

Aleksi Partanen

COMPARATIVE ANALYSIS OF 3D- DEPTH CAMERAS IN INDUSTRIAL BIN PICKING SOLUTION

Master of Science Thesis
Faculty of Engineering and
Natural Sciences
Niko Siltala
Jyrki Latokartano
May 2023

ABSTRACT

Aleksi Partanen: Comparative analysis of 3D- depth cameras in industrial bin picking solution
Master of Science Thesis
Tampere University
Master's Degree Programme in Automation Technology
May 2023

Machine vision is a crucial component of a successful bin picking solution. During the past few years, there has been large advancements in depth sensing technologies. This has led to them receiving a lot of attention, especially in bin picking applications. With reduced costs and greater accessibility, the use of machine vision has rapidly increased. Automated bin picking poses a technical challenge, which is present in numerous industrial processes. Robots need perception from their surroundings, and machine vision attempt to solve this by providing eyes to the machine. The motivation behind solving this challenge is the increased productivity, enabled by automated bin picking.

The main goal of this thesis is to address the challenges of bin picking by comparing the performance of different 3D- depth cameras with illustrative case studies and experimental research. The depth cameras are exposed to different ambient conditions and object properties, where the performance of different 3D- imaging technologies is evaluated and compared between each other. The performance of a commercial bin picking solution is also researched through illustrative case studies to evaluate the accuracy, reliability, and flexibility of the solution. Feasibility study is also conducted, and the capabilities of the bin picking solution is demonstrated in two industrial applications.

This research work focuses on three different depth sensing technologies. Comparison is done between structured light, stereo vision, and time-of-flight technologies. The main categories for evaluation are ambient light tolerance, reflective surfaces, and how well the depth cameras can detect simple and complex geometric features. The comparison between the depth cameras is limited to opaque objects, ranging from shiny metal blanks to matte connector components and porous surface textures. The performance of each depth camera is evaluated, and the advantages and disadvantages of each technology are discussed.

Results of this thesis showed that while all of the technologies are capable of performing in a bin picking solution, structured light performed the best in the evaluation criteria of this thesis. The results from bin picking solution accuracy evaluation also illustrated some of the many challenges of bin picking, and how the true accuracy of the bin picking solution is not dictated purely by the resolution of the vision sensor. Finally, to conclude this thesis the results and future suggestions are discussed.

Keywords: Bin Picking, Depth Camera, Machine Vision, Stereo Vision, Structured Light, Time-of-Flight

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Alexi Partanen: 3D- syvyyskameroiden vertaileva analyysi teollisen kasasta poimintasovelluksen näkökulmasta

Diplomityö

Tampereen yliopisto

Automaatiotekniikan diplomi-insinöörin tutkinto-ohjelma

Toukokuu 2023

Konenäkö on keskeinen osa automatisoitua kasasta poimintasovellusta. Syvyyskamerateknologiat ovat kehittyneet paljon kuluneiden vuosien aikana, joka on herättänyt paljon keskustelua niiden käyttömahdollisuuksista. Kustannusten alenemisen, sekä paremman saatavuuden myötä konenäön käyttö, erityisesti kasasta poimintasovelluksissa onkin lisääntynyt nopeasti. Automatisoitu kasasta poiminta kuitenkin omaa teknisiä haasteita, jotka ovat läsnä lukuisissa teollisissa prosesseissa. Motivaatio automatisoidun kasasta poiminnan taustalla on tuotettavuuden kasvu, jonka konenäkö mahdollistaa tarjoamalla dataa robotin ympäristöstä.

Tämän diplomityön tavoitteina on vastata kasasta poiminnan haasteisiin vertailemalla erilaisten 3D-syvyyskameroiden suorituskykyä tapaustutkimusten sekä kokeellisen tutkimuksen avulla. Syvyyskameroiden toimintaa arvioidaan erilaisissa ympäristöissä sekä erilaisilla kappaleilla, jonka seurauksena 3D-kuvaustekniikoiden suorituskykyä vertaillaan keskenään. Työn aikana arvioidaan myös kaupallisen kasasta poimintasovelluksen suorituskykyä, jossa tutkitaan tapaustutkimusten avulla sovelluksen tarkkuutta, luotettavuutta sekä joustavuutta. Tämän lisäksi sovelluksen toimintaa pilotoidaan, ja ratkaisun ominaisuuksia demonstroidaan kahdessa teollisessa sovelluksessa.

Tämä diplomityö keskittyy kolmeen eri syvyyskameratekniikkaan. Vertailu tehdään strukturoidun valon, stereonäön sekä Time-of-Flight tekniikoiden välillä. Arvioinnin pääkategoriat ovat ympäristön valoisuus, geometrinen muotojen havainnointikyky, sekä heijastavat pinnat. Syvyyskameroiden välinen vertailu rajoittuu läpinäkymättömiin kappaleisiin, jotka vaihtelevat kiiltävistä metalliaihiosta mattapintaisiin liitinkomponentteihin ja huokosiin pintarakenteisiin.

Tutkimuksen tulokset osoittivat, että vaikka kaikki tekniikat kykenevät automatisoituun kasasta poimintaan, strukturoitu valo suoriutui tutkituista teknologioista parhaiten. Kasasta poimintasovelluksen tarkkuuden arviointi havainnollisti myös sen monia haasteita, sekä kuinka sovelluksen todellinen tarkkuus ei riipu ainoastaan syvyyskameran resoluutiosta. Loppupäätelmien lisäksi työ päätetään ehdotuksilla tutkimuksen jatkamiseksi.

Avainsanat: Kasasta poiminta, Konenäkö, Stereonäkö, Strukturoitu valo, Syvyyskamera, Time-of-Flight

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

PREFACE

I could not have undertaken this journey without my mentors, Niko Siltala and Jyrki Latokartano. Your expertise and feedback helped me push forward and get the best possible outcome for this thesis.

Many thanks also to the TeknoHub- project (ESR, project S22569) and all the companies involved with this thesis work, who enabled me to work with a truly interesting topic of mine. I am also grateful for the TREDIH- project (EAKR, project A78368), which provided research equipment for this thesis.

Lastly, I'd like to thank my family and friends who have been there to support me, especially through the long weeks of the writing process with proofreading and editing help. I could not have made it so far without all your great advice.

All my years of study until this moment, starting from a basic degree in automation back in vocational school have been some of the best times of my life. I hope that you, the reader can find this thesis as interesting as I did, and may it be useful in future research.

Tampere, 22th of May 2023

Aleksi Partanen

CONTENTS

1.INTRODUCTION	1
1.1 Research objectives.....	1
1.2 Research problem and research questions	2
1.3 Limitations.....	2
1.4 Research methods.....	2
2.LITERATURE REVIEW.....	5
2.1 3D- Machine vision system	5
2.2 3D- imaging technologies.....	7
2.3 Bin picking	13
3.RESEARCH PLAN.....	18
3.1 3D- depth camera performance evaluation	18
3.2 Proof-of-concept- demonstration in industrial environment.....	20
3.3 Model-based bin picking solution accuracy evaluation	21
4.RESEARCH HARDWARE AND SOLUTION DESIGN	23
4.1 Research objects	23
4.2 Research equipment.....	25
4.3 Solution design	28
4.4 Depth camera calibration	29
5.RESEARCH WORK.....	31
5.1 Comparison between 3D- depth cameras	31
5.2 Pilot scale demonstration in industrial environment.....	45
5.3 Photoneo Bin Picking Studio accuracy.....	51
6.RESULTS AND DISCUSSION.....	59
6.1 Depth camera performance with reflective materials	59
6.2 Depth camera performance with different object properties.....	60
6.3 Depth camera performance with ambient light	61
6.4 Performance of Photoneo bin picking studio	63
7.CONCLUSIONS.....	67
REFERENCES.....	69

LIST OF FIGURES

Figure 1.	<i>Research plan</i>	4
Figure 2.	<i>3D- data structures (Based on Ahmed et al., 2018, p. 3, Figure 1)</i>	6
Figure 3.	<i>Stereo vision system (Based on DAQRI, 2018, Figure 1)</i>	8
Figure 4.	<i>Structured light vision system (Based on DAQRI, 2018, Figure 1)</i>	9
Figure 5.	<i>Structured light projection patterns (Geng, 2011, p. 133, Figure 3)</i>	10
Figure 6.	<i>ToF- depth camera system (Based on DAQRI, 2018, Figure 1)</i>	11
Figure 7.	<i>Laser triangulation system (Based on Mohammadikaji, 2020, p. 10, Figure 1.3)</i>	12
Figure 8.	<i>Traditional bin picking application (Photoneo, 2018, p. 1)</i>	13
Figure 9.	<i>Bin picking application tasks (Based on Ojer et al., 2022, p. 4, Figure 1)</i>	14
Figure 10.	<i>Model-based bin picking solution (Based on Photoneo, 2018, p. 9)</i>	15
Figure 11.	<i>Coordinate reference systems (Torres et al., 2022, p. 8, Figure 7)</i>	16
Figure 12.	<i>Ferromagnetic objects</i>	23
Figure 13.	<i>Connector components</i>	24
Figure 14.	<i>Subassembly components</i>	24
Figure 15.	<i>Universal Robots UR5 robot manipulator (Universal Robots, 2016)</i>	25
Figure 16.	<i>Photoneo Bin Picking Studio (Photoneo, 2020, p. 8, Figure 4)</i>	25
Figure 17.	<i>Photoneo PhoXi scanner</i>	26
Figure 18.	<i>Basler Blaze 101 ToF- camera (left) and SICK Visionary-S Stereo vision camera (right)</i>	26
Figure 19.	<i>Basler acA1300-60gm area scan camera</i>	27
Figure 20.	<i>Research environment</i>	27
Figure 21.	<i>Robot program framework</i>	28
Figure 22.	<i>Bin Picking Studio path planning restrictions</i>	29
Figure 23.	<i>Photoneo depth camera calibration</i>	30
Figure 24.	<i>Simple geometrical shapes imaged with different depth cameras</i>	32
Figure 25.	<i>Rectangular shapes imaged with different depth cameras</i>	33
Figure 26.	<i>Polygon shapes imaged with different depth cameras</i>	34
Figure 27.	<i>Cylinder shapes imaged with different depth cameras</i>	35
Figure 28.	<i>Cylinder shapes imaged with different depth cameras. Cylinders orientated parallel to the structured light patterns</i>	36
Figure 29.	<i>Cylinder shapes imaged with different depth cameras. Cylinders orientated perpendicular to the structured light patterns</i>	37
Figure 30.	<i>Point clouds from semi-finished products</i>	38
Figure 31.	<i>Point clouds from connector components</i>	39
Figure 32.	<i>Point clouds from subassembly components</i>	40
Figure 33.	<i>Point clouds from imaging multiple objects at once</i>	41
Figure 34.	<i>Different bin configurations</i>	42
Figure 35.	<i>Bin configuration test results in ambient light conditions</i>	43
Figure 36.	<i>Ambient light test results with SICK Visionary-S</i>	44
Figure 37.	<i>Ambient light effects on the point cloud density</i>	45
Figure 38.	<i>CAD- model matching to the point cloud of the scene</i>	46
Figure 39.	<i>Workflow of the first bin picking solution</i>	47
Figure 40.	<i>Workflow comparison of two different approaches</i>	48
Figure 41.	<i>Workflow of the second pilot demonstration</i>	49
Figure 42.	<i>Localization with oily and dirty objects</i>	50
Figure 43.	<i>Metal flakes and magnetic gripper</i>	50
Figure 44.	<i>System configuration for bin picking solution accuracy evaluation</i>	51
Figure 45.	<i>Binary images of the objects</i>	52
Figure 46.	<i>Effects of the robot gripper to the grasp accuracy</i>	53

Figure 47.	<i>Results of random and fixed orientation grasps.....</i>	<i>54</i>
Figure 48.	<i>Displacement of a cylindrical metal blank.....</i>	<i>55</i>
Figure 49.	<i>Displacement of semi-finished product.....</i>	<i>56</i>
Figure 50.	<i>Displacement of large connector component</i>	<i>57</i>
Figure 51.	<i>Displacement of small connector component</i>	<i>58</i>
Figure 52.	<i>Resource intensity & impact graph.....</i>	<i>66</i>

LIST OF SYMBOLS AND ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
API	Application Programming Interface
BLOB	Binary Large Object
BPS	Bin Picking Studio
CAD	Computer Aided Design
CCS	Camera Coordinate System
FoV	Field of View
HDR	High Dynamic Range
IR	Infra-Red
LED	Light Emitting Diode
MAE	Mean Absolute Error
MPI	Multi Path Interference
NIR	Near-infrared
OOTB	Out-Of-The-Box
PLA	Poly Lactic Acid
RANSAM	Random Sample Matching
RGB	Red Green Blue
RGB-D	Red Green Blue Depth
SDK	Software Development Kit
SL	Structured Light
ToF	Time-of-Flight

1. INTRODUCTION

Traditionally, moving randomly oriented parts from one place to another has required human resources to complete (Torres et al., 2022, pp. 1–2). This process, generally defined as bin picking, consists of locating, picking and orientating randomly placed parts in a specific manner (Ojer et al., 2022, pp. 1–2). This is a monotonous task, that in the context of smart factories and industry 4.0 is no longer suitable. This brings up the “Bin picking problem” - What is a simple and easy task for humans, requires an external visual sensing system for a robot to complete (Torres et al., 2022, pp. 1–2).

The problem of bin picking is primarily focused on locating and moving randomly orientated objects from a bin (Pochyly et al., 2012, p. 1). Even after being a research topic for years (Martinez et al., 2015, pp. 1–2), and with today's advances in machine vision technology (Malik et al., 2019, pp. 1228–1229) – Pure random bin picking have yet to be achieved (Boschetti et al., 2023, pp. 1–2). From the perspective of a bin picking solution, the machine vision system has to perform in the presence of textureless surfaces, varying lighting conditions and occlusions. Modern technology has come up with several approaches for robot guidance, each with different advantages and drawbacks to approach these challenges (Pérez et al., 2016, pp. 10–17).

With the fourth industrial revolution, Industry 4.0 ongoing, higher productivity and efficiency is expected (Lydon, 2016). The expectations are high, because the task is not to just match, but to outperform human capabilities across a wide span of applications (Carroll, 2021). While the reliability and accuracy of the bin picking solution are crucial for success (Tipary et al., 2021, pp. 1–2), the flexibility of the solution is often overlooked. The picking solution should also be reusable, when the type of the parts change (Rebhouh, 2022, p. 3).

1.1 Research objectives

The objective of this thesis is to research different 3D- depth camera technologies to find the strengths and weaknesses of each technology from a bin picking applications point of view. To compare the technologies, case studies and experimental research is performed with objects ranging from the metal industry to different sub assembly components. The goal of this evaluation is to find out what object properties are hard to detect

and potentially problematic for a bin picking application. This thesis also aims to evaluate the performance of a commercial bin picking solution, based on structured light technology. The goal of this research is to present the capabilities of modern bin picking solution in different industrial applications.

1.2 Research problem and research questions

Automated bin picking systems should be versatile and be able to adjust for the diverse and evolving industrial environment. This requires a vision system, capable of performing in varying environments and with different object properties. By analysing different 3D-depth camera technologies, this thesis attempts to compare and evaluate their performance with different object types and environments. The research questions this thesis attempts to answer are:

1. What are the object properties, or a combination of properties that enable or limit the use of a specific 3D- depth camera technology?
2. How different type of 3D- depth camera technologies perform in an industrial bin picking environment?
3. How well does a commercial bin picking solution perform, and are the benefits justifiable for the increased initial cost?

1.3 Limitations

There are three limitations in this study, that could be addressed in future research:

- The object properties and materials of the 3D- depth camera comparison were limited by the selection of products from participating companies.
- The 3D- depth camera selection was limited by the hardware available for the research work.
- Only structured light was evaluated in terms of accuracy and precision because bin picking solution was not provided for other depth camera technologies.

1.4 Research methods

The research plan was based on the research onion framework, with positivism philosophy and deductive approach. This approach was chosen, because it produces quantifiable data by objectively observing the results (Tengli, 2020). After the literature review,

the first research question is researched through illustrative case studies. These are descriptive studies, that examine the interplay of different variables in attempt to explain the outcome of the research. Case studies require a problem that seeks understanding through logic, and they answer questions such as “why” and “how” (Bronwyn et al., 2005). Because the resulting data is quantifiable, this research method was chosen for the first research question. As described by (Bronwyn et al., 2005), the questions to define a case study are:

1. What questions the study attempts to answer?
2. What are the subjects of the study?
3. How is the data collected during the study?
4. What data is collected during the study, and what data is relevant for the study?
5. What are the units of analysis and how is the data analysed?
6. What is the logic that links the resulting data to the questions of the study?

The second research question is approached through experimental research. This research method studies the relationship how an independent variable affects a dependent variable. Experimental research method usually has hypotheses and the experiments either confirms or disproves them. This requires comparable subjects, where only the independent variable is altered (Bronwyn et al., 2005). When the research question studies singular variables, such as effects of ambient light, this research method is a good approach to the problem. As described by (Bronwyn et al., 2005), experimental research is defined by the following steps:

1. Identify the research problem.
2. Formulate hypothesis of the results, causal relationships, and the confounding variables.
3. Select the control and treatment groups and conduct the experiment.
4. Collect and analyse the data.
5. Discuss the results.

The third research question is researched through illustrative case study and a feasibility study. Feasibility study is a small-scale project consisting of series of interlinked questions. The goal of a feasibility study is to disclose if a proposed project could be successful, if implemented in full scale (McLeod, 2021, pp. 1–4). Feasibility study was chosen as

a research method, because the companies involved with this thesis work all had questions related to feasibility of a bin picking solution. The common steps of a feasibility study, as described by (McLeod, 2021, p. 4) are:

1. Explore viable proposals and recognise fatal flaws.
2. Concept development in laboratory environment.
3. Feasibility demonstration in practise.

The research methods described above will produce sub-results from the depth cameras and a structured light-based bin picking solution. The first two research questions provide sub-results from the strengths and weaknesses of each depth camera technology. The third research question provides sub-results from the performance of a structured light-based bin picking solution. These sub results are finally combined to answer the research questions and produce the main results of this thesis, a comparative analysis of 3D-depth cameras. An overview of the research plan is presented below in figure 1, and the research plan is explained in more detail in chapter 3.

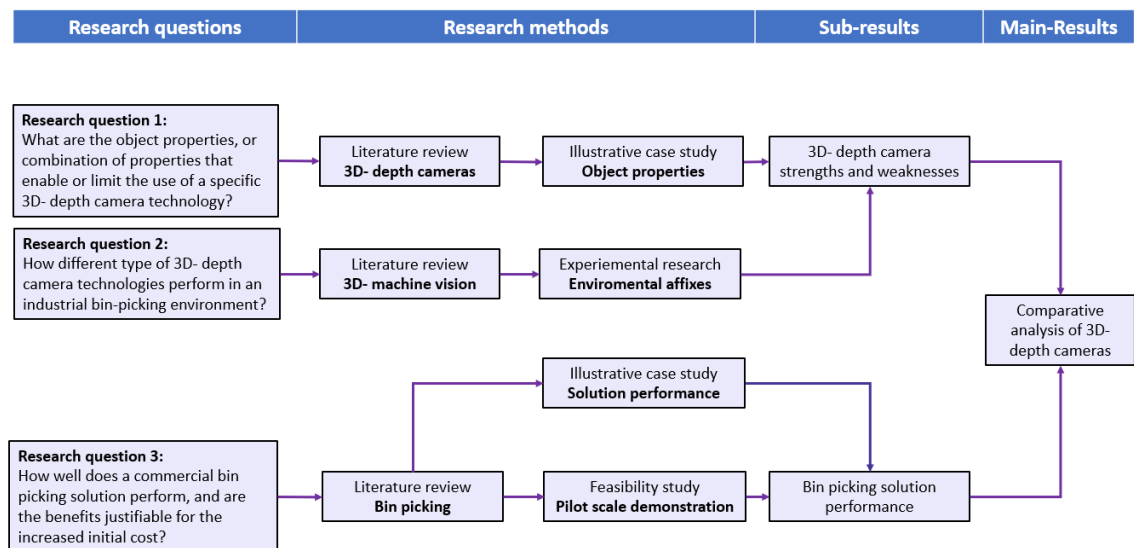


Figure 1. Research plan

2. LITERATURE REVIEW

This chapter introduces the major topics of the Thesis. Chapter 2.1 focuses on 3D- machine vision and presents the fundamentals of 3D- data structures. Chapter 2.2 presents different 3D- depth camera technologies, focusing on how they function and produce depth data. Finally, chapter 2.3 presents the industrial bin picking and the concepts of object localization and pose estimation.

2.1 3D- Machine vision system

Machine vision system is at the core of bin picking, among many other applications. Machine vision can be considered as a combination of hardware and software, that enables the operation of devices based on the captured and processed images (Cognex, 2018, pp. 3–10). Robots need perception from their surroundings in order to navigate within a 3D- environment, and machine vision attempts to solve this by providing eyes to the machine (Pérez et al., 2016, pp. 2 and 10). 3D- machine vision system furthers this concept of being the eyes of the machine, as the addition of third dimension allows even more similarity to the human vision (Lin et al., 2020, p. 551). The core of any 3D- machine vision system is based on the camera model. The inherent idea is, that the optical device, i.e., camera captures light reflected from the scene. The individual image points are then reformed to generate a single 3D- image of the scene (Giancola et al., 2018, pp. 5–12). The depth information of the resulting image is computed by mathematical models. This process requires parameters, describing the mapping between two data points. The process of mapping these two points is called calibration (Pérez et al., 2016, pp. 3–5).

The raw data produced by 3D- machine vision systems come in various forms, with different properties and structure. The amount of data preserved varies between the formats, and point clouds are a commonly used format (Dumic et al., 2020, pp. 595–596). Point clouds are one of the preferred formats, because they preserve the original geometric information of the images (Guo et al., 2020, p. 1). Point clouds are hard to process because the images may be incomplete due to occlusion, or the connectivity between the data points can be disrupted because of noise (Ahmed et al., 2018, p. 4). The important aspects, such as the geometrical shapes also needs to be filtered, structured, or segmented to provide more accurate results (Pérez et al., 2016, pp. 10–11). The different 3D- data structures are presented in figure 2.

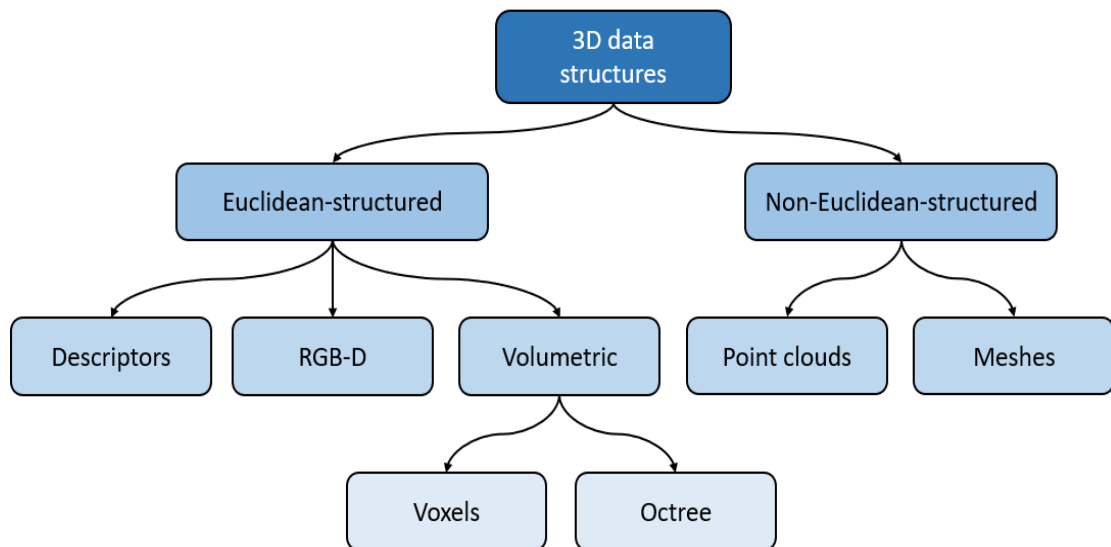


Figure 2. 3D- data structures (Based on Ahmed et al., 2018, p. 3, Figure 1)

Machine vision has enabled the use of robots in various branches of the industry, and 3D- imaging has expanded the possibilities even further. This has enabled machines to complete industrial tasks, which has allowed machine vision to be integrated into several industrial processes. Besides just robotic guidance, machine vision is also used in quality control and inspection tasks, where the vision equipped machines are used to detect and identify quality related issues (Javaid et al., 2022, pp. 1–5).

There are several benefits with 3D- machine vision compared to the traditional two-dimensional (2D) cameras. The main problems 2D- cameras have compared to the 3D- depth cameras are contrast, lighting, and lack of depth information. Lack of contrast is a significant challenge to traditional 2D- cameras, as the features it depends on are very difficult to detect from either shiny, dark or bright surfaces (Qualitas, 2011). Lighting of the scene is also a frequent problem of 2D- cameras, as implemented incorrectly it can cause shadows or reflections (Kleppe et al., 2017, p. 1). Lack of depth on the other hand becomes an issue, because it is impossible for the vision system to estimate the depth of a scene from a single 2D- image (He et al., 2018, pp. 1–2). As a by-product of better accuracy, 3D- machine vision also enables the collaborative operation between robots and humans. This includes tasks in material handling, assembly processes and industrial tasks (Borboni et al., 2023, pp. 18–19)

2.2 3D- imaging technologies

The term 3D- imaging refers to different imaging technologies that can capture 3D- data, including the depth of the scene (Geng, 2011, p. 130). Different depth- sensing 3D- technologies have many different use cases in various fields of technology. These technologies include stereo vision, structured Light, Time-of-Flight and laser triangulation (Fu et al., 2019, p. 1).

2.2.1 Stereo vision

Stereo vision system consists of two or more cameras, imaging the same scene from different points of view. Stereo vision is inspired by human vision, where the depth information is reconstructed from the disparities occurring between the captured images (Giancola et al., 2018, pp. 12–14). To compute the depth of individual data points, feature points are extracted from the scene and corresponding pixels are located between the images. These feature points can only be extracted from parts that are visible to all the cameras of the system (Zanuttigh et al., 2016, pp. 9–14). Disparity map is then computed by calculating the disparities between the images (Jafari Malekabadi et al., 2019, p. 630). The disparity map is finally converted to depth information by triangulation methods and the spatial relationship between the cameras (Lü et al., 2013, pp. 1–2).

Stereo vision systems can be an active or a passive system, depending if the system relies on external illumination of the scene or not. Passive stereo vision uses two or more cameras, but the system does not provide any additional illumination into the scene. In general, more than two cameras provide better results due to more pixel correlations. Passive stereo vision requires that the distance between the cameras and the object is known, which restricts the vision system as it needs to remain static (O’Riordan et al., 2018, pp. 178–179). The primary difference between passive and active stereo vision systems is a projector, which is present in active stereo vision systems (Jang et al., 2022, pp. 1–4). The purpose of the projector is to project light into the scene and provide additional texture, which provides more pixel correlations and leads to better accuracy. Active stereo vision also enables dynamic scene imaging, because the cameras are able to detect motion from distortions in the light patterns (O’Riordan et al., 2018, p. 179). An overview of passive and active stereo vision systems is presented in figure 3.

