Technical University of Dortmund

Stereo camera de-calibration detection based on observing kinematic attributes of detected objects and the camera rig

by

Venkata Rama Prasad Donda

A thesis submitted in partial fulfillment for the
degree of Masters in Automation and Robotics

in the
DEPARTMENT OF ELECTRICAL ENGINEERING

July 2015

Declaration of Authorship

I, Venkata Rama Prasad Donda, declare that this thesis titled, 'Stereo camera decalibration detection based on observing kinematic attributes of detected objects and the camera rig' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

"Take up one idea. Make that one idea your life - think of it, dream of it, live on that idea. Let the brain, muscles, nerves, every part of your body, be full of that idea, and just leave every other idea alone. This is the way to success."

Swami Vivekananda

Technical University of Dortmund

Abstract

DEPARTMENT OF ELECTRICAL ENGINEERING

Masters in Automation and Robotics

by Venkata Rama Prasad Donda

A common approach for calibration of the relative orientation between the camera heads of a stereo camera is inferring the rotation parameters from correspondence analysis between the two camera images. This method uses the information from the images and is dependent on scene characteristics. Thus, the image evaluation has to cope with different real world scenario challenges [1] [2]. Adding more information from different sensors has potential to enhance or stabilize the calibration.

"Stereo camera de-calibration detection based on observing kinematic attributes of detected objects and the camera rig" aims to investigate if motion information of the stereo camera rig and the measured kinematic attributes of the objects detected enables de-calibration detection or even enhances the calibration. In this thesis an appropriate algorithm for the detection and estimation of de-calibration due to yaw angle error has been presented. This algorithm is further tested on a simulation system and on real world data.

This algorithm requires less processing power and can be easily added as an additional system to check for de-calibration in the stereo camera due to yaw angle error. It is found out that the proposed algorithm is able to estimate the de-calibration due to yaw angle well with in the tolerance required for real world application.

Acknowledgements

I wish to express my sincere thanks to Robert Bosch GmbH for providing me with all the necessary facilities for the research. I am extremely thankful to Mr. Marcel de Jong, supervisor of the thesis, for his sincere and valuable guidance and encouragement extended to me.

I am also grateful to Prof. Dr. Jian Jia Chen, lecturer TU Dortmund, in the Department of Information technology. I am indebted to him for sharing expertise.

I take this opportunity to express gratitude to all of the members of Robert Bosch GmbH for their help and support.

I also place on record, my sense of gratitude to one and all, who directly or indirectly, have lemt their hand in this venture

Contents

List of Figures

Chapter 1

Introduction

Stereo vision systems are widely used in the fields of robotics and in automotive industry for object recognition, autonomous navigation, driver assistance or safety systems. A stereo vision system is similar to that of a human vision system, it provides 3-D information about the scene by using the two images obtained for the scene from two different positions. This 3-D information consists of depth information in addition to the 2-D information from a single camera

There are other systems to get the 3-D information about the scene, which are laser range finders. In automotive applications generally stereo camera systems are preferred over these because they can provide visual data which provides information about the scene like road signs.

The two cameras of stereo camera are related to each other by some rotation and translation parameters, through which the 3-D reconstruction of the scene is possible. A stereo setup in the figure 1.1 shows these rotation (yaw, pitch and roll) and translation (X,Y,Z) parameters.

Stereo camera calibration involves finding these relative orientation and translation parameters between two cameras. Any error in these parameters has erroneous effect on the estimation of 3-D measurements of the scene.

The aim of this thesis is to detect the de-calibration in a stereo camera due to error in yaw angle used for driver assistance systems in automobiles, by using the image information and knowledge about motion of the stereo camera and the objects detected by the stereo camera.

Similar approaches has been investigated for calibration of stereo camera using ego velocity estimation by Visual Odometry [6].

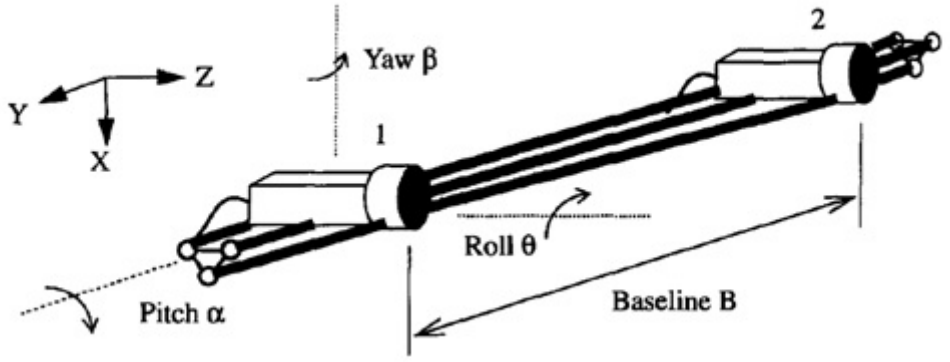

FIGURE 1.1: Rotation and translation parameters of stereo camera [3]

1.1 Motivation

It is not the goal of the thesis to develop a best possible algorithm to solve for online calibration. This thesis presents, how good an online calibration approach would be, when it is only based on motion information from detected objects and the ego vehicle. The benefit of such algorithm is light weight data processing in contrast to the general approach, which requires to analyze over all pixels of the image. Especially the former is also an important factor for embedded use.

1.2 Background

1.2.1 General setup used for driver assistance systems

A stereo camera is attached to the ego vehicle which is moving with known motion attributes (velocity, acceleration) in a scene. Based on the images obtained from this camera inferences about the world (like detection of an object, distance to the object etc.) are made. As the stereo camera is attached to the ego vehicle, it is subjected to mechanical shocks and thermal effects which can result in de-calibration of the camera over time. Thus a online calibration for the stereo camera is required, because static calibration is not possible (as it requires a known calibration object under controlled conditions).

1.2.2 Effect on 3-D measurements by relative rotation parameters

Even though the camera calibration of individual cameras is good, a poor alignment of the cameras of a stereo camera can lead to a bad estimation of 3-D measurements about

the scene [3]. Of all the rotation parameters, error in yaw angle has the most effect on the estimation of depth. This will be discussed in Chapter 2.

In safety critical systems like driver assistance, this effect is very important and any misalignment should be compensated. This thesis studies the effect of error in yaw angle and presents an algorithm to estimate the error present in it.

1.2.3 General approach to stereo camera calibration

A general way of stereo camera calibration involves finding the stereo correspondence between the two images obtained from it [7]. This involves finding corresponding image points in two images, by looking for similar locations in two images (block matching). Figure 1.2 shows stereo correspondence between two images. Once the corresponding image points are known, rotation and translation parameters are then determined using 7 or more such corresponding image points. [8]

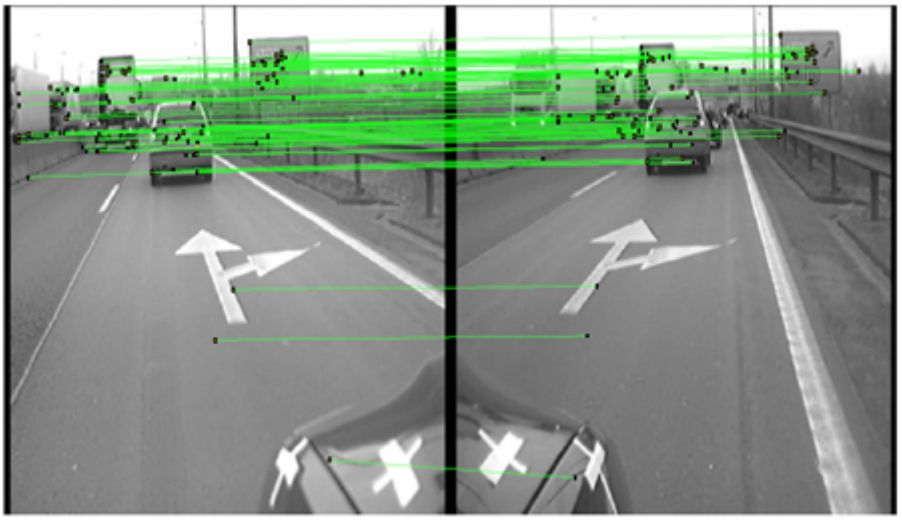

FIGURE 1.2: Stereo correspondence of a stereo image [4]

Disadvantage:

The disadvantage of this method is that it only uses the information from the images and is thus dependent on the scene characteristics, like weather conditions, day or night etc. This results in problems with establishing stereo correspondence between two images and thus further affects the calibration of the camera. [8] [2]

1.3 De-calibration detection from kinematic attributes

The idea of de-calibration detection from kinematic attributes is not only to use the image information but also use the motion information from the ego vehicle and that of the objects which are detected by stereo camera. This information is then used to detect the de-calibration due to incorrect estimation of relative orientation of the camera heads of a stereo camera.

When the ego car to which the stereo camera is attached is moving towards an object that is detected , a simple definition of relative velocity would be the difference between absolute velocity of object and ego car. Absolute velocity is the velocity of an object with respect to a static 3-Dimensional frame. Relative velocity is observed by the stereo camera as a function of depth between the objects detected and the ego car. If the object is assumed to be static then the relative velocity should be the same as the absolute velocity of the ego car but in the opposite direction. If the camera is de-calibrated then this relative velocity magnitude is observed not be equal to magnitude of ego velocity as shown in figure 1.3. If the assumption that relative velocity calculated by considering all the objects are static, is considered to be the ground truth, then the measured relative velocity from the stereo camera is compared to the ground truth and the deviation can be expressed in terms of error in yaw angle of the stereo camera. This simple affect is used by this thesis to determine the de-calibration in the stereo camera.

1.3.1 Goals of the thesis

The goals of this thesis are to:

- Design an algorithm to detect the de-calibration due to misalignment of yaw angle in stereo camera using motion attributes.
- Simulation of algorithm (in MATLAB) using introduced know error.
- Implement and evaluate the algorithm on the real world data from the existing stereo camera system.
- Analyse and compare the results against reference data.

1.3.2 Scope

The object detection, disparity and tracking information is available from the existing system. So, this thesis does not look in to these aspects. The ego velocity information is

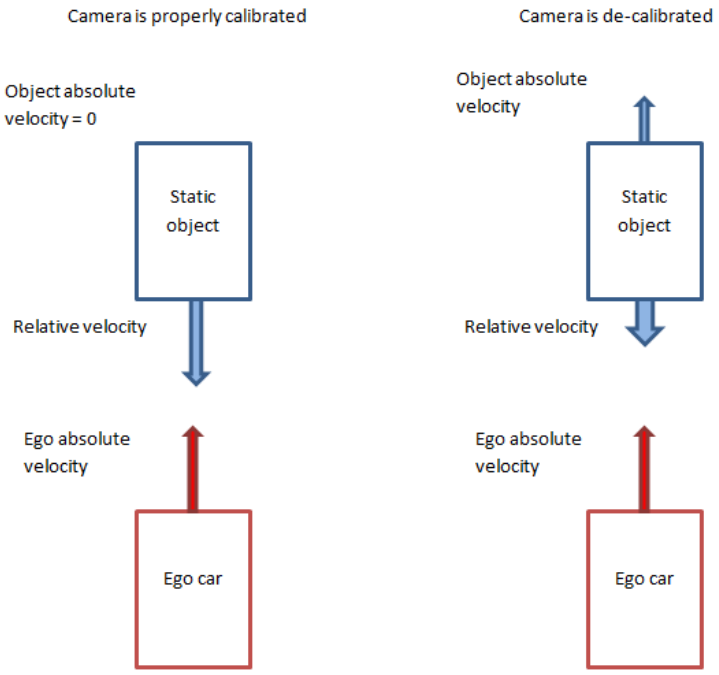

FIGURE 1.3: Relative velocity between the static object and the ego car when the stereo camera is calibrated and when it is de-calibrated

also available and it is assumed that there are no errors in it for designing the algorithm. The effects of errors in ego velocity will be discussed in Chapter 6. Furthermore this thesis does not look into classification of the detected objects and assumes a robust classification is already present.

1.4 Overview

Chapter 2 will present the state of the art. Chapter 3 will discuss about the theoritical background the de-calibration detection using kinematic attributes and establishes how motion information can be used for de-calibration detection. Chapter 4 will explain the algorithm, its mathematical derivation and its constraints. Chapter 5 will present simulation scenario (in MATLAB) and its results. Chapter 6 will present how the algorithm is adapted for real world data. Chapter 7 will present the results of the application of the algorithm and discuss the significance of the results. Chapter 8 will conclude the thesis.

