Shape from colored structured light and polarization

Jaime van Kessel 3855856
Supervisors:
Dr. Robby T. Tan (Utrecht University)
Erik de Bruijn MSc. (Ultimaker B.V.)

November 12, 2013

Acknowledgement

This master incorporates results from a project done in cooperation with Utrecht University and Ultimaker. Ultimaker is the developer of the Ultimaker Original and Ultimaker 2 3D printer. The project is carried out at Ultimakers' headquarter located in Geldermalsen, The Netherlands.

I would like to thank my supervisors Erik for giving me the oppertunity to realize my ideas at Ultimaker and Robby for doing the same at the University. Even though their specialisations did not always lie exactly in the direction of this research, their insights and support have proven to be invaluable.

My thanks go out to my colleagues at the research department of Ultimaker. David, Coen, Harma, Menno and Marvin, thanks for being a listening to my ideas and providing inspiration from your fields of experience.

Ultimaker

Universiteit Utrecht

Abstract

In this thesis data from two separate methods, *shape from diffuse polarization* and *rapid shape acquisition using color structured light* are combined into a single point cloud. Several improvements to the structured light method are proposed which decrease the negative impact of textured surfaces and enable sub-pixel accuracy. The improvements of the structured light method are tested by comparison with a ground truth model. The results fall within a 0.8 mm standard deviation of the ground truth, with outliers to 0,6 mm accuracy for some models when all improvements are used.

The accuracy of the polarized light method is also tested in several situations. The accuracy of this method increases in areas that are almost parallel to the light direction, leading up to 30% less noise when compared to areas where the light is perpendicular to the light direction.

The combination of the methods results in a ten times higher resolution pointcloud than would be possible with only the rapid shape acquisition. Unfortunately, there is a significant increase in noise as compared to the unmodified method, because if one of the methods provides incorrect data, the combination propagates these mistakes. Better results should be obtainable with a different gradient to depth algorithm and multiple viewpoints for the polarised light. The implemented system is modular, enabling easy modification and switching of methods.

Contents

Chapter 1

Introduction

1.1 Introduction

Technological developments are often inspired by solutions found in nature. A considerable number of high tech solutions are based upon biological phenomena or derived from observations of certain animals and plants. Seeing that animals and humans alike are experts at recognizing shapes in 3d space, it becomes a logical step to look at how this ability is achieved. Humans combine several depth cues in order to achieve this reconstruction. The most accurate of these cues as used by the human vision is the stereoscopic effect. By analyzing which points are the same in both images the depth can be triangulated. Due to its accuracy and its simple basic principle, this is also the process which drives the majority of 3D scanning algorithms.

Using multiple image sources in the form of cameras in order to obtain depth information, even though this is most similar to human perception, is difficult due to problems matching corresponding parts in the images. These difficulties are caused by the different lighting conditions, occlusions, lack of matchable features and differences between the devices used to capture the image. However, we only need to take inspiration from how the problem is solved in nature. A computer algorithm does not need to constrain itself to using similar methods as human perception. In order to simplify the process of matching the two images, one of the two cameras can be removed and replaced by a lightsource.

Much of the principles of a (digital) camera can be found in a projector, such as the usage of lenses and image planes. The main difference between these two is that the projector projects light instead of capturing it. A projector can therefore be treated as the inverse of a camera. This has one notable advantage in the usage of scanning applications, as it allows the complete control of at least one of the observed images. The images used for the projector are the so called patterns, which can vary greatly between structured light this kind of methods. Subsequently, the projected image will also influence the scene, which can be used to create easy to match feature points. The same process of matching both images or image points that is used for normal stereographic vision, in order to triangulate depth, can be used. This area of 3D scanning is referred to as structured light (SL).

The usage of different patterns has been explored intensively and so has using multiple images

for a single depth estimation. The pattern types that are used can be subdivided into 1D patterns; such as line lasers or phase shift patterns [14] and 2D patterns; such as used by the Microsoft Kinect [13] and m-arrays[10]. In all methods there is a tradeoff between accuracy, number of measured points, computational time, number of images required and handling of sub-surface reflection and or inter-reflection[1].

Aside from using triangulation other cues to obtain depth information can be used. Human vision, for instance, does not only rely on the disparity between the two eyes, but also combines this with the usage of shading (photometric stereo), vagueness (depth from focus) and parallax (depth from motion) in order to obtain depth information.

Another potential source of information can be found in polarized light. It is believed that some insects already use the polarization of light as an aditional source of information [23]. The polarization of light changes when it is reflected from a surface and its change is dependent on the material of the surface and its orientation. Once the type of surface is known, or an assumption is made how it reflects light, the surface orientation can be determined. This method, shape from polarization, does not directly result in depth information. A second step is required, wherein the surface orientation is converted into relative depth. Because the polarization angle can only be measured in the range of [0,180] degrees, there is an ambiguity in the estimated surface normal.

Several solutions have been suggested in academic literature to solve this ambiguity, such as using multiple views [3], combining information from shape from shading [24] and smoothness constraints [2].

There has been some research on shape from polarization for 3D scanning [4] and polarization in the field of robotics [5], but compared to methods such as photometric stereo and structured light, little research is being performed in this field.

1.2 Rationale

The last few years have been marked by significant technological progress in the field of consumer 3D printing [20]. These printers have improved in print quality, reduced in cost and can print larger objects. As a direct result of the technological improvements and greatly reduced costs, the 3D printer has shifted from being used exclusively workshops and research laboratories to being used by tech savvy early adopters. The recently released 3D printers by Ultimaker [21] and Makerbot [22] indicate that a further shift towards mainstream consumers is in progress.

In order to use these printers a 3D model is required. To enable full replication of objects a method to obtain such a model or at the very least a 3D point cloud[2] from a physical model is needed.

We've specifically opted to combine two areas of research; structured light (SL) and shape from polarization (SFP). Because of the fundamentally different nature of both methods a framework needs to be constructed that utilizes both methods. Ideally, this framework is a

[1] Light reflected by the object onto another part of the object
[2] A set of data points in some coordinate system.

modular system. This modularity enables the switching of one or both of the modules with a method that results in the same type of data, such as switching the shape from polarization with photometric stereo. This flexibility is especially important for sustainable open source development, as each module can be handled as a separate smaller project. As these projects are smaller, they should also be easier to resolve, as less specialized in-depth knowledge is required. The development of packages such as PCL and openCV have proven that open source development can increase the speed of complex software. The modularity will also prove important for further research, as the situations for which the scanning is performed will vary and can benefit from different methods.

The depth information obtained from structured light can be used to assist in the ambiguity solving of the surface normals, as even from a sparse point cloud it is possible to estimate rough surface normals.

1.3 Overview

Our work is based on two methods; the structured light method as proposed by Zhang et al.[1] and the shape from polarization method as proposed by Miyazaki et al.[2]. In its most basic form, the method of Zhang et al. employs a 1D color coded pattern which requires only a single image. A more sophisticated method is also provided by Zhang which has increased accuracy but requires multiple images and has a significantly increased running time. Because the method uses edges of the 1D pattern to match the camera and projector image, the resulting point cloud is rather sparse. Increasing the resolution of the projected pattern or the captured image can be achieved by using the more sophisticated method or by projecting multiple shifted patterns.

The method of Miyazaki uses a rotating linear polarization filter mounted in front of the camera. By taking three images of the object under the same lighting conditions where the filter is rotated by 45 degrees per image, polarization data is collected. Based on the intensities observed in the three images the rotation of the polarization of the reflected light can be interpolated. This resulting degree of polarization (DOP) can be used to calculate surface normals, up to an ambiguity of 180 degrees. In order to obtain depth information, the integrability assumtion is made, so the gradient field can be converted to a depth map.

The impact of combining SL and DOP are twofold. First, we expect that the impact of the noise is diminished, due to the usage of the SL data as anchor points. Secondly, we expect that less assumptions are required. Most notably, the continuity assumption that is used in the reconstruction of depth from gradient fields can be lessened. Because a far greater number of points can be retrieved from the shape from polarization, the overall resolution of the point cloud will be greatly increased.

1.4 Contributions

We have implemented both methods into a single framework that is capable of merging the results into a single point cloud. Several experiments have been performed to measure the

accuracy of both methods.

- Method to solve normal ambiguity.
- Simple calibration for camera / projector setup.
- Improved the results of rapid shape acquisition.
- Method to merge the two datasets.
- Experiments on the accuracy of shape from polarization and structured light.

1.5 Organization

This thesis is organized in several chapters; In chapter 2 we discuss the related work and provide general background information required for understanding and implementing both methods. Chapter 3 explains the methods as used by the main reference papers and states the drawbacks and advantages of both methods. At the end of the chapter we explain several improvements to each method and explain how both methods are integrated into one solution. Chapter 4 gives an evaluation of the accuracy of the methods and improvements and in chapter 5 several conclusions are drawn and recommendations for further research are made.

Chapter 2

Related work & background

Some background knowledge on polarization and structured light is required before the methods from both reference papers and improvements to these can be discussed. Section 2.1 provides background information into (digital) cameras. In section 2.3 the basic principles of structured light are provided and finally in 2.2 an overview of polarization of light and how it changes upon reflection is given.

2.1 Camera

Both methods use cameras, either to measure the change in intensities in the case of shape from polarization or to observed the projected pattern in the case of structured light. As cameras are the workhorse of most computer vision algorithms the mathematical models to represent them are quite sophisticated. The Pinhole model is covered in section 2.1.1, which is used to describe both the used cameras and the projector. The standard calibration of pinhole devices is explained in 2.1.2. Due to the importance of color for the used structured light method, the color perception of CCD sensors is explained in 2.1.3.

2.1.1 Pinhole model

The pinhole model is a simplification of a model for a contemporary camera. The definition as used by openCV [9] is used for each point in the scene, one ray enters the camera through a single point (the pinhole). This light is then projected onto an image plane behind the pinhole. An example of such a scene can be seen in figure 2.1. The distance of the pinhole to the image plane is the focal length (f_x and f_y), where the center of the image plane is expressed by c_x and c_y. These parameters form the intrinsic matrix A.

There are two focal length parameters (one for each axis) as in most cases the pixels on the ccd sensor and thus the image plane are not perfectly square. The principal point formed by c_x and c_y is introduced as the ccd sensor is not always perfectly aligned with the optical axis.

In practice, lenses will also have a certain distortion, which is not included in the intrinsic matrix. In order to account for these distortions a distortion model is included. The two

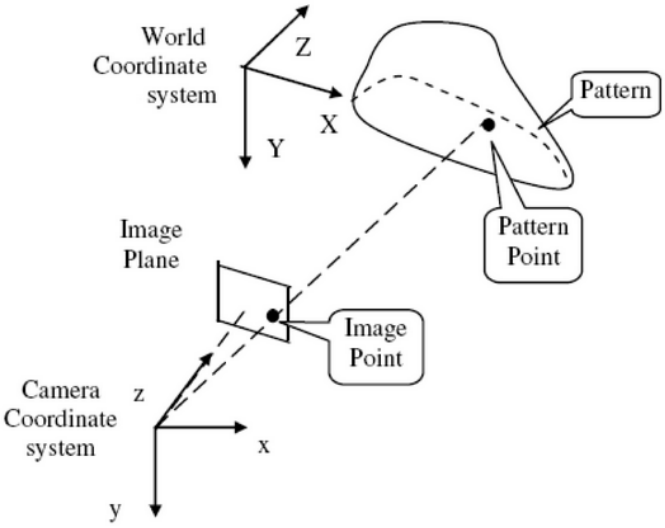

Figure 2.1: Example of single captured chessboard pattern..

types of distortion which are non-negligible are radial distortions (barrel and fisheye), which are cause by the (imperfect) shape of the lens and tangential distortion, which is caused by the lens not being exactly parallel to the image plane.

The radial distortion can be corrected by using

$$\begin{aligned} x_{corrected} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)\,, \\ y_{corrected} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)\,. \end{aligned} \tag{2.1}$$

The tangential distortion can be corrected by

$$\begin{aligned} x_{corrected} &= x + [2p_1 xy + p_2(r^2 + 2x^2)]\,, \\ y_{corrected} &= y + [p_1(r^2 + 2y^2) + 2p_1 xy]\,. \end{aligned} \tag{2.2}$$

Where x and y are the original coordinates in the input image and $x_{corrected}$ and $y_{corrected}$ are the undistorted coordinates. This means that by using 5 parameters (k_1, k_2, p_1, p_2 and k_3) we can account for these distortions.

The intrinsic matrix, combined with the distortion coefficients convert a 2D image point into a 3D coordinate on the image plane of the camera. In order to project these coordinates to world coordinates an extrinsic matrix is needed. This intrinsic matrix holds a 3x3 rotation matrix and a 3x1 translation matrix.