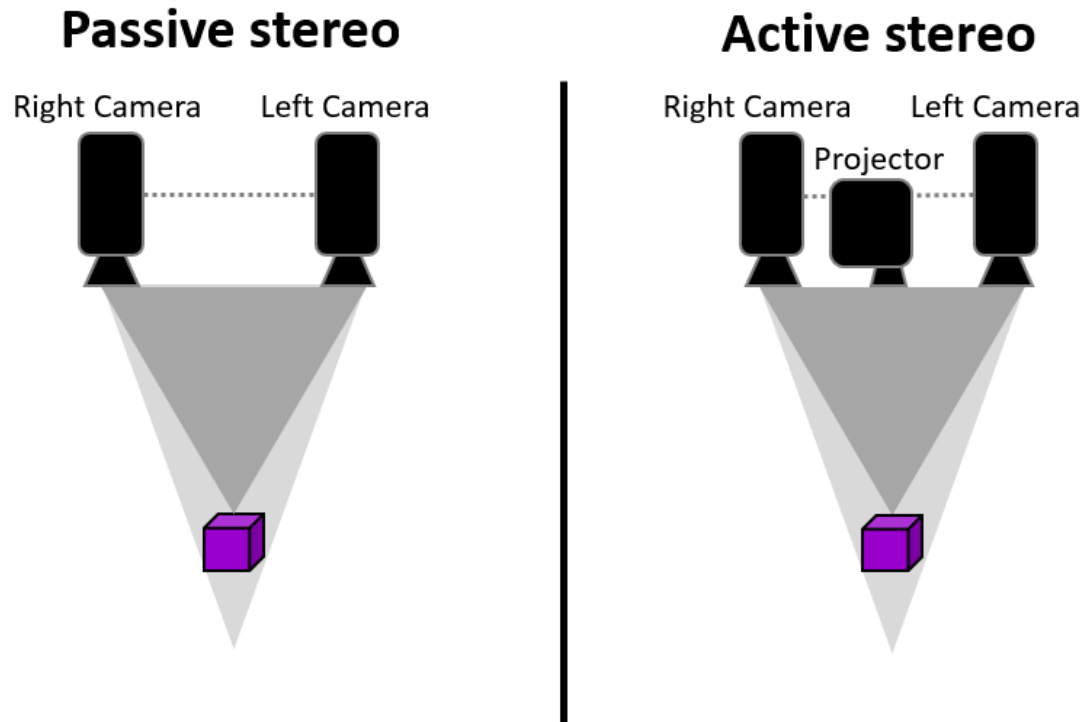


Figure 3. Stereo vision system (Based on DAQRI, 2018, Figure 1)

Stereo vision system cameras are able to capture 3D- images in real time, providing the cameras are calibrated correctly (Schrödl et al., 2010, p. 3). Another benefit of stereo vision systems is the capability of performing well in broad range of different ambient light conditions (Nagel, 2021, p. 5). One challenge stereo vision systems face is caused by the surface texture of the target objects. Smooth, reflective surfaces are known to be problematic and they can cause incorrectly computed depth data (O’Riordan et al., 2018, p. 181). Stereo images from a scene without identifiable features are also problematic for passive stereo systems, because of the correspondence problem (Zanuttigh et al., 2016, pp. 17–18). Since the stereo images from such a scene are uniform, corresponding pixels cannot be found and depth information cannot be obtained. This problem can be solved with an active stereo system. The projected patterns of light produce the texture needed for finding the corresponding pixels from both images. The downsides of active stereo systems are the increased cost and longer processing times, compared to passive stereo systems (Dal Mutto et al., 2013, pp. 9–11).

2.2.2 Structured light

Structured light system consists of a single camera and a projector, that projects structured light patterns to the scene (Giancola et al., 2018, pp. 18–20). The patterns projected to the scene are coded and designed in a way, that each of the pixels of the scene

can be individually identified. The projections essentially give each of the pixels an individual codeword, depending on the light values received by the pixel. Based on these codewords, correspondences can be established between the projected patterns and image points. The location of individual data points and the depth of the scene can then be computed with triangulation methods (Salvi et al., 2003, pp. 1–2). The depth computation of a structured light system is based on the light reflected from the scene, back towards the cameras (Giancola et al., 2018, pp. 18–20). From projective geometry point of view, the light projector can be considered as a standard pin-hole camera (Zanuttigh et al., 2016, pp. 22–27). Because of this, the triangulation principle of a structured light system is functionally equivalent to a stereo vision system (Zanuttigh et al., 2016, pp. 22–27). An overview of a structured light vision system is presented below in figure 4.

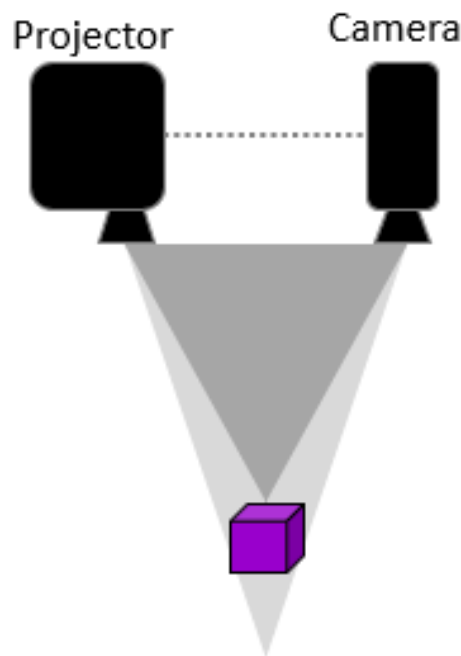


Figure 4. Structured light vision system (Based on DAQRI, 2018, Figure 1)

Structured light cameras are very accurate, even capable of sub-millimetre accuracy (Giancola et al., 2018, pp. 18–20). This accuracy however comes with a long processing time and requires a stationary scene. Imaging of dynamic scenes can be enabled if a specific type of projected patterns are used. The downside of dynamic scene imaging is lower resolution, and it makes the vision system very sensitive to surrounding light (Zanuttigh et al., 2016, pp. 22–27). A common problem shared between the stereo vision systems and structured light systems are reflective surfaces (Dal Mutto et al., 2013, pp. 36–40). Besides reflective surfaces, another common problem of the structured light systems is the ambient light of the surrounding environment. The Projected light patterns have to be brighter than the ambient light, or the quality of the resulting point cloud can

drastically decrease (Gupta et al., 2013, p. 1). Structured light systems can also face problems with the scene geometry, if the camera cannot detect all of the projected light patterns because of occlusions (Dal Mutto et al., 2013, pp. 36–40). The performance of structured light systems is largely affected by the design of the projected light patterns. Projection patterns of different structured light methods are presented below in figure 5, with their advantages and drawbacks discussed in existing literature by (Geng, 2011, pp. 133–146).

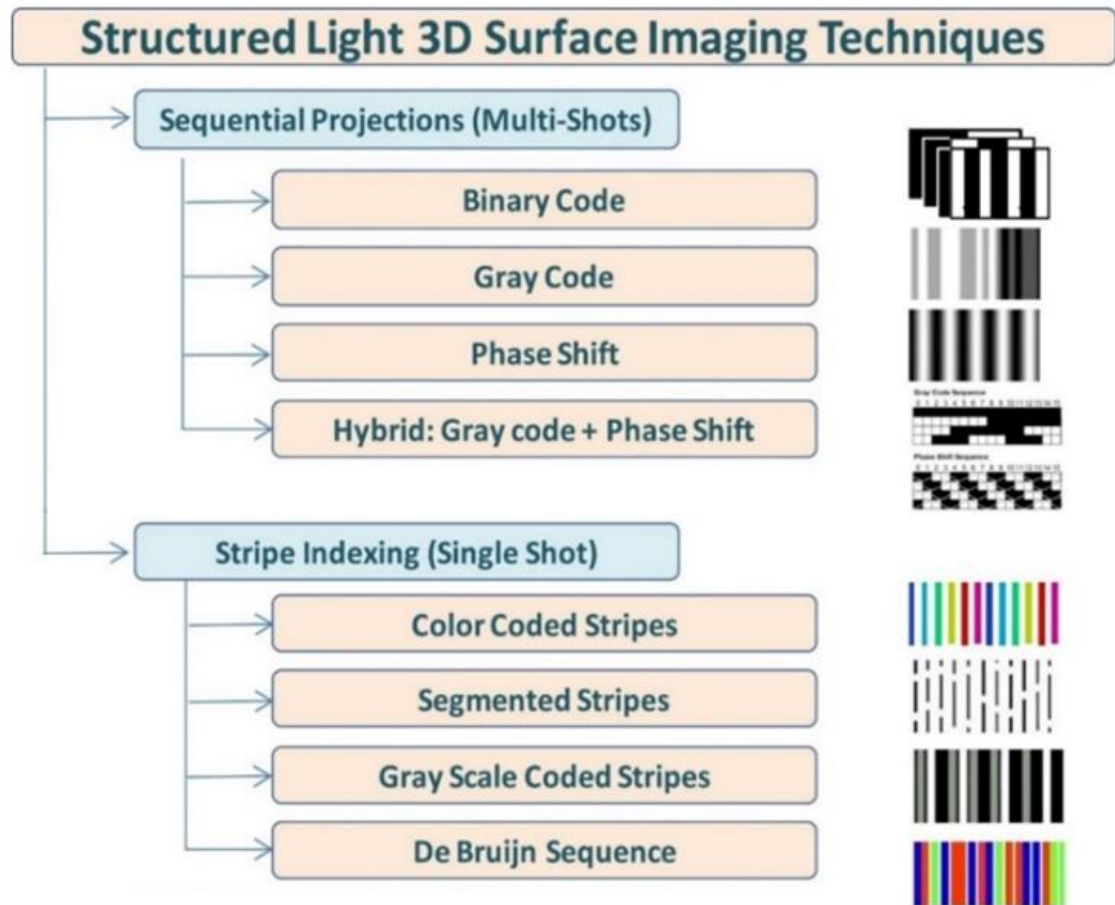


Figure 5. Structured light projection patterns (Geng, 2011, p. 133, Figure 3)

2.2.3 Time-of-Flight

A Time-of-Flight (ToF) camera is a 3D- depth camera, consisting of a transmitter and receiver. The transmitter illuminates the scene, by transmitting a modulated light- signal that is reflected back to the receiver from the objects in the scene. The transmitted laser beams are typically in near-infrared (NIR) range and emitted in very high frequencies of 10...100 MHz (Giancola et al., 2018, pp. 20–25). The receiver is a matrix of pixels used to collect the light that is reflected, where each pixel individually computes the delay between the transmission and reflection of the light rays (Zanuttigh et al., 2016, pp. 27–

32). The depth measurement of ToF- technology is based on the electro-magnetic radiation speed (speed of light). This technology measures the time for the light to leave the transmitter and being reflected back to the receiver (Dal Mutto et al., 2013, pp. 28–31). The transmitter of a ToF- camera is typically an array of laser emitters, that can illuminate the whole scene with just a single signal to produce a depth image of the scene (Zanuttigh et al., 2016, pp. 27–32). The two different approaches for ToF- camera depth computation are pulse modulated ToF (Direct method) and Continuous- wave ToF (Indirect method) (Syrjänen, 2021, pp. 10–14). Pulsed light operation is the commonly used method because the high-power laser pulsed in high frequency is more resilient to external noise from ambient light. This high power light also enables longer range measurements, and removes the need for a high sensitivity receivers (Zanuttigh et al., 2016, pp. 27–32). The structure and operating principle of an ToF- camera is presented below in figure 6.

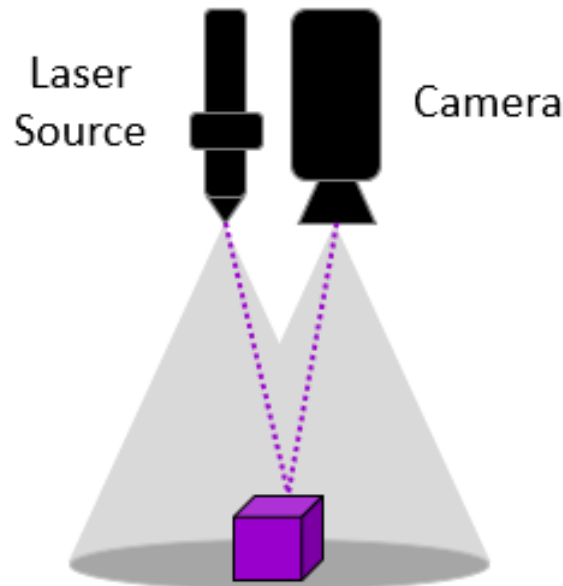


Figure 6. ToF- depth camera system (Based on DAQRI, 2018, Figure 1)

ToF- cameras are an efficient way to capture 3D- scenes and they are capable of real time measurements (Grzegorzek et al., 2013, pp. 3–12). Even though ToF- method sounds simple in nature, the major challenge of ToF- technology is simply the speed of light. Since the speed of light is known, it can be calculated that to achieve a 1 mm accuracy, the camera mechanisms must operate with reaction times faster than 6.67 ps. Even an accuracy of 1 cm requires reaction times of 70 ps (Zanuttigh et al., 2016, pp. 27–32). Another major error source for ToF- camera is the multi-path error, which is a phenomenon where the transmitted ray of light is reflected to the receiver from multiple different sources. This effect can lead to incorrectly estimated depth of the scene, and is hard to model because it is scene dependent (Dal Mutto et al., 2013, pp. 28–31).

2.2.4 Laser triangulation

Laser triangulation system consists of a laser emitter and a receiver. The laser emits a laser beam, which is projected onto the surface of the target object. This scattered light of the reflection is then captured by the camera. The depth is computed by triangulation methods and image processing algorithms (Huang and Kovacevic, 2012, pp. 1–4). The computation behind laser triangulation is based on locating and computing the vertical shift of the surface, from the displacement of the detected laser line captured by the camera. As presented in figure 7 below, a change in the surface height will lead to a lateral displacement of the image points (Mohammadikaji, 2020, pp. 8–11). Since the height profile acquisition requires object or laser movement, this method is also referred as a scanning technology (Nilsson and Murhed, 2015, p. 4).

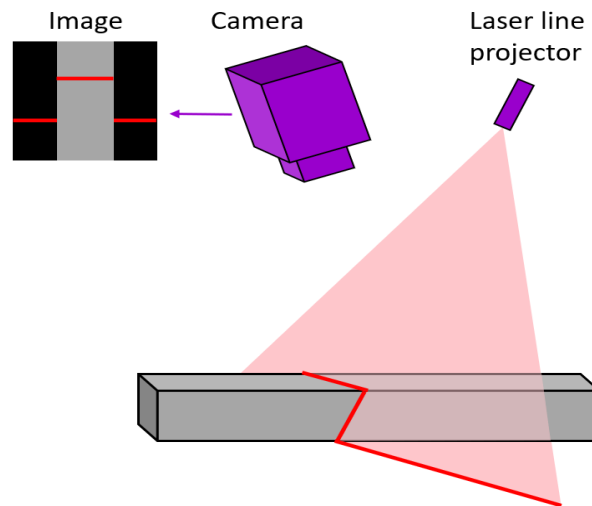


Figure 7. *Laser triangulation system (Based on Mohammadikaji, 2020, p. 10, Figure 1.3)*

The laser light is iteratively projected across the whole surface of the imaged objects, and individual data points are generated for each frame of the scene called scan images. While the laser projection is moved, subsequently scan images are generated. When the whole surface of the object has been scanned, a range image is created by combining all the scan images together. This resulting range image is a 3D- image, where the pixel values represent the surface coordinates of the object (Muhammad Amir and Thörnberg, 2017, pp. 2–4). Laser triangulation can be performed with two different methods. Fixed angle emission with variable distance, and fixed triangulation base with variable scanning angle (Lopez et al., 2008, pp. 400–410). Laser triangulation is a high precision method of 3D- imaging, but it is limited in range by the transducer. Laser triangulation also requires a specific surface with specific roughness and opacity to achieve the high resolution. Laser triangulation is sensitive to reflections, but one of its major advantages is the low cost of the system (Soave et al., 2020, p. 2).

2.3 Bin picking

Bin picking is a methodology used in vision guided robotics- systems, where a robot is used to pick objects from a bin to another part of the manufacturing process. In the recent years, the 3D- machine vision system have made it possible to implement robust bin picking applications, on the level of the smart factory concept (Torres et al., 2022, p. 2). The general bin picking solution is composed of a 3D- machine vision system, bin picking software, robot manipulator, and a gripper. The machine vision system is used to capture the scene, where the bin picking software localizes the objects and computes the trajectories. The robot manipulator then completes these trajectories, and grasps the object with the gripper (Rebbouh, 2022, p. 4). A traditional bin picking task is presented below in figure 8.

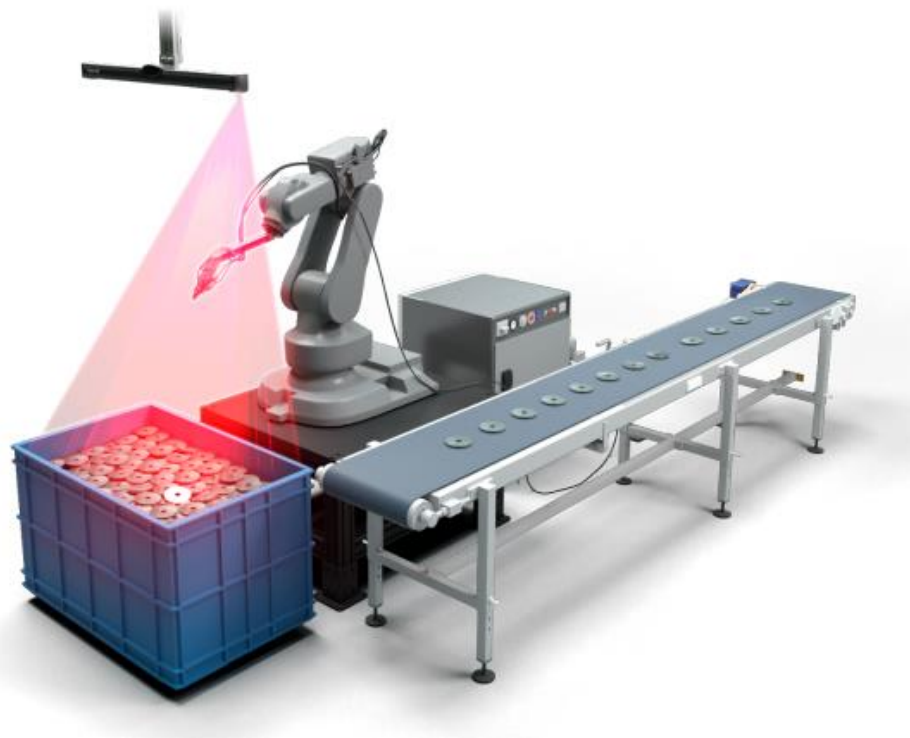


Figure 8. Traditional bin picking application (Photoneo, 2018, p. 1)

Efficient bin picking has many challenges. What makes bin picking so difficult, is to achieve the accuracy and flexibility that is required in the industrial environment (Sansoni et al., 2014, pp. 1–2). While robots are capable of high repeatability (Universal Robots, 2016), the difficulty comes with the recognition and pose estimation of randomly oriented objects (Sansoni et al., 2014, pp. 1–2). Besides just technical challenges, another challenge of automated bin picking is to meet all of the cycle times and optimization requirements, set by the industrial environment (Rebbouh, 2022, p. 3). Solving these challenges

requires a combination of machine vision, software, computing power and a gripping solution to perform (Anandan, 2016).

The general tasks of bin picking application are data acquisition, object localization, and path planning. Among all of these tasks, the accurate localization of the objects is considered to be the most challenging (Wnuk et al., 2017, pp. 1–2). This is challenging, because the success rate of a bin-picking solution is not only dependent on the resolution of the vision sensor, but also on the underlying algorithms to perform the localization (Torres et al., 2022, pp. 2–3). If objects have features that enable entanglement, localization becomes even more challenging because grasping an entangled object results in a failed grasp (Moosmann et al., 2020, pp. 1–2). The localization problem is approached by several steps of data handling, to increase the accuracy of the object localization. The raw point clouds, produced by the depth cameras are first filtered to reduce undesirable data. The pre-processed point clouds are then segmented to extract individual objects from the scene. To obtain the initial pose of the object, these potential objects are processed by matching tools such as descriptors or algorithms. Finally, the object pose is processed by the use of algorithms to achieve an accurate location (Li et al., 2019, pp. 149–150). The main tasks of a bin picking application are presented below in figure 9.

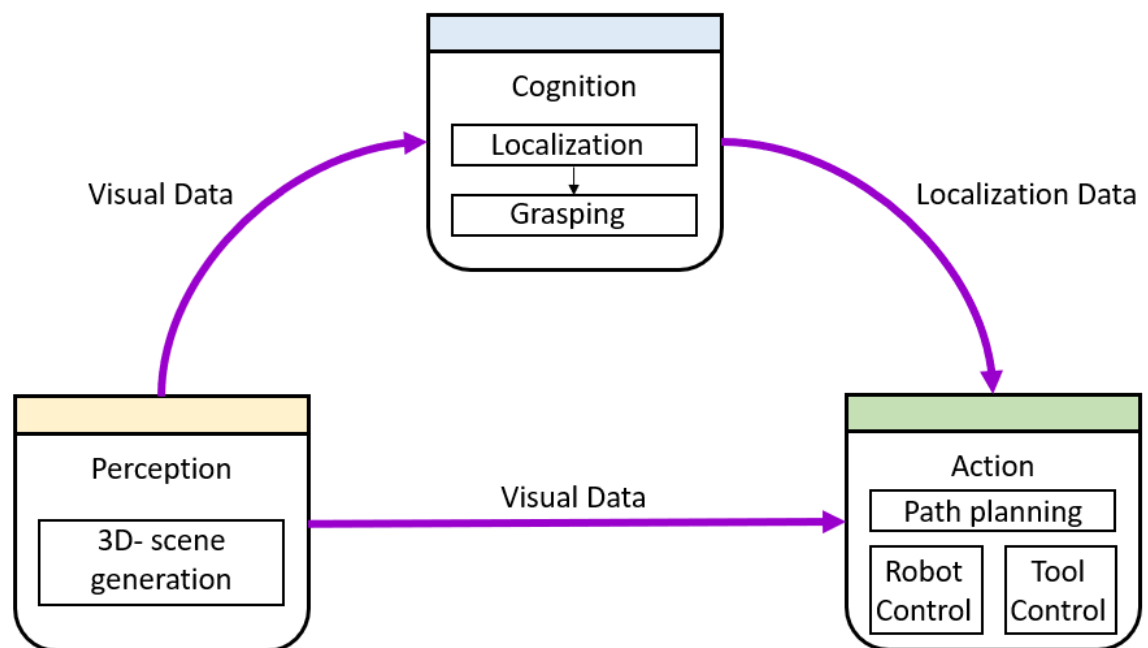


Figure 9. Bin picking application tasks (Based on Ojer et al., 2022, p. 4, Figure 1)

Recent advances in bin picking technology have come up with a model-based bin picking solutions, in attempt to provide easily configurable bin picking solutions. These commercial solutions come with preconfigured frameworks for the bin picking application, enabling bin picking without extensive experience with programming.

2.3.1 Model-based bin picking solution

Model-based bin picking solutions are a combination of hardware and software (Photoneo, 2018, p. 9). They search the scene for an explicit model, by comparing the data between the captured image and the reference model (Chen et al., 2018, pp. 1–2). Pose estimation of a model-based solution is based on 3D- point descriptors of the reference model, and the acquired points of the point cloud. The correspondences formed between features and depth edges result in initial pose estimation, which is refined for an accurate location of the object by using algorithms. The benefits of model-based solutions are the broad availability of 3D CAD- models, and they do not require the time-consuming training phases of machine learning-based solutions (Liu et al., 2012, pp. 1–6).

A general composition of a model-based bin picking solution is presented in figure 10. The main tasks of the computing unit is to deal with the data acquired by the 3D- depth camera and calculate trajectories for the robot manipulator as an output (Ojer et al., 2022, pp. 3–4). The conventional workflow of a model-based bin picking solution is to first generate a 3D- point cloud from the scene, and perform pose estimation for the object (Photoneo, 2018, p. 10). The gripping pose is then computed based on the localized object, which is aligned with the 3D- CAD model of the robot gripper (SICK, 2022a, pp. 12–16). Finally, a collision-free trajectory is generated for the robot to grasp the object (Photoneo, 2018, p. 11). This workflow is described in more detail by commercial solutions such as Photoneo (Photoneo, 2018, pp. 10–11) and SICK (SICK, 2022a, pp. 12–16).

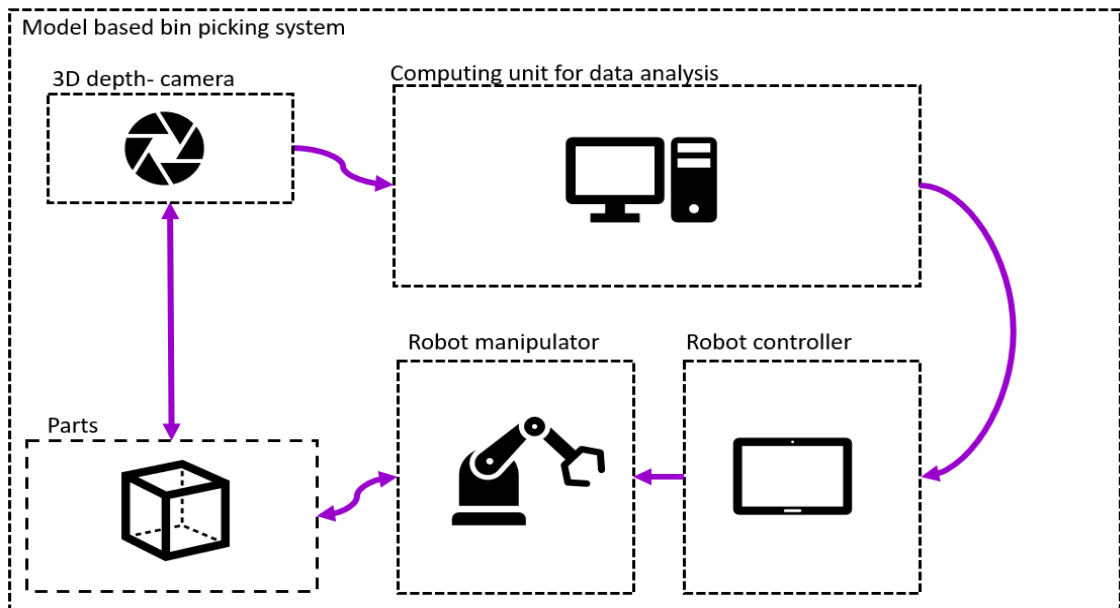


Figure 10. Model-based bin picking solution (Based on Photoneo, 2018, p. 9)

2.3.2 Pose estimation problem

The key prerequisites for successful bin picking application are accurate localization of the object, alignment of the world coordinate systems and pose estimation. These are crucial elements for the robot manipulator, to ensure collision free operation (Torres et al., 2022, pp. 7–8). To align the coordinate systems of the robot and the camera, both are calibrated relative to the same world coordinate system. To align the robot Tool Center Point (TCP) with the object, three important poses of the bin picking system need to be aligned. These poses are the object pose, gripper pose and gripping pose (Buchholz, 2016, pp. 10–16). Overview of these poses and coordinate systems is presented below in figure 11.