Chapter 2

Literature Review

This chapter establishes the theoretical base for the research of this thesis.

2.1 Pin hole camera

A pin hole camera is a simple camera without a lens. It has a small aperture size of a pin hole that is large enough to allow light ray pass through it. When a light source is present on one side of the camera, a light ray passes through this aperture and generates a inverted image on the other side of the camera. This can be ideally modeled as perspective projection [5].

2.1.1 Pin hole camera model and perspective projection

A pin hole camera model is used to find the relation between the 3-D co-ordinates in the scene and the 2-D co-ordinates in the image obtained. The figure 2.1 shows such a pinhole camera model, where the pin hole C is present on the focal plane F. A light ray from a world point M passes through the pin hole and is projected on to the image plane I as image point m. The world point, image point and the pinhole form a straight line. This kind of projection is called perspective projection.

The point C is the optical center. The plane passing through optical center and parallel to the image plane χ is the focal plane. The distance between the image plane and the optical center is the focal length. The line passing through the optical center and perpendicular to the image plane is the optical axis. The point at which it intersects the image plane is the principal point.

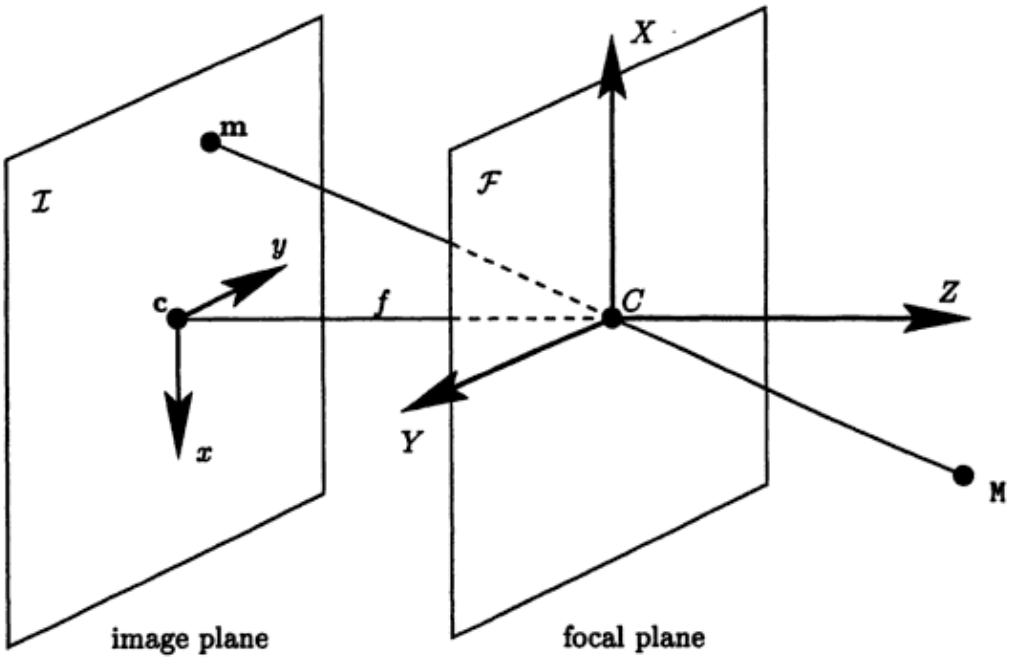

FIGURE 2.1: A pinhole camera model [5]

2.1.2 Geometry

In the figure 2.1, let (X,Y,Z) be the co-ordinates of the world point M and (x,y) be the planar co-ordinates of image point m. Since m, C, M are collinear (perspective projection) it can be written that [5],

$$\frac{x}{X} = \frac{y}{Y} = \frac{f}{Z}. \tag{2.1}$$

where f is the camera constant. It should be noted that the co-ordinate system

2.2 Stereo image analysis

Stereo image analysis deals with reconstruction of three-dimensional scene structure based on two images acquired from different positions and viewing directions [1]. A simple setup of stereo rig is show in the figure 2.2.

From the figure, let $W(X, Y, Z)$ be the point the world coordinate system whose center is at O. The project image co-ordinates of W are given by $\hat{u}_1, \hat{v}_1$ in the image plane of one camera and $\hat{u}_2, \hat{v}_2$ be the image plane co-ordinates in the image place of the other. Here f is the camera constant and b is the distance between the optical axis of two

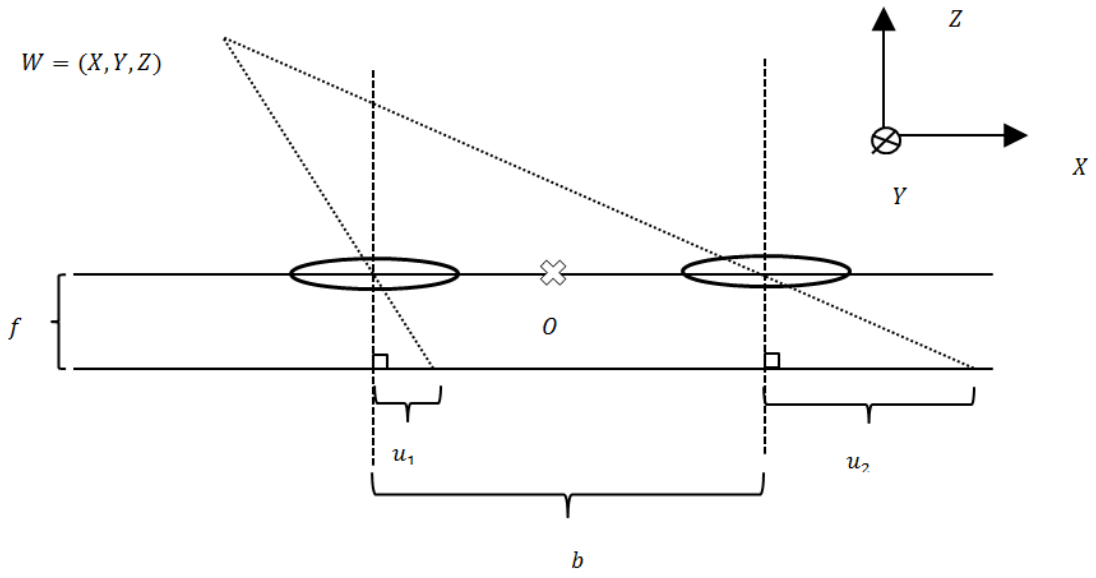

FIGURE 2.2: Stereo image analysis geometry

cameras (also called as stereo base). The relation between a world point and the two image points obtained from the projection of world point is given from the geometry as [1]:

$$\frac{\hat{u}_1}{f} = \frac{X + b/2}{Z} \tag{2.2}$$

$$\frac{\hat{u}_2}{f} = \frac{X - b/2}{Z} \tag{2.3}$$

$$\frac{\hat{v}_1}{f} = \frac{\hat{v}_2}{f} = \frac{Y}{Z}. \tag{2.4}$$

It can also be written as,

$$\frac{\hat{u}_1 - \hat{u}_2}{f} = \frac{b}{Z}. \tag{2.5}$$

Here the difference $\hat{u}_1 - \hat{u}_2$ is called as disparity. Which is the difference of the projected image co-ordinates on two image planes of the stereo camera. In a general way , it can be written as:

$$disparity = \frac{f \cdot b}{Depth} \tag{2.6}$$

From equation 2.5 it can be inferred that by knowing the disparity in the images of the stereo camera, the camera constant and the stereo base the depth information of point in the world co-ordinate system can be determined.

2.3 Camera calibration

In general camera calibration aims to determine the transformation parameters between the camera and the image plane (intrinsic parameters) and between the camera and the scene (extrinsic parameters). This is achieved by using the images acquired from the camera rig [1]. There are two types of camera calibrations, calibration based on images of a calibration rig of known geometry and self-calibration, where the calibration is based on the feature points extracted from the scene of unknown geometry.

2.3.1 Direct Linear Transform method of calibration

The DLT method is the simple method of camera calibration which assumes the pinhole camera model. It uses control points, the 3-D points in the world, whose co-ordinates are known precisely with respect to some origin, to determine the transformation parameters. The corresponding image points are measured from the principal point in the image plane. The basic principle involves expressing the image co-ordinates in terms of world co-ordinate system by means of a translation and rotation matrices. Then by using 6 known control points, the rotation parameters and the translation parameters are determined. This method is an example for camera calibration based on calibration rig of known geometry.

2.3.2 General method for self calibration of camera

The crucial step in the self calibration of the camera involves determination of the fundamental matrix between the image pairs. Fundamental matrix is the matrix that provides the representation of both the intrinsic and the extrinsic parameters of the two cameras. To determine the fundamental matrix the seven or more point correspondences is required. [1] [9]

These point correspondences are established by methods like block matching. In block matching the similarity between the images is measured by using pixel wise methods like finding sum of squared differences (SSD) or sum of absolute differences (SAD). The blocks with less value of SSD or SAD is are the established are corresponding points.[10]

Once the correspondences are established the fundamental matrix is then determined by using the linear constraint of

$$x_2^T \cdot F \cdot x_1 = 0. \tag{2.7}$$

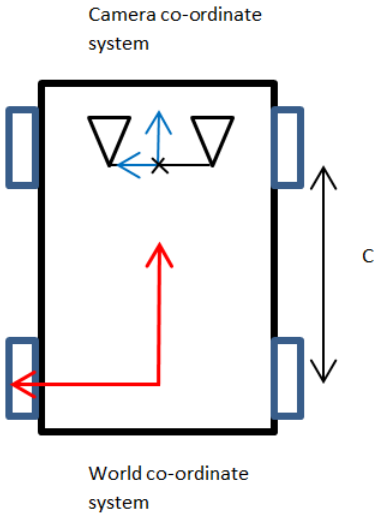

FIGURE 2.3: World and camera co-ordinate systems in the camera rig setup

Where x_1, x_2 are the corresponding points in camera sensor co-ordinate systems of two cameras. In sensor co-ordinate system measurements are made in pixels and the origin lies at the upper left corner. In contrast to the image co-ordinates where measurements are made in meters and the origin lies at the principal point.

2.4 Co-ordinate systems for the camera rig used in this thesis

In the camera setup used in this thesis there are 2 co-ordinate systems. The stereo camera co-ordinate system which is located at the center of the stereo camera. The world co-ordinate system which is located at the rear axle of the ego car. This is shown in the figure 2.3.

The linear transformations between the two co-ordinate systems can be given by :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} \tag{2.8}$$

where, $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ are the co-ordinates of the world point measured in meters in camera

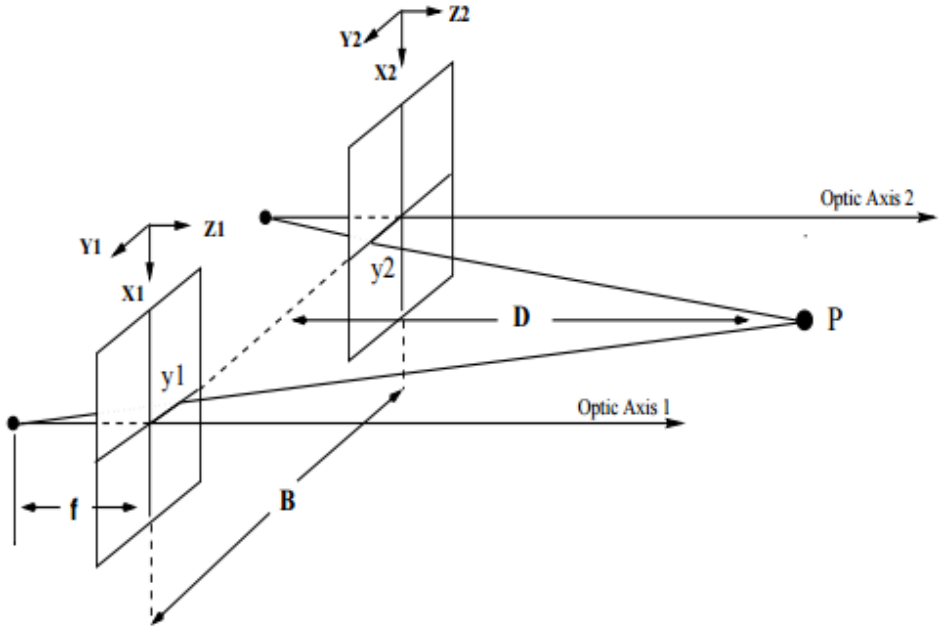

FIGURE 2.4: Stereo camera setup used for error analysis [3]

co-ordinate system. $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ are the co-ordinates of the world point measured in meters in world co-ordinate system. c is the distance between the stereo camera and the rear axle of the car.

