. Let $|\vec{x} - \vec{f}|^2 - R = 0$ be the standard form for a sphere with radius R and centre f.

$$\|\vec{e} + t\vec{d} - \vec{f}\|^2 = R^2$$
$$(\vec{e} + t\vec{d} - \vec{f}) \cdot (\vec{e} + t\vec{d} - \vec{f}) = R^2$$
$$t^2(\vec{d} \cdot \vec{d}) + 2t(\vec{d} \cdot (\vec{e} - \vec{f})) + (\vec{e} - \vec{f}) \cdot (\vec{e} - \vec{f}) - R^2 = 0 \qquad (8)$$

With t as the only unknown, the quadratic formula can be recognised. This has the form $at^2 + bt + c = 0$, with a, b and c as follows.

$$a = \vec{d} \cdot \vec{d} = \|\vec{d}\|^2 = 1 \qquad (9a)$$
$$b = 2(\vec{d} \cdot (\vec{e} - \vec{f})) \qquad (9b)$$
$$c = (\vec{e} - \vec{f}) \cdot (\vec{e} - \vec{f}) - R^2 = \|\vec{e} - \vec{f}\|^2 - R^2 \qquad (9c)$$

The solution for this quadratic equation is trivial using the quadratic formula.

$$\begin{aligned}
t &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
&= \frac{-2(\vec{d} \cdot (\vec{e} - \vec{f})) \pm \sqrt{(2(\vec{d} \cdot (\vec{e} - \vec{f})))^2 - 4(\|\vec{e} - \vec{f}\|^2 - R^2)}}{2} \\
&= -(\vec{d} \cdot (\vec{e} - \vec{f})) \pm \sqrt{(\vec{d} \cdot (\vec{e} - \vec{f}))^2 - \|\vec{e} - \vec{f}\|^2 - R^2} \qquad (10)
\end{aligned}$$

The resulting value t can be put into the formula for a line, $\vec{x} = t\vec{d} + \vec{e}$, to get the point of intersection. The formula allows up to two solutions for t. If $\vec{e}$ is set to the camera position and $\vec{d}$ is set along the viewing direction of the camera, the intersection with lowest t is where the ray hits the sphere first and thus where the ray will be reflected. If the determinant is negative, there are no solutions and the ray will have missed the sphere.

The direction of the reflected ray can be computed using the following formula for mirror-like reflection.

$$\vec{r_o} = \vec{r_i} - 2\vec{N}(\vec{r_i} \cdot \vec{N}) \qquad (11)$$

Where $\vec{r_o}$ is the outgoing ray direction, $\vec{r_i}$ the incoming ray direction and $\vec{N}$ the direction of the normal vector, as depicted in Figure 8. The normal vector $\vec{N}$ can be found by taking the difference vector from the sphere's centre to the point of intersection and normalising it.

Given three lines, the point of intersection must be found to obtain the supposed position of this light. The lines rarely intersect precisely, but a point can be found that is close to all three, minimising the squared distance to each line with a least squares algorithm.

To this end, define a projection matrix P_l for the line l: $\vec{x} = t\vec{d_l} + \vec{e_l}$ that projects any vector onto a plane perpendicular to $\vec{d_l}$.

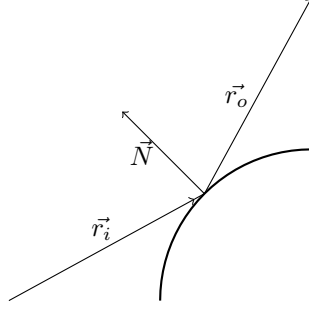$$P_l = I - \vec{d_l} * \vec{d_l}^T \qquad (12)$$

Figure 8: Mirror-like reflections on a sphere, computed in Equation 11

Using the projection matrix found in Equation 12, the point where l intersects this plane can be found by projecting any point on l to the plane, such as $\vec{e_l}$.

$$\vec{i} = P_l \vec{e_l} \qquad (13)$$

Then the distance from any vector $\vec{x}$ to l can be found by projecting $\vec{x}$ and $\vec{e}$ both onto the plane and computing the distance between the projections.

$$||\vec{x} \to l|| = ||P_l \vec{x} - P_l \vec{e_l}|| \qquad (14)$$

Therefore the sum of squared distances to lines $L = \{l_1, ..., l_n\}$ can be found as follows.

$$\epsilon = \sum_{l \in L} ||P_l \vec{x} - P_l \vec{e_l}||^2 \qquad (15a)$$

$$= \vec{x}^T \sum_{l \in L} P_l \vec{x} - 2(\sum_{l \in L} P_l \vec{e_l})^T \vec{x} + \sum_{l \in L} ((P_l \vec{e_l})^T \cdot P_l \vec{e_l}) \qquad (15b)$$

$$= \vec{x}^T A \vec{x} - 2\vec{b}^T \vec{x} + \vec{c}, \text{ where} \qquad (15c)$$

$$A = \sum_{l \in L} P_l \qquad (15d)$$

$$\vec{b} = \sum_{l \in L} P_l \vec{e_l} \qquad (15e)$$

$$\vec{c} = \sum_{l \in L} ((P_l \vec{e_l})^T \cdot P_l \vec{e_l}) \qquad (15f)$$

Equation 15c can be minimised for $\vec{x}$ by solving its derivative $A\vec{x} = \vec{b}$ using least squares.

$$\vec{x} = (A^T A)^{-1} A^T \vec{b} \qquad (16)$$

This results in the 3-dimensional position $\vec{x}$ which minimises the sum of squared distances to all lines in L and allows for simple sampling of the error by plugging $\vec{x}$ into Equation 15c.

The resulting hill-climbing technique makes minor adjustments to the height and angle of the calibration statue to minimise this error. A sample of the search space is drawn in Figure 9.
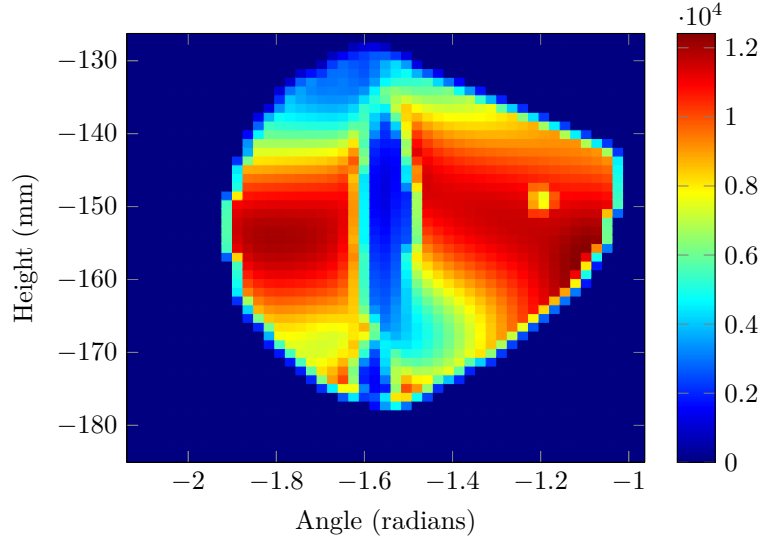

Figure 9: A measurement of the search space for hill-climbing

Once the most optimal orientation of the calibration statue is found, the error is sampled once again with the reflected rays and Equation 15c. The resulting vector $\vec{x}$ becomes the estimated position of that light source.

3.5.6 Rotating the Platform

Having a rotating platform introduces numerous advantages for calibration. One of these has already been described in Section 3.5.4: It allows the scanner to find the axis of rotation and allows a calibration statue of known dimensions to be fixed to this axis, which greatly reduces the degrees of freedom of the calibration problem. This section will describe a second advantage, being able to make multiple measurements without user interaction.

When calibrating a 3D scanner, it is a feasible idea to make multiple measurements. With multiple measurements, noise can be reduced by observing patterns in the data. Even just taking the average of the measurements reduces noise due to the Law of Large Numbers, because any outliers will eventually be countered with an outlier in the other direction. However, this law only applies if the measurements are independent and identically distributed [HR47]. Taking multiple measurements of a static scene results in measurements that are very dependent. They may all suffer from the same coincidental fault, caused for instance by environmental light being reflected in a specific way.

Making the platform rotate drastically reduces the dependence of the measurements. They will still not be completely independent, since they are made with the same calibration statue at the same height in the same room with the same environmental lighting conditions, but they are sufficiently independent that the Law of Large Numbers comes into play. A larger number of measure-

ments should then increase the accuracy of the calibration, but will take more time.

The technique proposed here is to take a small number of measurements, while between these measurements the platform is rotated slightly. A measurement is taken as described in Section 3.5.5. A parameter α_{max} is supplied to the program, which denotes the maximum deviation from a neutral angle, and a parameter N_α is supplied to denote the number of measurements to take. The platform first rotates to the angle $-\alpha_{max}$ and a measurement is taken. Then, repeated $N_\alpha - 1$ times, the platform is rotated with angle $\frac{\alpha_{max}}{N_\alpha - 1}$ and another measurement is taken. This results in N_α measurements in total, varying in angle from $-\alpha_{max}$ to α_{max}.

For each light, let μ_0 be the average position of all measurements and σ the standard deviation. The final result will be the average of all measurements that deviate less than σ from μ_0. To be precise, let $\mu = \sum_{m \in M} \frac{m}{||M||}$ and $M = \{m \in D | m - \mu_0 < \sigma\}$, where μ is the refined light position estimate and D is the measurement data for a light.