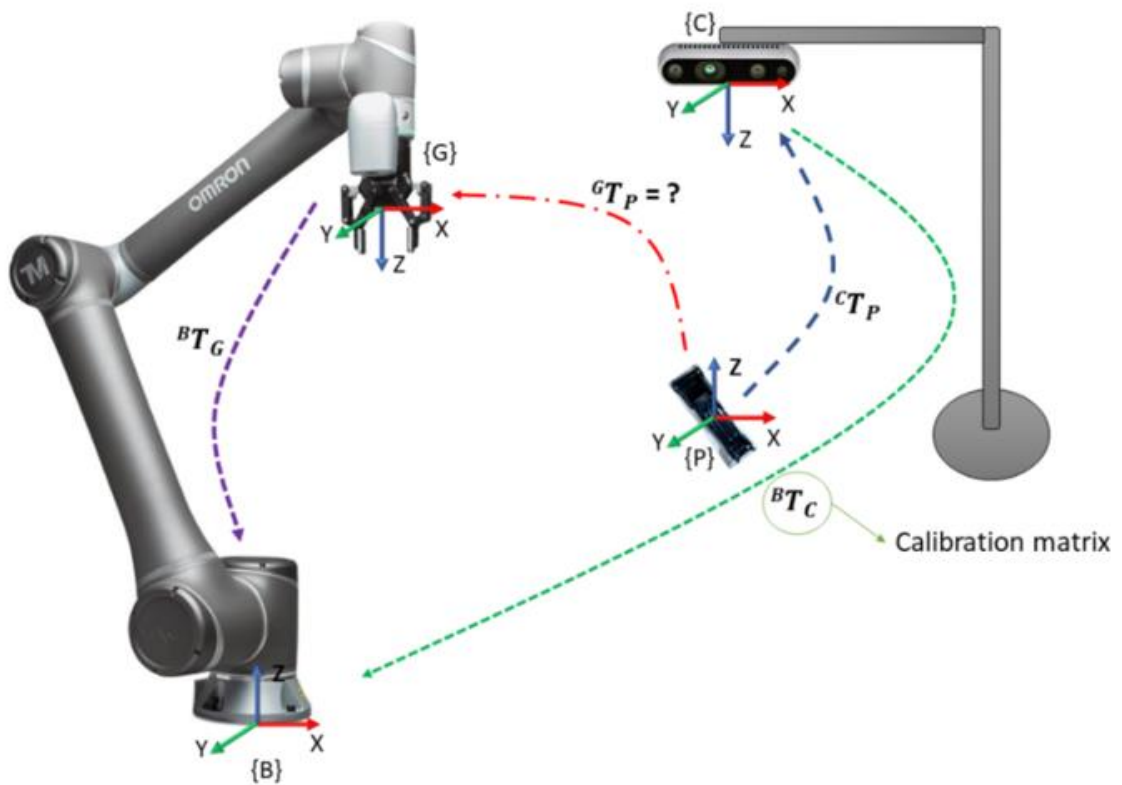


Figure 11. Coordinate reference systems (Torres et al., 2022, p. 8, Figure 7)

Alignment of the world coordinate systems leads to the most important problem of bin picking: Pose estimation. Bin picking has several methods for pose estimation, which are based on either point clouds, depth maps, or normal maps (Buchholz, 2016, pp. 10–16). To align the gripper with the object and grasp it, the gripping pose T_P^G needs to be computed. This pose describes the pose of the object in relation to the gripper coordinate system. In order to compute it, three components of the transformation chain presented in formula 1 need to be solved. The gripper pose T_G^B describes the pose of the gripper in relation to the robot base. The object pose T_P^C describes the orientation of the object in

relation to the camera reference frame. A calibration matrix T_C^B represents the camera orientation in relation to the robots reference frame (Torres et al., 2022, pp. 7–8).

$$T_P^G = (T_G^B)^{-1} \cdot T_C^B \cdot T_P^C \quad (1)$$

Pose estimation can be solved by various algorithms designed for this problem, such as the Random Sampling Algorithm (RANSAM). These algorithms estimate the transformation between the sets of data and provide the pose of the object as an output (Buchholz, 2016, pp. 10–16)

3. RESEARCH PLAN

The main focus of this chapter is to present a detailed research plan to achieve the objectives of this thesis and answer the research questions. To answer the first two research questions, the depth camera performance is evaluated to see how they perform against each other. To answer the third research question, the bin picking solution is evaluated in different bin picking tasks and a feasibility study.

Chapter 3.1 presents the illustrative case studies and experimental research common between all the depth cameras. Chapter 3.2 presents the feasibility study and pilot scale demonstrations of the bin picking solution, and finally, chapter 3.3 presents the illustrative case study to evaluate the accuracy of the bin picking solution.

3.1 3D- depth camera performance evaluation

Well-refined framework and evaluation criteria are needed to objectively compare the performance of different 3D- depth cameras (Stoyanov et al., 2012, p. 2). The common, comparable property between all the depth cameras of this thesis is that they produce 3D- point clouds. Common data format among all the hardware was the Polygon file format (PLY), which was chosen as the basis for evaluation to have comparable results. To achieve high quality point clouds, the camera Software Development Kit (SDK) parameters needs to be carefully tuned. To ensure the reliability of the results, guidelines from the manufacturers and manuals are followed and different variables are experimented with before the final data collection.

First comparison is an illustrative case study, with the goal to research the camera technologies to see how they perform with reflective surfaces. Reflectivity was chosen as a research topic because highly reflective surfaces and specular reflections are one of the major sources of error in depth estimation (Feng et al., 2023, pp. 1–2; Li et al., 2022, pp. 1–4; Tan et al., 2021, pp. 1–2). The problem this research attempts to answer, is *“how well is a specific depth camera technology capable of performing with reflective surfaces”*.

This research is conducted on simple geometrical shapes, such as polygons, squares, and cylinders. The first set of data is acquired from reflective metal blanks. These results are compared against the second set of data from identical 3D- printed objects from Poly Lactic Acid (PLA) plastic. This research is conducted on objects of different sizes to see if there is a limit when an object can reliably be recognized. The data for this research is

collected directly from the depth cameras, located directly on top of the bin and the objects. The relevant data of this comparison are the raw point cloud data generated by the depth cameras, where the focus is on the edge and planar fidelity of the objects. The results from this research are analysed by comparing the results of the two sets of data with point cloud analysis tools. The key properties of comparison are sharpness of the edges, convex and planar surfaces, and the density of the point clouds. These properties were chosen as the points of comparison, because having identical objects with different surface types, the results can be directly linked to the question of this research.

Second comparison is an illustrative case study, with the goal to research the camera technologies to see how accurately they can detect complex geometrical shapes. The detection of complex surfaces was chosen as a research topic because accurate detection of feature points is a crucial part of pose estimation and bin picking as a whole. If the depth camera cannot accurately capture the properties of the object, then the localization algorithms also fail to perform. The problem this research attempts to answer, is *“how well is a specific depth camera technology capable of performing with different object properties?”*.

This research is conducted on objects with angled-, uneven surfaces and objects with height differences and special properties. Objects of varying sizes, materials and reflectivity are compared to see if there is a property or a combination of properties that limits the use of any of the depth camera technologies. The data for this research is collected directly from the depth cameras, located directly on top of the bin and the objects. The relevant data of this comparison are the raw point cloud data generated by the depth cameras, where the focus is on the complex features of the objects. The results from this experiment are analysed with point cloud analysis tools. The important aspect of this comparison is the identification of a property or a combination of properties, that are problematic for the depth cameras to detect. The key features of analysis are combination of different surface types. The results of this research are quantifiable, visible results as point clouds. Therefore, the resulting point clouds can be directly linked to the question this research attempts to answer.

Third comparison is experimental research, with the goal to see how ambient light, or the lack of it affect the resolution of the depth- measurement. This test is completed with different bin configurations, with varying bin backgrounds. Ambient light and bin configuration were chosen as one of the criteria, as both variables can change during the operation of the bin picking solution. If the machine vision system is not capable of dealing with these variances, it can cause issues with the performance of the application. Based on the literature review, the following hypothesis are being researched:

- High intensity ambient light will have a negative effect on the point cloud quality of the structured light system.
- ToF and stereo vision technologies are resilient to external, ambient light and ambient light does not have an effect to the quality of the point clouds.
- All of the depth camera technologies are capable of accurately imaging non-reflective, uniform surfaces.
- Highly reflective surfaces degrade the quality of the point clouds with all depth camera technologies

The results from this research provide answers to the hypothesis presented above through two causal relationships. First, the results from varying ambient light conditions provide answers to the first two hypothesis through the causal relationship between the quality of the point clouds and ambient light. Secondly, the results from different bin configurations provide answers to the final two hypothesis through the causal relationship between the quality of the point clouds and the type of bin surface. The confounding variables of this research are the positions of the cameras, positions of the bins and ambient light. In order to have only one independent variable, both the positions of the bins and cameras are kept constant between the comparison. The control group in this study is the point cloud of an empty bin in ambient light. The treatment groups are the cases with varying amounts of ambient light and different bin configurations.

The comparison is completed with varying degrees of ambient light and different bin configurations, to see how prone the technology is for disturbance. The relevant data of this comparison are the raw point cloud data generated by the depth cameras, where the irrelevant data outside of the bins is cut off. The results from this experiment are analysed from the point clouds generated by the cameras. The effects of the environmental variables are evaluated based on the completeness of the point clouds between different conditions. The key properties of comparison are incomplete parts of the point cloud, caused by reflections, interreflections or ghost artifacts in the point cloud.

3.2 Proof-of-concept- demonstration in industrial environment

Functionality of the Photoneo Bin Picking solution is piloted in industrial environment by two different applications. These pilots are conducted in the premises of the companies involved, where both the robotic cell and depth camera are transported for the duration of the demonstrations. The goal of these demonstrations is to present how the bin picking solution can be utilized in different tasks and how does it perform. These presentations

also aim to present the flexibility of an OOTB- bin picking solution and contribute to answer the final research question.

The first pilot-scale demonstration presents a case where matte connector components are picked from a bin and placed for the next step of the process. This demonstration aims to present the capabilities of a bin picking solution, when the objects are small with many identifiable features. The current implementation of this task is done by hand and the goal of this pilot scale demonstration is to automate this task.

The second pilot-scale demonstration is a machine tending case, where the system is used to localize and palletize shiny metallic blanks. This demonstration also aims to present a more realistic scenario for the application, which is presented with objects that are covered in band saw cutting oil and metal flakes. The goal of this application is to present how the manual palletization of metal blanks could be automated with a bin picking solution.

Both applications are first approached from the object and gripper design point of view, to ensure that the objects can be properly grasped. The result from this step need to conclude that a gripper is available and able to grasp the objects. The objects also need to be recognisable by the machine vision system, which needs to be verified with the vision system. The next step of the study is to verify that the system is capable of accuracy and reliability, through laboratory testing and system parameter tuning. After successful laboratory testing, the system is ready for a feasibility demonstration in the premises of an industrial environment. In this case, the industrial environments are free of ambient variables such as dust or moisture, and the results from this pilot are expected to be similar to laboratory testing. The questions this feasibility study attempts to answer are the capabilities, implementation time and economic viability of the solution. Other valid questions would include sustainability and efficiency, but these variables would require further planning outside the resources of this study.

3.3 Model-based bin picking solution accuracy evaluation

The performance of the depth camera is evaluated by first localizing an object, grasping it with the robot manipulator and then placing the object in front of a 2D- camera. The 2D- camera is used to capture an image from each object, which is then analysed with an industrial vision application. The goal of the image analysis is to compute the deviation between the reference location and current location of the object. The main goal of this research is to evaluate the relative picking accuracy and precision of the bin picking

solution. The secondary goals are to see how errors accumulated from calibration, localization, and gripping the parts affect the performance of the solution. This evaluation is completed with Photoneo Phoxi scanner S- and M- models, to see if there are differences between two models of the same scanner. Both of the depth cameras will be placed on their respective, optimal heights during this test. Due to the low operation height of the S- model (442 mm), the bin will be removed to provide enough room for the robot to operate underneath the depth cameras.

This evaluation is an illustrative case study, with the goal to research the accuracy and precision of the bin picking solution. These properties were chosen as evaluation criteria because they are quantifiable metrics of the performance of the solution. The problem this research attempts to answer, is *“how accurate and precise a commercial bin picking solution is with different types of objects?”*. This research topic was chosen because accuracy in an industrial application, such as machine tending is remarkably high. If the vision application of a bin picking solution is accurate enough, a secondary system to compensate for the errors might not be needed.

This research is conducted on several different types of objects. Objects with different properties are compared to see if there are properties that increase or decrease the performance of the bin picking solution. The data for this research is collected from the industrial vision application, which provides the displacement of the object as an output. The data collected, and the relevant data of this study during this evaluation is the rotational, X-, and Y- axis displacement of the objects. The results from this experiment are analysed with graph analysis tools to visualize the displacement in millimetres and degrees. The important aspect of this comparison are outliers, or patterns in the inaccuracies. The key features of analysis are the identification of a problematic properties and potential error sources of the research. To achieve reliable results, proper calibration of the 2D- camera, scene lighting and image analysis tools setup is essential. Because the industrial vision application directly outputs the results in engineering units, the results can be directly linked to the question this research studies.

4. RESEARCH HARDWARE AND SOLUTION DESIGN

This chapter focuses on presenting the research objects, hardware, and solution design. Chapter 4.1 focuses on presenting the objects used in the research work, and chapter 4.2 presents the research equipment. Chapter 4.3 presents the solution design, and the robot program. Finally, chapter 4.4 presents the camera calibration procedures.

4.1 Research objects

The objects included in this thesis have been provided by companies involved in different branches of industry. These objects have different shapes, sizes and surfaces that provide a broad range of properties for the research work. All the objects have been assigned an object ID, which will be referred from now on for the remainder of this thesis. Figure 12 below presents the first category of objects made of metal. These objects compose from metal blanks and semi-finished products, used in the metal industry. The size of the metal blanks vary between $20 \times 20 \text{ mm}$ and $70 \times 60 \text{ mm}$, while the size of the semi-finished products range between $50 \times 35 \text{ mm}$ and $130 \times 115 \text{ mm}$. The common properties between object IDs 1...7 are that they are ferromagnetic and have shiny surfaces of varying degrees. The surfaces also have irregularities between the objects in terms of rust and cutting marks.



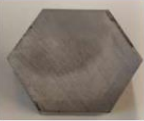




ID	Object description	Image	ID	Object description	Image
1	Metal blank Rectangular shape		5	Semi-finished product #1	
2	Metal blank Polygon shape		6	Semi-finished product #2	
3	Metal blank Cylindrical shape #1		7	Semi-finished product #3	
4	Metal blank Cylindrical shape #2				

Figure 12. Ferromagnetic objects

The second category of objects is presented in figure 13 below. These objects are connector components, used in cable manufacturing and electrical assemblies. The size of these objects vary between $30 \times 15 \text{ mm}$ and $80 \times 70 \text{ mm}$. Object IDs 8...12 are non-magnetic, matte and are composed of both plastic and aluminium. The common properties between object IDs 8...12 are the combination of convex and planar surfaces. These planar surfaces also have varying heights, which provide a comparison point for the research, in terms of how accurately they are detected. These objects provide the more complex features to the research, with a combination of different properties.






ID	Object description	Image	ID	Object description	Image
8	Connector component #1		10	Connector component #3	
9	Connector component #2		11	Connector component #4	
			12	Connector component #5	

Figure 13. Connector components

The third category of objects is subassembly components, presented in figure 14 below. The smallest feature to detect in this group is the diameter of the spring, which is $< 2 \text{ mm}$ in diameter. Object IDs 13...16 are composed of wool, aluminium, and plastic. These objects provide more complex features to the research, in the form of the porous surfaces and complex object geometry. These objects also provide small features to be detected, from the thin structure of the spring and small holes of the aluminium bracket.





ID	Object description	Image	ID	Object description	Image
13	Aluminium bracket		15	Plastic tube	
14	Wool piece		16	Spring	

Figure 14. Subassembly components

4.2 Research equipment

The hardware for this thesis consists of a Universal Robots UR5- robot manipulator, Photoneo Bin Picking Studio, and depth cameras from Basler, Photoneo and SICK. The camera used in bin picking solution accuracy evaluation is an area scan camera from Basler. Universal Robots UR5 is presented below in figure 15. It suits the parameters of the research work well, because as a collaborative robot, it has inbuilt torque sensors to react for possible collisions. It also has a maximum payload of 5 kg, which is more than enough for all of the objects being researched (Universal Robots, 2016).



Figure 15. *Universal Robots UR5 robot manipulator (Universal Robots, 2016)*

Photoneo Bin Picking Studio (BPS) is a Combination of hardware and software, designed to configure and run a bin picking process. The Application Programming Interface (API) processes the data from the depth camera, and the localization engine processes the point cloud for objects matching the reference model. The software then determines the location of the object in 3D- space as its output. After the object has been localized, the system computes a trajectory for the robot manipulator to grasp the object (Photoneo, 2018, pp. 9–11). Photoneo BPS is presented below in figure 16.



Figure 16. *Photoneo Bin Picking Studio (Photoneo, 2020, p. 8, Figure 4)*

Photoneo PhoXi Scanner is a depth camera based on structured light technology. It is a laser **class 3R** device, which is considered safe when handled carefully. Photoneo Phoxi Scanner used in the research work is presented below in figure 17. The depth camera produces point clouds with 3.2 megapixel resolution and can operate on heights between 458 *mm* to 1118 *mm* (Photoneo, 2021, p. 2). Compared to the other cameras used in the research, Photoneo has the highest resolution, but is the only camera unable to produce real time data.



Figure 17. *Photoneo PhoXi scanner*

Basler Blaze-101 is a depth camera based on ToF- technology, presented in figure 18 on the left. It is a **Class 1** laser product, which means it is safe under all reasonably foreseeable conditions of normal use. It produces point clouds with 0.32 megapixel resolution, capable of real time (30*fps*) operation. This depth camera has an operational range of 0.3 *m* ... 10 *m* (Basler, 2022a). SICK Visionary-S is a depth camera based on Stereo Vision technology, presented in figure 18 on the right. This depth camera is an active stereo device and classified as a **Class 1** laser product. The output of the camera is in 0.32 megapixel resolution, also capable of real time (30 *fps*) operation. The operational range is also similar to Basler Blaze-101, limited to 0.5 *m* ... 6.5 *m*. What separates this model from other cameras of this research, is the capability of colour imaging. With an RGB- camera, SICK Visionary-S is able to apply colour to the depth images it generates (SICK, 2022b).



Figure 18. *Basler Blaze 101 ToF- camera (left) and SICK Visionary-S Stereo vision camera (right)*

The 2D- camera used in the research work is presented in figure 19 below. Basler acA1300-60gm is an area scan camera, used in the accuracy evaluation of Photoneo bin picking solution. It is capable of producing images at 60 *fps* with a resolution of 1.3 megapixels (Basler, 2018).



Figure 19. *Basler acA1300-60gm area scan camera*

Figure 20 represents the work cell and research environment. The depth cameras are attached to a sturdy frame, with a clear view to the whole volume of the bin. The dimensions of the bin are 400 x 300 x 120 mm (width, length, depth). The perpendicular frame, bin and a small pallet are located in a way to allow easy placement of the 2D- camera, and external illumination between different tests. This configuration also allows large range of motion for the robot.



Figure 20. *Research environment*

4.3 Solution design

The robot program was developed on top of a bin picking framework provided by Photoneo. This framework for the robot controller includes a very basic bin picking application and communication between the vision system. A flowchart of the functions of this framework and the authors additions are presented below in figure 21. The yellow steps of the flowchart present modifications by the author, and blue steps represent the framework provided by Photoneo. To run the scripts provided by Photoneo, the robot software was updated to version 3.15.4.

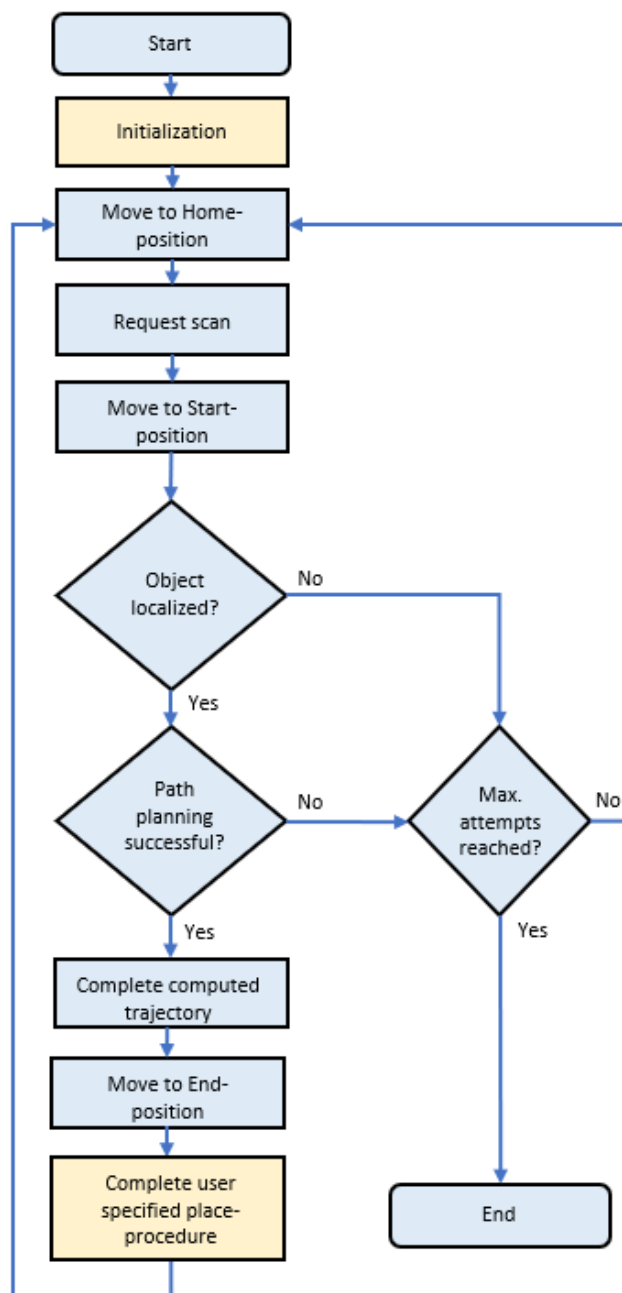


Figure 21. Robot program framework

To ensure safe operation of the bin picking solution, the movement of the robot manipulator was limited. This was done to prevent collisions with the research hardware. These restrictions are presented in figure 22 on the left with red volumes, restricted from the robot to access in path planning. The joint values of the robot were also limited to optimize the movement and avoid singularities. The joint values were set by visually inspecting the range of movement the robot requires to reach the entire volume of the bin. These joint value limitations are presented below in figure 22 on the right.

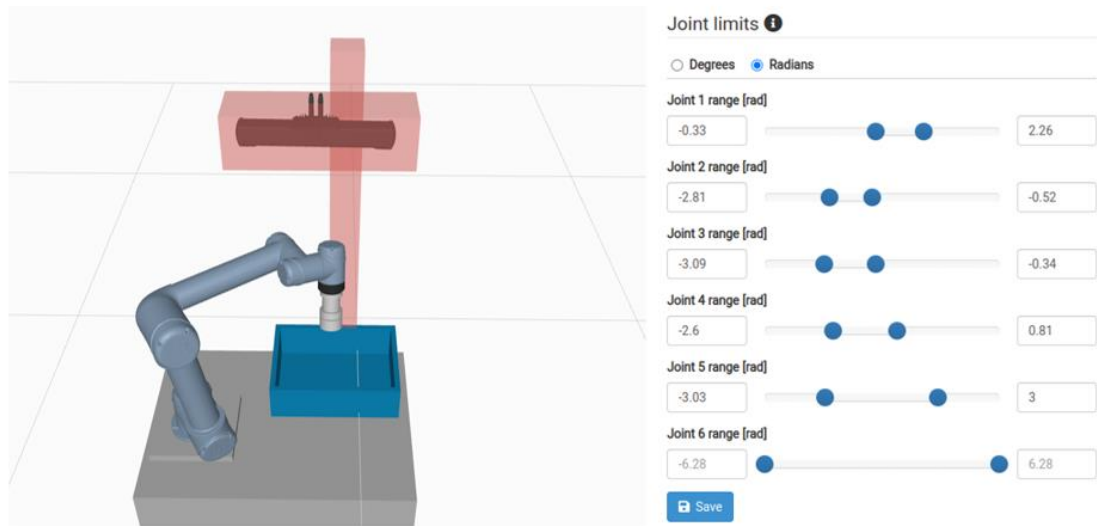


Figure 22. *Bin Picking Studio path planning restrictions*

4.4 Depth camera calibration

Camera calibration was not required for the Basler (Basler, 2022b) and Sick (SICK, 2022b) depth cameras, as the camera sensors were factory calibrated and they are not used in trajectory planning. The Photoneo scanner had to be calibrated however, to transform the default coordinate space of the scanner to the coordinate space of the robot controller. The calibration of the camera sensor is completed with a calibration tool provided with the vision system, which was attached to the tool flange of the robot manipulator. The goal of the calibration is to provide nine calibration points, by imaging the calibration tool in different poses around the bin volume. To achieve high calibration accuracy, these nine points should cover the whole volume of the bin. Wide range of robot manipulator joint values is needed to achieve high calibration accuracy. The calibration tool and calibration process are presented in figure 23.

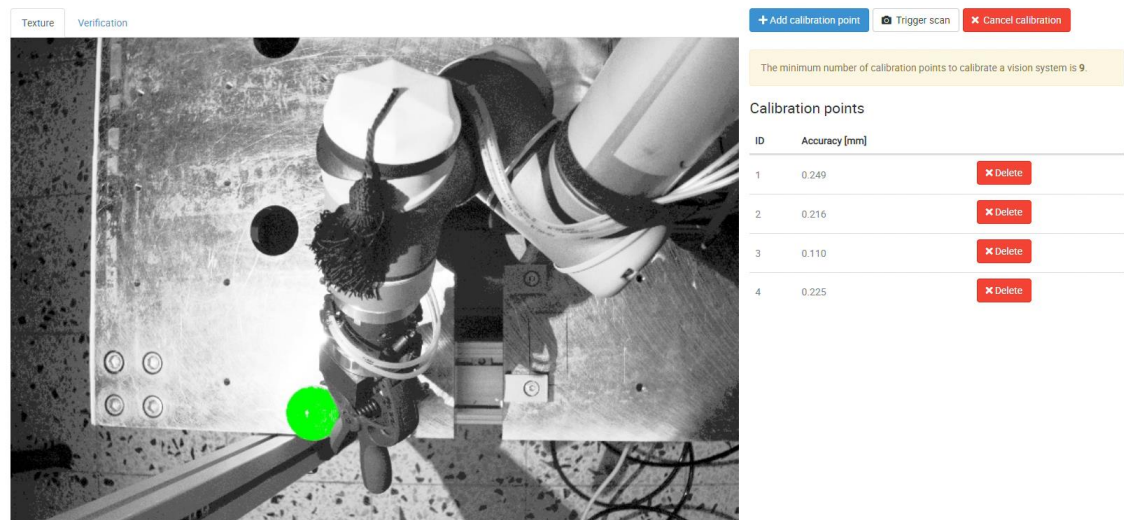


Figure 23. Photoneo depth camera calibration

5. RESEARCH WORK

This chapter presents the research work completed during the thesis. Chapter 5.1 presents and analyses the results of the research work common between all the technologies, and chapter 5.2 presents the results of pilot scale demonstrations. Finally, chapter 5.3 presents the results of Photoneo BPS accuracy evaluation.

5.1 Comparison between 3D- depth cameras

The comparison was completed with 750 *mm* distance between the surface of the bin and the camera sensors. This distance was selected with the constraints of the testing environment in mind. Too short distance would limit the movement of the robot due to potential collisions with the vision system, and too high distances were prevented by the operational limits of the structured light scanner (458 *mm* ... 1118 *mm*) (Photoneo, 2021, p. 2). All cameras under comparison were turned on for half an hour before taking the depth images. This was done to provide enough camera warm-up time, to guarantee stable operation temperature. This was only required by the Basler Blaze-101 (Basler, 2022a), but the same procedure was followed with all cameras.

The camera parameters were adjusted according to application manual recommendations. The parameter tuning was done by observing the point clouds generated by the depth cameras, until the undesired noise of the point clouds were minimized. The key features to adjust were the interreflection filter with the structured light scanner and High Dynamic Range (HDR) with the ToF- camera. Both of these options were highly effective in removal of incorrectly computed points in the point cloud. The stereo vision camera had an automatic tuning option that was utilized in the optimal parameter selection.

5.1.1 Reflective surfaces

The reflective surface recognition of the cameras was evaluated with simple geometrical shapes, consisting of rectangular, cylindrical and polygon shapes. The objects for this comparison are listed in figure 12 with object ID's 1...4. In order to compare how shiny surfaces affects plane and edge detection, these objects were reproduced by 3D- printing them for a matte- comparison. The bin configuration for this evaluation included a backplate at the bottom of the bin. This enables the research to focus more on the objects themselves, as there is less noise coming from the reflective background of the bin. The results of this evaluation are presented as PLY point clouds, edited by CloudCompare-

program. Modifications to the point clouds include common height mapping and point size.