2.5 Analysis of depth estimate error in stereo camera due to various alignment errors

The paper on stereoscopic depth estimates by Wenyi Zhao and Nandhakumar [3] presents the effect of alignment errors on depth estimates. The error analysis is divided in to 3 categories:

1. Depth error due to rotation between the two cameras
2. Depth error due to pitch between the cameras
3. Depth error due to yaw between the cameras

The camera setup used has a focal length of $f = 8.5mm$ and $b = 1330$ mm. This setup is shown in the figure 2.4.

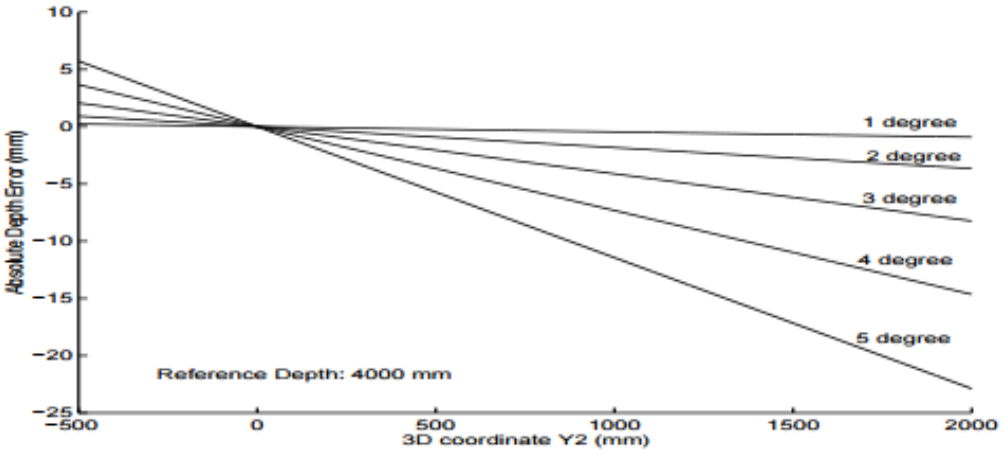

FIGURE 2.5: 3D error due to roll between the cameras [3]

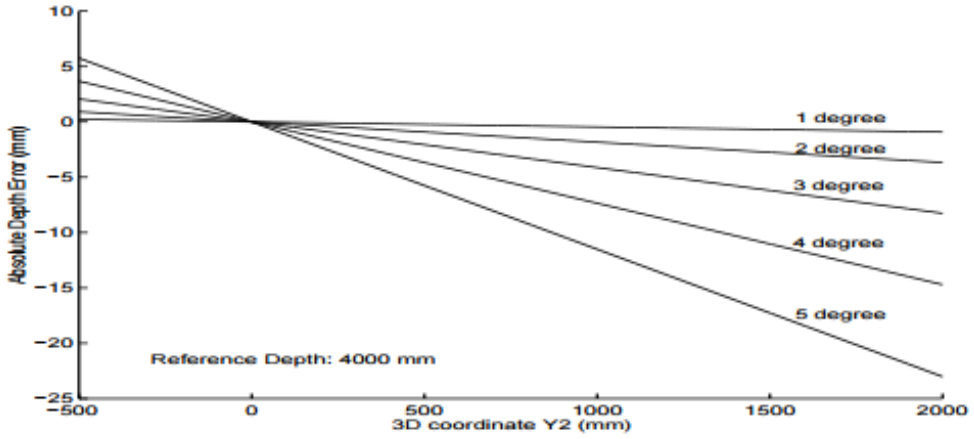

FIGURE 2.6: 3D error due to pitch between the cameras [3]

It is assumed that the camera 1 of the stereo rig is perfectly aligned but the camera 2 is mis-aligned for all the above stated error analysis. The graph of depth error vs y2 co-ordinate in the image plane. The results for each analysis is presented below:

From figures 2.5, 2.6, 2.7 it can be observed that and misalignment in the yaw angle has the most effect on the estimation of the depth when compared to roll and pitch between the cameras.

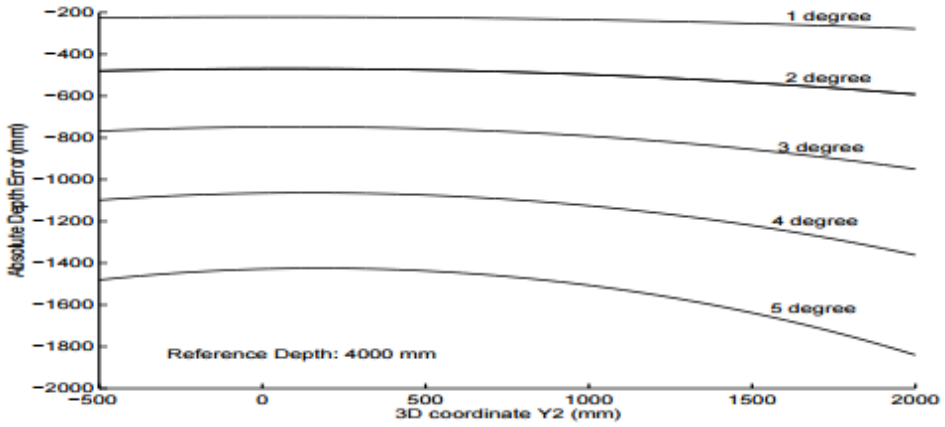

FIGURE 2.7: 3D error due to yaw between the cameras [3]

Chapter 3

De-calibration detection from kinematic attributes

This chapter establishes the basic idea of whether kinematic attributes can be used to detect de-calibration of the stereo camera and background knowledge leading to algorithm for de-calibration detection.

This thesis assumes that a de-calibration can occur only due to error in yaw angle between the camera heads. The motivation for this assumption is, as mentioned in the chapter 2, an error in yaw angle orientation between the camera heads is the most critical parameter and the reason for most depth error.

Relative velocity can be expressed as the velocity of the object detected when ego car is considered to be at rest. Relative velocity is used as a key parameter by this thesis to determine the yaw angle error between the cameras. Relative velocity between the object and the ego car has the motion information of both ego vehicle and object detected. It also can be related to the depth between the ego vehicle and objects, as a first derivative of depth with respect to time.

In order to determine whether relative velocity information can be used to find yaw angle error in the stereo camera the effects of yaw angle error on disparity and depth has to be studied first.

3.1 Relation between yaw angle error and error in disparity

Assuming all other rotation parameters are aligned such that there is de-calibration in them except for the yaw angle of the camera as shown in figure 3.1. Here camera 2 is misaligned with camera 1 by an yaw error of γ, the relationship between yaw angle error (γ) and error in disparity (ϵ_d) can be given from the geometry as shown in figure 3.2,

where, I is the image plane when the camera 2 is calibrated properly, I' is the image plane when it is de-calibrated
γ is the yaw angle error
O is the optical center of the camera 2
P is the point in world which is projected onto image planes I and I' at a distance from principal point of y_2 and $y_2\prime$

From the geometry in the figure 3.2, it is known that [3]

$$\alpha = \gamma + \tau \tag{3.1}$$

$$\tan\alpha = \frac{y_2}{f} \tag{3.2}$$

$$\tan\tau = \frac{y_2'}{f} \tag{3.3}$$

Solving for y_2',

$$y_2' = f\left[\frac{y_2 - f\tan\gamma}{y_2\tan\gamma + f}\right] \tag{3.4}$$

Error in disparity is given as,

$$\epsilon_d = y_2 - y_2' = \frac{\tan\gamma\cdot(f^2 + y_2^2)}{y_2\tan\gamma + f} \tag{3.5}$$

where, f is the camera constant and y_2 is the distance between the image plane co-ordinate projected from point P and the camera center O.

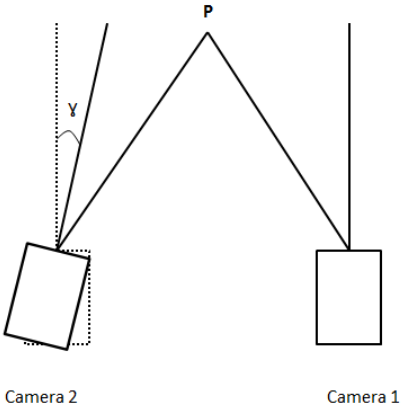

FIGURE 3.1: Camera alignment with yaw angle error [3]

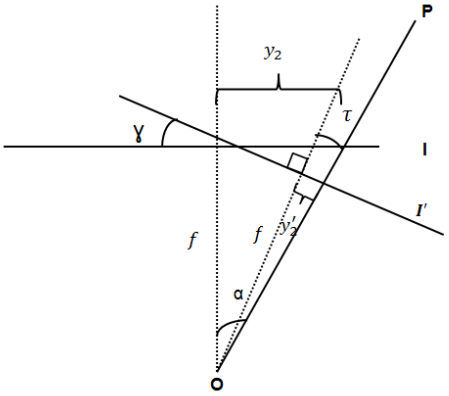

FIGURE 3.2: Effect of yaw angle error on image point [3]

If the center of the image is considered, then $y_2 = 0$ and this can be approximated as,

$$\epsilon_d \approx \tan\gamma \cdot f \tag{3.6}$$

As the camera constant f is does not change over time, it can be assumed that the error in disparity caused by yaw angle error is constant.

3.2 Behaviour of depth error when the error in disparity is constant

As stated above, it is assumed that the disparity error caused by a misalignment in the yaw angle is constant. From the pinhole camera model explained in the Chapter 2,

disparity and depth are inversely proportional to each other by the relation,

disparity $= \frac{f.b}{\text{depth}}$, where f is the camera constant, b is the stereo base.

Let d_t be the true disparity, i.e the disparity when the camera is not de-calibrated and let d_m be the disparity measured when the camera is de-calibrated by an error of ϵ_d. Then measured disparity is given by $d_m = d_t + \epsilon_d$

Let D_t be the true depth and the observed depth be D_m then the depth error would be,

$$\begin{aligned} D_{error} &= D_m - D_t \\ &= \frac{f \cdot b}{d_m} - \frac{f \cdot b}{d_t} \\ &= \frac{f \cdot b}{d_m} - \frac{f \cdot b}{d_m - \epsilon_d} \\ &= f \cdot b \frac{-\epsilon_d}{(d_m)(d_m - \epsilon_d)} \end{aligned} \tag{3.7}$$

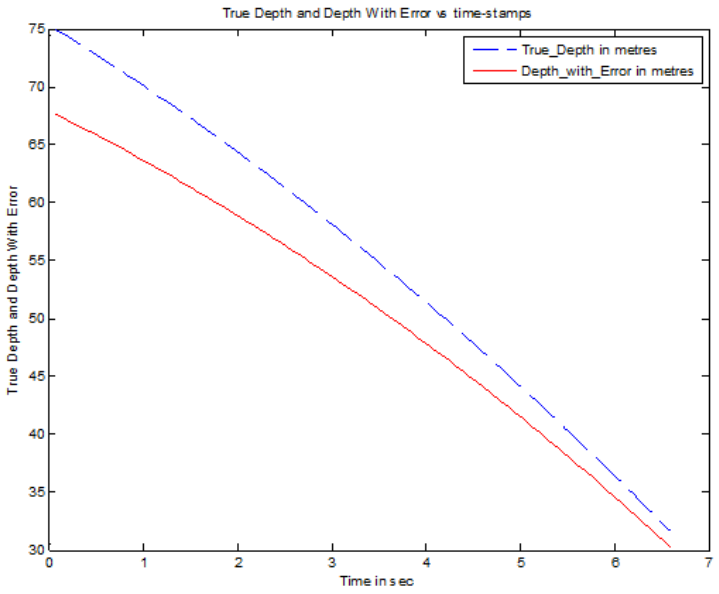

FIGURE 3.3: Variation of depth when the camera is calibrated and when it is de-calibrated by 0.25px disparity error

The figure 3.3 shows the behavior of the depth error when a known error in disparity(0.25 pixel) is introduced. This is for a case when depth between the ego car and the object detected is decreasing over time.

