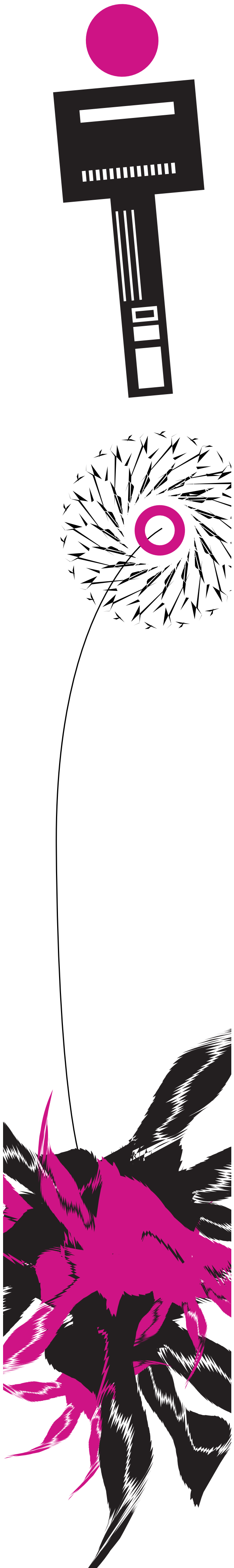

ROBOTICS AND MECHATRONICS

# Design of a 3D imaging system for psoriasis assessment

S. (Sander) Grimm

MSC ASSIGNMENT

**Committee:**
dr. ir. F. van der Heijden
dr. F.J. Siepel
prof. dr. ir. W. Steenbergen

August, 2020

034RaM2020
Robotics and Mechatronics
EEMathCS
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

UNIVERSITY OF TWENTE. | TECHMED CENTRE
UNIVERSITY OF TWENTE. | DIGITAL SOCIETY INSTITUTE

**Abstract**

Currently medical professionals rely on qualitative and subjective methods to assess the severity of psoriasis in patients. In this thesis a 3D stereo camera imaging system that can be used for more quantitative assessment of psoriasis severity is designed and implemented. The imaging system additionally includes an optical projector for pattern projection that is used to increase texture in the images and provide robustness to specular reflection. The results of this thesis are intended to be used as a stepping stone to designing an imaging system for at home use by patients in future work. With this in mind, care was taken to use techniques that would be suitable for patients to use in an at-home environment. Initial experiments to validate the system have shown promising results.

# Contents

# 1 Introduction

Psoriasis is a chronic skin condition that can cause red scaly patches of skin to appear. Correct assessment of psoriasis severity in patients is important to evaluate their response to treatment. Currently medical professionals rely on qualitative and subjective methods to assess the severity of psoriasis in patients. Performing accurate 3D reconstruction of human skin affected with psoriasis can help provide a more quantitative assessment of psoriasis severity. In addition, at home imaging performed by patients can reduce workload of medical staff while at the same time preventing the need for patients to travel to clinics or hospitals. In light of the recent COVID-19 pandemic and the possible emergence of social distancing as a permanent part of society the benefits of at home imaging are even more apparent.

The goal of this work is to design a stereo camera imaging system for 3D reconstruction of human skin to be used for quantitative the assessment of psoriasis severity. In addition to the stereo cameras the imaging system should include a projector for pattern projection to improve the systems robustness to specular reflection and low texture. This imaging system should designed to be low cost and easy to use in order to facilitate widespread at home use by patients.

Once the design is finalized a prototype of the imaging system is realized and its performance for psoriasis assessment evaluated. The benefit of pattern projection to 3D reconstruction quality in the presence of low texture and specular reflection is also evaluated.

This report starts providing a brief overview of the literary context of the conducted work. It continues by identifying and discussing the stakeholders in the design of the imaging system. Subsequently the specific requirements of the imaging system are presented. Based on these requirements a design space exploration is conducted to evaluate possible design approaches and make a final decision. The prototype implementation of this design is discussed next. After that, experiments are presented that were conducted to evaluate the results. This is followed by a discussion of the achievements and limitations of the project as well as recommendations for future work. Finally the conclusion is presented.

# 2 Related Work

This work investigates the feasibility of implementing high quality 3D reconstruction of human skin affected by psoriasis using a stereo optical setup with cheap commercial camera components. In this section we discuss the literary context and motivation of the work, as well as the contributions that are made.

## 2.1 Psoriasis assessment

A variety of research has been carried out in the area of Psoriasis assessment. A commonly used system for assessing the severity of psoriasis is scoring according to a PASI (Psoriasis Area Severity Index). The PASI score is evaluated through estimation of three parameters: Erythema (redness), Induration (thickness) and Desquamation (scaling) as well as the area of the affected skin. Generally, medical professionals assess each of these parameter subjectively by assigning them a severity of none, mild, moderate, severe or very severe. The goal of this work is to design a novel imaging system to assist in all in one objective assessment of the three parameters PASI score in order to provide better assessment and treatment for patients.

A number of studies have been conducted on objective assessment of psoriasis. The study in [1] explores a method for a more objective assessment of erythema. This study explored various machine learning approaches automated assessment of the erythema score from images. The Authors of [2] have ventured to perform objective assessment of induration by 3D scanning of psoriasis affected skin from a number of patients and comparing the results to PASI scoring done by medical professionals. Finally [3] presents an method to more objectively measure desquamation using an adhesive tape to collect flakes of skin rather than optically estimating the PASI score. Additionally [4] proposes a system integrating internet of things technology to objectively asses the erythema, this work is of particular interest because it presents an architecture that remotely connects patients and doctors to support assessment and imaging from the patients home. The imaging system created in this thesis could be integrated in such an architecture.

## 2.2 3D reconstruction of human skin

While quite some work has already been done regarding the 3D reconstruction of human skin for medical purposes. No effort has yet been made to investigate feasibility of obtaining high quality 3D reconstructions specifically using cheap commercial components for the purpose of at home imaging. In [5] a laser triangulation 3D scanner is used to obtain high quality 3D scans of human skin. The advantage of using a setup with a laser projector rather than a passive camera setup is that lack of surface texture has no influence on the quality of 3D reconstruction. In [6] the 3D reconstruction is achieved by estimating the depth of skin lesions from individual high quality 2D dermascopy images.

## 2.3 3D surface reconstruction

In [7] a narrow baseline stereo vision setup with state of the art components is used to obtain high quality 3D images of human skin. Use of a wider stereo baseline can be used to increase the accuracy of depth reconstruction from stereo images. This makes the DAISY descriptor[8][9] particularly interesting as a dense matching technique because it was designed specifically for wide baseline dense stereo matching.

While this work only considers a stereo view setup, another 3D reconstruction method that should be mentioned is multi-view reconstruction. These methods combine images from different (stereo or mono) camera views to obtain a single dense reconstruction. In order to perform such a reconstruction the pose of each of the views has to be accurately estimated. The act of tracking a camera sensor while creating a 3D reconstruction of the environment is known as SLAM (simultaneous localization and mapping). SLAM is an active area of research specifically within the field of robotics where SLAM methods are implemented with the goal of mapping the environment and tracking robotic vehicles, for instance in order to aid in navigation.

Methods such as ORB-SLAM2 [10] track the movement of a camera by finding easily identifiable matching points in a pair of frames, known as keypoints, and using those to obtain an estimate of the translation and rotation of the camera between the frames. This can be done by using for example the RANSAC algorithm. Quite a number of different methods for finding these keypoints have been developed such as SIFT[11], ORB[12], BRIEF[13] and SURF[14]. Unfortunately they rely on the presence of identifiable features such as sharp edges and corners in the image. Since these sharp edges and corners are typically not present in images of human skin these methods are likely not as suitable for our purpose unless such features are introduced to the scene manually, such as by adding a pattern or markers.

Other SLAM methods, such as LSD-SLAM[15] and ElasticFusion[16] have found success by using 3D data from individual frames to track the movement of the camera. Because these methods do not make use of keypoints they might be more suitable for application in the case of tracking a camera while it observes human skin.

# 3 Stakeholders

There are a number of stakeholders in de development of the imaging system.

First there are the **patients** affected by psoriasis who have a stake in the emergence of any technology that can help provide better treatment. In order for the imaging system to be able to do this it is important that it is not only capable of accurate imaging but also that it performs imaging in a patient friendly manner. This means that there should be no skin contact during imaging and patients should not be required to assume uncomfortable positions for prolonged periods of time, imaging should be non invasive and fast. The eventual goal is to produce a product that can be used by patients for at-home imaging, the prototype is therefore designed with this goal in mind and should be easy to use at home by patients as well as cheap to produce. In addition the system should be designed so that imaging is robust to the varied conditions found in patient homes.

Another stakeholder are the **medical staff** involved in the assessment of psoriasis. The reconstructions produced by the imaging system should be accurate enough so that medical staff can make a correct assessment of the severity psoriasis based solely on the reconstructions, this way imaging can be performed at home and medical staff can perform assessment remotely.

**The Robotics and Mechatronics group** (RAM) of the University of Twente that is researching imaging of skin disease such as psoriasis is another stakeholder in this project. As part of this research, interest was expressed in also evaluating the possible benefits of pattern projection on stereo reconstruction quality. As mentioned before, the objective of this work is not yet to create a polished product that is ready for deployment but instead to explore options for stereo imaging of skin diseases and to provide a prototype for proof of concept. The results of this work and the prototype are also intended to be used as a stepping stone for further research and refinement. In consideration of this RAM has required that the image acquisition of the prototype is done using software used by their researchers, namely Matlab.

Finally there is me, the author of this work. I have a stake in this project insofar as the time frame of the project should fall within the timeframe allotted to the completion of my thesis work (a period of 7-8 months). At the end of this project a working prototype of the imaging system must be completed and evaluated.

# 4 Design requirements

The requirements that the prototype should meet follow directly from the needs of the stakeholders. Summing up the previous section these are the following:

- The system should provide images accurate enough for correct psoriasis assessment.
- Imaging should be non invasive and fast.
- The system must be employable under varied environmental conditions.
- There should be a projector included in the prototype imaging system to evaluate the benefit of pattern projection.
- Image acquisition must happen in Matlab.
- The total cost of the system should be low.
- The prototype should be producible within the time frame of a thesis project.

It should be noted that the prototype, which is intended for research purposes and to provide a proof of concept, and the final product, which is intended for at home assessment of psoriasis are two different things. In this section, only the design requirements of the prototype are considered. Nevertheless it is desired that the general technology and components used in prototype are also suitable to be used in the final product. Therefore the requirements of the final product that are closely related to the choice of technology and components, such as the requirement for non invasive and fast imaging, as well as the requirement that total cost is low, are also added to the list of requirements for the prototype.

Since some of these requirements are quite vague they will be defined more concretely before moving on to the design process.

## 4.1 The system should provide images accurate enough for correct psoriasis assessment

The principal goal of the prototype imaging system is to assist in the diagnosis of various stages of psoriasis. For psoriasis these stages are defined by the PASI (psoriasis area and severity index). This scoring method defines five stages for the severity of psoriasis; none, mild, moderate, severe and very severe [17]. This is the system used by medical professionals and therefore the images provided by the imaging system should be accurate enough to allow them (or possibly an automatic system) to distinguish between the various stages of the PASI score.

The PASI system works by assessing three factors for the affected skin separately. These factors are Erythema (redness), Induration (thickness) and Desquamation (scaling). The assessment for these factors are combined to obtain a single score for psoriasis severity.

### 4.1.1 Erythema

The study in [1] explores a more objective assessment of erythema in psoriasis. This study explored various machine learning approaches for automatic assessment of the erythema score from 2D images. The most successful of these approaches, a two-layer feed forward Neural Network, achieved an accuracy of 92% in producing same the Erythema score as given by a medical expert for a given image. The input information used for this classification method included texture patterns, color space information, and difference in brightness levels. In addition a color correction, by including a pattern with known colors, was used on the images in order to improve robustness to differences in image sensors. If the prototype image acquisition system is to be used to also accurately estimate erythema the reconstruction should therefore also contain color information.

### 4.1.2 Induration

The Study in [2] has ventured to objectify the categories of induration by measuring thickness of psoriasis affected skin through 3D scanning and comparing the results to PASI scoring done by medical professionals. They found the following values for thickness of skin corresponding to each of the factors of the PASI index for induration (gaps in the ranges of thickness are due to lack of measurement data for those ranges):

| PASI score for thickness by medical expert | Measured thickness of affected skin tissue |
|---|---|
| 1 (mild) | 0.032mm to 0.202mm |
| 2 (moderate) | 0.208mm to 0.410mm |
| 3 (severe) | 0.463mm to 0.689mm |
| 4 (very severe) | 0.911mm to 2.268mm |

The data shows the first three categories of Induration spanning about 0.2mm in thickness each. For correct assessment the imaging system must be able to measure thickness accurately enough to differentiate between these categories.

### 4.1.3 Desquamation

No studies were found that explored an objective evaluation of the Desquamation score of the PASI system through imaging. The descriptions for PASI define the degree of desquamation by “roughness” “thickness” and “coverage”. The roughness in this case indicates the size of the scaling texture, where larger scales result in a higher desquamation score. In the same way thicker scales indicate a higher desquamation score. The coverage refers to the degree to which the scales cover the total lesion area, where a higher degree of coverage results in a higher desquamation score. The roughness and coverage concern 2D texture information, it should be possible to estimate these from a 2D image from one of the stereo cameras provided that the resolution and sharpness of the image is sufficient to do so. Based on sample images from a manual for PASI scoring, the thickness difference between classifications of desquamation is estimated to be in the order of 0.1mm

The Induration score is based on the thickness of the entire psoriasis area, and the thickness of the scales considered for the desquamation score is also taken over all the scales in the area. This means that the standard deviation in depth measurement of individual stereo matched pixels does not need to be extremely low provided that the image is of sufficient resolution. This is because the measurement can be averaged over multiple measurements in the psoriasis affected region.