The total conversion from image points to world coordinates is given by:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r23 & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2.3}$$

Where:

- (X, Y, Z) are the coordinates of a 3D point in the world coordinate space.
- (u, v) are the coordinates of the projection point in pixels. These are the image coordinates
- (c_x, c_y) is the principal point that is usually at or near the image center.
- (f_x, f_y) are the focal lengths expressed in pixel units.
- $(R_{11}, \ldots, R_{33})$ form the extrinsic rotation matrix.
- (t_1, t_2, t_3) form the extrinsic translation vector.

2.1.2 Calibration

OpenCV provides a simple to use calibration method based on the work of Zhang et al. [8]. This method uses multiple images of a printed checkerboard (or other pattern with several easily distinguishable points). In order to obtain an as accurate as possible calibration, its required to take as many calibration images as possible, each with different position and rotation. For most uses it is sufficient to use 10-20 patterns. An example of a detected chessboard pattern can be found in figure 2.2. A more in depth explanation of the calibration procedure as used by openCV can be found in the openCV book [9].

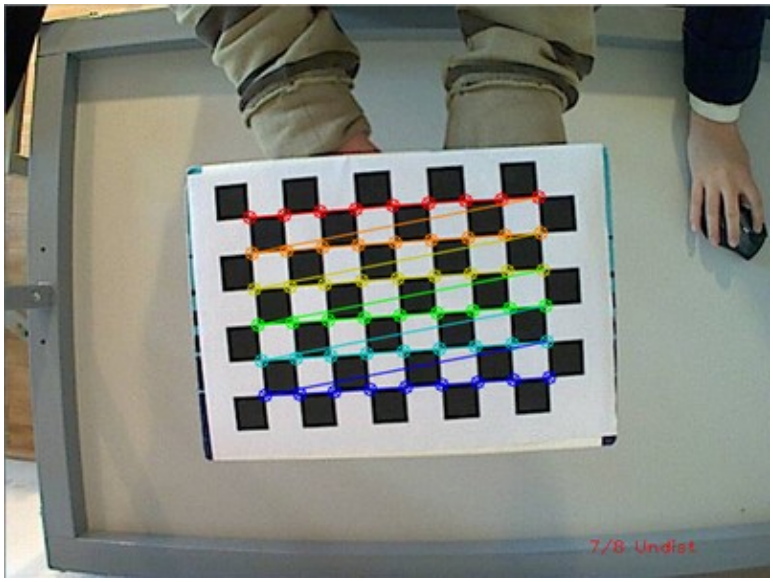

Figure 2.2: Schematic representation of the pinhole model.

2.1.3 Bayer pattern

Most color CCD chips use a mosaic color filter on a grid of photo sensors [12]. In theory, all but a single color should be filtered out by the colored filters. Each pixel has 4 of these

photo sensors, two with a green filter, one with a red filter and one blue filter. Because of this composition, the CCD chip is more sensitive to green light. This was implemented to better mimic the human vision, which is also more sensitive to green light. A schematic overview of this layout can be seen in figure 2.3.

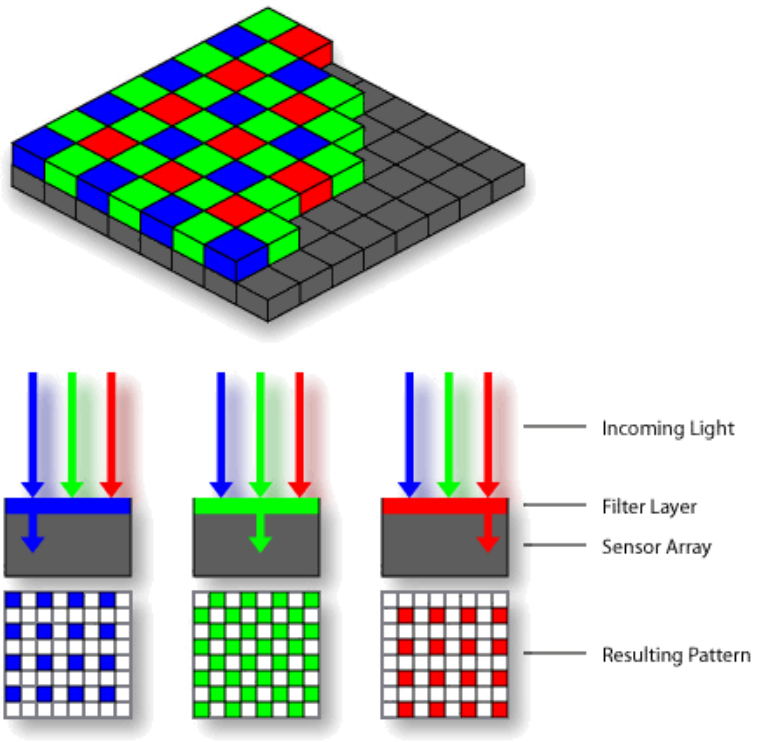

Figure 2.3: Ideal case Bayer pattern filter.

In practice, the used filters are far from perfect filters, especially with lower quality CCD sensors. Imperfect filters allow small amounts of differently colored light to pass. This effect is known as colorcrosstalk. Fortunately, the influence of this effect can be partially removed by a procedure aptly called color crosstalk removal, which is discussed more in chapter 4.

2.1.4 Gamma correction

In order to compensate for differences between human vision and CCD sensors, most images are gamma encoded. Instead of using a linear function to encode the perceived values of light an approximate power function is used. This function is more alike to the human vision, which ensures that more data that is perceivable for humans can be encoded while using the same amount of data. If an image is not gamma encoded, it needs to have more bits/bandwidth in order to maintain the same perceived image quality.

In the case of measuring the amount of reflected light, the calculated gamma encoded values are less suitable for our purposes, since specifically the original amounts of reflected light are of interest. As most cameras automatically convert images to the gamma corrected s-RGB[1]

[1]Standard RGB color space

format, the data needs to be converted back into the linear RGB space.

The s-RGB format is slightly different from simple gamma correction. Where standard gamma correction only uses a power to encode the intensities, s-RGB encodes the first few values linear before switching to a power. The effective difference between the two encoding schemes can be seen in figure 2.4.

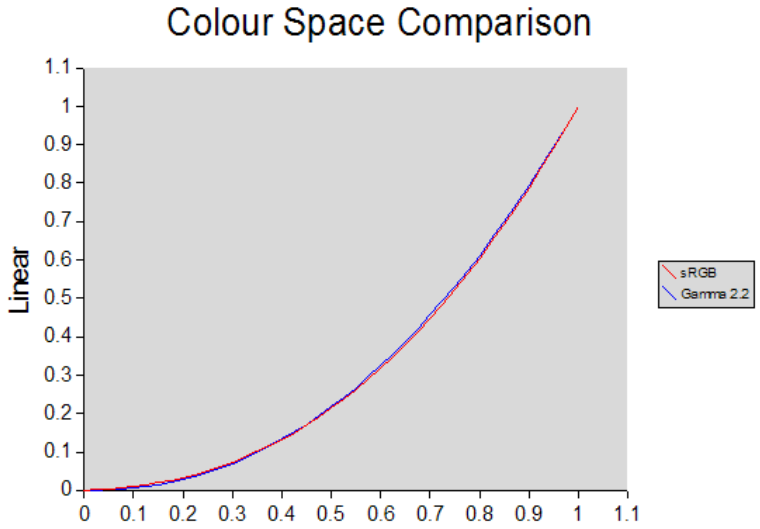

Figure 2.4: s-RGB versus gamma with normalized data.

The following pseudo code converts s-RGB into l-RGB:

function CONVERTTOLINEAR(int c)
  if $c \leq 0.03928$ **then**
    $lin \leftarrow c/12.92$
  else
    $lin \leftarrow (c + 0.055)^{2.4}$
  end if
return lin
end function

2.2 Polarized light

Polarization is a property of light and by extension of all electromagnetic radiation which refers to the orientation of the transverse electric field. We will provide a global overview of polarized light and explain the Fresnel theory, which is used to describe how light is polarized (eg; how its polarization *changes*) when it is reflected from a surface [11].

2.2.1 Introduction

There are three types of polarization; *Linear polarization*, *circular polarization* and *elliptical polarization*. When the wave of the electric field can be place on a single plane it is called

linear polarization. This linear polarization is expressed by a single rotation. If the plane rotates while the field travels, it is called rotational or elliptical polarization, depending on type of rotation. An example of these waves can be seen in figure 2.5.

The focus of this research will rest solely on linear polarization. Not only is this type of polarization easy to measure by utilizing a simple rotating linear polarization filter. All previous methods, such as those by Atkinson et al. [3] and Miyazaki et al. [2], only measure the linear polarization.

We assume that the used linear filters are ideal filters; only the component that is parallel and every rotation + 180 degrees, to the orientation of the filter is let through. Linear polarization can't distinguish between two angles 180 degrees apart, therefore, it is required to calculate both possible outcomes and resolve the resulting ambiguity in a later stage.

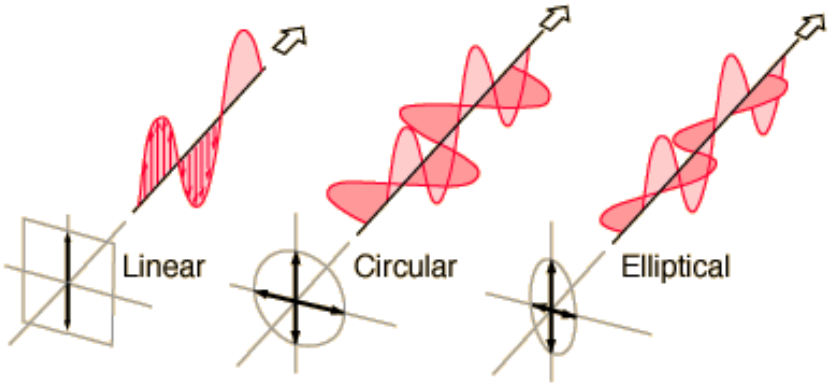

Figure 2.5: Different polarization states of light.

Most light sources will emit a large number of waves. When the orientation of their electric fields have no correlation, the light is said to be *unpolarized*. When a partial correlation can be found, the light is *partially polarized*. The degree of polarization (DOP) is used to differentiate between the two; The DOP is 0% for perfectly polarized light and 100% for perfectly polarized light. The DOP is calculated by using the total power of the polarized component of the wave.

2.2.2 Fresnel theory

The Fresnel equations give the ratios of the observed light compared to the projected light. This ratio is dependent on angle of incidence i and the refractive index n_t of the surface, as we know that the object is placed in air, we can use $n_i = 1$.

As can be seen in figure 2.6, when the light is polarized perpendicular to the surface, the resulting light will also be (somewhat) perpendicular. When the light is polarized parallel to the surface, the light will not be polarized perpendicular. The degree of polarization ρ is the amount of polarized light reflected from the surface. This degree of polarization is dependent on the angle of incidence. Because the light can always be resolved into two perpendicular components, the Fresnel equations can be used for all polarization states.

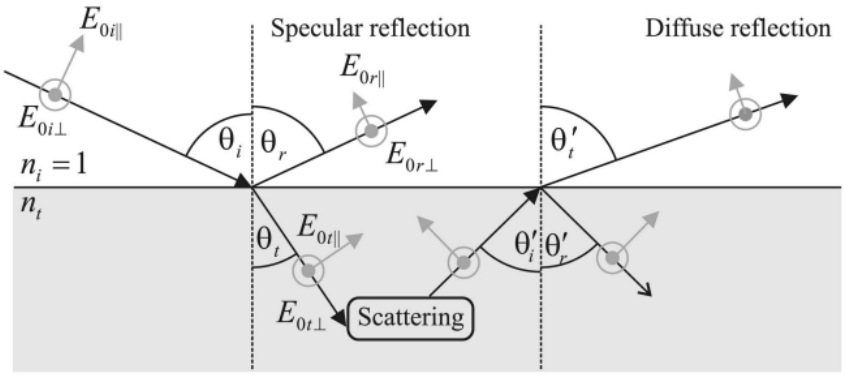

Figure 2.6: Refraction and scattering

The Fresnel equations, as applicable to figure 2.6, are given by the following two formulas;

$$
\begin{aligned}
r_{\perp}(n_i, n_t, \theta_i) &\equiv \frac{E_{0r\perp}}{E_{0i\perp}} = \frac{n_i cos\theta_i - n_t cos\theta_t}{n_i cos\theta_i + n_t cos\theta_t} \\
r_{\|}(n_i, n_t, \theta_i) &\equiv \frac{E_{0r\|}}{E_{0i\|}} = \frac{n_i cos\theta_i - n_t cos\theta_t}{n_i cos\theta_i + n_t cos\theta_t}
\end{aligned}
\tag{2.4}
$$

Where the first gives the reflection ratio for the light that is polarized perpendicular to the plane of incidence and the second for parallel to the plane.