It can be observed that the depth error is not constant over depth (in this case as the depth is decreasing over time). It has significance as it shows that the yaw angle error

can be observed in the relative velocity between camera and the object which is detected, as relative velocity is a function of change in depth between them.

$$v_{rel} = \frac{\delta D}{\delta t} \tag{3.8}$$

Where δt is the change in time in which the change in depth occurred.

3.3 Relative velocity between ego car and object detected

The origin of the reference co-ordinate system is at the center of the rear axis of the ego car as shown in figure 3.4. Consider a simple case when the ego car is moving with an absolute velocity of $\vec{v_{ego}}$ towards the object with absolute velocity of $\vec{v_{obj}}$, then the relative velocity can be given as

$$\vec{v}^{rel} = \vec{v}_{obj} - \vec{v}_{ego}. \tag{3.9}$$

Generally ego velocity can be accompanied by some yaw rate $\vec{\omega}$. By including this the relative velocity can be given as [11]

$$\vec{v}^{rel} = \vec{v}_{obj} - \vec{v}_{ego} + \vec{\omega} \times \vec{r}. \tag{3.10}$$

where $\vec{r}$ is the position vector of the object that is detected given by $\vec{r} = r_x\hat{\imath} + r_y\hat{\jmath}$ with respect to the co-ordinate system at rear axis of the car. r_x denotes the depth between the object and the ego car and r_y denotes the distance between the object and ego car along y direction as shown in figure 3.4.

As the depth between the ego car and the object corresponds to the x component of position vector i.e. r_x, consider the x component of the equation 3.10,

$$\vec{v}_x^{rel} = \vec{v}_{obj_x} - \vec{v}_{ego_x} - \omega r_y \tag{3.11}$$

3.4 De-calibration detection from relative velocity

It is established from the previous section 3.2 that the relative yaw angle error between the two cameras of stereo camera will affect the measured relative velocity between the

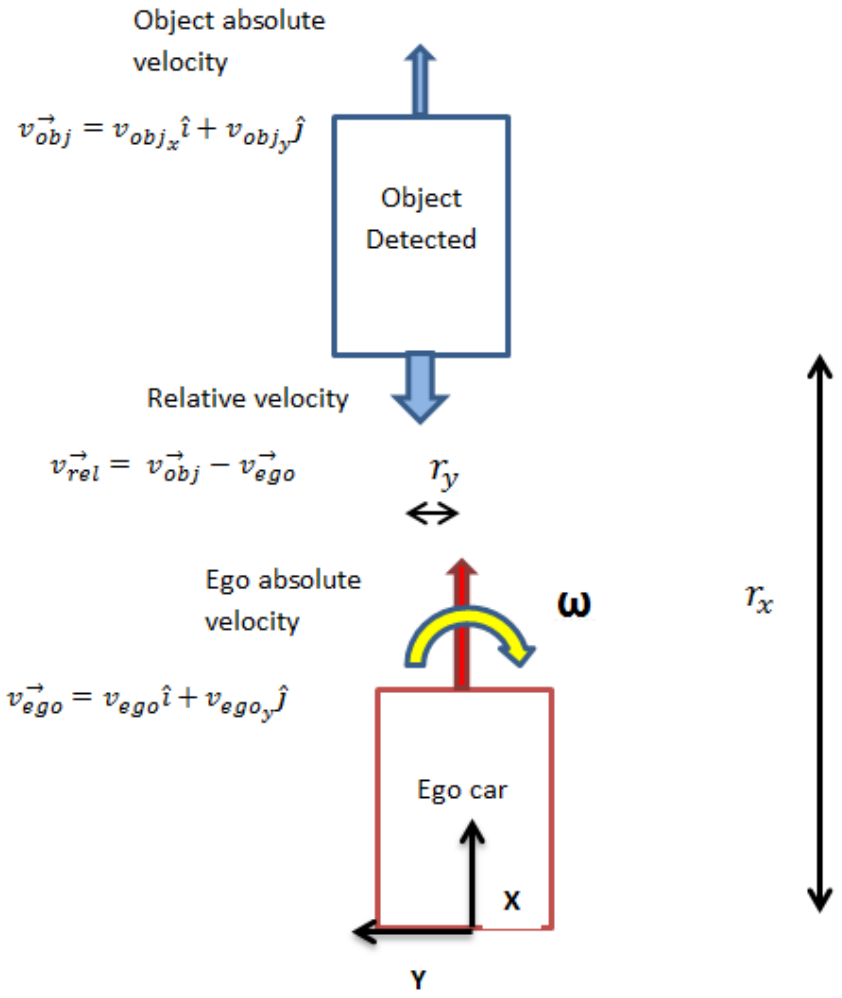

FIGURE 3.4: Depiction of relative velocity

ego car and the object detected. Assume a scenario where the ego car is travelling through a scene where all the objects that are present are static i.e. the absolute velocities of the objects are zero, then the equation 3.11 can be written as

$$\vec{\tilde{v}}^{rel}{}_x = -\vec{v}_{ego_x} - \omega r_y. \tag{3.12}$$

This relative velocity is considered as the ground truth $\tilde{v^{rel}}_x$ i.e. if the camera is calibrated properly all the objects that are detected must have produced a relative velocity of $\tilde{v^{rel}}_x$.

In general relative velocity (measured from stereo camera) can be given as a function of the measured depth of the objects as,

$$v^{rel}_{m_x} = \frac{D_m(t) - D_m(t - \delta t)}{\delta t} \tag{3.13}$$

Where δt is the change in time in which the change in depth occurred. Also, this depth signifies the distance between the ego car and the object that is detected along x-axis as shown in figure 3.4.

If the camera is de-calibrated then the measured relative velocity is not equal to the true relative velocity i.e. $v_m^{rel} \neq \tilde{v}^{rel}$. This deviation of measured relative velocity from ground truth is then used to find out the error in disparity. This will be explained in Chapter 4.

Although, in general the assumption that all objects are static does not hold true. As when the ego car is present in a city scenario where most of the objects present are static (road side objects like trees, signals etc.), there are almost always vehicles that are moving. In the further chapters it will be explained how this problem is solved.

In this chapter it has been shown that the relative velocity can be used to determine the de-calibration in the stereo camera. How the de-calibration can be detected from relative velocity will be explained in the chapter 4.

Chapter 4

Algorithm and Implementation

This chapter explains how motion attributes can be used to detect de-calibration in the camera and the mathematical derivation for it.

4.1 Notation

This is the notation used for the derivation of the algorithm:

D – Depth

d – Disparity

v – Velocity

t – Time stamp

δt – small change in time

ϵ_d – Error in disparity

$\tilde{v}^{rel}$ – Ground truth relative velocity value from assumption that all objects are static

v_m^{rel} – Measured relative velocity value obtained from the stereo camera which is de-calibrated

v_{ego} – Absolute velocity of the ego car

v_{obj} – Absolute velocity of the object

4.2 Assumptions

Following are the assumptions that were considered for the derivation:

- De-calibration is only because of incorrect yaw angle estimation of the cameras of a stereo camera.
- Error in disparity caused by the yaw angle error is constant over depth.
- Objects are already detected and the measured motion information is available.
- Ego velocity of the car is available and there is no error in it.

4.3 Algorithm Derivation

4.3.1 Relative velocity as measured by the stereo camera and ground truth

Relative velocity can be written in terms of change in depth. This is obtained as follows: after the objects are detected, the disparity corresponding to each object is obtained from the stereo camera. By assuming a pin hole camera model, this disparity is converted into depth as mentioned in the Chapter 2. As discussed in section 3.4, the relative velocity as measured by the stereo camera is given by equation

$$v_{m\ x}^{rel} = \frac{D_m(t) - D_m(t - \delta t)}{\delta t}. \tag{4.1}$$

The ground truth assumption that is made that the absolute velocity of the objects is zero.
True relative velocity as mentioned in section 3.4 is given as,

$$\tilde{v^{rel}}_x = -v_{ego_x} - \omega r_y. \tag{4.2}$$

Consider the relative error in relative velocity in terms of velocity

$$\text{RERV}_{velocity} = \frac{v_{m\ x}^{rel} - \tilde{v^{rel}}_x}{\tilde{v^{rel}}_x}. \tag{4.3}$$

From equation 4.2, this can be written as,

$$\boxed{RERV_{velocity} = \frac{v_{m\ x}^{rel} + (v_{ego_x} + \omega r_y)}{-(v_{ego_x} + \omega r_y)}}. \tag{4.4}$$

The $RERV_{velocity}$ value is known, as the objects the measurements are available from the stereo camera (de-calibrated) and the ego velocity. The yaw rate is available from the wheel sensors.

4.3.2 Relative velocity as a function of disparity

Let $\tilde{d}$ be the disparity the corresponds to ground truth and d_m corresponds to the measured disparity by the stereo camera which is de-calibrated.

A general relation between depth and disparity, as discussed in chapter 2 can be given as

$$D = \frac{f \cdot b}{d}, \tag{4.5}$$

where f is the camera constant and b is the stereo base.

So, measured depth can be given as

$$D_m = \frac{f \cdot b}{d_m}. \tag{4.6}$$

Similarly true depth is given as

$$\tilde{D} = \frac{f \cdot b}{\tilde{d}}. \tag{4.7}$$

Then true and measured relative velocity at time t is given by

$$v_{m\ x}^{rel} = \frac{\frac{f \cdot b}{d_m(t)} - \frac{f \cdot b}{d_m(t - \delta t)}}{\delta t} \tag{4.8}$$

$$\tilde{v}_x^{rel} = \frac{\frac{f \cdot b}{\tilde{d}(t)} - \frac{f \cdot b}{\tilde{d}(t - \delta t)}}{\delta t} \tag{4.9}$$

where $d(t - \delta t)$ is the disparity at time $t - \delta t$ and $d(t)$ is the disparity at t .

Let δd be the change in disparity from time $t - \delta t$ to t then,

$$d_m(t) = d_m(t - \delta t) + \delta d_m \tag{4.10}$$

$$\tilde{d}(t) = \tilde{d}(t - \delta t) + \delta\tilde{d} \tag{4.11}$$

Also, measured disparity at time t $d_m(t)$, can be written in terms of ground truth disparity at time t $\tilde{d}(t)$ as

$$d_m(t) = \tilde{d(t)} + \epsilon_d. \tag{4.12}$$

Since it is assumed that the error in disparity is constant, this holds true for $d_m(t - \delta t)$ as well.

By using equations 4.10, 4.11 and 4.12 it can be shown that

$$\delta d_m = \delta\tilde{d} = \delta d. \tag{4.13}$$

So, true relative velocity can be written as

$$\tilde{v^{rel}}_x = \frac{\dfrac{f \cdot b}{\tilde{d}(t - \delta t) + \delta d} - \dfrac{f \cdot b}{\tilde{d}(t - \delta t)}}{\delta t}. \tag{4.14}$$

and measured relative velocity can be written as

$$v^{rel}_{m_x} = \frac{\dfrac{f \cdot b}{\tilde{d}(t - \delta t) + \delta\tilde{d} + \epsilon_d} - \dfrac{f \cdot b}{\tilde{d}(t - \delta t) + \epsilon_d}}{\delta t}. \tag{4.15}$$

By substituting equations 4.14 and 4.15 in equation 4.3, relative error in relative velocity in terms of disparity can be written as

$$RERV_{disparity} = \frac{\dfrac{\dfrac{f \cdot b}{\tilde{d}(t - \delta t) + \delta d + \epsilon_d} - \dfrac{f \cdot b}{\tilde{d}(t - \delta t) + \epsilon_d}}{\delta t} - \dfrac{\dfrac{f \cdot b}{\tilde{d}(t - \delta t) + \delta d} - \dfrac{f \cdot b}{\tilde{d}(t - \delta t)}}{\delta t}}{\dfrac{\dfrac{f \cdot b}{\tilde{d}(t - \delta t) + \delta d} - \dfrac{f \cdot b}{\tilde{d}(t - \delta t)}}{\delta t}}. \tag{4.16}$$

This can be simplified as

$$\boxed{RERV_{disparity} = -\frac{\alpha^2 + 2 \cdot \alpha + \alpha \cdot \beta}{1 + \alpha^2 + 2 \cdot \alpha + \alpha \cdot \beta + \beta}} \tag{4.17}$$

where α is $\dfrac{\epsilon_d}{\tilde{d}(t - \delta t)}$ and β is $\dfrac{\delta d}{\tilde{d}(t - \delta t)}$.