The first set of depth images were taken from large objects that match the minimum object size recommendations of the stereo vision and ToF- cameras ($100\text{ mm} \times 100\text{ mm} \times 100\text{ mm}$). The goal of this comparison was to see how accurately the different depth cameras can record the edge fidelity of the objects, with different camera resolutions. The higher resolution of the structured light (3.2 Mpx) compared to the ToF and stereo vision (0.32 Mpx) is clearly seen in the edge fidelity, which is presented in figure 24. The vastly greater resolution of the structured light scanner highlights, why the results of this evaluation cannot be based on the edge fidelity of the objects alone. Even with large objects, the difference between 0.3 MP and 3.2 MP resolution is easily visible. This result was expected, which is why the analysis of the results is based on the recognition of geometrical shapes such as planes, corners, and angles. The depth images also clearly present the interpolation feature of the ToF- camera for missing data. This can be seen from data generated on the sides of the objects, compared to the black areas, present in both stereo vision and structured light. These areas are occluded from the camera view, leading to missing data with both the structured light and stereo vision cameras.

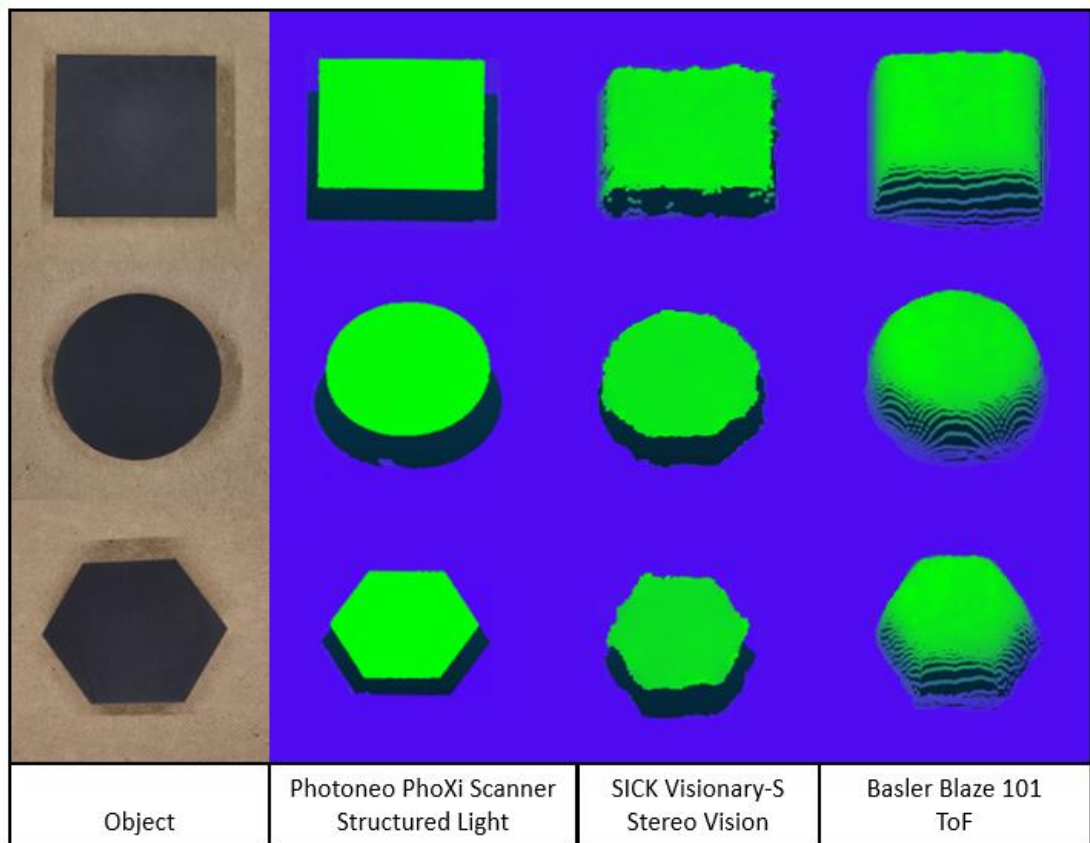


Figure 24. Simple geometrical shapes imaged with different depth cameras

The next sets of images were taken from the shiny metallic blanks, together with matte 3D- printed objects. Figure 25 represents object ID 1 in three different sizes ($20 \times 20 \text{ mm}$, $40 \times 30 \text{ mm}$, $60 \times 40 \text{ mm}$). The main goal with this object was to observe how well the cameras can recognise the planar surface and sharp corners of the object.

The results of the stereo vision camera are similar between the matte and reflective objects. There is however minor loss of data, visible at the edges of the objects. The results of the ToF- camera are the same, with ToF having slightly better edge fidelity out of the two. In this case, the interpolation for missing data is able to smoothen the corners by fill in some of the missing data. Common result between stereo vision and ToF is the inability to recognise the shape of the smallest object. This result however is mainly because the object is far below the recommended minimum size of objects for these cameras. The high resolution of the structured light camera is able to accurately capture the planes and edges of all of the object variations. This applies also for the smallest object, even when it falls under the recommended minimum object size of $40 \times 40 \text{ mm}$. By inspecting the point clouds between the matte and shiny objects, the results are equal without data loss between the two sets of data.

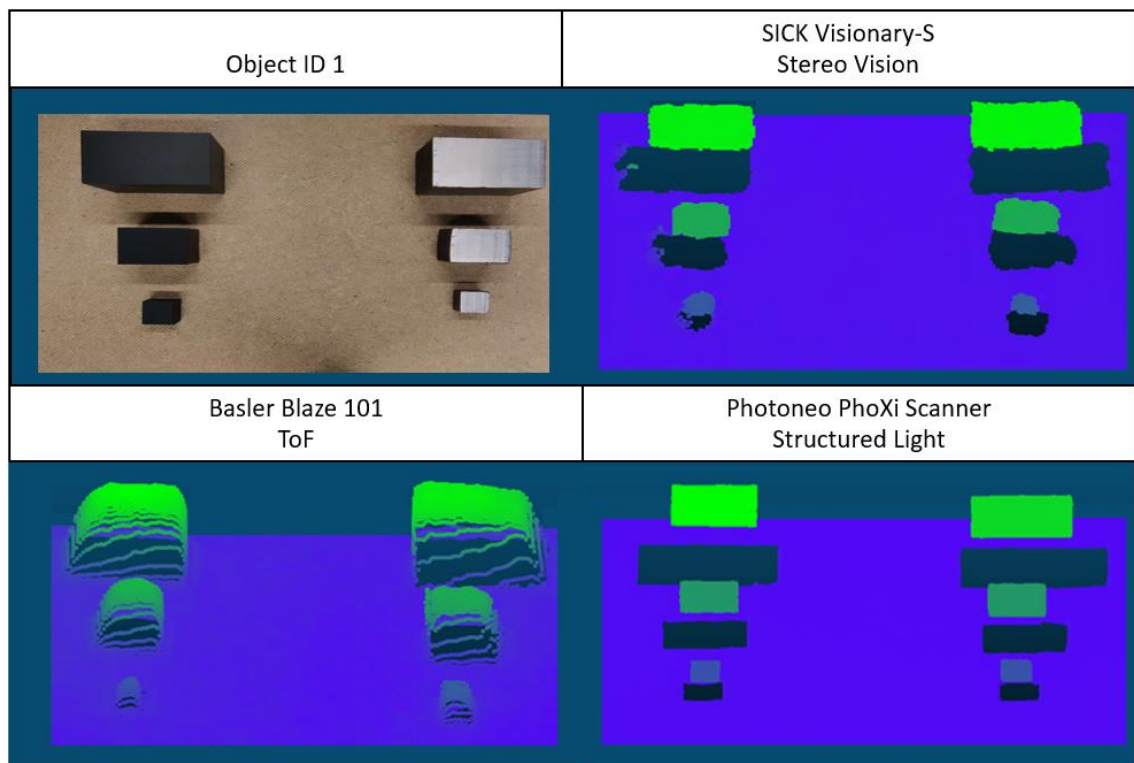


Figure 25. *Rectangular shapes imaged with different depth cameras*

The next comparison was completed with object ID 2, also in three different sizes ($D22, D37, D53$). The main goal with this object was to observe how well the cameras can recognise 45° convex angles of the polygons, and if it is mistaken for a round object. The results of this are presented in figure 26.

The results of the polygon shape share similarities with object ID 1. Both the stereo vision and ToF- cameras have a minor loss of data with the reflective objects, compared to the matte counterparts. This is most visible at the corners of the objects, which deforms the shapes to look more like cylinders. With polygon shapes, the stereo vision performs better than ToF in terms of point cloud quality. The poor surface quality of ToF could be explained by the uneven cutting markings of the object, which cause measurement noise and cause the surface to seem like non-planar. This is most notable with the largest polygon shape, where the noise is visualized by the height map colour coding the surface with varying shades of green. The second factor for the lower edge fidelity could also be noise originating from lost and mixed pixels. This phenomenon originates from the projected light hitting the edges of the object, resulting in a mixed measurements (Kim et al., 2013, p. 681). The results of the structured light scanner are in line with the results of object ID 1. The planes of the objects and corners have sharp details, and there are no visible differences between the two sets of data.

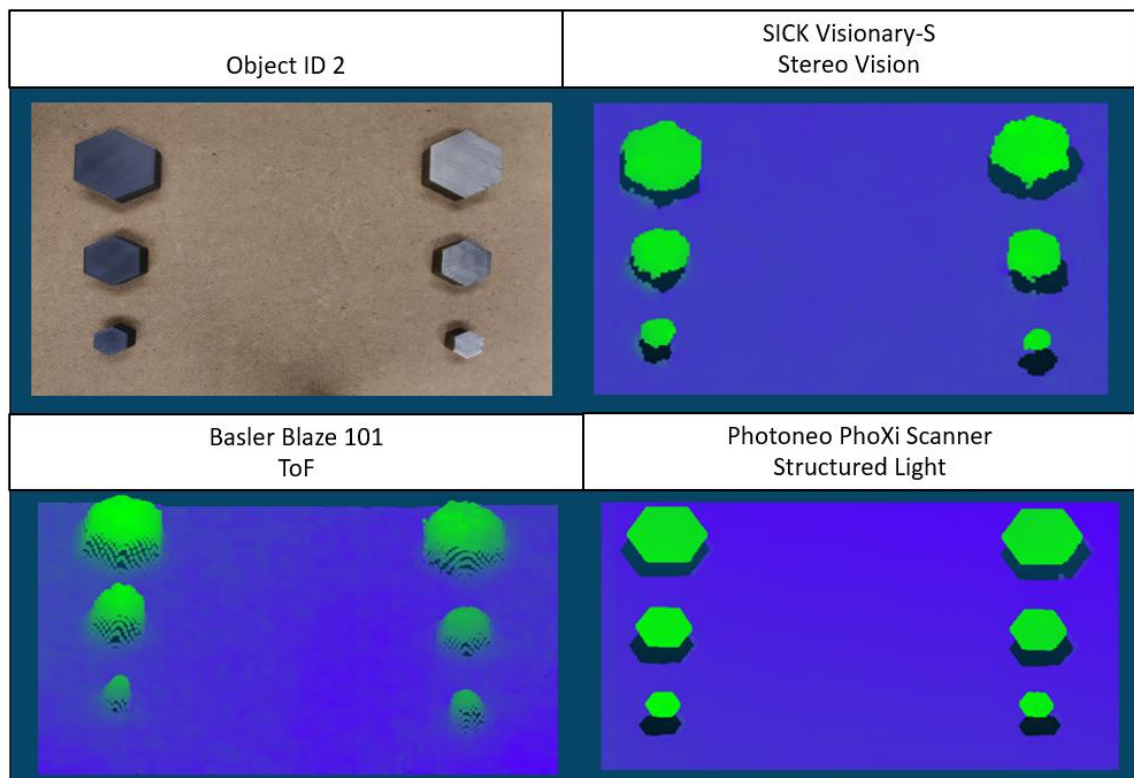


Figure 26. *Polygon shapes imaged with different depth cameras*

Finally, the comparison was completed with cylindrical shapes. This comparison was repeated in three different configurations, in order to evaluate both round shapes from the top and convex surfaces from the side. The main goal with this object was to observe how well the cameras can detect circular edges and convex surfaces. Figure 27 represents objects ID 3 and 4 in two different sizes ($D25, D70$).

The results of both stereo vision and ToF show that the edge fidelity of round edges is higher, compared to object IDs 1 and 2. By comparing the matte and shiny counterparts, there is also less lost data on the reflective objects. The interpolation feature of ToF has great results with round objects, where the edges look visibly better than stereo vision. The results of the structured light scanner are in line with the results of object IDs 1 and 2. The plane of the object and circular edges are clearly recognisable. With high objects, the different projection angles of the projected patterns can also be recognised from the large areas of occluded bin background represented by black colour.

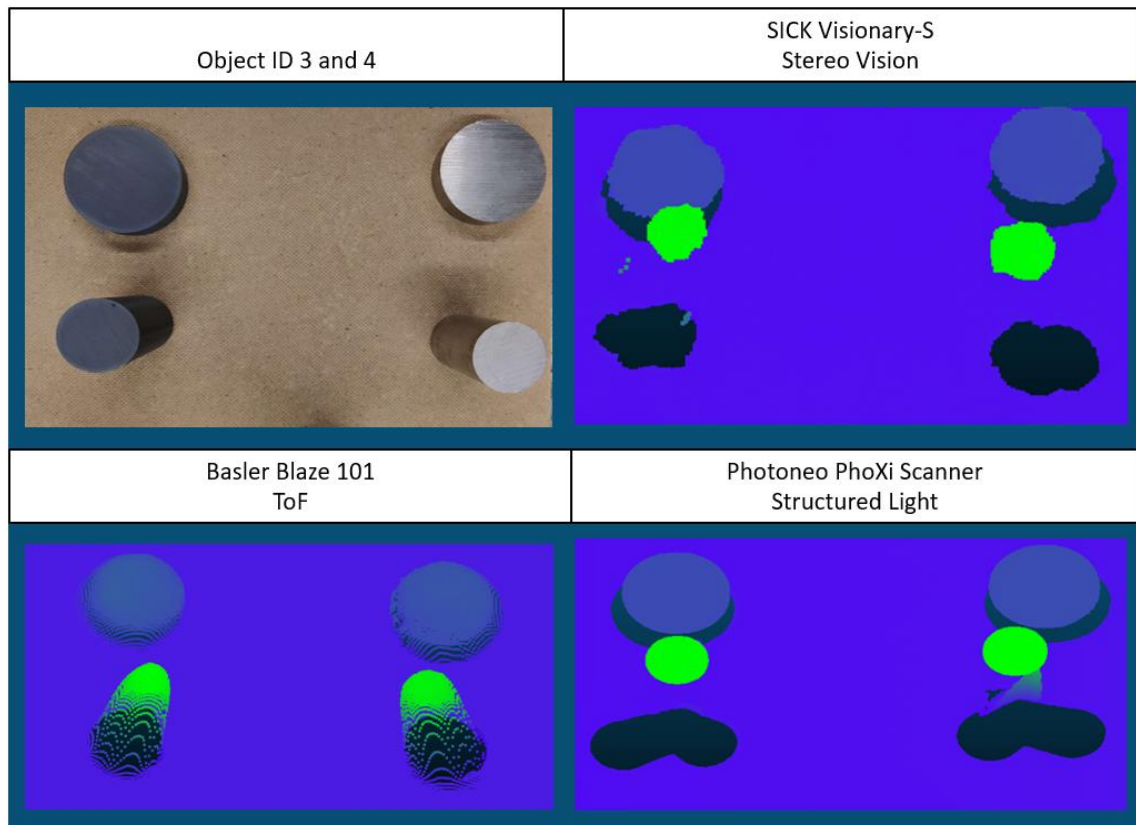


Figure 27. *Cylinder shapes imaged with different depth cameras*

The angular surface detection was evaluated in two different configurations, where the objects were rotated 90° between the comparisons. The reason behind the two different configurations was to observe how this affects the results of the structured light scanner. Figure 28 presents the results of a case, where the cylinders are parallel towards to the projected structured light patterns. The results show a common challenge between all

the technologies, a combination of convex, shiny surfaces. With stereo vision, this is visible as incorrect depth computation of the reflective cylinders. This phenomenon was viewing angle dependent, caused by poorly correlated corresponding regions. Results of the ToF- camera present the same challenges. The shiny cylinders are captured poorly, and even the matte, convex surface is problematic. The results of ToF- camera also warps the depth of the scene, as a results from Multi Path Interference (MPI). ToF- cameras work under the assumption that each light ray is reflected only once. When parts of the emitted light reflect multiple times, this leads to incorrectly estimated depth. The MPI- effect and methods to counter it are presented in more detail by (Agresti and Zanuttigh, 2019, pp. 355–371). The result of the structured light scanner are better, without errors in the depth values of the point cloud. The convex, shiny surface however causes loss of data because most of the projected light is reflected away from the camera (Song et al., 2013, p. 1). This effect is clearly visible by comparing the matte and reflective objects.

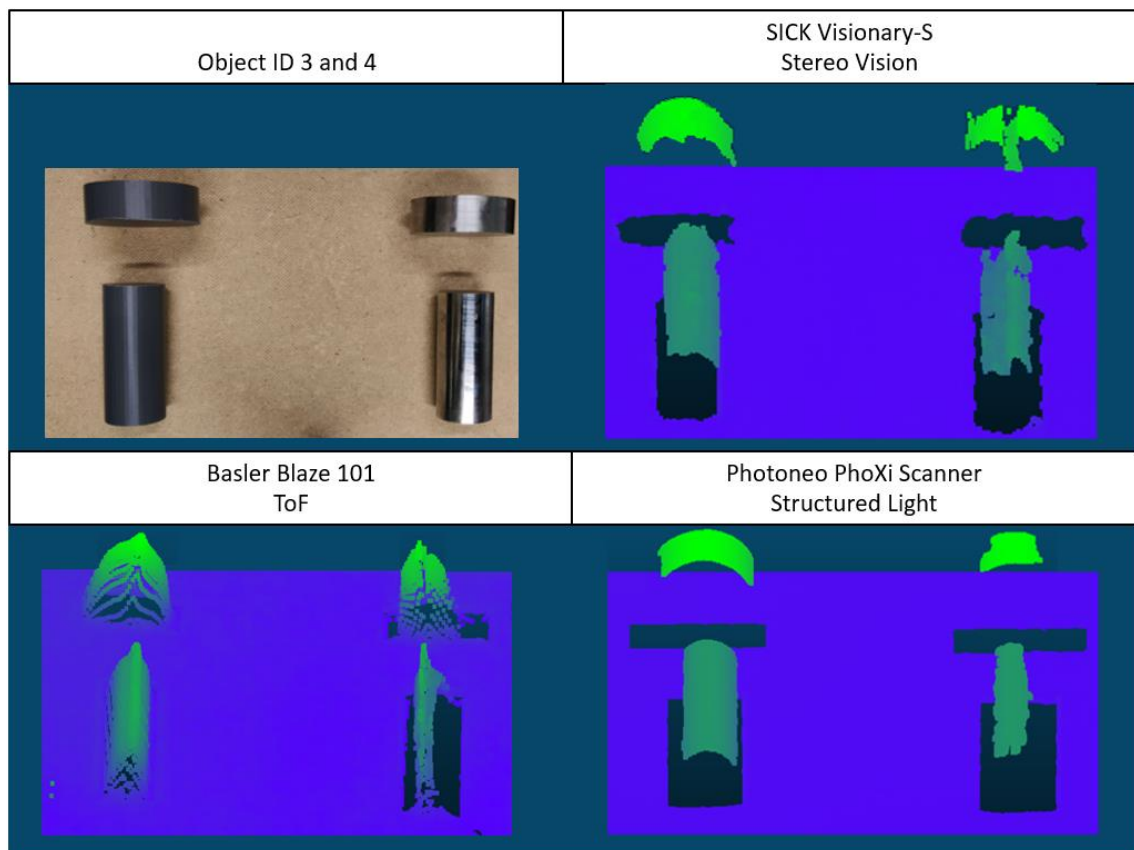


Figure 28. *Cylinder shapes imaged with different depth cameras. Cylinders orientated parallel to the structured light patterns*

Figure 29 presents the results of a case, where the cylinders are perpendicular to the projected patterns of the structured light scanner. The results are comparable to the results of the previous test with different cylinder orientation. The only visible difference between the two sets of data are the point clouds of the shiny cylinders, produced by the stereo vision camera. This is a result of different viewing angle, caused by the orientation of the cylinders. This has resulted in better results with object ID 3, but at the same time decreased performance with object ID 4. With both the ToF- and structured light technologies, the orientation of the cylinders does not have significant changes between the results.

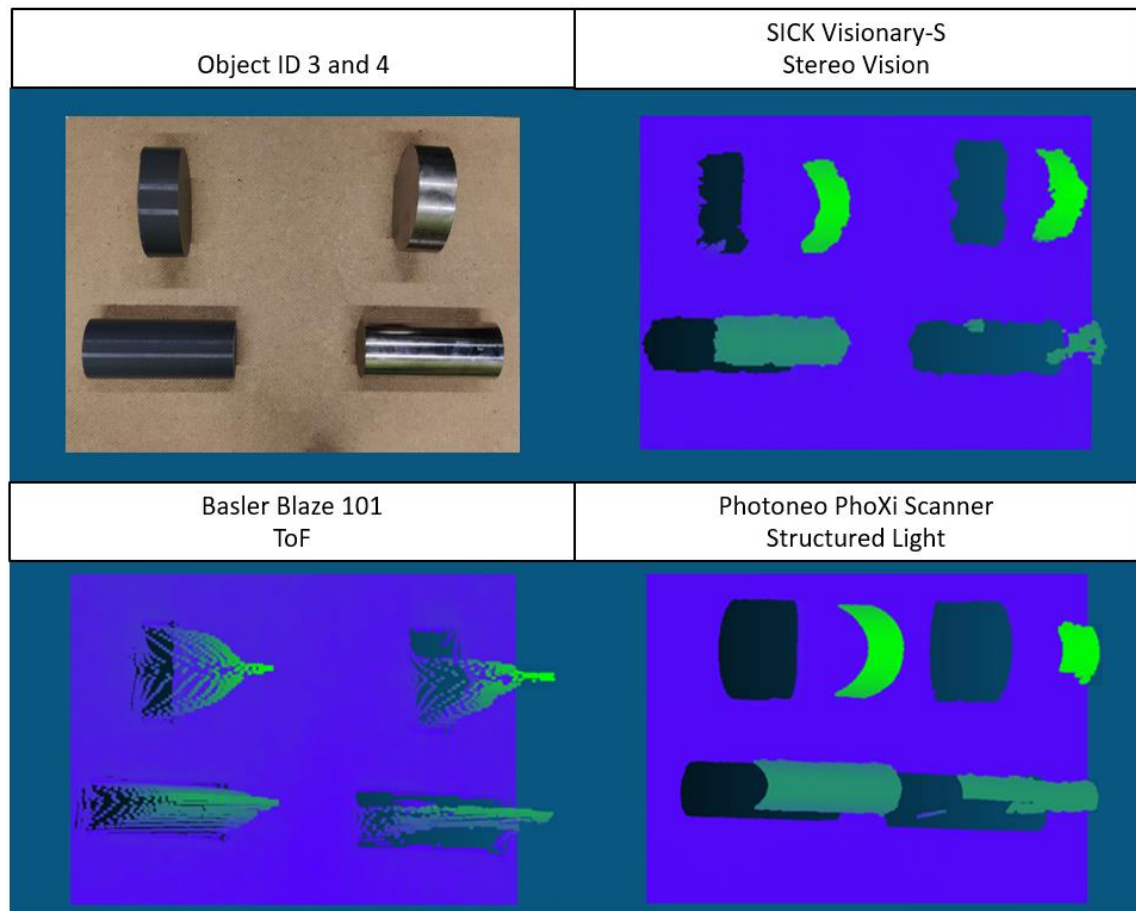


Figure 29. Cylinder shapes imaged with different depth cameras. Cylinders orientated perpendicular to the structured light patterns

5.1.2 Complex geometrical shapes

This part of the research focuses on complex geometric features. The bin configuration and point cloud modifications described in chapter 5.1.1 are applied. The first set of objects are object ID's 5...7. These objects are semi-finished products, where the focus is on rounded edges, planar surfaces, and the holes in the objects.

Figure 30 presents the results from the semi-finished products. The structured light scanner is capable of recognising all of the objects in high detail, with clear edge fidelity and surface details. Even the convex surfaces of object ID 5 are captured in high detail, which proved to be challenging with highly reflective objects. Stereo vision has comparable results, where the shapes of the objects are clearly visible. The lower resolution is detectable from the edge fidelity, but the small hole of object ID 7 and convex surfaces of object ID 5 are still clearly visible. ToF- has similar results as the simple geometrical features from chapter 5.1.1. Object IDs 6 and 7 with planar surfaces, round edges and holes are captured well, comparable to the results of the stereo vision. The problematic properties are the convex surfaces of object ID 5. These surfaces reflect the light away from the camera, resulting in incorrect depth estimation and some of the data being lost.

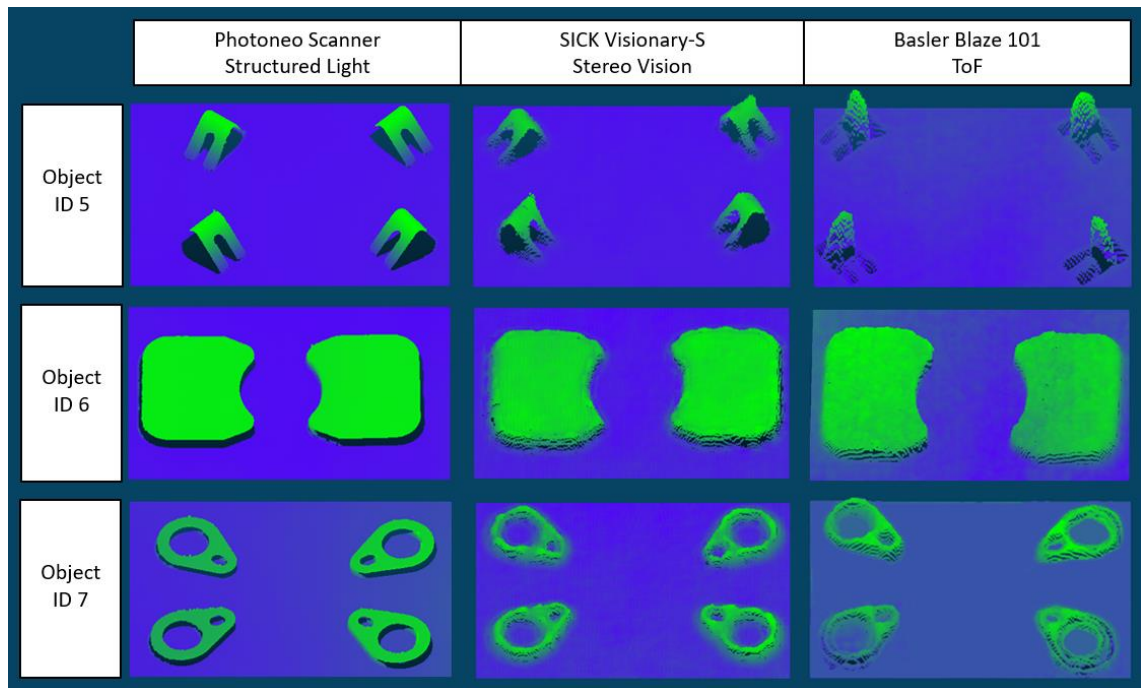


Figure 30. Point clouds from semi-finished products

Figure 31 presents the results from the connector components, where the main focus of comparison is the combination of convex and planar surfaces. The structured light scanner is again capable of recognising all of the objects in high detail. The benefit of high resolution enables the camera to capture even the small height differences of the smallest of components. With a close inspection of the point clouds, even the imprints and small features of the objects are recognisable. With stereo vision, the results are noticeably worse. The height differences of the objects are not recognised properly, and the objects seem to have flat surfaces. This phenomenon is caused by the combination of several edges, causing edge erosion. Because the edges of the target are viewpoint dependent, this causes the edges to be smoothed, resulting in loss of data (Kadambi et al., 2014, p. 10). By comparing the ToF- camera against the stereo vision camera, the results are noticeably better. The height differences are visibly noticeable, and even some of the details of the objects can be observed, regardless of the low resolution.