The diffuse reflection needs less strictly controlled lighting conditions, but has the disadvantage of being noisier and lacking the ability to be used on glossy surfaces such as metal. In order to use specular reflection, the object needs to be placed in a enviorement where all light sources can be controlled. Diffuse reflection does not require such strict enviorement to be correctly meassured. As the the method is intended to work together with structured light, the lighting conditions need to be similar as much as possible. The used structured light method is not capable of working with very glossy objects, so losing this information will have little impact. In all cases where polarization or reflection is mentioned, diffuse reflection is meant.

2.3 Structured light

2.3.1 Introduction

One of the most commonly used techniques for 3D object scanning is the so called active stereoscopic vision. Where passive stereoscopic vision uses at least two capture devices and uses the disparity between the two images to calculate the depth of the scene, active stereoscopic vision uses at least one light source and one capture device. This allows for much simpler detection of correspondence between the two or more images, as the features needed to match the images are created in the scene. It also provides the added benefit of being able to modify the projected pattern based on the scene that is being observed.

Compared to passive stereoscopic vision, active stereoscopic vision has the option to project several images in sequence and use the combined data to compute a depth map. However, dependent on the capturing interval, it may not be feasible to use this method. When objects move and/or deform during scanning, assumptions about consistency will not hold, resulting in increased complexity in data processing. In these cases it would be impractical to use multiple images taken in sequence.

2.3.2 Projector calibration

Because the projector is unable to capture images as a camera, the calibration method as implemented by openCV is insufficient. However, several steps can be taken to create a dataset that can be used to calibrate the projector, as if it were a camera. For this to work, only the intrinsic parameters of the camera need to be known beforehand.

In order to calibrate the camera, a checkerboard pattern is projected on a physical checkerboard pattern. To simplify the extraction of the two checkerboards from the camera image, a red checkerboard is projected on a blue printed checkerboard. The corner points can then simply be extracted by looking at the red and blue color channels as obtained from the camera feed.

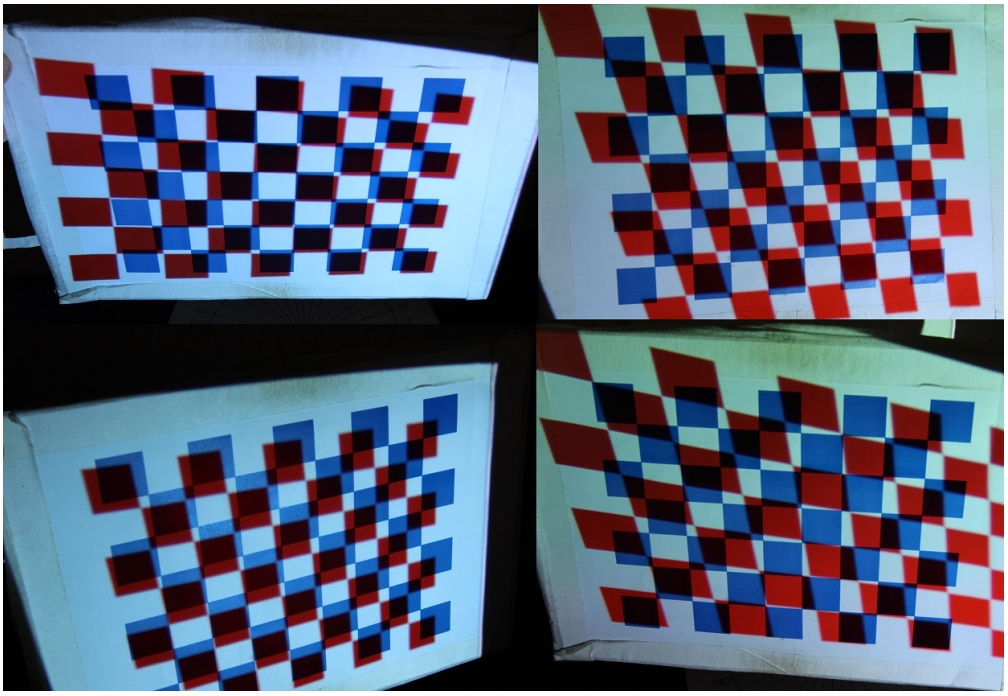

Figure 2.7: Red checkerboard pattern projected onto blue printed checkerboard.

Because the corner points of the red projected checkerboard are projected on a flat plane, they are transformed in a certain way. Examples of this effect can be found in figure 2.7.

If the projector was capable of taking images, the red checkerboard image would be untransformed (eg; exactly as it was projected). Because its known what the captured image

looks like and should look like, an perspective transformation can be performed on the image to convert the camera image into a projector image.

This perspective transformation results in image as if it were taken from the location of the projector. The (transformed) blue printed corner points can then be used for the standard calibration as implemented in openCV. Figure 2.8 displays an uncorrected image and a warped / transformed image.

This method is far less complex than comparable methods [19] and requires similar number of images to calibrate.

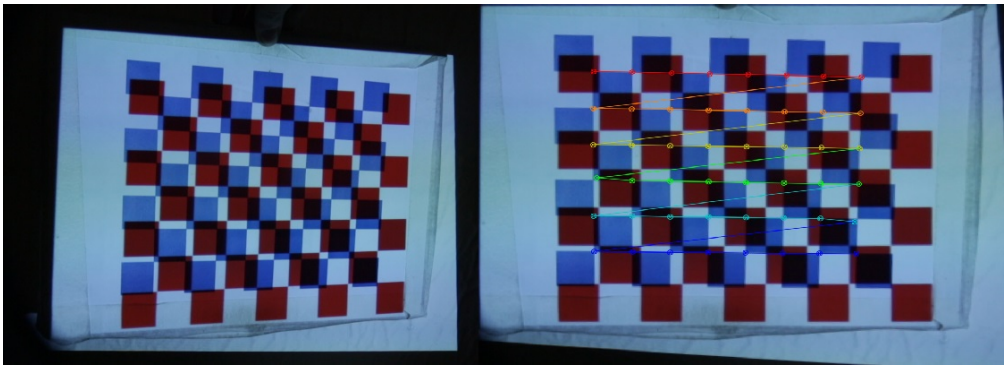

Figure 2.8: Original image and warped image. *Left:* Original image of printed checkerboard (blue) with projected checkerboard (red). *Right:* Image transformed so the red checkerboard corners are aligned. Detected (blue) corner points are drawn.

2.3.3 Optical triangulation

The main problem that is solved by structured light is the matching of two points on two different viewpoints. Once such a correspondence is found, they need to be triangulated in order to find a depth. The camera matches a single pixel, which can be seen as a vector from the center of the camera to the point on the image plane. Because the projector projects vetical stripes, we can not distinguish a single point that is matched. However these stripes can be seen as planes that are formed from the center of the projector and all image points of that vertical stripe. A least square fitting algorithm is used on the image points to find the planes for the projector. This method ensures that the possible distortion of the projector is accounted for. The intersection point between the projector plane and camera vector is the resulting point, which can be placed in the point cloud. The image points are generated using the pinhole model.

The projected vertical lines can be seen as planes projected from the center of the projector through all points.

Chapter 3

Theory

In this chapter the two reference papers *Rapid shape acquisition using color structured light and multi-pass dynamic programming* and *Polarization-based Inverse Rendering from a single View* are discussed. The methods from these papers are analyzed and improvements are proposed to the structured light method. Finally the combination of the data-sets is discussed and a method to handle this is proposed.

3.1 Rapid shape acquisition

3.1.1 DeBruijn pattern

The structured light system as described by Zhang et al. uses colored bands as a pattern. This allows for multiple edges to be detected from a single projected image, but introduces the problem of matching the right projected edge with the detected edges. By only using unique colors, only a handful edges are projected, but ensures that the matching is a trivial task.

If only globally unique edges are used, only a very limited resolution can be achieved, as fully saturated colors are needed to decrease the impact of the texture and color of the object. In order to increase this resolution, instead of unique colored stripes, a number of unique sets are projected.

Each set of n stripes needs to be unique in order to be identified. Such a pattern can be generated by using a deBruijn sequence, which ensures that every set of n stripes that can have k values per stripe that are unique for the entire pattern. Because the algorithm uses edges, an unique set of stripe transitions is required. The direct use of the deBruijn sequences does not guarantee this, so an additional step is needed.

Figure 3.1: Section of the generated de Bruijn pattern with $k = 5$ and $n = 3$.

In order to construct this pattern, a tuple consisting of 7 colors $\{001, 010, \ldots, 111\}$ is created, where a 0 in a channel means that RGB color is not in the stripe and a 1 means its at the maximum value. By starting at a certain color and performing an XOR operation on that color with another color a third color can be obtained. For example, an XOR operation of $\{1, 0, 0\}$ performed on $\{0, 1, 0\}$ (green) would result in $\{0, 1, 1\}$ (yellow).

The following formula is used to compute the final deBruijn projection pattern:

$$p_{j+1} = p_j \oplus d_j \tag{3.1}$$

Where p_j is the color of a single projector stripe and d_j is the color of the deBruijn sequence. This ensures that every subsequent stripe is different from its direct neighbours and still satisfies the local window uniqueness property.

3.1.2 Crosstalk removal

In an ideal world a certain ray of colored light would only agitate corresponding colored camera pixels. In practice the Bayer filter is not perfect and will let small amounts of differently colored light through. This phenomenon is known as color crosstalk. As we have no prior knowledge of the surface from which the light reflected from, which could have influenced the spectrum of the light its quite difficult to remove this crosstalk. Even if such information would be known, objects could have multiple types of surfaces which further complicate the removal.

Caspi et al [7]. have formed a model that relates the observed color s to the projected color p as;

$$\underset{A}{\begin{pmatrix} s_r \\ s_g \\ s_b \end{pmatrix}} = \underset{X}{\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}} \underset{F}{\begin{bmatrix} \rho_t & 0 & 0 \\ 0 & \rho_g & 0 \\ 0 & 0 & \rho_b \end{bmatrix}} \underset{p}{\begin{pmatrix} p_r \\ p_g \\ p_b \end{pmatrix}} + \underset{o}{\begin{pmatrix} o_r \\ o_g \\ o_b \end{pmatrix}} \tag{3.2}$$

Where X is the color crosstalk matrix, F is the scene albedo at a single point of the scene and O is the ambient light for the same point. By pre-multiplying the observed colors with the inverse of the crosstalk matrix, we can largely factor out the influence of the crosstalk. Even though this is not a perfect crosstalk removal, it moves the observed colors much closer to the actual colors of the projected image.

$$\tilde{\mathrm{s}} = X^{-1}s = \begin{pmatrix} \rho_r p_r + \tilde{\mathrm{o}}_r \\ \rho_g p_g + \tilde{\mathrm{o}}_g \\ \rho_b p_b + \tilde{\mathrm{o}}_b \end{pmatrix} \tag{3.3}$$

The color crosstalk matrix can be obtained by taking three images of a white plane on which red, green and blue light is projected. The averaged observed RGB values form the nine values of X. An example of an image with the crosstalk removal can be seen in figure 3.2 and 3.3. Note the far greater intensity of the colors, even if they are almost indistinguishable in the original image. The image with the crosstalk removed is, due to the linearity assumption, not perfect. The assumption results in oversaturated colors. The matching benefits from this assumption, as it allows for easier matching.

Figure 3.2: Before and after color crosstalk removal; Original pattern.

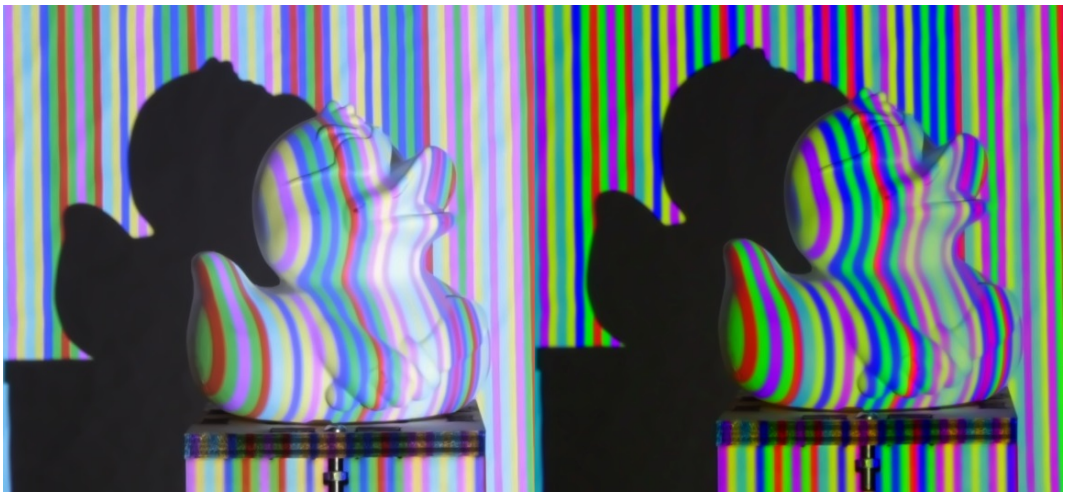

Figure 3.3: Before and after color crosstalk removal; Original pattern.