Based on these considerations it is estimated that a standard deviation of 0.1mm or better in depth measurement should be sufficient for correct assessment of Induration and scale thickness. In addition it is estimated, based on reference images used to illustrate desquamation assessment, that a point density of 20 pixels per square mm should be more than sufficient to show the texture of desquamation. Finally, as discussed before, color information is required for correct assessment of erythema.

To summarize, the following concrete requirements for the prototype imaging system are defined:

- The standard deviation of depth measurements of individual pixels should be at most 0.1 millimeter.
- The 3D reconstruction should include color information.
- the 3D surface reconstruction should have a point density of 20 pixels per square millimeter.

## 4.2 Imaging should be non invasive and fast

The focus is to develop a system for accurate 3D imaging of skin conditions such as psoriasis. In order to do this safely there should be no medical risks involved with the imaging itself. This means there should be no physical contact between the imaging equipment and the affected skin. Additionally there should be no strong airflow or optical radiation targeting the skin. Finally, patients should not be required to hold still for long periods of time and therefore imaging should be fast. A maximum imaging time of 3 seconds is taken as a reasonable amount of time to expect a patient to hold still without experiencing discomfort. To summarize: the following detailed requirements are defined:

- There should be no physical contact required with affected skin during imaging.
- There should be no strong (optical) radiation or airflow directed at the affected skin.
- The imaging process should not require more than 3 seconds to complete.

## 4.3 The system must be employable under varied environmental conditions.

The intention is to eventually develop a system that can be used in the home of a patient. As a result the system should be robust to the varied lighting conditions found in peoples homes. The lighting condition in which the system should function is hereby defined as a "well lit room without being exposed to direct sunlight", it is assumed that these lighting conditions can reasonably be expected to be present in most homes. This leads to the following requirement:

- The imaging system should function in a well lit room while not exposed to direct sunlight.

## 4.4 The total cost of the system should be low.

Traditionally the cost of medical imaging devices range in the thousands or tens of thousands of euros. However, in order for the imaging system to be given to patients for extended time to be used for at home imaging it must much cheaper than this to produce since a larger number of devices will be required (one per patient). After discussion with RAM this budget was defined to be 500 euros.

- The budget for the prototype imaging device is set at 500 euros.

## 4.5 Summary of concrete design requirements

The design requirements for the prototype imaging system have now been defined more concretely. The final list of requirements is shown below.

- The standard deviation of depth measurements of individual pixels should be at most 0.1 millimeter.
- The 3D reconstruction should include color information.
- The 3D surface reconstruction should have a point density of 20 pixels per square millimeter.
- There should be a projector included in the prototype imaging system to evaluate the benefit of pattern projection.
- There should be no physical contact required with affected skin during imaging.
- There should be no strong (optical) radiation or airflow directed at the affected skin.
- The imaging process should not require more than 3 seconds to complete.
- The imaging system should function in a well lit room while not exposed to direct sunlight.

- The budget for the prototype imaging device is set at 500 euros.
- The prototype should be producible within the time frame of a thesis project.
- Image acquisition must happen in Matlab.

# 5 Design Space Exploration

In this chapter an overview is given of the design choices that were made and the motivation behind them. First all functional parts that must exist within the imaging system to be able to meet the design requirements are named. The most important design choices are identified and possible options are then explored. Each option is assessed through scoring at the hand of a number of criteria. Finally a number of complete design configurations are evaluated based on the total score of their component parts and a motivated choice is made for a final design.

**Design questions and assessment criteria**
The design requirements lead to the following functional parts that must be present in the imaging system:

- A PC running Matlab (from the requirement that matlab must be used as acquisition software)
- A set of camera sensors in order to realise stereo imaging.
- A projector in order to evaluate the possible benefits of pattern projection.
- A way to connect components together, either through an embedded plaform or directly.
- A frame to hold the components, excepting the PC, together in a rigid stereo configuration.
- Software that manages acquisition, calibration, communication and image processing as well as a user interface for the operator to interact with the system.

The realisation of these parts requires a number of design choices to be made. The most important of which are listed below and will be explored in more detail.

- Choosing baseline distance and stereo matching method
- Choosing a camera sensor
- Choosing how to combine the components

Each of these choices has a number of possible options. In order evaluate which option is most suitable they will be scored, through best effort reasoning, on a scale of -3 to +3 for a number of weighted criteria. These criteria are the following:

**Accuracy (weight 12):** The most important goal of the imaging system is to be able to provide accurate 3D reconstruction of human skin so that psoriasis severity can be assessed correctly. As a result this criterion representing the accuracy of the 3D surface reconstruction compared to the ground truth will have the highest weighting factor of 12.

**Implementation time (weight 7):** The time that is available for the design, implementation and evaluation of a working prototype spans that of a single Msc. thesis project which is 1000 hours. In order to ensure that the project can be finished within this limited time it is important to somewhat limit the implementation complexity of the design. The weight for this criterion was determined to be 7 because it is considered somewhat less important than the reconstruction accuracy.

**Budget (weight 5):** Keeping the monetary cost of the the working prototype limited is important for this project because the goal is to eventually reach a product that is cheap enough to be used for at home imaging. This means that a patient must keep the device in their home for extended periods of time which means that many of these devices are likely needed, this is not feasible if the device is very expensive. While the budget is relevant it is not considered as important as being able to produce an accurate imaging device within the project time and therefore the weight is set at 5.

**Robustness (weight 5):** The final product is intended for at home use by patients. To facilitate the development process of reaching this product from the prototype. The prototype should be designed to be somewhat usable in varying environmental conditions and subject shapes. This criterion represents the degree to which the imaging device is robust to such changing conditions and can produce consistent good results. This criterion is considered of similar importance to the budget and the weight is set at 5.

## 5.1 Choosing baseline distance and stereo matching method

The choice of baseline distance between the cameras is of vital importance. A large baseline distance can lead to a good depth resolution of the measurement (a detailed explanation of this can be found in the section on Determining the baseline in the Implementation chapter). A large baseline distance also means that the viewpoints of the cameras will differ from eachother far more than in the case of a narrow baseline. This causes perspective distortion and possibly occlusions which negatively impacts the accuracy of stereo matching. Some stereo matching techniques are more suited to dealing with these effects than others which is why the choice of baseline distance and stereo matching method have been grouped together here in one design choice. A number of approaches will be presented and scored according to the criteria. Because this choice does not involve physical components the budget criterion is not relevant and marked as 0 for each approach.

**Wide baseline + Daisy descriptor:** The first aproach that will be considered here is a wide baseline approach with matching using the DAISY descriptor

[8]. This approach boasts the accuracy benefit of having a wide baseline while limiting the negative effects of perspective distortion and occlusion by using a stereo matching method that is robust to these effects. The software for the DAISY descriptor is openly available which makes the ease of implementation high. The DAISY descriptor is reported to be robust to occlusions resulting from wide baseline which should also make it more robust to possible occlusions and depth discontinuities that could be present in the subject.

**Narrow baseline + NCC:** The second approach to be considered is a narrow baseline approach with stereo matching on rectified images using normalised cross correlation. Normalised cross correlation is a common stereo matching technique that has shown numerous times in literature to be able to produce accurate stereo matches. A known downside of normalised cross correlation is that it does not perform well around depth discontinuities but considering that these should be limited for human skin surfaces this method should still be viable. Because NCC is a common matching algorithm implementations are readily available in Matlab and image processing libraries such as OpenCV, making the implementation time trivial.

| Platform | Accuracy (x12) | Implementation time (x7) | Budget (x5) | Robustness (x5) | Total score |
|---|---|---|---|---|---|
| Wide baseline + DAISY descriptor | 3 | 0 | 0 | 1 | 41 |
| Narrow baseline + NCC | 1 | 2 | 0 | 0 | 26 |

## 5.2 Choosing a camera sensor

The choice of camera sensor is especially relevant because its properties are so important to achieving accurate reconstruction. First of all a higher resolution of the camera directly leads to more accurate reconstruction. The same is true for a narrower field of view (results in higher angular resolution which leads to more accurate stereo matching) as well as a low minimum focus distance. In order for a camera sensor to be viable it must also be able to lock its focus for an extended period of time, so that the intrinsic parameters of the camera do not change between the calibration images and the measurement images. Some investigation was done into what camera sensors were available and those considered the best candidates are evaluated below according to the criteria.

**2x Waveshare IMX179 8MP USB Camera:** This camera module has a resolution of 8.2 MP(3288x2512), a field of view of 145 degrees. The resolution is good but the field of fiew is quite wide giving this option an average score for accuracy. The camera focus is controlled by physically rotating the lens on top of the module therefore locking the focus should straightforward. This, combined with its plug-and-play USB cable should make implementation time low. The cost of a single module is 50 euros making it one of the more expensive options here. Finally, the module has integrated hardware for automatic white balance, exposure and gain control which should make it robust to varying lighting conditions. It should be noted that these functions can also be detrimental if they cannot be turned off or controlled since they can cause different white balance and exposure levels between the two camera modules.

**2x Friendly Elec CAM1320:** This camera module has a high resolution of 13.2MP (4224x31336) with a relatively narrow viewing angle of 60 degrees, these properties result in a good score of 3 for accuracy. The focus of this camera module is controlled by a small motor that moves the lens through a magnetic force, this motor is software controlled and there is not much documentation about interfacing with it, resulting in a potentially high amount of time to get locked focus implemented. This camera module is quite cheap at 29 euros each so it scores high on the budget criterion. Finally, similar to the Waveshare, this camera boasts automatic white balance, exposure and gain functions resulting in a good score for robustness. This camera module has a CSI (Camera Serial Interface) connector instead of a USB connector, which would greatly limit the potential choice of embedded platforms since two of these connectors must be present to use this option.

**OV9750 USB stereo webcam:** This prebuilt stereo module features two cameras with a resolution of 2.5MP (2560x960) each and a field of view of 60 degrees. While the field of view is good the resolution is quite low, furthermore since this is a prebuilt stereo camera module designed mostly for 3D Reconstruction of larger scenes the baseline distance is quite small ( 60mm) and cannot be changed resulting in a very low score for accuracy. On the other hand, because of it is a plug and play prebuilt stereo camera this option is the easiest to implement. The stereo camera costs 70 euros making it a good choice budget wise. It boasts the same automatic exposure, gain and white balance options as the other cameras putting it on par with them when it comes to robustness.

| Platform | Accuracy (x12) | Implementation time (x7) | Budget (x5) | Robustness (x5) | Total score |
|---|---|---|---|---|---|
| 2x Waveshare IMX179 | 0 | 1 | -1 | 2 | 12 |
| 2x Friendly Elec CAM1320 | 3 | -2 | 2 | 2 | 42 |
| OV9750 USB stereo webcam | -3 | 3 | 2 | 2 | 5 |

## 5.3 Choosing how to combine the components

For the choice of how to combine the components three options will be considered.

**No embedded platform:** The first option is to connect the cameras and projector directly to the pc with long cables. The advantages of this approach are firstly that implementation time will be short since there are much better libraries and drivers available for a pc compared to most embedded platforms.

Another advantage is that, since no embedded platform has to be purchased at all, this is the cheapest option. A disadvantage is that this approach necessitates that long cables exist between the imaging device and the pc and that the pc must be placed close to the imaging device. This disadvantage is reflected in a low robustness score since the system will not be robust to being used in situations where a pc is not nearby and use of long cables can make the system harder to use.

**Single-Board-Computer:** The next option is the use of a single-board-computer (SBC) that connects to the cameras and projector. This is essentially a miniature pc running an embedded operating system (e.g. a raspberry pi). The advantage of using a SBC is that it also has access to many of the same high level libraries and drivers that are available to a pc, keeping implementation time within bounds although slightly higher than when not using an embedded platform since communication interfaces between the PC and SBC would have to be implemented. The single board computer is quite expensive, often in the 30-70 euro range, meaning it scores a bit lower on budget than the other options presented here. Finally this solution can be truly wireless, most SBC's have out of the box WiFi and Bluetooth support and can be battery operated, allowing the system to be small and portable, making it robust to be used in most home situations.

**Micro-controller board:** The final option is the use of a Micro-controller board that connects to the cameras and projector. The advantage of micro-controller boards is that they are quite cheap. Micro-controller boards do however have limited computational power and driver and library support which makes their implementation time the worst of all the options considered, if implementation is even feasible. Finally, through the use of a cheap WiFi module, this option could potentially be used wirelessly on battery power, earning it a similar robustness score to that of the SBC.

| Platform | Accuracy (x12) | Implementation time (x7) | Budget (x5) | Robustness (x5) | Total score |
|---|---|---|---|---|---|
| No embedded platform | 0 | 1 | 3 | 0 | 22 |
| Single-Board-Computer | 0 | 0 | -2 | 2 | 0 |
| Microcontroller-board | 0 | -2 | 0 | 2 | -4 |

## 5.4 Configurations

A number of configurations were created using the possible options for each of the design choices.

**Configuration 1:** This configuration is designed to determine the viability of creating the imaging system without an embedded platform. Since the best camera option, the FriendlyElec cam1320, which has the highest score for the camera sensor choice, requires two CSI connections that are not available on a desktop computer, the Waveshare IMX179 is chosen as the next best option. Finally a wide baseline + DAISY descriptor is chosen since this was the best option for the choice of baseline distance and stereo matching algorithm.