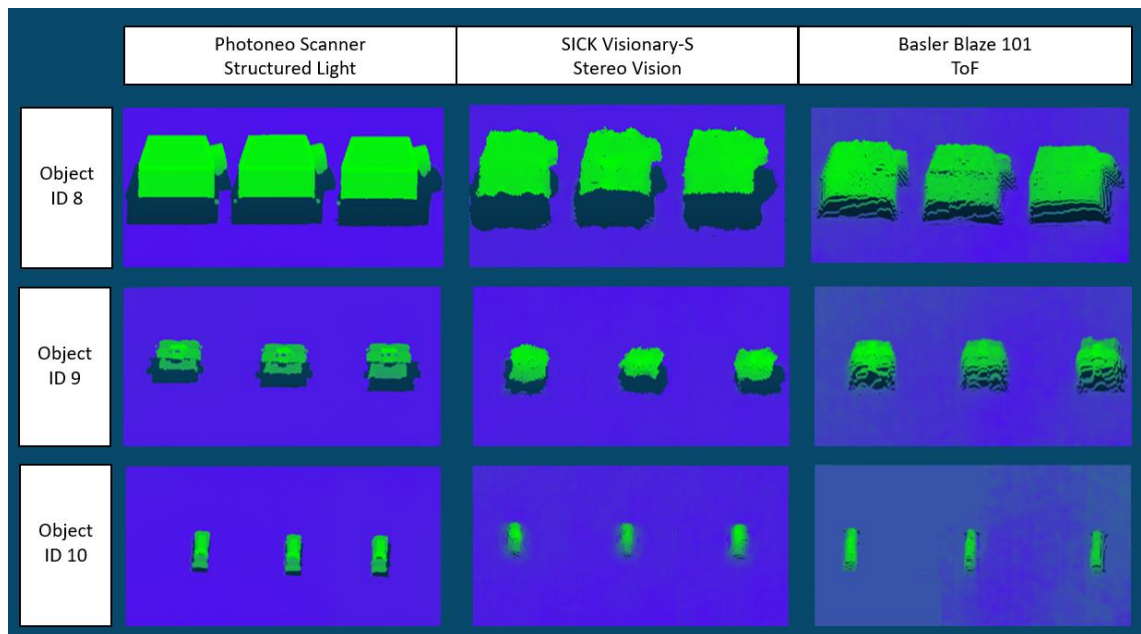


Figure 31. *Point clouds from connector components*

The next set of comparison was made with subassembly components. These objects have a wide range of different sizes, materials, and surface textures. A common discovery between all of the technologies were the importance of proper camera parameters. This was noticed with the object ID 14, where the porous surface initially had very poor quality in the point clouds. The camera filters mistaken the porous surface as unwanted reflections, which led to a lot of lost depth data. Figure 32 presents the results from the subassembly components. The structured light scanner was able to accurately detect the small holes and height of object ID 13, and the porous surface of object ID 14. The complex surface of object ID 15 was also accurately captured in high detail. The only

problems the structured light scanner faced came from object ID 16. The combination of shiny paint coating, convex surface and very small size caused the scanner to be unable to accurately detect the object. Some parts of the object were still visible, but not enough for an accurate localization. With stereo vision, the details of object ID 13 are lost due to the low resolution, but the porous surface of object ID 14 is successfully captured. The complex surface of object ID 15 is also captured successfully, similar to structured light. The object ID 16 was practically invisible to stereo vision, where only a few data points were successfully captured. By comparing the results of stereo vision and ToF, there are some differences. The small holes of object ID 13 are detectable with ToF, while they were practically invisible with stereo vision. The most notable difference however was with the porous surface of the object ID 14. The porous surface is problematic and is captured as a concave surface, instead of planar. This phenomenon can be seen by the colour variations of the height map with different shades of green. The results of objects ID 15 and 16 are comparable to stereo vision. The complex surface of object ID 15 is captured in slightly less detail and details of object ID 16 are not visible due to the low resolution.

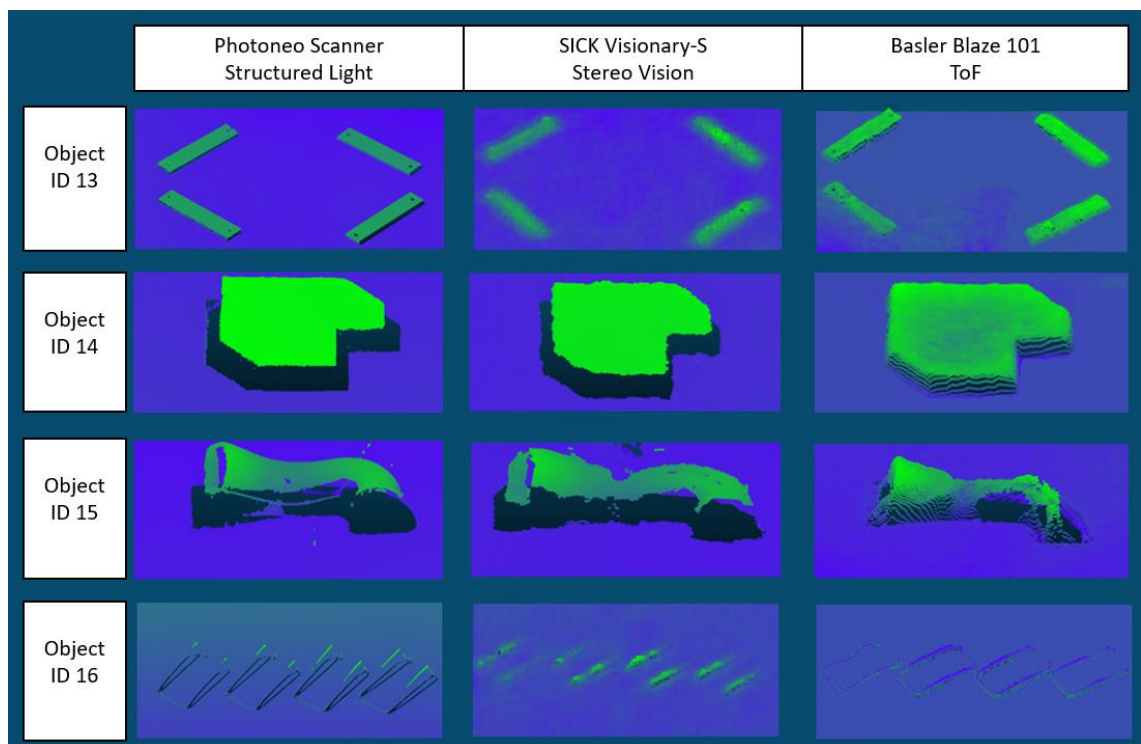


Figure 32. *Point clouds from subassembly components*

Final test with complex objects was to image all of the subassembly components at once. This test aims to evaluate how wide range of different textures can the cameras detect at once. This is done by comparing how well the technologies can simultaneously detect objects of different materials and surface textures. The results of this comparison are

presented in figure 33. The results show that the structured light scanner performs same as before, maintaining all of the surface details of all of the objects. The structured light scanner was also the only depth camera, where camera parameters did not need to be changed between individual and multiple objects. The results from the ToF- camera show how some of the data acquired from individual objects has now been lost, as a result from the exposure time settings of the camera. It was noted that single exposure time setting is not enough to simultaneously capture the whole scene, as matte and reflective materials require different camera settings for accurate detection. In order to achieve a good result from the plastic tube, a higher exposure time ($> 600 \mu s$) was required. This however lowered the quality of the rest of the scene, which required lower exposure times ($< 300 \mu s$). The stereo vision camera faced similar problems as the ToF- camera, but to a lesser extent where only the reflective aluminium part suffered data loss.

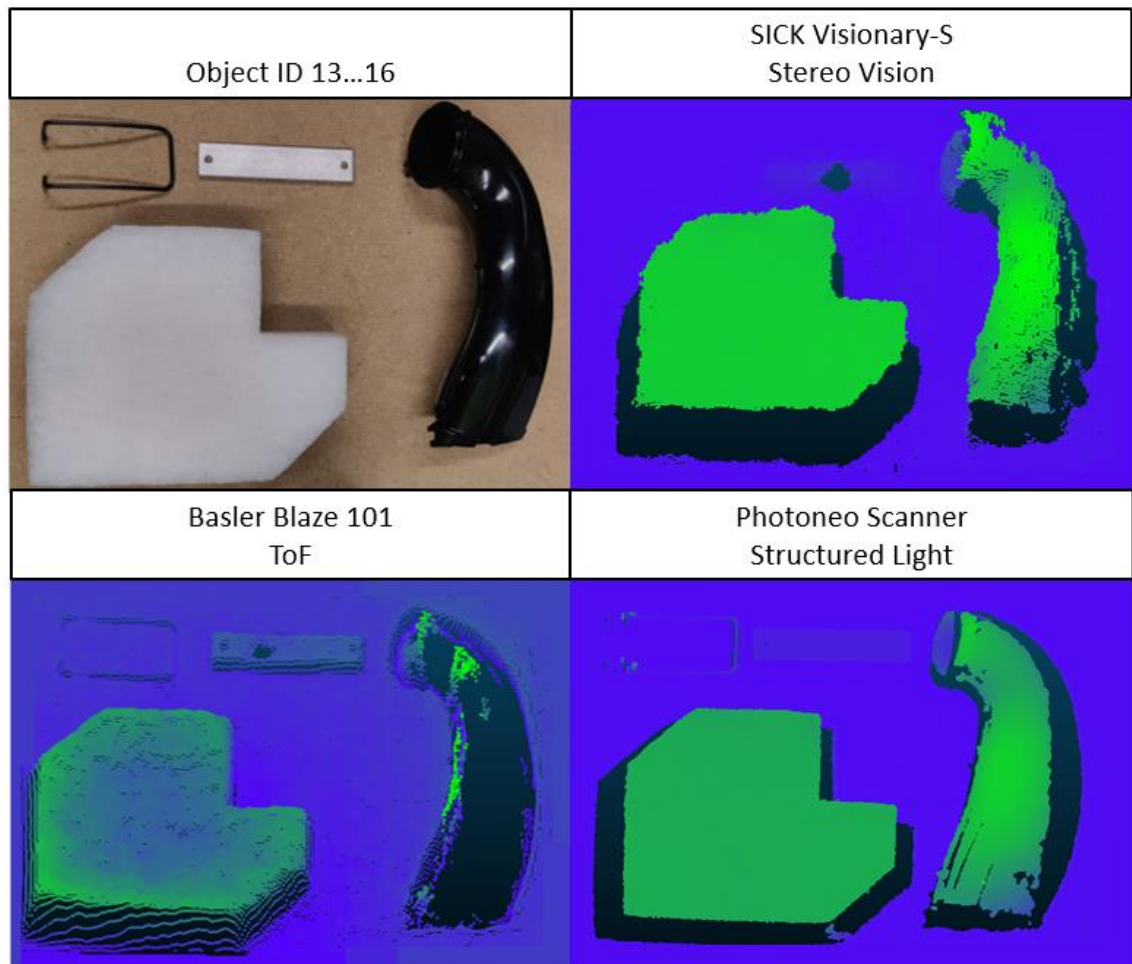


Figure 33. Point clouds from imaging multiple objects at once

5.1.3 Tolerance to ambient light

The ambient light tolerance of the cameras was evaluated with different bin configurations and varying amounts of ambient light. The comparison was done with an empty bin with no backplate, rough wooden backplate, painted wooden backplate, and a reflective metal sheet. The two wooden backplates were chosen to have non-reflective surface types. The empty bin was chosen to represent the more standard, reflective bin. Another goal of selecting these surface textures was to see if any of them cause the correspondence problem for the stereo vision camera. The reflective metal sheet was included to see how well the cameras perform with highly reflective materials, and if strong ambient light combined with reflective surfaces is problematic. These different bin configurations are presented in figure 34. The empty bin is presented in the top left corner, rough wood in the top right corner, painted wood in the bottom left corner and reflective metal sheet in the bottom right corner.

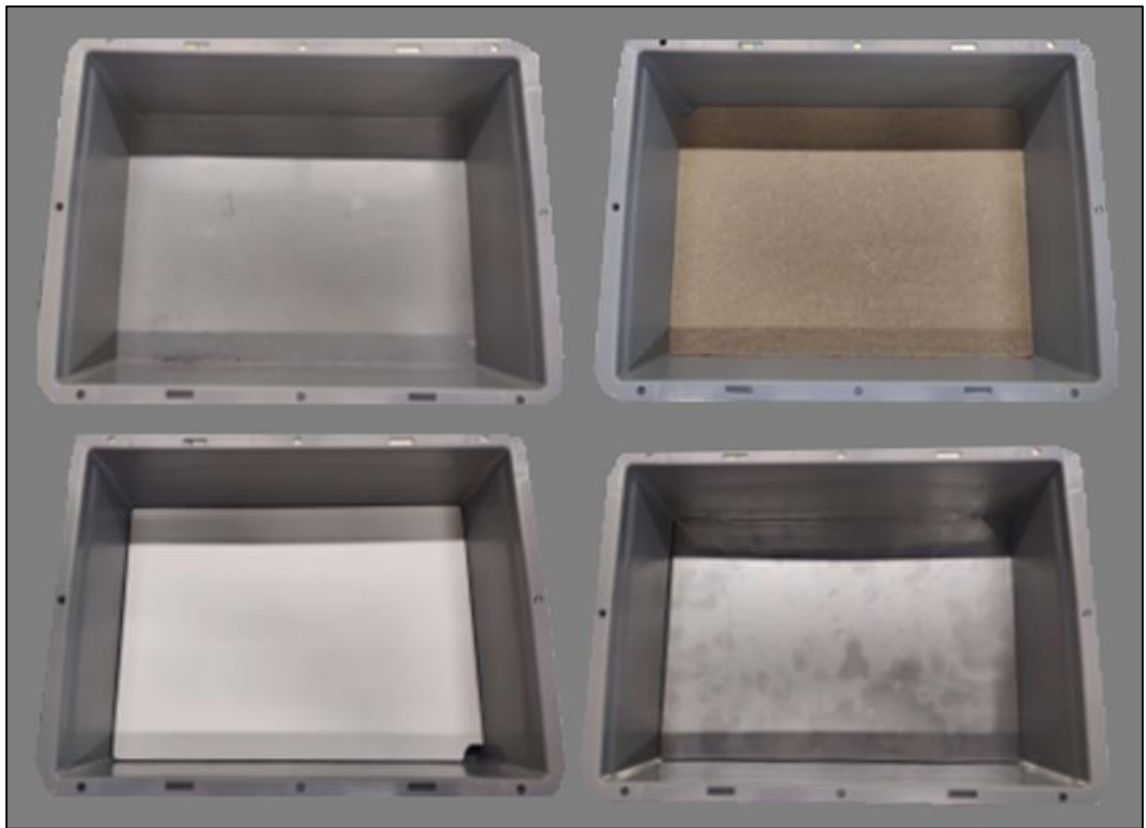


Figure 34. *Different bin configurations*

To produce comparable results, a complete set of images was captured from one bin configuration at a time without moving the bin or the camera. The only variable was the ambient light, which was controlled by the lighting of the room and a LED- illuminator next to the vision system. The LED- illuminator was placed at 250 mm distance from the edge of the bin at a height of 550 mm. This location was chosen to have it as close to

the bin as possible, without obstructing the Field of View (FOV) of the cameras. The angle of the light was alternated between $45^\circ \dots 90^\circ$ to adjust the amount of light reflected to the surface of the bins. The LED illuminator has brightness of 1206 *lm* (Sangel, 2018, p. 4) and operates on a different wavelength than the depth cameras.

The results in ambient light conditions are presented in figure 35. The results are presented as point clouds, in native format of each camera. With both wooden backplates, all of the depth cameras were able to image the whole content of the bin, without ghost artifacts or missing data from reflections. By having the cameras perpendicularly above the bins, the stereo vision camera performs the poorest with reflective surfaces. The reflections from the bin and the metal sheet causes correspondence problem, as the reflections are dependent on the angle of the cameras. This assigns invalid disparity values to the algorithm, and it is seen by the holes in the point clouds. These results are in line with similar research conducted by (Nair et al., 2015, p. 1). This problem could be partially resolved by tilting the camera, but with shiny surfaces this phenomenon is hard to entirely remove.

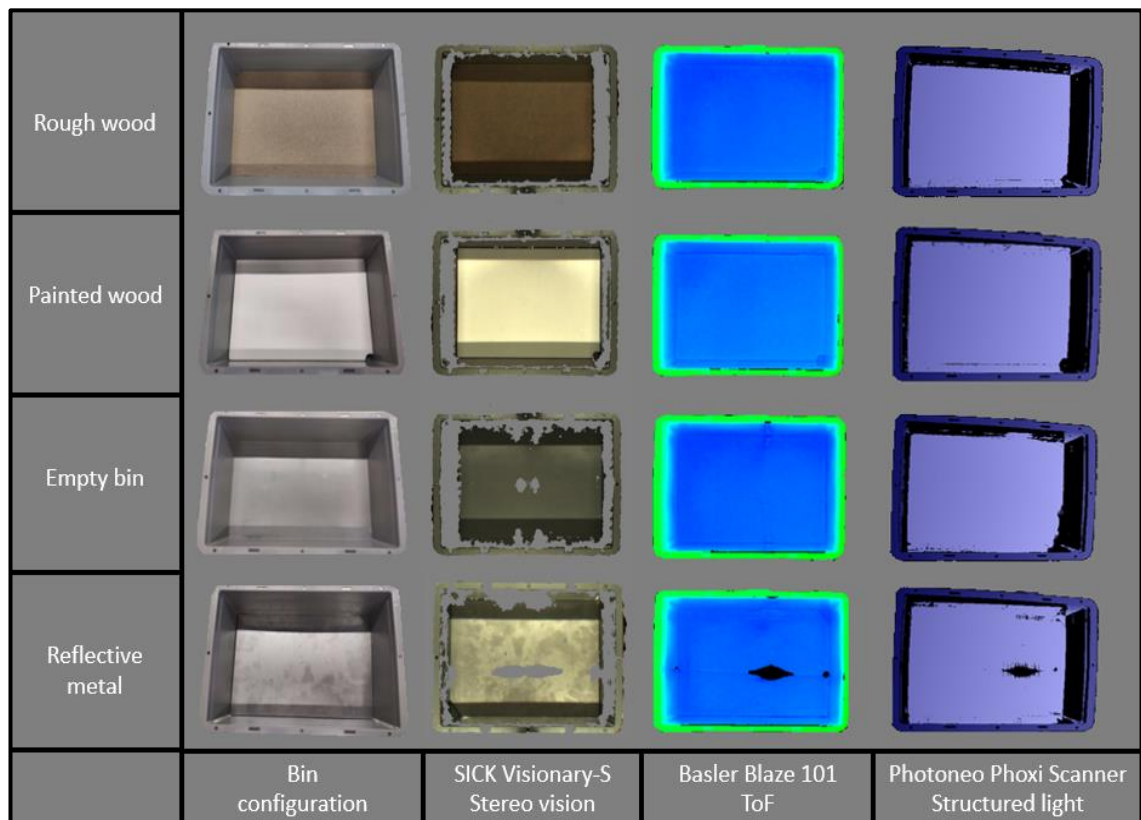


Figure 35. *Bin configuration test results in ambient light conditions*

The ToF- camera performs well with the empty bin but shares similar results with the shiny metal sheet as stereo vision. This result is also based on the angle of the camera, similar to results from stereo vision. When the camera and target surface are parallel to

each other, all of the projected light is reflected back to the sensor, causing over exposure. This phenomenon is also researched by Laukkanen with various ToF- cameras (Laukkanen, 2015, pp. 39–42), having similar results. Structured light has comparable results with the ToF. It performs well with the empty bin, but the specular reflections from the reflective metal sheet causes overexposure at the camera. This makes it not possible to distinguish the structured light patterns in the scene, which leads to missing data in the point cloud. This issue could potentially be solved by interpolating for the missing points, as researched by (Milani and Calvagno, 2016, pp. 31–33).

The experiment was repeated with varying ambient light conditions, and all of cameras shared the same results. The amount of ambient light ranging from a completely dark room, all the way to strong ambient light next to the bin has very little affect to the quality of the 3D- point cloud. This is illustrated below in figure 36 with the results from stereo vision camera. The changes in the ambient light are presented on the top row of the image. This however does not affect the quality of the point clouds, as presented on the corresponding images on the bottom row.

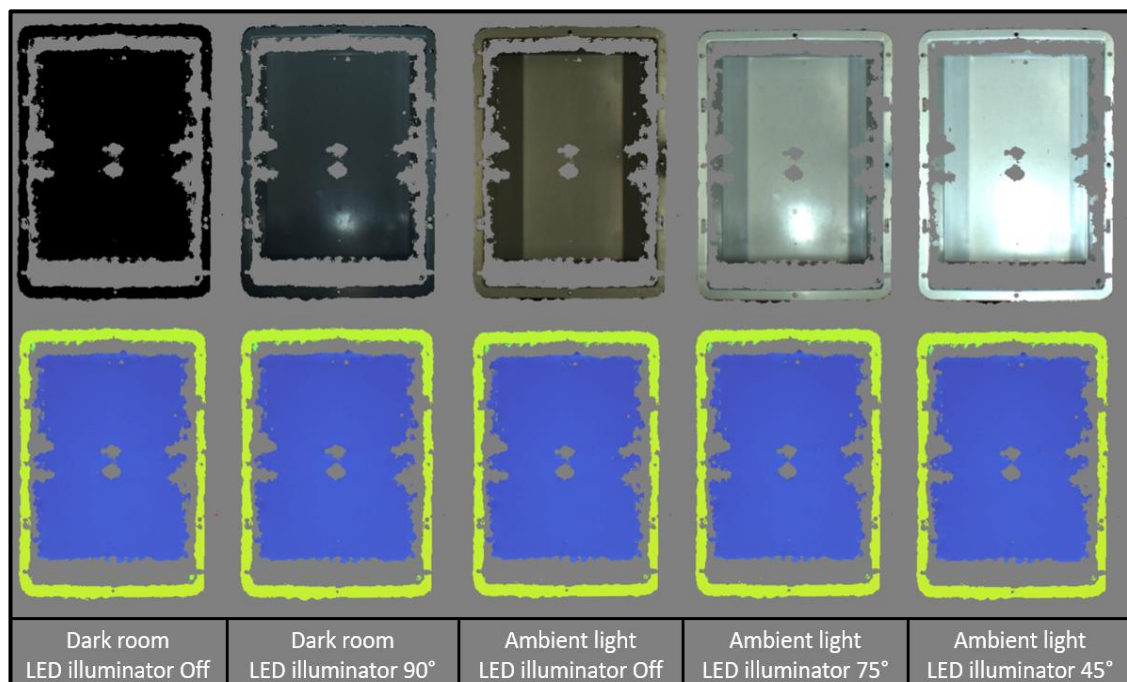


Figure 36. *Ambient light test results with SICK Visionary-S*

Finally, the results were inspected with CloudCompare- software to further analyse the resulting point clouds. The results were collected to a graph, presented in figure 37. This graph presents the densities of the point clouds with different bin configurations and ambient light conditions. The results show that while there were no visible changes in the point clouds, ambient light still had minor effects to density of the point clouds. The clustered bars are used to present the point cloud densities compared against the reference

point cloud. The painted wood configuration was chosen as the reference point because it had the densest point cloud with all of the technologies. The effects of varying ambient light conditions are presented with the error bars at the end of the bar graphs. A common result with the technologies was that ambient light had little effect to the point cloud quality. The results were also similar between the different bin configurations, so the effects of ambient light are also not related to the surface of the bins. The results also show how the different technologies performed with different bin configurations. ToF- camera had the best overall results, with structured light having comparable results. It should be noted that the performance of the ToF- camera was also improved by the integrated interpolation feature, which enabled the camera software to interpolate for missing data. Stereo vision performed similarly with non-reflective materials but had the most problems with both reflective bin configurations due to correspondence problem.

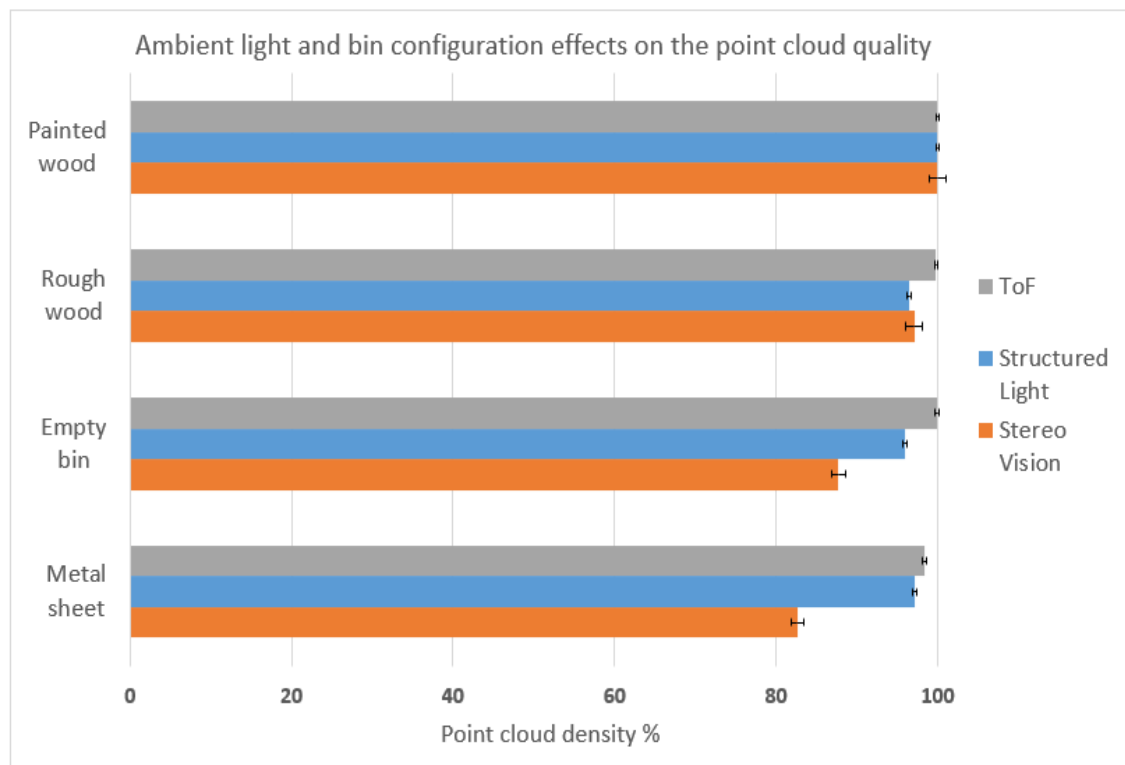


Figure 37. *Ambient light effects on the point cloud density*

5.2 Pilot scale demonstration in industrial environment

The first step of the pilot scale demonstrations was to evaluate the compatibility of the grippers and objects. The grippers chosen for the demonstrations were vacuum- and magnetic grippers, which limited the choice of objects to ferromagnetic objects and light objects with smooth surfaces. As the object IDs 1...4 all have large enough surface area, they were chosen for the magnetic gripper. Respectively object ID's 9...11 have smooth

surfaces, hence they were chosen for the vacuum gripper. The subassembly components were excluded from the demonstrations because the surface types and materials do not support either of the grippers available.