3.1.3 Edge detection & scoring

In order to detect the edges, a gradient function is computed for every horizontal row of the camera. This gradient function consists of the sum of the squares of all the color channels. Edges are placed at the maxima of this gradient function. An user-defined threshold is used to ignore small peaks in this gradient field.

Correctly identifying the correspondence between the projection and the captured image hold two main difficulties; mislabeling and occlusions.

The mislabeling of edges can occur because of the surface properties of the scanned object. The color, surface reflectance, viewing direction, crosstalk and noise of the CCD all influence the observed color. It is therefore impossible to fully match all edges with 100% certainty.

Occlusions provide another problem of the edge matching, because it can result in (large) parts of the projected pattern not being visible to the camera and can introduce inconsistency in the observed sequence. The first step is to perform general edge detection on the captured

image. For each of the color channels a 1D gradient is created. These gradients are combined and the local maxima of this combined gradient are used to obtain the edge candidates. Note that this set might contain false detections due to texture and occlusion. A fixed threshold can be used to filter out very small changes, as they are not likely to be real edges in the first place.

In order to match the edges, a scoring function is used, which calculates the match of any given observed edge with any given projected edge. This is done by defining the function $score(q, e)$ of a projected edge q and an observed edge e. This function calculates the score match between the provided edges, with the score being -1 for a very poor match and 1 for a perfect match. These values are calculated by taking the lowest value of the consistencies of each channel;

$$score(q, e) = \min_{c \in r,g,b} \{consistency(q_c, e_c)\}$$

Let $e = (er, eg, eb)$, where $e_c \in [-1, 1]$, which is the 1D intensity gradient of e in a single color channel and $q = (q_r, q_g, q_b)$, where $q_c \in \{1, 0, -1\}$. The $consistency(1, e_c)$ should be 1 if e_c is sufficiently large, 0 if $|e_c|$ is sufficiently small and negative when e_c is negative. The consistency is expressed by:

$$\begin{aligned} consistency(1, e_c) &= CLAMP(\frac{e_c - \alpha}{\beta - \alpha}; -1, 1) \\ consistency(0, e_c) &= CLAMP(1 - \frac{|e_c| - \alpha}{\beta - \alpha}; -1, 1) \\ consistency(-1, e_c) &= consistency(1, -e_c) \end{aligned} \tag{3.4}$$

Where

$$CLAMP(x; x_0, x_1) = \begin{cases} x_0 & \text{if } x < x_0 \\ x & \text{if } x_0 < x \leq x_1 \\ x_1 & x_1 < x \end{cases}$$

and $\alpha \leq 0 \leq \beta$ are used as soft thresholds which are chosen based on the uncertainty of the edge measurements. All measurements in the domain of $[\alpha, \beta]$are classified by values that reflect the uncertainty, where sufficiently small or large values are classified with -1, 0 or 1. The consistencies lead to graphs in figure 3.4.

3.1.4 Multi pass dynamic programming

By using the probability of two edges being a match as calculated by the score function, we can create a $n * m$ sized grid that hold all the possible matches between the projection and camera image.

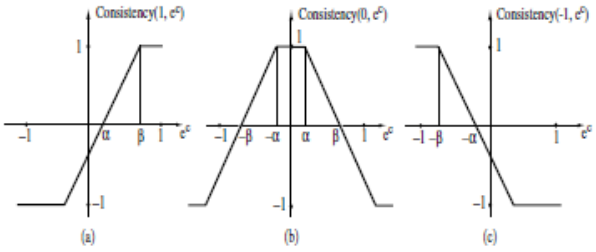

Figure 3.4: Consistency graph.

A final matching between e and q is found by finding a path through the score grid, where each edge can only be matched once. Without additional constraints, there are far too many possible paths to calculate. In order to decrease this, we adopt the monotonicity assumption. This assumption holds that all matches must be depth ordered. This assumption could lead to dropouts in the reconstructed point cloud as occlusions can violate the depth ordering. We can reconstruct these dropped areas by using multiple passes of the dynamic programming that is used to find the optimal path.

The optimal path of the sub-grid G_{ji} that is defined by $[0, j]X[0, i]$ must be either of three possible configurations;

- It holds the $vertex(j, i)$ and the optimal path $(j - 1, i - 1)$.
- The optimal path is the same as the optimal path of $(j - 1, i)$.
- The optimal path is the same as the optimal path of $(j, i - 1)$.

The calculation of every optimal path to every possible match takes $O(nm)$ space and time and a final path of all possible paths can be backtracked in $O(n + m)$ time.

0.4	0.8	-0.4	0.9
-0.2	1	0.3	0.6
0.3	0.8	0.9	0.1
0.5	0.4	0.1	-0.7

Score

0	1	1.1	2
0	1	1.1	1.5
0	0.8	0.9	0.9
0	0	0	0

Path

The highlighted path value in the table is 1.1 because the path value of $(1, 1)$ plus the score of $(2, 2)$ is higher than the path value of $(1, 2)$ and $(2, 1)$ respectively.

3.1.5 Drawbacks

The system has several major drawbacks. The encoding only allows for a limited number of stripes to be used, which decreases the numbers of retrieved points. This can be partially addressed by projecting multiple patterns that are shifted in between, but this increases the

number of images required. While it is possible to take a large number of images, a less complicated pattern with a less complex consistency matching, is preferred.

Another drawback is the usage of color in the projected pattern. The correct identification of the projected colors plays an important role in the correct depth estimation. When scanning a white object, this can mostly be solved by using color calibration, such as the proposed color crosstalk matrix. This method however, fails when the object has bright colors, as the camera is no longer to pick up the stripe pattern. Such objects would be retrievable by a binary scan or phase shift pattern, as only the intensity of the observed pixels is of importance.

The algorithm only matches a 1D coordinate for the projector, only a ray-plane intersection is possible. This gives less accurate results when compared with ray-ray intersections, as the plane is created using a least square matching, which generalizes the distortion of the lens. While this effect is barely noticable on the overal shape of the object, it could influence small details.

The total time that the algorithm needs, including the capture time, is larger than that of phase shift of binary pattern. Especially the proposed multi shot method suffers from long running time, in the order of 10-15 minutes for a single scan when tested on the setup as described in 4.1.

Aside from these method specific drawbacks, there are several cases in which almost all structured light algorithms fail. Its not possible or very difficult to scan highly reflective surfaces (such as shiny metals), transparent objects (such as glass and to a lesser extent wax or polystyrene) or objects with much internal scattering (such as the inside of bowls).

3.2 Polarization-based inverse rendering from a single view

3.2.1 Degree of polarization

From the Fresnel theory, we know that the intensity of polarized light reflected from an object will vary based on the refractive index of the object and the orientation of the surface. By observing the object with a linear polarizer mounted camera, we can obtain a sinusoidal curve of the intensity with respect to the orientation of the linear polarizer. A schematic overview of this can be seen in figure 3.5.

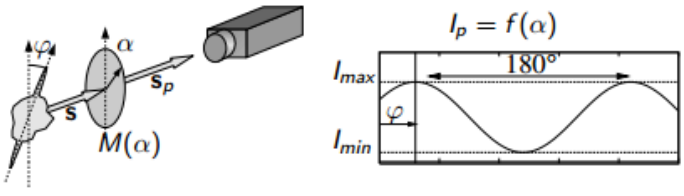

Figure 3.5: Schematic overview of Linear polarizer and observed intensities.

As we are looking for a 3D orientation, we must find two different rotations; the zenith angle and the azimuth angle. For the zenith angle, we must first retrieve the DOP (degree of

polarization). The DOP represents how much of the light has been polarized, where 0 means that the light is unpolarized and 1 means the light has been perfectly polarized. The degree of polarization is given by;

$$\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{3.5}$$

However, this assumes that we would have a large set of images, in which the filter was rotated by small steps. As this is not feasible and will not result in discrete results, we use three images taken at three distinct polarization filter angles ($I_0, I_{45} and I_{90}$). More images could be used to obtain results, but this will be much more time intensive and prone to small shifts of the setup. The following equation is used to calculate the DOP from the three images.

$$\rho = \frac{I_{90} - I_0}{(I_{90} + I_0) cos 2\phi} \tag{3.6}$$

3.2.2 Refraction index

The refraction index describes how much of the light that is reflected by the object is refracted and polarized in the internal scattering. This refraction index is dependent on not only the material of the object, but also on the smoothness of the object. Rougher surfaces tend to have much higher refraction indexes. This can be explained by the fact that higher degrees of polarization are found on edges that are aligned more parallel to the light rays. A rough surface has much more of these surfaces and a higher refraction index is required to compensate for this.

The DOP is related to the zenith angle by the following equation;

$$\rho_d = \frac{(n - 1/n)^2 sin^2\theta}{2 - 2n^2 - (n + 1/n)^2 sin^2\theta + 4cos\theta\sqrt{n^2 - sin^2\theta}} \tag{3.7}$$

Where ρ is the DOP of diffuse light, is the zenith angle and n is the refraction index of the object to air. By numerically solving this equation we can retrieve the zenith angle. In order to get correct zenith estimations a correct refraction index is needed. For the sake of simplicity, all results are generated with the assumption that the refractive index is 1.5. Figure 3.6 maps the DOP versus the zenith angle and shows the effect of different refractive indexes.

3.2.3 Phase

The phase is defined to be the angle at which the maximum intensity is observed. From the Fresnel coefficients it follows that the azimuth angle is either the same as the phase or the

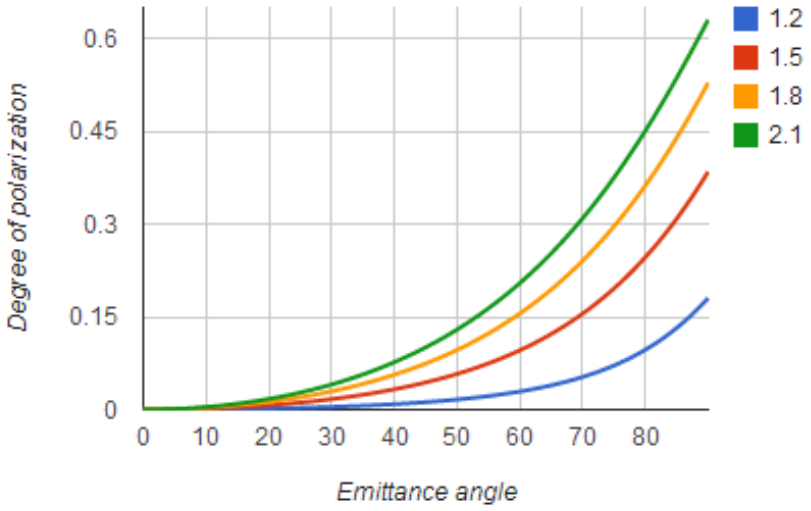

Figure 3.6: Graph that maps the degree of polarization versus the emittance angle with different refraction indexes.

phase + 180 degrees, due to the ambiguity. Using the three intensity images, we can calculate the phase by using the following formula:

$$\phi = \frac{1}{2} arctan(\frac{I_0 + I_{90} - 2I_{45}}{I_{90} - I_0}) + 90° \tag{3.8}$$
$$if(I_{90} < I_0)[if(I_{45} < I_0)\phi = \phi + 90°else\phi = \phi - 90°]$$

Figure 3.7: Degree of Polarization and Phase of a rubber duck. *Left:* Degree of polarization. Dark areas indicate a higher DOP. *Right:* Phase

3.2.4 Drawbacks

As can be seen in figure 3.6, the zenith angle is quite different when another reflection index is chosen. Even if a correct reflection index can be obtained (by knowing the material of the object beforehand), the reflection index assumes a smooth surface. A rough surface composed of the same material will have a different reflective index.

In figure 3.7, another of the drawbacks of the system is visible. The resulting normals suffer from noise where the degree of polarization is low. This is caused by the small differences of the intensity images when this is the case.

A final drawback is caused by the angle ambiguity. In the case of smooth objects this can be solved with a relative simple normal propagation and a smoothness assumption. In most cases however, there will be discontinuities in the scanned object and the smoothness constraint cannot be used.

3.3 Improvements

Based on our implementation and subsquent use of the SL algorithm, several improvements haven been made to the method. These improvements include simplifications to the method, an improved pattern and several noise removal operations.