4.3.3 Comparing RERV in terms of velocity and RERV in terms of disparity

By comparing equation 4.4 and equation 4.17

$$-\frac{\alpha^2+2\cdot\alpha+\alpha\cdot\beta}{1+\alpha^2+2\cdot\alpha+\alpha\cdot\beta+\beta}=\frac{v_{m}^{rel}{}_{x}+(v_{ego_x}+\omega_x r_y)}{-(v_{ego_x}+\omega_x r_y)}. \tag{4.18}$$

Here there are the two unknowns α and β and one equation. But, the value of β (i.e. the ratio $\frac{\delta d}{\tilde{d}(t-\delta t)}$) is comparatively low and can be neglected. This is because when the depth between the ego car and the object detected is less then the disparity $\tilde{d}(t-\delta t)$ is high which means that β is low. When the depth is more then the change in disparity is very less which also further means that β is low. In the next chapter a simulation is done to validate this. Section 4.3.4 explains how β can be approximated if it is not neglected. The idea about whether neglecting or approximating β will be explained in the chapter 5 from the results of simulation.

By neglecting β in 4.18

$$-\frac{\alpha^2+2\cdot\alpha}{1+\alpha^2+2\cdot\alpha}=RERV_{velocity}. \tag{4.19}$$

As mentioned in the section 4.3.1 the value of $RERV_{velocity}$ is known and for simplicity, assume $RERV_{velocity}=\kappa$.

Then equation 4.18 can be written as

$$(1+\kappa)\alpha^2+2\alpha(1+\kappa)+\kappa=0. \tag{4.20}$$

The roots of this equation are

$$\alpha=\frac{-2(1+\kappa)\pm\sqrt{4(1+\kappa)^2-4\kappa(1+\kappa)}}{2(1+\kappa)} \tag{4.21}$$

$$\boxed{\alpha=-1\pm\frac{1}{\sqrt{1+\kappa}}}. \tag{4.22}$$

See section 4.4 for discussion on the roots of the equation.

4.3.4 Approximating β from measured disparities

The value of β the relative change in true disparity $\left(\frac{\delta d}{\tilde{d}(t-\delta t)}\right)$ is unknown. As the change in disparity δd is known (as $\delta d_m = \delta\tilde{d} = \delta d$), the only unknown is denominator $(\tilde{d}(t-\delta t))$. This can be approximated with measured disparity values as

$$\hat{\beta} = \frac{\delta d}{d_m(t-\delta t)}. \tag{4.23}$$

The impact of this approximation will be discussed in chapter 5.

Rewriting equation 4.18 without neglecting β gives

$$(1+\kappa)\alpha^2 + 2\alpha(1+\kappa) + \alpha\hat{\beta}(1+\kappa) + \kappa\hat{\beta} + \kappa = 0. \tag{4.24}$$

The roots of this equation are

$$\boxed{\alpha = \frac{-(2+\hat{\beta})(1+\kappa) \pm \sqrt{[(2+\hat{\beta})(1+\kappa)]^2 - 4\kappa(1+\kappa)(1+\hat{\beta})}}{2(1+\kappa)}}. \tag{4.25}$$

4.3.5 Calculating error in disparity from α

The value α is the relative error in disparity $\frac{\epsilon_d}{\tilde{d}(t-\delta t)}$. This value is known from the above procedure. Error in disparity can be calculated from α as follows:

The measured disparity equation 4.12, $d_m(t) = \tilde{d(t)} + \epsilon_d$ can also be written at time $t-\delta t$ as

$$d_m(t-\delta t) = \tilde{d}(t-\delta t) + \epsilon_d. \tag{4.26}$$

From equation 4.26 and $\alpha = \frac{\epsilon_d}{\tilde{d}(t-\delta t)}$ error in disparity can be given as

$$\boxed{\epsilon_d = \frac{\alpha d_m(t-\delta t)}{1+\alpha}}. \tag{4.27}$$

4.3.6 Relation between error in disparity and error in relative yaw angle between the cameras of stereo camera

As mentioned in the section 3.1, from equation 3.6
($\epsilon_d \approx \tan\beta \cdot f$), error in yaw angle γ can be given as,

$$\boxed{\gamma = \arctan\frac{\epsilon_d}{f}} \tag{4.28}$$

Thus, if ego absolute velocity, measured disparity and measured absolute velocity of the detected static object are known, then the de-calibration of the stereo camera due to error in yaw angle can be determined.

4.4 Roots of α

From equations 4.22 and 4.25 it can be noticed that there are two roots for α. The question which root of α to be considered for calculating error in disparity needs to be answered.

A simulation system (will be explained in chapter 5) is used to find out which root to be considered. Few known error in disparities (0.25 pixel, 0.4 pixel, 0.55 pixel) are introduced to a system where the ego vehicle to which the stereo camera is attached, is approaching an object that is static. The estimated error in disparity ϵ_d is calculated from the above algorithm using both positive and negative roots, using different absolute velocities of the ego vehicle. The figure 4.1 shows the behaviour of estimated error in disparity when found out using positive and negetive roots of alpha.

It can be seen from figure 4.2, that always the positive root of α is estimating the introduced error in disparity. It can also be observed that even if the ego velocity is varied it has no effect on the estimated error in disparity ϵ_d if positive root is considered. While on the other hand if affects the estimated error in disparity using the negative root.
Hence, it can be determined that the positive root of α must be considered for estimating a de-calibration of the stereo camera.

4.5 Constraints of the algorithm

There are some mathematical constraints to the algorithm which are to be noted. This algorithm cannot work:

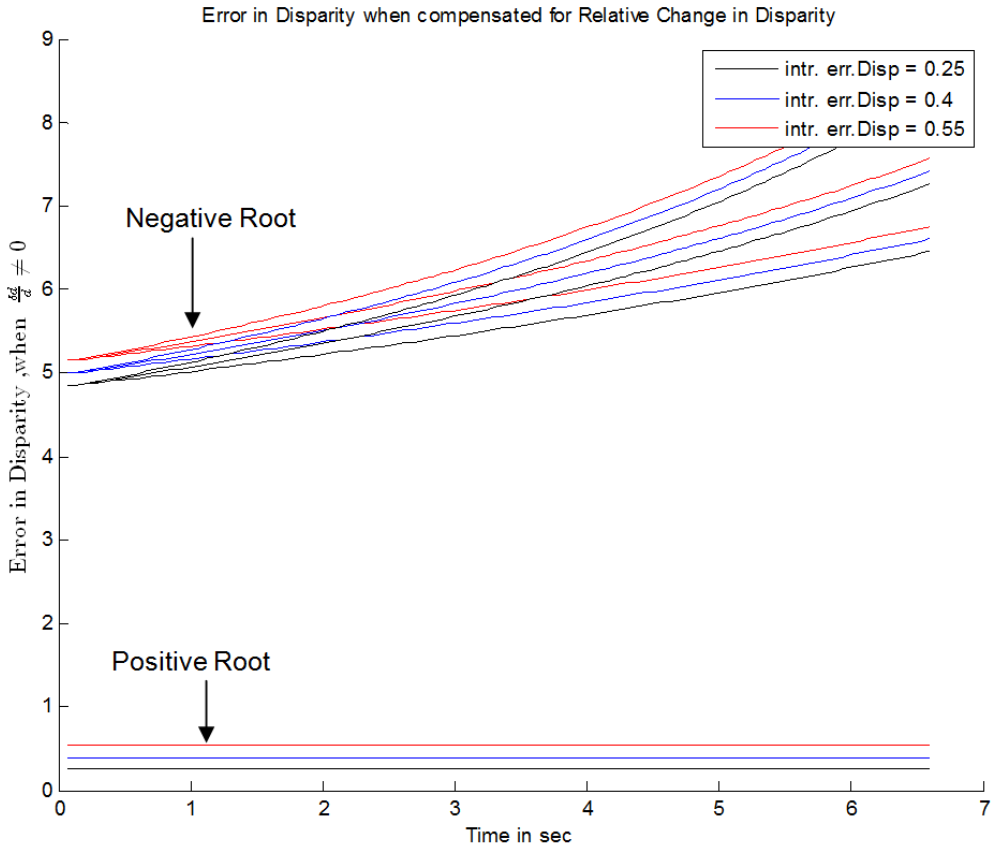

FIGURE 4.1: Variation of estimated error in disparity when using positive root and negative root and at different ego velocities

- When the ego velocity is equal to zero or close to zero ($v_{ego} \approx 0$).
- When $1 + \kappa$ is negative.

From equation 4.4 it is evident that when there is no yaw rate of ego car and when ego velocity tends to zero ($v_{ego} \to 0$), then the value of κ tends to infinity ($\kappa \to \infty$). From equation 4.22 it is known that α tends to -1, this in turn leads to the measure of error in disparity tending to infinity ($\epsilon_d \to \infty$) from equation 4.27.
Hence the algorithm cannot output a Yaw angle error when ego car is at rest.

From equation 4.22 and 4.25 is can be seen that when $\kappa <= -1$, there will be no real root of α. This means that there is no real solution for ϵ_d and error in yaw angle. That means κ has to be greater than -1 in order that the error in disparity ϵ_d is real and to detect de-calibration in the stereo camera.

The next chapter will explain the application of this algorithm on a simulation environment.

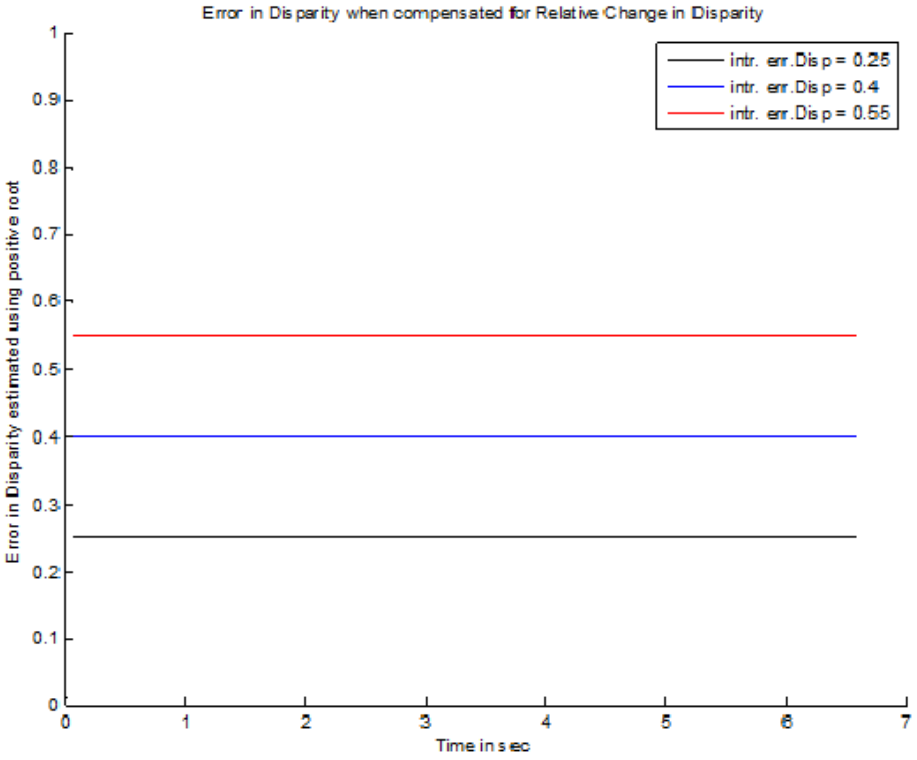

FIGURE 4.2: Estimated error in disparity when calculated using positive root of *alpha* and at different ego velocities

Chapter 5

Simulation

This chapter explains the simulation scenario used for testing the algorithm and gives an idea how good the algorithm able to estimate under ideal conditions. For the purpose of simulation the MATLAB environment is used.

The main objective of the simulation is to check the ability of the algorithm to estimate the error in disparity introduced. The simulation system is designed so that it follows all the assumptions that were considered for the algorithm (as mentioned in section 4.2). The advantage of doing a simulation is that behavior of different variables can be studied and better understood for various scenarios. It also helps to test the algorithm under different scenarios. For example in this chapter the result of the algorithm is tested for cases of β neglected and approximated from measured values.

5.1 System setup

An ego car with a known initial velocity is considered to be moving towards a detected object. The yaw rate ω of the ego car is neglected for the simulation. It is assumed that the ego car is a point object (i.e. length, width and height of the car are negligible). Also the stereo camera is located at this point object.