**Configuration 2:** This configuration is designed to achieve the best reconstruction accuracy. The FriendlyElec cam1320 was chosen along with a Wide baseline and DAISY descriptor for stereo matching since these options achieve the highest accuracy score. Finally, since the FriendlyElec cam1320's CSI connectors are not compatible with a desktop PC, the Single-Board-Computer was selected as next best choice for the embedded platform.

**Configuration 3:** This configuration is designed around being cheap and easy to implement. The prebuilt OV9750 USB stereo webcam was chosen. This camera sensor automatically comes with a narrow baseline. Finally the choice of using no embedded platform is made for this configuration since it is the cheapest and easiest to implement option.

The Table below shows an overview of each of the configurations and their component parts.

| Configuration | baseline and matching | Camera sensor | Embedded platform |
|---|---|---|---|
| Configuration 1 | Wide + DAISY descriptor | Waveshare IMX179 | No embedded platform |
| Configuration 2 | Wide + DAISY descriptor | FriendlyElec cam1320 | Single-Board-Computer |
| Configuration 3 | Narrow baseline + NCC | OV9750 USB stereo webcam | No embedded platform |

The table below shows the total scores for the configurations for each criterion as well as a total score. Configuration 2 is shown to have the highest total score. This configuration is therefore chosen as final configuration for the imaging prototype. This choice should result in accurate reconstructions with good robustness to environmental conditions, the downside being that implementation will be harder and parts more expensive than the other configurations.

| Configuration | Accuracy (x12) | Implementation time (x7) | Budget (x5) | Robustness (x5) | Total score |
|---|---|---|---|---|---|
| Configuration 1 | 3 | 2 | 2 | 3 | 75 |
| Configuration 2 | 6 | -2 | 0 | 5 | 83 |
| Configuration 3 | -2 | 6 | 5 | 2 | 53 |

# 6 Implementation

This section covers implementation of all the parts of the prototype. First the selection of suitable hardware is covered. Then the physical design and dimensions of the prototype imaging device are explained. Finally an the implementation of the various software components in the imaging system is discussed. For an overview of the theory of wide baseline stereo reconstruction using the DAISY descriptor refer to Appendix A.

## 6.1 Remaining Hardware selection and total cost

The single board computer selected to be used with the camera modules is the **NanoPi M4v2**, also by Friendly Elec. This board has connectors for two of our CSI camera modules and the software from Friendly Elec has built in support for the CAM1320 modules used which helps make software development easier. Furthermore the memory and processing speeds of this single board computer should be more than enough for the purpose of image acquisition. The NanoPi M4v2 costs 70 euros.

Finally for the projector a **picopix ppx5110** is used. This projector was readily available and supplied by the RAM group and a new one costs about 290 euros, although similar projectors are available nowadays for far less. This small battery powered projector is ideal for an embedded setup because of its short operating range, it has a projection resolution of 854x480.

This puts the total cost of the components at 428 euros. In addition to this about 40 euros of 3D printed models and assorted components such as cables and memory cards was used to produce the final prototype. This puts the total cost of the prototype at 468 euros, which is within the determined budget of 500 euros.

## 6.2 Physical Setup

### 6.2.1 Determining physical parameters

The operating distance, the distance between the cameras and the object, is an important parameter for achieving accurate 3D reconstruction. It would be preferable to have a short operating distance simply because having the sensor closer to the object means more pixels per surface area and therefore better stereo matching and higher depth resolution. After some experimentation with the camera modules, it was found that one of the two modules could focus correctly on an object at a distance of 160mm or further while the other could only perform a sharp focus if an object was at least 230mm away.

This means that for the modules used for the prototype the operating distance cannot be less than 230mm because a sharp image is required for accurate stereo matching. It would have been worthwhile to try and obtain a second camera module that could focus at 160mm but at the time of this project replacement camera modules were unfortunately not available within the required timeframe of the project. This was mostly due to the effects of the covid-19 pandemic on international shipping and manufacturing in China. As a result the decision was made to design the prototype around the requirement of an operating distance of 230mm.

With the operating distance known, a suitable baseline must be selected. A bigger baseline can, up to a certain point, lead to more accurate depth estimation. However, a high baseline causes occlusion and perspective deformation between the images because the viewpoints of the two cameras are far apart. These effects make it more difficult to perform accurate stereo matching.

Since the imaging system is intended for use on patches of human skin, which is are generally flat surfaces, it is assumed that there will be little to no occlusion even if a high stereo baseline is chosen. Since the DAISY descriptor is used to perform pixel matching. And this descriptor has been shown to be especially suitable for wide baseline dense stereo matching [9], it should be possible to still perform accurate stereo matching.

### 6.2.2 Determining the baseline

In order to a large overlapping region of the camera images the cameras are rotated towards each other so that the centers of each image converge at 230mm from the cameras (=the operating distance). The goal is to choose a baseline such that a good depth resolution is achieved, while still keeping the perspective distortions and occlusion within bounds. We define the depth resolution here as the difference in depth caused by shifting the stereo match by a single pixel along the epipolar line.

Consider the depth resolution of a stereo match with pixels exactly at the center of of each image. The geometry of this situation is shown in Figure 1.

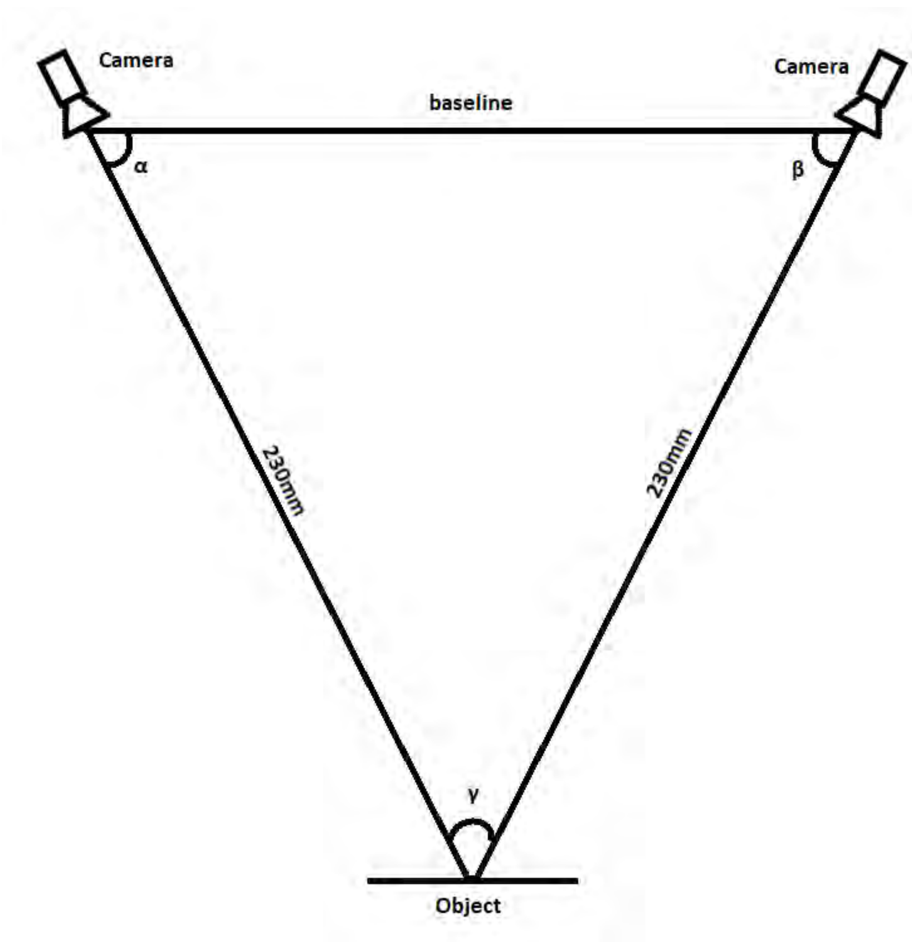


Figure 1

In this situation, the depth for the stereo match is 230mm, which is identical to the operating distance. The angles $\alpha$, $\beta$ and $\gamma$, are calculated as follows:

$\alpha = acos(1/2 * baseline/operatingdistance), \beta = \alpha, \gamma = 180 - \alpha - \beta$

The left camera here is taken as a reference, the depth resolution can be determined by observing the change in depth resulting from shifting the stereo match in the right camera by a single pixel. The cameras have a field of view of 60 degrees. It is assumed here that the epipolar line is horizontal in the image of the right camera, this is a reasonable assumption since this runs parallel to the baseline and the pixel matches are in the center of each image. Shifting the stereo match by a single pixel along the horizontal axis would also shift the angle of the ray projected out of the camera. Given 4224 pixels along the horizontal axis of the image with camera specifications of a 2.94mm focal length and a width of 4.8mm of the image sensor, the shift in angle $\beta$ due to a shift of a single pixel along the horizontal direction is calculated to be 0.0221 degrees, which we will call $\delta$.

The depth for this new situation can be calculated with:

$$Depth = baseline * sin(\alpha)/sin(\gamma) = baseline * sin(\alpha)/sin(180 - \alpha - \beta \pm \delta)$$

The depth resolution is defined as the average difference in depth resulting from shifting a single pixel along the horizontal axis in either direction, this is calculated as follows:

$$depth\_res = 1/2 * abs(baseline * sin(\alpha)/sin(180 - \alpha - \beta + \delta) - baseline * sin(\alpha)/sin(180 - \alpha - \beta - \delta))$$

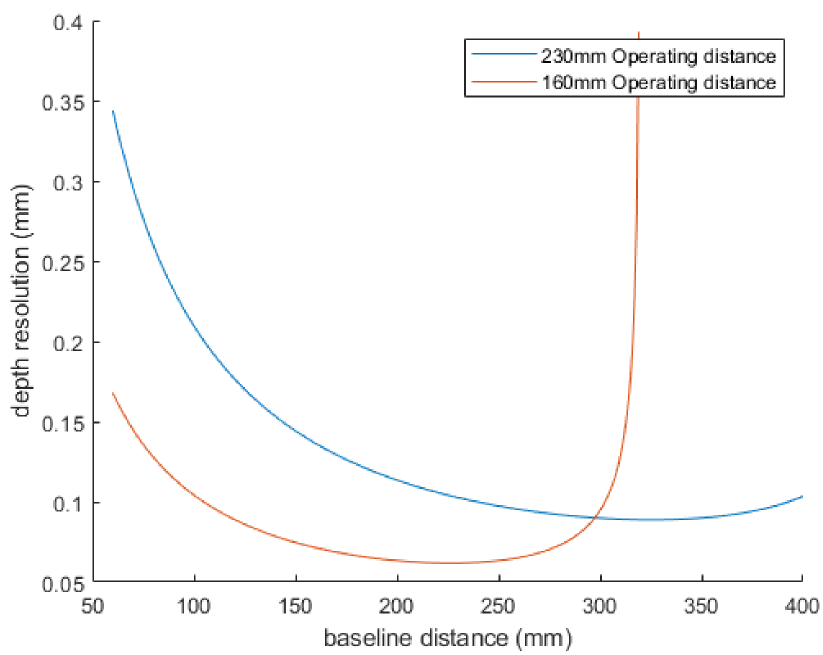


Figure 2: depth resolution as a function of baseline distance, for the situation of a pixel match at the center of each image at the operating distance. Plotted for operating distances of 230mm and 160mm

Figure 2 shows the depth resolution for the situation plotted as a function of the baseline distance. Looking at the plot for a 230mm operating distance. We can see that the best resolution is achieved when the angle $\gamma$ between the camera rays is exactly 90 degrees, which happens at a baseline of 325mm. The depth resolution here is 0.089mm. Because a baseline of 325mm might cause too much perspective distortion a baseline of 210mm is chosen instead. The depth resolution for a baseline of 210mm is 0.1mm. With this choice, and by implementing a sub-pixel resolution in stereo matching, the depth resolution can be further improved so that the standard deviation of depth measurement of actual measurements should be below 0.1mm as per the requirements set forth. It should be noted that the depth resolution will be worse when pixel matches are not in the center of the image as is considered here because the angle between the camera rays becomes smaller. It is therefore important when using the final system to make sure that the area of interest is as centered as possible in both images, and that objects are placed at the correct operating distance from the cameras.

Figure 2 also shows a plot of depth resolution versus the baseline distance for the case of an operating distance of 160mm. This curve shows that using a 160mm operating distance and a 150mm baseline would bring the depth resolution to 0.074mm while causing a similar amount of perspective distortion, it might therefore be interesting to order another camera module once the covid-19 pandemic is over and they are available again to see if this could be achieved. The asymptote where the depth resolution at 160mm operating distance approaches infinite seen in the figure occurs in the moment when the baseline distance becomes twice as large as the operating distance, the angle between the camera rays is 0 degrees at this point which no longer represents a realistic situation.

The final setup with all the measurements worked out can be seen in Figure 3. The imaging system will use a baseline of 210mm with the cameras rotated 62.8 degrees away from it, this means that the angle between the cameras is 54.4 degrees. For accurate reconstruction objects should be 204mm below the middle of the imaging system.