There is some information yet unused. The lights are known to be fixed onto an arc that was cut from a flat piece of wood. The positions of the lights should be approximately planar. This gives rise to the final tweak of the light positions: Fit a plane through the refined light positions and project all lights onto that plane. These projected light positions are the final calibration results that are used in the scanning algorithm. This concludes the description of the calibration algorithm.

The only part of this algorithm that scales in time complexity with the number of pixels is the first, which finds the positions of the reflections in the camera image. This is a matter of finding the three brightest pixels and can be performed in linear time complexity and can be parallelised. The other parts of the algorithm all scale with the number of spheres or the number of lights, which is negligible compared to the number of pixels. The calibration requires the user to place down the calibration statue, activate the calibration sequence and remove the statue, but these interventions are necessary for all calibration techniques involving calibration objects. For the rest, everything is automatic. No technical knowledge is required. This satisfies the preliminary requirements to the calibration algorithm.

3.5.7 Verification

To verify if the calibration is correct, it is useful to have some measurement of the error in the calibration. This can be communicated to the user, possibly warning them that something might have gone wrong. They might then check to see if all lights are not occluded, the hardware is properly working and the calibration statue is properly positioned with the spheres firmly on top.

During every step in the algorithm, information has been thrown away. This information can be used to measure the error. Possibilities for error measurement are listed below.

- Information is discarded when finding the positions of the reflections in the camera images, when a grid of pixel values is converted into just three 2-dimensional coordinates. The information that is discarded here is the

brightness of places where there is no reflection. If there are more reflections than the number of spheres, that means that either the lights reflect on some other surface or that there is a changing environment. If some of these other reflections or light sources are very bright, there is no distinction to the software between the actual light source and this noise, and the wrong coordinate might have been chosen. This sort of information is difficult to analyse well and not used in the implementation of this technique, but is a possibility for future research. The other information that is discarded, the brightness of the light sources that are selected, could be used to determine the relative brightness of the lights. For the purpose of this thesis, the lights are assumed to be equally bright.

- After refining the positions of the spheres, some error still remains after the hill-climbing has converged. This is the result of compressing the positions of the reflections into the two degrees of freedom for the calibration statue's position. This error could be used directly for verification.

- Information is discarded when multiple measurements from different platform rotations are combined. The deviation from the average is a direct measurement of error that can be communicated to the user.

- When finally the lights are projected onto a plane, this discards more information: the difference of the refined light position from the plane. That means that the projection distance to the resulting plane is also a measure of error. This error is the easiest to measure, since there are no multiple measurements involved any more at this point.

If one of these is out of order (for instance there are an unusual number of reflections in the camera or the lights are very far from the final plane), a warning is given to the user. For the scientific verification of the calibration method in Section 6.1, a conventional ruler is used rather than these error measures, since an invalid method would invalidate these error measures too.

3.5.8   Assumptions

Some assumptions are made for the calibration of the lights that will interfere with the calibration quality if they do not hold.

- The lights are point lights. If they are, every light would have a specular reflection that falls in a single pixel. In truth, the lights have a surface from which they emit light. This surface is several pixels wide in the camera image. In [SW13], a technique is explained for fitting an oval to the specular reflection of a light in a sphere, to obtain a more accurate light direction, but this is not implemented for the experiments in this thesis due to time constraints.

- Every light is equally bright. For this thesis, six lights from the same series of fabrication have been used. They are approximately equally bright. Some slight variation may occur however and other set-ups might combine different types of lights or lights of different colours. This can be calibrated for by observing the intensity of the reflections in the spheres, but the intensity of the reflections falls above the intensity range that can be measured by the camera that was used in this research.

24

- The lights emit an equal amount of light in every direction. In reality, a LED emits most light along its axis and this gradually descends to darkness along its perpendicular. This results in slight distortions in the resulting scan. This is difficult to model because it requires the calibration sequence to learn the orientation of the lights.

- The lights do not move. This assumption is inherent in the calibration of the lights. The wooden arc used in these experiments was wobbly and the lights inevitably wobbled a few tenths of millimetres when the scene was changed.

- Lights are assumed to be white. The specifications however specify that they emit a slightly yellow light. This means that yellow surfaces would appear more reflective than blue surfaces, even if they are in reality equally reflective. This should not affect the final scan if the lights all have the same colour.

4 Scanning

As described in Section 1, photometric stereo is a technique to find the normal vectors of a surface by observing the scene under different lighting conditions. The basic idea is to measure the amount of light that gets reflected on the surface from each direction and constrain by the reflectance properties of the surface what the possible orientations of the surface could be. This section describes in detail how the basic photometric stereo technique works that is used in these experiments. The photometric stereo technique that is used here is no new research. It is not the focus of this research to develop an advanced photometric stereo technique but to fuse its results with the other technique. Therefore this section only illustrates the simplest technique to give an idea of how to replicate the results and what sort of data is expected in the fusion with the structured light scan.

4.1 The Lambertian Model

The photometric stereo method used here makes the assumption that the surface of the item to scan is perfectly diffuse. This means that the surface scatters an equal amount of light in every direction. To model this, the Lambertian model is used. The Lambertian model rests on Lambert's Cosine Law [PP93, p. 11], stating that the radiance L_e of a surface patch is directly proportional to the cosine of the angle α_χ between the incoming light χ and the normal vector of the surface. For these formulas the superscript χ is added to denote a metric for the case when only the light $\chi \in X$ is activated. The subscript χ is added to denote a property of the light χ itself.

$$L_e^\chi \propto \cos \alpha_\chi \tag{17}$$

Lambert's model also takes the intensity of the light source into account. The radiance of the surface is also directly proportional to the intensity of the incoming light, E_e^χ.

$$L_e^\chi \propto E_e^\chi \cos \alpha_\chi \tag{18}$$

The cosine of the angle between the light source χ and the normal vector can be computed with the dot product of the two vectors, which brings us to the most common definition of Lambert's model. Let $\vec{N}$ be the normal vector of the surface and $\vec{\lambda_\chi}$ the direction of the incoming light χ.

$$L_e^\chi \propto E_e^\chi (\vec{N} \cdot \vec{\lambda_\chi}) \tag{19}$$

The model that is assumed in this thesis also takes the albedo of the surface into account. Albedo is not originally included in the Lambertian model, but a constant albedo is simple to implement and drastically increases the variety of surfaces that can be scanned. A scalar albedo per pixel allows for scanning of diffuse textures. In the case of constant albedo, the albedo is a scalar, indicating the fraction of light that gets reflected by the surface (and by extension the fraction of light that gets absorbed). This makes the albedo ρ also directly proportional to the radiance of the surface.

$$L_e^\chi \propto \rho E_e^\chi (\vec{N} \cdot \vec{\lambda_\chi}) \tag{20}$$

Equation 20 is the model of surface reflectance assumed in this thesis. Any deviation from this model is detrimental to the quality of the scan. More complex models exist, allowing for a wider variety of surfaces to be scanned. For an overview of the most influential surface reflectance models, refer to Section 2.1.

4.2 Inverting the Lambertian Model

Photometric stereo is a technique to find the normal vectors of a surface. In Equation 20, this vector is indicated by $\vec{N}$. To this end, the reflectance equation is solved for $\vec{N}$. This section explains how to invert the reflectance equation to solve for $\vec{N}$.

The radiance of a patch on the surface is measured by the camera. The camera measures the irradiance of the camera $E_{e,C}^\chi$, how much light is received by the camera *indirectly* from a certain patch of surface when illuminated with light χ. However, this is not all of the light that is emitted by that surface. Most of the reflected light does not enter the camera. How much light enters the camera is governed by the inverse-square law. This law states that the irradiance of the light decreases proportional to the distance squared, $E_{e,C}^\chi \propto \frac{L_e^\chi}{(d_C)^2}$, where d_C is the distance of the surface to the camera. This brings the reflectance equation to the following.

$$E_{e,C}^\chi (d_C)^2 \propto \rho E_e^\chi (\vec{N} \cdot \vec{\lambda_\chi}) \tag{21}$$

The intensity of the incoming light also decreases with the inverse-square law, proportional to the distance from the surface to the light source. Let this distance be d_χ and let $L_{e,\chi}$ be the inherent radiance of the light source χ. Then $E_e^\chi \propto \frac{L_{e,\chi}}{(d_\chi)^2}$. The reflectance equation becomes the following.

26

$$E_{e,C}^{\chi}(d_C)^2 \propto \rho \frac{L_{e,\chi}}{(d_\chi)^2}(\vec{N} \cdot \vec{\lambda_\chi}) \tag{22}$$

Of these variables, d_C, d_χ and $\vec{\lambda_\chi}$, can be found only when the position of the surface is already known. This poses a circular problem: To find the surface geometry, the surface geometry must be known. Luckily, an estimate of the surface geometry is readily available. As discussed in Section 5, the primary function of the photometric stereo scan is to refine the low-frequency data of the structured light scan with the high-frequency data of the photometric stereo scan. With this intention in mind, it is a logical decision to use the structured light data as initial estimate. The initial estimate can later be refined in a second run of the algorithm, eventually converging to an accurate surface.

$L_{e,\chi}$, the inherent radiance of the light source, is assumed to be a constant. This intensity is assumed to be the same for all lights (see Section 3.5.8). Therefore, the light radiance can be denoted as a single constant L_e.

$E_{e,C}^{\chi}$ is directly measured by the camera, using the assumptions listed in Section 3.2.3.