After the objects were chosen, the localization of the objects was tested with the vision system tools. The objects were placed in a bin underneath the depth camera, and the vision system parameters were adjusted until high overlap between the 3D- model and the object was reached. The goal of this stage was to verify that the vision system is capable of aligning the reference models with the point cloud of the scene. Localization of all of the objects were successful, confirming the feasibility of the bin picking task. The process of object localization is presented below in figure 38, where the reference CAD-models aligned to the point cloud are presented in yellow and blue models.



Figure 38. *CAD- model matching to the point cloud of the scene*

After confirming the objects can be localized, the bin picking task was then completed in laboratory conditions. This was done to verify the results of the localization tests. At this stage, the vision system localization parameters were also finetuned and the robot program finalized. This part of the process included localizing the objects and simply moving the objects from one bin to another. This verified that the objects can be successfully localized, and the robot is capable of placing the objects in set orientation. After successful laboratory testing, the last task was to design the pilot scale demonstrations in industrial environment. The environments and processes were familiarized through company visits and the resulting demonstrations were designed according to observations made at the site.

The first application was inside an indoor assembly area, where the collaborative robot and machine vision system were used in a component labelling process. The current

manual completion of this task involves placing the components on a conveyor belt, where they are marked as they travel beneath a spray marker. Both the placing of the unmarked parts and collection of the marked parts is done by hand. The pilot demonstrated two different solutions for this task. In both solutions, the objects are first localized with the vision system and gripped by the robot using a vacuum gripper. In the first solution, the robot utilizes a conveyor belt to move the objects underneath the spray marker. After localizing the objects from the bin, the parts are placed down on the conveyor belt in a set orientation where they travel beneath the spray marker to an unsorted bin. The workflow of the first solution is presented below in figure 39. This method solves the automatic picking and placing of the parts, but the parts still needs to be sorted by hand after they have been marked.

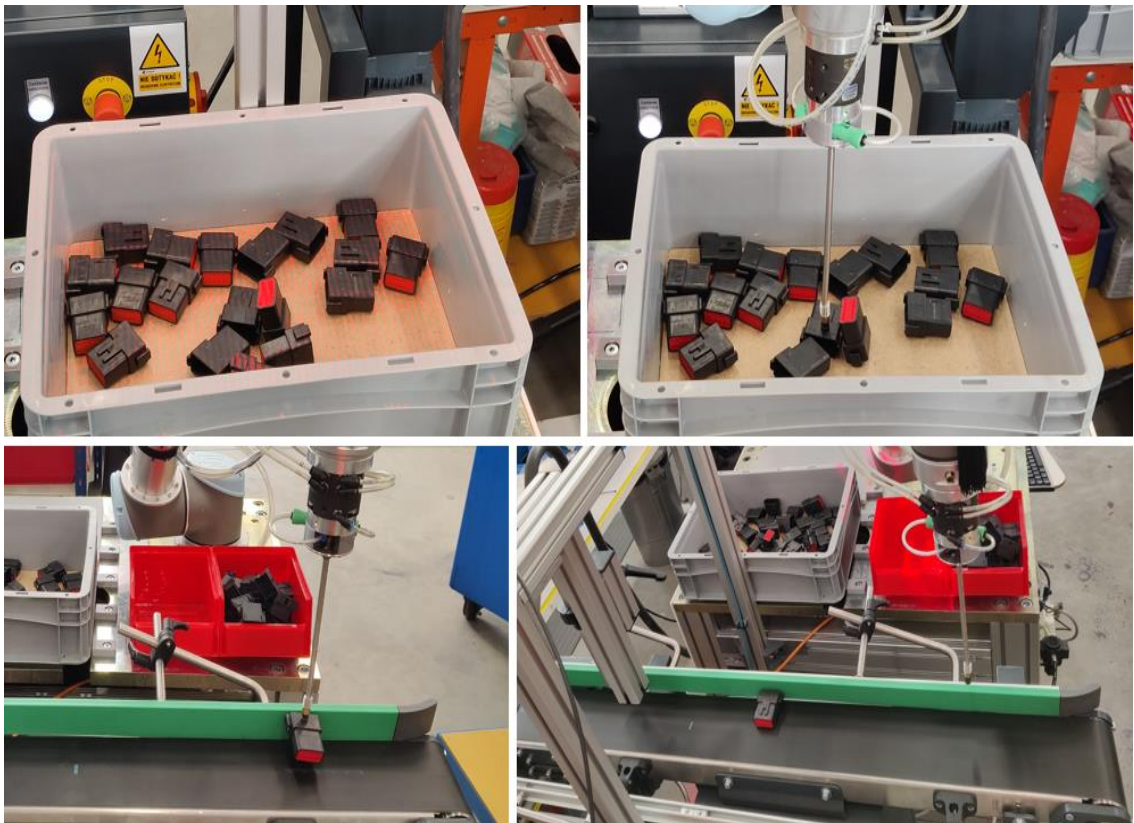


Figure 39. *Workflow of the first bin picking solution*

In the second solution, the task of marking the components is fully automated. This is done by placing the objects beneath the spray marker directly by the robot. This method requires relocation of the spray marker in a way that the robot can reach beneath it, which was simulated by adjusting the trajectory of the robot. The workflow of the second solution is very similar to the first. After gripping the object from a bin, the robot moves the parts directly under the spray marker for the duration of the labelling. After the part has been marked, the robot proceeds to sort the object. The benefit of this approach is

that a conveyor belt is not needed, and the marked components do not need to be sorted by hand. A comparison between the two solutions is presented in figure 40.

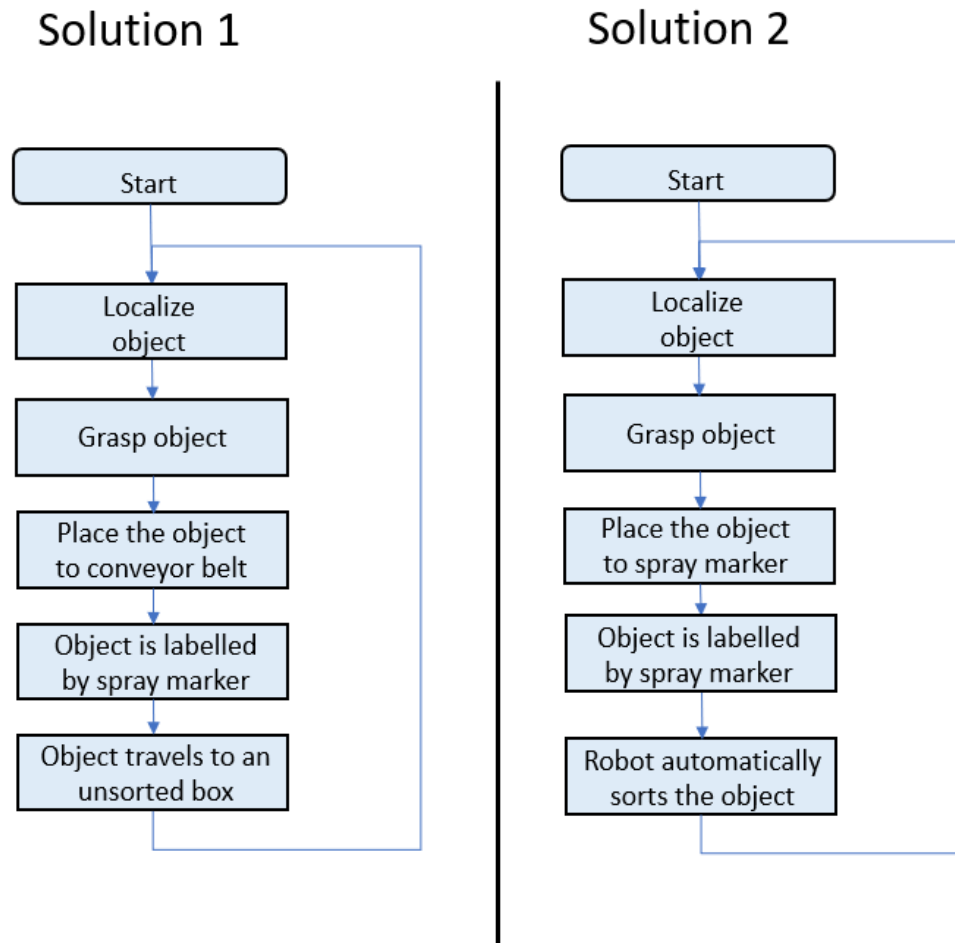


Figure 40. Workflow comparison of two different approaches

To conclude the first pilot demonstrations, the goals were achieved with functional demonstrations of the two solutions presented above. These solutions were presented with three different connector components and two different grippers, to demonstrate the flexibility of the bin picking solution. The vision system was able to successfully localize all of the objects, with no issues even with the objects under the recommended minimum size. During the part placement to the conveyor belt, minor error with the orientation of the parts was noticeable. This happened because of friction, as the conveyor belt was constantly running at a set speed. This could be remedied by having the robot control the conveyor belt, after the parts have been placed down.

The second application was inside an outdoor warehouse, where the collaborative robot and machine vision system were used in a machine tending task. The goal of this pilot was to demonstrate, how bin picking could be utilized in the palletization of metal blanks. The current, manual completion of this task is to palletize a set of metal blanks by hand

and place them within a reach of an industrial robot. The pilot demonstrated an automatic palletization with several different metal blanks (Object IDs 1...4). The objects are first localized with the vision system and grasped by the robot using a magnetic gripper. After localizing the objects from the bin, the parts are placed down on a pallet next to the robot. The workflow of the solution is presented below in figure 41.

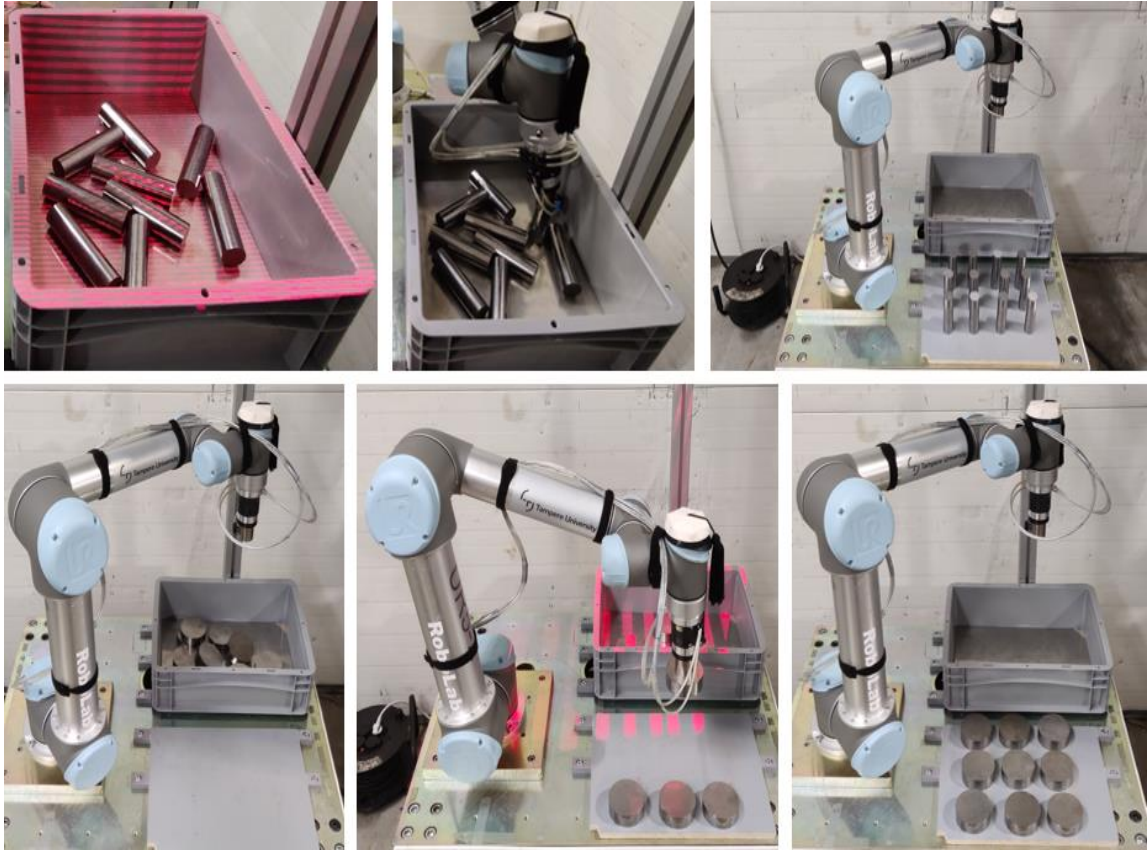


Figure 41. *Workflow of the second pilot demonstration*

The demonstrations with object IDs 1...3 were successful without any failed localizations or collisions. When the bin was filled with object ID 4, there was a collision between the robot gripper and the objects. This happened due to the robot attempting to pick an object from an angle, where the depth camera failed to detect an object. This issue could be avoided by having more strict collision avoidance settings in the vision system, which allowed 3% volumetric collision at the time of the pilot. This result highlighted the importance of a good quality point cloud, as collisions are possible without accurate knowledge of the surroundings. The robustness of the vision system to measurement noise was also demonstrated during the pilot. This was demonstrated with non-laboratory conditions, when there are band saw cutting oil, dirt and metal flakes present in scene. This demonstration also provided valuable information about the performance of the vision system in a more industrial environment, where environmental conditions vary. The objects for this demonstration were covered with cutting oil and metal flakes, as

presented on the left in figure 42. This did not affect the localization of the objects, as seen by the localization results on the right in figure 42. This scene shows how the localization algorithms were able to align the yellow reference CAD-model on top of all five metal blanks in the bin. The reflections caused by the oil, or the piles of metal flakes did not have an effect to the results unless the whole surface of the object was covered.

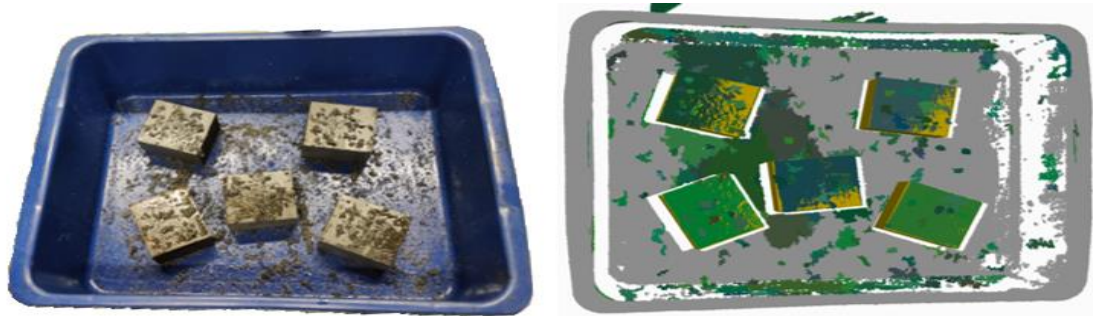


Figure 42. *Localization with oily and dirty objects*

However, a problem was identified while using the magnetic gripper. The magnetic metal flakes started to accumulate in the base of the gripper over time, because the residual magnetism held them in place. This layer of metal flakes eventually prevented the gripper from grasping the objects. This is presented below in figure 43, where the surfaces between the magnet and the object do not have a contact because of metal flakes. In an application where metal flakes are present, a system to keep the gripper clean is required. In the case of magnetic grippers, this could be accomplished as an example with pneumatic air between the grasps.

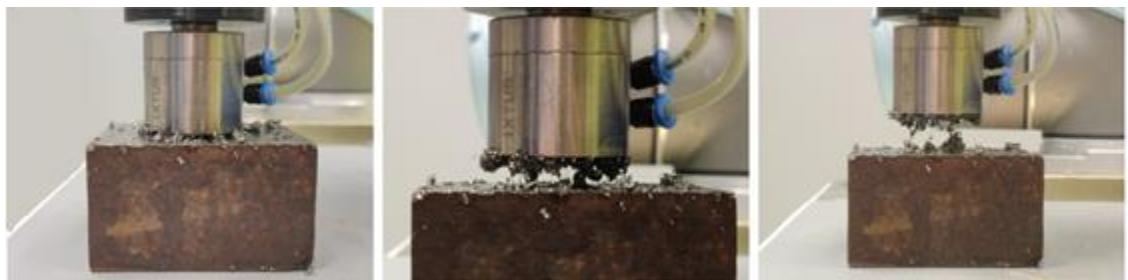


Figure 43. *Metal flakes and magnetic gripper*

To conclude the second pilot demonstration, the goals were achieved with functional demonstrations of the applications presented above. The pilot was able to present how even very shiny objects can successfully be localized and how metallic flakes are not a problem for the vision system. The flexibility of the bin picking solution was also successfully presented with different objects.

5.3 Photoneo Bin Picking Studio accuracy

The true accuracy of a bin picking- system is a combination of multiple factors, consisting of error sources from the robot manipulator, gripper, and vision system error sources. To measure the repeatability accuracy of a Photoneo BPS, a system presented in figure 44 was configured. The goal of this system is to localize and grasp objects with the Photoneo bin picking solution. The accuracy and precision of the grasp is then recorded with a 2D- camera and an industrial vision application, by measuring the displacement of the object against a reference position. The 2D- camera was accompanied by a ring light, as proper illumination is essential for accurate image analysis. The purpose of the ring light was to produce high intensity light, which is reflected back to the camera. To achieve the best contrast between the object the background, the parts of the robot manipulator visible to the camera were also covered with black tape. This approach provided the contrast needed to remove the background from the images and enabled the vision application to form binary images of the objects.

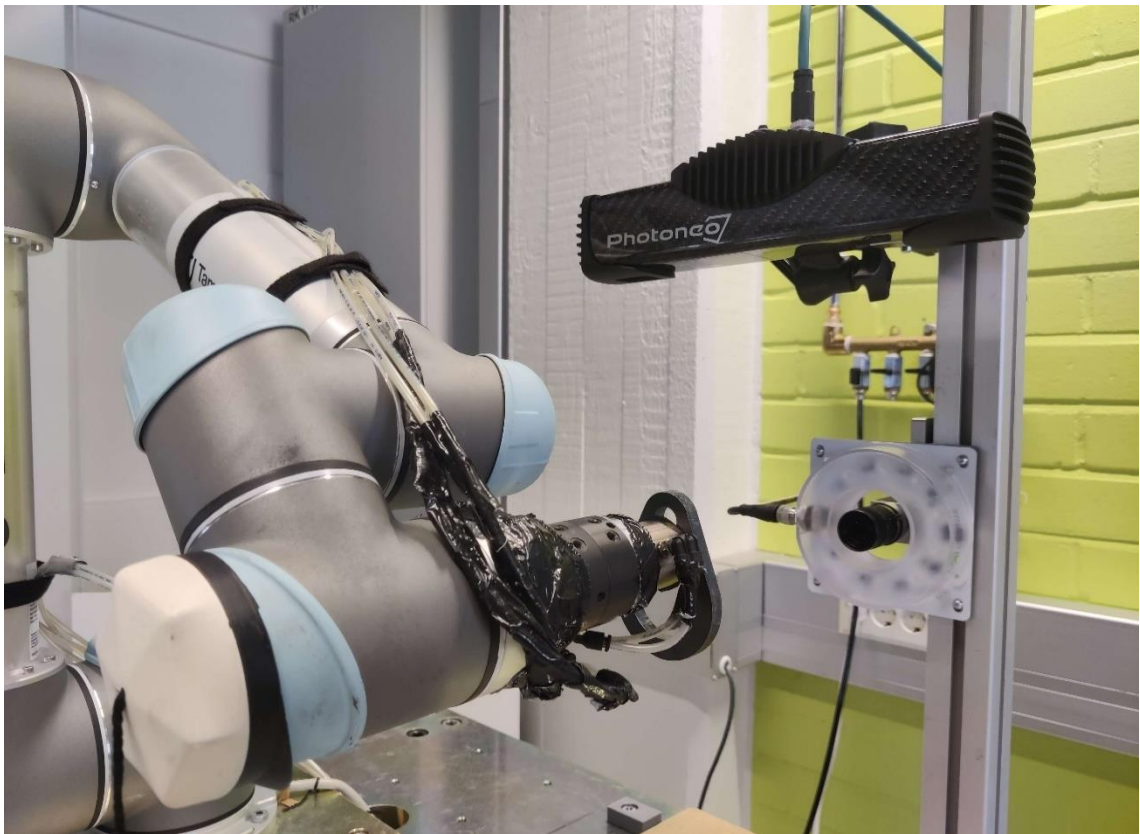


Figure 44. System configuration for bin picking solution accuracy evaluation

The Photoneo vision system was configured with a goal of minimizing localization related errors. By removing small point cloud clusters below 500 voxels and activating the inter-reflection filter most of the noise in the point clouds was removed. Objects with the high-

est overlap with the reference model were preferred to be picked up first to avoid incorrectly localized objects. This approach also increased the overlap of remaining objects, as removal of one source of reflections usually increased the quality of the remaining point cloud. Finally, the control of the magnetic gripper was activated when the distance between the surfaces of the magnet and the target object was very small. This was done to lessen the effect of the object being drawn towards the magnet resulting in positional errors. The industrial vision application (NI Vision builder for Automated inspection) was also designed in a way to minimize localization related errors. The images taken by the camera were first edited with image analysis tools and converted to binary images. These binary images are presented in figure 45. The binary images were inspected by locating a distinguishable feature, which were the round edges of object ID 2, large hole of object ID 7, and the Binary Large Object (BLOB) of the object IDs 10 and 11. The displacement of the objects were then computed by computing the distance between the reference point and the location of the found feature. With object IDs 2 and 7, the distance was computed from the centre of the circles and with object IDs 10 and 11, the distance was computed from the centre of mass. The localization principle of these tools is also visualized below in figure 45.

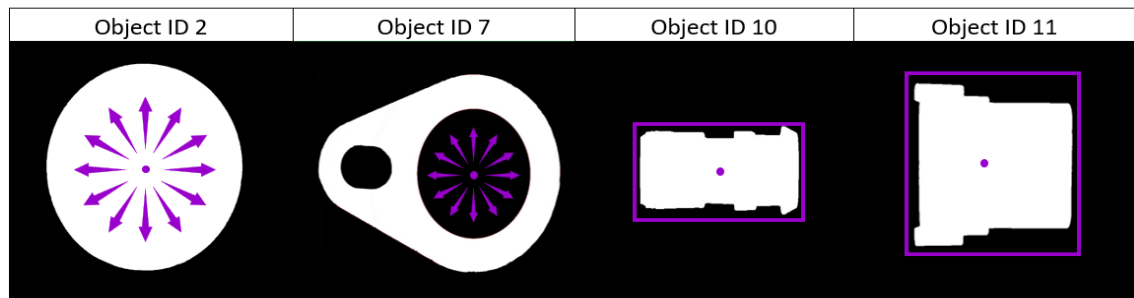


Figure 45. *Binary images of the objects*

To ensure high accuracy of the vision system, both the Photoneo structured light scanner and Basler area scan camera were calibrated before recording the data. The objects (IDs 2, 7, 10 and 11) were selected because the shapes can be accurately recognised by the vision application, and they are easy to grasp with the grippers available. These properties help to limit the errors caused from an unsuitable gripper because of too complex object geometry. The error sources for the performance testing were evaluated and approximated in order to have a baseline for results analysis. The identified error sources for the evaluation are listed in table 1. These error sources were acquired from manufacturer manuals and approximated through testing whenever applicable. This evaluation results in an expected accuracy of approximately $\pm 0.485 \text{ mm}$, with the Photoneo S-model and $\pm 0.590 \text{ mm}$ with the Photoneo M- model. This estimation does not include

the non-measurable error sources, originating from the localization algorithms and gripper.

Table 1. *Accuracy evaluation error sources*

Error source	Description	Values
Calibration	Standard deviation of the measurement	± 0.280 mm (S- model) ± 0.385 mm (M- model)
Vision application	Deviation measured from imaging the same object multiple times	± 0.05 mm
Robot manipulator	Repeatability of the robot manipulator	± 0.1 mm
Robot gripper	Deviation of the TCP. Measured by rotating the gripper 360°	± 0.1 mm
Object drawn towards gripper	Object grasped moves due to the vacuum or the magnet	Not measurable
Localization algorithms	Misaligned model matching due to localization parameters.	Not measurable

The errors from localization algorithms are expected to be minor, based on the visual inspection of the point clouds and the alignment of the reference CAD- models. The effects of the vacuum- and magnetic grippers were visibly noticeable, where grasping the objects caused the object to dislocate. This phenomenon was noticeable with both grippers, with all the objects being tested. The displacement was most noticeable with object ID 11, which was not parallel with the bottom of the bin. The effects of the errors originating from the grippers are illustrated in figure 46 below.

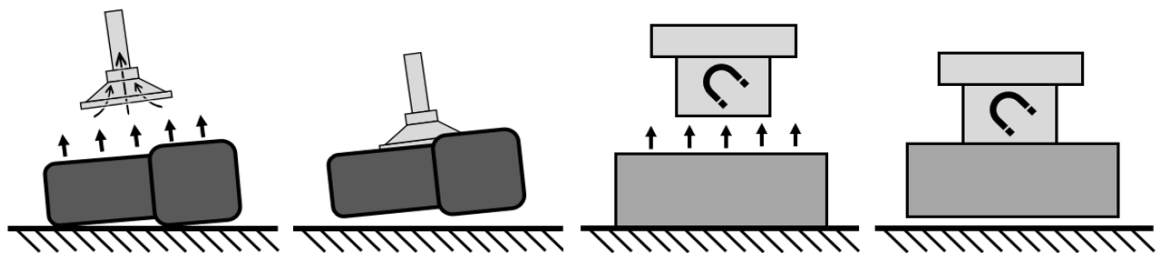


Figure 46. *Effects of the robot gripper to the grasp accuracy*

After repeating the bin picking task multiple times, the results clearly show how the vacuum gripper displaces the object as it is being grasped. When the object is eventually grasped in all orientations, the end results show a systematic error. The results of grasping object ID 11 with a vacuum gripper in both random orientations and a set orientation

are presented in figure 47. The graphs present the displacement of the objects both in X- and Y- axis over the whole sample size, where each measurement is represented as one data point. The results of a case where the objects were in random orientations is represented by blue markers and the results of a case where the objects all had same orientation are represented with red. The same effect applied with the magnetic gripper, where grasping an object slightly moved the object when it was grasped. The magnitude of these errors can only be approximated, as there is not a direct method of measuring them. The results of this error source evaluation present the importance of proper gripper design and just how much it affects the end result. In this case, it can be argued that the main error source of this application originates from the grippers.