In a real world scenario, the stereo camera is present at a distance from the rear axle of the ego car (as the world co-ordinate system is present at it). The distances of the detected objects are only known from the stereo camera. Thus the knowledge of the distance between rear axle and stereo camera is needed to find out the object distances from the world co-ordinate system. As the ego car is considered to be a point object in the simulation this distance can be ignored.

An object that is detected by the stereo camera is assumed to be present at a known distance $D_{initial}$. The detected object is also a point object and is static (i.e. the absolute velocity of the object $v_{obj} = 0$). This setup is shown in figure 5.1

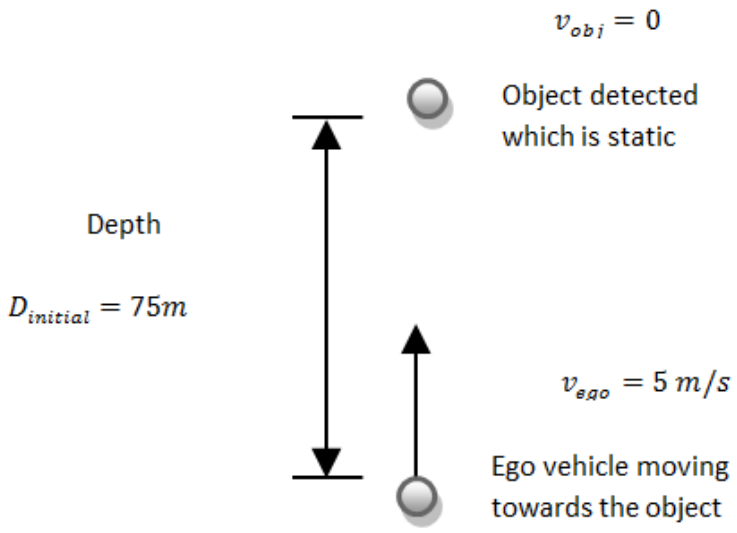

FIGURE 5.1: Simulation setup showing the ego car and object detected

The processing frequency of the stereo camera is assumed to be 15 Hz. That implies the change in time δt between two time stamps $t - \delta t$ and t will be approximately 0.066 sec.

5.2 Simulation of the scenarios

The simulation must be done in such a way that it tests the estimation of error in yaw angle under different scenarios. To achieve that, several scenarios of an ego car approaching the detected object are generated. This is done by introducing different motion types for the ego vehicle. For this system, ego motion is generated at constant velocity (i.e. acceleration $a_{ego} = 0$), constant acceleration and varying acceleration. This ensures that estimation of error in yaw angle is tested for different functions of depth between ego car and detected object.

If a_{ego} is the acceleration of the ego vehicle, then the kinematic equations for the above three cases can be given as

Constant velocity:

$$v_{ego}(t) = v_{ego}(t - \delta t) \tag{5.1}$$

Constant acceleration:

$$v_{ego}(t) = v_{ego}(t - \delta t) + a_{ego}t \tag{5.2}$$

Varying acceleration:

$$v_{ego}(t) = v_{ego}(t - \delta t) + \int_{t-\delta t}^{t} a_{ego}(\tau)\mathrm{d}\tau \tag{5.3}$$

5.2.1 Generation of true disparity and measured disparity information

For each motion type, the true depth information between the ego vehicle and the detected object is retrieved from the absolute velocity of the ego vehicle as follows

$$\tilde{D}(t) = \tilde{D}(t - \delta t) - v_{ego}(t) \cdot \delta t. \tag{5.4}$$

This depth information is then converted to disparity to give true disparity as follows

$$\tilde{d} = \frac{f \cdot b}{\tilde{D}}. \tag{5.5}$$

A known error in disparity $\hat{\epsilon}_d$ is introduced to the above true disparity $\tilde{d}$ to get measured disparity:

$$: d_m = \tilde{d} + \hat{\epsilon}_d. \tag{5.6}$$

Measured depth D_m is then generated from the measured disparity values d_m as

$$D_m = \frac{f \cdot b}{d_m}. \tag{5.7}$$

5.2.2 Inputs to the algorithm

The true and measured disparity, true and measured depth generated for each motion type are inputted to the algorithm along with the velocity information. The estimated error in disparity ϵ_d is then compared to the introduced error in disparity $\hat{\epsilon}_d$.

For the results mentioned in the section 5.3 the following initial values are used

$$\begin{aligned} \hat{\epsilon}_d &= 0.25 \text{ pixels introduced error in disparity} \\ D_{initial} &= 75 \text{ metres} \\ \delta t &= 0.066 \text{ seconds} \\ v_{ego} &= 5\ m/s \\ a_{ego} &= 1.5\ m/s \text{ (for constant acceleration case)} \\ \frac{da_{ego}}{dt} &= 0.05\ m/s^3 \text{ (for varying acceleration case)} \end{aligned}$$

5.3 Results of the simulation

5.3.1 Depth and disparity variation

The aim of this section is to give a brief idea of variation of depth and disparity for one of the motion cases. Following are the depth vs time plot, disparity vs time plot for the case ego vehicle is moving with constant acceleration towards the detected object as given by the equation 5.2. The results of depth/disparity corresponding to other motion models for ego vehicle are present in the appendix section A.1.

Behaviour of depth with respect to time is shown in the figure 5.2.

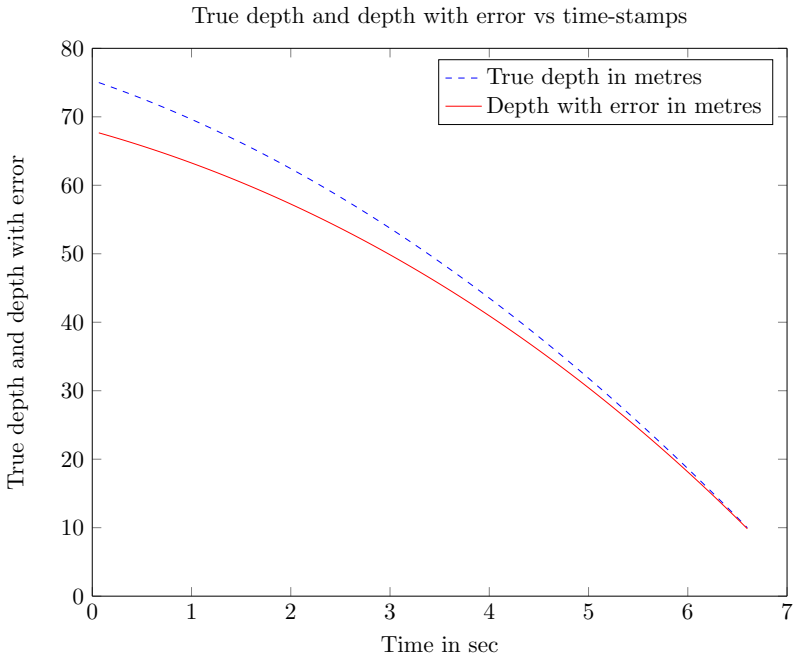

FIGURE 5.2: Variation of true depth and measured depth when ego vehicle is moving with constant acceleration

The corresponding disparity vs time plot is shown in figure 5.3.

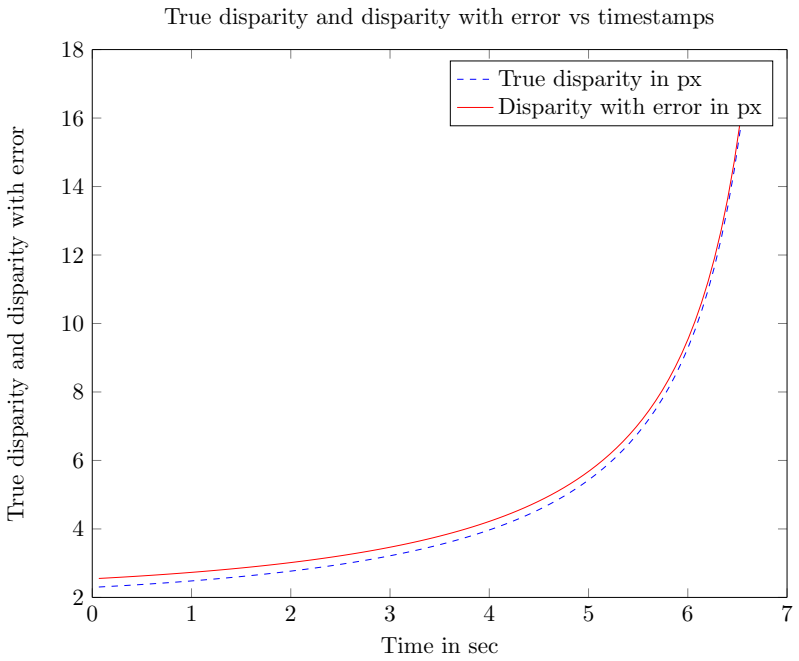

FIGURE 5.3: Variation of true disparity and measured disparity when ego vehicle is moving with constant acceleration

Here depth/disparity with error corresponds to measured depth/disparity. From figure 5.3 it can be observed that the introduced error in disparity is constant over time. Also it can be seen from the figure 5.2 that the depth error generated is not constant when the depth is decreasing as mentioned in section 3.2 in chapter 3.

5.3.2 Estimation of error in disparity if the detected object is static

This section gives the results of the estimation of error in disparity when β is neglected and when β is approximated from measured disparities as mentioned in the sections 4.3.3 and 4.3.4. As discussed earlier the question about whether β should be neglected or approximated from the measured disparities is still open. This section provides the answer to it. The results are presented for the case where ego vehicle is moving with constant acceleration towards detected object. The results for other motion types are presented in appendix section A.1.

5.3.2.1 Estimated error in disparity when β is neglected

Figure 5.4 shows behaviour of estimated error in disparity when the ego vehicle is moving with constant acceleration and when β is neglected. It can be seen from figure 5.4 that

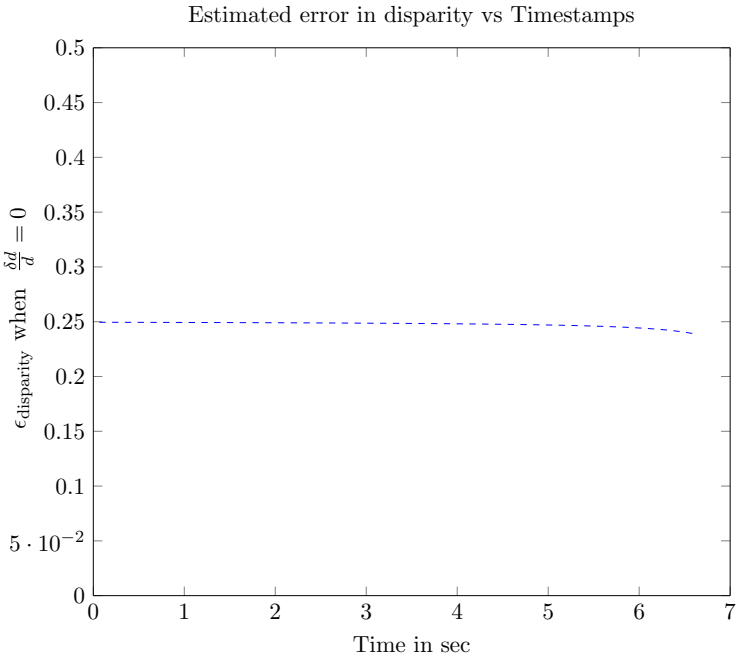

FIGURE 5.4: Behaviour of estimated error in disparity when the ego vehicle is moving with constant acceleration and β is neglected

the algorithm is able to sufficiently estimate the introduced error in disparity of 0.25 pixels .

The deviation in the error in disparity from introduced disparity error can be seen in the figure 5.5. It can be observed that the deviation of estimated error in disparity is increasing as the depth decreases (time increases). This is because the change in disparity (δd) increases as the depth between ego and object decreases and even though d true disparity also increases, the value for the term $\frac{\delta d}{d}$ increases in the equation 4.18. As $\beta = \frac{\delta d}{d}$ is neglected, the estimation of error in disparity deviates from the introduced error in disparity as the depth decreases.

From the figure it can be observed that even though β is neglected the estimated error in disparity is not affected significantly.

5.3.2.2 Estimated error in disparity when β is approximated

Figure 5.6 shows the behavior of the estimated error in disparity when the ego vehicle is moving with constant acceleration and when β is approximated with measured disparities as mentioned in section 4.3.4.