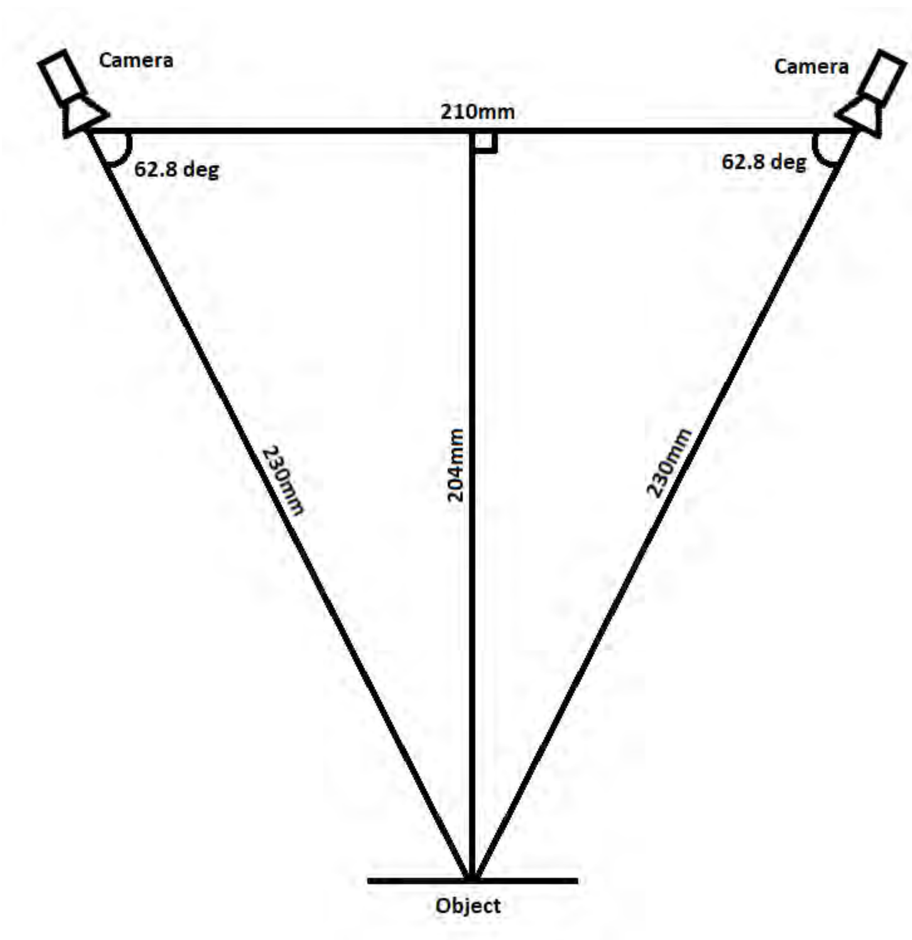


Figure 3: Stereo configuration

A physical design was 3D modeled and printed according to these specifications. The 3D model of the imaging system can be seen in Figure 4. The model is comprised of 9 components that are each designed with 3D printing in mind (no or minimal overhang and a flat base) which can be assembled together to form the prototype setup. The components (camera modules, embedded processing board and projector) have also been modeled and are mounted to the setup using screws. The final 3D printed result is shown in Figure 5. The legs are added separately and made out of threaded metal rods that can be rotated to control the height of the camera setup, this allows for objects of different heights to be placed under the system and imaged while still maintaining the operating distance of 230mm from the camera modules to the object surface.

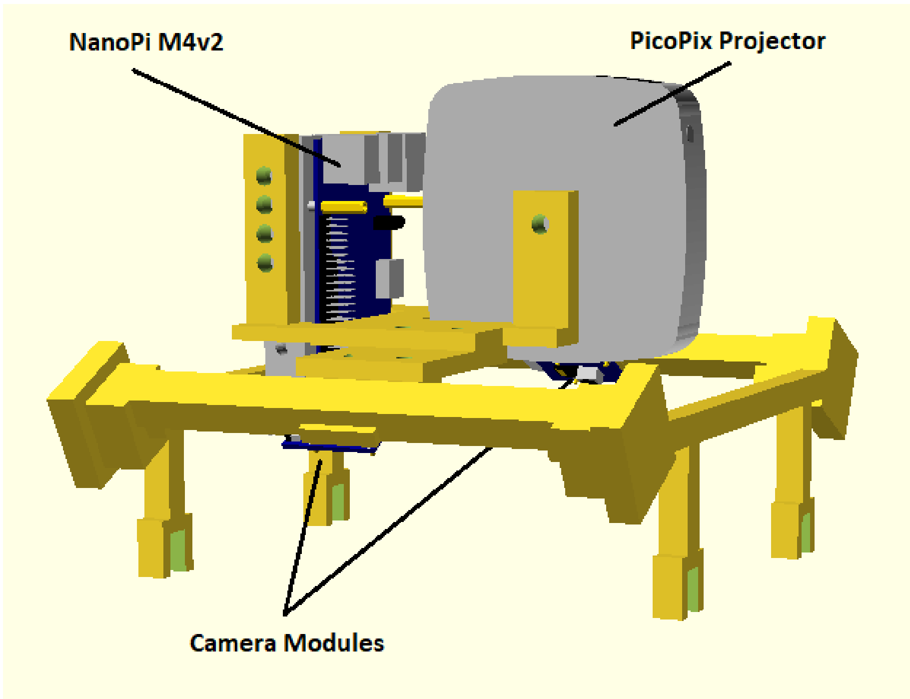


Figure 4: 3D model of the the imaging system

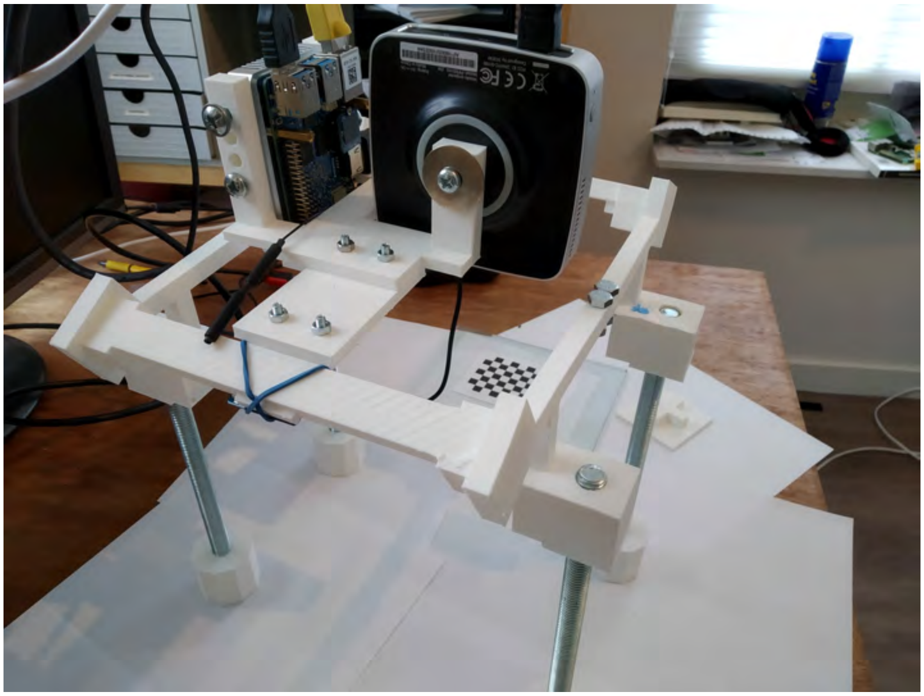

Figure 5: image of the the imaging system

### 6.3 Embedded Software

The NanoPI M4v2 that is going to run the acquisition software is compatible with two operating systems, Linux (Ubuntu) and Android. Unfortunately the Ubuntu distribution has no driver support for controlling the focus of the camera modules, which means Android, which does support controlling camera focus using the default Camera API, is used instead.

Given the requirement that image acquisition must be controlled within Matlab this means the embedded android application must be capable of the following things:

- Transmitting image data to and receiving commands from a pc running Matlab remotely.
- Focusing individual cameras and keeping the focus stationary afterwards for acquisition and calibration.
- Capturing images with both cameras simultaneously.
- Controlling the projector image.

An android application was developed in Java that realizes all of these functionalities. The data transmission was achieved by implementing a TCP/ip server for WiFi / Ethernet communication with Matlab. The individual focusing of cameras was achieved through use of the default Camera API of Android. Unfortunately it was not possible to directly control the focus (i.e. manually set a focus distance), instead it is only possible to perform an auto-focus operation on a single camera and then to keep this focus stationary. A drawback here is that the focus is lost whenever the application is closed or the system is powered down which means that calibration has to be repeated.

Simultaneous image capture was realized by running the capture software for each camera on separate parallel threads. The time between image captured of the separate cameras was measured by capturing images of a running stopwatch (which was updating at 60Hz), an average difference in capture time of 0.12s was obtained by averaging over 10 measurements.

Finally, the projector is connected to the main HDMI port of the NanoPi m4v2, allowing it to function as the main display of the android OS. This allows full control of the projected image using the UI of the android application. A UML diagram of the android application is shown in Appendix B (Figure 21).

The application works by launching a number of parallel threads. The network communication, image acquisition and UI display parts of the application all run on parallel threads with a single class “IOhandler” performing communication and synchronization between these threads. This parallelism means that the application can capture new images while the TCP server is transmitting previously captured images. The application also contains a TCP write buffer which can hold hundreds of images waiting to be transferred so the the user does not have to stop capturing images to wait for transmission.

The application is controlled through commands transmitted over the network from a Matlab application.

## 6.4 Matlab Application

The matlab application allows the user to connect to and control the imaging system remotely from a pc and acquire images. The matlab application was created with the Matlab App Designer and has a UI to help the user control the imaging system.

The matlab application UI is shown in Figure 6. A connection to the imaging system is opened by entering its IP address and pressing “Connect”, if connection is successful (some handshake messages are sent) the red circle next to the connect button will turn green and the imaging system can be used for image acquisition. There are buttons for focusing individual cameras, acquiring calibration images, data images and for controlling the projector mode. There is a button that will open the Matlab Stereo Calibration tool for all acquired calibration images. The area at the top (showing "image placeholder" in Figure 6) allows the user to see a preview of any images received from the imaging system, allowing the user to inspect the images as they are downloaded from the imaging system.

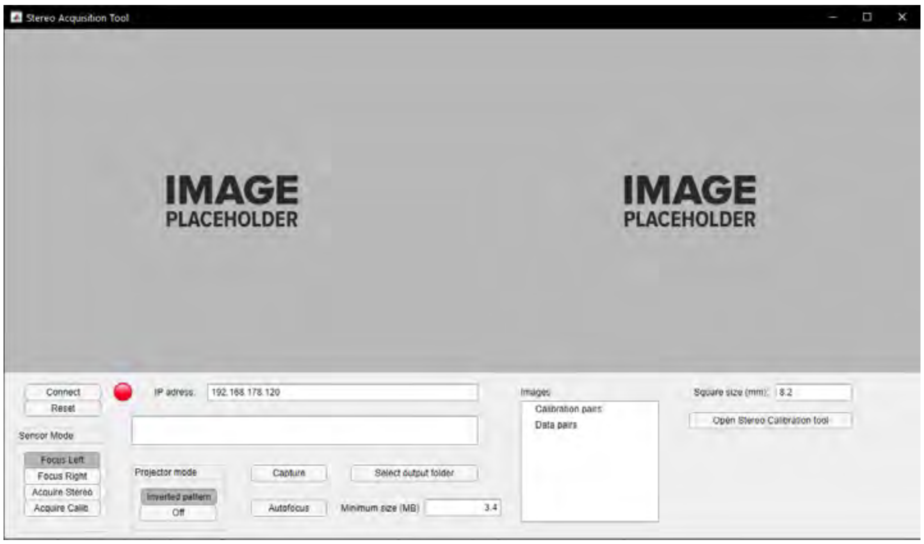


Figure 6: image of the UI of the Matlab application

## 6.5 Stereo Matching and Triangulation

This section discusses the implementation of stereo matching and triangulation. Before stereo matching all images are first undistorted using the Matlab function "*undistortImage*" and the results of the calibration. Then for each pixel in the reference image a search for a match in the other image is carried out along its epipolar line. Since there are approximately 13 million pixels in each image this process is quite resource intensive and time consuming. In order to speed up the process somewhat the initial search considers every 4th pixel, after which a more detailed search of every 2nd pixel is carried out near the best result. This is repeated, halving the pixel resolution each time, until a final match is found with a sub-pixel resolution of 1/4 pixel. The calculation is further sped up by making use of multiple CPU cores, this is done by splitting the reference image into multiple horizontal slices. The matching of each slice is then performed by a separate CPU thread.

Pixel comparisons are carried out by calculating the difference (error) between DAISY descriptors computed at the various pixel locations in the images. DAISY descriptors calculated along the epipolar line are rotated to be parallel to the epipolar line to improve matching The epipolar lines are not horizontal since no rectification is performed. The comparison that results in the lowest difference is considered the best match. DAISY descriptors are computed the same way as in [8][9] using C++ code provided by the authors.

Once pixel matches are found for each pixel in the reference image the 3D world locations are determined using the Matlab function "*triangulate*". Filtering and visualization of the results are implemented using the C++ library PCL (Point Cloud Library). The program CloudCompare can also be used for filtering and visualization.

# 7 Experiments

Due to the situation surrounding the covid-19 pandemic it was not possible to perform any imaging on psoriasis patients during this project. For the same reason it was also not possible to access lab equipment such as high quality 3D laser scanners to obtain ground truth reconstructions, this project was entirely performed at home as a result of quarantine restrictions.

Despite these limitations, a number of experiments were performed with the intention of making an effort of evaluating the performance of the prototype imaging system using available materials and tools, it is recommended that a more thorough validation of the system is performed in future work when quarantine restrictions are lifted.