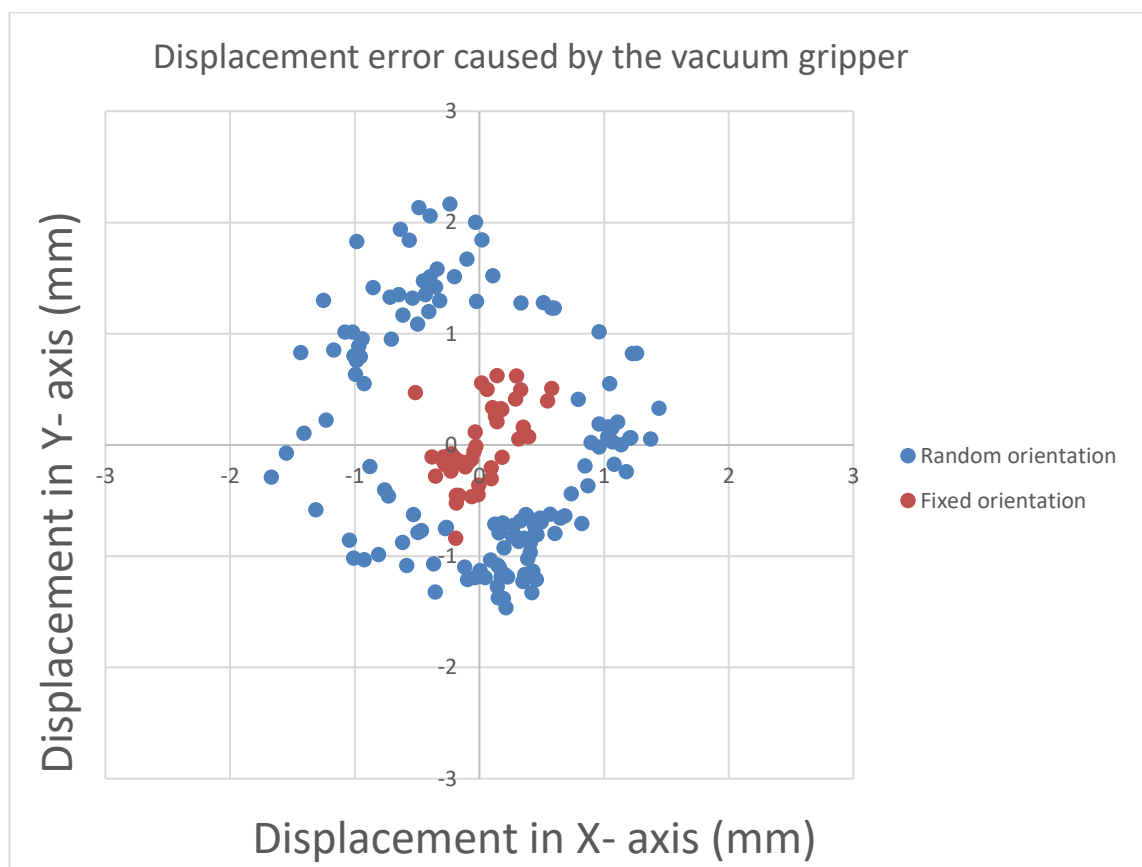


Figure 47. *Results of random and fixed orientation grasps*

After the error sources of the evaluation were identified, the testing between different objects and two camera models was completed. The results are presented as the displacement of the objects in both in X- and Y- axis over the whole sample size. The reference point [0,0] represents the average location of all localizations of the data set. The displacement of a single data point is presented in millimetres, compared to this reference point. The results from object ID 2 are presented in figure 48, where the object was grasped by using a magnetic gripper. The results clearly show a systematic error in the

results, where the magnetic gripper has affected the grasping of the object. With a symmetrical shape, such as a cylinder, the object is eventually detected in rotations between $0 \dots 360^\circ$. This makes the systematic error in one direction appear as errors in all directions. The results between the two models are similar, with the S- model having slightly better results. The Mean Absolute Error (MAE) of the S- model was 0.76 mm , with M- model having 0.90 mm . The maximum deviation between the samples was also smaller with the S- model, having a 1.96 mm compared to the 2.23 mm of the M- model. With a cylindrical, symmetric object the orientation of the grasp was not measured.

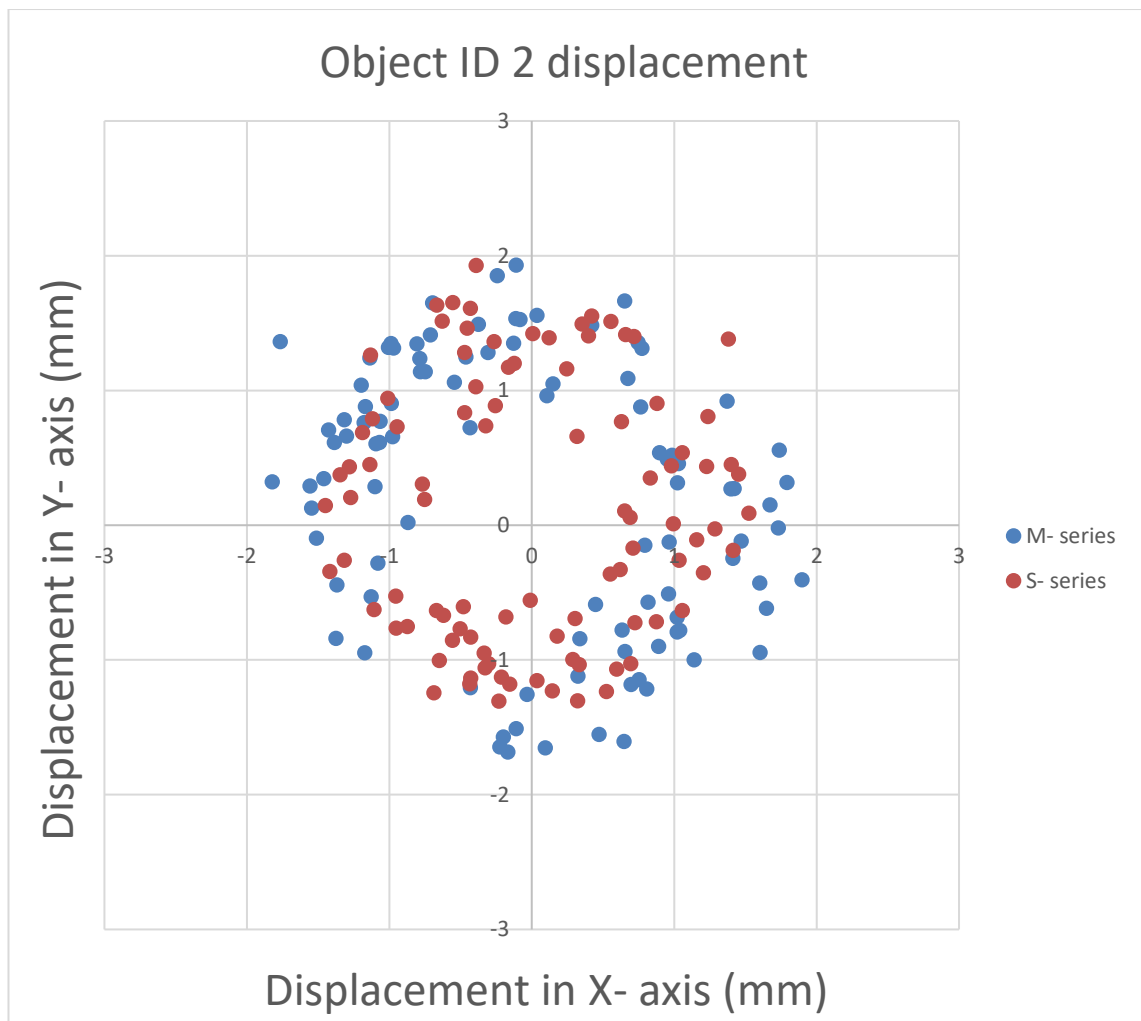


Figure 48. Displacement of a cylindrical metal blank

The results from object ID 7 are presented in figure 49, where the object was grasped by using a magnetic gripper. The results also show the systematic error of the magnetic gripper. The results between the two models are mixed, where the M- model had better MAE of 0.50 mm over the 0.66 mm of the S- model. This can be explained by the effects of the magnetic gripper and grasping the object from different orientations. The results show that the S- model had a broader set of samples, while the M- model had more grasps from smaller set of orientations. This hypothesis is also supported by the rotational accuracy of the grasps and the maximum deviation between the data set. The S- model had a better rotational accuracy out of the two having an average accuracy of 0.35° compared to the 0.47° of the M- model. The maximum deviation between the samples was also smaller with the S- model with a result of 1.54 mm compared to the 1.76 mm of the M- model.

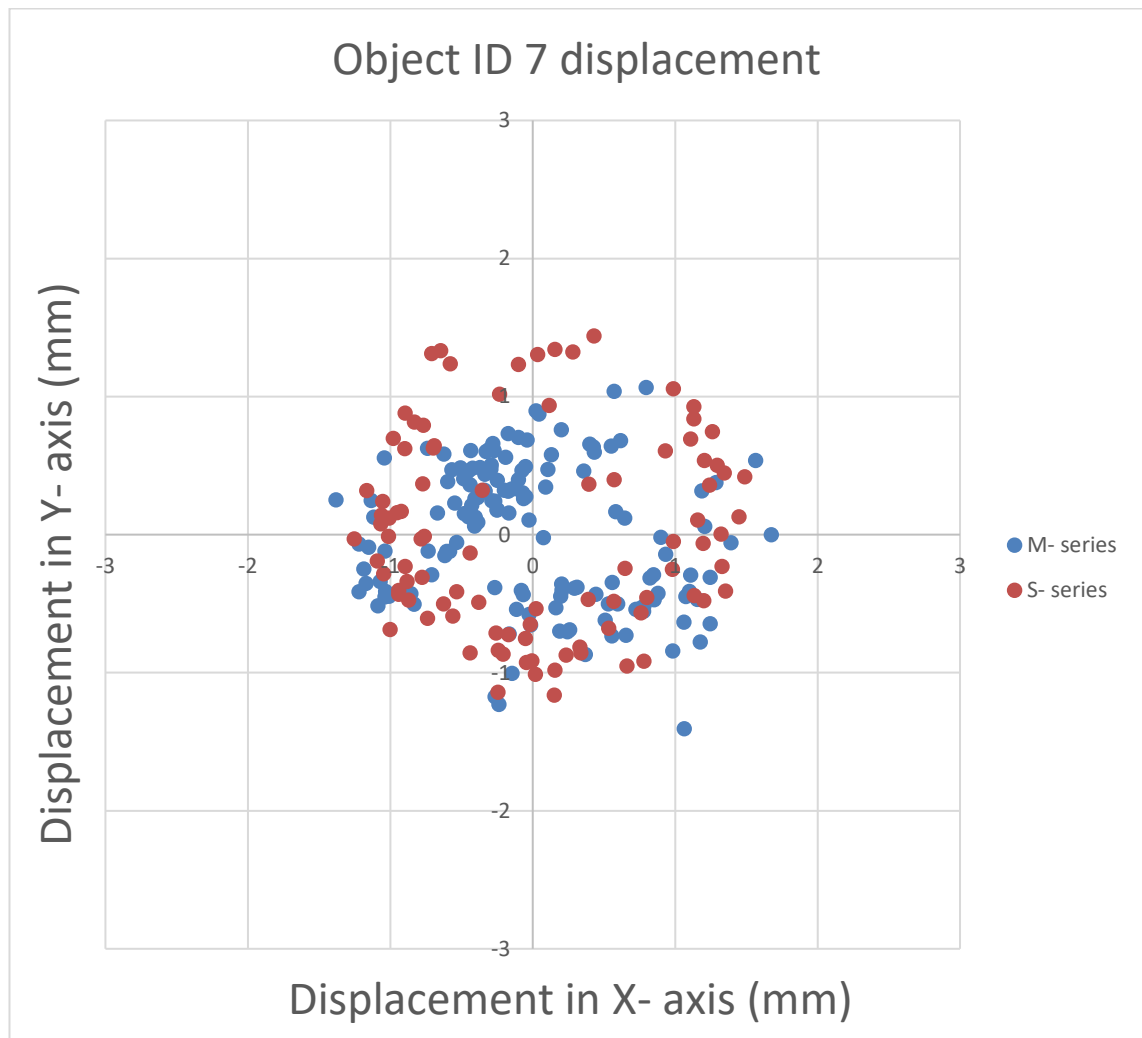


Figure 49. *Displacement of semi-finished product*

The results from object ID 11 are presented in figure 50, where the object was grasped by using a vacuum gripper. The results of vacuum gripper shares similar results as the magnetic gripper, where the grasping of the object dislocates the object from the initial grasping pose. This is again clearly visible in the results, where the localizations have a systematic error in all directions. The differences between the S- and M- models are more apparent however, with the S- model having visibly better results. These results can be explained by the S- model being able to better align the models due to the higher resolution of the camera. The MAE of the S- model was 0.55 mm compared to the 0.76 mm of the M- model. The maximum deviation between the data set is also clearly better with the S- model. The S- model had a maximum deviation of 1.42 mm , compared to the 2.18 mm of the M- model. With a non-symmetrical object, the orientation of the localized object was also measured. The orientational accuracy of the two units was almost identical, with an average rotational displacement of 0.21° for the S- model and 0.22° for the M- model.

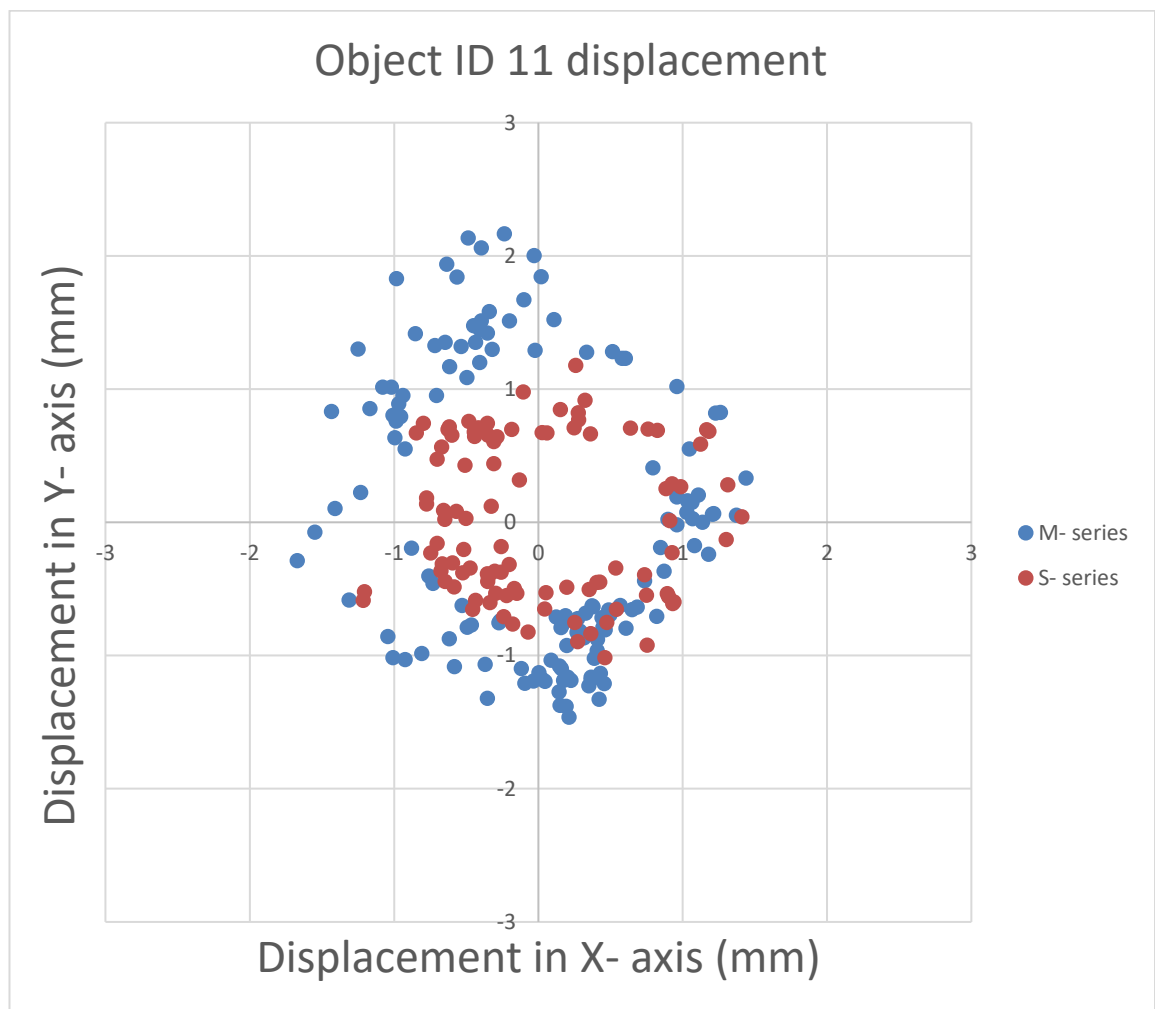


Figure 50. Displacement of large connector component

The results from object ID 10 are presented in figure 51, where the object was grasped by using a vacuum gripper. The results of the small connector component shared the similar effects of the vacuum gripper. This effect was noticeably smaller however, which can be explained by the orientation of the objects. By comparing the two objects, the object ID 10 was more parallel to the bottom of the bin than object ID 11. The differences between the S- and M- models are very similar, with the S- model having only marginally better results. These results can be explained by the object having more feature points than object ID 11, which enabled the M- model to accurately localize the object regardless of the small size. With more feature points to align the models, the M- model scanner performed better than with object ID 11. The MAE of S- model was 0.53 mm compared to the 0.65 mm of the M- model. The maximum deviation between the data set is also better with the S- model. The S- model had a maximum deviation of 1.41 mm , compared to the 1.72 mm of the M- model. With a non-symmetrical object, the orientation of the localized object was also measured. The results of the evaluation were an average rotational displacement of 0.21° for both scanner models.

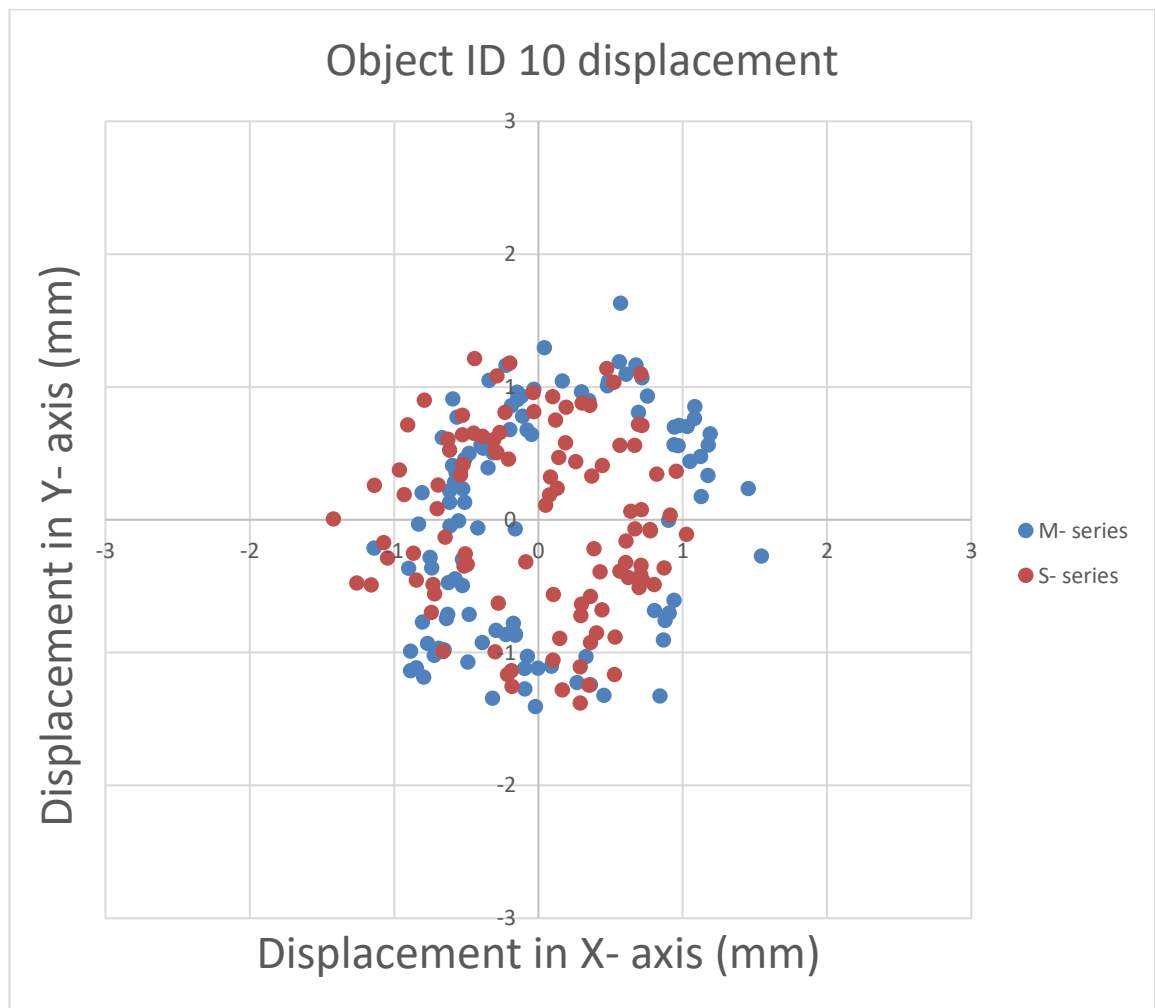


Figure 51. Displacement of small connector component

6. RESULTS AND DISCUSSION

This chapter presents and summarises the research work results of this thesis. Chapter 6.1 summarises the performance of technologies with reflective materials, and Chapter 6.2 the results from complex object geometry evaluation. Chapter 6.3 focuses on presenting a summary of the results from ambient light tolerance, and Chapter 6.4 presents the results from bin picking solution performance testing.

6.1 Depth camera performance with reflective materials

The depth camera performance with reflective materials was tested by comparing results between shiny metal objects and 3D- printed matte counterparts. Common results between all technologies were planar surfaces, which were easily detectable even with highly reflective surfaces. Common result between the ToF and stereo vision was the loss of edge fidelity with sharp corners that was not present with round objects. The main issue with reflective materials was found when it was combined with convex surfaces. These reflective surfaces caused problems with all the technologies. The combined results from comparison with reflective materials are presented below in table 2.

Table 2. *Depth camera performance with reflective materials*

Technology	Performance comment
Time-of-Flight	ToF performed well with reflective materials of planar surfaces. The ability to interpolate for missing data also enabled the camera to perform well with different object types. The issues of ToF were most notable with a combination of convex, reflective surfaces. Compared to other technologies, the ToF performed poorly with these types of surfaces.
Stereo Vision	Stereo vision performed similarly to ToF with reflective materials of planar surfaces but had less noise on the surface texture. The issues of stereo vision were also with convex, reflective surfaces which were based on the correspondence problem. This caused some lost depth data, but the effects were smaller than ToF.
Structured Light	Structured light performs the best with reflective materials and even objects under the recommended minimum object size were captured accurately. Planar surfaces cause little to no problems, but convex, shiny surfaces still cause some loss of the depth data.

Considering the reliability of the results, the problems encountered with stereo vision (Yang and Waslander, 2022, pp. 1–2), structured light (Li et al., 2022, pp. 1–3) and ToF-technologies (Sasaya et al., 2021, pp. 329–332) were in line with recent studies. It must be noted however that the stereo vision and ToF- cameras of this thesis were not in their optimal environments with optimal object sizes. A more fitting environment for these technologies would have been long range measurements, with large object sizes. It should also be noted that the sample size of depth cameras provided for the thesis was small. This limited the data analysis, with only one depth camera per technology. The answer to the first research question is answered in chapter 6.2, as both chapters 6.1 and 6.2 contribute to this research question.

6.2 Depth camera performance with different object properties

The depth camera performance with different object properties was evaluated by imaging different properties and combination of properties. This research had different results between the technologies and properties were identified where some technologies excelled at, while others performed poorly. The combined results from object detection testing are presented below in table 3.

Table 3. *Depth camera performance with different object properties*

Technology	Performance comments
Time-of-Flight	ToF performed well with planar surfaces and small surface details. ToF had issues with porous surfaces, which appeared as concave instead of planar. Both stereo vision and ToF had loss of data with simultaneous detection of multiple different textures, but this was more noticeable with ToF.
Stereo Vision	Stereo vision performed well with both planar and convex surfaces. The problem of stereo vision was noticed with small surface details of the objects. This phenomenon was most apparent with connector components, where the details of holes and surface levels were lost.
Structured Light	Structured light performed noticeably the best in all cases. Structured light was able to accurately detect all of the objects, even those under the recommended minimum size of the camera specifications. Structured light also performed the best when multiple different object types were imaged simultaneously.

The potential error sources of this evaluation are common with the error sources discussed in chapter 6.1, with more focus on the resolution difference of the depth cameras. This was more noticeable with the small objects, where features such as small holes or

surface details were affected more by the resolution difference. Potential error sources of this evaluation also include the camera parameters of the ToF- camera, which were manually tuned for the different objects. These parameters were noticeably harder to get correct, compared to the predefined profiles of the structured light or the autotune function of stereo vision.

The results of chapters 6.1 and 6.2 both contributed to the first research question “What are the object properties, or a combination of properties that enable or limit the use of a specific 3D- depth camera technology?” The research concluded that structured light performed the best with no limitations from a single property. The technology performed well with all surface types, which was also demonstrated in pilot scale demonstrations. The results with shiny metal cylinders however showed that a combination of convex, highly reflective surfaces cause loss of depth information. If there are high accuracy requirements, this can be a limiting factor. The performance of stereo vision shared similar limitations, originating from reflective cylindrical shapes and a combination of multiple convex surfaces. Recent studies in stereo vision (Yang and Waslander, 2022, p. 2; Zhang et al., 2021, pp. 7–9) support these results, which describe the influence of angles and reflections to the measurement accuracy. Common limitation with both stereo vision and ToF was noticed when there was a large number of different surface textures imaged at once. The cameras were unable to fully image the scene due to exposure time limitation. The performance of ToF was also limited by porous surfaces, which were captured as concave instead of planar. Porous surfaces are not highly researched topic with ToF- cameras, but the results are similar to ToF- comparison researched by (Laukkanen, 2015, pp. 39–42), where foam also had depth deviation.

The wide range of different surface types and object properties make the results of chapters 6.1 and 6.2 applicable with large quantity of products. This enables the results to be used as a tool in camera technology selection or feasibility study with similar object types. The results of this comparison could also be expanded upon by researching transparent and translucent objects, which were excluded from this thesis.

6.3 Depth camera performance with ambient light

The depth camera performance with ambient light was evaluated in varying indoor conditions and with different bin configurations. The results from ambient light interference were common between all of the tested technologies. It was hypothesized, that ToF and stereo vision perform well with external ambient light and that high intensity ambient light has a negative effect with the structured light scanner. None of the technologies however had visible errors in the resulting point clouds. Summary of the results is presented below

in table 4. The results of stereo vision and ToF were in line with the expectations, but structured light performed better than initially hypothesized.

Table 4. *Summary of ambient light testing*

Technology	Performance with ambient light
Time-of-Flight	ToF performed well in all ambient light conditions, and the hypothesis for this technology was confirmed. ToF performed the best with no data loss in any of the testing scenarios.
Stereo Vision	Stereo vision performed well in all ambient light conditions, and the hypothesis for this technology was confirmed. Stereo vision performed well in all cases, with only a minor $\pm 1\%$ data loss between the best and worst cases.
Structured Light	Structured light performed well in all ambient light conditions, and the hypothesis for this technology was disproven. The structured light scanner unit was equipped with a high power laser unit, which was able to illuminate the scene in both dark room and overpower the external LED illuminator.

The performance with different bin configurations had varying results between the tested technologies and the results are presented below in table 5. It was hypothesized that all camera technologies perform well with the non-reflective materials and that the highly reflective surfaces are problematic. The results were in line with the expectations and the non-reflective materials had similar results between all the technologies. There was no loss of data, and the point clouds were complete in the whole volume of the bin. The exception to the expected results was with the stereo vision camera which struggled with the empty bin. The shiny metal sheet caused problems with all the technologies, where the reflections from the metal caused visible errors in the point clouds with all the cameras.

Table 5. *Summary from different bin configurations*

Technology	Performance with ambient light
Time-of-Flight	The ToF- camera performed the best in this comparison, and the interpolation feature was able to fill in any small holes of the point clouds.
Stereo Vision	Stereo vision performed the poorest with reflective surfaces, being the only technology having problems with the empty bin. The reflective bin caused the correspondence problem, causing small areas of the bin to be not detected.
Structured Light	Structured light performed comparatively with ToF, with only slightly higher loss of data with the reflective surfaces.