Figure 5.7 shows the deviation of the estimation of error in disparity from introduced disparity error.

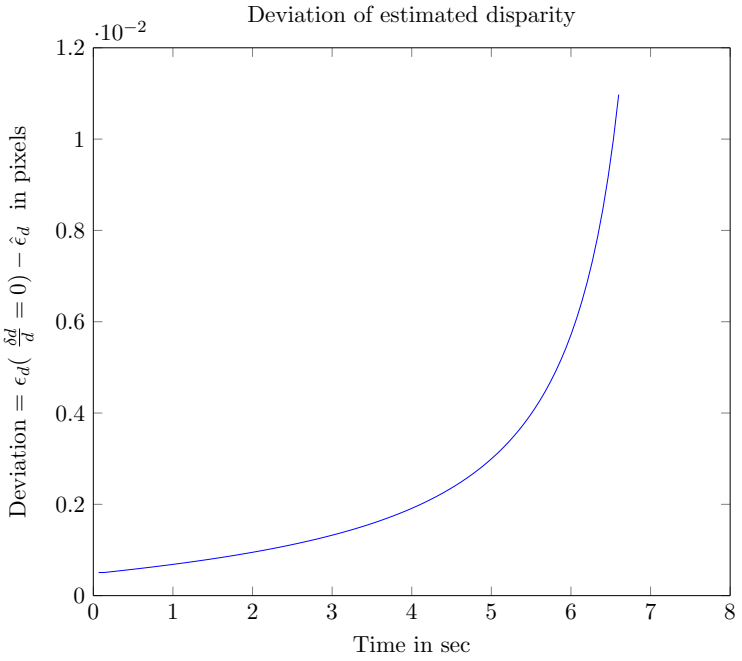

FIGURE 5.5: Deviation of estimated error in disparity from introduced error in disparity when β is neglected

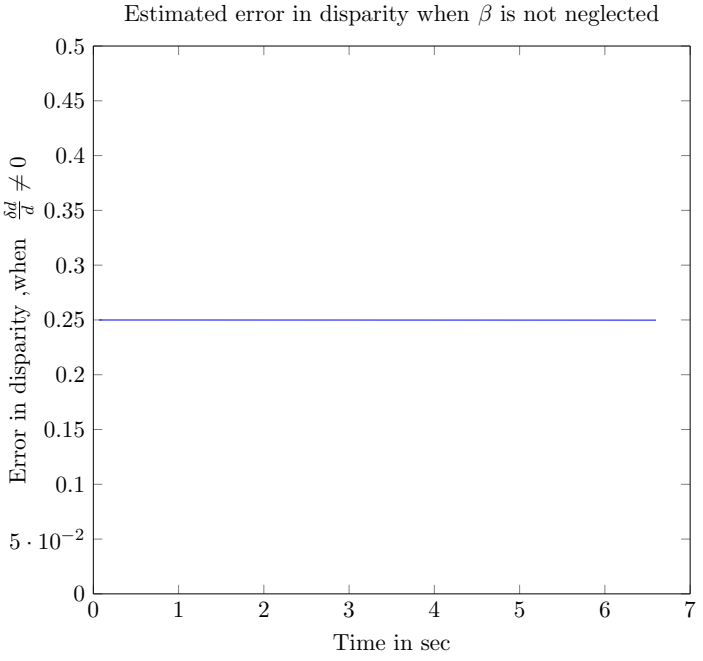

FIGURE 5.6: Behaviour of estimated error in disparity when ego vehicle is moving with constant acceleration and β is approximated

From figures 5.6 and 5.7, it can be inferred that:

- Approximation of β from measured disparities is plausible as the estimation of error in disparity stays close to the introduced error in disparity.

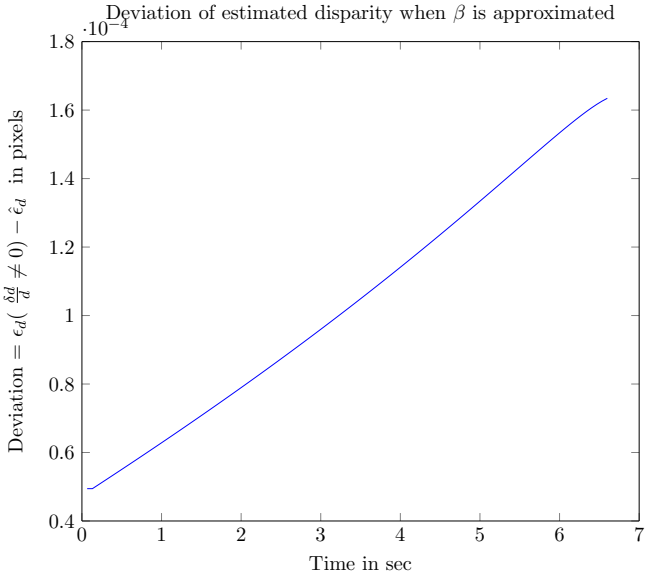

FIGURE 5.7: Deviation of estimated error in disparity from introduced error in disparity when β is approximated

- The estimation of error in disparity when β is approximated is better than estimation when β is neglected as even when the depth is decreasing(time increases), the deviation of error in disparity is less compared to the deviation when the β is neglected.

5.3.3 Estimation of error in disparity if the detected is not static

It is assumed that the detected object is static for the before mentioned simulation results. But in a general scenario there are always objects that are non-static. Hence it is needed to check the performance of the estimation of error in disparity when the object is not static.

For the same simulation environment used before, an absolute velocity for the detected object of 0.5 m/s is introduced. Hence, it no longer confines to the assumptions made. The figure 5.8 shows the behaviour of estimation of error in disparity when the detected object is not static and β is approximated.

From figure 5.8 it can be observed that the estimation is no longer good as the deviation from introduced disparity error is significant. So it can be determined that the estimation results from a non-static object are not desirable and the algorithm should be applied for static objects only.

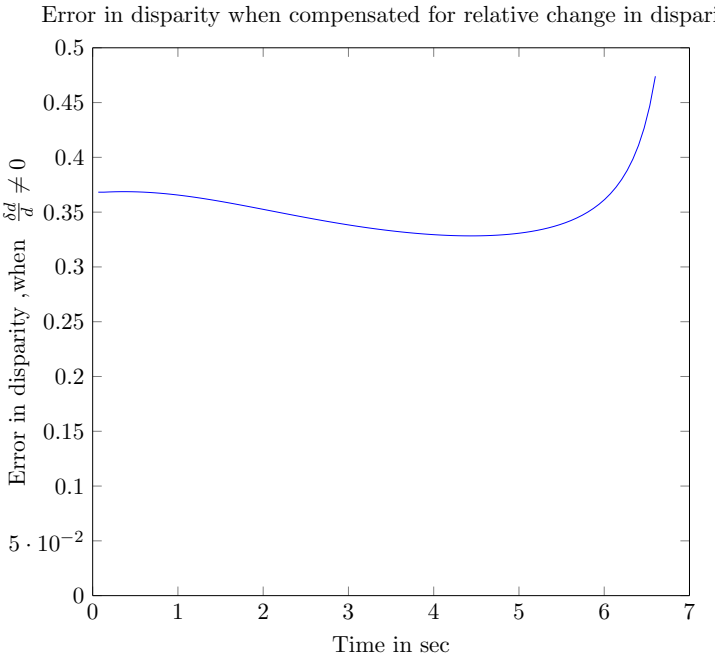

FIGURE 5.8: Behaviour of estimated error in disparity when object is not static

5.4 Summary

Under the simulated conditions, it is observed that the error in disparity is sufficiently estimated even when β is approximated from measured disparities.

But a real world scenario is different from the simulated environment, where there can be multiple objects present. Not all the objects detected are static. It can also have measurement errors in the values of ego velocity, yaw rate. In the next chapter, it will be presented how these challenges are confronted.

Chapter 6

Application to real world data

From the previous chapters it known that from a ground truth assumption that all the objects detected are static, it is possible to detect the de-calibration present in the stereo camera. It is also known that the estimation of the de-calibration (i.e. yaw angle error) of the stereo camera is good under the simulated conditions.

The aim of this chapter is to find out whether the algorithm can be used for determining the de-calibration of stereo camera in a real world scenario and how the algorithm can be extended to the real world data. This chapter also presents the results of testing the algorithm on an active system used for driver assistance.

6.1 Real World scenario

The stereo camera attached to the ego vehicle collects and processes the image information at a particular frequency. So for every specific time interval the measured motion attributes about the objects and motion information of the ego car are known.

Unlike the simulation environment, a general traffic scene consists of vehicles that are moving, pedestrians, road side objects like road signs, guard rails, traffic signals etc. Hence there will be multiple objects that might be detected for each frame. These objects can be static or non-static. Also, from frame to frame the number of objects changes. The algorithm must be adapted for this scenario.

In order to approximate $\beta = \frac{\delta d}{d(t-\delta t)}$ from measured disparities (for estimation of error in disparity mentioned in section 4.3.4), the measured information from two consecutive frames for the same object is required. Since in a real world scenario, the objects that are present in one frame can be different from another frame the tracking information

of the object like corresponding objects present in the two frames is required. This information is available from the existing system. The details of how tracking is done is not the focus of this thesis.

A depiction of a general scenario with the different types of objects is shown in the figure 6.1.

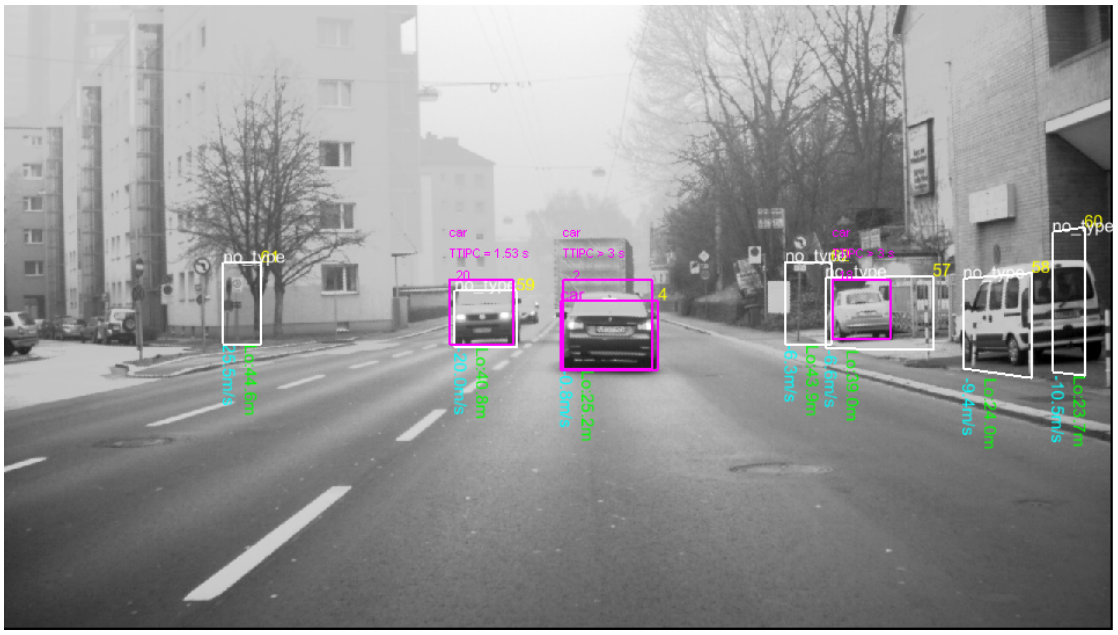

FIGURE 6.1: General scenario in real world with detected objects

6.2 Error in disparity in terms of distance error

Error in disparity ϵ_d can be expressed in terms of distance error D_{error} ($D_{error} = D_m - \tilde{D}$) with respect to a reference depth D_{ref}. It is given by

$$D_{error} = -\frac{D_{ref}^2 \cdot \epsilon_d}{f \cdot b + \epsilon_d \cdot D_{ref}}. \tag{6.1}$$

Consider camera constant $f = 1400$,stereo base $b = 0.12$ m and a reference depth of $D_{ref} = 30$m,
then for an error in disparity of 0.25 pixels there exists an approximate distance error of 1.28 m at 30m.
Depending on the requirements of the safety system, there is a tolerance to this distance error. From now on this thesis assumes that a distance error of 1.28 m (i.e 0.25 pixel disparity error) is tolerable. That implies, given reference data (explained in section

7.1), the difference between estimated error in disparity and reference error in disparity must be less than 0.25 pixels.