This leaves two variables unknown, ρ and $\vec{N}$. Solving these with only one light source is an underconstrained problem. Therefore, the solution must be obtained by solving a system of equations, one for every light source $\chi \in X$. The photometric stereo method with diffuse surfaces requires at least 3 light sources [Woo79], at least 3 equations to solve (inherent in the invertibility of matrices composed of 3-dimensional vectors). This system of equations can be solved using linear algebra. To this end the metrics of each light is combined:

$$\forall_{\chi \in X}(L_e = L_{e,\chi}) \tag{23a}$$
$$\vec{E_{e,C}} = (E_{e,C}^{\chi_1}, ..., E_{e,C}^{\chi_n}) \tag{23b}$$
$$\vec{d} = (d_{\chi_1}, ..., d_{\chi_n}) \tag{23c}$$
$$\Lambda = (\vec{\lambda_{\chi_1}}, ..., \vec{\lambda_{\chi_1}}) \tag{23d}$$

The system of linear equations becomes as follows.

$$\vec{E_{e,C}}(d_C)^2 \propto \rho \frac{L_e}{\vec{d}^2}\vec{N}\Lambda \tag{24}$$

This system can be solved for its three unknowns, $\rho L_e \vec{N}$.

$$\rho \frac{L_e}{\vec{d}^2}\vec{N}\Lambda \propto \vec{E_{e,C}}(d_C)^2$$
$$\rho L_e \vec{N}\Lambda \propto \vec{E_{e,C}}(d_C)^2 \vec{d}^2$$
$$\rho L_e \vec{N} \propto \Lambda^{-1}\vec{E_{e,C}}(d_C)^2 \vec{d}^2 \tag{25}$$

The normal vector $\vec{N}$ has length 1. Therefore, $\vec{N}$ can be found by normalising the vector $\rho L_e \vec{N}$, c.q. $\vec{N} = \frac{\rho L_e \vec{N}}{||\rho L_e \vec{N}||}$. The actual albedo cannot be obtained

this way, but since L_e is a constant a relative albedo can be obtained that allows comparison of the albedo of different patches of surface, by letting $\rho L_e \propto ||\rho L_e \vec{N}||$.

Equation 25 requires Λ^{-1}, which only exists if Λ is invertible. Λ must be square in order for it to be invertible, meaning a solution can only be found if there are 3 lights in the system. Therefore, rather than using the inverse matrix of Λ the Moore-Penrose Pseudoinverse is used, $(\Lambda^T \Lambda)^{-1} \Lambda^T$. Therefore, the normal vector $\vec{N}$ is finally computed as follows.

$$\vec{N} = \frac{(\Lambda^T \Lambda)^{-1} \Lambda^T \vec{E_{e,C}} (d_C)^2 \vec{d^2}}{||(\Lambda^T \Lambda)^{-1} \Lambda^T \vec{E_{e,C}} (d_C)^2 \vec{d^2}||} \tag{26}$$

4.3 The Initial Estimate

In Equation 26, the variables d_C, $\vec{d}$ and Λ depend on an initial estimate of the position of the surface. If the surface position is known, the distance to the camera, distance to the light and direction to the light can be computed trivially. For this, a different scanning technique is used.

Range imaging techniques based on triangulation have been known to exhibit high-frequency noise, while their low-frequency signal is correct [NRDR05]. This is also the case for structured light [SC04]. With this observation, the structured light scan can be used as initial estimate for the photometric stereo scan.

The initial estimate is a render of the depth map for the point cloud that results from a structured light scan, as seen from the camera. All points of the point cloud are projected onto the camera image using the camera calibration matrix, which is available from the camera calibration sequence. Then for every point the distance to the camera is computed. The smallest distance is stored in every pixel where a point is mapped.

Since the structured light scan contains high-frequency noise, the high frequencies are removed from this data. This is done with a simple Gaussian blur. Care must be taken not to include pixels where no points are projected to in the weighted average of this blur.

4.4 Shadow Handling

One other deviation from the Lambertian model has been included to greatly increase the number of items that can be scanned correctly: Drop shadows. Shadows are nearly unavoidable in any scene. Lights are often occluded at places in the scene, mostly by the scene itself. At these places the Lambertian model is unreliable, since the surface is dark and no longer correlates much with the surface orientation. These shadows must be detected.

For shadow detection, a simple threshold is often not enough. Shadows may not be completely dark due to environmental light conditions, interreflections and other deviations from the Lambertian model. Barsky and Petrou propose a method in [BP03] that detects highlights and shadows by observing the difference in chromaticity, but this method is not applicable since the camera in this set-up measures only luminance. More recently a method was proposed that uses a graph cut to classify shadows [CAK07]. This was deemed too computationally intensive. Another common technique for shadow detection is to use

RANSAC [HVC08] [SZP10]. However, probably the most common technique when a superfluous number of lights are available is to simply mark the pixel in the shadow based on the relative luminosity of that pixel among the different lights [PB01]. This technique is chosen for the implementation of this thesis.

To detect for which lights a surface patch falls in the shadow, the average intensity of the patch is measured with the different lights activated. When the intensity of the patch with a light is below 20% of the average intensity, the patch is deemed to be in the shadow of that light.

The lights for which a patch is in the shadow are not included in the photometric stereo model. If there are fewer than 3 lights illuminating a patch of the surface, no photometric stereo data is available for that part and the resulting surface is wholly dependent on the structured light scan. How many lights are included in the photometric stereo model is stored for later use; it serves as initial confidence for the photometric data in the fusion stage.

In reality, shadows are not binary. Contrary to the assumptions, the light source is never a true point light. This causes the scene to cast penumbrae along the edges of every shadow. Increasing the threshold of the shadow detection will eventually remove too much signal. A better option is to extend the reach of the shadows by dilating the shadow map. This discards only the data in the penumbra and little else.

5 Merging

When both the structured light data and the photometric stereo data have been obtained, the two must be combined. This section details the process by which they will be combined. The requirements of this algorithm are as follows.

- The algorithm must run in linear time.

- The algorithm must be parallelisable.

- The process must run with as little user intervention as possible.

- The work flow of the process must be simple enough to be used by a layman.

- The resulting mesh must be more accurate than the mesh returned by the structured light scan.

The last requirement is the most important. It has been shown that the structured light scan primarily has high-frequency noise [SC04] and that the photometric stereo scan primarily has low-frequency noise [NRDR05]. Therefore, the low-frequency data of the structured light scan is combined with the high-frequency data of the photometric stereo scan. To achieve this, an approach is chosen that is similar to a Kalman filter and bears ideas from the field of information theory. The structured light scan has some uncertainty, but adding information to it will refine the data and reduce the uncertainty. This can be modelled by the convolution of probability density functions, which map the depth of the scene (the distance of a surface patch to the camera) to a probability density. The data of the photometric stereo scan is applied on a local level to the smoothed data of the structured light scan.

5.1 Probability Density Functions

Each camera pixel is assigned a probability density function. Initially, the probability density function is set equal to the probability density function of the structured light scan.

The assumption is made that the noise in the structured light scan is normally distributed and thus can be modelled with a Gaussian function. This simplification is not entirely correct, since the noise in the structured light scan with a Gray pattern seems to be caused mostly by aliasing effects, making it structural rather than random. In reality however, it comes close and this simplifies the calculations greatly. The probability density function for the depth of the surface behind a pixel according to structured light will therefore be a normal distribution. The mean of this normal distribution is the depth of the surface. The variance of the normal distribution should match the actual variance in the data. This is left as a parameter for the user to tweak the influence of the structured light scan. In the experiments conducted for this thesis, a variance of 5mm was used, since that seemed to give the best results.

New information enters the model via the normal vectors returned by the photometric stereo technique. These normal vectors have some noise in them as well, which is modelled with a probability density function. The normal vector conveys information on the direction of the surface. Combined with the position of the camera, this translates into depth displacement. The depth displacement defined by a normal vector is also assumed to be normally distributed. This assumption stems ultimately from the assumption made in Section 3.2.3 that the noise in the camera image is normally distributed. Therefore the depth displacement is also modelled with a Gaussian function.

The depth displacement is relative between two neighbouring pixels (denoted here as pixels 1 and 2). Every pair of neighbouring pixels is treated as one piece of information which implies a probability density function on the depth of both patches of surface under each pixel. To find the depth displacement according to the normal vectors of these pixels, first the two normal vectors of these pixels are combined by averaging them.

$$\vec{N} = \frac{\vec{N_1} + \vec{N_2}}{||\vec{N_1} + \vec{N_2}||} \tag{27}$$

Next, the tangent vector is found, which is perpendicular to the normal vector $\vec{N}$ and goes through both pixels. This is computed using cross products.

$$\vec{T_x} = \vec{N} \times \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{28}$$

$$\vec{T_y} = \vec{N} \times \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{29}$$

This way, the tangent vector in the x direction $\vec{T_x}$ is guaranteed to be perpendicular to $\vec{N}$ and to the y-axis. The tangent vector in the y direction $\vec{T_y}$ is guaranteed to be perpendicular to $\vec{N}$ and to the x-axis. The depth difference

implied by the normal vector is the depth-component of the size of the line segment in the direction of $\vec{T}$ which reaches from pixel 1 to pixel 2. Refer to Figure 10a for an illustration.
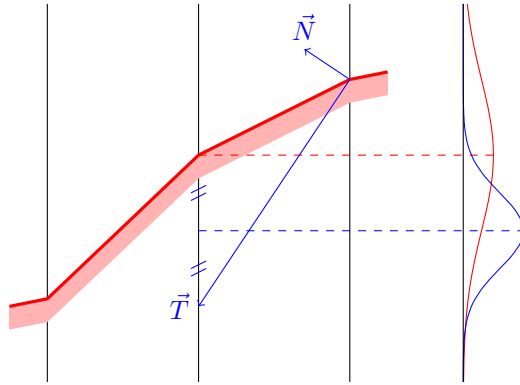
(a) Cross-section of a surface (red) where new information is added (blue)
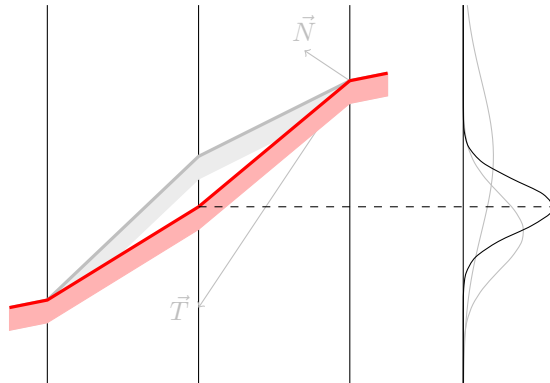
(b) New surface after incorporating new information

Figure 10: Merging new information into an initial estimate.

