Master Thesis

Vision Technology for Robotic Material Feeding in Primary Packing

submitted by

Dipl.-Ing. Volha Leusha
Matriculation Number: 21043638

Supervisor:

Prof. Dr.-Ing. Rolf-Rainer Grigat
Institute of Vision Systems
Hamburg University of Technology

Prof. Dr.-Ing. Andreas Timm-Giel
Institute of Communication Networks
Hamburg University of Technology

Dipl.-Ing. Thomas Lübcke
Peter Breidenbach
Department for Global Packing and Supply Chain Excellence
Baby Care Engineering, Europe
Procter and Gamble

Hamburg, January 2014

Declaration

Hamburg, 20 January 2014

I, Volha Leusha (student of Masters Programme in Information and Communication Systems, Matriculation Number: 21043638, hereby swear that I single-handedly compiled this document. I did not make use of any other sources than those listed here. Furthermore I swear that this document was never handed in to another department in this or any deviated manner. Every passage referring to or quoted from another source is identifiable as such.

Volha Leusha

Abstract

Continuously growing demand of robotic applications in industry pushes the development of machine vision to the direction of more sophisticated technologies. However, complicated 3D technologies are only entering the market and the simplification of currently used vision systems is still predominant. There is no need in complete 3D reconstruction of the object in the cases where only a specific portion of information is required to perform the task.

Machine vision can be used in a variety of tasks such as quality control, classification, gripping, etc. In this thesis the gripping task has to be performed by a production robot. The vision system needs to detect an object inside the pallet, calculate its coordinates and rotation angles, and provide it to higher-level tasks. The robot has to grip this object, namely, wicket pins, and place it on production line. This is not a typical industrial application, as pins are randomly distributed inside the pallet and, consequently the coordinates of the pin are random in all three directions of the coordinate system X, Y, and Z.

In this thesis a simple way of object detection and coordinate calculation is presented. A special algorithm of image analysis is developed and statistically tested in laboratory. The objective of the algorithm is the usage of information that is known priory about the object structure in order to be able to calculate portion of information that is needed for a specific task.

The logic of developed algorithm and mathematical models, derived in this thesis, can be reapplied in similar applications and for further research.

Contents

List of symbols **iv**

1 Introduction **1**

1.1 Background 1

1.2 Scope of Work 2

2 Related Literature and Theory **4**

2.1 Outlook for Vision Inspection Projects 4

2.2 Optical Foundations 5

2.2.1 Thin Lens Camera Model 5

2.2.2 Pixel Coordinates 8

2.2.3 Depth of Field 9

2.3 Calibration of Camera for Robotic Applications 10

2.3.1 Camera Calibration 11

2.3.2 Robot Hand-Eye Calibration 13

2.3.3 Cartesian Robot Hand Calibration 13

2.4 3D Vision 14

2.4.1 Marr's Theory 15

2.4.2 Shape from X 16

2.5 Illumination Technology 17

2.5.1 Front Lighting 17

2.5.2 Back lighting 20

2.5.3 Aperture and Light Intensity 21
2.6 Image Preprocessing 21
2.6.1 Pixel Operations 21
2.6.2 Local Operations 23

3 Problem Statement and Proposed Approach 24
3.1 Problem Description 25
3.2 Specifications and Constraints 27
3.3 Proposed Approach 29
3.3.1 General Approach 29
3.3.2 Image Capture 30
3.3.3 Calibration 35
3.3.4 Assignment of Pins to Wickets 36
3.3.5 Calculation of Pins 39
3.3.6 Shape Reconstruction of Pins 40
3.3.7 Extraction of Common Gripping Point and Eulerian Angles for Gripper Rotation 42
3.3.8 Illumination 42

4 Mathematical Models 44
4.1 Calculation of Z Coordinate 44
4.2 Recalculation of X and Y Coordinates from Scaling 45
4.3 Calculation of Search Region 48
4.4 Image Distance From Motion 51
4.5 Calculation of Gripping Point 52

5 Tests and Results 56
5.1 Design of Experiments 56
5.2 Evaluation of Mathematical Models 58
5.2.1 First Test: Image Distance From Motion 58

5.2.2 Second Test: Reliability Study 61
5.2.3 Third Test: Calculation of Pin Coordinates from First Capture Distance 65
5.2.4 Fourth Test: Calculation of Pin Coordinates from Second Capture Distance 69
5.2.5 Fifth Test: Calculation of Gripping Point Coordinates from Second Capture Distance 72
5.3 Proposed Solutions for Error Improvement 75
5.4 Results Summary 77

6 Conclusions **80**
6.1 Summary 80
6.2 Outlook 81

Bibliography **83**

List of Symbols

Abbreviations

DOF	Depth of Field
FOV	Field of View
IR	Infrared
SD	Standard Deviation

Symbols

f	Focal length
i	Image distance
R	Rotation Matrix
T	Translation Vector
(x, y, z)	Coordinates in 3D coordinate system
(X, Y, Z)	Axes in 3D coordinate system
(x_c, y_c, z_c)	Coordinates in camera coordinate system
(x_g, y_g, z_g)	Coordinates in gripping point coordinate system
(x_p, y_p, z_p)	Coordinates in pin coordinate system
(x_{pt}, y_{pt}, z_{pt})	Coordinates in pallet coordinate system
(x_t, y_t, z_t)	Coordinates in pin circular top coordinate system
(x_w, y_w, z_w)	Coordinates in world coordinate system
(u, v)	Pixel coordinates
(ϕ, θ, ψ)	Eulerian angles

Chapter 1

Introduction

There is an increasing demand in the industry for automation of processes by using industrial robots. Robots can be applied in variety of tasks, such as packing, gripping objects, placing product or raw material on production line, etc. For such applications machine vision system is often required. This makes robotics vision technology an important topic of discussion. The complexity of the task, variety of objects to be visualized and industrial conditions widen the scope and open a lot of possibilities for development and research. However, findings on this topic in one industrial or scientific application can be reapplied in similar applications.

1.1 Background

Global Packing Group of Procter and Gamble is targeting to save manpower through automation by using robot for bag supply to the primary packing system. Currently this process is done manually by an operator. Bags are delivered as wicket stacks of up to one hundred fifty pieces which are hooked on to plastic spools (pins) in stackable, reusable pallets. Wickets lie on each other forming layers inside pallets, as can be seen in Figure 1.1. The wicket is fixed on wicket conveyer using pins. Therefore, a strategy for robot to grip pins, but not the whole wicket, needs to be developed. The operator takes wickets from pallets and puts it on wicket conveyer as is shown in Figure 1.2. The stack of bags is used to pack diapers inside later. The fastest consumption of a single stack takes approximately one minute for the smallest wicket. However, in the future, when production lines speed up, the process of picking up wickets from pallet will need to be performed faster. This has to be taken into account while designing the strategy for pins detection and robot movements.

Figure 1.1: Pallet with Wickets

Figure 1.2: Wicket Conveyer Belt

1.2 Scope of Work

Not every industrial robotic system requires integrated vision technology. However, for wickets detection, described in this thesis, it is necessary. Wickets are not fixed inside pallets and the exact location of pins can not be predicted without image processing or other detection strategy. Therefore, in this thesis, definition of vision application interfaced with robotic system is described. Specifications of a gripper to pick wickets of bags from a pallet are investigated and taken into account in vision system design.
In Chapter 2 the overview of theories, implemented in the thesis, is provided. It targets to form reader's basic knowledge that will help to understand algorithms designed in subsequent chapters. It starts with the outlook for industrial vision projects where a structural way to organize the workflow in such projects is explained. Next, optical foundations follow with description of basic camera models and metrics that are used in computer vision. The subsequent section targets to provide theoretical information about one of the most critical issues that can be faced in similar projects - calibration of camera. Last

sections are dedicated for 3D vision, where popular techniques of 3D shape reconstruction are listed, as well as illumination technology and image preprocessing techniques.
In Chapter 3 the algorithm of pin detection to calculate the coordinates of gripping point and orientation of pin in space is proposed. Several different strategies are described. After option analysis the most promising ones are chosen for further implementation. Namely, since all pins have same dimensions and circular top, scaling is used to calculate coordinates of pins, and projection of the circle is used to calculate the gripping point of the pin. Moreover, in this chapter camera settings, image capture and illumination setup are discussed in details, being necessary part of every machine vision application.
All the mathematical models used to extract information from images, required for gripping, are gathered in Chapter 4. It provides an easy access to equations used in tests described in Chapter 5. This chapter is about the organization of experiments and there structure. Results from different stages of algorithm implementation are collected and analyzed. Conclusions and suggestions, based on tests and results, are made in conclusion part of this thesis.

Chapter 2

Related Literature and Theory

In this chapter, theoretical aspects used in this thesis are presented. In Section 2.1, steps of industrial vision inspection projects are discussed. This section targets to show important aspects that have to be taken into account in image processing. Moreover, challenges that can be faced by researcher and effective solutions to them are highlighted. Section 2.2 forms the base for a reader in optical foundations such as pinhole and thin lens camera models, depth of focus and depth of field. In Section 2.3 calibration techniques for vision applications are discussed with focus on camera calibration. Section 2.4 provides an overview of the main approaches that are applicable for 3D shape reconstruction from 2D image projections. In the last sections of this chapter, common illumination setups for collecting qualitative images and image preprocessing methods are discussed.

2.1 Outlook for Vision Inspection Projects

Machine vision is playing a significant role in design of intelligent machine systems. Mostly it targets to use image data analysis to provide required information for a high-level task [1].

As is mentioned in [2], design of image processing system should start with detailed assessment of specifications and restrictions of the process under observation. This is often overlooked in image processing applications as they are more associated with software design than mechanical design of machine. The study of specifications should be done at customer site in order to take into account all possible restrictions and issues that can be further faced by vision system designer.

The solution should be checked using sufficient samples that will allow to form reliable statistics and cover variations inherent in the production process. The probability to see good and faulty cases should be taken into account. In the case of unavailability of sam-

ples, parts that are already produced by existing machine, scientific models or prototypes can be used. However, they should be modified to satisfactorily reflect the real situation. Solution provided by designer should also be well-matched with electrical, mechanical and computer systems already existing on production line. Polluted environment, for example, can contaminate or damage equipment, placing restrictions on image capturing procedure. Knowing that, one of the most crucial aspects for image processing applications, to get a good quality image, can not be neglected.
Another important aspect that allows collection of good quality images, is proper illumination technology. Image processing has a direct dependency on information being accentuated through illumination. Therefore, creating an optimal shooting condition and lighting setup are first steps towards successful vision system concept.

2.2 Optical Foundations

The theory presented in this section gives brief overview of optical foundations. For more detailed information about this topic the reader should refer to literature sources [2, 3, 4].

2.2.1 Thin Lens Camera Model

Reconstruction of three-dimensional object from two-dimensional image has been investigated in various past researches. Projection of three-dimensional object into two-dimensional image plane can be described by idealized models of perspective (central) projection or orthographic (parallel) projection [4]. Already in 19th century, researches, such as Laussedat and Meydenbauer, have discussed the first photogrammetric approaches based on perspective projection of real 3D world situation on 2D image plane [5]. The image created by perspective projection is associated with pinhole camera model. In this model light passes through an infinitesimal hole. Assuming the model shown in Figure 2.1 with an image plane having coordinate z equal to f (e.g. focal point is the projection center), the point with real world coordinates $(x, y, z)^T$ can be mapped to the image plane point with coordinates $(fx/z, fy/z, f)^T$ using equation [4, 6]

$$(x, y, z)^T \rightarrow (\frac{fx}{z}, \frac{fy}{z})^T. \tag{2.1}$$

In homogeneous coordinates equation (2.1) may be written as

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fx \\ fx \\ z \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}. \tag{2.2}$$

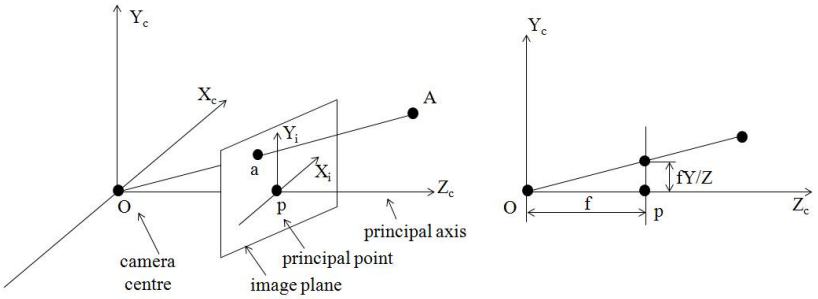

Figure 2.1: Pinhole Camera Geometry [6]

These equations describe mapping between world and image coordinates assuming central projection. The line that is perpendicular to image plane and originates from camera centre is known as principal or optical axis of the camera. The point where the principal axis crosses the image plane is called principal point. If principal point is not at the origin of image plane, Equations (2.1) and (2.2) can be modified into quations as follows:

$$(x, y, z)^T \rightarrow (\frac{fx}{z} + p_x, \frac{fy}{z} + p_y)^T, \tag{2.3}$$

and

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} fx + zp_x \\ fy + zp_y \\ z \end{pmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \tag{2.4}$$

where p_x and p_y are the coordinates of the principle point, as is shown in Figure 2.2.

It should be taken into account that in practice, pinhole camera is supplemented with thin lens for light collection purposes. In reality, an optical camera system combines several lenses and forms a complex assembly. However, in most cases the situation can still be approximated by thin lens model [7].

Thin lens model is an idealized model that is based on the assumption that lens dimensions are irrelevant for the image. The model consists of an ideal lens with focal length f. The focal length is the distance between the central point of a lens and its focal point. The

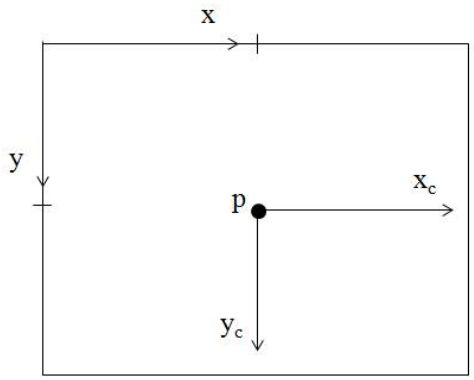

Figure 2.2: Image (x, y) and Camera (x_c, y_c) Coordinate Systems [6]

geometry of imaging by a thin lens is shown in Figure 2.3. This model is based on following laws [2]:

1. If a ray is parallel to the optical axis on one side of lens it crosses a point F(focus) on the other side of the lens.
2. If a ray travels through the center of the lens it does not change the direction of travel.
3. If a ray crosses the focus point on one side of the lens it is parallel to the optical axis on the other side of the lens.

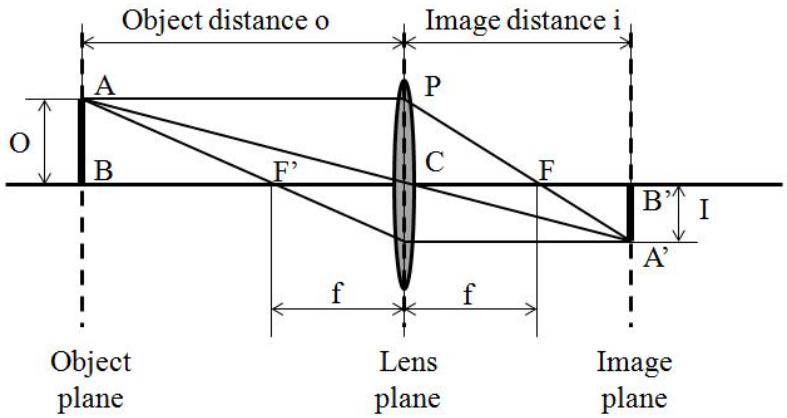

Figure 2.3: Geometry of Imaging by a Thin Lens [2]

Both camera models discussed in this section, pinhole and thin lens, can be considered geometrically the same. Therefore, central projection model to map coordinates of real

world objects into image plane is applicable for both cases [4]. Moreover, from the similarity of the triangles ABC and A'B'C z-coordinate can be estimated by using the equation [2]

$$\frac{I}{O} = \frac{i}{o}. \tag{2.5}$$

In spite of the fact that thin lens and pinhole model are similar, there is a significant difference between them. The fact specific to a thin lens model is that only those points of real world scene that are located at a defined special distance from the lens center can be projected sharply on the image plane [4].

2.2.2 Pixel Coordinates

The formation of pixel coordinates is explained in [8] and summarized in this section. Light falls at the image plane on sensors and irradiance is changed into electrical signal at specific discrete positions. These discrete image points form pixel coordinates. In each pixel coordinate the signal value is stored, for example, for grey images, the grey value corresponds to each pixel position. The pixel positions are specified by the row and column number they are located in. The columns are counted from left to right and rows are counted from top to bottom. Assuming that the coordinates of first left top pixel $x_{0,0}$ are (x_0, y_0) and the distance between the center of two subsequent pixels are $\triangle x$ and $\triangle y$ it follows that the position of pixel in image coordinates is

$$x_{m,n} = \begin{bmatrix} x_0 + m \triangle x \\ y_0 + n \triangle y \end{bmatrix}. \tag{2.6}$$

Four parameters highlighted above, x_0, y_0, $\triangle x$, and $\triangle y$ are related to intrinsic camera parameters, as they are not influenced by the location and orientation of camera in real world scene. Pixels form the pixel array with size MxN pixels. In perfect case, the center of pixel array $(M/2, N/2)$ and the principle axis of the camera should overlap. In reality it is not always true. However, geometric distortions are almost never required to be corrected between image and pixel coordinate systems. It happens due to the fact that detectors of modern semiconductors are almost ideally aligned on a rectangular grid. The coincidence of optical axis with the centre of image plane can be assumed.

2.2.3 Depth of Field

From the basic laws presented in Section 2.2.1 it can be concluded that all the rays coming from point A of the object are meeting at point A' that is at image distance i from the lens plane. For this distance the point appears focused showing the point on the image plane. For distances close or farther away from the lens the point is defocused and appears as a blurring circle or the so called circle of confusion [2]. This effect is presented in Figure 2.4.
The radius of the circle of confusion ϵ is increasing with increasing distance from the image plane and can be expressed by formula [8]

$$\epsilon = \Delta d' \frac{1}{N_f} \frac{f}{f + d'}, \tag{2.7}$$

where N_f is f-number of optical system that is related to focal length f and diameter of lens aperture $2r$, d' is distance from focal point to image plane, $\Delta d'$ is the distance from the focused image plane to the plane of interest. N_f is given by

$$N_f = \frac{f}{2r}. \tag{2.8}$$

The range of positions $\pm\Delta d'$ distant from the focused image plane is called depth of focus. Rearranging equation (2.7) and solving for $\Delta d'$ the following formula can be derived:

$$\Delta d' = N_f \left(1 + \frac{f}{d'}\right) \epsilon. \tag{2.9}$$

If the circle of confusion is smaller than a single sensor element of the camera chip, the point will not blur, as only one element will illuminate [2].
The more practical concept is the range of distance from the camera to the object within which the object appears sharp if image plane is fixed. This range is known as depth of field. This is illustrated by Figure 2.5 and can be expressed by formula [8]

$$d \pm \Delta d' = \frac{f^2}{d' + N_f \left(1 + \frac{f}{d'}\right) \epsilon}. \tag{2.10}$$

From Figure 2.5 it can be seen that depth of field is dependent on aperture size. With the reduction in aperture size, the depth of the field increases.

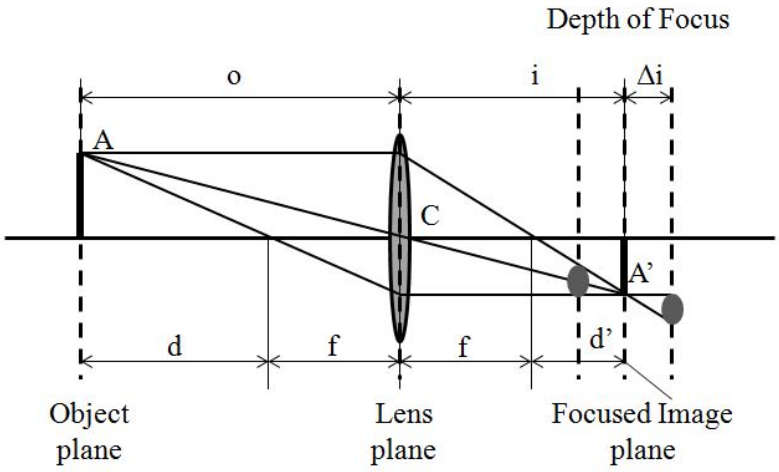

Figure 2.4: Focused vs Defocused Object [8]

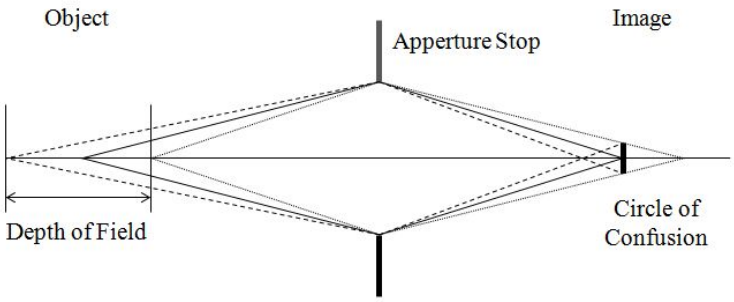

Figure 2.5: Depth of Field [8]

2.3 Calibration of Camera for Robotic Applications

In this thesis, an industrial application that requires collaborative work of vision system, robot, and gripper is discussed. The task of the vision system is to detect the location and orientation of an object in 3D scene, wicket pin inside a pallet, and supply a robot with this information. However, the robot should know where the object is relatively to its coordinate system, in order to be able to send the gripper to proper gripping point. Namely, in order to get the coordinate and orientation of the object in robot base coordinate system, it is necessary to learn the relationship between coordinate systems of robot base and robot hand, hand and camera, and camera and object. Tsai in [9] points out that three calibration procedures, listed below, can solve these three tasks:

1. Camera calibration.
2. Robot Hand-eye calibration [9, 10].

3. Cartesian robot hand calibration [11].

The relationship between these procedures is shown in Figure 2.6.