3.3.1 Consistency threshold

The consistency stage of the algorithm uses two thresholds (α and β) in order to clamp the consistency values. This clamping has an impact on the global optimisation stage as it could lead to a different path being chosen. By setting these values to 0 and 1 respectively, all scores will have a gradient confidence. This also removes the need to tweak these parameters for different observed scenes. Despite that very low and high values will no longer receive a discrete consistency, the pathfinding will still try to find the most optimal path, which will likely ignore low values and match high values.

3.3.2 deBruijn Pattern

The original pattern includes both white and black stripes, which both lead to false classifications. The edges where one of the stripes is black are indistinguishable from shadows caused by occlusion, thus leading to false classifications near shadows.

The white and black edges both also suffer from intensity differences when compared to the other stripes. The white stripes are much brighter than the colored edges and the black stripes are much darker. In order to capture both these edges, the range of values that need to be captured is much greater, leading to less accurate results for the values in between.

The more difference between two neighboring stripes, the easier the resulting edge can be detected. By taking this into account, we constructed a pattern that uses no white and black edges and two successive stripes always change in at least two color channels. This leads to

a much higher local uniqueness property as each stripe can only be followed by three other colors.

Figure 3.8: Extract of the improved deBruijn pattern.

The difference between the results as obtained by both the new and old pattern are shown and explained in chapter four.

3.3.3 Single shot sub-pixel accuracy

One of the main disadvantages of the single shot method is the lack of subpixel accuracy. As only discrete pixel values are used to calculate the depth, a strong anti-aliasing effect can be seen in the point cloud. The main reference solves this by using space-time analysis to calculate a subpixel match. This space-time analysis is done by smoothing the projected pattern with a gaussian blur, combined with a horizontal shift between patterns. Instead of a consistency function a cost function is used. This cost can then be used to calculate a score, which can be used by the path finding algorithm. Due to the very greedy nature of this algorithm, this increases the computational time from 30-60 seconds towards 20-60 minutes.

Due to the greedy nature of the algorithm, we not only have a score of how well an edge matches with its final match, but also a score of how well that edge matched with the neighbouring camera pixels. If the score to the right of the matched camera pixel is much higher than the score of the matched pixel to the left, it is most likely caused that the true value of the match is in between the matched pixel and the right pixel. This is the same as done for the space-time analysis, but only using a single image. The usage of multiple images will theoretically give a more accurate match, but the truthful retrieval of color is also much more important. Especially areas with (intensely) colored texture influence the space-time analysis.

In order to find this match, we match a parabola through the three values. We can then use the maximum of the parabola as the new sub-pixel value of the match.

3.3.4 Noise removal

As digital cameras rely on the counting of photons, which always results in an discrete value, the measured RGB values are subjected to some noise. This noise can be decreased by using long exposure times, low ISO values or even average multiple images, but all of these assume either a reasonably complex camera system and/or the ability to take multiple images.

The most simple noise reduction is to use the surrounding pixels, combined with a certain weight, to calculate the new pixel value. This method assumes that nearby pixels are also nearby in the scene and thus influence each other. This assumption does not hold for our specific case and would also lead to blurring of the image, which is problematic for the stripe edge detection.

A more elegant solution is used by Antoni Buades et al. in Non-local means denoising [6]. Instead of assuming that nearby pixels are related, they assume that similar pixels are related.

This results in noise removal in which the edges are blurred far less, but is significantly more computationally expensive. For a more detailed explanation of the algorithm, we refer to the original paper.

Weve used the implementation of this algorithm as provided by the openCV library, using the default parameters.

3.3.5 Point cloud

Even with the improved noise removal, there will still be a significant amount of noise in the resulting point cloud. Some of this noise can be handled by performing statistical outlier removal (SOR). This method looks at the n neighbours of each point and removes the point if it does not meet a certain criteria. In this case, we calculate the mean distance from the point to its n neighbours. By assuming that this results in a gaussian distribution with a mean and standard deviation, we can remove all points of which their standard deviation to its neighbours is more than a user-defined threshold.

3.3.6 Subtraction method

The method suffers from influences of texture and object color. In order to remove the effects of this influence, we need to know what part of the observed image is caused by the projected image and what part is caused by the original color and texture of the scene. In other words, we need to look at the change of the object being illuminated by the pattern and without it.

Taking an image completely without projection and one with the pattern projected does not always give accurate results, as the camera range is set to perform best with the object brightly illuminated. Taking an image without illumination in a dark environment will therefore have very little effect on the overall result. To solve these problems, we propose the subtraction method.

Instead of taking a picture without illumination, we take an image with full white illumination and look at the absolute difference between the two images. This subtraction results in an inversion of the color pattern, so instead of comparing with the original pattern, we compare the subtracted image with the inverse pattern. The improvements of this method are discussed in more detail in chapter four.

3.4 Ideas

The only way to retrieve normals from a point cloud is through neighbouring point estimation. By using the nearby points, a normal can be estimated for each point. This, however, requires a user defined neighbourhood. A large neighbourhood leads to over smoothed normals, but a too small neighbourhood leads to incorrect normals.

Another issue with this method is that its not well equipped to handle sparse point clouds. Due to the sparse data points, there can be too little data for correct normal estimation. An example of such a situation can be seen in figure 3.10, where the black lines are the ground

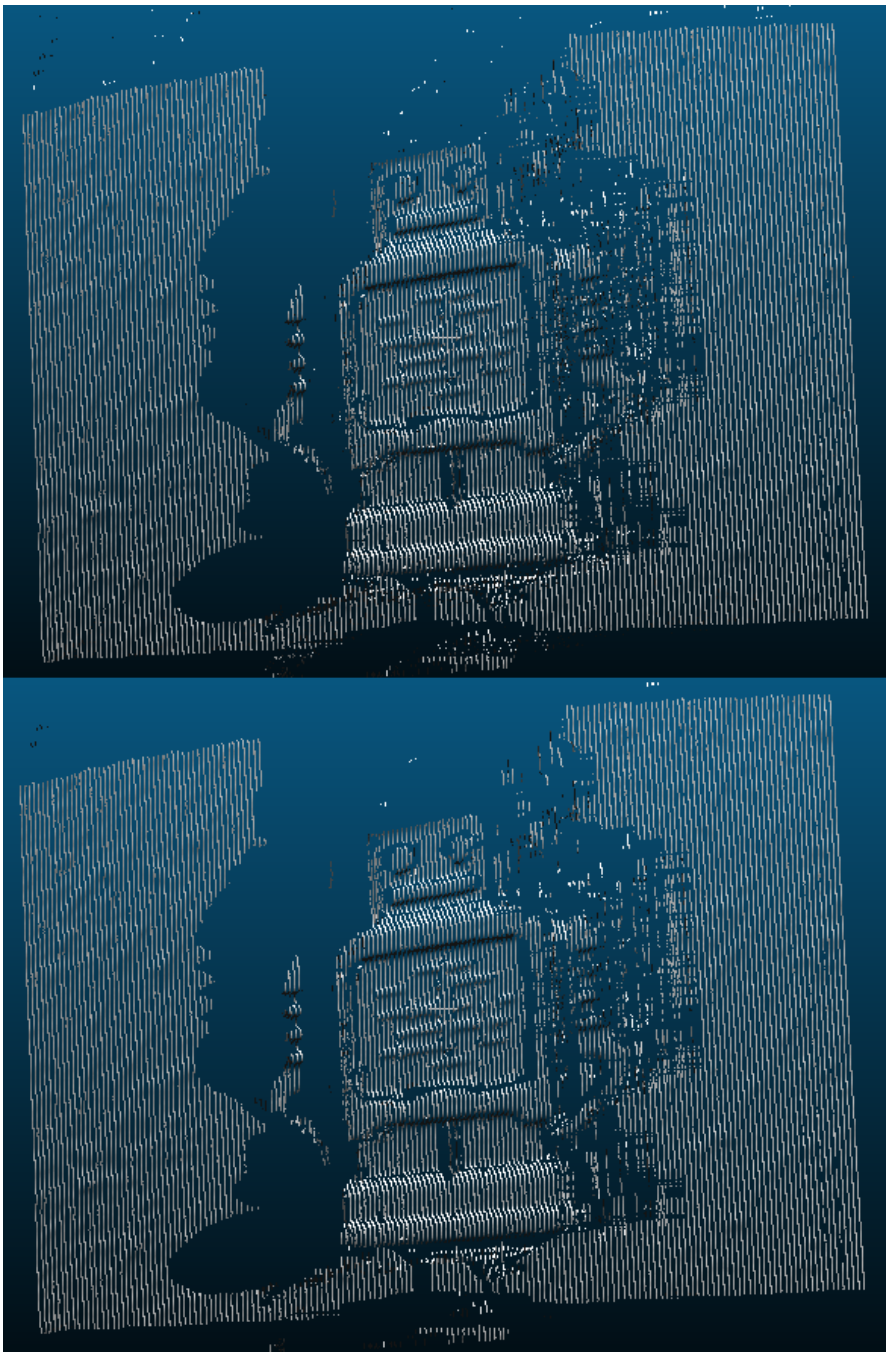

Figure 3.9: Pointcloud before and after SOR.

truth of the surface, the red points the calculated depth, the red lines the estimated normals (using the neighbours) and the green is the estimated surface. The estimated normals in the figure do not represent the ground truth normals, which will lead to an incorrect mesh reconstruction.

The shape from polarization method suffers from angle ambiguity, which could be solved by using the estimated angles from the nearest point. This does not completely remove the

Figure 3.10: Schematic drawing of failed normal estimation.

smoothness constraint, but does lessen the impact of this constraint considerably, as only a local smoothness is used instead of a much stricter global smoothness.

3.5 Proposed method

The combined method is composed of two stages; the polarization normal ambiguity solver and depth reconstruction.

The normal ambiguity can be solved by using the structured light points as a scaffold. The point cloud itself holds no normal estimation, so it needs to be estimated from the neighbouring points. For each point in the cloud a covariance matrix needs to be created using:

$$c = \frac{1}{k}\sum_{i=1}^{k} \cdot (p_i - \bar{p}) \cdot (p_i - \bar{p})^T, c \cdot \overrightarrow{v}_j = \lambda_j \cdot \overrightarrow{v}_j, j \in \{0, 1, 2\} \tag{3.9}$$

Where k is the number of points considered in the neighbourhood of p_i,which is user defined. $\bar{p}$ represents the centroid of of the nearest neighbours, λ_j the j-th eigenvalue and v_j is the j-th eigenvector. There is no mathematical way to solve the sign of the normal, an ambiguity exists in the calculated normals. Because the normals must face the viewing direction, the ambiguity can be solved by orienting the normals toward the viewpoint.

When the normals of the (sparse) point cloud are known, they can be used as a scaffold to solve the ambiguity in the normals from polarization. The polarization normal that has the lowest angular difference between the average normal of the nearest estimated normals is used.

A local, patch based method is used to reconstruct the depth. The patches are generated by using known depths of the structured light as the corner points and the normals obtained by the polarization method.

The patches are generated by starting at the left topmost pixel of the image that holds the detected edges. Once a pixel is found with a known depth, all pixels to the right are checked until another pixel with a known depth is found. Finally all pixels directly below the rightmost found pixel are checked until a pixel with a depth value is found. This subdivides the image into patches where at least three of the corners have a depth. An example of this can be seen in figure 3.11.

Each patch is used as a local area of which the depth needs to be calculated. In order to calculate the depth we use trapezoidal integration [15]. The polarized normals are used for

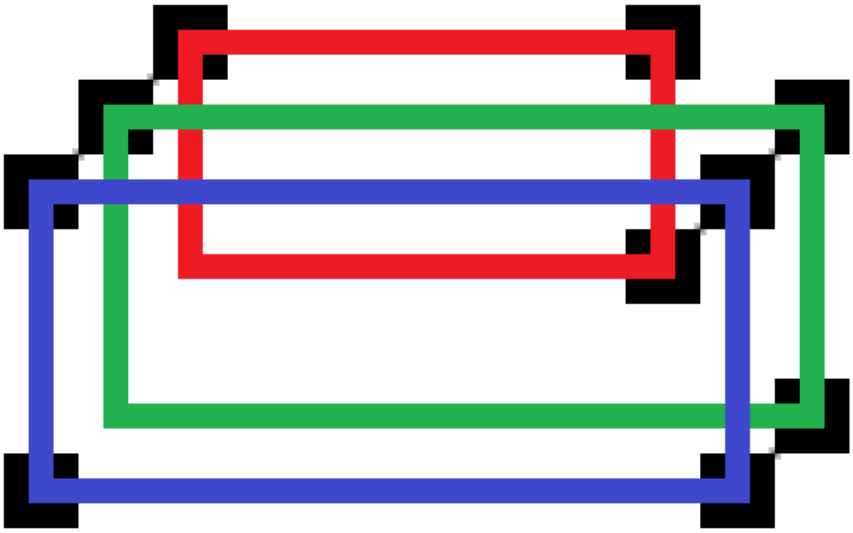

Figure 3.11: Schematic representation of the patch creation. The black squares represent pixels with a known depth and the colored lines the extracted patches.

the gradient field. When starting from the leftmost top corner the following algorithm is used to fill the horizontal and vertical depth;

$$\begin{aligned} Z_{i,1} &= Z_{i-1,1} + p_{i-1,1} \\ Z_{1,j} &= Z_{1,j-1} + q_{1,j-1} \end{aligned} \tag{3.10}$$

Where Z_{ij} is the depth of the image coordinate (i, j). When a single line of horizontal and vertical depth of the patch are filled, the rest of the patch can be populated by using:

$$Z_{i+1,j+1} = \frac{1}{2}(Z_{i,j+1} + Z_{i+1,j}) + \frac{1}{4}(p_{i,j+1} + p_{i+1,j+1} + q_{i+1,j} + q_{i+1,j+1}) \tag{3.11}$$

Because the depths are propagated through the grid, noise in the normals from polarization will result in error propagation. To lessen the impact of the noise on this propagation the algorithm is repeated for each corner of which a depth is known. If for example the algorithm starts at the top right corner, the horizontal depth is first propagated from right to left and the vertical depth from top to bottom. The average of the three or four calculated depths are used as the final depth.