The first experiment attempts to assess whether the imaging system can be used to measure and identify structures with a height difference in the order of 0.1mm, similar to those expected to be found on psoriasis affected skin. A second experiment was performed to assess the standard deviation in depth measurement. Thirdly, an experiment was performed to test the hypothesis that projection of inverted noise patterns can be used to improve reconstruction performance in the presence of specular reflection. Finally, 3D reconstruction is performed for human skin (not affected by psoriasis) in order to assess imaging systems capability of performing skin reconstructions.

All reconstructions performed in the experiments were performed with the same DAISY descriptor with a radius of 31 pixels, Radius Quantization of 8, Angular quantization of 8 and a histogram quantization of 12. The parameters of this DAISY descriptor were reached after some experimentation with different parameters and qualitative assessment of the results.

## 7.1 Calibration

Calibration for the experiments was performed using a 10x7 checkerboard pattern with squares of 8.4mm x 8.4mm. The re-projection error for the first experiment (7.1 Identification of small structures) was 0.92 pixels, all other experiments share the same calibration with a re-projection error of 1.11 pixels.

The extrinsic parameters obtained through the calibration show a baseline of 208mm and an angle between the cameras of 47.6 degrees. The original design specification of the setup was a baseline of 210 mm and an angle between cameras of 54.4 degrees. This means the physical setup differs from the design with an error 2mm ($\sim 1\%$) in the baseline distance and an error of 6.8 degrees ($\sim 14\%$) in camera angle. The 3D printed parts that make up the setup should be quite true to their model and the differences with the design are likely due to the necessary spacing and tolerances in the areas where parts lock into each other. It should be considered that possible measurement error of the calibration square size could also impact the baseline distance reported by the calibration. A $\sim 1\%$ error in the baseline distance is quite small relative to the total baseline distance should have little impact on reconstruction accuracy. The error in camera angle will reduce the overlap between the two camera field of views and should not impact the reconstruction quality.

## 7.2 Identification of small structures

The goal of this experiment is to identify whether small structures, with height differences in the order of 0.1mm can be measured and identified in the surface reconstruction created by the prototype imaging system. If these structures can indeed be measured and identified this bodes well for the measurement of Induration (thickness) of psoriasis affected skin.

For this experiment the height of small structures will be measured using the imaging system and the results evaluated. In order to create a sample containing such height structures small patches of regular printer paper are layered in top of each other and glued to a flat piece of cardboard. Five structures are created with each a different number of layers (1 to 5 layers), a texture pattern is printed onto the top layer to aid stereo matching. The thickness of a single sheet of printing paper is measured at 0.12mm, this measurement was performed by hand measuring a stack of 100 sheets of printing paper. An image showing the sample can be found in Figure 7.

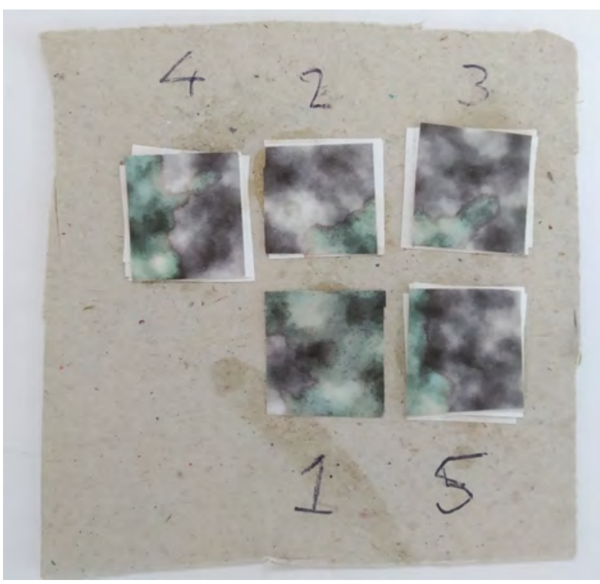


Figure 7: The sample featuring stacks of paper glued to piece of cardboard

Imaging and 3D reconstruction of the sample is performed using the imaging system resulting in the surface mesh shown in Figure 8.

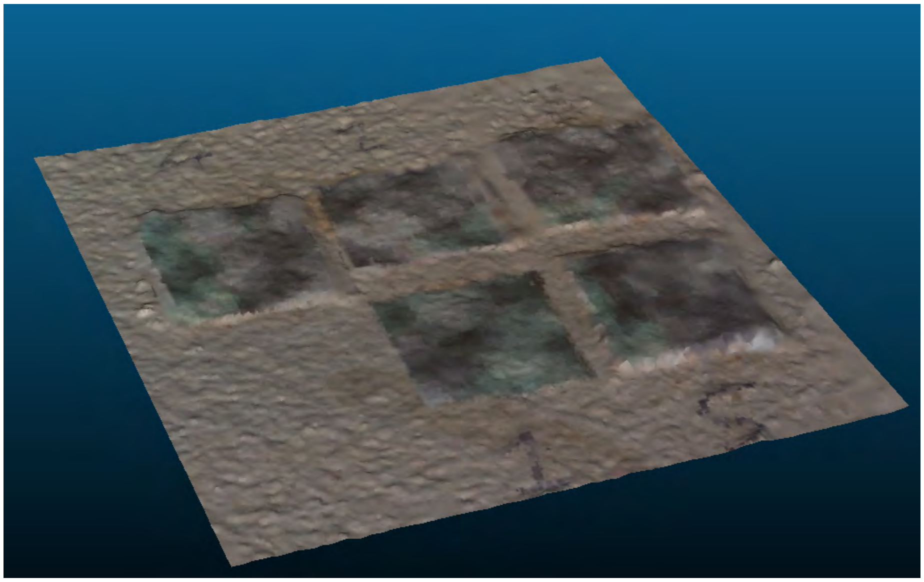

Figure 8: Mesh of reconstructed surface of the sample

Measurement of the height for each of the stacks of paper is performed by first segmenting the individual paper stacks as well as the cardboard background. Then a plane is fit to each of the segmented parts. The height of each stack of paper is then calculated by calculating the distance from the center of its plane (at the center of the stack) to the plane of the cardboard background.

The measured heights for each of the paper stacks are reported in the following table:

| Layers | Measured Thickness (mm) | Measured thickness per layer(mm) |
|---|---|---|
| 1 | 0.19 | 0.19 |
| 2 | 0.33 | 0.17 |
| 3 | 0.48 | 0.16 |
| 4 | 0.51 | 0.13 |
| 5 | 0.79 | 0.16 |

One thing to note is that the thickness for a single layer of paper is measured at being between 0.13mm and 0.19mm, this is more than the previously measured paper thickness of 0.12mm. Although care was taken to use as little glue as possible it is likely that the glue between layers adds to the thickness of the stack which can explain the measured thickness being higher than expected. The measurement of the stack containing four layers of paper is the biggest outlier with a lower layer thickness than the other measurements. Upon taking a closer look at the sample it was discovered that the cardboard base is very slightly curved downwards towards the side of the stack of four layers which could explain the outlier.

Due to the lack of resources as previously discussed there was no success in obtaining a measurement sample more representative of psoriasis affected skin. The results of this experiment nevertheless do show an upwards progression of stack thickness as the number of layers increase, which bodes well for the reconstruction accuracy of the system.

## 7.3 Standard deviation of depth

The goal for this experiment is to measure the standard deviation of depth measurements of the prototype imaging system.

In this experiment a 3D reconstruction is made of a sample featuring two flat surfaces with a height difference of 10mm between them. The point cloud from this reconstruction is segmented into two clouds, one containing pixels of the top surface and one containing the pixels of the bottom surface. A plane is then fitted to the pixels of the bottom surface. Finally the distance from each individual pixel of the top surface to this fitted plane is calculated.

The sample used for this experiment is a 3D print of a model that features a flat cross shaped structure that is 10mm higher than the base to which it is attached. The surface area of the cross shaped structure in the model is 700mm$^2$. After 3D printing the surface of the sample was 3D reconstructed using the imaging system. Figure 9 shows images of these stages. Imaging was done by placing the sample under the imaging system so that it appears centered in each camera image at a distance of 230mm in order to achieve maximum reconstruction accuracy.

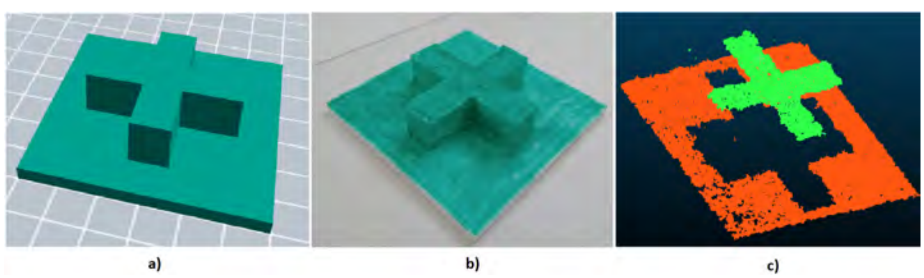


Figure 9: a) 3D model of the sample b) image of the 3D printed sample c) segmented point cloud of the surface reconstruction of the sample

For each of the 70070 pixels in the top cloud (green in Figure 9c) the distance to a plane fitted to the bottom cloud was calculated. The results are plotted in a histogram in Figure 10. the measurements follow a normal distribution around 10.20mm with a standard deviation of 0.085mm.

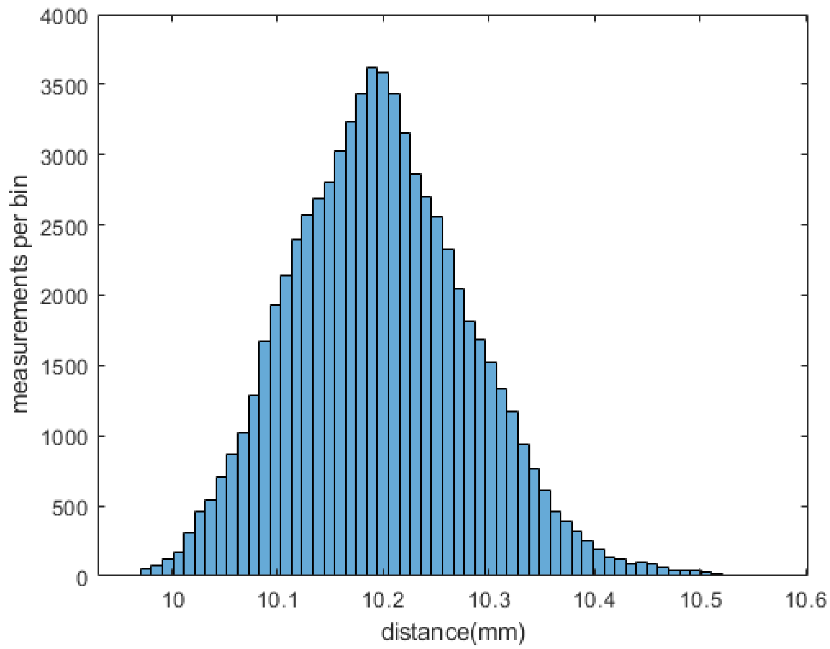

Figure 10: Histogram of distance measurements between the top and base of the sample, bins have a width of 0.01mm.

The measurement of 10.20mm differs 0.20mm from the height of 10mm to which the 3D model was designed. It is assumed that this difference is a result of the scaling factor introduced resulting from the error in calibration due to measurement error of the calibration square size or due to the error introduced by the 3D printing process. The surface of the sample is assumed as perfectly flat for this experiment, in actuality the surface has a very slightly rough texture resulting from its 3D printed nature, it should be noted that this could cause the measured standard deviation to be higher.

Stereo matching was performed with a sub-pixel resolution of 1/4 pixel. The effects of imperfect calibration, image noise, sub-pixel interpolation, and possibly lack of texture and sharpness in the image all introduce error in stereo matching which is therefore expected to be less accurate than 1/4 pixel.

During the design process, the requirement was set forth that the standard deviation of depth measurements should be below 0.1mm. This requirement has been met for the reconstruction performed in this experiment.

The thickness measurement required for Induration assessment is the mean thickness of a psoriasis affected patch of skin. A standard deviation of 0.085mm was measured for individual pixel reconstruction over 70070 samples. In order to calculate the standard deviation of the mean, representing the standard error for thickness measurement over an entire patch of skin, the standard deviation must be divided by the root of the number of independent repeated measurements. The 70070 individual points do not however constitute independent measurements. This is because each measurement not only considers one pixel in the reference image, but an entire region of the image by calculating the daisy descriptor, this means that measurement of pixels close to each other share an overlap in the region over which the daisy descriptor is calculated and are therefore strongly correlated. In order to determine a good estimate of what can be considered the number of independent measurements in the data, it is first considered that a circular daisy descriptor with radius 31 has an area of $A = \pi * 31^2 \approx 3019 pixel^2$. We now consider that given an area of 70070 pixels in the reference image, it is possible to perform about $70070/3019 \approx 23.2$ uncorrelated measurements (i.e. no overlap in daisy descriptor). This leads to a standard deviation of the mean of $0.085/\sqrt{23.2} = 0.018mm$.

It should be noted here that the use of a smaller daisy descriptor, i.e. one with lower radius, might lead to a more accurate mean which is something that could be considered in future work.

The mean thickness of categories of Induration severity were shown in [2] to be about 0.2mm apart. Since $0.018mmmm << 0.2mm$ the conclusion is drawn that the quality of depth measurement seen in this experiment is good enough for induration assessment according to the PASI score. While this is certainly a hopeful indication of the quality of the imaging system, it is important to note here that this experiment was conducted on a sample with a flat surface, little to no reflection, and fine texture. Conditions will likely not be as good for actual skin measurement. Therefore additional experiments, focusing on evaluating the quality of measurements of actual affected skin, will be required to draw a more concrete conclusion of the feasibility of induration assessment using the prototype imaging system.