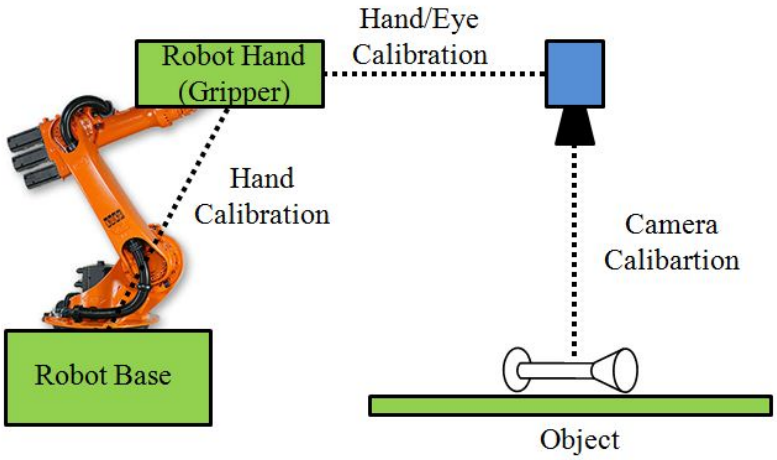

Figure 2.6: Calibration Trio [9]

2.3.1 Camera Calibration

According to [3], camera calibration is the task of computing the relationship between a camera lens and an image plane, as well as between the camera lens and 3D scene. The objective of this task is to extract intrinsic and extrinsic camera parameters. If camera settings are not changed after calibration, these parameters can be used for image analysis throughout the entire application.
The camera lens is connected with the real world scene by extrinsic parameters, namely rotation matrix and translation vector [12] according to equation

$$s\tilde{m} = A[R \quad T]\tilde{M}, \tag{2.11}$$

where camera model is pinhole, $\tilde{m} = [u, v, 1]^T$ is augmented vector of image projection, $\tilde{M} = [X, Y, Z, 1]^T$ is augmented vector of point coordinates in 3D scene, s is arbitrary scale factor, R is rotation matrix, T is translation vector. According to [8], translation vector is the vector by which the origin of real world coordinate system is shifted from the origin of camera coordinate system. Rotation between camera coordinate system and real world coordinate system can be performed in three steps as it contains rotation about three coordinate axes X, Y and Z, and forms three Eulerian angles (ϕ, θ, ψ) as can be seen in Figure 2.7. From mathematical point of view, rotation is multiplication with a matrix and translation is a vector subtraction, as it is shown in the following formula:

$$S_c = R(S_w - T). \tag{2.12}$$

Here $S_w = (x_w, y_w, z_w)$ are world coordinates, $S_c = (x_c, y_c, z_c)$ are camera coordinates. The procedure of rotation from the world coordinate system into the camera coordinate system includes following three steps:

1. Rotation about Z_w axis by ϕ, given by

$$R_\phi = \begin{bmatrix} \cos\phi & sin\phi & 0 \\ -sin\phi & cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.13}$$

   where $S_1 = R_\phi S_w$.

2. Rotation about X_1 axis by θ, given by

$$R_\theta = \begin{bmatrix} & 0 & 0 \\ 0 & cos\theta & sin\theta \\ 0 & -sin\theta & cos\theta \end{bmatrix}, \tag{2.14}$$

   where $S_2 = R_\phi S_1$.

3. Rotation about Z_2 by ψ, given by

$$R_\psi = \begin{bmatrix} \cos\psi & sin\psi & 0 \\ -sin\psi & cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.15}$$

   where $S_c = R_\phi S_2$.

Formulas (2.13-2.15) are valid for right-hand coordinate system and rotation angles are defined positive in counterclockwise direction.

Parameters describing the relationship between the camera lens and the image plane are known as intrinsic camera parameters or the camera intrinsic matrix A, given by [12]

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.16}$$

where camera model is pinhole, (u_0, v_0) are pixel coordinates of principle point, α and β are scale factors in u and v axes, and γ is skew parameter. Many different techniques

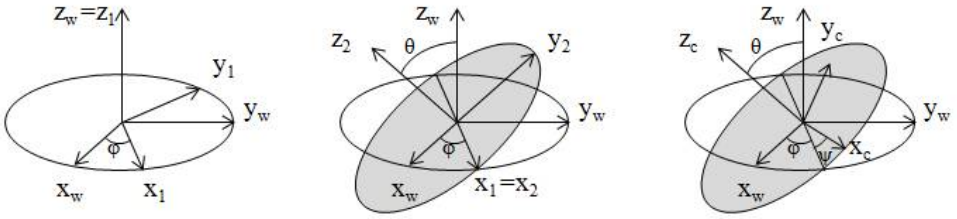

Figure 2.7: Rotation of World Coordinates S_w and Camera Coordinates S_c Using Eulerian Angles [8]

exist to perform camera calibration. In some of them, the image of calibration object with known spatial parameters is used, e.g. chess board. These techniques are known as photogrammetric. Some of them, namely self-calibration, use several images of the same scene from different positions and angles. An overview of most popular techniques used in computer vision application are provided in [3]. Wöhler starts with the outline of camera calibration approaches that target to extract only intrinsic camera parameters and proceeds with description of more advanced techniques that perform simultaneous extraction of intrinsic and extrinsic parameters. The most popular of them are Tsai [13], Zhang [14] and Bouguet [15]. An example of self-calibration technique can be investigated by the reader in [16].

2.3.2 Robot Hand-Eye Calibration

Robot hand-eye calibration is the calibration between camera coordinate system and robotic hand coordinate system, when the camera is attached to the robot. By sending the robot tool to the points that are known in 3D scene and detecting these points in the image taken by the camera, the transformation between these two coordinates can be found. This approach is explained in detail in [9]. However, it should be taken into account that this calibration will introduce some error in the whole system as it is difficult to send robot exactly to the desired point. Consequently, it is not easy to get an accurate knowledge about position of these points.

2.3.3 Cartesian Robot Hand Calibration

Cartesian robot hand calibration is the calibration between robot hand and robot base [9]. Cartesian robot hand calibration algorithm is provided by the robotic company and is left out from this thesis as it is not directly connected with discussed vision application. For more details about this topic the reader should refer to [11].

2.4 3D Vision

This section presents a short overview of existing techniques for reconstruction of three-dimensional (3D) real world scene or objects from two-dimensional (2D) images. For more details regarding this topic the reader should refer to [3, 4, 17, 18, 19].
3D computer vision is considered to be a young and still developing topic in image processing science. The complexity of 3D scene reconstruction from 2D image makes the task ill-conditioned. Sonka shortly describes the difficulties related to 3D vision in [18]:

1. As camera system performs perspective projection of the real world scene, all the points along one ray passing through the optical center, A_1, A_2, and A_3 as is shown in Figure 2.8, are projected as a single image point A'. Therefore, several points that are separated in real world scene can appear as the same point of the image. This leads to a loss of information when reconstruction of 3D coordinates has to be performed and additional information has to be added to reconstruct the scene properly.

2. The pixel intensity of the point that is projected on the image plane is dependent on a lot of factors, such as surface reflectivity, illumination angle, orientation of the surface, etc. Therefore, reconstruction of surface orientation and depth, influenced by the listed factors, is a difficult task.

3. The intersection and occlusion of objects and presence of noise from environment are also critical restrictions in 3D shape reconstruction.

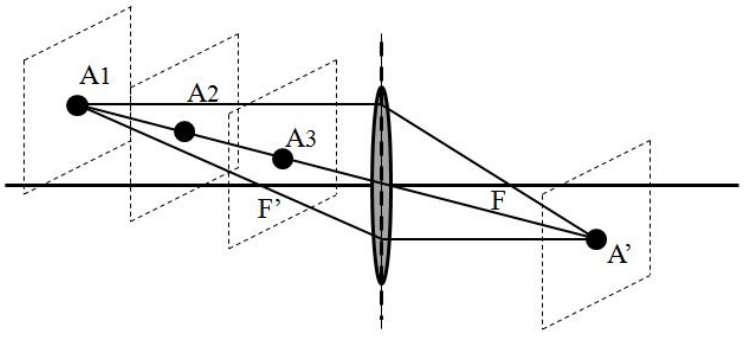

Figure 2.8: Projection of Several PointS Along the Ray Passing Through Optical Center [4]

Sonke in [18] states that there are several approaches to solve 3D reconstruction task such as Marr [20], Aloimonos and Shulman [21], Wechsler [22], Aloimonos [23]. The Marr

theory, presented in Section 2.4.1 is still the commonly accepted paradigm.
The theories can be categorized as reconstruction bottom-up or model-based vision top-down approaches. Bottom-up approach allows reconstruction of 3D geometrical descriptions of the object from one or more intensity or range images. Top-down approach assumes that there exists the model of the object that is based on priory knowledge about this object.

2.4.1 Marr's Theory

In this subsection, a short summary of Marr's approach, based on [20] and [18] is presented. Marr's theory is a strict case of bottom-up approach for 3D geometric reconstruction of objects from intensity images. For the reconstruction very little information about objects in real world scene is required.
Marr's approach is more abstract and theoretical, and restricted to a single static scene. Marr divides the task into several steps.

Transformation of input intensity image into primal sketch

The target of this step is to extract significant edges from the image in viewer-centered coordinates.

Derivation of 2,5D sketch from primal sketch

It can also be called as a depth map as the target is to extract the distances between object and camera. This stage does not provide complete 3D shape information, but depth and local orientation of surface. It is considered to be intermediate step between 2D and 3D representations. The main strategies to get 2,5D sketch from 2D representation are known as 'shape from X' and are described in Section 2.4.2.

Full 3D representation of object geometry

Facts, extracted in first two steps, are used to identify the 3D structure of objects. Namely, in this stage the conversion to object-centered coordinate system is performed. This requires information about features that can describe objects.
3D reconstruction of the object is a very complex task and can not be solved for general case. The solution for this task is very difficult to obtain and requires complex algorithms.

2.4.2 Shape from X

Shape from X are the techniques that use information about intensity images to extract shape of the objects. Many of these techniques are used to find local surface orientation. Some of these techniques are listed and described in literature sources [3] and [18]: shape from contour [24, 25, 26], shape from shading [27, 28, 29, 30], shape from stereo [4, 18], shape from motion [31, 32, 33], shape from optical flow [18], shape from texture [34], shape from focus and defocus, etc. Some of them, that are considered to be the most interesting for this thesis, are briefly explained below.

Shape from Shading

This method is based on reconstruction of three-dimensional information from pixel intensities of a single image. Intensity of pixels depends on light sources, surface reflection properties and surface orientation. Photometric information allows extraction of information about these surface gradients. However, some assumptions have to be made about illuminations, surface reflections and smoothness in order to be able to reconstruct full height map surface projected on image plane [3].

Shape from Contour

Shape from contour approach is the technique of 2D shape reconstruction from one or several images based on contour extraction. However, this approach is not always reliable because of some limitations. First of all the contour should be distinguishable (form an edge on image plane) and projection is not always invertible, as two different shapes can have the same projection [18].

Shape from Stereo

Stereo vision is based on binocular image acquisition technique and has the same principle of three-dimensional scene reconstruction as visual systems of humans. Human vision is based on reconstruction of depth combining images from two eyes. This is known since middle of 19th century.
In order to be able to reconstruct 3D shape from two different views, two images should be matched with each other. Namely, it should be learnt which point from one image corresponds to a point in a second image. The main issue of stereo vision is, that usually for a selected point in one image there exist more than one point in the other image that can be a candidate for correspondence [18].
Assuming that there is no object movements between acquisition of two images, correspondence calculations will not be influenced by motion. Such stereo vision is known as static. This technique, in contrast to dynamic stereo vision that is aimed to understand

motion, targets to reconstruct depth or shape of the real scene [4].

Shape from Motion

According to [18], shape from motion technique is very similar to shape from stereo in terms of approach structure. Several images are taken in different time frames and combined together by finding correspondence between same points on both images. However, for motion analysis, correspondence problem is considered to be less complicated as position of point in neighboring image can be predicted by estimating the trajectory of motion. In most cases even this prediction is not needed as time separation between neighboring images is too small and corresponding points are too close to each other.
The structure from motion theorem says that three orthographic projections of four non co-planar points have a unique 3D interpretation as belonging to the same rigid body. This theorem is also applicable for perspective projection and is proved by Ullman in [31]. The derivation and further explanation of this theorem together with formulas for shape reconstruction from motion are provided in [18].

2.5 Illumination Technology

The proper setup of illumination is one of the basic requirements in image capture process. Usage of different types of light allows one to highlight or hide specific features of observable object or scene, making illumination techniques a powerful mechanism in vision applications. Jähne in [8] classifies illumination techniques according to the applications that they are applicable for.
However, in this section, the described classification is based on [2], where Demant divides illumination techniques in two main basic setups: front and back lighting.
Further classification is presented in Figure 2.9 and is summarized in following subsections using [2] and [8] literature sources.

2.5.1 Front Lighting

Front lighting means that the light source is located on the same side as camera. Choosing different angle of incidents, type of light source, its shape and location, designer of vision application can simulate different conditions and get various effects.

Diffuse Front Lighting

Diffuse front lighting can be used for applications where it is required to avoid specular reflections or to eliminate shadows. The diffusion is achieved by radiation from large

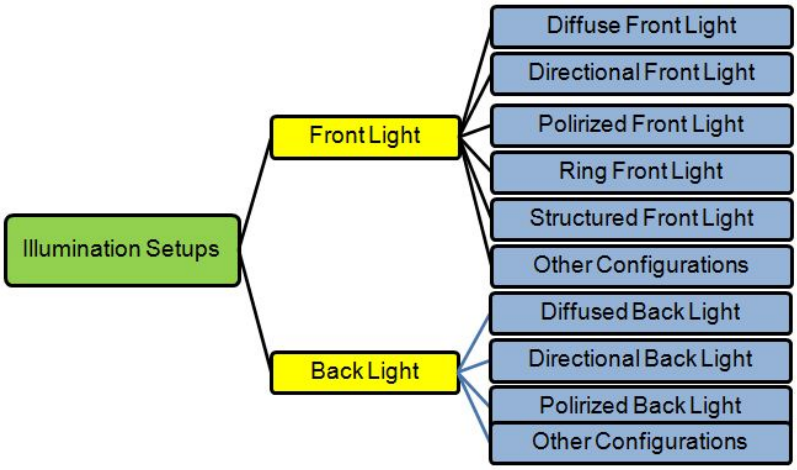

Figure 2.9: Illumination Technology Classification

uniform area or by indirect light rays that are first reflected from white screens. The setup of such illumination is shown in Figure 2.10.

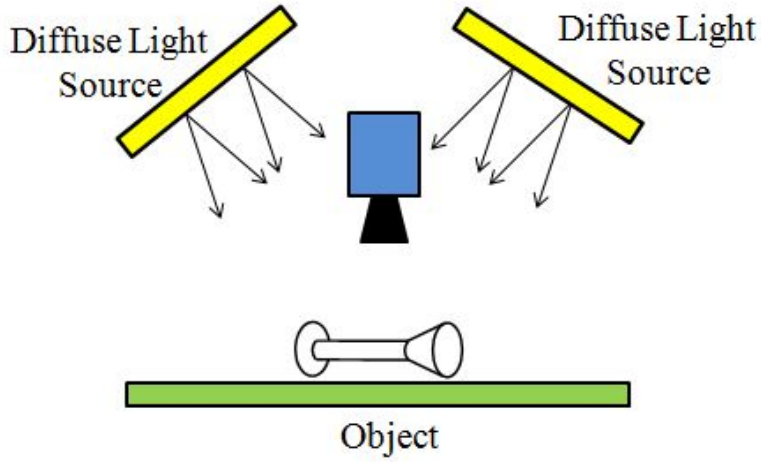

Figure 2.10: Diffuse Front Lighting Setup [2]

Directional Front Illumination

In directional front illumination, shown in Figure 2.11, the incident angle of light is very small. It creates the effect of dark field, namely reflections from smooth surface will not enter the lens and will appear dark in the image. If the surface is structured it will appear light in the image.
This type of illumination is often used for identifying marker indentations, physical stamps or imperfections on smooth surfaces.

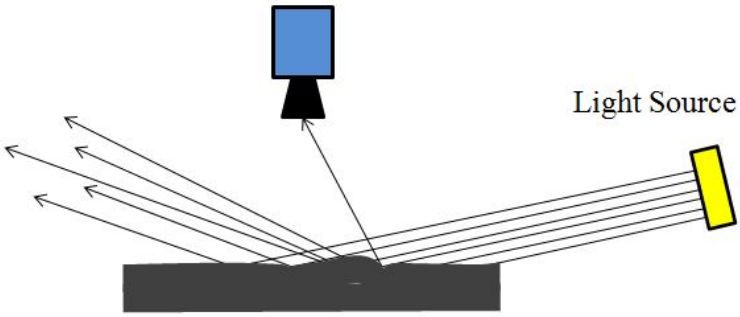

Figure 2.11: Directional Front Lighting Setup [2]

Polarized Light

Polarized light is used for cases when it is needed to eliminate reflections. Polarization of incident light can be achieved by applying polarizing filters between light and object. The polarized light has a specific feature that it vibrates only in a direction that is perpendicular to the direction of propagation. Therefore, radiation from specular reflections will be fully polarized. By placing a polarizing filter between camera and object at right angle to polarization direction, the radiation from specular reflections will not pass this filter. Consequently, these reflections will not appear in the image as is shown in Figure 2.12. However, it should be taken into account that filters in general absorb some light and therefore higher light intensity is required.

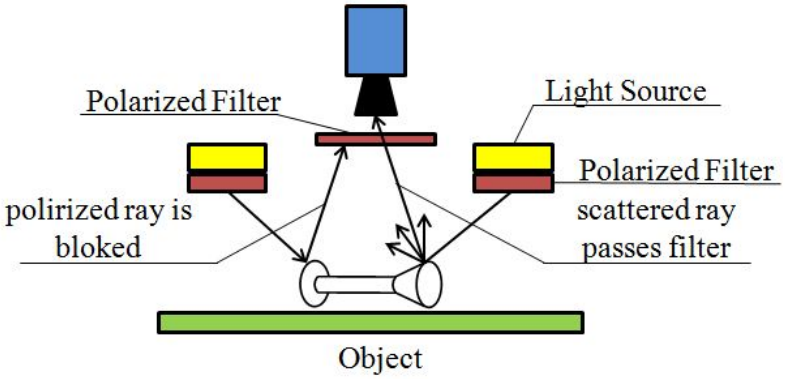

Figure 2.12: Polarized Lighting Setup [2]

Ringlight

Ringlights are often used in combination with polarizing filters. They provide intensive illumination of surface and are used in applications where it is necessary to avoid shadows.

The camera is mounted inside the hole of the ringlight that provides light distribution along optical axis of camera.

Structured Light

Structured light technique is based on the analysis of patterned light with which an object is illuminated. This light can have various forms such as circles, grid, lines, etc. The three-dimensional characteristics of the object can be extracted using information of how the pattern light is deformed when it touches the surface of the object. For example, if there is a curved surface under observation the line pattern will be transformed to a curve [35]. This technique is also useful for presence verification and depth estimation [35] as is shown in Figure 2.13. For the first case, if there is no object in camera field of view the pattern will not change. However, if the object is placed under the light source, the pattern will appear deformed. For the second case, the method is known as contrast-free height gauging. Objects that have different depth coordinate will appear as different lines of light in the image. The distance between these lines can be used to calculate the height [2].

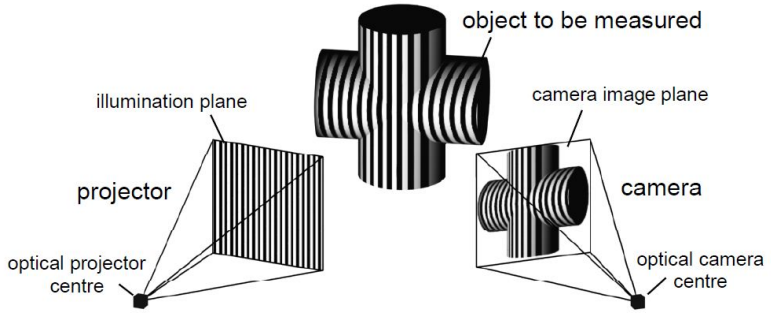

Figure 2.13: Structured Lighting Setup [35]

2.5.2 Back lighting

In back lighting, the object is placed between the camera and the source of light so that it creates a shadow contour in the image. According to [2], main back light setups are diffuse back lighting, directional back lighting and polarized back lighting. However, as back lighting is considered to be not feasible for the application developed in this thesis, the theoretical information about it is not further discussed in this section.

2.5.3 Aperture and Light Intensity

As was mentioned in Section 2.2 one of the possible solutions to increase depth of field is to reduce the aperture. However, the smaller is the aperture the less is the light passing through it. Consequently, a proper source of light should be chosen in order to supply the system with enough light intensity.
Not all the light that is falling on the surface of the object is participating in image formation. There is a considerable amount that is lost due to different reasons. For calculating optimal settings for camera shutter speed and aperture, luminance (reflected light) of the object should be measured. However, according to [36], it is complicated to determine it accurately. Therefore, approximation methods are preferred in practice, in spite of the fact that they are less accurate. Namely, measurement of total scene luminance, measurement of incident light, measurement of darkest shadow or lightest highlight, etc. The author also derives formula, connecting illuminance E_s with shutter speed t and aperture as follows:

$$E_s = \frac{CN^2}{tS}, \tag{2.17}$$

where C is the calibration constant, with typical value of 20.83 for a mid-grey tone, S is the ISO value, and N is the relative aperture.

2.6 Image Preprocessing

As is discussed in [2], preprocessing is a processing step transforming a captured image into a new image. The new image is essentially analogous to the source image, but specific aspects of these two images differ from each other. This processing step helps in analysis of the image by extracting or highlighting an important for a certain application features of a scene or an object. Preprocessing functions can be generally divided into two groups: pixel operations, such as image arithmetic, and local operations, such as linear and non-linear filter, etc.

2.6.1 Pixel Operations

In this subsection pixel operations are discussed using literature sources [2] and [37]. Pixel operations include image arithmetic and grey scale transformation. In image arithmetic, the output image is the result of pixel-wise combinations of two or more images. The resulting value of the output image depends on values of this pixel in all the images

participating in the operation. Addition, averaging, maximization are the examples of image arithmetic operations.
Grey scale transformation has a general form of

$$o_{m,n} = f(i_{m,n}), \tag{2.18}$$

where $i_{m,n}$ is the grey level of pixel $x_{m,n}$ of the input image and $o_{m,n}$ is the grey level of the output image pixel after it is trasformed with function $f()$. The transformation function $f()$ is indifferent to the pixel position. Among grey scale operations are such operations as look-up tables, histogram equalization, contrast enhancement, etc. This section focuses on contrast manipulations that are used in this thesis as the technique to eliminate reflections from surface of bags in the image.

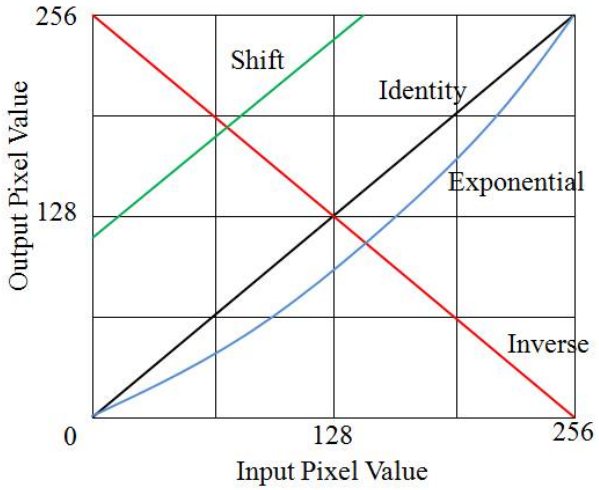

Figure 2.14: Examples of Several Transfer Functions [37]

Transfer function is a mechanism by which the manipulation of image contrast can be explained. It relates the input pixel value with the output pixel value. The contrast of the image can be manipulated in different ways. As it is shown in Figure 2.14, the dependency can be exponential, inverse, shifted, etc. The most popular technique is contrast enhancement. It highlights some image details, at the same time hiding others. However, as the objective of the procedure is to modify the transfer function until the image shows the features that are of most interest for the target application, the type of the used function is application specific.