The depth difference of the old estimates is measured and subtracted from the new depth difference information. This results in a depth error. This error is in part due to the depth of the current pixel and in part due to the depth of the neighbouring pixel. Therefore, half of it is applied to the current pixel and half to the neighbouring pixel. The result is a new depth estimate for both pixels.

The certainty of this estimate, and thereby the variance of the resulting normal distribution, is dependent on how certain the accuracy of the normal vector $\vec{N}$ is and on the accuracy of the old depth difference. The variance of the normal distribution will therefore be the average variance of the two samples plus some constant variance for the normal vector. This constant is meant to depict the accuracy of the depth difference estimates and can be adjusted by the user. For the experiments in this thesis, it was estimated at 0.6mm.

Lastly, the old information is combined with the new. This means that the probability density functions are convolved. One of the advantages of the normal

distribution is that convolving two Gaussian functions is very simple [Bro03].

$$\frac{1}{\sigma_{f\otimes g}^2} = \frac{1}{\sigma_f^2} + \frac{1}{\sigma_g^2} \tag{30a}$$

$$\mu_{f\otimes g} = (\frac{\mu_f}{\sigma_f^2} + \frac{\mu_g}{\sigma_g^2})\sigma_{f\otimes g}^2 \tag{30b}$$

Using Equations 30a and 30b, the variance and mean of the new probability density function can be computed. Conveniently, the expected value of the normal distribution is its mean. Therefore $\mu_{f\otimes g}$ is the new best estimate for the depth of the surface.

5.2 Belief Propagation

Updating the depth of a surface patch, as described in Section 5.1 results in a different depth difference with its neighbours. This would change the difference in depth those neighbours have with their further neighbours. This gives rise for the need to propagate the new information forward across multiple pixels, using a technique called belief propagation.

Belief propagation has been used more often in the fusion of photometric stereo scans with other range imaging as a way to apply photometric stereo data on a local level [ZWYD08] [ATD+08]. This technique involves producing a Markov random field, which is a grid of random variables where neighbouring variables can have some correlation in their probability density functions. Commonly, this correlation is defined and the MAP-estimate of the Markov random field is then solved using an algorithm such as *L-BFGS-B*. These MAP-estimators are complex algorithms however, which is in conflict with the linearity requirement. They don't parallelise well. The algorithm explored in this thesis aims to find a different solution to see whether feasible results can be obtained in linear time.

In this algorithm, a Markov random field is constructed first with only the information from the structured light scan. This serves as the initial estimate. Next, the operation described in Section 5.1 is applied using the new information from the photometric stereo scan, changing the probability density functions of each pixel. The depth differences in this operation are computed using the expected values of the random variables before applying the information, so the order in which the new information is added is irrelevant. Next, the process is repeated, again applying the information from the photometric stereo scan for each pair of neighbouring pixels. This time however, the variance of the random variables has been reduced due to the convolution with another normal distribution. This causes the random variable to be more rigid when new information is added. In a natural way, the Markov random field converges to a state where the normal vectors have a locally enhancing effect while the global geometry remains that of the structured light scan. This phenomenon is also observed in practice: With the experimental set-up, all pixels had converged to a standard deviation of less than 1 m within approximately 12 iterations. When replicating this experiment, the convergence rate may vary according to the accuracy of both scans and the resolution of the camera image.

The algorithm described in this section performs only linear operations on the number of pixels. The time complexity is multiplied by a constant number of

propagations, arguably making it non-linear. However, the linearity constraint is intended to allow the scanning technique to scale well with camera resolutions of the future. Since the number of propagations does not need to scale along, the number of propagations can still be interpreted as a constant, making the algorithm linear. All operations where pixels are involved are local operations, involving only one pixel or a pair of neighbouring pixels. This allows for efficient parallelisation. Apart from placing the item onto the platform, no interaction or technical knowledge is required from the user. This satisfies three of the requirements on the scanning algorithm. The last requirement, whether the result is more accurate than the original range imaging technique, will be tested with the experiment described in Section 6.2.

6 Validation

To verify the validity of these techniques, two types of tests are performed, one aimed at verifying the calibration technique and one aimed at verifying the merging technique. This section details the experiments that were performed.

6.1 Calibration

The validity of the calibration technique is measured by comparing the final positions of the lights against a ground truth. The experiment was performed with different parameters to measure the influence of these parameters.

6.1.1 Ground Truth

The ground truth for the positions of the light is measured with rulers and markers. The lights are fixed on a large wooden arc. The dimensions of this arc are known due to the way it is fabricated with a laser cutter. The positions of the lights on it are measured. The arc is placed over the scanner such that it aligns with markers on the hull. The distance of these markers to the centre of the rotating platform is measured with a ruler. The position of the centre of the rotating platform is known from the calibration of the platform.

This fixes the position of the lights in the camera's coordinate space. The ground truth can then be computed via trigonometry with the known dimensions. The position is verified by plotting the positions of the lights in a 3D space alongside a scan of the calibration board that was made with the structured light technique.

6.1.2 Experiments

The calibration technique proposed in this thesis is designed with the idea that having a rotating platform allows the scanner to find the light sources in 3D coordinates rather than just the direction. To test the merit of this idea, the main experiment on the calibration technique involves making multiple measurements with different numbers of rotations. The expectation is that more rotations will cause the measured calibration to converge towards the ground truth.

The number of rotations is varied from 1 to 20. The experiment is repeated 50 times in order to obtain statistical information on the variance and mean of
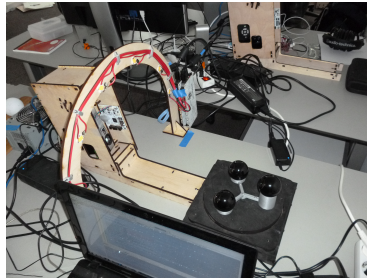
Figure 11: The set-up of the calibration experiments.

the error. The experiment always has a maximum angle of 20°from neutral, since that is approximately the largest angle at which all spheres are still completely in view of the camera. The angle convergence step and height convergence step size were set to 0.01, the exposure time factor set to 3 and the signal threshold to detect lights at is set to 150, since those parameters were found to give the best calibration results. The experiment is conducted in a room with low lighting conditions: The windows are blinded and the lights are kept off. This minimises the influence of weather and of people moving through the room. The room could not be kept closed.

A second experiment (of smaller scale) is conducted to measure the influence of the exposure time factor on the results. This experiment is conducted since the exposure time factor heavily influences the duration of the calibration sequence. Changing this parameter to something other than 1 causes the scanner to take multiple pictures with the camera for every light. These pictures should in theory be identical but due to measurement errors they will often differ. This removes some noise and allows the scanner to work with very weak light sources.

The exposure time factor is varied from 1 to 20 and the experiment is repeated 10 times. The settings are kept identical to the earlier test, except that the number of platform rotations is set to 10 and the exposure time factor now varies. The expected result of this experiment is that an increased exposure time results in slightly lower error, up to a point where noise from the camera is no longer a leading issue.

The results of these two experiments are detailed in Section 7.1.

6.2 Scanning

The validity of the combined scanning technique is measured by comparing it with results from scanners that are known to work as well as by scanning items with shapes that are known. This measurement is made by fitting the scan onto a ground truth and computing the average deviation of the points.

6.2.1 Comparisons

The scans from the technique described in this thesis are compared with a ground truth and 3 other techniques. The three other scanners are described here. All scanners are handled by an expert user to obtain the most accurate scan that can be achieved with the respective 3D scanners.

The first of the compared scanning methods is the *David SLS 2* 3D scanner, produced by David Group in 2014. This is a commercial product with wide recognition for its accuracy. It is also a structured light scanner, with a beamer and a camera on top of a ruler, allowing for adjustment of the parallax. It is equipped with an *Acer K132* projector, model *CWX1147*, and a stock *DAVID-CAM-3.1-M*, which has a 12mm *Computar* lens and an *Aptina AR0134* sensor. The scanner was set up with a vertical triangulation angle of 0.27°, a horizontal triangulation angle of 18.73°, resulting in a total triangulation angle of 18.73°. Furthermore, camera and projector are set at a distance of 376mm from each other. The exposure time was set to $\frac{1}{60}$s. The pattern size was calibrated at 240mm. The set-up of these measurements can be seen in Figure 12, except that the room was darkened before making the scans to eliminate ambient light.
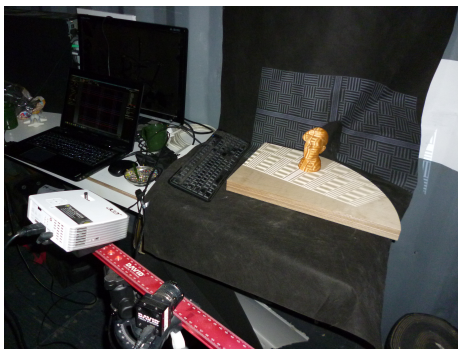

Figure 12: The set-up of the scans with the David SLS 2.