The reconstructed point cloud of the cross shaped surface consisted of 70070 pixels, Since the surface area of this surface was 700mm$^2$, this results in a surface density of $\sim$100 $pixels/mm^2$, meeting the estimated requirement of $20 pixels/mm^2$ for assessment of desquamation.

## 7.4 Specular reflection with pattern projection

The goal of this experiment is to evaluate the following hypothesis:

**If the difference images are calculated between two sets stereo images of an object, one where a noise pattern is projected onto the object and one where a negative of that noise pattern is projected onto the object, stereo reconstruction of that difference image is more robust to specular reflection than the individual stereo image sets.**

The experiment is conducted by capturing images of a white household ceramic plate. The surface of this plate is highly reflective, to introduce specular reflection to the scene a desk lamp is positioned above the imaging system. Three sets of stereo images are captured. This first stereo image has the projector turned off and functions as a control image. For the second image the projector projects a perlin noise pattern onto the surface of the ceramic plate, for the third image the projector projects an identical noise pattern except inverted. The use of perlin noise patterns is motivated because perlin noise has fine grained non repeating texture. Fine grained texture is desired for stereo matching because it allows for accurate measurements. Repeating patterns in texture can cause multiple local maxima in stereo matching error which can lead to an incorrect match being selected. Figure 11 shows a magnified 100x100px section of the total 854x480px (=resolution of the projector)sized pattern that was used for this experiment.

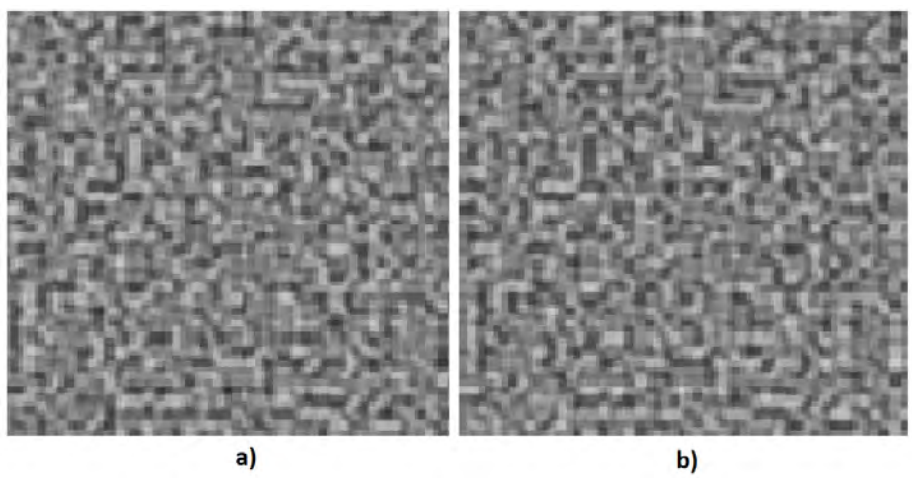


Figure 11: 100x100px section of projected perlin noise patterns, a) pattern b) inverted pattern

The idea is that, since both of the pattern images contain the same specular reflection, substracting one image from the other will effectively remove this reflection while at the same time amplifying the texture added by the noise pattern since the patterns are inverted between the images, thus improving reconstruction quality.

In order to test this the number of successfully reconstructed pixels in a specular reflective region of each image is compared.

Figure 12 shows the control images captured with the projector turned off, the area exhibiting specular reflection can be seen in the image from the right camera. Figure 13 shows the area of specular reflection for all three images as well as the difference image in which both pattern projections have been subtracted from each other. While the specular reflection is still visible in the in the difference image it is not nearly as pronounced as in the other images and does exhibit more texture, which is in line with expectations.

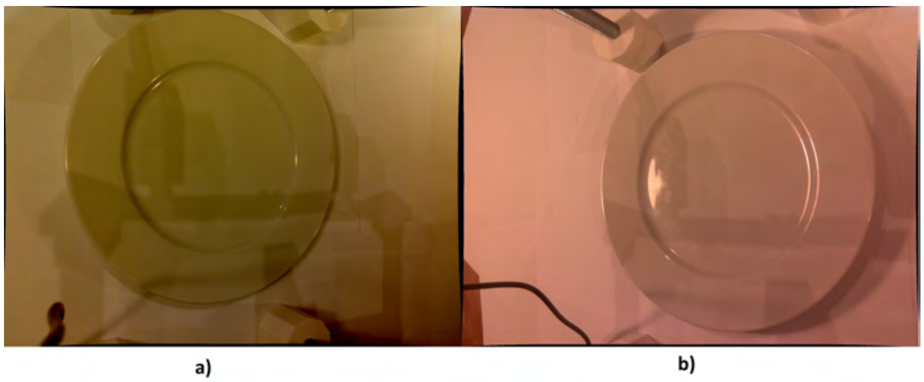


Figure 12: Stereo image of ceramic plate without pattern projection: a) image from left camera, b) image from right camera (stereo reference)

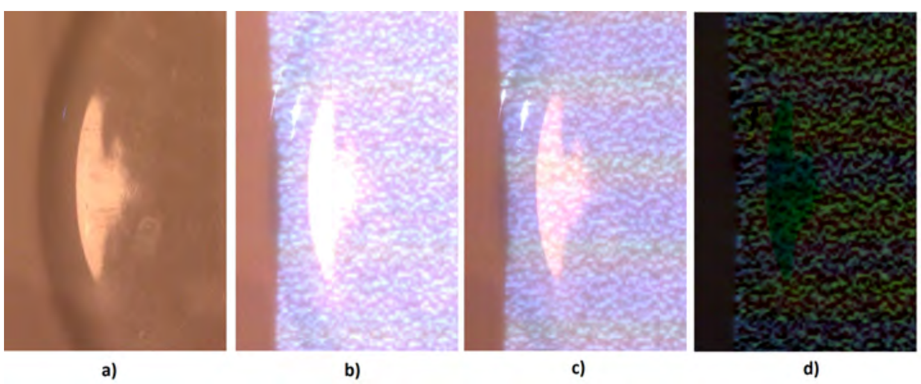


Figure 13: Area surrounding the specilar reflection in the image from the right camera: a) no projection b) pattern projection c) inverted pattern projection d) difference between the pattern projection images

Surface reconstruction was performed for each of the small image patches displayed in Figure 13. Figure 14 shows the resulting unfiltered point clouds of this reconstruction. The image patches are taken from an identical region of size 559x854 in each image. The point cloud reconstructions are at full density and each contain exactly 477.386 points.


Figure 14: Unfiltered point cloud reconstruction of each of the stereo images: a) no projection b) pattern projection c) inverted pattern projection d) difference between pattern images

The specific area featuring specular reflection is a mostly flat surface. A bounding box of 1mm thick and 15mm long and 15mm wide is manually positioned over the area in the point clouds where specular reflection occurs. For the purpose of this experiment, any points that fall within this bounding box are considered "accurate" points on the ceramic plate surface at the area of specular reflection. The cameras were not moved between capturing images and the bounding box is placed identical locations for each of the four point clouds that are considered. By counting the number of points for each of the point clouds that fall within this box some indication of the quality of reconstruction of the area of specular reflection is obtained.

Figure 15 shows the combined resulting point clouds for each image featuring only pixels that fall within the bounding box.

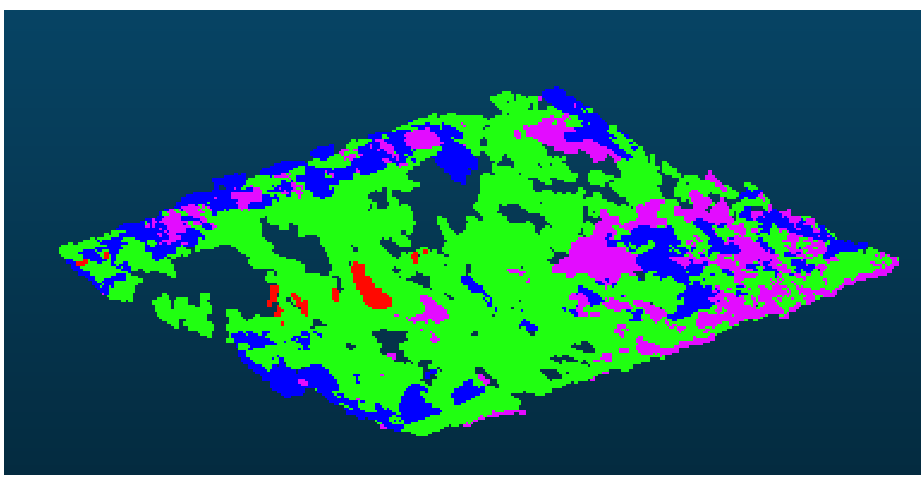

Figure 15: Point cloud reconstruction of the area of specular reflection: no projection (red), pattern projection (pink), inverted pattern projection (blue), difference between patterns (green)

The following table shows the number of pixels that fall within the bounding box for each point cloud reconstruction:

| Image | Number of points within bounding box |
|---|---|
| No pattern | 342 |
| Pattern | 13459 |
| Inverted pattern | 12985 |
| Difference image | 36918 |

The reconstruction from the image without pattern projection shows almost no points within the bounding box. Likely this is due to the fact that there is almost no texture present on the ceramic plate used in this experiment making stereo matching wildly inaccurate. The data for both pattern projections shows close to thirteen thousand points within the bounding box, however, looking at figure 15 it should be noted that most of these points are in the edge region of the box, where the specular reflection is least pronounced. Finally the data for the difference image shows the greatest number of points at about 3 times that of the individual pattern images. This, in addition to the fact that figure 15 shows the difference image having correctly matched points even in the area of most reflection indicates that the hypothesis is likely correct and the use of inverted pattern projection does indeed improve stereo surface reconstruction quality of objects with large specular reflection.

### 7.5 reconstruction of human skin

In this experiment 3D skin reconstruction is performed on a human hand in order to obtain an impression of the imaging systems ability to perform 3D skin reconstruction.

For this experiment the same projection procedure is used as in the previous experiment. A human hand is placed under the imaging system and 3 sets of stereo images are captured in rapid succession. One set of stereo images with the projector turned off to capture color information of the skin surface. The other two sets of stereo images are captured with the projector projecting a perlin noise pattern and its inverse respectively. The images featuring the noise pattern and its inverse pattern are then subtracted from each other to obtain a difference image. Figure 16 shows the result.

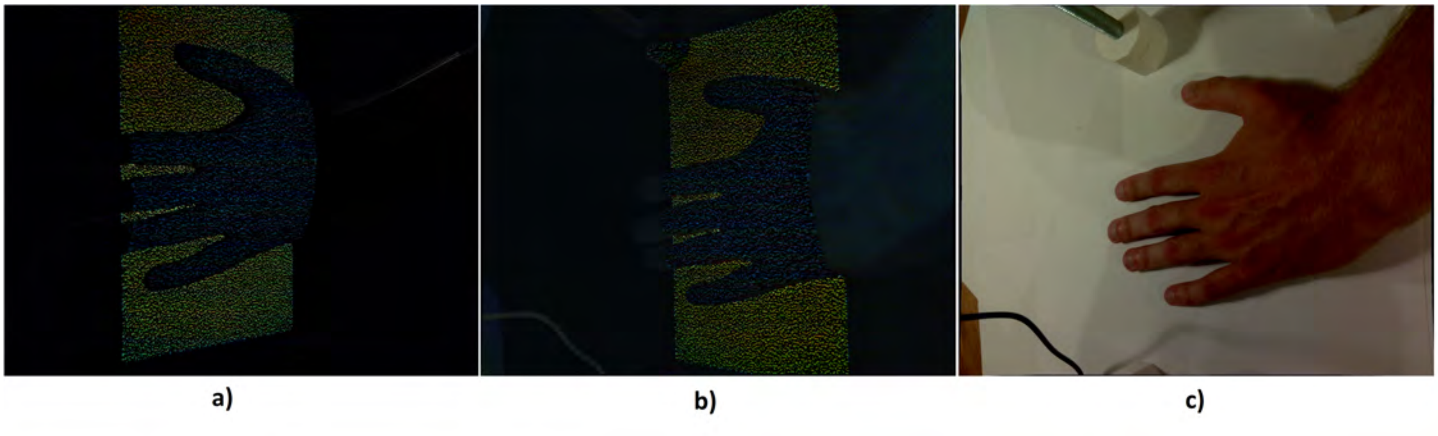


Figure 16: Stereo images of human hand, a) left camera difference image, b) right camera difference image, c) right camera image with no projected pattern

The difference images are used as input for stereo matching and the image without pattern projection is used to obtain color data for all the resulting 3D points. The resulting point clouds (with and without color data) are shown in Figure 17.