2.6.2 Local Operations

Local operations, in calculation of output grey value of pixel, take into account the neighborhood of the current pixel. Such preprocessing operations are also called filtering [18]. In filtering operations the neighborhood can also be called a mask or filter kernel. Filtering operation can be formulated as follows [2]:

$$g(x^*, y^*) = f(\{g_i(x, y)|(x, y) \in N^*\}), \tag{2.19}$$

where $g(x^*, y^*)$ is the grey value of pixel at a position (x^*, y^*) in resulting image, $f()$ is the mapping function applied to all grey values g_i of pixels in input image that are located in neighborhood N^*.

According to [18], local preprocessing methods can be divided into two main groups dependent on the target of the processing: smoothing and gradient operators. The aim of smoothing is to eliminate noise or small fluctuations in the image. However, sharp edges can appear blurred in this case, and useful information can get lost. Gradient operators are formed by local derivatives of image functions. The principle of these operations is to find parts of the image where the image function is quickly changing. These locations can be found by having bigger derivatives.

Another classification of local operations can be done by transformation properties: linear and non-linear filtration. Linear filters use the weighted sum of grey values in the neighborhood and the output is linearly dependent on input, whereas for non-linear filter the output is not linearly dependent on input.

Different filters are useful in different situations. For example, Prewitt and Laplacian are commonly known filters for edge detection operations. Gaussian and rectangular filters can be used for smoothing. More detailed information about filters and their possible applications can be found in [2, 18, 37].

Chapter 3

Problem Statement and Proposed Approach

In this chapter, the problem statement is discussed in more detail and an approach to solve this problem is proposed. The workflow structure for this project is designed according to the outlook for vision inspection projects that was provided in Section 2.1 and is summarized in Figure 3.1.

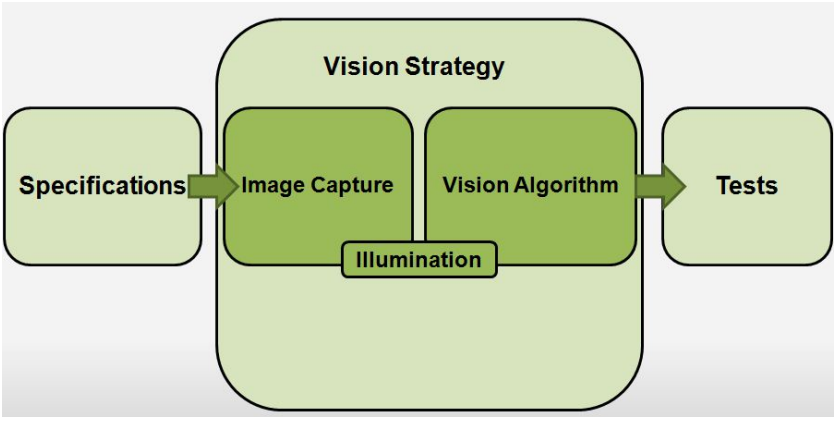

Figure 3.1: Workflow Structure

The work begins with the analysis of process specifications that have to be extracted by vision system. These specifications are needed by the robot in order to send the gripper to wicket pins. The second step is the design of a vision strategy that is subdivided into several tasks: image capture, vision algorithm, and illumination. Vision algorithm is the detailed step by step analysis of the image captured by camera. This allows to extract specifications required for higher-level tasks. To perform image analysis, designer of the vision strategy should think about capture conditions and camera settings. This is necessary to obtain the quality of image that is enough for specific analysis. As was stated in

Section 2.2, light rays form the image. Therefore, in parallel to image capture and vision algorithm strategy, illumination of the scene should be defined. After the vision strategy is decided, it should be tested using design of experiments and statistic tools.
In Section 3.1 the problem of manual wicket loading is explored. The study of the process is done on a production line, in order to understand conditions and constraints affecting extraction of specifications.
The available gripper design is presented in Section 3.2. Specifications, that have to be extracted by vision algorithm are defined according to this design. Moreover, possible issues and capture conditions are discussed in this section.
The proposed vision strategy is explored in Section 3.3. The sections starts with brief overview of the strategy. In Section 3.3.2 image capture and camera setup are discussed. Field of view and depth of field are calculated for the application. The calibration procedure is explained in Section 3.3.3. In subsequent subsections the vision algorithm is discussed in more details. Procedure of image analysis is explained step by step starting from assignment of pins to wickets and finishing with calculation of coordinates of gripping point and shape reconstruction of pins. The illumination setup is discussed in the last subsection.

3.1 Problem Description

The target of this thesis is to develop a vision system that can detect wicket pins and send this information to higher-level tasks. The application overview has been given in the introduction section. This section provides more detailed description of the process and issues that can be faced in design.
According to information, provided by P&G Techpack group, there are more than two hundred types of wickets. They have different sizes and artworks depending on product type and number of diapers in the bag. The pallet and pins are always of the same size and shape regardless of the wicket type (see Figure 3.2). The surface of the wicket is shiny and slippery. This creates undesirable reflections from the surface of bags. Moreover, wickets move inside the pallet during transportation and the layout can be displaced.

The distance between two pins of one wicket for the same wicket type is fixed. This comes from the process specifications. Nevertheless, the wicket can be wrinkled between pins and the distance between pins can be projected smaller on the image plane than it is in reality. This is reflected in Figure 3.3. The distribution of wickets inside the pallet, together with transportation conditions, result in chaotic distribution of pins. Some pins lie close to the wall of the box. This creates a difficult situation for gripper design. As

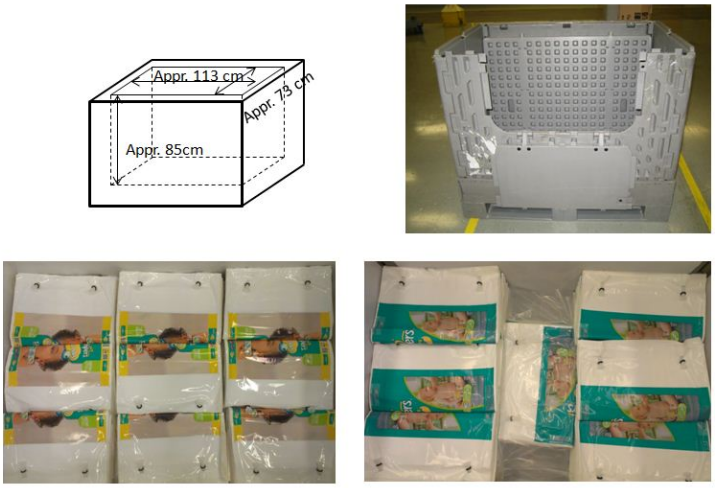

Figure 3.2: Wicket Pallet Dimensions and Wicket Types

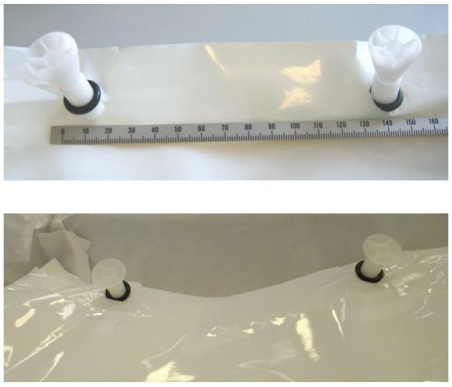

Figure 3.3: Deviation in Fixed Distance between Two Pins of One Wicket

pins are not fixed with any additional equipment inside the pallet they can be differently oriented in X ,Y, Z direction (shown in Figure 3.4). Therefore, x, y, and z coordinates of pins can not be predicted without any additional technology. In this thesis, the vision strategy was proposed and investigated as such additional technology.

Figure 3.4: Chaotic Orientation of Wicket Pins

Dimensions and shape of wicket pins are standard with a slight deviation insignificant for

the vision strategy. The old wicket pin will be substituted by the new one in the near future (see Figure 3.5). Consequently, new pin design is used in tests. The new wicket pin has a different plastic type and shape. Moreover, there is a deepening in the structure of new pin which keeps the black rubber ring in fixed position. It can create less rotation of wicket pin relative to the wicket stack, making orientation of pins less chaotic. However, this should be tested in the future, when new wicket pins are available in mass production.

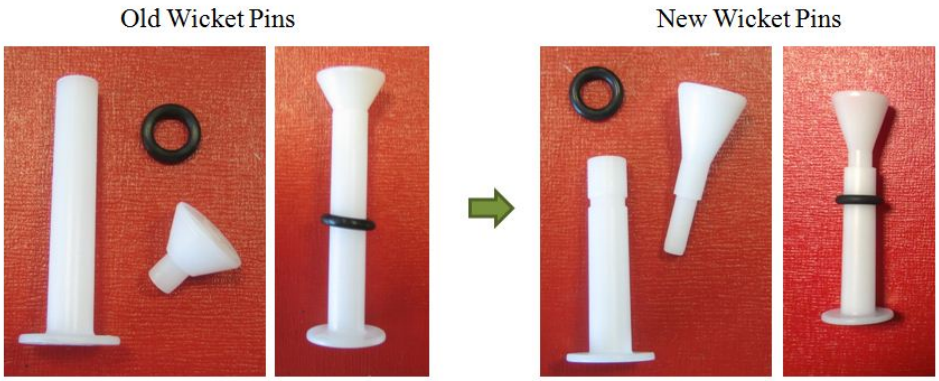

Figure 3.5: Wicket Pin

The top of the pin is circular. The pin is molded from white plastic, however, the possibility to mold it in different colors is discussed with supplier and can be implemented.

3.2 Specifications and Constraints

The setup of coordinate systems for current application is shown in Figure 3.7. X_w, Y_w, Z_w are axes of world coordinate system, X_c, Y_c, Z_c are axes of camera coordinate system, X_{pl}, Y_{pl}, Z_{pl} and X_p, Y_p, Z_p are axes of pallet and pin coordinate systems. Angles (ϕ, θ, ψ) are three Eulerian angles explained in Section 2.3.1.

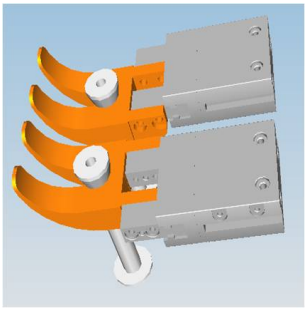

Figure 3.6: Draft Design of Gripper

According to the process study presented in previous section, the gripper draft design was developed and is presented in Figure 3.6. Available gripper design is not final and can be modified in the future. The developed vision application takes into account specifications required for this design. However, these specifications are general and with a slight modification are applicable for any gripper design. From the gripper design, following specifications for gripping strategy and vision system are derived:

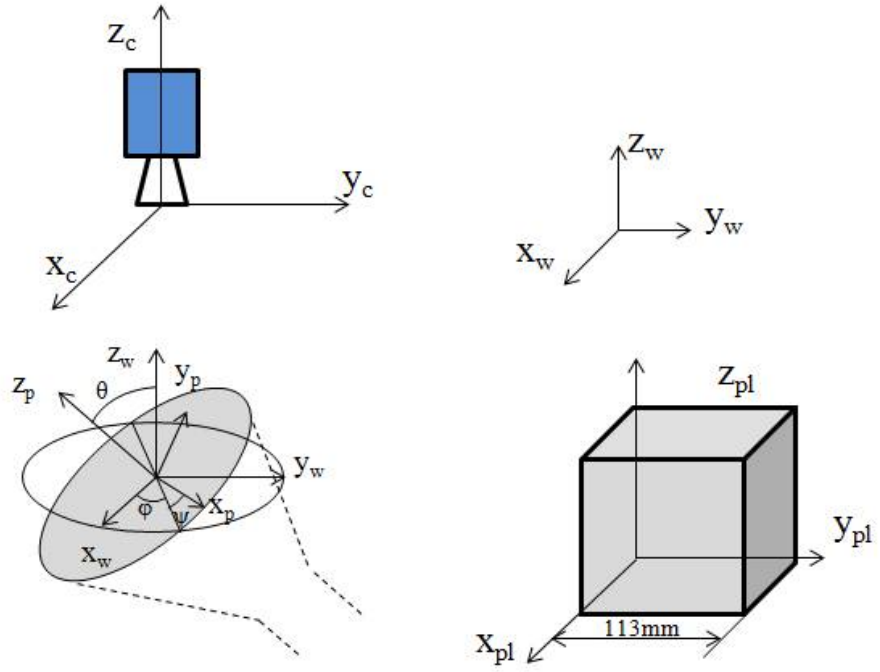

Figure 3.7: The Setup of Coordinate Systems

1. The gripper has to pick two pins simultaneously. Therefore, the vision system should detect both pins before gripping.

2. The gripping area is defined to be between the rubber ring and the head of the pin. Therefore, the gripping point calculated by vision strategy should be the center of the gripping area. From pin structure, the maximum allowable deviation from gripping point is up to 5mm in every coordinate.

3. The gripper has two fingers that are static relative to each other, and are attached to the robot arm. Thus, space coordinates (x_o, y_o, z_o) of common point O for the whole gripper should be defined. The common point is the middle point of the line connecting gripping points A and B of two pins belonging to the same wicket (see Figure 3.8).

4. Eulerian angles $(\phi_{cg}, \theta_{cg}, \psi_{cg})$ for gripper rotation can be calculated from orientation of vector AB connecting gripping points of two pins belonging to the same wicket as is shown in Figure 3.8.

From the first image analysis several constraints are observed. The pin is currently white and is normally lying on white background. This makes it almost impossible for camera to detect it. Therefore, it is decided to mold the heads of pins in black in the future. Additionally, the edges of wicket are not always visible. Sometimes white edges of one wicket

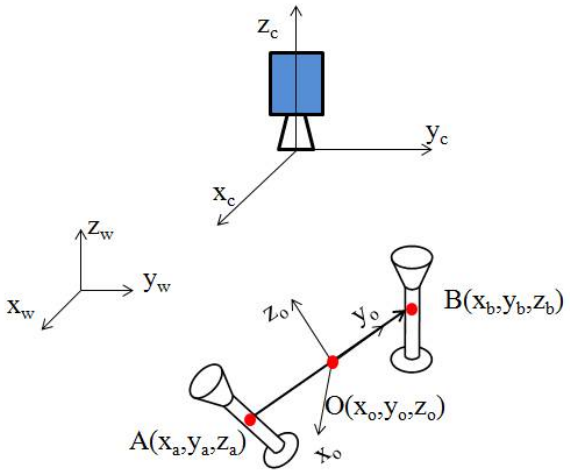

Figure 3.8: The Common Gripping Point and Gripper Coordinate System

are covered with white edges of another wicket. Consequently, an alternative strategy to separate wickets from each other is required, and a proposal is presented in Section 3.3.4. Reflections from the surface of bags create undesired edges in image analysis. This problem is discussed in more details in Section 3.3.5.

3.3 Proposed Approach

3.3.1 General Approach

The first question arising in such applications is: where to place the camera? The camera can be static, attached to an additional device, or mounted on robot. In the second case, the camera can be moved closer to pin, allowing to get the bigger projection on the image plane. It should also be taken into account that the coordinate system of camera will move together with the coordinate system of robot hand. Consequently, every time when the robot hand moves to a new location, coordinate system of camera also should be recalculated.

In static approach, the camera location should be chosen so that resulting image has enough resolution for analysis. At the same time it is critical to take into account that the camera should never crash with the robot arm. Shadows of the robot can also influence the image because they create unstable environment. Performing an option analysis by analyzing all discussed pros and cons of both strategies, it is decided to mount the camera on the robot.

As has been already discussed in Section 3.2, each wicket has two pins that have to be detected and calculated before the gripping. This creates the second question: how to

detect which pin belongs to which wicket?
Taking this into account, the vision strategy is divided into two steps: assignment of pins to wicket and calculation of coordinates and angles of pins.
For the first step, field of view, or the area of object space imaged at the focal plane of a camera [8], can be bigger as the target is to detect a wicket and pins with a certain approximation that is enough for assignment of pins to wickets and approximate calculation of wicket coordinates. There are more than two hundred different types of wickets with widths varying from $190mm$ to $750mm$. This should be taken into account while designing the strategy for pin assignment. It should be as much universal and standard as possible, in order to avoid complication of software and additional work during roll-out stage of the project.
The objective of the second step is to extract the coordinates of pins with higher precision. Namely, with error of less than $5mm$ according to gripper design. For this purpose, laser line scanner can be used. This requires additional equipment, but gives a precision in micrometers. From financial point of view it is cheaper to use the same camera for the second step. However, it can result in some problems that have to be solved. The bigger the pin appears on the image the smaller the error is in calculation of its coordinates. This means, smaller field of view, and results in closer distance to the pin for the second image capture. The main issue of this strategy is to choose appropriate camera settings that provide good quality for image analysis in both steps. Namely, the aim is to understand how to assure that camera is focused enough for both distances.

3.3.2 Image Capture

It was shown in previous subsection, that two steps algorithm proposed in this thesis can be approached by using different strategies. For the first strategy, in case if the camera and line scanner are used together, it is easy to focus the camera for one distance. However, in case if a single camera is used the difficulty with focusing the camera appears. If the camera is out of focus for the distance of observation, the object appears blurred. If the object is blurred more than acceptable for the application, image analysis does not provide reliable information.
In example, shown in Figure 3.9, for the second distance the pin appears focused and for the first distance the picture is indistinct. However, as discussed in Section 2.5.3, closing the aperture increases the depth of the field. At the same time, as there is less light passing through the lens the time of light collection should be increased and a slower shutter speed needs to be used. Because of unstability of environment slowing the shutter speed is possible only to a certain extent. Instead of slowing the shutter speed an additional

amount of light can be added. In this thesis, the real world scene is static. Therefore, there are no specific restrictions on shutter speed. However, the shutter speed is synchronized to the frequency of the current source for artificial light of oscillating type. This need has been talked about in [38, 39]. For this work shutter speed is assumed to be 20ms.

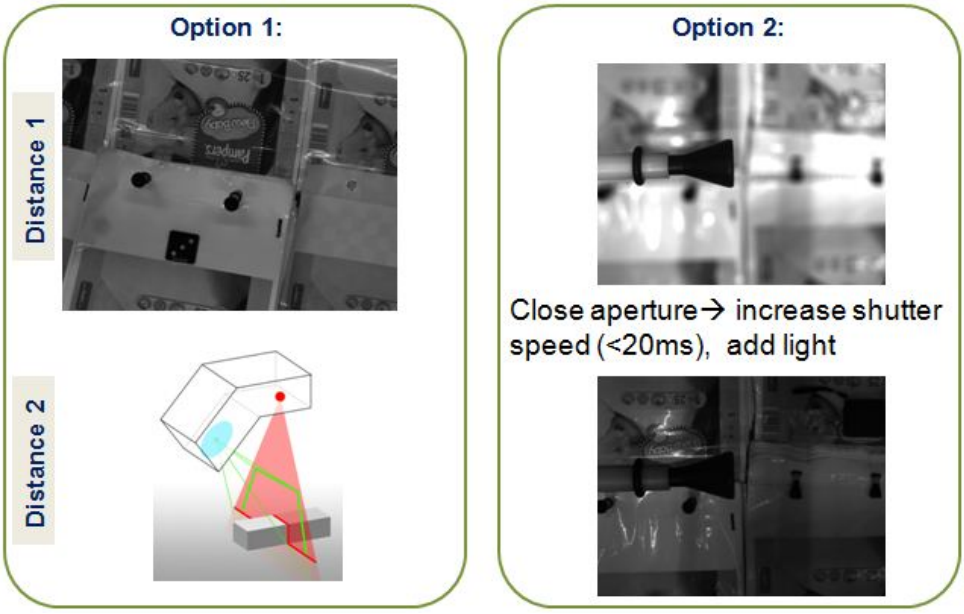

Figure 3.9: Two Strategies for Two Step Algorithm of Pin Detection

The camera used in this thesis has an array length of 1/1.8″. The dependency of field of view (FOV) on working distance is provided for 1/1.8″ camera in Figure 3.10. Where working distance is the distance from the front lens of an optical system to the object plane [8]. The dependency in Figure 3.10 is shown for different focal lengths of the lens, starting from $4mm$ and up to $75mm$. The highlighted information is taken from the data sheet of the camera. As we are interested to get an image from two different distances, the data from the graph is analyzed in more detail only for the two FOV under interest.

The FOV for the first distance, when the assignment of pins to wicket has to be done, is assumed to be approximately one fourth of the pallet area in x,y-pane. The

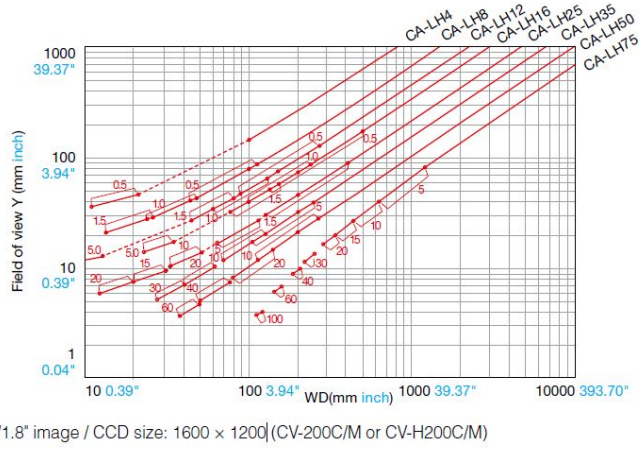

Figure 3.10: Field of View vs Working Distance

segmentation of the first shot into four images allows to decrease the FOV and, consequently, to achieve better resolution. Moreover, the closer the first distance for image capture is to the second distance the narrower is the depth of field. It means that the aperture can be closed lesser and higher amount of light passes through the lens. It simplifies the setup of camera and light for the application. The size of the pallet area is shown in Figure 3.2 and is 113x73cm. One fourth of this size with slight allowance is 60x40cm as is shown in Figure 3.11. Assuming the FOV width to be 400mm and using the graph shown in Figure 3.10 first working distance for 8mm lens is approximately 600mm.