The second of the compared scanning methods is *3D Digital Corp.*'s *RealScan USB Model 100*. This is a laser scanner that is among the most widely used scanners in research. Due to this, its results have been validated often in other research over the years [MCC$^+$03] [SSBS04], making it a useful comparison. The gain of the laser is set to 129 and the laser stripe threshold to 40. Scans are made with 255 points per line and 300 lines per scan.

The third of the compared scanning methods is a 3D scanner made by *Ultimaker* in 2013 through 2015. It is as of this writing yet unreleased and no name for the device has been decided upon. The scanner is based on the structured light technique. The version used in these experiments has only the Gray pattern available, making it slightly less accurate than the *David SLS 2*. It uses the same hardware as described in Section 3, as the technique proposed in this thesis has been implemented as extension of this scanner. This is a useful comparison since it directly measures the influence of the combination of the photometric stereo method on the original results of this scanner.

The new scanner is calibrated first using the ground truth of the calibration in the previous experiments.

6.2.2 Items

The scans are compared using 8 different items with diverse properties. The items are listed below and their special properties with respect to scanning are described. The ground truth is obtained in different ways for different items, as

the nature of some items allows for better ways to obtain a ground truth than it does for others.

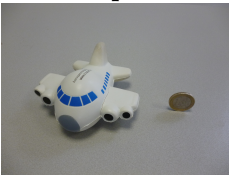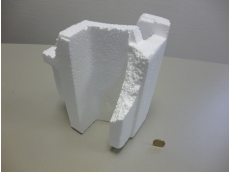|	
Cutting Board	A wooden cutting board. It has never been used and is flat and smooth. Its surface reflectance is matte, well approximating the Lambertian model, but it has the texture of wood, so proper separation of albedo and surface orientation is required to scan properly. The ground truth of this item is a plane. The resulting scans are cropped such that the corners are not included in the results, to make it fit to this ground truth better. This item was chosen for its simple ground truth.
Hockey Ball	A plastic ball for field hockey. It is sprayed with white *Crick 130* paint to make it matte. The ground truth for this item is a sphere. This item was chosen for its simple ground truth, while still being somewhat more difficult to scan without distortion.
Beethoven	A bust of Ludwig van Beethoven, printed in PLA by an Ultimaker 2 3D printer. The model has been sprayed with white *Crick 130* to reduce the specular reflections on its surface. The ground truth for this item is the digital 3D model that was originally used to print this item. The item was chosen because other research on photometric stereo has often used this model, though the 3D model used to print this item comes originally from a scan of the original bust, which inherently loses some detail. Nevertheless it is a useful test since it features good relief but no steep sides.
Colonel	A bust of a fictitious military figure called The Colonel. It is printed in golden PLA by an Ultimaker 2 3D printer. The ground truth for this item is the digital 3D model that was originally used to print the item. This item is chosen since it has a darker albedo than the rest and the print is very close to the ground truth.
Robot	The logo of Ultimaker printed with an Ultimaker Original 3D printer and sprayed with white *Crick 130*. It is a very rough print. The white paint has worn off in some places. Since the print is so rough, the original model that was used to print it is not deemed accurate. Therefore, the ground truth for this model is the scan from the *David SLS 2* scanner. This item is selected since it is the principal testing item of the developer of Ultimaker's original scanner.

Aeroplane	A model of an aeroplane made out of foam rubber as a stress reliever. It is matte and mostly white, but has dark blue windows and a grey nose. The shape of this item is not known a-priori, so the scan from the *David SLS 2* scanner is used as ground truth. This item is selected for its sharp contrasting albedo but matte surface.
Rock	A natural rock with quartz crystals inside, found in a river in France. Its surface is matte and has uneven albedo. It is chipped on one side, revealing a highly specular and slightly translucent cluster of quartz crystals inside. The ground truth for this item is again the scan of the *David SLS 2*. This item was chosen for its combination of matte surface and specular crystal.
Styrofoam	A block of styrofoam with many corners. It is very white. Since it consists of many small balls of foam, there is an interesting relief on its surface that is just at the limit of the amount of detail that can be scanned and it is broken such that it just fits within the view of the camera and projector. The shape of this item is not known a-priori, so the scan from the *David SLS 2* scanner is used as ground truth. This item is selected for its texture and geometric shape.

6.2.3 Experiments

To conduct the experiment to validate the scanning technique, the following steps are taken.

1. Every item is scanned from one side with every scanner. The same side of the item is scanned every time, taking care to maximise the overlap between scans. Only one side of the item is scanned because the laser scanner and the *David SLS 2* are not equipped with a rotating platform and the stitching method is not under scrutiny.

2. Each scan is manually aligned to the ground truth.

3. The alignment is refined using the Iterative Closest Point algorithm.

4. Any parts of the scans that don't overlap with the ground truth are removed. These cannot be compared if there is no data in the ground truth. This was only a concern for the scans of the rock.

5. A depth map is generated for each scan by rendering the resulting scan and writing the depth buffer to a file. This depth map is subtracted from the depth map of the ground truth in order to create a depth difference map, taking care not to include pixels where no points of either model are projected to. The resulting image is normalised.

6. The distance of the scan to the ground truth is measured as the average distance to the ground truth per point.

The last two steps produce two results for each scan: A depth difference map that gives a visual impression of the difference between the scan and the ground truth, and a numerical error value. The depth difference map is useful for visualisation, allowing visual inspection to get an idea of the sort of error that is obtained. The numerical error value is a more objective error measurement.

The expectation of this experiment is that the addition of data from a photometric stereo scan will strictly improve the accuracy of the scan. This means that the scans from this hybrid technique would be more accurate than the scans from Ultimaker's structured light scanner.

The results of these experiments are detailed in Section 7.2.

7 Results

In this section, the results are described of the experiments that were described in Section 6. The results are summarised and graphed and an interpretation is given to indicate how well the results are congruent with the premises and what this means for the applicability of the techniques.

7.1 Calibration

As described in Section 6.1, two experiments are conducted for calibration. The first is to measure the effect of multiple rotations on the calibration accuracy. The accuracy was measured with up to 20 rotations and each experiment was repeated 50 times, yielding a set of 20 box plots. These are depicted in Figure 13.

The result of this experiment is as predicted. The mean error gradually decreases with increasing numbers of rotations, from approximately 500mm to approximately 200mm. The variance in the error also gradually decreases. One notable result, which isn't depicted in Figure 13 is that with 2 rotations, the calibration failed in 31 out of 50 experiments (62%). This is likely because these experiments feature only the two most extreme angles of rotation, where the spheres may have been slightly occluded.

It must be noted however that the calibration error is still fairly high. Even with 20 rotations, the most inaccurate of the experiments had a total error of 506.200mm. This is the sum of the distances from calibrated position to ground truth for each of the six lights, meaning that the calibration is off by 84.3667mm on average in this particular case. Looking at the actual light positions, the calibrated light positions are found to often be too close to the platform, but have approximately correct direction. This would cause the scan to warp, were it not that the fusion algorithm guards against such low-frequency errors. This high error seems to be caused by accumulation of error at multiple stages throughout the calibration sequence. Even a slight error in the positions of the spheres causes a large error in the calibration. This is detrimental to the calibration quality.

The second experiment measures the effect of increasing exposure time on the calibration accuracy. The results are depicted in Figure 14.

The results from this experiment are not congruent with expectations. The error is indeed correlated with the exposure time, but seems to increase with
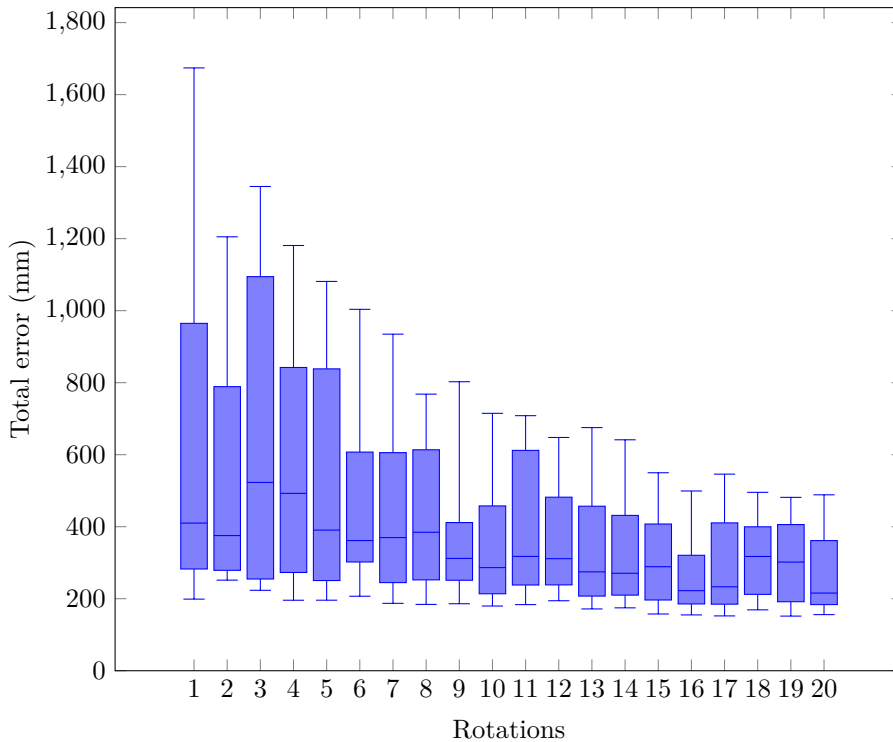
Figure 13: The error decreases with increasing number of rotations

longer exposure time. One possible explanation for this phenomenon is that the increased exposure time could cause a larger patch of the camera image to be classified as reflection from a light, making the false assumption that the light is a point light more apparent. This could be remedied by increasing the threshold at which a pixel is classified as a light reflection, but sensor noise does not seem to be a problem with the set-up of this experiment.