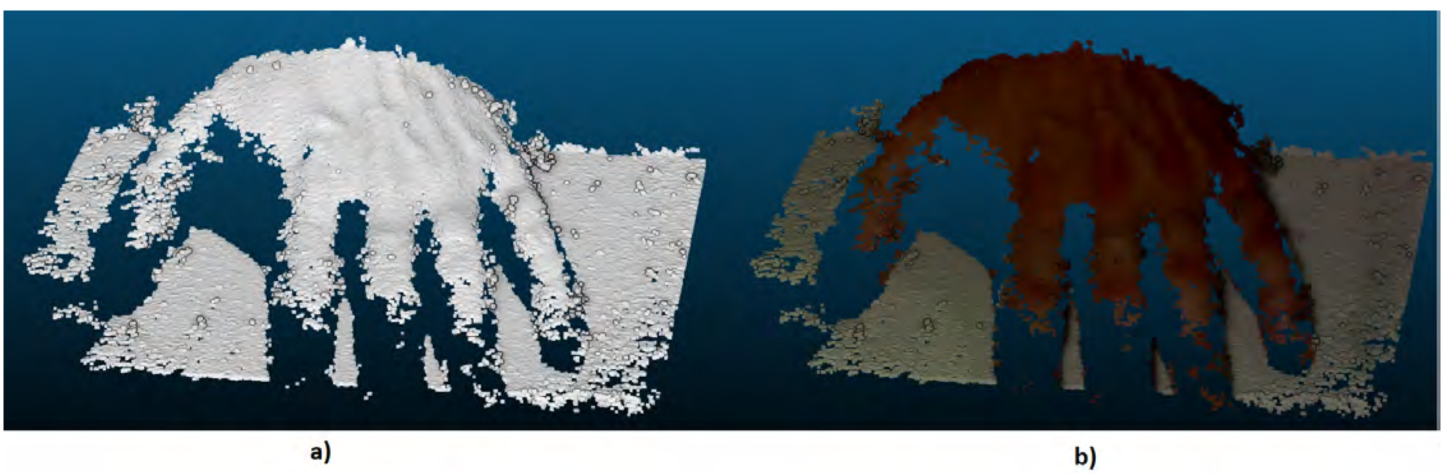


Figure 17: Reconstructed point cloud of human hand and background surface, a) without color, b) with color.

The point cloud is quite dense and the reconstructed surface is smooth and clearly shows the 3D features like tendons and veins that are present on the hand. This gives an impression of the systems ability to perform 3D skin reconstruction.

# 8 Discussion

This chapter presents a general discussion of the project and its results. In the first section the original design requirements are reexamined to evaluate if they were sufficiently met. Next the biggest limitations that were encountered during the project are discussed. Finally some recommendations for future work are presented.

## 8.1 Evaluation of the results

First the list of design requirements of the imaging prototype is reconsidered to determine which requirements were met. This list is reiterated below:

- The standard deviation of depth measurements of individual pixels should be at most 0.1 millimeter.
- The 3D reconstruction should include color information.
- The 3D surface reconstruction should have a point density of 20 pixels per square millimeter.
- There should be a projector included in the prototype imaging system to evaluate the benefit of pattern projection.
- There should be no physical contact required with affected skin during imaging.
- There should be no strong (optical) radiation or airflow directed at the affected skin.
- The imaging process should not require more than 3 seconds to complete.
- The imaging system should function in a well lit room while not exposed to direct sunlight.
- The budget for the prototype imaging device is set at 500 euros.
- The prototype should be producible within the time frame of a thesis project.
- Image acquisition must happen in Matlab.

During experimentation 3D surface reconstruction was achieved with a standard deviation in depth of 0.085mm, a surface pixel density of $\pm 100\ pixels/mm^2$ and color information. While this is a promising result that indicates that the imaging quality might be sufficient for the first three requirements of the imaging system to be met. Further work would be required to definitively conclude that these requirements were indeed met since no experimentation was performed on actual human skin.

During the experimentation process it was found that the imaging quality was more dependent on lighting conditions than expected. In particular, imaging without the projector was found to require a strong light source such as a desk lamp (e.g. at 100cm) pointed at the object in order to correctly auto-focus and obtain good texture information. In addition, it was found that for imaging with the projector turned on excessive light, e.g. from indirect sunlight through drapes on a bright day, would overexpose the image, causing texture information to be lost. While the imaging system can certainly be used effectively, it must be concluded that the requirement of robustness to lighting conditions was not met.

The imaging system that was implemented within the time frame of the thesis project and includes a projector. Imaging requires no physical contact with skin. The complete imaging process, where 3 stereo pairs are captured (two with inverted pattern projection and one without) takes about two seconds. The total cost of the imaging system prototype was within budget. And Image acquisition was implemented in Matlab. In short, all other requirements were met.

In summary, out of the eleven requirements originally set forth, seven requirements were met, the evaluation of three requirements showed promising but as yet inconclusive results, and one requirement was not met.

An additional experiment was conducted to evaluate the hypothesis that inverted pattern projection could be used to improve the reconstruction quality of surfaces exhibiting large specular reflection. During this experiment a single image featuring specular reflection was reconstructed and it was observed that inverted pattern projection caused an increase of three times as many points correctly stereo matched in the area of the image where the specular reflection occurred. While the scope of this experiment was not extensive enough to conclusively prove that the hypothesis was correct, it does show promising results to indicate that it is.

## 8.2 Limitations

A number of limitations and setbacks were encountered during this project. The most significant of these were due to the effects of the COVID-19 pandemic that occurred during this project. As a result of this pandemic, the manufacture and shipping of electronics suffered huge delays of several months. In addition, access to laboratory equipment for ground truth measurement as well as access to psoriasis patients to use as test cases was not available due to the quarantine in effect.

Another limitation was encountered in the driver quality of the camera modules that were used in this project. Because of the limited functionality of these drivers it was not possible to directly control the focus distance of the cameras using software, it was only possible to use the auto focus function. Since the motors in the camera module that control the focus distance are magnetic and require constant power to maintain focus, the focus is also lost whenever the module is powered down. This means that every time the imaging system is rebooted, the camera focus has to be set again using the auto focus function, because doing this also changes the intrinsic parameters of the cameras the calibration also has to be repeated. While this is already very inconvenient when the imaging system is used for experimentation by researchers, it would be unacceptable in a product used for at home psoriasis assessment by patients since they cannot be expected to have the knowledge to perform correct focusing and calibration.

## 8.3 Recommendations for future work

The first recommendation for future work is to perform a more exhaustive evaluation of the prototype imaging systems capabilities. This could be accomplished by using the system to reconstruct the surface of test samples of which accurate ground truth data is available for comparison. In addition, reconstruction of psoriasis affected skin should be performed using actual patients as test cases. The measured PASI assessment criterion of thickness should then compared to assessment by medical professionals. Finally medical professionals should be presented with the obtained 3D images of psoriasis affected skin and consulted on whether they consider these images of sufficient quality to perform PASI score assessment.

In order to improve the quality of reconstruction it is recommended to replace the magnetic motors controlling the focus of the camera modules with lenses of which the focus can be manually adjusted. This would prevent the need for refocusing and recalibration every time the system is powered down.

Finally, further work should be conducted to build on the results of this thesis and eventually realize an imaging device for at-home imaging and assessment of psoriasis by patients.

# 9 Conclusion

A low cost stereo camera imaging system for 3D reconstruction of human skin for quantitative assessment of psoriasis severity was designed and a prototype implemented. Initial experiments looking into the performance of the the prototype for general surface reconstruction have shown promising results that indicate that the system is might be accurate enough to be used for quantitative psoriasis assessment. The imaging system was not yet validated for actual psoriasis affected skin tissue and it is therefore recommended that further experimentation is performed to confirm this indication. In addition, a short range projector was included in the prototype imaging system and experimentation was conducted to evaluate whether the projection of noise patterns could be used to improve the 3D reconstruction of stereo optical systems on subjects exhibiting specular reflection and low texture. Results of initial experimentation indicate that noise pattern projection can indeed be used to improve 3D reconstruction in the presence of specular reflection. The results of this work can be used as a stepping stone in the development of an at-home imaging device for psoriasis assessment to be used by patients.

# References


[1] S. Banu, G. Toacse, and G. Danciu, "Objective erythema assessment of psoriasis lesions for psoriasis area and severity index (pasi) evaluation," *2014 International Conference and Exposition on Electrical and Power Engineering (EPE 2014), 16-18 October, Iasi, Romania.*

[2] A. F. M. Hani, H. Fitriyah, E. Prakasa, H. Nugroho, S. Hussein, and A. M. Affandi, "Objective assessment of psoriasis lesion thickness for pasi scoring using 3d digital imaging," *In Proceedings of World Academy of Science, Engineering and Technology, vol. 63, pp. 109–115 (2010).*

[3] J. H. Kim, B. Y. Kim, J. W. Choi, S. O. Kim, H. S. Lee, K. C. Park, and S. W. Youn, "The objective evaluation of the severity of psoriatic scales with desquamation collecting tapes and image analysis.," *Skin research and technology, vol. 18, pp. 134-150, 2012.*

[4] S. Banu, L. Dascălu, and G. Toacşe, "Tnternet of things in psoriasis assessment and treatment," *4th Doctoral Conference on Computing, Electrical and Industrial Systems Costa de Caparica, Portugal. p, pp.172-179, Apr 2013,.*

[5] M. Callieri, P. Cignoni, P. Pingi, R. Scopigno, M. Coluccia, G. Gaggio, and M. Romanelli, "Derma: monitoring the evolution of skin lesions with a 3d system," 2003.

[6] T. Y. Satheesha, D. Satyanarayana, M. N. G. Prasad, and K. D. Dhruve, "Melanoma is skin deep: A 3d reconstruction technique for computerized dermoscopic skin lesion classification," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. PP, pp. 1–1, 01 2017.

[7] H. Skvara, P. Burnett, J. Jones, N. Duschek, P. Plassmann, and J.-P. Thirion, "Quantification of skin lesions with a 3d stereovision camera system: validation and clinical applications," *Skin Research and Technology 2012;0: 1–9.*

[8] E. Tola, V. Lepetit, and P. Fua, "Daisy: A fast local descriptor for dense matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence Volume 32 Issue 5*, 2010.

[9] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 5, may 2010.*

[10] R. Mur-Artal and J. D. Tardos, "Orb-slam2: an open-source slam system for monocular, stereo and rgb-d cameras," *IEEE Transactions on Robotics (Volume: 33, Issue: 5, Oct. 2017).*

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,vol.50, No. 2, 2004, pp.91-110.*

[12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: and efficient alternative to sift or surf," *IEEE International Conference on Computer Vision, 2011.*

[13] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," *European Conference on Computer Vision, 2010.*

[14] H. Bay, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding, vol.110, No.3, 2008, pp. 346-359.*

[15] J. Engel, T. Schops, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," *European Conference on Computer Vision 2014 pp 834-849.*

[16] T. Whelan, S. Leutenegger, R. F. Salas-Moreno, B. Glocker, and A. J. Davison, "Elasticfusion: Dense slam without a pose graph," *Robotics: Science and Systems 2015.*

[17] R. G. Langley and C. N. Ellis, "Evaluating psoriasis with psoriasis area and severity index, psoriasis global assessment, and lattice system physician's global assessment.," *Journal of the American Academy of Dermatology, vol. 51, no. 4, pp. 563 – 569 (2004).*

[18] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence — December 2000, Vol 22: pp. 1330-1334.*

# 10 Appendix A: Theory

Explanation of relevant theory.

## 10.1 The pinhole camera model

The most common model used to represent a passive optical sensor (camera) is the pinhole camera model. A representation of this model can be seen in Figure 18. This model presents what is essentially a camera obscura where light from an object passes through a tiny hole so that a projection is formed on the image plane behind the hole.

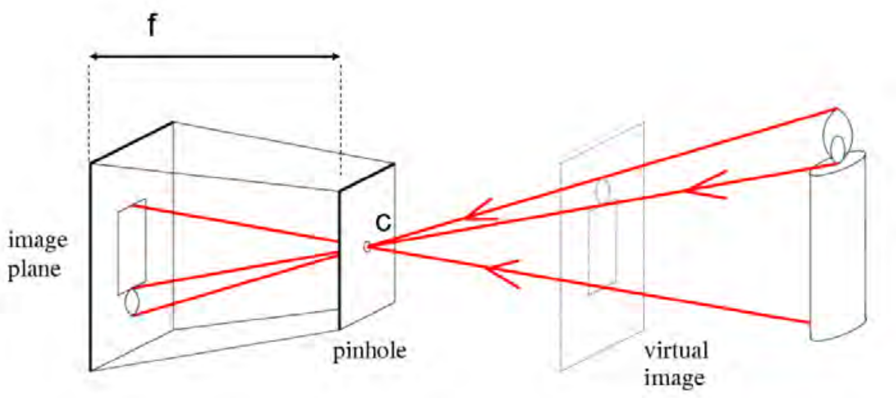


Figure 18: Pinhole Camera (source: http://vision.stanford.edu)

The coordinate frame is chosen such that the origin is precisly in the center of the pinhole $\boldsymbol{c}$, this results in the image plane being located at $Z = -f$ where $f$ is the focal length of the camera. If the real world coordinates $X, Y$ and $Z$ of a point $P$ are known, its projection $P'$ onto the image plane will have the coordinates:

$x = -f\frac{X}{Z}$ and $y = -f\frac{Y}{Z}$

Because image coordinates are often given in pixel coordinates scaling factors $s_x$ and $s_y$ should be added to convert to pixel distances. In addition, because pixel coordinates are always in the positive quadrant rather than centered around the origin, a translation in the coordinate frame ($u_0$ and $v_0$) needs to be added. This results in the following equations for the pixel coordinates of $P'$ in the image:

$x = s_x f\frac{X}{Z} + u_0$ and $y = s_y f\frac{Y}{Z} + v_0$

The x and y coordinates here have been mirrored along the origin to obtain the coordinates on the virtual image as can be seen in Figure 18. This has the advantage of providing a non mirrored image of the object similar to that produced by a modern camera. Finally a factor $\alpha$ is added to account for any skew between the coordinate systems.