Most pixels will have depths calculated by multiple patches due to overlap. This adds an aditional smoothing of the error propagation, both of the errors due to incorrect normals as errors in the estimated depth from structured light.

Chapter 4

Experimentation

Several experiments have been conducted in order to test the accuracy of the methods sperately. Each of the improvements to the SL method have been tested sperately in order to quantify the improvements. The combined method has been tested with a simple model.

4.1 Setup

The computer system used in all tests is an Alienware M11x R3 laptop with 6GB of RAM and an Intel I5-2467M cpu that is clocked at 1.6GHz. The software does not utilize threads, so only a single core is used.

All images are taken with a Canon D650 DSLR camera that is controlled through the software using the EDSDK API provided by Canon. The rotating linear polarizer filter in front of the camera is a glass filter that is manually rotated to the correct position.

The beamer is an Optoma DLP beamer, with a maximum supported resolution of 1024 X 786. If a polarized image source is mentioned, this beamer is used with a plastic linear filter placed in front of this projector. Because of the perfect filter assumption, it is also assumed that this results in a perfect polarized lightsource.

4.2 Structured light

4.2.1 Improved pattern

In section 3.3.2 we propose a new encoding scheme for the structured light pattern. This pattern attempts to solve some of the problems with the old pattern. The difference between a scan made with the old pattern and the new pattern

Because of the new constraint that between the two colors of an edge there need to be two changes in a color channel, there is far less noise in the image. Even though the window of colors is much larger ($n = 5$ instead of 3), there is only a very small area near the foot that

could not be reconstructed. The lack of any notable difference can be attributed to the global optimization scheme, as it fixes most of mistakes caused by the larger window.

The new pattern also lacks white stripes, as they are brighter than their neighbours. This brightness influences nearby stripes, which could lead to false classifications. Because the difference between two neighbouring stripes is guaranteed to be at least two color channels edges more edges are detected, especially in regions that are almost perpendicular to the projector image plane.

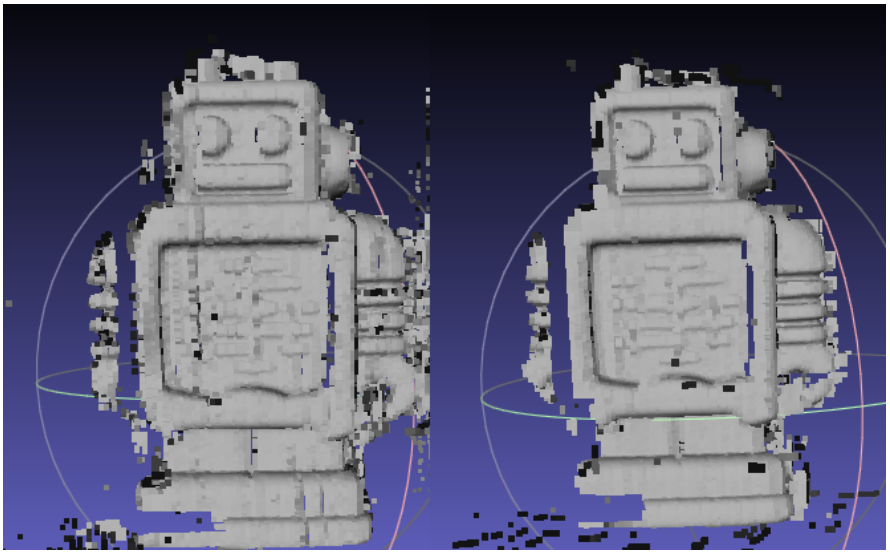

Figure 4.1: Resulting point clouds of two different projection patterns. *Left:* Old pattern. *Right:* New pattern.

Because the robot is 3D printed, we have the original 3D model that was used to fabricate it in the first place. This allows for easy comparison between the scans and the ground truth. The improved pattern results in 1.65 mm standard deviation with a sigma of 5.88422. The old pattern has a standard deviation of 2.1 and a sigma of 8.1. Part of this standard deviation can be attributed to limitations of geometrical tolerances of the manufacturing process.

Figure 4.1 shows a comparison of the point clouds created with both projection patterns and figure 4.2 shows the pattern matched to the ground truth. Blue colored points have little to no difference to ground truth and red points have a large difference when compared to the ground truth.

4.2.2 Polarized filters

Polarization is an excellent method to reduce the specularities of a surface. When using an unpolarized light source, a larger part of the specular light will be filtered when compared to the diffuse reflection. If a polarized light source is used, which is obtained by placing a linear polarizing filter in front of the source, the angle of the filter in front of the camera becomes important. If both hold the same orientation more specular light is captured. When the angle is at 90 degrees difference, almost all specular light is filtered. The difference between these two states can be seen in figure 4.4.

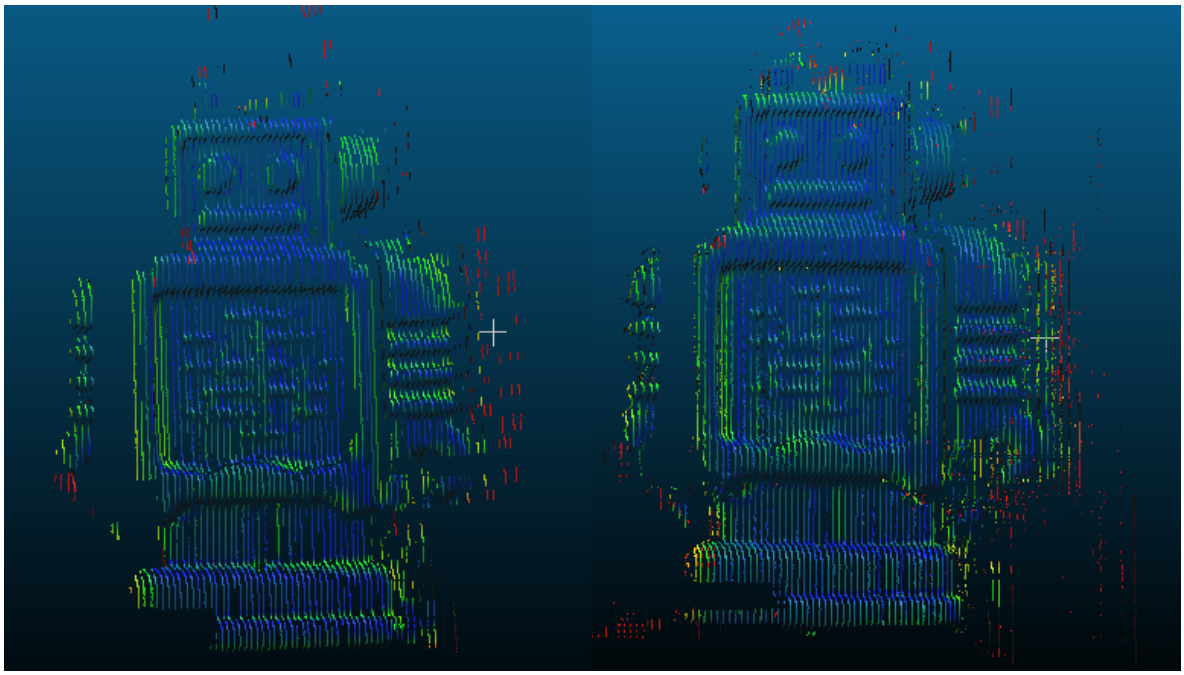

Figure 4.2: Differences to ground truth *Left:* New pattern. *Right:* Old pattern.

	0 Degrees	45 Degrees	90 Degrees	No polarization
Mean dist	0.605	0.739	0.851	0.727
Sigma	1.447	1.770	2.6	1.549

Table 4.1: Standard deviation and sigma of point clouds to ground truth in mm

Because specularities tend to result in oversaturated pixels, these specularities need to be avoided as much as possible. Almost all common objects have at least a small amount of specular reflection.

In order to test the impact of the polarization of light on the structured light method, four different scans of the same object (the robot) have been made. These have been compared with the ground truth of the robot model.

The results in table 4.1 indicate that using the polarization increases the accuracy of the results when compared to no polarization. The accuracy decreases when the polarization is set at 90 degrees, which is most likely caused by less light reaching the camera. The decrease of the pixel intensities leads to lower matching certainties, which in turn lead to poorer global matching.

4.2.3 Substraction

The subtraction method mostly has influence on edges of objects, where the intensity of the projected pattern is much less. In these areas, the projected light has much less influence on the perceived color. As the subtraction method scales the observed colors, the effect of the object color is factored out somewhat.

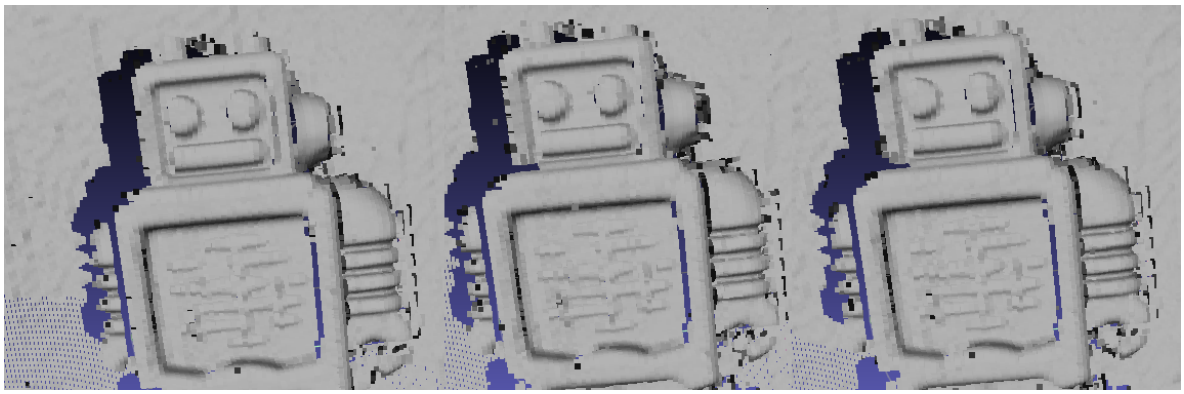

Figure 4.3: Point cloud results with different polarization filter ortientations. *Left:* 0 degrees. *Middle:* 45 degrees. *Right:* 90 degrees.

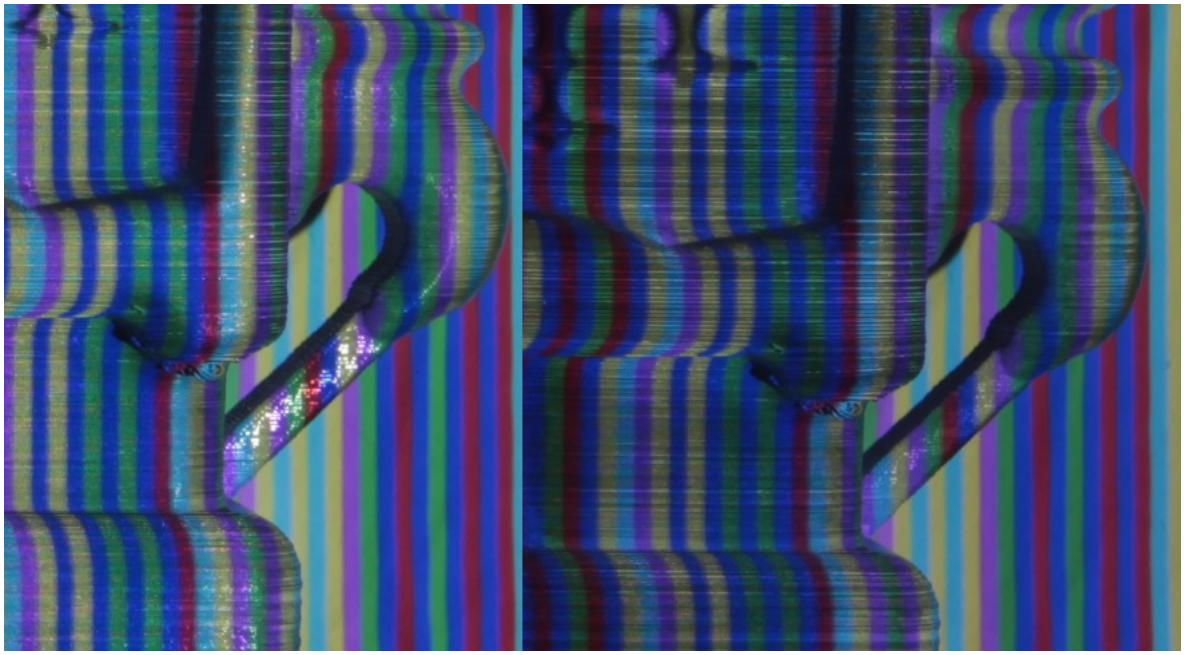

Figure 4.4: Images taken with polarized light. *Left;* filter at 0 degrees. *Right:* Filter at 90 degrees.