7.2 Scanning

The last experiment is to measure the accuracy of the scans. The result of that experiment are listed in Table 2.

The even numbered rows in Table 2 show the depth difference of the scans with the ground truth. Naturally, the difference of the ground truth to itself is always zero, causing those images to be black. For the robot, the aeroplane, the rock and the Styrofoam, the scan from the *David SLS 2* is used as ground truth, causing the difference to be zero too. It must be noted that in some cases, the alignment had to be repeated for the visualisation of these results, in order to be able to visualise them from the same angle or because the original alignment had been lost. This causes some shifts in the difference image of the Beethoven figurine for the *David SLS 2* as well as the Colonel and the sphere. The numerical results are accurate.

Upon visual inspection, the *David SLS 2* seems to produce the most accurate results. Especially the scans of the robot and the Styrofoam capture their details

Ground Truth	Laser	David SLS 2	Ultimaker	Hybrid

Table 2: The results of the comparisons of the scans. The odd numbered rows of this table show renders of the scans. The even numbered rows show the difference with the ground truth. 40
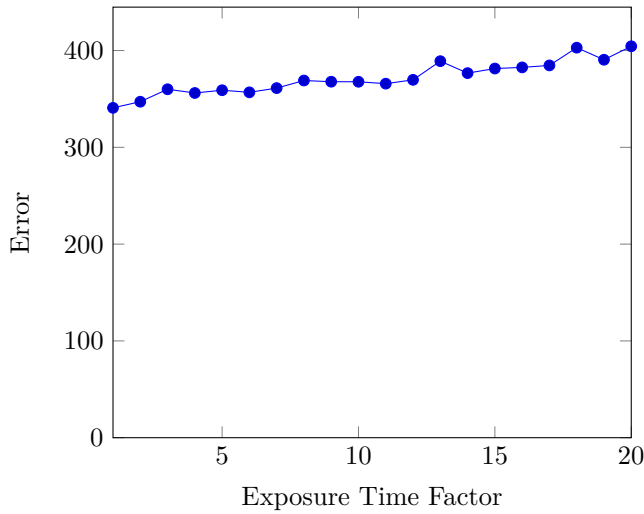
Figure 14: The effect of increasing exposure time on calibration accuracy.

well. This observation is the reason to choose the *David SLS 2* as ground truth for those items where there is no pre-defined model available. The *David SLS 2* does exhibit some noise on the surface, causing little bumps on all surfaces. The scans from the laser scanner are all fairly coarse, but it was the only scanner that performed well on the darker surface of the Colonel. Ultimaker's structured light scanner produces comparatively noisy scans, causing some details to be lost, but the scans were fairly accurate overall. The scans from the hybrid technique, the one described in this thesis, produces very smooth and accurate results but also exhibits Moiré patterns. These patterns are present in the structured light scan as well but become apparent when the high-frequency noise is removed, since Moiré patterns typically have low frequency. The hybrid technique also shows slim lines at the border of the shadows of the lights, indicating that the penumbrae have not completely been eliminated.

The difference images show that all scans perform well on the plane, but none perform well on the sphere. A possible explanation is that the ball is not completely round, invalidating the ground truth, but the automatic alignment algorithm that was used to align the scans to the sphere also does not perform well when matching parts of a sphere to a full sphere. Further, the depth images show a slight warping of the hybrid technique towards the sides of many scans. The sides are typically less well illuminated, causing the normal vectors to be off, which might explain this warping effect. Lastly, an unexpected result is that all scans seem to perform better on the farther surfaces of the Styrofoam rather than those in front, where the opposite would be expected.

Aside from visual inspection, the experiment produces numerical comparisons as well, listing the average distance of a scan to the ground truth. Since this is very dependent on the alignment, great care has been taken to ensure a good alignment in cases where the Iterative Closest Point algorithm does not perform well (mainly the sphere and the Colonel). The results of this are visualised in Figure 15.

Note again that the *David SLS 2* is the ground truth for the robot, aeroplane,
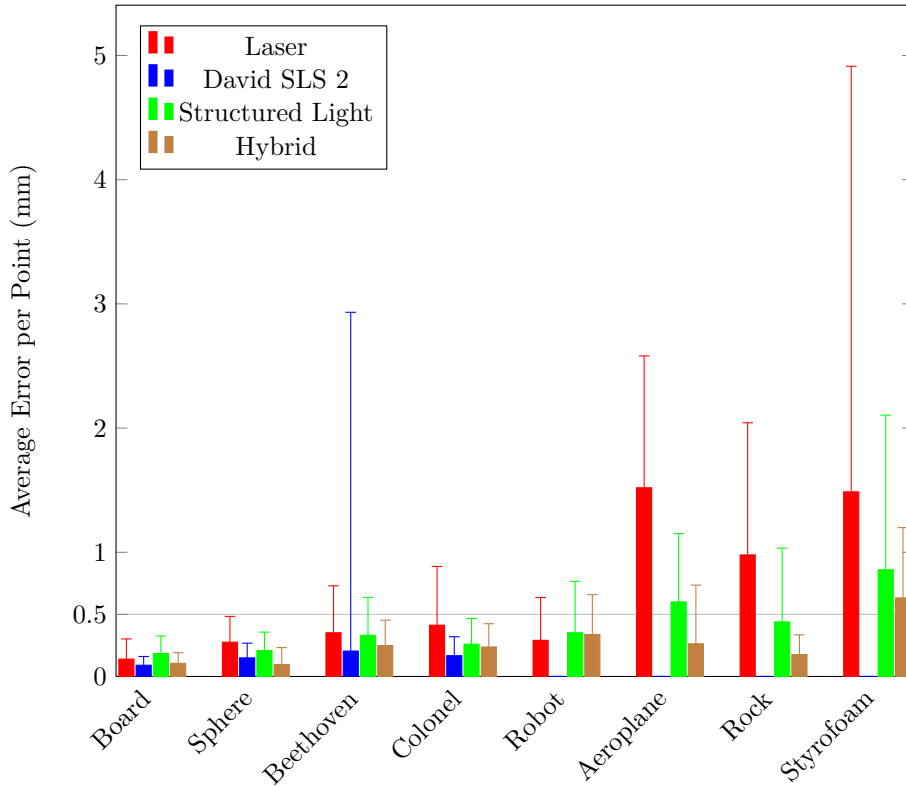
Figure 15: Numerical measurement of error in the scans.

rock and Styrofoam, which is why the error for those tests is 0. According to the numerical results in Figure 15, the laser scan generally performs worst, followed by Ultimaker's structured light scanner, the hybrid scan and finally the *David SLS 2*. In all cases, the addition of photometric stereo data to the scan reduced the error.

8 Conclusion

This thesis introduced two new techniques. One meant to increase the accuracy of the calibration by making multiple measurements from different angles. The other meant to efficiently combine a photometric stereo scan with a structured light scan. The process of scanning is described in its entirety, including the calibration of the hardware, the new calibration of the lights, the photometric stereo scan and the new fusion technique. Three experiments have been set up and the results have been discussed extensively.

In Section 3.5, the new technique for calibration was described. The technique involves a statue with three spheres on top of a rotating platform. The lights are turned on and recorded in the reflections of the spheres. The positions of the spheres are first estimated, then refined by looking at how close the reflections converge into one point. Then the reflections are traced to find the

position of the lights. The calibration statue can then be rotated to produce a new independent test in order to improve the accuracy of the calibration. In Section 7.1, the results of the calibration experiments showed that calibration from multiple angles does indeed reduce the error in the calibration. However, the technique requires a complex set-up with 3 spheres that is extremely sensitive to noise at multiple stages in the process. This caused the general calibration quality to decrease such that the new calibration technique no longer produces better calibrations than more traditional techniques.

In Section 4, a photometric stereo scanning technique was described, which uses results from the structured light scan to produce an initial estimate, allowing photometric stereo to work with point lights rather than directional lights.

In Section 5, a new technique is described to merge normal vectors obtained from photometric stereo with position data such as from structured light. This involves a Markov random field with a random variable for each pixel. The probability density function for each variable was then continuously refined by adding information from the normal vectors to them, causing them to converge to a combination of the low-frequency data from the range image and the local information from the normal vectors. In Section 7.2, the results of the scanning comparisons showed that adding photometric stereo data to the scan with the new technique improves the quality of the scan in all cases. The algorithm still runs in linear time, meaning that it scales well with increasingly higher resolution cameras. It can also be parallelised to a thread per pixel, making it a good candidate for running on a graphical processor. This completes the goals that have been set out for the scanning algorithm, since all requirements have been met.

8.1   Future Research

The calibration method has not been tested against other calibration techniques. No other implementations of calibrating light positions for photometric stereo were available, since most publicly available implementations seem to calibrate only the direction of the light or produce a look-up table to map intensity to incident angles. No time was available to make a custom implementation of a proven technique. A follow-up of this idea should test against such implementations, if these are made available in the future.

To make the calibration method feasible, the first order of business is to reduce or eliminate the proliferation of errors throughout the process. One option is to find the positions of the spheres using optical markers, such as QR codes, on the calibration statue or the spheres themselves.

The scanning technique as it stands now is completely ignorant of the actual range imaging technique used, as long as it produces little low-frequency noise. To further test this technique, it could be used to improve the scans of other range imaging techniques as well, such as stereo triangulation. This could produce different types of noise and have different results. The technique can be tested further by implementing more sophisticated photometric stereo techniques to allow scanning of different types of surfaces, as long as the accompanying range imaging technique is also able to scan these surfaces, however inaccurately.