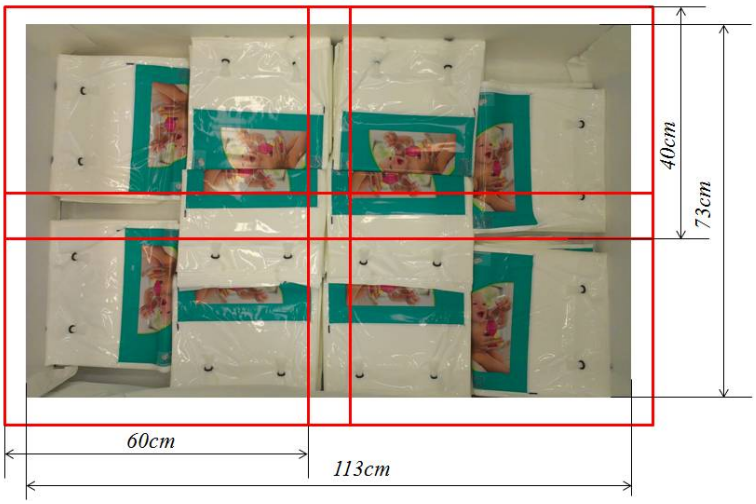

Figure 3.11: Segmentation of the Wicket Layout

The FOV width for the second distance, the distance closer to the pin, assumed to be 120mm in y-direction, results in 120/1200mm for one pixel. Consequently, 198 pixels are equivalent for 19.8mm circular top of the pin. From a theoretical point of view this resolution of image is enough for detecting coordinates of the pin with error less than 0.5mm. For example, for a lens with focal length of 8mm and a specified FOV of 120mm the working distance is approximately 200mm. This model is illustrated in Figure 3.12. z coordinate of the pin is obtained from the size of the pin (for further explanation refer to section 3.3.4).

From the geometrical model presented in Figure 3.12 and similarity of triangles the following can be derived

$$Ri = z_{second}p, \tag{3.1}$$

where R is real world pin size in cm, i is image distance, z_{second} is the distance from lens center to the pin for the second image capture, and p is the size, in pixels, of pin projection on the image plane, or

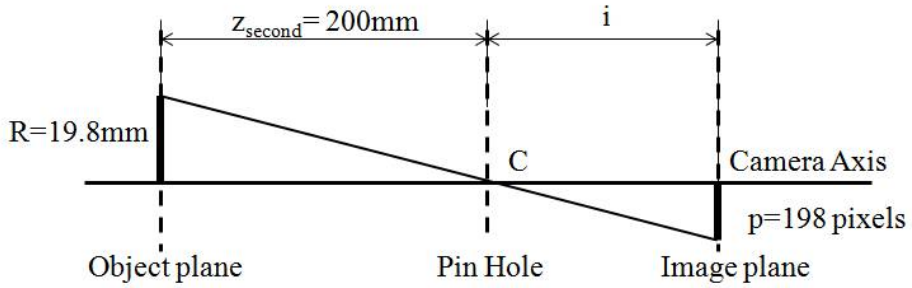

Figure 3.12: Approximate Projection of an Object with 19.8mm Size from Working Distance of 200mm Using 8mm Lens

$$19.8\dot{i} = 200 * 198, \tag{3.2}$$

and

$$i = \frac{200 * 198}{19.8} = 2000[pixels]. \tag{3.3}$$

Assuming the situation, when instead of 198 pixels the pin width is detected as 197 pixels, for the same image distance i, the error, namely, the difference of detected z coordinate from real z coordinate, is given by

$$error = 200 - z_{detected} = 200 - \frac{Ri}{p}, \tag{3.4}$$

or

$$error = 200 - \frac{19.8 * 2000}{197} = 200 - 201 = -1[mm]. \tag{3.5}$$

From equation (3.5) it is clear that an error of one pixel in projection of the pin can result in an error of approximately 1mm in z coordinate for the camera settings, described above.

As is already mentioned for the single camera approach, the coordinates should be obtained out of the image taken from the second distance. Therefore, the camera needs to be focused for this distance, but blurring should be taken into account for the first distance. In addition, the image for the second distance can not be always taken precisely from second capture distance z_c (200mm for 8mm lens) over the pin. As from the first step of the algorithm only approximate coordinates of the pin can be calculated, there is a deviation in second working distance. According to tests, described in Chapter 5, the deviation of $\pm 4.5\sigma$ is approximately 40mm. For the second working distance circle of confusion should

be not more than one pixel for precise calculation of coordinates. For the first working distance circle of confusion can be up to five pixels. According to software requirements, this is enough for reliable pattern detection and rough estimate of pin coordinates with acceptable error. Summarizing all discussed above following requirements can be made for $8mm$ lens:

1. Camera should be focused for working distance of $200mm$.
2. The circle of confusion should be not more than one pixel for deviation from focused distance up to $40mm$.
3. The circle of confusion should be not more than five pixels for deviation from focused distance up to $500mm$.

The same analysis is made for $12mm$ lens with following requirements:

1. Camera should be focused for working distance of $300mm$.
2. The circle of confusion should be not more than one pixel for deviation from focused distance up to $40mm$.
3. The circle of confusion should be not more than five pixels for deviation from focused distance up to $930mm$.

Further, these values are compared with values calculated from Equation (2.10) for depth of field in Tables 3.1 and 3.2. Depth of field is calculated for different f-numbers. The calculation is done for $8mm$ and $12mm$ lenses as for higher focal length the distance between first and second capture distances grows. Consequently, the required depth of field also grows and these lenses are not applicable for current application. The circle of confusion for closer border of depth of field region is assumed to be one pixel and for the farther border is assumed to be five pixels. Namely, for the camera that has $4.4\mu m$ size these values are $4.4\mu m$ and $22\mu m$. Field of view is assumed to be $120mm$. For this field of view, working distance for $8mm$ lens is approximately $200mm$ and for $12mm$ lens is approximately $300mm$.
From Table 3.1 it follows that only for f-number 32 or 64 the requirements are fulfilled for $8mm$ lens as near border deviation from working distance is $200 - 140.61 = 59.39mm$, that is bigger than calculated to $40mm$. The far border, in this case, also fulfils the requirements. The same conclusions can be made for a $12mm$ lens. Therefore, both lenses can be used in application.

Table 3.1: Depth of Field for 8mm Focal Length and 200mm Working Distance for Different f-numbers

Parameter	f-number				
	4	8	16	32	64
Near border for DOF [mm] (COF=1 pixel)	189.97	180.90	165.13	140.61	108.41
Far border for DOF [mm] (COF=5 pixels)	271.74	423.73	infinity	infinity	infinity

Table 3.2: Depth of Field for 12mm Focal Length and 300mm Working Distance for Different f-numbers

Parameter	f-number				
	4	8	16	32	64
Near border for DOF [mm] (COF=1 pixel)	298.80	280.27	262.97	234.08	191.91
Far border for DOF [mm] (COF=5 pixel)	364.08	462.97	1013.51	infinity	infinity

3.3.3 Calibration

As has been discussed in Section 2.3 of this thesis, calibration is an integral part of every machine vision application. Robot-hand calibration and hand-eye calibration are left out of the scope as they are integrated procedure in robotics software and is not the objective of this thesis.

After the hand-eye calibration is performed, the external parameters of the model are known. However, for detection of pin coordinates it is critical to understand how to relate 2D pixel image coordinates with real world 3D coordinates. Namely, the camera calibration, described in Section 2.3, has to be performed. The camera lens and the image plane are connected with intrinsic camera parameters, namely, pixel coordinates of principle point (u_0, v_0), scale factors in u and v axes α and β, and skew parameter γ (refer to Equation (2.16)). These parameters are calculated automatically by software and can be applied to every subsequent image if the camera settings are not changed.

The calibration technique used in this thesis is a state of the art of Keyence corporation and can not be discussed in details in this thesis.

According to [40], calibration tool enables the correction of captured images for linear and non-linear distortion caused by lens distortion and camera mounting. The calibration parameters are obtained via a captured image of a known grid. These calibration parameters form the calibration function that allows to generate a corrected image and, at the same time, to recalculate pixel coordinates into real world coordinates. However, the camera is calibrated for the fixed z-coordinate (calibrated distance). Namely, these real coordinates are calculated always for the same fixed calibration distance. This technique is useful when the object is lying on production line that has fixed z-coordinate in space. However, z-coordinate of a pin always deviates from this distance as pins are randomly distributed inside the box. Therefore, the recalculation of z-coordinate should be done for every pin. The mathematical model for such recalculation is presented in Section 4.1.

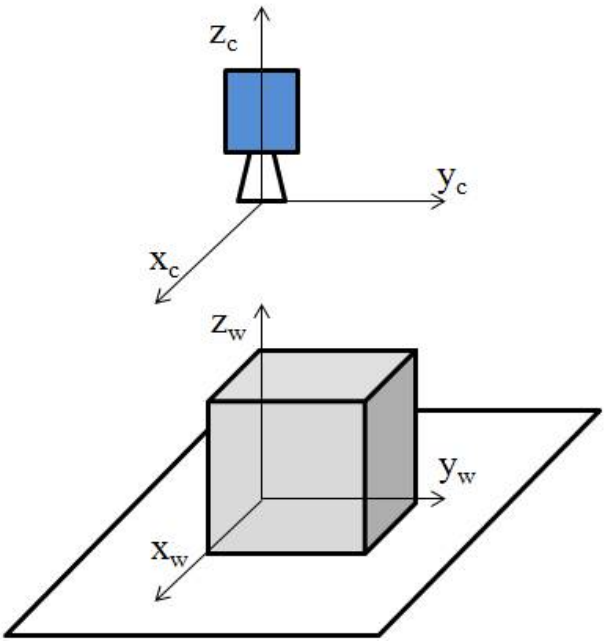

Figure 3.13: Camera vs Real World Coordinate System

To simplify the mathematical models used in this thesis for pin calculation, it is assumed that camera performs only pure translation movements before any subsequent image capture. Namely, camera X_c, Y_c and Z_c axes are parallel or overlapping with X_w, Y_w and Z_w axes of real world coordinates. X and Y axes are chosen to be parallel to virtual lines where walls of the pallet will be placed. This configuration is presented in Figure 3.13. Z_c axis is assumed to perpendicular to the floor. The camera initial position for calibration is assumed to be so that it points vertically down on the center of the calibration grid lying on the floor. The advantage of such a configuration is the simplification of calibration procedure. In case of process change or failure, the pallet can be removed and the system can be easily recalibrated.

3.3.4 Assignment of Pins to Wickets

According to specifications extracted from the process, two pins of one wicket should be taken from the box simultaneously. If two pins of different wickets are placed on the wicket conveyer, the process collapses. Thus, it is critical to understand which pin belongs

to which wicket. Furthermore, it is required to detect the orientation of the wicket as it can be placed on the conveyer only with a long side out as it is shown in Figure 1.2.

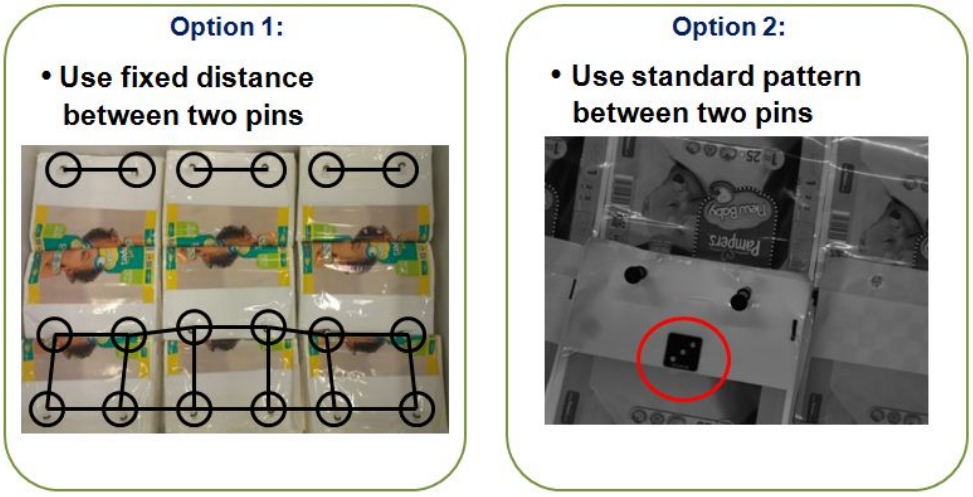

Figure 3.14: Two Options of Alternative Strategy for Assignment of Pins to Wickets

From first observations it is concluded that the edges of wickets are not reliable enough for wicket detection. In some cases white edge of one wicket overlaps with white edge of another wicket. This makes it impossible to see the border. For this reason an alternative strategy is required for pin assignment.

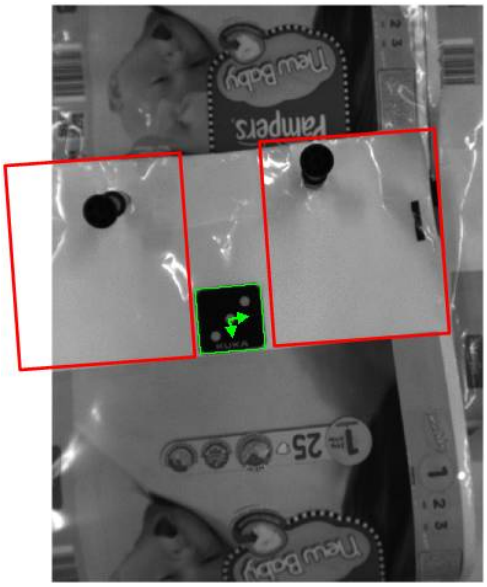

Figure 3.15: Search Region for Pins Search

Two strategies are proposed and tested in this thesis. The first option is to separate pins from the background by using fixed distance between two pins of the same wicket type in order to understand that these two pins belong to the same wicket. However, this strategy is rejected as it requires a lot of programming and assumptions. As has been mentioned in Section 3.1, the bag is not rigid and there are wrinkles on the bag between pins. Additionally, z coordinate of the wicket is not fixed in space. These issues lead to wrong detection of the distance. Moreover, there are wicket types for which it is impossible to use this strategy. As shown in the Figure 3.14, the same distance can be observed for two pins of different wickets.

The second strategy is about usage of a standard mark (pattern) on the wicket between pins as is presented in Figure 3.15. After detection of this pattern the pins can be found

in defined search regions from this pattern. The mathematical model for calculation of search region coordinates and size is provided in Section 4.3.

This strategy is more standard as it can be used for all two hundred different types of wickets. The only difference is that search regions need be defined separately for every wicket type. This strategy also makes it possible to detect wicket orientation if the pattern is not symmetric and is chosen to be oriented so that it points towards the outer edge of the bag. The possibility to add the pattern should be discussed with other departments of the company as it affects the artwork of the bag. In this thesis, it is assumed that such a pattern will be used in future. The pattern is chosen to be assymetric and comparable in size with pins for reliable detection.

As discussed above, the first image is captured so that the FOV is one fourth of a pallet area in x,y-plane, namely 60x40cm . The segmentation of wicket layout into four parts can lead to an issue. Even if the pattern is detected in one image, pins belonging to this pattern can be detected in another image as is shown in Figure 3.16. To solve this problem four images can be stitched together and analyzed as a complete set. The alternative strategy is to use an information about pattern orientation in x,y-plane extracted from the first image. Knowing the wicket type, approximate x_p and y_p coordinate of two pins belonging to this wicket can be calculated from standard dimensions of the wicket that is known priory for every wicket type. Then, moving the camera in x,y-plane, images of two pins can be captured separately, for big wickets, or together. for smaller wickets. As estimation of pin location is too rough, both images should be again taken from the first image distance. This should be done to avoid the situation that pin is again out of camera visibility. This strategy requires one or two additional shots, but allows to avoid image adhesion.

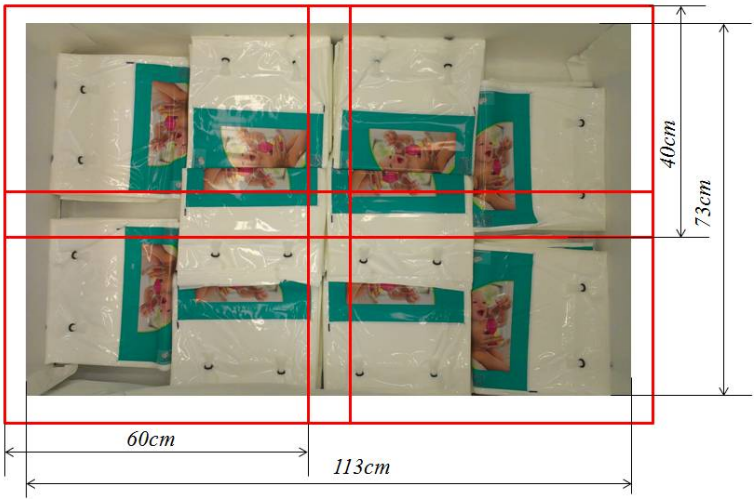

Figure 3.16: Segmentation of the Wicket Layout: Two Pins of the Same Wicket Appear in Different Images

After pins are found on the image and assigned to the same wicket, the calculation of their coordinates should be performed. The Marr's Theory presented in Section 2.4 is used for this purpose. For simplification, it is assumed that camera always points vertically down. The camera set up is shown in Figure 3.13 and the camera always performs pure translation. Namely, Eulerian angles (ϕ, θ, ψ) of camera coordinate system relatively to world coordinate system are $(0, 0, 0)$.

3.3.5 Calculation of Pins

Pins are symmetrical across their central axis and the widest part of the pin is the circular top with diameter of $19.8mm$ (see Figure 3.17). The object is made of rigid material and is always of the same size. The contour of the pin can be used in shape reconstruction. As the circular top is the widest part of the pin, the pin width represents the diameter of the circular pin. The closer the pin is to the camera, the wider the pin is in the image.

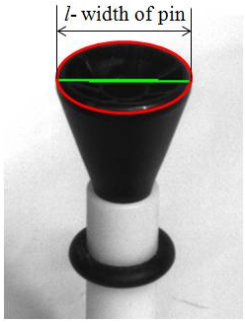

Figure 3.17: Pin Width

Since contours of pins are used for analysis it is necessary to detect edges of the pin and avoid undesirable edges in the image. The wicket is shiny and it creates reflections from the surface of the bag. These reflections can be seen in the image. However, they appear as white lines on grey background, whereas pins appear as black spots on grey background. In order to eliminate reflections, the contrast conversion technique, explained in Section 2.6.1, can be used. In Figure 3.18, the difference between pin analysis, using contrast manipulations and filtering techniques, is shown. From the Figure it can be observed that contract conversion prior to the edge filter allows elimination of undesired reflections from wicket surface.

For detection of pin z_t coordinate special procedure is designed. Objective of this procedure is to use a deviation from known pixel size of projected object on the image plane for deriving deviation in z coordinate from calibration plane. The projected object in current application is the width of top of the pin in pixels. Because pin has a symmetrical structure, pixel width of the pin is formed by a line passing through the center of the circle. Each point of this center line lie on the same z_t coordinate if the center of the pin is directly under the center of camera lens as is shown in the Figure 3.19. It allows to avoid an error from perspective projection. However, if these two centers are shifted the error from perspective projection influences results, because points of the center line of the pin are not lying on the same distance from the calibration plane. The more the centers are shifted towards each other, the bigger is the error.

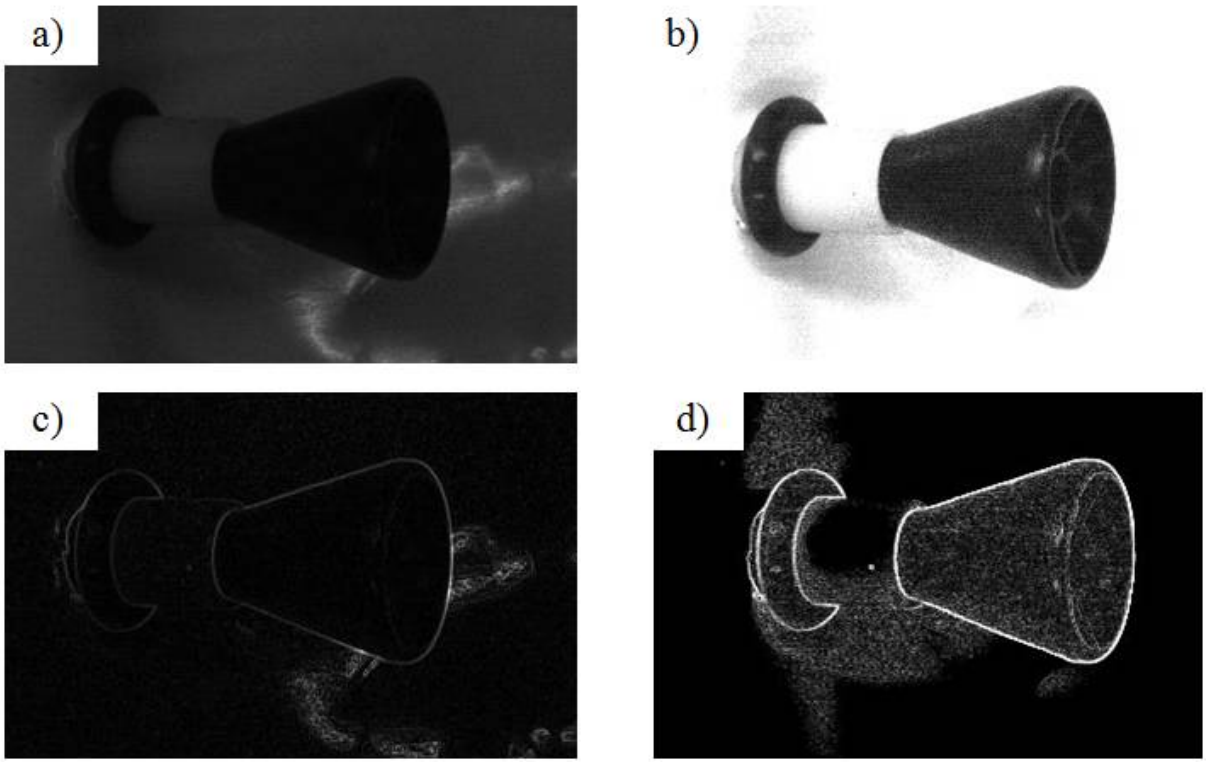

Figure 3.18: Pin Contours Detection under Different Conditions: a) without filters, b) with contract convension, c) with Prewitt filter, d) with contract convension and Prewitt filter

Further, if the size in pixels of the pin is known for the calibrated distance, the deviation in z_t coordinate from calibrated distance can be calculated, from deviation in width of the pin from fixed value in pixels. This technique is named scaling in this thesis and is presented in detail in Section 4.1.
After z_t coordinate of the pin is calculated from scaling it is necessary to recalculate x_t and y_t coordinates. In Section 4.2 the mathematical model for recalculating x and y coordinates of the point in real scene from pixel coordinates of the image using scaling is presented.