The method improves the results for both the original and the improved pattern. The results of the scans are displayed in figure 4.5 and 4.6. No quantified test were performed, as the method results in previously unmatched edges being matched correctly.

The images have been taken with the polarization filter at 90 degrees and a ≈30 degrees angle between camera and projector. The distance between both is ≈20 cm.

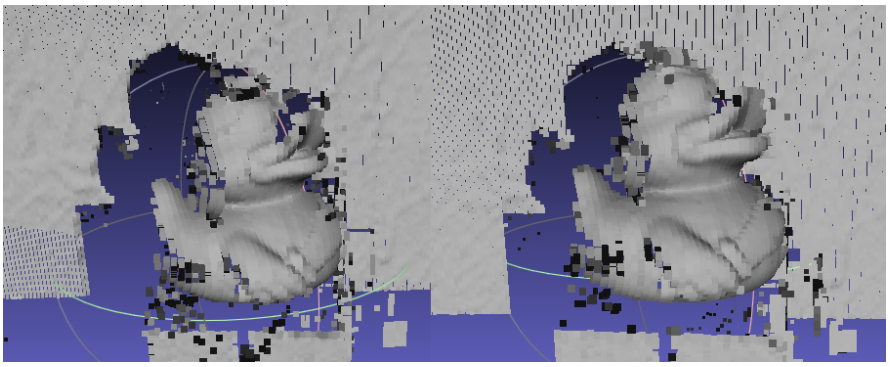

Figure 4.5: Scan of rubber duck with the new pattern. *Left:* Original method. *Right:* With subtraction method.

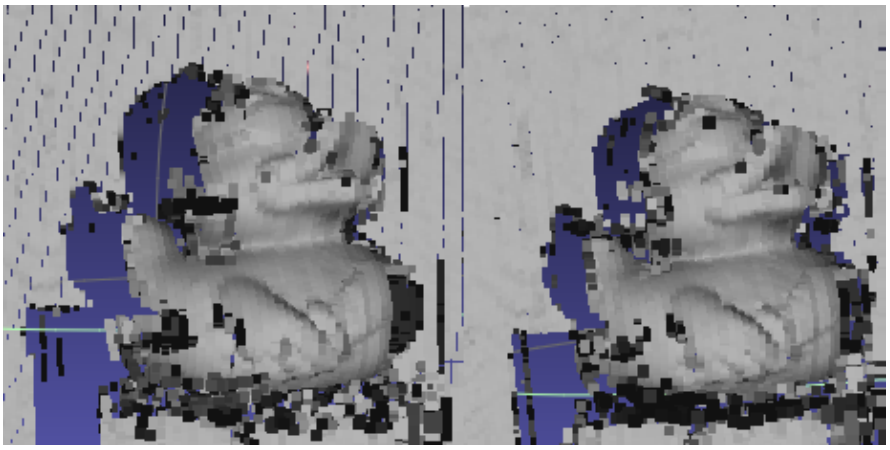

Figure 4.6: Scan of rubber duck with the old pattern. *Left:* Original method. *Right:* With subtraction method.

4.3 Polarization

All polarization images are taken with an exposure time of 0.25 seconds, an aperture of 9.0 and an iso value of 100.

4.3.1 Accuracy

By scanning a flat sheet of paper that has been attached to a wooden board, we've created a surface in which all normals should point in the same direction. In order to measure the accuracy of the method, we made several scans of that image. Each of the three intensity images used per scan are combined averages of 3 images taken in succession.

In order to test the observation that the degree of polarization has impact on the noise of the results, two cases are compared. In the first case, the flat plane is placed in such a way that the degree of polarization is very low. In the second case, the plane is rotated around the X axis, which results in higher degree of polarization. If the higher DOP results in less noise,

	First	Second	Third	Fourth	Fifth
Avg normal 1	0.935046 -0.154976 0.318859	0.922792 -0.162958 0.349128	0.919447 -0.185264 0.346836	0.93509 -0.180557 0.304968	0.91891 -0.167726 0.357034
Avg normal 2	0.933759 0.14678 -0.32642	0.921369 0.154734 -0.356564	0.918192 0.17697 -0.353309	0.934102 0.172335 -0.312656	0.917452 0.159431 -0.364505

Table 4.2: Average normals measured with an zenith of 36 degrees

	First	Second	Third	Fourth	Fifth
Avg normal 1	0.534614 0.637949 0.554264	0.557549 0.620982 0.550927	0.565878 0.62377 0.539159	0.585443 0.614287 0.529064	0.548423 0.623408 0.557311
Avg normal 2	0.529559 -0.639834 -0.556937	0.552352 -0.623142 -0.553715	0.560657 -0.62596 -0.542073	0.580144 -0.61665 -0.532143	0.543277 -0.625494 -0.560006

Table 4.3: Average normals measured with an zenith of 88 degrees

the differences between the scans should be less.

A patch of 250x250 pixels is taken and the average normals of that patch are used as the normal of the surface. The normals are ordered as ZYX. Table 4.2 shows the average normals for a low degree of polarization and 4.3 shows the average normals for a higher degree of polarization.

In order to see the actual difference between these angles, the dot product was used to calculate the angle between the normals. The average difference between the angles is 2.3 degrees for the first test and 1.5 for the second.

This displays the influence the noise has on the smoothed normals of an entire patch. In order to see what the spread of the individual normals are, their angle to the average of the patch needs to be calculated. For the first test the angle each normal to the total average is 13.4 degrees, with a standard deviation of 6.3 degrees. The second test has an average difference of 9.3 degrees with a standard deviation of 6.8. The low degree of polarization results in 30.6% more noise.

The noise has a big impact on the estimated normals, even when all methods to decrease the noise as proposed by [3] are used. Low degrees of polarization are the main cause of this noise, as these areas only have small differences in their intensity values. Because the changes are smaller, they are far more susceptible to noise. A good example for this can be seen in figure 3.7. The areas where there is a high degree of polarization, such as around the beak of the duck, there is little noise in the phase. Areas with low DOP, such as around the wing, have a large amount of noise.

4.4 Merging

An as simple as possible object was used to show the results of the combination method. A picture where the object is lighted by a single image source can be seen in figure 4.7. The shape of the cone results in multiple areas with low degree of polarization, which lessens the impact of the noise on the normals obtained from polarization.

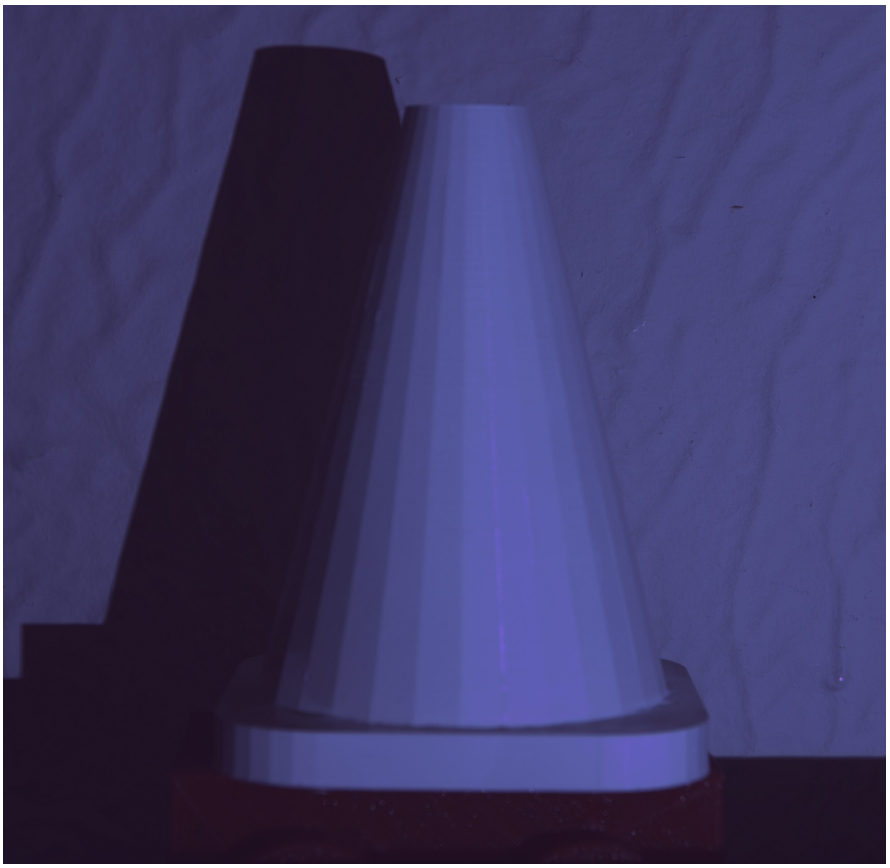

Figure 4.7: Picture of cone as lighted by lightsource

The method is not perfectly capable of solving the ambiguity, as can be seen in figure 4.8. The most notable drawback of the method is the lack of solved normals for areas where no depth is estimated, as is visible for the flat area on which the cone resides. Noise in the results from structured light also influences, as can be seen in the red patches to the right of the cone. Due to the depth being incorrectly measured, the normals are incorrectly estimated, which leads to the wrong choice of normals.

The number of estimated depths used for a single point varies greatly as the patching method combines a large numbers of patches for areas that are composed of a diagonal, especially in occluded regions. This has the advantage that the depth of large patches is calculated from more sources. This lessens the negative impact the larger distance to a known normal somewhat, due to the averaging of multiple calculated depths. The number of depths used can be seen in figure 4.9. At the brightest points 20 patches overlap.

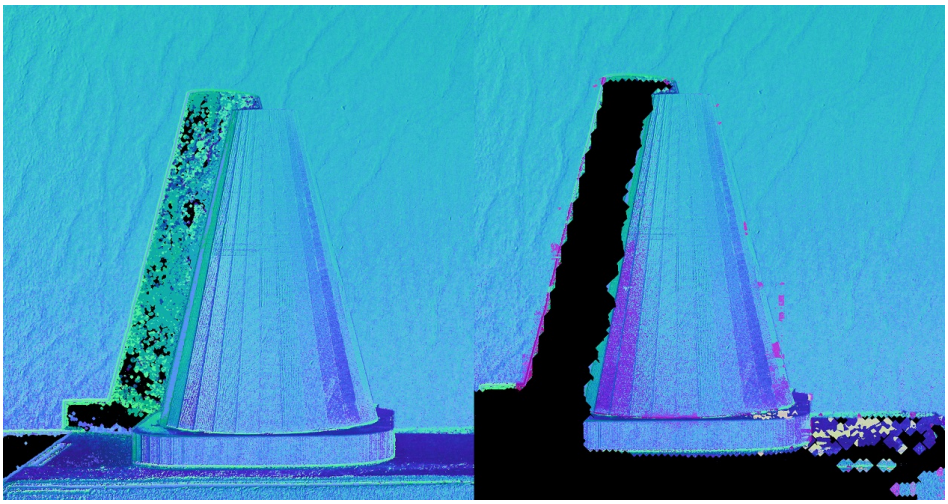

Figure 4.8: Normal ambiguity solving. Normals are displayed in RGB channel. *Left:* One of the normal candidates set. *Right:* normals resulting from ambiguity solver.