The fusion technique makes some unrealistic assumptions to allow information to be combined efficiently with Kalman filters. Higher accuracy can be

achieved by switching to a more complex model. Notably, a particle filter is a similar estimator which is also linear, but allows for greater flexibility and does not converge to a local optimum. The current belief propagation depends on convergence of the probability density functions via the natural reduction of the variance of the normal distributions, so some new idea is required to ensure that the algorithm converges.

Since this algorithm scales well to high-resolution imagery and the camera is stationary, a further improvement to this technique is to use super-resolution [PPK03]. This could greatly increase the quality of scans at virtually no costs other than the difficulty of implementation, and the method discussed in this thesis is an excellent candidate to handle those amounts of data. It remains up to future research to discover the role of the rotating platform in this.

8.2   Personal Impression

While conducting this research, the fusion algorithm changed drastically numerous times and with each iteration the results became more promising. While the calibration technique was not successful, the scanning technique shows great promise. Furthermore, the implementation of the fusion technique is fairly simple and its source code is made available online[1]. This also includes an implementation of standard photometric stereo, of which few working implementations have been made public. I hope the technique will be of use to many future ideas.

Glossary

3D scanner

>   A device that measures the geometry of a 3-dimensional scene. 2–4, 6–8, 34, 35

albedo

>   A property of surfaces which specifies how much light is reflected in each direction. 4, 5, 11, 26, 27, 36, 37

anisotropic

>   Literally "not the same in every direction", but in computer vision the property that a surface may reflect light coming from different directions differently (e.g. along the grain of wood or perpendicular to it). 4

belief propagation

>   An algorithm for performing inference on Markov random fields that works by passing on beliefs between different nodes of the model. 5, 6, 32

blooming

>   The phenomenon in digital cameras where a very intense signal bleeds over to neighbouring pixels. 10

---

[1]Currently the repository is private, but it will be made available on `https://github.com/Ultimaker`

Brown-Conrady model

A model for distortion of a camera lens that accounts for radial distortion and tangential distortion. This accounts for lenses that are not perfectly centred. 10, 11

Cook-Torrance model

A model of surface reflectance properties where the surface consists of specular microfacets, which accounts for colour shifting. 4

depth from motion

A method to measure depth by measuring optical flow in a video to estimate the parallax in the scene. 4

diffuse

A property of surfaces that the surface reflects incoming light equally in all directions. 4, 5, 25

dilate

An operation on binary images, where every pixel within a certain range of a TRUE-marked pixel is also marked true. The opposite of erode. 29

directional light

A light source, like the sun, which illuminates the scene from a specific direction (rather than a specific position), such that all light rays are parallel. 4, 15, 43

expected value

The theoretical average value of a random variable if it were tested an infinite amount of times. 32

fast Fourier transform

An algorithm to compute the discrete Fourier transform, which transforms the input data into a weighted sum of a list of base functions, usually sine waves (because sine waves will model the explicit frequencies present in the data). 5

Gaussian blur

A filter on images that blurs the image with a 2D Gaussian function, computing a weighted average of pixels in the neighbourhood by applying the Gaussian function on the distance of that pixel to determine its weight. 28

Gaussian function

A function of the form $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, which represents a normal distribution with mean μ and standard deviation σ. 6, 30, 32

Gaussian Markov random field

A Markov random field where the random variables are all related with normal distributions. 6

graph cut

An operation on graphs that breaks the graph in two. One commonly used graph cut is the minimum cut that breaks a minimum number of edges, which has some desirable properties in optimisation algorithms. 28

high-frequency noise

A type of noise that consists of short deviations from the signal, while the baseline of the signal remains intact. In the frequency spectrum of the signal, only the high frequencies are wrong. 28, 29

hill-climbing

A type of optimisation that, starting at some initial state, searches nearby states and keeps climbing to the best of these states, until no better state can be found in the local neighbourhood. 19, 22, 24

identically distributed

Two random variables are identically distributed if they draw from the same probability distribution. 22

independent

Two events are independent from one another if the occurrence of the first does not affect the probability of occurrence of the other. This is the true if and only if their joint probability equals the product of their separate probabilities: $P(A \cap B) = P(A)P(B)$. 22

interreflection

Diffuse reflection on a surface illuminating other parts of the surface. 5, 28

inverse-square law

The inverse-square law describes (among other things) the reduction of light intensity with increasing distance if the light is evenly radiated in all directions. Since the same energy gets spread out over the surface of a sphere, and the surface of a sphere increases quadratically with increasing radius, the intensity is reduced quadratically. 26

Iterative Closest Point

An algorithm to minimise the distance between two point clouds, by iteratively performing linear operations until the mean squared error is minimised. 37, 41

Kalman filter

An algorithm that combines normally distributed measurements to a new estimate by multiplying their probability distributions. 2, 6, 29, 43

Lambertian model

A simple model of surface reflectance that specifies that the surface is perfectly diffuse. 5, 13, 25, 26, 28, 36

Law of Large Numbers

A theorem of probability theory that the average results of a large number of experiments should approach the expected value, or in terms of noise reduction, should approach the signal. 22

least squares

A technique that minimises the squared error in the data by a specified model, by solving a linear system of equations. 5, 6, 20, 21

LED

Light-emitting diode, a semiconductor light source which is often very small and very energy-efficient. 6, 7, 10, 25

MAP

Maximum A-Posteriori. 2, 6, 32

Markov random field

A graphical model, similar to a Bayesian network, but undirected and cyclic, which models a graph of probabilistic variables that may be dependent on each other. 2, 3, 6, 32, 43

Moiré pattern

A low-frequency pattern that is the result of super-imposing multiple high-frequency patterns or rounding a high-frequency signal to regular intervals. 41

Moore-Penrose Pseudoinverse

A common pseudoinverse of a matrix such that if A^+ is the Moore-Penrose Pseudoinverse, then $AA^+A = A$. It is computed as $A^+ = (A^T A)^{-1} A^T$. 28

nodal equations

A class of equations that models the deviation from a base function by fitting elastic curves to the base functions and measuring the energy required to stretch them to the data. 5

normal distribution

Also called the bell curve, a normal distribution is a probability distribution which represents the probability of random variables that are independently drawn from independent distributions.. 6, 10, 30–32, 44

normal vector

The vector that is perpendicular to the surface, pointing outwards from an object. 4–6, 13, 20, 25–28, 30–32, 41, 43

offset-error problem

A problem where small errors are allowed proliferate, affecting large segments of the output. These errors then stack up, causing the resulting scan to deviate further and further as errors are summed. 2

penumbra

A segment of shadow that is partially illuminated, because the light source is only partially occluded. 29, 41

photometric stereo

A method to measure the normal vectors of a surface by shining lights on it from different angles and comparing the amount of light that is reflected from each light. 2–7, 9, 10, 13, 15, 25–30, 32, 35, 36, 38, 42–44

point light

A light source that is modelled as if all light originates from one point in three-dimensional space. 15, 24, 29, 43

Poisson

A collection of techniques that involves solving Poisson's equation, $\Delta\,\phi\,=\,f$, for ϕ (where Δ is the Laplace operator). 5, 6

precision matrix

The inverse of the covariance matrix, which describes the correlation between random variables. 6

pre-knee circuit

A circuit in many modern cameras that captures a greater dynamic range by compressing the high end of the signal. 9

probability density function

A function mapping the values of a continuous random variable to a probability density. The value with the highest probability density is the most likely outcome of the random variable. 29–32, 43, 44

projection

An operation on vectors, mapping the vectors to a subspace. In the context of geometry, this is usually a projection of 3-dimensional space to 2-dimensional space such that difference in distance between every vector and its projection is minimised. 15, 17, 28

RANSAC

RANdom SAmple Consensus, a technique to filter noise by selecting a subset of the samples which agree on the model to select. 29

shape from shading

A method to measure depth by looking at how light is reflected on the object. 4, 5

slope-intersect form

A standard way of representing lines as an equation involving the slope of the line and a point on the line. 17–19

specular

A mirror-like reflection of light on a surface, where the angle of the incoming light ray to the surface normal is approximately equal to the angle of the outgoing light ray. 4, 19

standard form

A standard way of representing a geometric shape as an equation involving the coordinates, where the shape is precisely the set of coordinates where the equation holds. 17–19

stereo triangulation

A method to measure depth by matching parts of pictures from different viewpoints and measuring their disparity. 4, 43

structured light

A method to measure depth by projecting patterns of light on the surface and measuring how these patterns are deformed. 2–4, 7, 9, 10, 25, 27–30, 32, 33, 35, 38, 41–43

subsurface scattering

The phenomenon of light penetrating the surface, scattering diffusely below the surface and exiting the surface at another position. 5

super-resolution

A technique that captures multiple shots of the same scene, and from minute differences in the alignment of pixels manages to produce a single image with much higher resolution than the originals. 44

textured

Having different reflectance properties on every point on the surface. 4

time-of-flight

A method to measure depth by measuring how long light or sound takes to reflect back from the surface. 4

Torrance-Sparrow model

A model of surface reflectance properties where the surface consists of distributions of perfectly specular microfacets. 4

References

[AAT81]   Martin D. Altschuler, Bruce R. Altschuler, and John M. Taboada. Laser electro-optic system for rapid three-dimensional (3-d) topographic mapping of surfaces. *Optical Engineering*, 20(6), December 1981.