3.3.6 Shape Reconstruction of Pins

To reconstruct the shape of the pin, techniques for shape reconstruction, described in Section 2.4, are analyzed and optimized for the current application. It is decided to use the shape from contours approach.
As the top of the pin has a circular shape the projection of the circle can be used in order to reconstruct the 2.5D shape. There is a variety of literature where the circle reconstruction from elliptical projection is discussed, such as [41, 42, 43]. In this thesis,

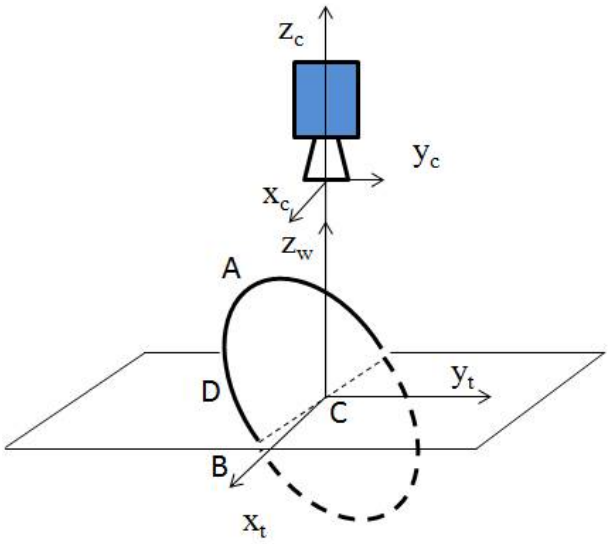

Figure 3.19: Projection of Circular Top of the Pin on X,Y-Plane

the technique is simplified assuming that the camera always points vertically down on the circular top and the center point of the circle is situated in the middle of the image. In the described case, the circle appears as an ellipse on the image and the relation between the width and the length of this ellipse shows θ rotation angle of the central axis of the pin. It is shown in the Figure 3.20. The difficulty of this approach is to assure that the camera center axis is placed on the center of the ellipse. For this purpose the coordinates of the ellipse center should be calculated first. The error in placing the camera on top of the circle affects the angle detection. The conducted tests and respective results are presented in subsequent chapter.

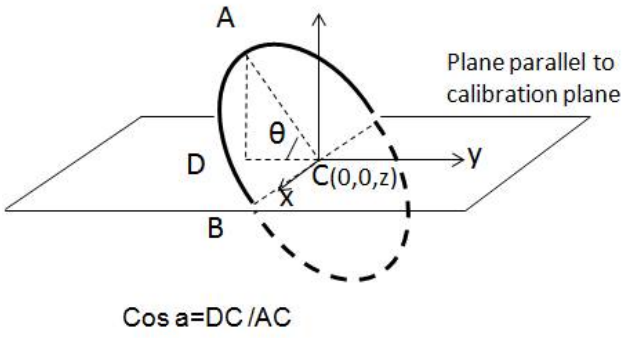

Figure 3.20: Ellipse Projection of the Circular Top

Knowing the coordinate of the ellipse center, θ angle of the pin, and the shape of the pin in the real world, the gripping point can be calculated. The mathematical model for this calculation is presented in Section 4.5. However, this model can be extended using more advanced software for complete 3D shape reconstruction of the pin. For this purpose, precise angles and coordinates of pin can be extracted using a line scanner.

3.3.7 Extraction of Common Gripping Point and Eulerian Angles for Gripper Rotation

From specifications, extracted from gripper design, the common gripping point, where the gripper should be sent in order to grip two pins simultaneously, is the middle point O of the line connecting gripping points A and B of two pins belonging to one wicket(see the Figure 3.8). The middle point of two points can be calculated using formula

$$\begin{cases} x_o = \frac{x_a + x_b}{2} \\ y_o = \frac{y_a + y_b}{2} \\ z_o = \frac{z_a + z_b}{2} \end{cases} \tag{3.6}$$

The Eulerian angle for gripper rotation can be calculated from vector AB, connecting points A and B. The calculation of this angle is not considered in this thesis.

3.3.8 Illumination

In this thesis, ring front light is chosen. Such a light configuration allows to avoid unnecessary shadows and provides light distribution along the optical axis. It is an advantage for discussed application as the objective of the algorithm, discussed above, is to get the pin as close to the image center as possible, in order to decrease the error from perspective distortion.

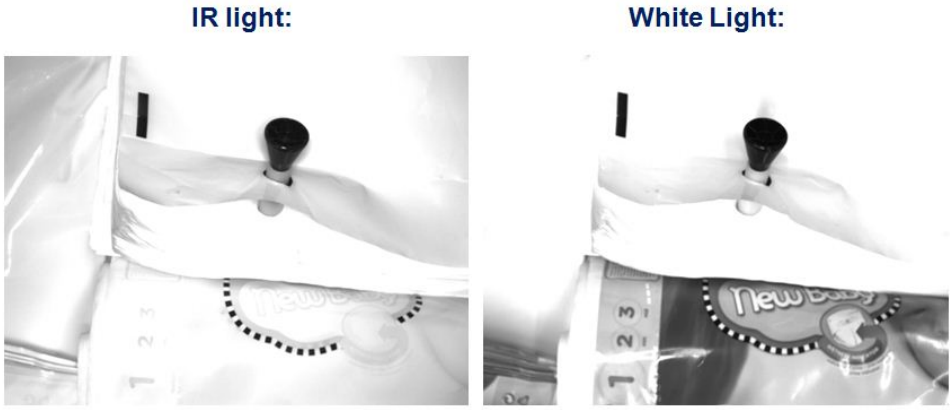

Figure 3.21: Two Strategies for Two Steps Algorithm of Pin Detection

The application is tested with visible white light and invisible infrared light (IR)(as shown in Figure 3.21). These two light configurations have different wavelengths. In case of IR light, it is observed that the color printing on wickets is eliminated. Because of the higher wavelength the light penetrates deeper inside the material under

the print. Using the right wavelength of IR there is a possibility to complettely eliminate the artwork in the image, for particular type and thickness of paint. This technique is known as infrared reflectography and is developed by J.R.J. van Asperen De Boer [44]. This technique requires additional IR filters and amount of light reaching the lens decreases. For almost closed aperture, in the single camera strategy, IR is not attractive as it requires light of high intensity. However, for the strategy with one camera and line scanner, IR light can be used. It gives an opportunity to separate the pin from the background when the colorful artwork on wickets overlaps with the pin as shown in Figure 3.22.

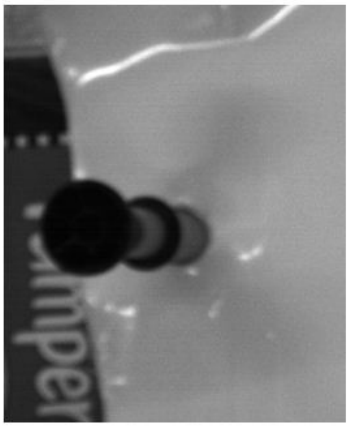

Figure 3.22: Pin Overlaping with Artwork

In Section 3.3.3, it is calculated that for the required application f-number of minimum 32 should be used. In Section 2.5.3, Formula 2.17, connecting illuminance E_s with shutter speed t and aperture, is provided. In Table 3.3, the illuminance is calculated for different ISO values. Values are calculated for shutter speed of 20ms. From the table it can be seen that for higher f-number more light is required as the aperture is closed more.

Table 3.3: Illuminance for 32 and 64 f-number and for Different ISO Values

Parameter	ISO value				
	3200	6400	10000	20000	40000
Illumunance for f-number 32 [lux]	4000	2000	1280	640	320
Illuminance for f-number 64 [lux]	16000	8000	5120	2560	1280

Chapter 4

Mathematical Models

In this chapter, mathematical models, referenced in previous chapter, are summarized and explained in detail. The chapter starts with a model derived for calculation of z coordinate of an object from scaling. Namely, the property that size of the pin changes as the distance between camera lens and pin deviates from the calibration distance is used. In Section 4.2, mathematical equations to recalculate x and y from scaling are presented. In Section 4.3, the model of calculating search region coordinates and size is described. This is needed to find two pins of the same wicket after the pattern is detected.
Next, in Section 4.4, a special procedure to find image distance for pinhole camera model is designed. The camera is translated purely in camera axis direction and the change in size of calibration grid is analyzed. It is an additional procedure that has to be performed once for each new camera settings. After the image distance is found, it can be used to find the distance between the object and center of camera lens if pixel size of projection and real size of object are known. This distance is further applied in the model from Section 4.1 for calculation of z coordinate from scaling.
In the last model, presented in this chapter, the property of the pin circular top, to appear as an ellipse on the image, is used to calculate θ rotation angle of the pin. It is also shown how the angle and coordinates of the center point of pin circular top are used to calculate the coordinates of gripping point.

4.1 Calculation of Z Coordinate

As stated in Section 3.3.6, pins are always of the same dimensions and the change in pixel size of the pin can be used to estimate its z coordinate.
The assumption of pinhole camera model is valid for derivation of mathematical models, as calibration tool of the software performs correction for linear and non-linear distortions

caused by lens imperfections of any captured image.
Figure 4.1 shows the capture of the same object with assumption that the camera is moved purely in camera axis direction. In this application the camera initial position is so that it points vertically down on the calibration grid. Namely, the objective of this procedure is to find an image plane with recalculated z coordinate that is parallel to calibration plane.
From similarity of triangles the following equation can be obtained

$$\begin{cases} \frac{R}{z_1} = \frac{p_1}{i} & (1) \\ \frac{R}{z_2} = \frac{p_2}{i} & (2) \end{cases} \tag{4.1}$$

where R is size of the object in real world in mm, p_1 is the size of the object from calibration distance z_1 in $pixels$, p_2 is the size of the object from observation distance p_2 in $pixels$, and i is image distance.
Dividing (1) by (2) the following can be derived:

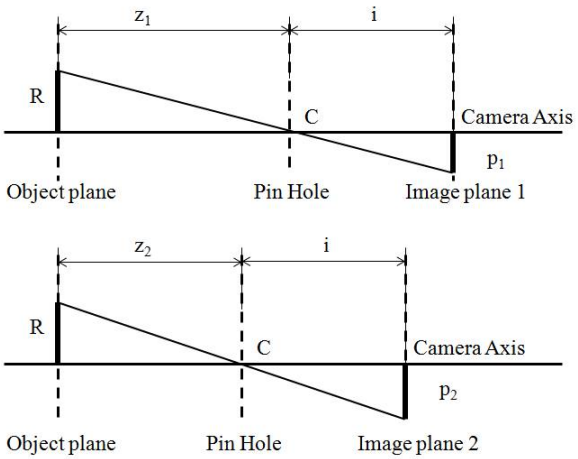

Figure 4.1: Camera Movement in Pure Camera Axis Direction

$$\frac{z_2}{z_1} = \frac{p_1}{p_2}. \tag{4.2}$$

As is shown in Equation (4.2), z coordinate of every new image can be recalculated from scaling, if p_1 and z_1 are known. p_1 can be defined easily from the image of calibration grid captured from calibration distance. If the chess board with cell size $10mm$ is used as calibration grid, p_1 can be calculated multiplying cell size in pixels with 1.98 (the pin biggest diameter is $19.8mm$). z_1 is difficult to estimate as it is almost impossible to measure where the center of the lens is relatively to the calibration object. For this purpose the procedure, explained in Section 4.4 is designed.

4.2 Recalculation of X and Y Coordinates from Scaling

The production vision software automatically recalculates the real world coordinates from pixel coordinates of each point in the image. However, this calculation is done for a fixed calibrated z coordinate. Therefore, after z coordinate of the plane, cutting the pin top (as it is shown in the Figure 3.20) in the middle, is extracted, it is necessary to

recalculate also x and y coordinates for the newly detected z coordinate. The mathematical model of this situation for pinhole camera model is presented in Figure 4.2. It can be seen from the image that the same point D with pixel coordinates (u, v) on the image plane can have different (x, y, z) coordinates in real world (points A and B). It happens, because two points lie on the same ray of light. Assuming that z coordinate is derived from the pin width, as it is shown in Section 4.1, x and y - coordinates can also be recalculated. Figure 4.2 can be simplified to the model shown in Figure 4.3 looking at x and y coordinates separately.

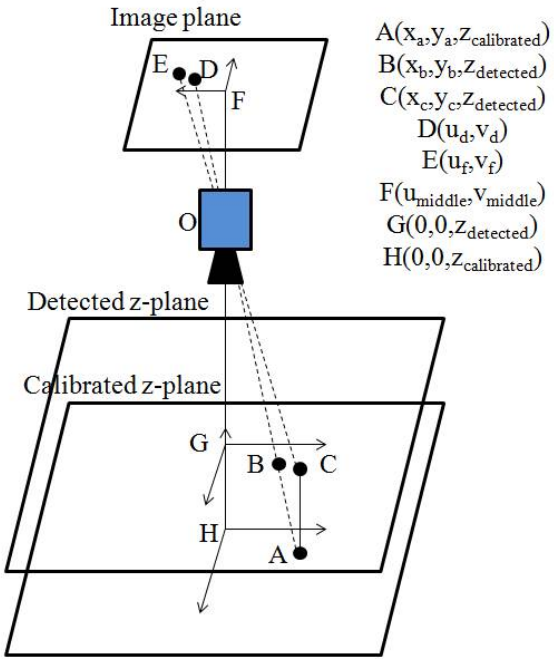

Figure 4.2: Recalculation of X and Y Coordinates for Different Z-Plane

Having the same x and y coordinates, two points C and A with different z coordinates have different projection points on the same image plane. As calibration is done for the distance where point A is located, the pixel coordinates of point C should be recalculated so that they correspond to the coordinates of point A. After this recalculation, the transformation of pixel coordinates into the world coordinates (x, y) can be performed using internal parameters of camera, calculated automatically by software for the calibrated plane.

From the geometrical properties of the model and similarities of triangles AHO with DFO and CGO with EFO the following can be derived:

$$\begin{cases} \frac{OF}{FD} = \frac{OH}{HA} \\ \frac{OF}{FE} = \frac{OG}{GC} \end{cases}, \tag{4.3}$$

where OF is image distance i, OH is $z_{calibrated}$, OG is $z_{detected}$ and $HA = GC$ as x and y coordinates of two points are identical. Therefore,

$$\begin{cases} \frac{i}{FD} = \frac{z_{calibrated}}{HA} & (1) \\ \frac{i}{FE} = \frac{z_{detected}}{GC} & (2) \end{cases} \tag{4.4}$$

From equation (2)

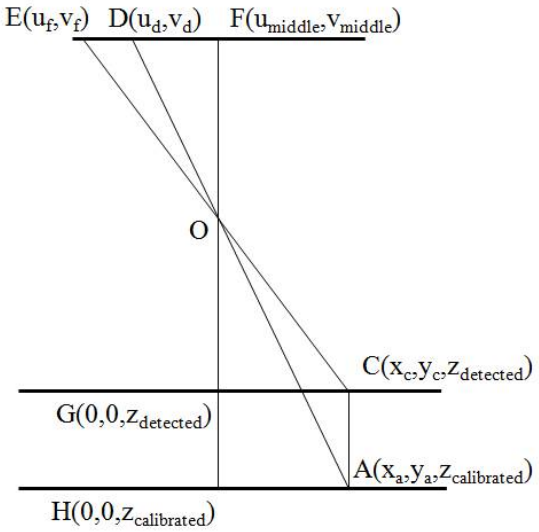

Figure 4.3: Simplified Model for Recalculation of X and Y Coordinates for Planes with Different Z Coordinates

$$i = \frac{z_{detected} FE}{GC}. \tag{4.5}$$

From equations (1) and 4.5

$$FD = \frac{HAi}{z_{calibrated}}, \tag{4.6}$$

$$FD = \frac{HA}{z_{calibrated}} \frac{z_{detected} FE}{GC}, \tag{4.7}$$

$$FD = FE \frac{z_{detected}}{z_{calibrated}}. \tag{4.8}$$

Equation (4.8) is valid in case where point F has pixel coordinates (0,0). However, point F is the middle of the image plane. For example, for image size of 1600x1200 pixels point F has coordinates (800,600). Taking it into account, the final formulas are derived for x and y coordinates in equation as

$$\begin{cases} u_{final} = (u_d - u_{middle}) \frac{z_{calibrated}}{z_{detected}} + u_{middle} \\ v_{final} = (v_d - v_{middle}) \frac{z_{calibrated}}{z_{detected}} + v_{middle} \end{cases}, \tag{4.9}$$

where (u_d, v_d) are detected pixel coordinates of the point, and (u_{middle}, v_{middle}) are pixel coordinates of the center point of image plane. This situation is illustrated in Figure 4.4.

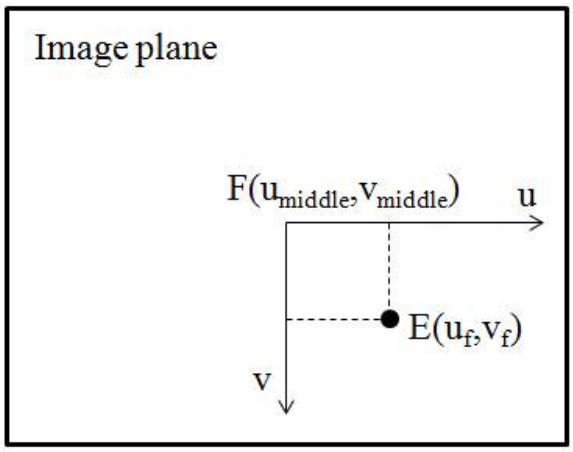

Figure 4.4: Recalculation of X and Y Coordinates Taking Into Account Middle Point of the Image

4.3 Calculation of Search Region

This section is dedicated for the first step of vision algorithm in which pins need to be assigned to the same wicket. After the pattern is detected in the image, the pattern coordinates can be approximately calculated using scaling strategy developed in previous sections of this chapter. Moreover, as the placement of search region is not required with a high precision, the calculation from scaling can be skipped if to assume that the laser point is used to capture the first image of pattern detection. It allows to take the image of the pattern from the distance close to the first capture distance. Therefore, all the values in pixels can be transformed into real world coordinates using calibration parameters extracted for exact first capture distance.

After coordinates of the pattern are calculated and rotation angle ϕ in x,y-plane around Z axis is extracted automatically by software, the search region center point can be calculated using geometrical model, presented in the Figure 4.5. In this calculation, it is assumed that the camera is translated only in x,y-plane and z coordinate is not changing. This is needed to get a field of view that is big enough for pin detection.
The derivation of equation, for finding center points of search region, is divided into several steps. Figure 4.5 can be simplified to Figure 4.6, assuming that pattern is located in the origin of image coordinate system and is not rotated. Namely, ϕ is equal 0^o. In this case

$$\begin{cases} (x_1, y_1) = (a, -b) \\ (x_2, y_2) = (-a, -b) \end{cases} \tag{4.10}$$

where (x_1, y_1) are coordinates of the center of first search region, (x_2, y_2) are coordinates of the center of second search region, a and b are distances from the center of the pattern

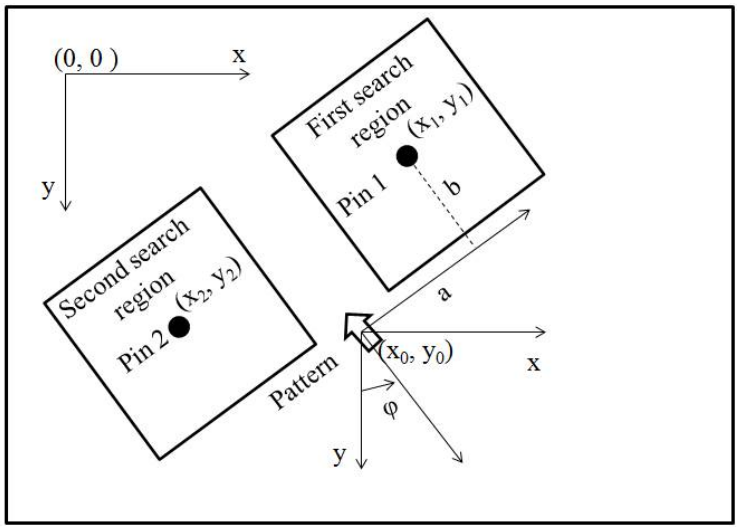

Figure 4.5: Simplified Model for Recalculation of X and Y Coordinates for Different Z-Plane

to the center of search regions along X and Y axes. a and b values depend on the wicket type.

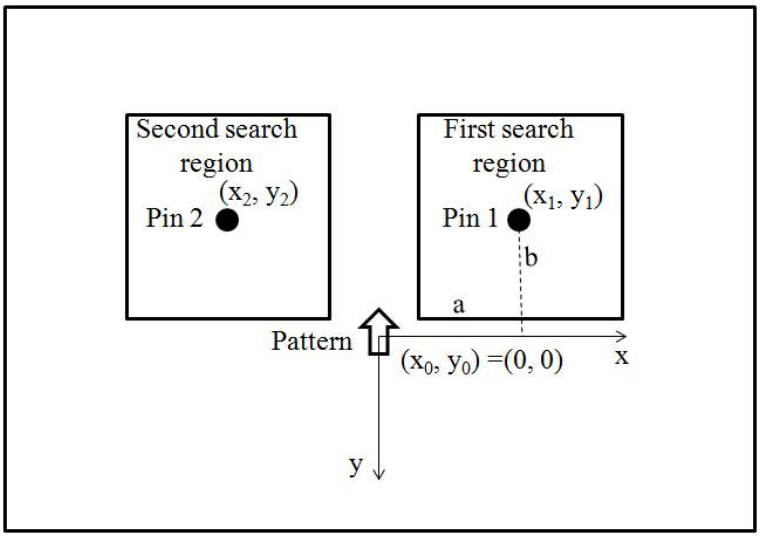

Figure 4.6: Simplified Model for Recalculation of X and Y Coordinates for Different Z-Plane with ϕ is Equal to 0^o

The next step is the rotation of the pattern and pins so that ϕ angle is not equal to 0^o. This is shown in the Figure 4.7. Using Formula 2.13 from Section 2.3.1 for counterclockwise rotation, coordinates of the center of first search region after rotation can be modified to equation

$$\begin{cases} x_1' = x_1\cos(\phi) + y_1\sin(\phi) = a\cos(\phi) - b\sin(\phi) \\ y_1' = y_1\cos(\phi) - x_1\sin(\phi) = -b\cos(\phi) - a\sin(\phi) \end{cases}, \tag{4.11}$$

where (x_1', y_1') are coordinates of the center of first search region after rotation, and for the second search region

$$\begin{cases} x_2' = x_2\cos(\phi) + y_2\sin(\phi) = -a\cos(\phi) - b\sin(\phi) \\ y_2' = y_2\cos(\phi) - x_2\sin(\phi) = -b\cos(\phi) + a\sin(\phi) \end{cases}, \tag{4.12}$$

where (x_2', y_2') are coordinates of the center of second search region after rotation.

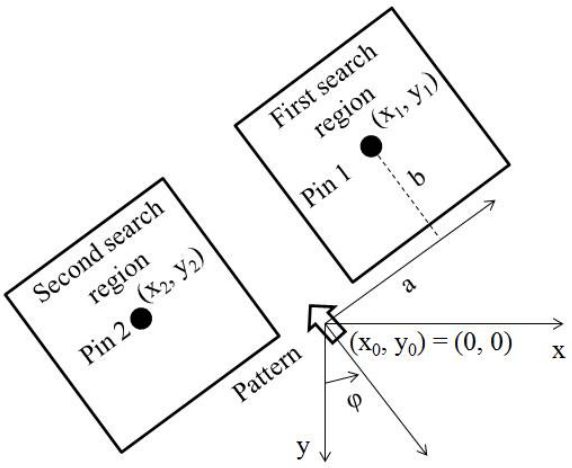

Figure 4.7: Simplified Model for Recalculation of X and Y Coordinates for Different Z-Plane ϕ not Equal to 0^o

The last step is modification of calculated in Equations (4.11) and (4.12) coordinates taking into account that pattern center is not located in the origin of image coordinate system. Therefore,

$$\begin{cases} x_1'' = x_0 + a\cos(\phi) - b\sin(\phi) \\ y_1'' = y_0 - b\cos(\phi) - a\sin(\phi) \end{cases}, \tag{4.13}$$

where (x_1'', y_1'') are coordinates of the center of first search region after rotation and translation, and for the second search region

$$\begin{cases} x_2'' = x_0 - a\cos(\phi) - b\sin(\phi) \\ y_2'' = y_0 - b\cos(\phi) + a\sin(\phi) \end{cases}, \tag{4.14}$$

where (x_2'', y_2'') are coordinates of the center of second search region after rotation and translation.