There were no notable error sources in this comparison as the only variable, ambient light did not have a notable effect in any of the testing scenarios. The results of the stereo vision camera could potentially be better, if the comparison is remade and tested with different camera angles. This however has an effect on the camera FoV, which can make parts of the bin not visible. The wide range of different ambient light conditions and bin configurations make the results applicable in many industrial environments. The results of this research topic and pilot scale demonstrations both contribute to the second research question “*How different type of 3D- depth camera technologies perform in an industrial bin picking environment?*” The research had a clear outcome, where all of the tested technologies based on structured light, ToF and stereo vision perform well in different conditions. The environment itself is not a limiting factor for any of the technologies, but what is important, is the location and the orientation of the depth camera.

Reflecting on the results of the 3D- depth camera comparison, evaluation of different technologies against each other is a challenging task. Comparable results are hard to achieve when the different technologies have different resolutions, working distances and even integrated data enhancing tools such as interpolation. The testing environment also has an effect to the comparison. Different depth cameras can have different optimal viewing angles, which can also vary depending on the application. When reviewing the performance of different technologies, the evaluation should be completed in an environment relevant to the target application.

6.4 Performance of Photoneo bin picking studio

Photoneo bin picking solution performance was evaluated with a combination of accuracy testing and pilot scale demonstrations. The accuracy was evaluated with different objects between two different scanner models. The summary of accuracy testing is presented in tables 6 and 7. The results from different object types show that as the objects have more complex feature points, the accuracy of the model alignment increases. This is presented on the results with both Photoneo S- and M- model scanners, where the more complex object ID 7 can be localized more accurately than the metal blank (Object ID 2). The connector components also share these same results, where the more complex object ID 10 was localized more accurately than object ID 11.

Table 6. *Accuracy evaluation results (Photoneo Phoxi scanner S)*

Object	Rotational error	Mean absolute error	Maximum deviation
Object ID 2	-	± 0.76 mm	1.96 mm
Object ID 7	$\pm 0.35^\circ$	± 0.66 mm	1.54 mm
Object ID 10	$\pm 0.21^\circ$	± 0.53 mm	1.41 mm
Object ID 11	$\pm 0.21^\circ$	± 0.55 mm	1.42 mm

Table 7. *Accuracy evaluation results (Photoneo Phoxi scanner M)*

Object	Rotational error	Mean absolute error	Maximum deviation
Object ID 2	-	± 0.90 mm	2.23 mm
Object ID 7	$\pm 0.47^\circ$	± 0.50 mm	1.76 mm
Object ID 10	$\pm 0.21^\circ$	± 0.65 mm	1.72 mm
Object ID 11	$\pm 0.22^\circ$	± 0.76 mm	2.18 mm

There were several notable error sources in the bin picking solution performance evaluation. The measurable error margins from calibration, robot manipulator and vision application sums up to ± 0.485 mm with the Photoneo S- model and ± 0.590 mm with the Photoneo M- model. The MAE of the results was close to this estimation, but because of the errors caused by the gripper, the maximum deviation between all the results was far higher. The results of a study such as this are rarely public, so there is very little reference material available to compare the results of this research with. The results of this evaluation can be used as a tool to consider if the accuracy of a commercial bin picking system is enough on its own. By considering a machine tending task with very high accuracy requirements, the accuracy of ± 0.5 mm is not enough, even if the errors from the gripper are excluded. Task such as this requires either a mechanical design with better alignment of the object, or a compensation for the displacement with an additional camera. Tasks with more loose requirements, such as component labelling however could be completed without an external vision system.

The pilot scale demonstrations also produced valuable data regarding the performance of a modern bin picking solution. The solution was robust not only to the surface reflectivity but performed with both simple and complex geometric shapes. The solution was also able to function with noise originating from oil and metal flakes. While both of the demonstrations were successful, they also presented the challenges of highly reflective objects. If a surface of an object cannot be fully captured, these blind spots in the point cloud can cause collisions between the objects and the gripper. The results of these pilot scale demonstrations can be used in the consideration of a full scale deployment of a bin picking task. In a full scale deployment of machine tending, this solution could be located in a way that the metal blanks can be fed to the bin directly from the band saw.

The palletization of the objects could also be done in a way that the industrial robots have direct access to the pallets. This approach would fully automate the process and human interaction with the metal blanks would no longer be needed.

The results of accuracy evaluation and pilot scale demonstrations both contribute to the third research question “*How well does a commercial bin picking solution perform, and are the benefits justifiable for the increased initial cost?* “. The outcome of the research showed that a modern bin picking solution is capable of sub millimetre accuracy, but the accuracy is highly dependent on the gripper of the robot. The computation time of the solution is also fast, ranging between few seconds depending on the complexity of the localization. Considering the time it takes for the robot to complete the computed trajectory, this computation time was not a limiting factor. The benefits of a commercial bin picking solution also include fast commissioning and modification of the bin picking task. Relocating the system also takes a short time, where only recalibration of the depth camera was required before continuing the use of the system. This means that a single robotic cell could be relocated with very little downtime before it can continue working with a different task. Considering how long it takes to design and commission a standalone system from the ground up, a commercial OOTB- solution can very quickly compensate for the initial cost of the system. The downsides of commercial solutions are that they do have their limits, either with the localization algorithms or the maximum number of localized object types. These are something that cannot be changed, and in some specific cases these solutions might not be able to match the requirements of the system.

To expand upon the results of this thesis, future studies could continue with larger camera selection, gripper selection or more environmental variables, such as multi view camera interference. Another, more expensive option would be to explore the performance of different bin picking solutions from other manufacturers and evaluate the performance of the solutions. The resource intensity & impact graph of these proposals is listed in figure 52.

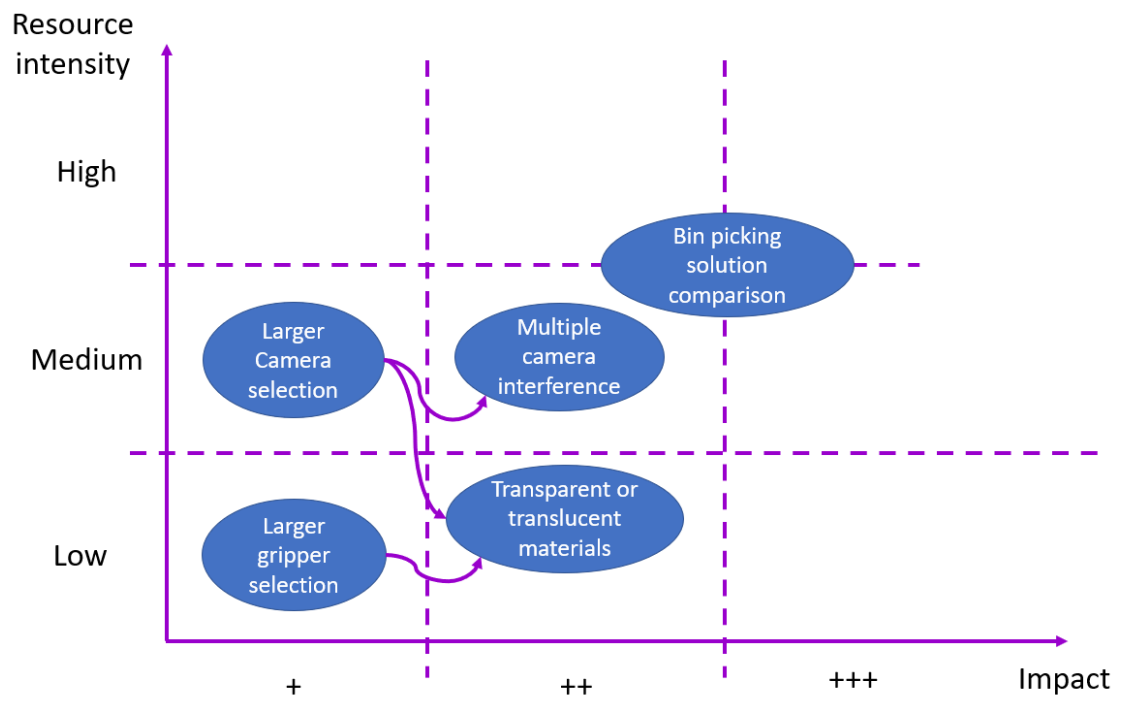


Figure 52. *Resource intensity & impact graph*

7. CONCLUSIONS

This thesis researched the challenges of automated bin picking and performance of a modern, commercial bin picking solution. This research aimed to identify the strengths and weaknesses of different 3D- depth camera technologies and evaluate the performance of each technology. The research objectives were achieved through tangible results between the technology comparison and pilot-scale demonstrations of a modern bin picking solution. The results of this thesis present the different advantages and disadvantages of different 3D- depth camera technologies and based on the perspective of a bin picking application; some are more significant than others. Summary of the strengths and weaknesses of stereo vision, structured light and ToF from the perspective of a bin picking system are presented below in table 8.

Table 8. *Summary of depth camera strengths and weaknesses*

Technology	Strengths	Weaknesses
Time-of-Flight	Capable of real time imaging, which enables bin picking with short cycle time requirements. Capable of performing bin picking, but most suited for longer range measurement with large objects. Best performing technology with ambient light.	Comparatively lower resolution to stereo vision and structured light. Convex, reflective surfaces are problematic due to Multi Path Error.
Stereo vision	Capable of real time imaging, which enables bin picking with short cycle time requirements. Wide range of different camera resolutions and dynamic scene imaging enables different bin picking tasks from short to long range. RGB- camera enables bin picking based on colour of the objects.	Reflective surfaces, especially from the bin background can be problematic and cause the correspondence problem.
Structured light	High resolution enables bin picking of even small objects and the best suited technology for close range measurements. Versatile to many different surface textures and the best performing technology with reflective surfaces.	Lower range compared to Stereo vision and ToF. Scanning the scene takes some time, which requires a stationary scene.

The approach for this thesis and the research problems were influenced by the common questions of the companies involved with the thesis work. The expected results from this research work were both practical and theoretical, mainly focusing on modern depth camera technologies. The results matched the expectations, where the performance of a modern bin picking solution even exceeded them. The research methods of this study fulfilled the goals set for the research work. The results presented the capabilities of modern depth cameras and how successful bin picking application depends on much more than simply an accurate vision system.

The results of this thesis provide valuable information from the performance of different depth camera technologies, which are not usually publicly available from similar research. The results can be used as a tool to help selecting the best depth camera technology for a specific task or used as a reference in a feasibility study with similar object types. As a conclusion based on the results of this thesis, structured light was the best performing technology for bin picking of small to medium sized objects. Considering the requirements of a successful bin picking task, structured light had clear advantages compared to stereo vision and ToF. Structured light had the versatility to perform with many surface textures and a broad range of reflective surfaces. Based on the results of the pilot-scale studies, the longer computation time also outweighed the perks of real time performance. This is because the most important part of bin picking, the pose estimation is highly dependent on the resolution of the point cloud. Accurate perception of the scene is also important from trajectory planning point of view, especially in occluded environments.

REFERENCES

- Agresti, G., Zanuttigh, P., 2019. Combination of Spatially-Modulated ToF and Structured Light for MPI-Free Depth Estimation, in: Leal-Taixé, L., Roth, S. (Eds.), *Computer Vision – ECCV 2018 Workshops*, ISBN: 978-3-030-11009-3. Springer International Publishing, pp. 355–371.
- Ahmed, E., Saint, A., Shabayek, A.E.R., Cherenkova, K., Aouada, D., 2018. Deep Learning Advances on Different 3D Data Representations: A Survey. ResearchGate. pp. 3-4.
- Anandan, T.M., 2016. Robotic Bin Picking – The Holy Grail in Sight. Association for advancing automation. <https://www.automate.org/industry-insights/robotic-bin-picking-the-holy-grail-in-sight>.
- Basler, 2022a. Product manual, Blaze-101. <https://docs.baslerweb.com/blaze-101>.
- Basler, 2022b. Product specification, Blaze-101. <https://www.baslerweb.com/en/products/cameras/3d-cameras/basler-blaze/blaze-faq/>.
- Basler, 2018. Product manual, Area Scan Camera ace acA1300-60gm. <https://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca1300-60gm/>.
- Borboni, A., Reddy, K.V.V., Elamvazuthi, I., AL-Quraishi, M.S., Natarajan, E., Azhar Ali, S.S., 2023. The Expanding Role of Artificial Intelligence in Collaborative Robots for Industrial Applications: A Systematic Review of Recent Works. MDPI. pp. 18-19. <https://doi.org/10.3390/machines11010111>
- Boschetti, G., Sinico, T., Trevisani, A., 2023. Improving Robotic Bin-Picking Performances through Human–Robot Collaboration. MDPI. pp. 1-2. <https://doi.org/10.3390/app13095429>
- Bronwyn, B., Dawson, P., Devine, K., 2005. Colorado State University - Designing and Conducting Case Studies. <https://writing.colostate.edu>.
- Buchholz, D., 2016. Bin-Picking - New approaches for a classical problem. Springer. pp. 10-16. <https://doi.org/10.1007/978-3-319-26500-1>
- Carroll, J., 2021. 3D Vision Technology Advances to Keep Pace With Bin Picking Challenges. Association for advancing automation. <https://www.automate.org/industry-insights/3d-vision-technology-advances-to-keep-pace-with-bin-picking-challenges>.
- Chen, Y.-K., Sun, G.-J., Lin, H.-Y., Chen, S.-L., 2018. Random Bin Picking with Multi-view Image Acquisition and CAD-Based Pose Estimation. IEEE Xplore. pp. 1-2. <https://doi.org/10.1109/SMC.2018.00381>
- Cognex, 2018. Cognex whitepaper - Introduction to machine vision: A guide to automating process & quality improvements. Cognex. pp. 3-10.
- Dal Mutto, C., Zanuttigh, P., Cortelazzo, G.M., 2013. Time-of-Flight Cameras and Microsoft Kinect™. Springer US. pp. 9-31. <https://doi.org/10.1007/978-1-4614-3807-6>
- DAQRI, 2018. Depth cameras for mobile AR: From Iphones to wearables and beyond. <https://medium.com/@DAQRI/depth-cameras-for-mobile-ar-from-iphones-to-wearables-and-beyond-ea29758ec280>.
- Dumic, E., Battisti, F., Carli, M., da Silva Cruz, L.A., 2020. Point Cloud Visualization Methods: a Study on Subjective Preferences. IEEE Xplore. pp. 595-596. <https://doi.org/10.23919/Eu-sipco47968.2020.9287504>

- Feng, W., Cheng, X., Sun, J., Xiong, Z., Zhai, Z., 2023. Specular highlight removal and depth estimation based on polarization characteristics of light field. *Optics Communications*. <https://doi.org/10.1016/j.optcom.2023.129467>
- Fu, B., Li, F., Zhang, T., Jiang, J., Li, Q., Tao, Q., Niu, Y., 2019. Single-Shot Colored Speckle Pattern for High Accuracy Depth Sensing. *IEEE Xplore*. p. 1. <https://doi.org/10.1109/JSEN.2019.2916479>
- Geng, J., 2011. Structured-light 3D surface imaging: a tutorial. Optica publishing group. pp. 130-146. <https://doi.org/10.1364/AOP.3.000128>
- Giancola, S., Valenti, M., Sala, R., 2018. A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies, ISBN: 978-3-319-91761-0. Springer.
- Grzegorzec, M., Theobalt, C., Koch, R., Kolb, A. (Eds.), 2013. Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications. Springer. pp. 3-12. <https://doi.org/10.1007/978-3-642-44964-2>
- Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2020. Deep Learning for 3D Point Clouds: A Survey. *IEEE Xplore*. p. 1. <https://doi.org/10.1109/TPAMI.2020.3005434>
- Gupta, M., Yin, Q., Nayar, S.K., 2013. Structured Light in Sunlight, in: 2013 IEEE International Conference on Computer Vision. Presented at the 2013 IEEE International Conference on Computer Vision (ICCV), IEEE, Sydney, Australia, p. 1. <https://doi.org/10.1109/ICCV.2013.73>
- He, L., Chen, C., Zhang, T., Zhu, H., Wan, S., 2018. Wearable Depth Camera: Monocular Depth Estimation via Sparse Optimization Under Weak Supervision. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2018.2857703>
- Huang, W., Kovacevic, R., 2012. Development of a real-time laser-based machine vision system to monitor and control welding processes. Springer. pp. 1-4. <https://doi.org/10.1007/s00170-012-3902-0>
- Jafari Malekabadi, A., Khojastehpour, M., Emadi, B., 2019. Disparity map computation of tree using stereo vision system and effects of canopy shapes and foliage density. *Science direct*. p. 630. <https://doi.org/10.1016/j.compag.2018.12.022>
- Jang, M., Yoon, H., Lee, Seongmin, Kang, J., Lee, Sanghoon, 2022. A Comparison and Evaluation of Stereo Matching on Active Stereo Images. *MDPI*. pp. 1-4. <https://doi.org/10.3390/s22093332>
- Javaid, M., Haleem, A., Singh, R.P., Rab, S., Suman, R., 2022. Exploring impact and features of machine vision for progressive industry 4.0 culture. *Sensors International*. pp. 1-5. <https://doi.org/10.1016/j.sintl.2021.100132>
- Kadambi, A., Bhandari, A., Raskar, R., 2014. 3D Depth Cameras in Vision: Benefits and Limitations of the Hardware: With an Emphasis on the First- and Second-Generation Kinect Models. Springer. pp. 3-26, *Advances in Computer Vision and Pattern Recognition*. https://doi.org/10.1007/978-3-319-08651-4_1
- Kim, S.-Y., Kim, M., Ho, Y.-S., 2013. Depth Image Filter for Mixed and Noisy Pixel Removal in RGB-D Camera Systems. *IEEE Transactions*. p. 681. <https://doi.org/10.1109/TCE.2013.6626256>
- Kleppe, A., Bjørkedal, A., Larsen, K., Egeland, O., 2017. Automated Assembly Using 3D and 2D Cameras. *MDPI*. p. 1. <https://doi.org/10.3390/robotics6030014>
- Laukkanen, M., 2015. Performance Evaluation of Time-of-Flight Depth Cameras. 2015. pp. 39-42.

- Li, B., Xu, Z., Gao, F., Cao, Y., Dong, Q., 2022. 3D Reconstruction of High Reflective Welding Surface Based on Binocular Structured Light Stereo Vision. MDPI. p. 159. <https://doi.org/10.3390/machines10020159>
- Li, D., Liu, N., Guo, Y., Wang, X., Xu, J., 2019. 3D object recognition and pose estimation for random bin-picking using Partition Viewpoint Feature Histograms. Science direct. pp. 149-150. <https://doi.org/10.1016/j.patrec.2019.08.016>
- Lin, X., Wang, J., Lin, C., 2020. Research on 3D Reconstruction in Binocular Stereo Vision Based on Feature Point Matching Method, in: 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE). Presented at the 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE), p. 551. <https://doi.org/10.1109/ICISCAE51034.2020.9236889>
- Liu, M.-Y., Tuzel, O., Veeraraghavan, A., Taguchi, Y., Marks, T.K., Chellappa, R., 2012. Fast object localization and pose estimation in heavy clutter for robotic bin picking. The International Journal of Robotics Research. pp. 1-6. <https://doi.org/10.1177/0278364911436018>
- Lopez, M., Sergiyenko, O., Tyrsa, V., 2008. Machine Vision: Approaches and Limitations. Research Gate. pp. 400-410. <https://doi.org/10.5772/6156>
- Lü, C., Wang, X., Shen, Y., 2013. A stereo vision measurement system Based on OpenCV. IEEE Xplore. pp. 1-2. <https://doi.org/10.1109/CISP.2013.6745259>
- Lydon, B., 2016. Industry 4.0: Intelligent and flexible production. International Society of Automation. <https://www.isa.org/intech-home/2016/may-june/features/industry-4-0-intelligent-and-flexible-production>.
- Malik, A.A., Andersen, M., Bilberg, A., 2019. Advances in machine vision for flexible feeding of assembly parts. Science Direct. pp. 1228-1229. <https://doi.org/10.1016/j.promfg.2020.01.214>
- Martinez, C., Boca, R., Zhang, B., Chen, H., Nidamarthi, S., 2015. Automated bin picking system for randomly located industrial parts. IEEE Xplore. pp. 1-2. <https://doi.org/10.1109/TePRA.2015.7219656>
- McLeod, S., 2021. Feasibility studies for novel and complex projects: Principles synthesised through an integrative review. ScienceDirect. pp. 1-4. <https://doi.org/10.1016/j.plas.2021.100022>
- Milani, S., Calvagno, G., 2016. Correction and interpolation of depth maps from structured light infrared sensors - ScienceDirect. ScienceDirect. pp. 31-33. <https://doi.org/10.1016/j.image.2015.11.008>
- Mohammadikaji, M., 2020. Simulation-based Planning of Machine Vision Inspection Systems with an Application to Laser Triangulation. KIT Scientific Publishing. pp. 8-11. <https://doi.org/10.5445/KSP/1000099225>
- Moosmann, M., Spenrath, F., Kleeberger, K., Khalid, M.U., Mönnig, M., Rosport, J., Bormann, R., 2020. Increasing the Robustness of Random Bin Picking by Avoiding Grasps of Entangled Workpieces. ScienceDirect. pp. 1-2. <https://doi.org/10.1016/j.procir.2020.03.082>
- Muhammad Amir, Y., Thörnberg, B., 2017. High Precision Laser Scanning of Metallic Surfaces. International journal of optics. pp. 2-4. <https://doi.org/10.1155/2017/4134205>
- Nagel, D.N., 2021. SICK AG White paper - A comparison of working principles for 3D Time of Flight, Stereo and active stereo. p. 5.
- Nair, R., Fitzgibbon, A., Kondermann, D., Rother, C., 2015. Reflection Modeling for Passive Stereo. IEEE, p. 1. <https://doi.org/10.1109/ICCV.2015.264>

- Nilsson, F., Murhed, A., 2015. SICK AG whitepaper - Select the best technology for your vision application. SICK. p. 4.
- Ojer, M., Lin, X., Tammaro, A., Sanchez, J., 2022. PickingDK: A Framework for Industrial Bin-Picking Applications. MDPI. pp. 1-4. <https://doi.org/10.3390/app12189200>
- O’Riordan, A., Newe, T., Dooly, G., Toal, D., 2018. Stereo Vision Sensing: Review of existing systems. IEEE Xplore, pp. 178–179. <https://doi.org/10.1109/ICSensT.2018.8603605>
- Pérez, L., Rodríguez, Í., Rodríguez, N., Usamentiaga, R., García, D.F., 2016. Robot Guidance Using Machine Vision Techniques in Industrial Environments: A Comparative Review. MDPI. pp. 3-10. <https://doi.org/10.3390/s16030335>
- Photoneo, 2021. Quick start guide, PhoXi Scanner M. p. 2.
- Photoneo, 2020. Quick start guide, Bin Picking Studio 1.5. p. 8.
- Photoneo, 2018. Instruction manual, Bin Picking Solution. 2018. pp. 1-11.
- Pochyly, A., Kubela, T., Singule, V., Cihak, P., 2012. 3D vision systems for industrial bin-picking applications, in: IEEE Xplore, ISBN: 978-80-01-04987-7. Proceedings of 15th International Conference MECHATRONIKA, p. 1.
- Qualitas, Q., 2011. What is 3D Machine Vision? Qualitas Technologies. <https://qualitastech.com/image-acquisition/3d-machine-vision/>.
- Rebbouh, A., 2022. Bin-Picking handbook. VISIO NERF. <https://content.visionerf.com/from-2d-to-3d-the-operating-principles-of-industrial-vision-systems-0?hsCtaTracking=59e2a72a-d663-450a-af5e-51d259af5ed1%7C0aa52600-51c0-4eb1-a0a5-d30cb78dc3a7>, 3–4.
- Salvi, J., Pagès, J., Batlle, J., 2003. Pattern codification strategies in structured light systems. ScienceDirect. pp. 1-2. <https://doi.org/10.1016/j.patcog.2003.10.002>
- Sangel, 2018. Product manual, LED illuminator. p. 4. https://www.ghv.de/files/led-industriebeleuchtung/sangel/downloads/ghv_led_aufbauleuchte_sl.pdf.
- Sansoni, G., Bellandi, P., Leoni, F., Docchio, F., 2014. Optoranger: A 3D pattern matching method for bin picking applications. ScienceDirect. pp. 1-2. <https://doi.org/10.1016/j.optlas-eng.2013.07.014>
- Sasaya, T., Watanabe, W., Ono, T., 2021. Depth Correction For Time-Of-Flight Camera Using Depth Distortion Dependency On Pulse Width Of Irradiated Light. IEEE, pp. 329–332. <https://doi.org/10.1109/ICIP42928.2021.9506225>
- Schraml, S., Nabil Belbachir, A., Milosevic, N., Schön, P., 2010. Dynamic stereo vision system for real-time tracking. IEEE Xplore. p. 3. <https://doi.org/10.1109/ISCAS.2010.5537289>
- SICK, 2022a. Operating manual, PLB engine 6.2. pp. 12-16.
- SICK, 2022b. Product manual, Visionary-S. <https://www.sick.com/my/en/machine-vision/3d-machine-vision/visionary-s/c/g507251>.
- Soave, E., D’Elia, G., Mucchi, E., 2020. A laser triangulation sensor for vibrational structural analysis and diagnostics. Research Gate. p. 2. <https://doi.org/10.1177/0020294019877484>
- Song, Z., Chung, R., Zhang, X.-T., 2013. An Accurate and Robust Strip-Edge-Based Structured Light Means for Shiny Surface Micromasurement in 3-D. IEEE Transactions on Industrial Electronics. p. 1. <https://doi.org/10.1109/TIE.2012.2188875>

Stoyanov, T., Mojtahedzadeh, R., Andreasson, H., Lilienthal, A.J., 2012. Comparative evaluation of range sensor accuracy for indoor mobile robotics and automated logistics applications. 2012 ScienceDirect. p. 2. <https://doi.org/10.1016/j.robot.2012.08.011>

Syrjänen, A., 2021. Experimental evaluation of depth cameras for pallet detection and pose estimation. 2021. pp. 1-10.

Tan, J., Lin, W., Chang, A.X., Savva, M., 2021. Mirror3D: Depth Refinement for Mirror Surfaces, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, pp. 1–2. <https://doi.org/10.1109/CVPR46437.2021.01573>

Tengli, M., 2020. Research Onion: A Systematic Approach to Designing Research Methodology. <https://www.aesanetwork.org/research-onion-a-systematic-approach-to-designing-research-methodology/>.

Tipary, B., Kovács, A., Erdős, F.G., 2021. Planning and optimization of robotic pick-and-place operations in highly constrained industrial environments. *Assembly Automation*. pp. 1-2. <https://doi.org/10.1108/AA-07-2020-0099>

Torres, P., Arents, J., Marques, H., Marques, P., 2022. Bin-Picking Solution for Randomly Placed Automotive Connectors Based on Machine Learning Techniques. *ResearchGate*. pp. 1-8. <https://doi.org/10.3390/electronics11030476>

Universal Robots, 2016. ur5_en.pdf. https://www.universal-robots.com/media/50588/ur5_en.pdf.

Wnuk, M., Pott, A., Xu, W., Lechler, A., Verl, A., 2017. Concept for a simulation-based approach towards automated handling of deformable objects — A bin picking scenario. *IEEE Xplore*. pp. 1-2. <https://doi.org/10.1109/M2VIP.2017.8211452>

Yang, J., Waslander, S.L., 2022. Next-Best-View Prediction for Active Stereo Cameras and Highly Reflective Objects. pp. 1–2. <https://doi.org/10.1109/ICRA46639.2022.9811917>

Zanuttigh, P., Marin, G., Dal Mutto, C., Dominio, F., Minto, L., Cortelazzo, G.M., 2016. Time-of-Flight and Structured Light Depth Cameras. *Springer*. pp. 9-32. <https://doi.org/10.1007/978-3-319-30973-6>

Zhang, M., Cui, J., Zhang, F., 2021. Research on evaluation method of stereo vision measurement system based on parameter-driven | Elsevier Enhanced Reader. *ScienceDirect*. pp. 7-9. <https://doi.org/10.1016/j.ijleo.2021.167737>