[ACR05]   Amit Agrawal, Rama Chellappa, and Ramesh Raskar. An algebraic approach to surface reconstruction from gradient fields. In *Computer Vision, Tenth IEEE International Conference on*, volume 1 of *ICCV 2005*, pages 174–181. IEEE, 2005.

[ARC06]   Amit Agrawal, Ramesh Raskar, and Rama Chellappa. What is the range of surface reconstructions from a gradient field? In *Computer Vision*, ECCV 2006, pages 578–591. Springer, 2006.

[ATD⁺08]   Naveed Ahmed, Christian Theobalt, Petar Dobrev, Hans-Peter Seidel, and Sebastian Thrun. Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry. In *Computer Vision and Pattern Recognition, 2008 IEEE Conference on*, pages 1–8. IEEE, 2008.

[Bos10]   Frédéric Bosché. Automated recognition of 3d cad model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. *Advanced Engineering Informatics*, 24(1):107–118, 2010.

[BP03]   Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1239–1252, 2003.

[Bro03]   Paul A. Bromiley. Products and convolutions of gaussian distributions. Technical Report 2003-003, Imaging Sciences Research Group, Institute of Population Health, School of Medicine, University of Manchester, 2003.

[BZK85]   Andrew Blake, Andrew Zisserman, and Greg Knowles. Surface descriptions from stereo and shading. *Image and Vision Computing*, 3(4):183–191, 1985.

[CAK07]   Manmohan Chandraker, Sameer Agarwal, and David Kriegman. Shadowcuts: Photometric stereo with shadows. In *Computer Vision and Pattern Recognition, 2007, IEEE Conference on*, pages 1–8. IEEE, 2007.

[CKC03]   Chia-Yen Chen, Reinhard Klette, and Chi-Fa Chen. 3d reconstruction using shape from photometric stereo and contours. Technical report, CITR, The University of Auckland, New Zealand, 2003.

[FC88]   Robert T. Frankot and Rama Chellappa. A method for enforcing integrability in shape from shading algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):439–451, 1988.

[FH06]   Roberto Fraile and Edwin R. Hancock. Combinatorial surface integration. In *Pattern Recognition, 18th International Conference on*, volume 1 of *ICPR 2006*, pages 59–62. IEEE, 2006.

[FL94]   Pascal V. Fua and Yvan G. Leclerc. Using 3-dimensional meshes to combine image-based and geometry constraints. Technical Report 536, SRI International, 1994.

[GCHS10]   Dan B. Goldman, Brian Curless, Aaron Hertzmann, and Steven M. Seitz. Shape and spatially-varying brdfs from photometric stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(6):1060–1071, 2010.

[HB88]   Gleen Healey and Thomas O. Binford. Local shape from specularity. *Computer Vision, Graphics and Image Processing*, 42(1):62–86, 1988.

[HCG14]   Sk. Mohammadul Haque, Avishek Chatterjee, and Venu Madhav Govindu. High quality photometric reconstruction using a depth camera. In *Computer Vision and Pattern Recognition, 2014 IEEE Conference on*, pages 2283–2290. IEEE, 2014.

[HLHZ08]   Michael Holroyd, Jason Lawrence, Greg Humphreys, and Todd Zickler. A photometric approach for estimating normals and tangents. In *ACM SIGGRAPH Asia 2008 Papers*, SIGGRAPH Asia 2008, pages 133:1–133:9. ACM, 2008.

[Hor70]   Berthold K.P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical Report 232, MIT Artificial Intelligence Laboratory, 1970.

[Hor90]   Berthold K.P. Horn. Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75, 1990.

[HR47]   Pao-Lu Hsu and Herbert Robbins. Complete convergence and the law of large numbers. *Proceedings of the National Academy of Sciences of the United States of America*, 33(2):25–31, 1947.

[HVC08]   Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Multi-view photometric stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):548–554, 2008.

[Ike81]   Katsushi Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-3(6):661–669, November 1981.

[KKOF04]   Rieko Kadobayashi, Nobuo Kochi, Hitoshi Otani, and Ryo Furukawa. Comparison and evaluation of laser scanning and photogrammetry and their combined use for digital recording of cultural heritage. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35(5):401–406, 2004.

[KKS98]   Reinhard Klette, Ryszard Kozera, and Karsten Schlüns. Shape from shading and photometric stereo methods. Technical Report CITR-TR-20, The University of Auckland, 1998.

[Kov05]   Peter Kovesi. Shapelets correlated with surface normals produce surfaces. In *Computer Vision, Tenth IEEE International Conference on*, volume 2, pages 994–1001. IEEE, 2005.

[LHY11]   Miao Liao, Xinyu Huang, and Ruigang Yang. Interreflection removal for photometric stereo by using spectrum-dependent albedo. In *Computer Vision and Pattern Recognition, 2011 IEEE Conference on*, pages 689–696. IEEE, 2011.

[LJ77]   R.A. Lewis and A.R. Johnston. A scanning laser rangefinder for a robotic vehicle. Technical Report 33-809, California Institute of Technology, 1977.

[MCC+03]   Michael I. Miga, David M. Cash, Zhujiang Cao, Robert L. Galloway Jr., Benoit Mr. Dawant, and William C. Chapman. Intraoperative registration of the liver for image-guided surgery using laser range scanning and deformable models. In *Medical Imaging*, pages 350–359. International Society for Optics and Photonics, 2003.

[MP79]   David Marr and Tomaso Poggio. A computational theory of human stereo vision. In *Proceedings of the Royal Society of London Series B: Biological Sciences*, volume 204, pages 301–328. The Royal Society, 1979.

[NIK91]   Shree K. Nayar, Katsushi Ikeuchi, and Takeo Kanade. Shape from interreflections. *International Journal of Computer Vision*, 6(3):173–195, 1991.

[NRDR05]   Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM Transactions on Graphics*, 24(3):536–543, 2005.

[PB01]   Maria Petrou and Svetlana Barsky. Shadows and highlights detection in 4-source colour photometric stereo. In *Image Processing, Proceedings of the 2001 International Conference on*, volume 3, pages 967–970. IEEE, 2001.

[PP93]   Frank L. Pedrotti and Leno S. Pedrotti. *Introduction to Optics.* Prentice-Hall International, Inc., 1993.

[PPK03]   Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: A technical overview. *Signal Processing Magazine*, 20(3):21–36, 2003.

[RH05]   Håvard Rue and Leonhard Held. *Gaussian Markov Random Fields: Theory and Applications.* Monographs on Statistics and Applied Probability 104. Chapman & Hall / CRC Press, 2005.

[Rin66]   Thomas C. Rindfleisch. Photometric method for lunar topography. *Photogrammetric Engineering*, 32:262–277, 1966.

[SC04]   Øystein Skotheim and Fred Couweleers. Structured light projection for accurate 3d shape determination. In *12th International Conference on Experimental Mechanics, Bari, Italy*, volume 29, 2004.

[SSBS04]   Sinal Shah, Geeta Sundaram, David Bartlett, and Martyn Sherriff. The use of a 3d laser scanner using superimpositional software to assess the accuracy of impression techniques. *Journal of Dentistry*, 32(8):653–658, 2004.

[SSL⁺11]   Rafael F.V. Saracchini, Jorge Stolfi, Helena C.G. Leitão, Gary Atkinson, and Melvyn L. Smith. Multi-scale integration of slope data on an irregular mesh. Technical Report IC-11-11, Instituto de Computação, Universidade Estadual de Campinas, 2011.

[SW13]   Dirk Schnieders and Kwan-Yee K. Wong. Camera and light calibration from reflections on a sphere. *Computer Vision and Image Understanding*, 117(10):1536–1547, 2013.

[SZP10]   Kalyan Sunkavalli, Todd Zickler, and Hanspeter Pfister. Visibility subspaces: Uncalibrated photometric stereo with shadows. In *Computer Vision, 2010 European Conference on*, pages 251–264. Springer, 2010.

[Td91]   Hemant D. Tagare and Rui J.P. deFigueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):133–152, 1991.

[Ter88]   Demetri Terzopoulos. The computation of visible-surface representations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):417–438, 1988.

[Ull79]   Shimon Ullman. *The Interpretation of Visual Motion*. Massachusetts Institute of Technology, 1979.

[VMC97]   Tamás Várady, Ralph R. Martin, and Jordan Cox. Reverse engineering of geometric models – an introduction. *Computer-Aided Design*, 29(4):255–268, 1997.

[WLDW11] Chenglei Wu, Yebin Liu, Qionghai Dai, and Bennett Wilburn. Fusing multiview and photometric stereo for 3d reconstruction under uncalibrated illumination. *Visualization and Computer Graphics, IEEE Transactions on*, 17(8):1082–1095, 2011.

[Woo79]   Robert J. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *22nd Annual Technical Symposium*, pages 136–143. International Society for Optics and Photonics, 1979.

[Woo80]   Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[YSL10]   Chanki Yu, Yongduek Seo, and Sang Wook Lee. Global optimization for estimating a brdf with multiple specular lobes. In *Computer Vision and Pattern Recognition, 2010 IEEE Conference on*, pages 319–326. IEEE, 2010.

[ZWYD08] Jiejie Zhu, Liang Wang, Ruigang Yang, and James Davis. Fusion of time-of-flight depth and stereo for high accuracy depth maps. In *Computer Vision and Pattern Recognition, 2008 IEEE Conference on*, pages 1–8. IEEE, 2008.

[ZYY+12] Quing Zhang, Mao Ye, Ruigang Yang, Yasuyuki Matsushita, Bennett Wilburn, and Huimin Yu. Edge-preserving photometric stereo via depth fusion. In *Computer Vision and Pattern Recognition, 2012 IEEE Conference on*, pages 2471–2479. IEEE, 2012.