4.4 Image Distance From Motion

In Equation (4.2) it is shown that z coordinate of every new image can be recalculated from scaling if z coordinate for the calibrated distance is known. An additional procedure is required to estimate the relation between the center of the lens and the calibration object. The objective of this procedure is to translate the camera purely in camera axis direction and observe how the projection of the same object on image plane changes its size. The mathematical model of such movement is presented in Figure 4.8.

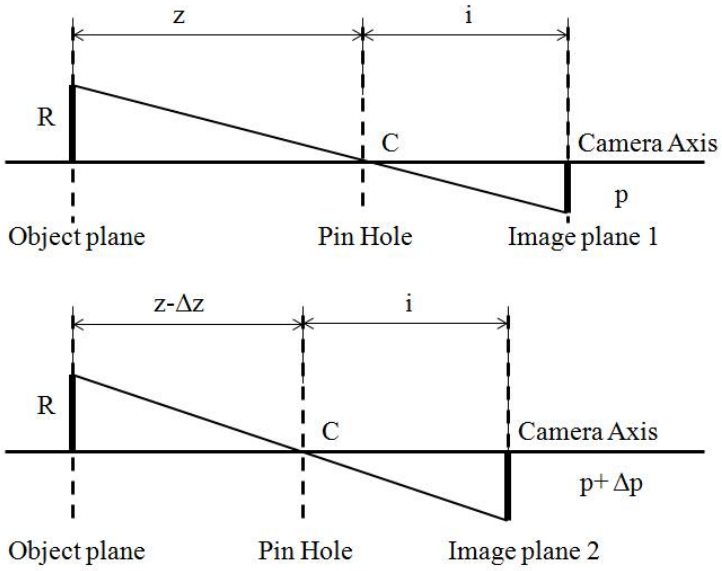

Figure 4.8: Pure Translation of Camera in Camera Axis

Several assumptions have to be made in order to make this model valid. First of all, in order to avoid undesired distortions from perspective projection, the camera should point vertically down on the object under observation and this object has to be flat and located at the middle of the image. As this procedure has to be performed once for any new camera settings, the calibration grid located on the floor can be used. The center cell of the chess board is selected as an object under observation, and the change of its size in pixels is analyzed in order to extract image distance.
From the distance, for which camera is focused, the camera is translated up and down by amount of $\triangle z$ and images are captured for each new z coordinate. Results are combined and image distance i is calculated. From image distance i and pixel size of the object, taken from any distance z, this distance z can be calculated with Formula (4.16(1)).
From geometrical properties of the model in Figure 4.8

$$\begin{cases} \frac{R}{z} = \frac{p}{i} & (1) \\ \frac{R}{z-\triangle z} = \frac{p+\triangle p}{i} & (2) \end{cases} \tag{4.15}$$

where R is the size in mm of an object under observation in real world, z is the z coordinate of the object relative to the center of camera lens, i is image distance, p is size in $pixels$ of projection of the object under observation if the object is at z coordinate, $\triangle z$ is translation of camera in Z axis direction, $\triangle p$ is the change in pixel size of the projection of the object if the camera is translated by $\triangle z$.
Or

$$\begin{cases} Ri = pz & (1) \\ Ri = (p + \triangle p)(z - \triangle z) & (2) \end{cases} \tag{4.16}$$

Equating left sides of Equations (4.16(1)) and (4.16(2))

$$pz = pz + z \triangle p - p \triangle z - \triangle p \triangle z. \tag{4.17}$$

Extracting i from Equation (4.15(1)) and z from Equation (4.17)

$$\begin{cases} i = \frac{pz}{R} & (1) \\ z = \frac{\triangle z(p+\triangle p)}{\triangle p} & (2) \end{cases} \tag{4.18}$$

Substituting z in (4.18(1)) by (4.18(2))

$$i = \frac{p \triangle z(p + \triangle p)}{r \triangle p}. \tag{4.19}$$

Image distance i, calculated from Formula (4.19), can be applied for further calculations, if camera settings (aperture and focus) are not changed. z coordinate can be recalculated for every object with known real world size R. Pixel size of the object p should be formed by points lying on the same z coordinate, in order to minimize an error from perspective projection. However, if the object is not flat, perspective projection introduces an error in the system that should be tested.

4.5 Calculation of Gripping Point

Target of the mathematical model, presented in this section, is to calculate coordinates of gripping point (x_g, y_g, z_g) from coordinates of the center point of circular top (x_t, y_t, z_t). For this purpose additional information about the structure of the pin and rotation angle in Z axis θ is needed.

According to gripper specifications, gripping point defined from gripper design is the point on the central axis of the pin between rubber ring and the head of the pin, as is presented in the Figure 4.9. From the structure of the pin distance from the center of circular top

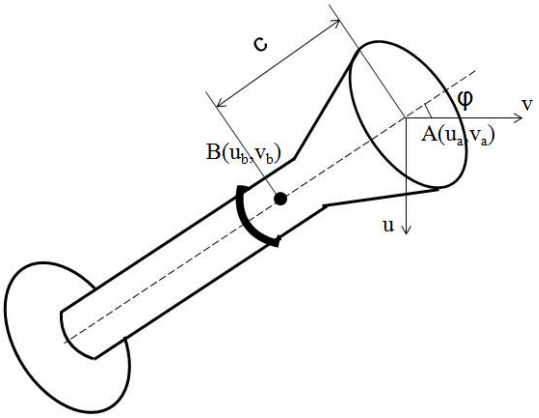

Figure 4.9: Projection of the Pin on Image Plane

to the gripping point is $l = 25mm$. In mathematical model, presented in Figure 4.10, an ideal case is assumed where the center of the circular top is located at the center of the image. In this case, undesired influence from perspective distortion can be eliminated. However, because of calculation errors in z_t coordinate from the image taken from first capture distance, shift of the point from the center of the image can occur. It introduces an error in 2,5D reconstruction and this error is tested in Chapter 5. If final error is smaller than the error, allowed by the gripper design, the following model is valid for current application.

The circle appears as an ellipse on the image and relation between the width R_1 and the length R_2 of this ellipse is rotation angle θ in Z axis of the pin. As AC is the radius of the circular top and R_2 is projection influenced only from translation of camera, but not from rotation in angle θ, the following Equation can be derived:

$$AC = R_2. \tag{4.20}$$

Consequently

$$\cos\theta = \frac{R_1}{R_2}. \tag{4.21}$$

Figure 4.11 shows projection of the pin on the plane that is formed by central axis of the pin and Z_w axis of world coordinate system. h is projection of distance l from the center of circular top to the gripping point. Angle AO_1D is equal to angle BO_1O_2 and this follows from equation

$$\begin{cases} angle(AO_1D) = 90^o - angle(DO_1O_2) \\ angle(BO_1O_2) = 90^o - angle(DO_1O_2). \end{cases} \tag{4.22}$$

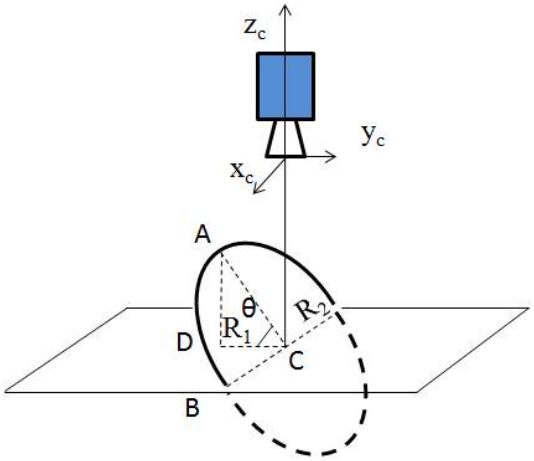

Figure 4.10: Projection of Circular Top of the Pin on X,Y-Plane

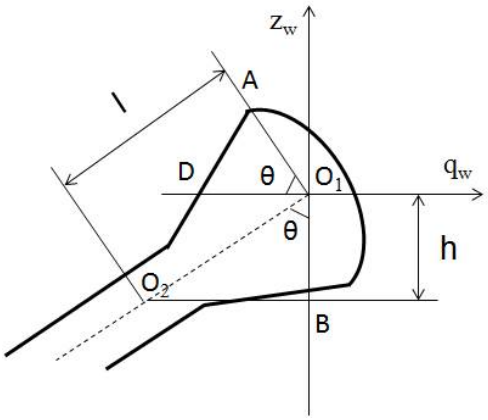

Figure 4.11: Projection of the Pin on Plane Formed by Z-axis of World Coordinate System and Center Axis of the Pin

Therefore,

$$h = l\cos\theta = l\frac{R_1}{R_2}. \tag{4.23}$$

z_g coordinate of gripping point can be calculated from z_t coordinate of the center point of the circular top with formula

$$z_g = z_t - h. \tag{4.24}$$

assuming that z_g is always lower than z_t. This assumption is valid as it follows from the process specifications. Knowing z_g and pixel coordinates (u_b, v_b) of gripping point (as can be seen in Figure 4.10 that shows the projection of pin on the image plane) (x_g, y_g) coordinates can be calculated using scaling.

In this model gripping point for available gripper design is calculated. However, the same logic can be applied in calculation of any point on the pin or near the pin.

Chapter 5

Tests and Results

In this chapter, tests and results are discussed. Tests are designed using design of experiments described in [45, 46] and analyzed using statistical tools. The models developed in Chapter 3 are used to generate the results. In Section 5.4, setup of experiments is described and the procedure for obtaining valid and statistically reliable results is explained. After the procedure for experiments is designed, results of mathematical models derived in Chapter 3 are checked using this procedure. Tests are divided into five parts. In the first part, the correctness of procedure, explained in Section 4.4, is tested. Image distance is found and results are compared to the results obtained for the same camera settings by Camera Calibration Toolbox for Matlab [15]. In the second subsection, the first step of the algorithm developed in this thesis is validated. The error in coordinate detection of pin from the first capture distance is tested in the third test. The results of this error are used to design experiments for the fourth and fifth test procedures. The fourth test procedure is about correctness of coordinate detection of the center point of circular top if the image is captured from the second distance. The fifth subsection provides results obtained from calculation of gripping point using the algorithm developed in Section 3.3.6.
In the last sections of this chapter, results are summarized and suggestions about error reduction are highlighted and evaluated. Results of total error are compared with specifications provided by gripper design. The analysis allows to conclude whether the vision strategy with single camera fulfills the requirements for pin detection or the strategy with camera and line scanner should be used.

5.1 Design of Experiments

According to [45], the purpose of tests and experiments is to study the performance of processes and systems. Figure 5.1 shows an example of a

model of such a process or a system. The output of the process is transformed from the input. This transformation is influenced by observable response variables that can be controllable and uncontrollable for purpose of tests. In this thesis, the input of the system is an image and the output is specification that vision system has to extract from this image and provide to the robot for reliable gripping. The controllable variables, decided as most influential to results, are position of lens center relative to the pin, light conditions and type of wicket used. Among uncontrollable variables are unexpected shadows, undesirable reflections from the surface of a bag and unstable environmental conditions. This is summarized in the Figure 5.2. Design of experiment is closely related to statistical analysis.

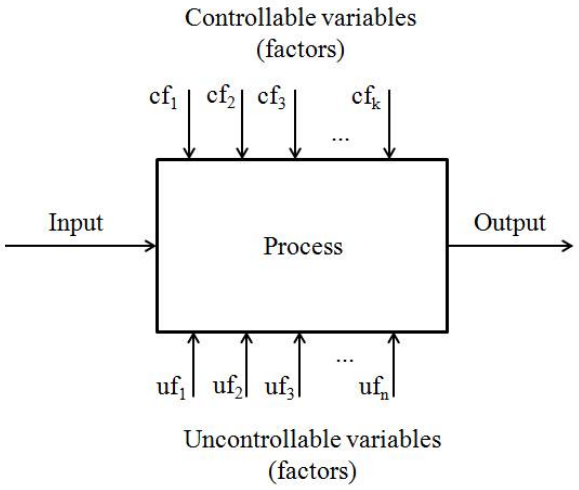

Figure 5.1: General Model of Process or System [46]

To draw meaningful conclusions appropriate data should be collected and analyzed by statistical methods.

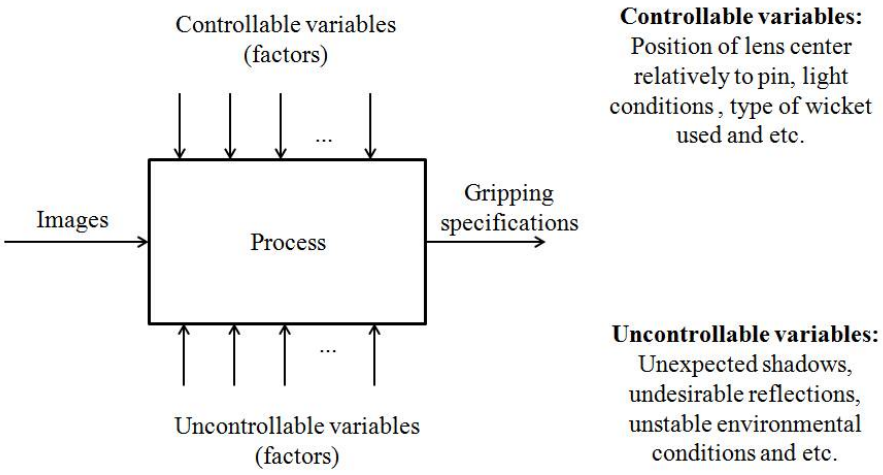

Figure 5.2: Model of Process under Experiment

Therefore, as statistical methods require observations to be independently distributed random variables, randomization in experiments is needed. In this thesis, experiments are designed taking into account randomization for samples. However, due to unavailability of materials and time, randomization is created to a certain extent. For test design the

following sequential activities are performed (for more imformation refer to [45] and [46]):

1. Problem study and declaration
2. Choice of the controllable and uncontrollable variable
3. Selection of factors, ranks, and ranges influencing variables
4. Experimental design
5. Experimental Performance
6. Statistical data analysis
7. Conclusions and recommendations

5.2 Evaluation of Mathematical Models

5.2.1 First Test: Image Distance From Motion

First test is performed in order to check the model that is developed in Section 4.4. The objective of this test is to calculate the image distance i from a set of images of the same calibration grid. Therefore, the input of the test is a set of images, the output is the calculated image distance. The model of this test is shown in the Figure 5.3.

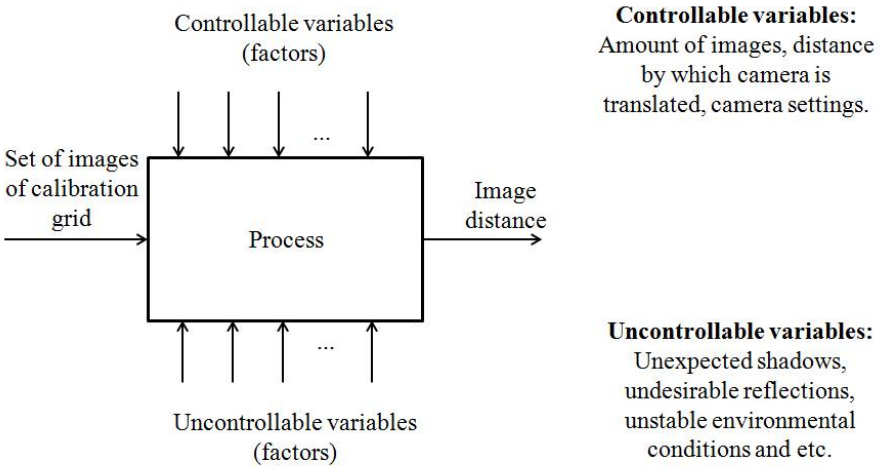

Figure 5.3: Model of Process for First Test

Every new image is taken by translating the camera purely in camera axis direction that is parallel to Z_w axis of the world coordinate system. Amount of images is chosen

to provide statistically reliable results with 95% confidence. Consequently, controllable variables influencing the process and output, are amount of images used for image analysis, and the distance by which the camera is translated between each subsequent image capture. Moreover, controllable variables are camera settings, namely aperture and the distance for which camera is focused. Uncontrollable variables are changes in environmental conditions, reflections, shadows, etc.

The camera is focused near the second capture distance for all tests. The calibration chess board is placed under the camera and coordinate axes of calibration board are parallel to axes of camera coordinate system as can be seen in Figure 4.8. The camera is translated purely in camera axis (or Z_c) direction with steps of $\triangle z = 10mm$. For every new z coordinate an image is captured and the size in pixels of the chess cell, situated in the middle of the image, is measured. In total thirty images are captured with the camera moving up from the second capture distance. These images are combined in different combinations. For Table 5.1 the variable that is changed is $\triangle z$ and amount of images is fixed. For Table 5.2 amount of images is varying variable and $\triangle z$ is fixed at $100mm$. The image distance for every combination is calculated using Formula (4.19). Results are compared with results from Calibration Toolbox for Matlab [15] presented in Figure 5.4. It can be concluded that calculation of image distance, done with the developed strategy, gives the similar results to the Matlab Toolbox, already developed and validated by scientists [15].

From Table 5.1 it can be seen that with the increase in the distance $\triangle z$ between two neighboring captures the error in calculation of i decreases. The same conclusion can be made from Table 5.2 for the amount of the images used. Increasing the number of images this error can be reduced.

Table 5.1: Image Distance for the Fixed Number of Images and Changing $\triangle z$

Number of Images	$\triangle z$ [mm]	Image Distance i [pixel] (with confidence 95%)			$\pm deviation$ [pixel]
10	10	2816.9	2866.1	2915.3	49.2
10	50	2856.5	2865.4	2874.3	8.9
10	100	2863.8	2866.7	2869.6	2.9
10	150	2864.3	2866.8	2869.2	2.5
10	200	2863.4	2865.1	2866.9	1.7

Table 5.2: Image Distance for the Fixed $\triangle z$ and Changing Number of Images

Number of Images	$\triangle z$ [mm]	Image Distance i [pixel] (with confidence 95%)			$\pm deviation$ [pixel]
5	100	2860.6	2864.8	2869.0	4.2
10	100	2863.8	2866.7	2869.6	2.9
15	100	2864.4	2866.5	2868.6	2.1
20	100	2863.3	2865.3	2867.2	2.0

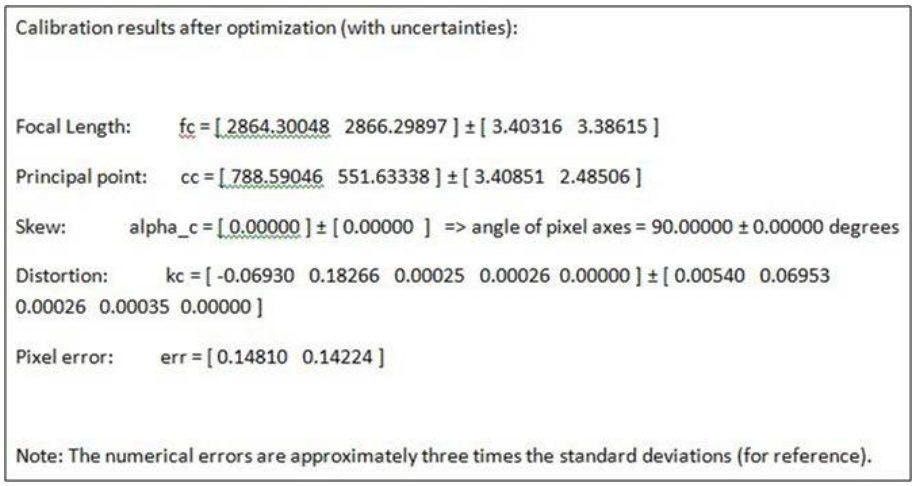

Figure 5.4: Image Distance Calculated by Camera Calibration Toolbox for Matlab

The test is repeated for changed camera settings and is presented in Tables 5.3 and 5.4. Results are also compared with results from Matlab Calibration Tool which showed approximately the same image distance of $2834 pixels$. Summarizing all the results, it is decided to use combination of $\triangle z$ equal to $200 mm$ and number of images equal to ten for obtaining image distance in calculations, as this provides a relatively small deviation for 95% confidence interval.

Table 5.3: Image Distance for the Fixed Number of Images and Changing $\triangle z$

Number of Images	$\triangle z$ [mm]	Image Distance i [pixel] (with confidence 95%)			$\pm deviation$ [pixel]
10	10	2819.8	2838.3	2856.8	18.5
10	50	2830.1	2835.0	2839.8	4.9
10	100	2832.4	2834.5	2836.6	2.1
10	150	2832.1	2833.6	2835.0	1.5
10	200	2831.8	2833.0	2834.2	1.2

Table 5.4: Image Distance for the Fixed $\triangle z$ and Changing Number of Images

Number of Images	$\triangle z$ [mm]	Image Distance i [pixel] (with confidence 95%)			$\pm deviation$ [pixel]
5	100	2831.0	2835.7	2840.5	4.7
10	100	2832.4	2834.5	2836.6	2.1
15	100	2831.7	2833.5	2835.3	1.8
20	100	2831.6	2833.1	2834.7	1.5

Image distance, derived from the procedure, is used in calculation of distance between the center point of camera lens and calibration grid. For this purpose Formula (4.2) is used. Moreover, tool attached to the robot was sent down to the calculated z coordinate. Visually, the z coordinate, calculated by software, and the z coordinate of the calibration grid in real scene, matched with each other as shown in Figure 5.5. Consequently, it is assumed that the image distance, calculated with the concept highlighted in this section, can be used in calculations for further tests.

Figure 5.5: Robot Tool at Calculated Z Coordinate

5.2.2 Second Test: Reliability Study

In the second test, reliability studies are done for assignment of pins to wickets and for pin detection. The assignment of pins is dependent on pattern detection and correctness of pin detections, namely pin separation from the background. Pin detection is dependent on reliable finding of pin contours on the image.

Test is divided into three steps. First step targets to check the reliability of pattern detection. The controllable variables in this case are the distance from camera to the pattern and the type of pattern used. Uncontrollable variables are shadows, environmental conditions, and wrinkles on the wickets that can create undesirable reflections, as is shown in Figure 5.6. The pattern detection is evaluated with the sticker, attached on the wicket, and the mark, printed on the surface of the wicket. Randomization in shadows and wrinkles is done manually due to unavailability of samples created by transportation. For this purpose wicket is moved and folded in different ways.