$x = s_x f \frac{X}{Z} + \alpha \frac{Y}{Z} + u_0$ and $y = s_y f \frac{Y}{Z} + v_0$

Using homogeneous coordinates this results in the following transformation matrix, also known as the internal camera matrix $\boldsymbol{K}$:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim Z \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} s_x f & \alpha & u_0 & 0 \\ 0 & s_y f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \boldsymbol{K} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

## 10.2 Calibration

In order to obtain the values of the various parameters making up the internal camera matrix a calibration procedure has to be performed. There are numerous ways of doing this but the most commonly used is the method as described by Zhengyou Zhang in [18]. A brief summary of this method is given below.

Zhangs method requires that a planar pattern of known dimensions, such as a checkerboard pattern printed on paper, is photographed from various different poses. And that certain points (such as the corners of the checkerboard) be located on the resulting image. The following equation describes the transformation from a point in 3D world coordinates (in m or mm) to 2D image coordinates (in pixels).

$$C \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \boldsymbol{K} \begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

The term $\begin{pmatrix} \boldsymbol{R} & \boldsymbol{t} \end{pmatrix}$ denotes a transformation matrix that translates and rotates a point from the world coordinate system to the camera coordinate system of the pinhole camera, this matrix is known as the extrinsic matrix. C is an arbitrary scale factor.

When one of the corners of the planar pattern are selected as the origin of the world coordinate system, and the plane of the pattern selected as Z = 0, this results in all calibration points having a Z coordinate of 0 and then the following equation can be used

$$C \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \boldsymbol{K} \begin{pmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{r_3} & \boldsymbol{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \boldsymbol{K} \begin{pmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix}$$

By solving the set of equations resulting from equating the 3D world coordinates of their corresponding 2D pixel coordinates the 3x3 matrix $\boldsymbol{K}\begin{pmatrix}\boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t}\end{pmatrix}$ is known upto an arbitrary scale factor for each calibration image.

$$\boldsymbol{H} = \lambda\boldsymbol{K}\begin{pmatrix}\boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t}\end{pmatrix} = \begin{pmatrix}\boldsymbol{h_1} & \boldsymbol{h_2} & \boldsymbol{h_3}\end{pmatrix}$$

$\boldsymbol{H}$ is known as an homography where $\lambda$ is the arbitrary scale factor. This results in two equations that provide constraints on the intrinsic camera matrix $\boldsymbol{K}$:

1) $\boldsymbol{h_1^T K^{-T} K^{-1} h_2} = 0$

2) $\boldsymbol{h_1^T K^{-T} K^{-1} h_1} = \boldsymbol{h_2^T K^{-T} K^{-1} h_2}$

Since each calibration image with known homography provides two constraints on the intrinsic camera matrix $\boldsymbol{K}$ which has 5 degrees of freedom atleast 3 calibration images are required to find a unique solution for $\boldsymbol{K}$, if the skew cooefficient $\alpha$ is assumed to be 0 only 2 images are needed. Once $\boldsymbol{K}$ is known the rotation and translation for each calibration image can also be calculated. See [18] for more details.

## 10.3 Stereo Geometry

Once both cameras of a stereo camera setup have been calibrated so that their intrinsic matrices as well as the rotational and translational transform between their camera centers is known, it is possible to provide a constraint on corresponding pixel points between the images of the stereo setup.

Figure 19 shows what is known as the epipolar geometry. In this image we can observe two cameras in a stereo setup, with $O_1$ and $O_2$ being the camera centers and $e_1, e_2$ the epipoles defined by the location where the image planes intersect the line intersecting $O_1$ and $O_2$ (the baseline). A single real world point $P$ is projected onto the left camera image as point $p_1$ and onto the right camera image as point $p_2$. It should be noted that all these points, $P, p_1, p_2, O_1$ and $O_2$ as well as the epipoles always lie on the same plane in 3D space, the epipolar plane. It is therefore possible that when only the camera centers and the location of a single pixel $p_1$ on one image is known, one can calculate the location and orientation of the epipolar plane. The projection of this plane onto the other image is a straight line that must contain the pixel matching $p_1$. This line is known as an epipolar line because it always intersects the epipole.

Because of the epipolar geometry of a stereo camera setup, it is only required to search along the epipolar line when pixel matching between the stereo images. The math required to find the epipolar line will now be explained.

We define the origin of the world coordinate frame as the origin of the first camera coordinate frame and the second camera coordinate frame, as being

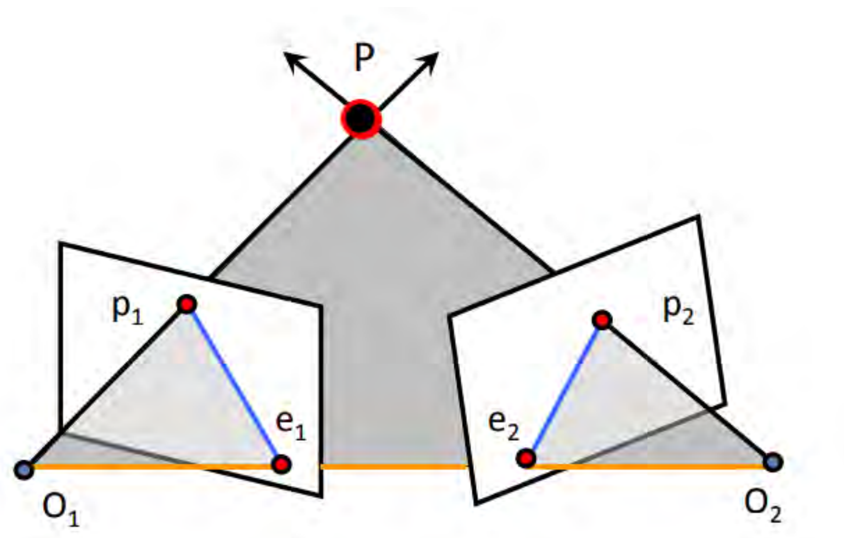


Figure 19: Epipolar Geometry

translated by $T$ and rotated by $\boldsymbol{R}$ within the real world coordinate system. The (homogeneous) image points $p_1$ and $p_2$ have to be transformed back to world coordinates by multiplying with the inverse camera matrix:

$\boldsymbol{K}_1^{-}1p_1 = p_1'$ and $\boldsymbol{K}_2^{-}1p_2 = p_2'$

This means that the vector $p_2'$ is $\boldsymbol{R}p_2$ in the world coordinate system. The vector $O_2$ is $T$ in the world system.

Since these vectors $\boldsymbol{R}p_2$ and $T$ both lie within the epipolar plane their corssproduct $T \times \boldsymbol{R}p_2$will produce a vector normal to the epipolar plane. Given that $p_1'$ also lies on the epipolar plane:

$p_1'T \cdot [T \times \boldsymbol{R}p_2'] = 0$

Using a matrix notation of the crossproduct we get:

$p_1'T \cdot [[\boldsymbol{T}x]\boldsymbol{R}p_2'] = p_1'^T\boldsymbol{E}p_2' = p_1^T\boldsymbol{K}_1^{-T}\boldsymbol{E}\boldsymbol{K}_2^{-1}p_2 = p_1^T\boldsymbol{F}p_2 = 0$

Where $\boldsymbol{E}$ is the Essential matrix and $\boldsymbol{F}$ is the Fundamental matrix. $\boldsymbol{F}$ is a 3x3 matrix with 7 degrees of freedom (one lost because $detF = 0$ and one lost due to the use of homogeneous coordinates), requiring at least 7 point correspondences to be determined if the cameras are uncalibrated. Alternatively if the cameras are already calibrated, determining $\boldsymbol{E}$ is enough to also determine $\boldsymbol{F}$, $\boldsymbol{E}$ only has 5 degrees of freedom because the internal camera parameters are already known and therefore requires fewer point correspondences to determine.

Once F is known it is possible to calculate the epipolar lines using the following simple equations:

$l_1 = \boldsymbol{F} p_2$, $l_2 = \boldsymbol{F}^T p_1$

Where $l_1$ is the epipolar line on the left image corresponding to $p_2$ on the right image and $l_2$ is the epipolar line on the right image corresponding to $p_1$ on the left image.

## 10.4 the DAISY descriptor

This section provides a brief summary definition of the DAISY descriptor that is used for stereo matching in this work, for a more detailed definition please refer to [9] and [8].

In order to compute the DAISY descriptor first orientation maps, containing the image gradient in a certain direction at each location, are computed over the image in a number of H quantized directions. Each of these H orientation maps is then convolved with Q Gaussian kernals, each with different $\sum$ values. This results in a total of Q * H convolved orientation maps of the image.

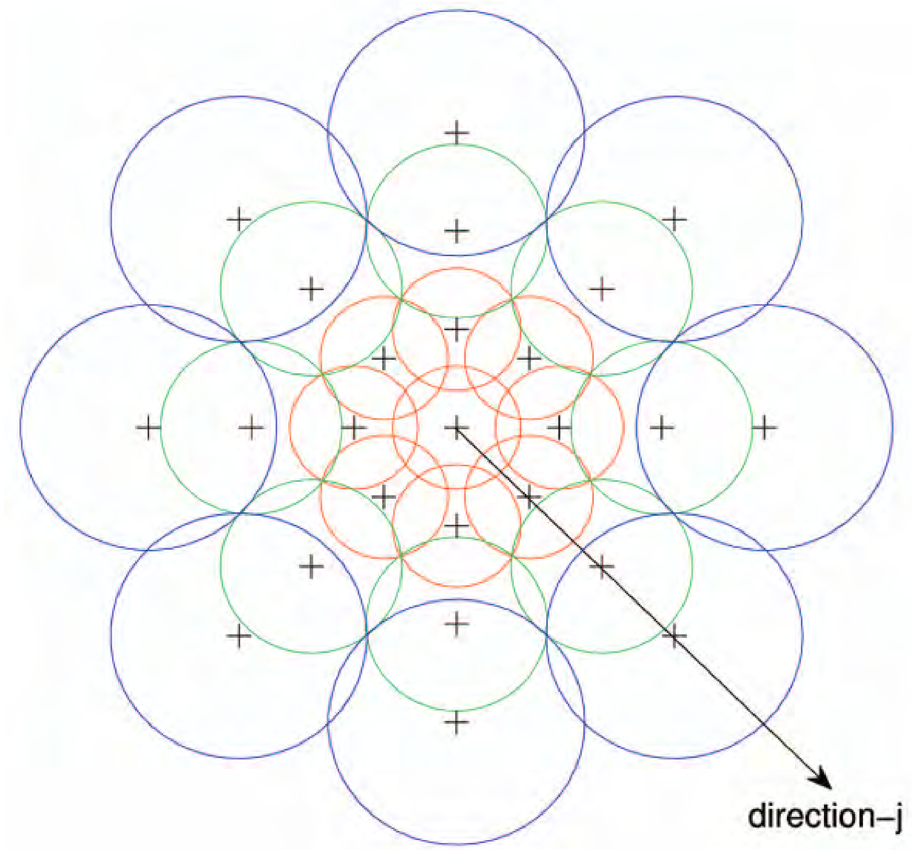


Figure 20: The DAISY descriptor, image source: [9]

Figure 20 shows an image of the DAISY descriptor. As can be seen in the Figure the DAISY descriptor is obtained by sampling the convolved orientation maps at different locations around a pixel (sampling locations are marked in the figure with a +). The orientation maps that were convolved with larger gaussian kernels represent a data over a larger image region and these are therefore sampled when at greater distance from the center pixel (kernel size is represented in the figure by the radius of the circles).

At every location a the convolved orientation is sampled in every of the H directions. The number of angular directions in which sampling is performed around the pixel is denoted as T (Angular Quantization). In the case of the DAISY descriptor in figure 20, sampling is done once at the center pixel and then at 3 distances; therefore Q=3. The sampling in the figure is done in 8 directions (e.g. direction-j in the figure) for each distance; therefore T = 8. Finally at each of these T*Q +1 = 25 sampling locations the convolved orientation maps are sampled for all the the H orientations. This means that the total number of data points in the descriptor = (Q*T +1) * H.

Besides, Q, T and H the last parameter of the DAISY descriptor is its radius. This is defined as the distance between the center pixel and the outermost sample point.

If the convolved orientation maps are pre-calculated for an image, the computation of a DAISY descriptor then only requires sampling of these maps. The DAISY descriptor can be easily rotated by rotating the sampling locations, and can be evaluated at sub-pixels by interpolating the convolved orientation maps.

## 10.5 triangulation

Once a set of matching points has been obtained through stereo matching, the real world location of the 3D landmark can be determined through triangulation. The triangulation method used in this thesis is Least Squares Error estimation. In short the algorithm works as follows:

Let $\mathbf{p}_1$ and $\mathbf{p}_2$ be the matching points in the stereo images. suppose an estimate $\hat{\mathbf{X}}$ of the 3D position of the landmark.

predictions of the pixel locations in the images of $\hat{\mathbf{X}}$ are then calculated:

$$\hat{}\underline{\mathbf{p}}_1 = \mathbf{K}_1({}^1\mathbf{R}_0\hat{\mathbf{X}} +^1 \mathbf{t}_0)$$

$$\hat{}\underline{\mathbf{p}}_2 = \mathbf{K}_2({}^2\mathbf{R}_0\hat{\mathbf{X}} +^2 \mathbf{t}_0)$$

after converting to inhomogenous coordinates mean squared reprojection error of the estimate is calculated with:

$$SE = \frac{1}{2}(\| \mathbf{p}_1 - \hat{\mathbf{p}}_1 \|^2 + \| \mathbf{p}_2 - \hat{\mathbf{p}}_2 \|^2)$$

The $\hat{\mathbf{X}}$ that minimizes the SE is known as the Least squares error estimate and is considered, according to by algorithm the best estimate of the 3D position of the landmark.

# 11 Appendix B: Images


Figure 21: UML diagram of the embedded android application