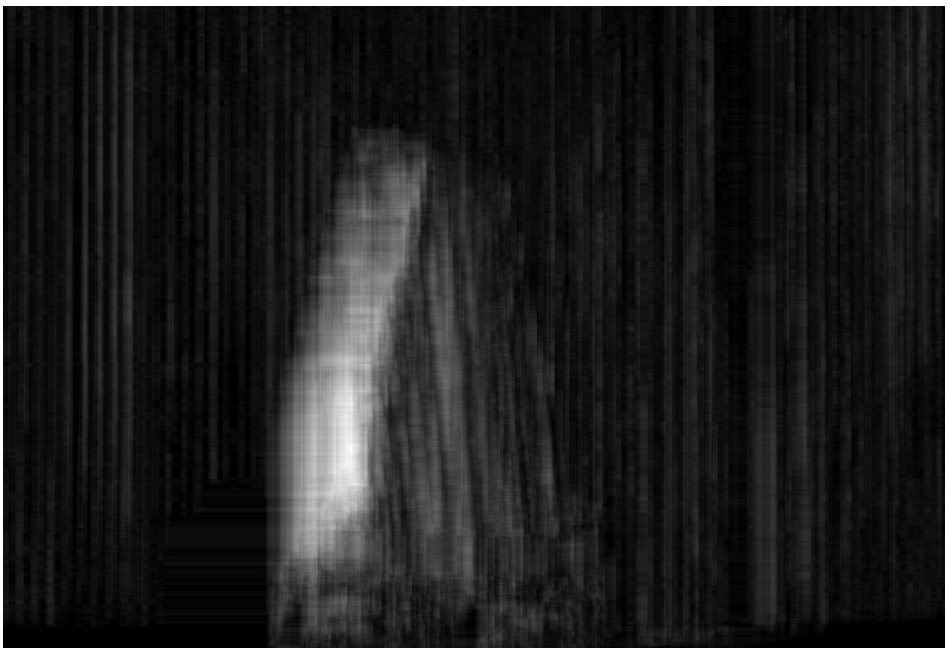

Figure 4.9: Number of depths used. Lighter areas indicate more used depths.

The resulting point cloud has a much higher resolution. Where the original point cloud contained ≈180.000 points, the point cloud resulting from the combination holds ≈2.000.000 points. The result is significantly more noisy than the original. This is due to the propagating of the gradient field, as a few erroneous results can influence the entire patch. If one of the anchor points is wrong, it is almost guaranteed that all results in that patch are wrong, despite the smoothing used. The patch method does ensure that the global shape is maintained to a certain degree.

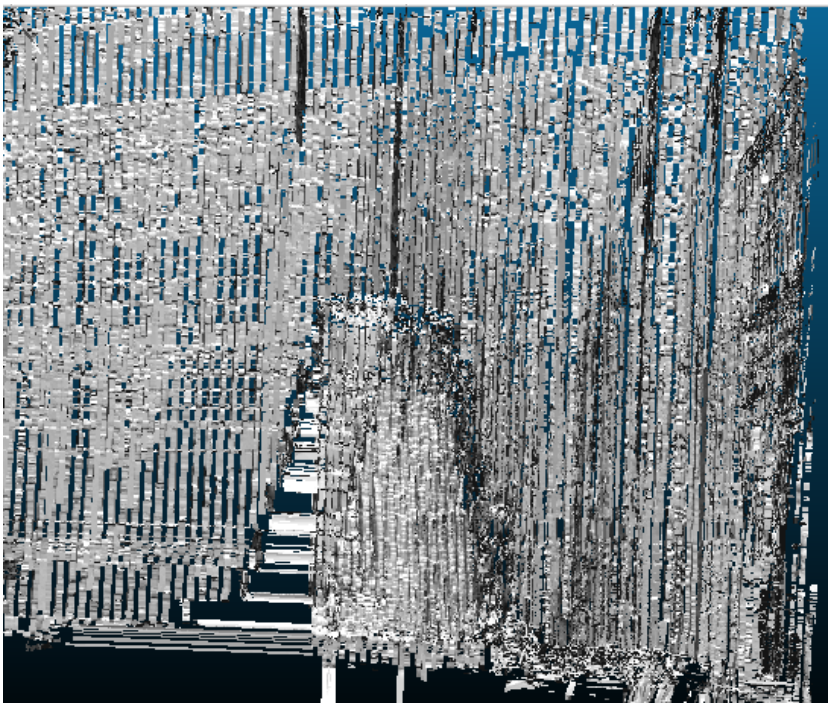

Figure 4.10: Point cloud resulting from combined method.

The current method has hardly any noise reduction or correction. Higher order methods such as Fourier transformations[17], wavelets[18] or shapelets[19] have far superior results when compared to this method, but they would require extensive modification to be used with known depths.

As there is no global optimization or smoothing employed, the patches can clearly be seen in the results. The areas with a large number of patches are much smoother, but not always correct. The large patches also result in shadowy regions, where no data is available to be incorrectly filled in with depth data.

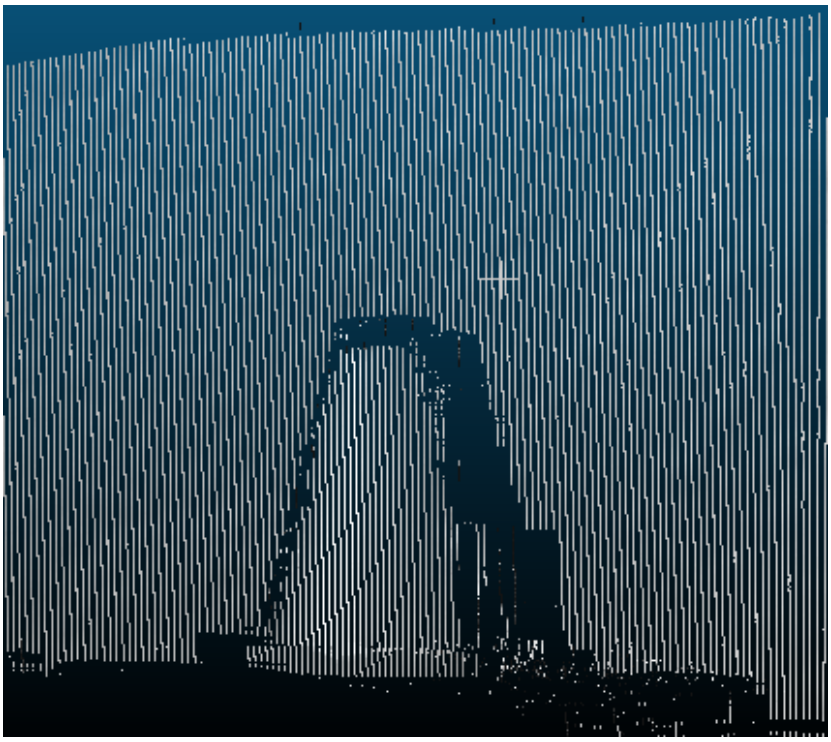

Figure 4.11: Point cloud as generated by structured light.

Chapter 5

Conclusions and future work

5.1 Conclusions

For this project, both the structured light method as proposed by Zhang et al. and the polarization method as proposed by Miyazaki et al. have been implemented. The motivation for this project was to develop an open source framework that would allow for accurate high resolution 3D scanning.

The scanning method in its current state is unsuited for general 3D scanning, due to its many constraints and noise even though improvements were made to the reference . However, it does hold potential, as it provides a solution for the scale retrieval that most gradient to depth algorithms suffer from.

5.1.1 Structured light

Several improvements have been made to the method of Zhang et al., most notably the sub-pixel accuracy and subtraction method. Both methods remove some of the drawbacks of the method, making it more robust and accurate. The method itself is still not usable for brightly colored objects, as this results in stripes no longer being visible or being identified as differently coloured stripes. The accuracy of the improved single shot method is less than the reported accuracy of the multi-shot solution of the reference paper (0.6 mm versus 0.25 mm), but requires only two images instead of seven and several orders of magnitude less running time.

5.1.2 Polarization

Based on the experiments we can conclude that the the shape from diffuse polarization is very susceptible to noise, especially with low degrees of polarization. Even with the lower degrees of polarization, there standard deviation of normals of a flat plane is about 7 degrees, which makes it not quite suited for the purpose of recovering details that would not be recoverable with structured light. Multiple measurements could be used to increase the accuracy, but this has a significant impact on the running time.

The selection of a correct refraction index is difficult and has significant impact on the estimated normals. Because of the assumption that the refraction index is the same for the entire model, only objects that consist of a single material with a known refraction index can be scanned.

5.1.3 Merging

The combination of both methods suffers greatly from noise due to the error propagation. If one of the methods has wrong results for a certain patch, the entire patch will have wrong results. Because of this "additive" nature of the algorithm, the final result sufers from a very large number of false estimations. Despite this problem, it is able to create very high resolution pointclouds that require only five images to be taken of the scene.

5.2 Future work

The normals estimated from the polarization could be improved by using multiple viewpoints, so that multiple estimations can be used to compute a single normal. This also allows for discarding normals estimated with a low degree of polarization, which suffer more from noise. Because the usage of multiple views also increases the number of points retrieved from SL, these normals should also increase in accuracy.

The refraction index could be calculated by minimizing the difference between the normals as estimated from SL and the polarization. This has an added benefit of providing with an indication of the type of material that is scanned.

The modularity of the system enables relatively easy switching of methods for each of the sub problems (Obtaining depth anchors, obtaining normals, combining,etc). It would be valuable to pursue several other combinations of methods.

Bibliography

[1] Zhang, Li, Brian Curless, and Steven M. Seitz."Rapid shape acquisition using color structured light and multi-pass dynamic programming." *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on.* IEEE, 2002.

[2] Miyazaki, Daisuke, et al. "Polarization-based inverse rendering from a single view." *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on.* IEEE, 2003.

[3] Atkinson, Gary A., and Edwin R. Hancock. "Shape estimation using polarization and shading from two views." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.11 (2007): 2001-2017.

[4] Huynh, Cong Phuoc, Antonio Robles-Kelly, and Edwin Hancock. "Shape and refractive index recovery from single-view polarization images." *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on.* IEEE, 2010.

[5] Thilak, Vimal, David G. Voelz, and Charles D. Creusere. "Polarization-based index of refraction and reflection angle estimation for remote sensing applications." *Applied Optics 46.*30 (2007): 7527-7536.

[6] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel, Non-Local Means Denoising, Image Processing On Line, vol. 2011

[7] Caspi, Dalit, Nahum Kiryati, and Joseph Shamir. "Range imaging with adaptive color structured light." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20.5 (1998): 470-480.

[8] Zhang, Zhengyou. "A flexible new technique for camera calibration." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.11 (2000): 1330-1334.

[9] Learning OpenCV: Computer Vision with the OpenCV Library (2008) by Gary Bradski, Adrian Kaehler

[10] Morita, Hiroyoshi, Kaanyasn Yajima, and Shojiro Sakata. "Reconstruction of surfaces of 3-d objects by m-array pattern projection method." *Computer Vision., Second International Conference on.* IEEE, 1988.

[11] Born, Max, and Emil Wolf. Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. CUP Archive, 1999.

[12] Bayer, Bryce E. "Color imaging array." *U.S. Patent No. 3,971,065.* 20 Jul. 1976.

[13] Kinect product specifications, [Retrieved on: 10 oct 2013]. Available: http://msdn.microsoft.com/en-us/library/jj131033.aspx

[14] Bergmann, Dirk. "New approach for automatic surface reconstruction with coded light." *SPIE's 1995 International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics,* 1995.

[15] Wei, Tiangong, Reinhard Klette, and Tamaki Campus. "On depth recovery from gradient vector fields." *Algorithms, architectures and information systems security* (2009): 75-95.

[16] Ng, Heung-Sun, Tai-Pang Wu, and Chi-Keung Tang. "Surface-from-gradients without discrete integrability enforcement: A Gaussian kernel approach." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32.11 (2010): 2085-2099.

[17] Frankot, Robert T., and Rama Chellappa. "A method for enforcing integrability in shape from shading algorithms." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 10.4 (1988): 439-451.

[18] Kovesi, Peter. "Shapelets correlated with surface normals produce surfaces." *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Vol. 2.* IEEE, 2005.

[19] Audet, Samuel, and Masatoshi Okutomi. "A user-friendly method to geometrically calibrate projector-camera systems." *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. IEEE, 2009.*

[20] Sells, Ed, et al. "Reprap: the replicating rapid prototyper: maximizing customizability by breeding the means of production." *Handbook of research in mass customization and personalization, Forthcoming (2010).*

[21] Ultimaker 2 specifications [Retrieved on 1 november 2013]. Available: https://www.ultimaker.com/pages/our-printers/ultimaker-2

[22] Replicator 2 specifications [Retrieved on 1 november 2013]. Available: http://store.makerbot.com/replicator2.html

[23] Wehner, Rdiger. "Polarization vision-a uniform sensory capacity?." *Journal of Experimental Biology* 204.14 (2001): 2589-2596.

[24] Atkinson, Gary A., and Edwin R. Hancock. "Surface reconstruction using polarization and photometric stereo." *Computer Analysis of Images and Patterns.* Springer Berlin Heidelberg, 2007.