The test is done for two hundred fifty samples. Samples include different combinations

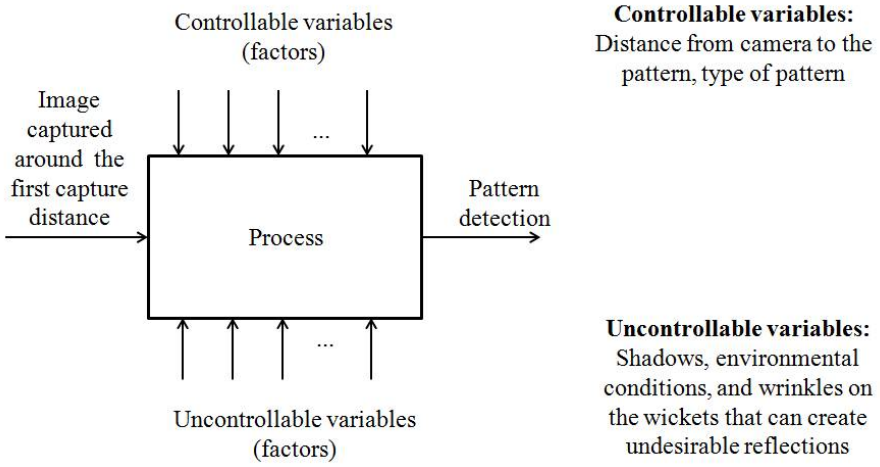

Figure 5.6: Model of Process for First Step of Second Test

of variables listed in Figure 5.6. Changes in uncontrollable variables is created manually. The software allows detection of patterns that are twice smaller or twice bigger than the pattern chosen to be a standard. It follows that the capture distance should not deviate from the standard distance more than twice. The standard pattern is decided to be the pattern taken from the first capture distance. It has to be flat without wrinkles and visible reflections. Assuming that the capture distance for all two hundred fifty samples is deviating from standard within permitted range, the sticker is detected for all samples. Moreover, if no extreme wrinkles are created on the wicket, the printed pattern is also detected in all the cases. However, the detection of printed pattern should be tested in real conditions, when wrinkles are formed during transportation.
For the second step, the reliability of pin separation from background is tested. Controllable and uncontrollable variables for this step are summarized in Figure 5.7. Due to unavailability of all range of wicket types it is decided to use the most colorful artwork available. It creates the conditions when the pin can overlap with background and separation of pin can be influenced. The pin produced with a black head is not available yet for tests. Therefore, it is decided to color the pin with opaque and shiny paint. It should also be taken into account that the paint layer slightly increases the width of the pin and adds an additional error, which can not be measured, to the total error.
For the second step of the test the same samples are used as for the first step. It is reasonable as test results are independent from each other. Issues assumed in approach description in Section 3.3.4 are observed in this test as well, and are summarized in the Table 5.5.

The issue when the pin is not visible in the same image as the pattern is observed in 13.2% of the cases. Therefore, solutions described Section 3.3.4 , namely, sticking the four

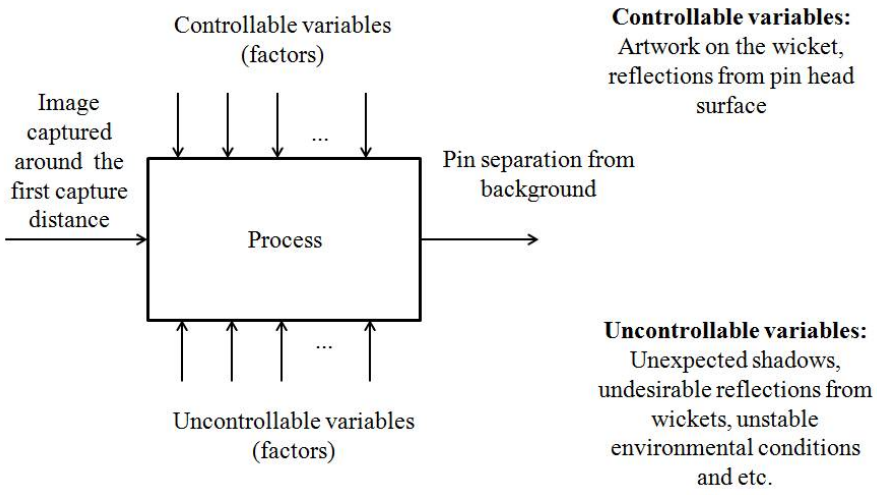

Figure 5.7: Model of Process for Second Step of Second Test

images of the complete layout, taken from first capture distance, or alternative strategy are the necessary requirement.

The issue when pin overlaps with artwork is observed in 6% of cases. This problem will not be observed for less colorful artwork. Even though the pin overlaps with artwork, the pin is detected. However, the contours of pin are detected wrongly in this case as it is shown in the Figure 5.8.

Figure 5.8: Detection of Pin Contours without filters, with Prewitt filter and with Binary Filter

Table 5.5: Results of Second Step of Second Test

Observed Issue	Total Amount of Samples	Number of Samples with Issue	% of Issue Samples
Pin is not visible on the same image as the pattern	250	33	13.2
Pin overlaps with artwork	250	15	6

In the last step of second test reliable contour finding is checked. Controllable variable for this part of the test is the distance between the camera and the pin, because contour will appear more blurred as the camera is moved away from the focused distance. Therefore, the images for the test are obtained from different distances. Moreover, the contour detection can be influenced by the problem described in the second step of this test, namely, pin overlapping with artwork of the wicket. The difficulty with this issue is that even though the pin is detected, the contours of the pin can be detected wrong and the scaling from the width of the pin will give an error. This error is not easy to detect. If the error is too big, it can be realized by looking at the scaling of the second pin belonging to the same wicket or to the scaling of the pattern. If deviation is more than $200mm$ in z coordinate, it is obvious that the detection is wrong, as two pins of the same wicket can not be that far away from each other in z coordinate. However, if the error is in this range, it is difficult to be detected. Therefore, this issue can not be ignored and requires to be solved. As is mentioned in Section 3.3.8 the artwork can be eliminated using the IR light and filter. For this case the depth of focus problem appears, as less light is passing through the lens and two cameras or camera and a line scanner are needed. Moreover, as IR eliminates the artwork, only the sticker on the wicket can be used as a pattern.

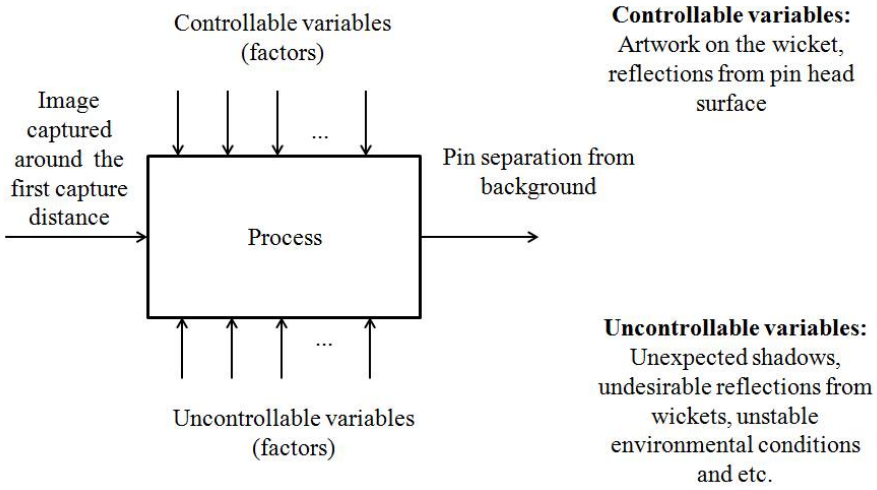

Figure 5.9: Model of Process for Third Step of Second Test

Uncontrollable variables are reflections from the surface of the bags as they create undesirable edges on the image. However, these reflections are eliminated using contrast conversion technique described in Section 3.3.5. In addition, uncontrollable factors are environment conditions and unexpected shadows. The model for current test is summarized in the Figure 5.9. Using all the techniques, described above, and assuming that the pin is not overlapping with artwork, contour detection is tested. The pin detection is checked for all the images used in next subsections and the contours of pins are found re-

liably for all the images tested. However, it should be taken into account that for capture distances near the focused distance the edge of the contour is sharp, and with the increasing deviation from the focused distance the edge of the contour appears more blurred. This means, that, for the steps of algorithm where the images are obtained from the first capture distance that is far away from under the focus (second) capture distance, the sensitivity of edge detection functions should be increased.

5.2.3 Third Test: Calculation of Pin Coordinates from First Capture Distance

After two pins are assigned to the wicket and contours of pins are found for the first capture distance, approximate first coordinates of pins should be calculated. This is in order to be able to place the camera on top of the pin circular top so that the center of the circular top is located directly under the camera center. This is a necessary requirement that is assumed in mathematical models for the calculation of pin and gripping point coordinates. The detection of pin coordinates from the first capture distance has an error due to low resolution of image, perspective distortion, etc. This means that, when the camera captures the image from the second image distance, the Z_c axis of camera is not pointing at the center point of circular top, but is shifted by some amount. Consequently, in the image used in analysis for the second step of the algorithm, the center point appears shifted from the middle of the image. This introduces the error in subsequent calculations. Due to perspective distortion, the more the point is shifted, the bigger is the error. Therefore, the target of the algorithm is to get the pin circular top as close to the center of the image as possible.
The objective of the third test is to understand the error in calculation of pins coordinates from the first capture distance. This error is analyzed and used for design of subsequent tests.
The controllable variables for this test are the relative coordinates of the pin and the camera in the space. Uncontrollable variables are reflections from the wicket, shadows, and changing environmental conditions.
Conditions changing relative positions of pins and camera are created by moving the camera relatively to the stable wicket. If the wicket pins are located at the same position in the space and the camera is translated by a known distance from one image to another, the analysis of images should show the same movement. The wicket with pins is placed under the camera approximately first capture distance away from the camera. Then, the camera is first translated purely along Z_c axis around the first capture distance by $20mm$. The images are taken for each new z_c coordinate of the camera and the approximate

coordinates of the pins in the image are found using the strategies described in Sections 4.1 and 4.2. The image analysis should show the same x_p, y_p coordinates of the pin and a change of $20mm$ in z_p coordinate.
In the second part of this test the camera is translated by a known amount purely in X_c direction so that pin location is changed from left edge of the image to the right edge.
Next, the camera is translated along X_c and Y_c axes simultaneously. Consequently, the image analysis should give results of the same z_p coordinate of pins and changing x_p and y_p coordinate. The camera is also translated in all the axes X_c, Y_c and Z_c at the same time and results, calculated from images, are compared to the known situation.
The last part of the test targets to check shadowing randomization. The camera and the wicket location in space are stable and the shadow conditions are changed, by adding additional light from different sides and by creating different shadows. In this part of the test it is observed that if the shadow is very intense the image appears too dark for reliable image analysis. Therefore, in real production conditions such a situation should be avoided.
The results of this test are collected in Table 5.7. The coordinates of the center point of pin projection are found for each image. The images are divided into sets, corresponding to the steps of this test as described above. Namely, the images with the same x_c, y_c and changing z_c coordinates form the first set. The images with the same y_c, z_c and changing x_c coordinates form the second set, and so on. As every image potentially has an error in calculation of every pin coordinate, the mean of coordinates is assumed to be the closest to reality and the allowable deviation around this value is calculated. The allowable deviation is the possible error in detection, and is assumed to be $\pm 4.5\sigma$.
However, this statistic strategy can be applied only if a coordinate stays the same during the test. If the coordinate is changing by a known distance from initial value, the mean calculation of z_p is not applicable. The amount the camera is translated should be found from images and compared with $20mm$ real world result. This is shown in Table 5.6. It can be seen from the table that the maximum deviation of $\triangle z_p$ coordinate from the real world result is $25.723mm$ and the mean of $\triangle z_p$-$\triangle z_c$ is $5.965mm$. That means that in this case the error, that can be observed in z_p, theoretically is $5.965 \cdot 4.5 = 26.841mm$. For x_p coordinate the standard deviation from mean is $4.1165mm$ and, consequently, $4.5\sigma = 18.5243mm$. Therefore, the shift of camera X_c axis relative to X_t axis of pin circular top coordinate system should be in range $\pm 18.5243mm$. In Y axis direction the same calculation gives results of $4.1672 \cdot 4.5 = 18.7524mm$.

Table 5.6: Analysis of Detected Pin Coordinate for the Case of Pure Camera Translation in Z_c axis

Parameter	coordinates [mm]			$\triangle z_p$ [mm]	$\triangle z_p$-$\triangle z_c$ [mm]
	x_p	y_p	z_p		
Detected pin coordinates	42.293	68.666	-260.172	41.242	21.242
	45.307	73.184	-218.930	23.413	3.413
	46.068	73.919	-195.517	41.986	21.986
	43.084	70.384	-197.503	45.723	25.723
	47.160	75.133	-151.780	20.338	0.338
	47.353	75.682	-132.118	23.593	3.593
	48.261	76.448	-108.525	26.612	6.621
	49.856	77.162	-81.904	24.512	4.512
	51.085	77.935	-57.392	25.812	5.812
	52.735	78.669	-31.580	20.806	0.806
	53.294	79.221	-10.774	22.578	2.578
	53.073	79.759	6.648	22.698	2.698
	52.737	80.283	23.950	20.029	0.029
	53.106	80.922	43.979	20.935	0.935
	53.756	81.448	64.914	20.472	0.472
	54.255	82.063	85.386	20.175	0.175
	54.595	82.709	105.211	20.468	0.468
	55.136	83.203	125.679		
Mean	50.175	77.599			5.965
SD	4.1165	4.1672			

Results, presented in Table 5.7, are the summarized results of all the steps of third test, using the same logic in analysis as is used in Table 5.6. The complete test is repeated three times. This amount is assumed to be reasonable due to time and resource constraints. In each test the real world picture is changed by changing the position of wicket and rotating the wicket and pins in the scene. The results of the biggest errors observed for all three tests are summarized in Table 5.8.

Table 5.7: Analysis of Detected Pin Coordinates: Third Test Results

$x_p$coordinate [mm]						
Changing variable	Pin 1			Pin 2		
	4.5σ [mm]	$4.5 \triangle x$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle x$ [mm]	Maximum error observed [mm]
z_c	6.060		2.306	18.524		7.882
x_c		4.269	2.658		3.034	1.999
x_c and y_c		3.369	2.441		7.380	4.235
x_c, y_c, z_c		4.365	2.848		7.851	4.192
shadow	1.7482		0.706	2.6091		0.823
$y_p$coordinate [mm]						
Changing variable	Pin 1			Pin 2		
	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]
z_c	14.503		5.083	18.753		8.933
x_c	0.746		0.403	0.669		0.208
x_c and y_c		8.812	6.337		4.766	2.651
x_c, y_c, and z_c		8.787		4.497	5.873	4.318
shadow	7.332		3.179	9.063		2.771
$z_p$coordinate [mm]						
Changing variable	Pin 1			Pin 2		
	4.5σ [mm]	$4.5 \triangle z$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle z$ [mm]	Maximum error observed [mm]
z_c		3.787	3.165		26.841	25.723
x_c	24.467		6.842	19.159		8.881
x_c and y_c	27.951		10.667	37.999		14.076
x_c, y_c, and z_c		32.043	22.924		25.394	9.567
shadow	6.261		4.183	22.852		6.468
where $4.5 \triangle x = 4.5 * (\triangle x_p - \triangle x_c)$, $4.5 \triangle y = 4.5 * (\triangle y_p - \triangle y_c)$ $4.5 \triangle z = 4.5 * (\triangle z_p - \triangle z_c)$						

Table 5.8: Summary of Results for Three Runs of the Third Test

Parameter	max(4.5σ) [mm]	Maximum error observed [mm]
x_p	18.524	7.882
y_p	18.752	8.933
z_p	40.650	25.723

5.2.4 Fourth Test: Calculation of Pin Coordinates from Second Capture Distance

The test presented in this section is designed to verify the possible error in calculation of coordinates of the center point of pin circular top (x_t, y_t, z_t) when the image is taken from the second capture distance. The same procedure is performed, as in the third test, with all the steps of the test repeated. The difference is that due to better resolution, the $\triangle x_c$, $\triangle y_c$, and $\triangle z_c$ are chosen to be smaller, specifically, $5mm$. The test is performed with a $12mm$ lens. Additionally, the reliability in finding the center point of the circular top in the image is tested. The top appears as an ellipse when it is projected on the image plane. The ellipse finding is tested for different rotation angles and it is concluded that ellipse can be found for θ angles up to 80°. For higher angles the ellipse can not be detected by software and an additional strategy is required. For example, the edge center point of the top can be taken to calculate x_t and y_t coordinates. Such a strategy results in a slight error as the edge point is located very close to the center of the circle in x,y-plane, however, it is not the same point. For z_t calculation the width of the pin is detectable regardless of the ellipse detection.

Results of first try of this test are presented in Table 5.9. Results are divided into two parts: without error and with error correction. In part without error correction, the camera is translated in x,y-plane so that pin can be projected even near the border of image plane. Results are analyzed taking into account images on which pin appears from one edge of the image plane to the opposite edge. For the second part, with error correction, the maximum possible shift of the pin center from the image center, calculated in the third test, is taken into account. Consequently, presented results with error correction are summarized from the images, on which the camera is translated around the center of the pin circular top approximately within the range, derived from the third test $(\pm x_c, \pm y_c) = (\pm 18.524mm, \pm 18.752mm)$. Translation along z_c axis is assumed to be in range $\pm 50mm$ for both parts of the test. The value is chosen to be close to the result of $\pm 40.650mm$ obtained for error in z_p.

Table 5.9: Analysis of Detected Center Point of the Circular Top Coordinates: Fourth Test Results for $theta = 45°$

$x_p$coordinate [mm]						
Changing variable	Without error correction			With error correction		
	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]
z_c	1.162		0.439	1.162		0.439
x_c		0.213	0.098		0.172	0.072
x_c and y_c		0.290	0.201		0.211	0.091
x_c, y_c, z_c		0.199	0.166		0.237	0.166
shadow	0.105		0.056	0.105		0.056
$y_p$coordinate [mm]						
Changing variable	Without error correction			With error correction		
	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]
z_c	1.728		0.638	1.728		0.638
x_c	0.313		0.130	0.239		0.082
x_c and y_c		0.318	0.307		0.275	0.149
x_c, y_c, and z_c		0.302		0.167	0.341	0.167
shadow	0.127		0.046	0.127		0.046
$z_p$coordinate [mm]						
Changing variable	Without error correction			With error correction		
	4.5σ [mm]	$4.5 \triangle z$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle z$ [mm]	Maximum error observed [mm]
z_c		0.8379	0.777		0.8379	0.777
x_c	1.216		0.547	0.385		0.256
x_c and y_c	1.852		1.000	1.156		0.378
x_c, y_c, and z_c		1.387	0.955		1.889	0.955
shadow	0.656		0.211	0.656		0.211
where $4.5 \triangle x = 4.5 * (\triangle x_p - \triangle x_c)$, $4.5 \triangle y = 4.5 * (\triangle y_p - \triangle y_c)$ $4.5 \triangle z = 4.5 * (\triangle z_p - \triangle z_c)$						

From Table 5.9 it can be observed that the main impact on error in detection of x_p

and y_p has deviation of z_p coordinate from the capture distance, that follows from wrong detection of the pin width. As it is shown in Section 3.3.2, even the slight error in contours detection can lead to a critical error in detection of z_p coordinate. Moreover, the software calibration tool automatically corrects lens distortions and, consequently, reduces the impact of changing x_p and y_p coordinates. This can be seen also from the test results. Impact of changing z_p is lower than impact of changing x_p and y_p. Therefore, the biggest portion of error in this case, mostly comes from the wrong calculation of z coordinate.
Comparing results with and without error correction it can be concluded that, with error correction, resulted error is generally smaller. It proves the statement from theoretical description to the algorithm that the shift of the circular top from the center of the image introduces an error in the system due to perspective distortion, and the more is the shift, the bigger is the error.
The test is repeated several times for different θ angles starting from 0° and up to 67°. Results of the biggest errors observed are summarized in Table 5.10. From these results follows that the maximum 4.5σ observed is $2.548mm$. Therefore, the center point of the pin circular top can be detected with minimum precision in (x_t, y_t, z_t) coordinates of approximately ($1.162mm$, $1.728mm$, $2.548mm$) without error correction, and ($1.162mm$, $1.728mm$, $1.707mm$) with error correction.

Table 5.10: Summary of Results of the Fourth Test for Changing θ Angle

θ	Parameter	max(4.5σ)[mm]	
		Without error correction	With error correction
0°	x_p	1.086	1.086
	y_p	1.677	1.677
	z_p	1.795	1.707
28°	x_p	0.945	0.945
	y_p	1.661	1.661
	z_p	2.132	1.608
45°	x_p	1.162	1.162
	y_p	1.728	1.728
	z_p	1.852	1.160
67°	x_p	0.958	0.958
	y_p	1.7203	1.7203
	z_p	2.548	0.933

5.2.5 Fifth Test: Calculation of Gripping Point Coordinates from Second Capture Distance

In the fifth test the gripping point coordinates are calculated from the mathematical model derived in Section 4.4. The same test procedure and images are used as in previous subsection. The results for the first run of fifth test are presented in Tables 5.11 and 5.12 for rotation angle of pin θ approximately equal to 45^o. From Table 5.11, where analysis of calculated coordinates of gripping point is presented, it can be concluded that the error, previously calculated for the fourth test, in general increases. It happens, because $cos\theta$ influences the calculation of gripping point coordinates (see Formulas (4.23) and (4.24)). Therefore, an additional error of wrong angle detection increases the total error.
Table 5.12 shows the analysis of precision in angle finding. As it is predicted in approach description, the error in angle estimation grows with a shift of pin circular top center from the center of the image. For example, the maximum deviation in detected angle from the angle, detected when the camera lens center is located directly on top of the pin, is 10.658^o. In this case the pin is located near the edge of the image plane. The more the pin is shifted to the middle of the image, the less is the error due to perspective projection, as is shown in Figure 5.10. For example, if the pin is $20mm$ away from the center, the ellipse analysis shows a deviation of 4.609^o from undistorted value. As the error in θ decreases when the circular top appears closer to the middle of the image, the error in calculation of gripping point coordinates also decreases. This explains the fact that results of the test with error correction are in general lower in error than results without error correction.

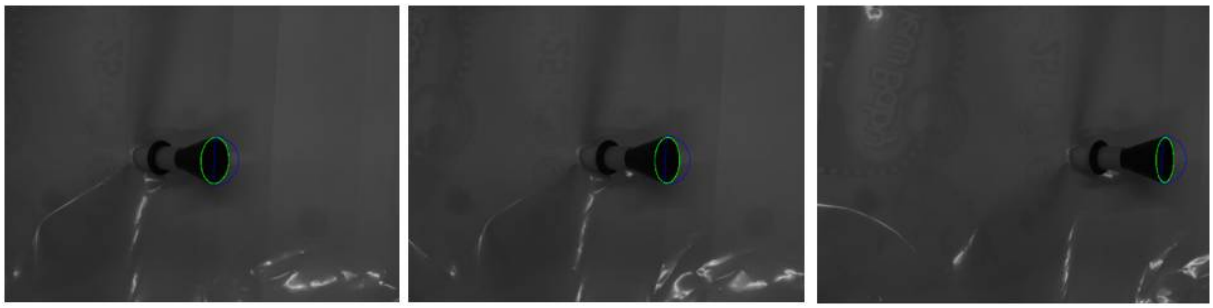

Figure 5.10: The Growth of Influence of Perspective Projection on θ Angle Detection with the Shift of Pin from the Middle of the Image Plane

Table 5.11: Analysis of Detected Gripping Point Coordinates: Fifth Test Results for $theta = 45°$

$x_p$coordinate [mm]						
Changing variable	Without error correction			With error correction		
	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]
z_c	1.120		0.531	1.120		0.531
x_c		1.388	0.480		1.382	0.437
x_c and y_c		1.114	0.627		1.008	0.543
x_c, y_c, z_c		1.310	0.493		1.074	0.421
shadow	0.670		0.341	0.7670		0.341

$y_p$coordinate [mm]						
Changing variable	Without error correction			With error correction		
	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle y$ [mm]	Maximum error observed [mm]
z_c	1.386		0.522	1.386		0.522
x_c	1.623		0.732	0.788		0.334
x_c and y_c		1.511	0.469		1.485	0.469
x_c, y_c, and z_c		1.274		0.626	1.443	0.626
shadow	0.458		0.102	0.458		0.102

$z_p$coordinate [mm]						
Changing variable	Without error correction			With error correction		
	4.5σ [mm]	$4.5 \triangle z$ [mm]	Maximum error observed [mm]	4.5σ [mm]	$4.5 \triangle z$ [mm]	Maximum error observed [mm]
z_c		0.804	0.720		0.804	0.720
x_c	9.185		3.489	4.451		1.696
x_c and y_c	5.126		2.182	3.027		1.631
x_c, y_c, and z_c		1.798	0.993		1.729	0.993
shadow	1.134		0.765	1.134		0.765

where $4.5 \triangle x = 4.5 * (\triangle x_p - \triangle x_c)$, $4.5 \triangle y = 4.5 * (\triangle y_p - \triangle y_c)$
$4.5 \triangle z = 4.5 * (\triangle z_p - \triangle z_c)$

Table 5.12: Analysis of an Error in Detected θ Angle: Fifth Test Results for $theta = 45°$

$x_p\theta[°]$				
Changing variable	Without error correction		With error correction	
	SD [°]	Maximum error observed [°]	SD [°]	Maximum error observed [°]
z_c	0.226	0.457	0.226	0.457
x_c	6.444	10.658	3.233	4.609
x_c and y_c	4.277	7.413	2.536	3.966
x_c, y_c, z_c	4.370	7.133	2.597	3.256
shadow	0.267	0.478	0.267	0.478

The test is repeated several times for different θ angles starting from 0^o and up to 67^o. Results of biggest errors observed in every run are summarized in Table 5.13. From these results it follows that the maximum 4.5σ is growing for higher θ values. It follows from Equation (4.23). From this equation h and, consequently, z_g is dependent on $cos\theta$. Cosine value falls faster for smaller angles. As $cos\theta$ is in the numerator of the formula, the same amount of error in angle detection for lower angles impacts calculation of coordinates less than for higher angles.

Analysis of results without error correction shows that for angles θ more than 30° 4.5σ is higher than, acceptable from gripping design, $5mm$ deviation in z_g coordinate. In results with error correction, if θ is less than 50° calculated coordinates of gripping point are acceptable. However, if the angle is higher, the error falls outside the border of permission. It signals the need in strategy change or additional strategy to be added for higher θ angles. Possible solutions for improvement of the strategy are proposed in the next section.

Results, presented in this subsection for gripping point calculation, are valid only for the gripping point assumed for the current gripper design. The gripping point in this case is situated on pin center axis and is located $l = 25mm$ away from the center of the circular top as is shown in Figure 4.11. The error grows in proportion to l as it follows from equation 4.23. Therefore, the change of gripping point location will decrease or increase the possible error in gripping point calculation.

Table 5.13: Summary of Results of the Fourth Test for Changing θ Angle

θ	Changing variable	Without error correction		Without error correction	
		max error in θ[°]	max(4.5σ)[mm]	max error in θ[°]	max(4.5σ) [mm]
0°		7.215		4.222	
	x_p		1.086		1.086
	y_p		1.677		1.677
	z_p		2.093		1.814
28°		11.021		4.450	
	x_p		3.571		1.285
	y_p		3.670		3.401
	z_p		4.473		3.438
45°		10.658		4.609	
	x_p		1.388		1.382
	y_p		1.623		1.485
	z_p		9.185		4.451
67°		11.983		5.010	
	x_p		2.126		2.065
	y_p		1.589		1.589
	z_p		10.990		5.949

5.3 Proposed Solutions for Error Improvement

As it is calculated from tests, designed for evaluation of mathematical models proposed in this thesis, the error in calculation of gripping point coordinates for rotation angle θ higher than 50^o is not acceptable for the proposed gripping design. However, there are cases when pin is rotated in scene more that 50^o from Z_w axis. In this situation, an alternative strategy is required in order to detect coordinates of pins with sufficient correctness. There are different solutions to solve this problem. One possible solution is to use line scanner together with camera. This approach is described in Section 3.3. After the image from the first distance is captured, pins are assigned to the same wicket and first coordinates of pins are calculated, the line scanner can be sent to the calculated positions of pin. The precision of this calculation is found to be $(\pm 18.524mm, \pm 18.752mm, \pm 40.650mm)$ minimum for the image focus with five pixels circle of confusion. However, even this precision is enough to be able to send a line scanner to the place where the pin is approximately

located and get more precise coordinates. The advantage of this strategy is that solution is standard regardless of the angles and situation in the scene. Moreover, the precision of coordinates calculation increases up to μm.
The alternative strategy is to capture one more image of the pin on which the center of circular top is closer to the image center. The error in calculation of the center point of a pin circular top is calculated from the fourth test and is ($1.162mm$, $1.728mm$, $2.548mm$) for 4.5σ. Therefore, the position of camera can be corrected accordingly, so that the center point of the pin shifts to the center point of the image plane, and one more image can be captured from newly calculated position. Consequently, an error due to perspective distortion will be further decreased. This procedure can be done several times in order to get higher precision in coordinate calculation. Theoretically, the lowest error, that can be obtained using this logic, is the same error that is calculated for the images, when randomization with shadowing and different environment conditions is created, and camera and pin are static in the scene. The results of this situation are summarized in Table 5.14. According to the results, the errors observed in coordinates detection are less than the acceptable error of $5mm$. Consequently, the described step procedure is relevant for the application of pin gripping.

Table 5.14: Analysis of an Error in Detected θ Angle and Coordinates of Pin with Shadow as Changing Variable

	4.5σ deviation in detected			
θ [°]	θ [°]	x_g [mm]	y_g [mm]	z_g [mm]
0°	4.327	1.086	1.677	1.685
28°	0.694	0.204	2.250	1.096
45°	1.202	1.134	0.458	0.670
67°	1.311	0.720	0.163	1.639

However, increasing the number of images required in application can be done to only a certain extent. As it is stated in process description the time between two subsequent placements of wicket on wicket conveyer can reach up to $1m$. This time is enough to capture and analyze more than twenty images, depending on translation amount, speed of robot and time needed for capture and analysis. However, while choosing the amount of images for image analysis the future speed up of line should be taken into account. The minimum number of images, needed for different strategies described in this thesis, is listed in Table 5.15.

Table 5.15: Minimum Amount of Images Needed for Image Analysis for Different Strategies Described in the Thesis

Strategy	Captures Required for Analysis	Total Amount
Camera and Line Scanner	Pattern detection (1) Pin Assignment (1 or 2,3 dependent on wicket size) Coordinates of pins with line scan (2,3 or 4,5)	3 or 5
Single Camera (with one image for each pin)	Pattern detection (1) Pin Assignment (1 or 2,3 dependent on wicket size) Coordinates of pins from one image (2,3 or 4,5)	3 or 5
Single Camera (step strategy)	Pattern detection (1) Pin Assignment (1 or 2,3 dependent on wicket size) Coordinates of pins from several images (2-5 or more)	min 5

5.4 Results Summary

In this chapter, tests to evaluate the algorithm, designed for calculation of coordinates required for gripping of wicket pins, are presented. All the tests are designed using design of experiments theory briefly discussed in Section 5.1.
In Section 5.4, experimental setup and results are shown. Tests are divided into five steps. In first test, the model presented in Section4.4 was tested. The image distance i from a set of images of the same calibration grid was found, and results were compared with scientifically reliable results of Camera Calibration Toolbox for Matlab. The test was repeated three times for different camera settings. In all runs, results calculated with the designed procedure coincide with results obtained from Matlab Toolbox. Moreover, the robot tool was sent to calculated, using found image distance, z coordinate, and the coordinate on which the tool stopped matched the coordinate of calibration grid in real scene. The calculated image distance was used in further tests.
The second test was designed in order to check the reliability of concepts designed for assignment of pins to wicket and contour finding of pins. The reliability of pattern detection, both sticker attached to the wicket and printed mark, was checked. For test two hundred fifty samples were used. It was found that if no extreme wrinkles were created the pattern was found on every wicket. However, as randomization of wrinkles and reflections from wickets was created manually, the test should be repeated in future with representative samples created by transportation. Moreover, more samples should be used for more reliable statistics.

It was also discovered that separation of pin from background can create difficulties for most colorful wicket artworks. The overlap of pin with artwork on the image was observed for 6% of cases, that can become an issue in detection of pin contours. Additionally, the issue, when the pin is not visible on the same image as the pattern, was found. However, in this case, the alternative strategy was proposed. When the pattern and two pins are not visible on the same image, for example, when wicket is too big in size, two additional images can be take for two pins separately. The details of this strategy are described in Section 3.3.

In the last step of second test, contour finding of pin is checked. It is proved that the influence of wicket reflections can be avoided in analysis, using contract conversion technique.

The third test is designed to calculate the error in detection of pins coordinates if the images are captured around the first capture distance. This procedure is needed in order to understand how close it is possible to place the camera center axis on top of the pin circular top for capture from the the second capture distance. The shift in this placement introduces an error in subsequent models due to perspective projection. It was calculated that the error in this case is approximately $(\pm 18.524mm, \pm 18.752mm, \pm 40.650mm)$ in (x_p, y_p, z_p) for 4.5σ.

In the fourth test, the error in coordinates detection of center point of pin circular top if the images are captured around second capture distance was checked. Results were calculated for two cases: with error and without error correction. Namely, taking into account or ignoring results from the third test. It was found that the minimum precision found in four runs of this test in (x_g, y_g, z_g) coordinates is approximately ($1.162mm$, $1.728mm$, $2.548mm$) without error correction, and ($1.162mm$, $1.728mm$, $1.707mm$) with error correction. These values can be used in any gripping design. For example, if the gripping point of the pin would be the center point of the circular top, the strategy, developed in this thesis, would assure that this point can be found with approximate precision of ($\pm 1.162mm$, $\pm 1.728mm$, $\pm 1.707mm$).

The fifth test is designed to calculate the possible error in finding the gripping point, specified for current gripping configuration. The gripping point is located on pin center axis $25mm$ away from center point of the circular top. Results, found in this test are valid only for highlighted specifications. The mathematical model for gripping point calculation should be modified in case if the gripper design will be changed.

The θ angle of pin was calculated for each image in this test. According to test results the error in θ increases as the pin moves away from the center of the image. This error influences the final coordinate calculation of gripping point. Moreover, the error in extraction of gripping point coordinates was calculated with and without error correction. Results show that this error is bigger for high θ angles than is accepted from gripping

design. Therefore, in last section of this chapter the solution for decreasing the error in calculation was proposed and tested. With the strategy, proposed as alternative for high θ angles, the error decreases to ($1.162mm$, $1.728mm$, $2.548mm$) for 4.5σ. This value is acceptable for the current gripping design.

Chapter 6

Conclusions

6.1 Summary

In this thesis, the machine vision strategy to detect the coordinates of wicket pins has been developed in order to provide them to higher-level tasks.
First, the current process on production line is studied. Process restrictions and possible issues are analyzed and taken into account for mechanical and vision system design of a gripper. For the proposed gripper design specifications, that vision algorithm should extract from captured images, are defined.
The vision strategy is divided into two steps. First step is about assignment of two pins to the same wicket. As the gripper has to take two pins simultaneously, place and fix them both on the wicket conveyer, this step can not be neglected in this application. The wrong assignment of pins to wicket can lead to collapse of the process.
The second step is dedicated for calculation of the gripping point specifications listed for current gripper configuration. The special step procedure was designed to get an error in calculation of gripping point coordinates that is still acceptable for the available gripper design. Moreover, several alternative strategies were proposed and tested. After the option analysis and first observation tests, the most promising of them were chosen for further analysis.
Mathematical models developed in this thesis were tested. Analysis of results statistically proves that the concept, developed in this thesis, can be applied for wicket pins detection. It was shown that increasing the amount of images for step procedure analysis the error in (x_g, y_g, z_g) coordinates can be approximately decreased up to ($1.162mm$, $1.728mm$, $2.548mm$) without error correction, and to ($1.162mm$, $1.728mm$, $1.707mm$) with error correction for 4.5σ statistics. This value is acceptable for current gripper design. Moreover, the strategy to extract more precise coordinates with line scanner is discussed for

the case the gripper design is changed.
Process restrictions and possible issues, that were listed in theoretical description to developed algorithm, were observed during tests. Reliability study was made to understand the importance of them for application. Possible solutions to decrease the amount of rejects due to these issues was discussed and tested. It was shown that the there is a need for further testing due to unavailability of representative amount of samples.
Mathematical models derived in this thesis, as well as the step procedure, designed to decrease an error due to perspective projection, can be considered as universal for similar production applications. Namely, this strategy can be used for applications when the object under observation has known dimensions and structure.

6.2 Outlook

Suggestions for further research that can be an extension of this thesis are presented below:

1. In this thesis, the gripping point calculation was checked for the strategy where a single camera is used in application. The error, calculated in tests, is smaller than the error acceptable by the available gripper design. However, it is assumed that even bigger precision can be achieved by using a combination of the camera with a line scanner. For this strategy, the error in obtaining approximate pin coordinates, extracted from the first capture distance, should be recalculated. This is needed as the camera in this case will be focused for the first capture distance, but not for the second, as is assumed in this thesis. Consequently, the edges of pin will appear sharper and deviation in detected width of the pin will be lesser. It can influence the type of line scanner needed for the second step of the algorithm. For example, in this thesis an error of $(\pm 18.524mm, \pm 18.752mm, \pm 40.650mm)$ in (x_p, y_p, z_p) for 4.5σ is calculated. Therefore, the line scanner should be chosen such that it covers this specified range. A smaller range will allow to choose a smaller line scanner.

2. In this thesis, the coordinates of two separate gripping points were calculated for two pins belonging to the same wicket. The model should be further extended to the calculation of common gripping point and gripper rotation angles, as is shown in Figure 3.8. The model to find the Eulerian angles from the vector connecting two gripping points of pins should be derived.

3. In this thesis it was assumed that the camera always performs pure translation between subsequent shots. The model can be extended to a model where the images

are captured with a camera that is not looking vertically down. In this case, the strategy to analyze circular top projection should be changed or modified.

4. In this thesis, for the model, calculating gripping point coordinates, it is assumed that the camera axis coincides with the Z_t axis that comes from the center of pin circular top and is parallel to Z_c axis. This allows to decrease the errors due to perspective distortion. For the further research, an extension of this model to a model in which the highlighted axes are not overlapping can be done.

As, due to samples unavailability reliability tests could not be performed fully, suggestions for further testing of the algorithm developed in this thesis are summarized below:

1. Pin detection should be tested with pins that are not painted, but molded with black heads.

2. Test on pin assignment should be repeated for different types of wickets. Controllable variables for this test should be size of the wicket and type of the artwork, from the most dark and colorful to the most light and bright. Most important uncontrollable variables are transportation conditions.

3. The gripper design should be tested in real conditions and changed accordingly. Vision strategy should be modified to the specifications as required by the final gripper design.

Bibliography

[1] Dongming Zhao, Songtao L.: A 3D image processing method for manufacturing process automation. In: *Computers in Industry* 56 (2005), p. 975–985

[2] C. Demant, B. Streicher-Abel ; Waszkewitz, P.: *Industrial image processing. Visual qualite control in manufacturing.* Springer Berlin Heidelberg, 1999

[3] Woehler, Christian: *3D computer vision. Efficient methods and applications.* Springer-Verlag Berlin Heidelberg, 2009

[4] Reinhard Klette, Karsten S. ; Koschan, Andreas: *Computer Vision. Three-dimensional Data from Images.* Springer-Verlag Singapore Pte. Ltd., 1998

[5] Luhmann, T.: *Nahbereichsphotogrammetrie.Grundlagen,Methoden und Anwendungen.* Wichmann, Heidelberg., 2003

[6] Richard Hartley, Andrew Z.: *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2003

[7] Jinwei Ye, Jingyi Y.: Ray Geometry in non-Pinhole Cameras: a Survey. (2013)

[8] Jaehne, Bernd: *Practical Handbook on Image Processing for Scientific Applications.* CRC Press LLC, 1997

[9] R.J.Tsai ; R.K.Lenz: A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration. In: *IEEE Transactions on Robotics and Automation* 5 (1989)

[10] J.Ilonen ; V.Kyrki: *Robust Robot-Camera Calibration.* EEEI, 15th International Conference on Advanced Robotics, 2011

[11] R.K.Lenz ; R.Y.Tsai: *Calibrating a Cartesian Robot with Eye-on-Hand Configuration Independent of Eye-to-Hand Relationship.* EEEI Computer Vision and Pattern Recognition Cont., 1988

[12] Z.Zhang: *A Flexible New Technique for Camera Calibration.* EEEI Transactions on Pattern Analysis and Machine Intelligence, 1998

[13] R.Y.Tsai: A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. In: *IEEE Jornal of Robotics and Automation* (1987), p. 323â344

[14] Zhang, Z.: Flexible Camera Calibration By Viewing a Plane From Unknown Orientations. In: *Proc. Int. Conf. on Computer Vision* (1999), p. 666â673

[15] J.Y.Bouguet: *Camera Calibration Toolbox for Matlab.* http://www.vision.caltech.edu/bouguetj/calib doc/, accessed September 04,2007

[16] L.Krueger, A. Wuerz-Wessel ; F.Stein: In-factory calibration of multiocular camera systems. In: *Proc. SPIE Photonics Europe, Optical Metrology in Production Engineering* (2004), p. 126â137

[17] B.Girod, G.Greiner ; H.Niemann: *Principles of 3D Image Analysis and Synthesis.* Kluwer Academis Publishers, 2000

[18] Milan Sonka, Vaclav H. ; Boyle, Roger: *Image Processing, Analysis, and Macine Vision.* Brooks/Cole Publishing Company, 1999

[19] Wang, G. ; Wu, Q.M.J.: *Guide to Three Dimensional Structure and Motion Factorization, Advances in Pattern Recognition.* Springer-Verlag London Limited, 2011

[20] D.Marr: *Vision- A computational Investigation into a Human Representation and Processing of Visual Information.* Freeman, San Francisco, 1982

[21] Aloimonos, Y. ; Shulman, D.: *Integration of Visual Modules- An Extension of the Marr Paradigm.* Academic Press, New York, 1989

[22] Wechsler, H.: *Computational Vision.* Academic Press, Academic Press, London-San Diego, 1990

[23] Y. Aloimonos, editor: *Active Perception.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1993

[24] R.Nevatia, M.Zerroug ; F.Ulupinar: Recovery of Three-Dimensional Shape of Curved Objects from Single Image. In: *Handbook of Pattern Recognition and Image Processing: Computer Vision* (1994), p. 101–129

[25] F.Ulupinar ; R.Nevatia: Shape from Contour: Straight Homogeneous Generalized Cylinders and Constant Cross Section Generalized Cylinders. (1995)

[26] S.Ullman: *High-Level Vision: Object Recognition and Visual Cognition.* MIT Press, Cambridge, MA, 1996

[27] B.K.P.Horn: *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View.* PhD Thesis, Department of Electrical Engineering, MIT, Cambridge, MA, 1970

[28] B.K.P.Horn: *Robot Vision.* MIT Press, Cambridge, MA, 1986

[29] B.K.P.Horn: *Hieght and Gradient from Shading.* 1990. – 37–75 p

[30] B.K.P. Horn, editors: *Shape from Shading.* MIT Press, Cambridge, MA, 1989

[31] S.Ullman: *The Interpretation of Visual Motion.* MIT Press, Cambridge, MA, 1979

[32] S.Maybank: *Theory of Reconstruction from Image Motion.* Springer,Berlin, 1993

[33] L.S.Shapiro: *Affine Analysis of Image Sequences.* Cambridge University Press, 1996

[34] K.Kanatyni: Detection of surface orientation and motion from texture by a stereological technique. In: *Artif Intell, 23* (1984), p. 213–237

[35] Winkelbach, S. ; Wahl, F. M.: Shape from Single Stripe Pattern Illumination. In: *Pattern Recognition, Lecture Notes in Computer Science 2449, Springer* (2002), p. 240–247

[36] R.E.Jacobson, G.G.Attridge ; N.R.Axford: *The Manual of Photography, Ninth Edition.* Focal Press, Reed Educational and Professional Publishing Ltd, 2000

[37] J.C.Russ: *The Image Processing Handbook (Sixth Edition).* Taylor and Francis Group, LLC, 2011

[38] Hursrev Taha Sencar, Editors: *Digital Image Forensics. There is More to a Picture than Meets the Eye.* Springer Science Business Media New York, 2013

[39] B.Andersson, J.L.Geyen: *The DSLR Filmmaker's Handbook: Real-World Production Techniques.* John Wiley Sons Inc., 2012

[40] CORPORATION, KEYENCE: *Vision Development Software for the XG-8000 Series XG VisionEditor. Reference Manual (Programming Edition).* KEYENCE CORPORATION, 2012

[41] Y. Zhehg, Y. L.: *Another Way of Looking at Monocular Circle Pose Estimation.* EEEI, 2008

[42] Jiang, N. ; Jiang, Z.: *Distance Measurement from Single Image Based on Circles.* EEEI, 2007

[43] R. Safaee-Rad, I. T.: *Three-Dimensional Location Estimation of Circlar Features for Machine Vision.* EEEI Transactions on Robotics and Automation, 1992

[44] Boer, J.R.J van Asperen de: Infrared Reflectography: A Method for the Examination of Paintings. In: *Appl. Opt. 7* (1968), p. 1711–1714

[45] D.C.Montgomery: *Design and Analysis of Experiments.* John Wiley Sons,Inc., Seventh Edition 2009

[46] Antony, Jiju: *Design of Experiments for Engineers and Scientists.* Elsevier Science Technology Books, 2003