6.3 Real world application

As discussed in the section 6.1, a real world traffic scenario will have objects that are static and that are non-static. For each of these objects in a particular frame, the error in disparity can be estimated by the algorithm. But from the previous chapter, section 5.3.3, it is known that non-static objects in the scene will lead to wrong estimation of error in disparity. Hence the question prevails, what should be done to avoid the wrong estimation of error in disparity due to non-static objects?

It is not reliable to consider the output of the algorithm from a single object as an estimate of the error in disparity as this object can be static or non-static. It is also known that there are several objects in the scene in a frame. There are exists the same objects for several frames. If all the objects are assumed to be static then the output of the algorithm for each object must be equal. This deduction can be used to solve the issue of wrong estimation.

6.3.1 Method

Assume that a scene contains more static objects than non-static objects. If a histogram is constructed for all the estimated errors in disparity for each object over some frames then the location of the peak of this histogram (Maximum frequency of error in disparity) gives an estimate of the error in disparity. This method is explained below.

- Collect the input data for each frame.
- For each frame calculate error in disparity for each object present in it.
- Collect this calculated errors in disparity until the amount of values reaches a certain threshold (explained in section 6.3.2). So if each frame has 5 objects and the threshold is 50 estimations of error in disparity, then 10 frames of data is needed.
- After the threshold is reached, construct a histogram over the estimated errors in disparity.
- As the scene is assumed to be static, the dominant the peak of the histogram should give the correct error in disparity. Convert this error in disparity to a yaw angle error between the camera to find the de-calibration of the stereo camera.

A flow chart for the above method is shown in figure 6.2.

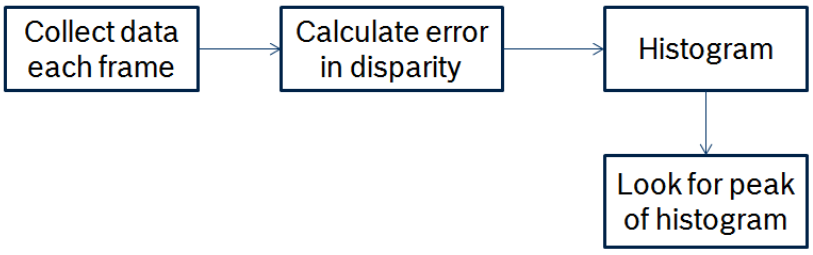

FIGURE 6.2: General method for applying algorithm to real world data

6.3.2 Collection of estimated error in disparity

In order to construct a histogram for finding de-calibration of stereo camera, as mentioned before collection of estimated errors in disparity is required. But the data on which the histogram is constructed must be a representation of the current scenario. So, collecting the estimated error in disparity data must be done in such a way that it allows room for new data and throws the old data without loosing the information from the past data. Also, the output for the above mentioned method must be available for every frame. Applying a sliding window to the data allows to achieve this. Sliding window to the data involves throwing one old value at a time and adding one new value at the same time. It maintains the same number of the overall data present in a window at all times. Figure 6.3 shows how windowing is done. The threshold mentioned in section 6.3 in this context will be the size of the window being used for windowing.

As the collection of data is needed, depending on the window size the first output for the method for calculating correct error in disparity is delayed. Choosing a window size will be explained in section 6.8 of this chapter. But during the explanation of the method in the further sections the histogram will be constructed over all the data that is available, this is to remove the dependency of window size on the output (will be explained in section 6.8).

6.4 Construction of the histogram

For the error in disparity data present in each instance of the sliding window explained above, a histogram is constructed. The peak of the histogram is given by the maximum of the frequencies in the histogram. The figure 6.4 shows such a constructed histogram for a real world scenario which is static object dominant. The selection of the bin size for the histogram will be explained in the section 6.9.

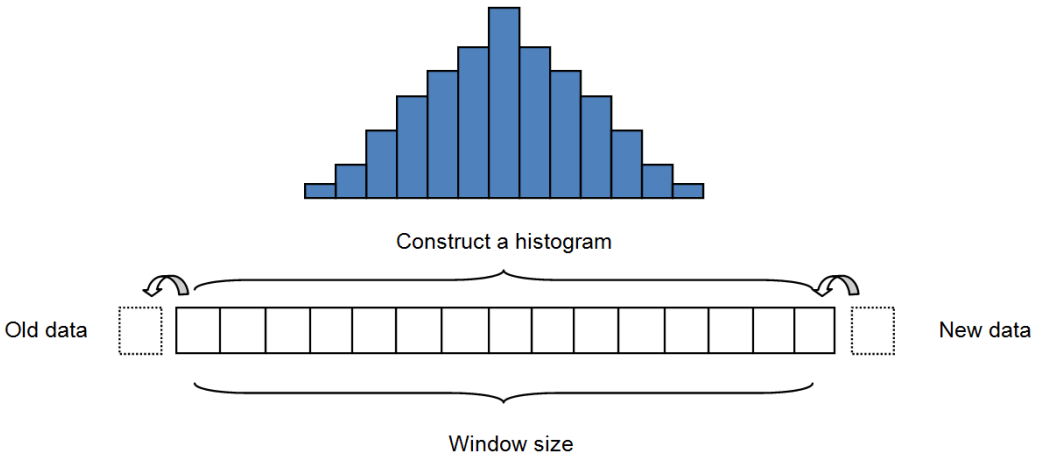

FIGURE 6.3: Applying a sliding window approach to the data

This histogram has the error in disparity data estimated both from static objects and non-static objects. As it is assumed that the scene has more static objects, it can be inferred that the data near to the peak of the histogram (shown in the figure 6.4 by *) is from the static objects (inliers) and the data that are away from the peak of the histogram are from the non-static objects (outliers). Because the estimation of the errors in disparity is almost same for all the static objects (as it considered assumption stays true) and form a peak in the histogram, while the errors in disparity estimated from the dynamic objects are varying (as it violates the assumption that objects are static) and are spread away from peak of the histogram. A range of [-10,10] px disparity error is considered and it is assumed that any error over this range is not plausible.

6.4.1 Smoothing of the histogram

The peak of the histogram is very important for the estimation of the error in disparity. But there is a possibility that the peak of the histogram can be varied significantly because of the noise in the measurement data. As small variations in the measurement of ego velocity or the measurement of object motion data may vary the estimated error in disparity.This in turn affects the peak of the histogram to vary from one bin to another. Hence smoothing of the histogram is needed. This thesis uses a Gaussian smoothing over the histogram. Here a 1-D Gaussian filter with size of 5 bins is considered. The Gaussian with a mean of 0 and the standard deviation of 5/6 of bin size is considered

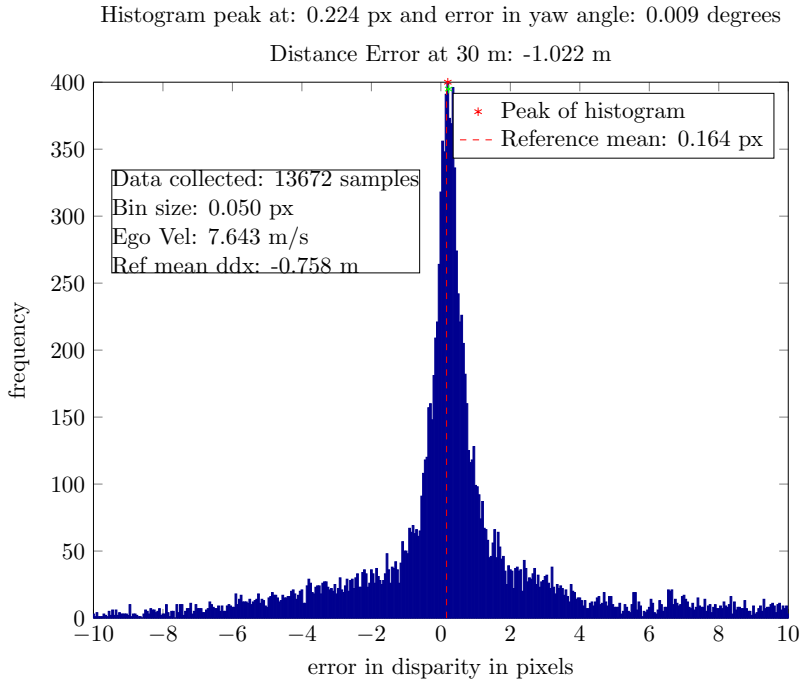

FIGURE 6.4: Constructed histogram over estimated error in disparities over a span of time

(3-sigma Gaussian). The smoothed frequencies of the histogram shown in figure 6.4 is shown in figure 6.5.

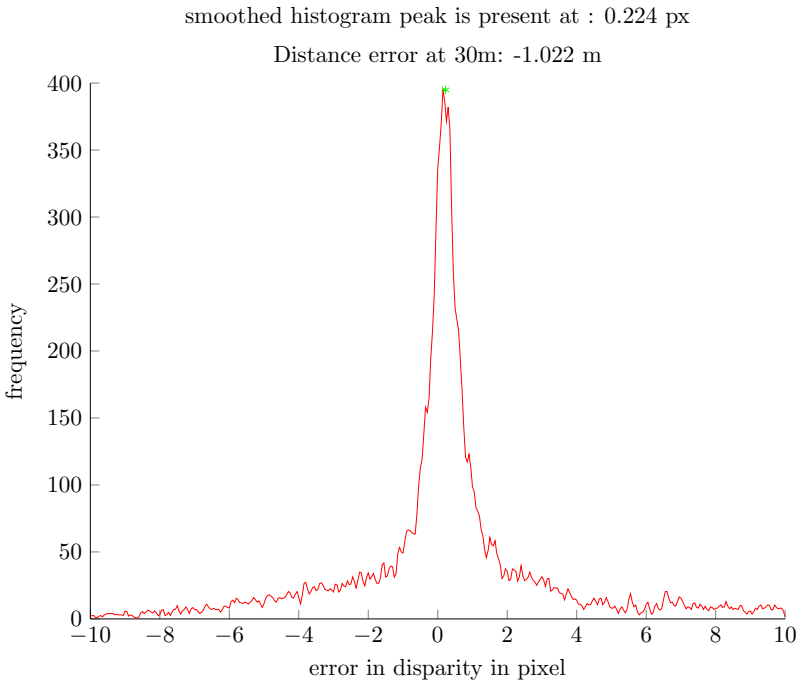

FIGURE 6.5: Smoothed frequencies of the histogram shown in 6.4 by using Gaussian filter

6.4.2 Weighing the peak bin and its neighbors

Even after smoothing and determining the peak bin of the histogram, the actual peak need not to be at the center of the bin. In order to estimate the location of the peak inside the bin, weighing the peak bin and its adjacent neighbors is required. In this thesis the peak bin and the immediate two neighbors are considered for weighing. Based on the frequencies and their bin center location, the centroid of these bins is considered as the approximate location of the peak. For example, if $f1, f2, f3$ are the frequencies of the bins and $b1, b2, b3$ are the bin centers correspondingly. Where $f2$ is the peak frequency and $b2$ is the bin center of the peak, then the approximated location of peak b' is given by:

$$b' = \frac{f1 \cdot b1 + f2 \cdot b2 + f3 \cdot b3}{f1 + f2 + f3} \tag{6.2}$$

6.5 Discussion on estimation of error in disparity

Reference error in disparity can be calculated from the data provided by a laser range finder (explained in section 7.1). From the figure 6.4 it can be seen that the mean of reference error in disparity is 0.164 pixel and the peak of the smoothed histogram is at 0.224 pixel. The difference between reference data and estimated peak is approximately 0.06 pixel or -0.28 m of distance error at a reference distance of 30 m. This is less than the tolerance which was discussed before, hence the estimation is good.

6.6 Error in disparity histogram when the scene is dominant with non-static objects

In contrast to the assumption made above that the scene is static object dominant, if the scene has more moving objects than the static objects, then the above histogram constructed might not be useful to estimate the error in disparity. It is expected that there are more outliers than the inliers (which are explained in section 6.4). That implies a proper peak in the histogram cannot be identified. Consider a case where a moving car is present in front of the ego car, in the field of view of stereo camera for a longer period of time. This allows the histogram constructed to have two significant peaks (a bimodal histogram). One peak corresponds to the static objects in the scene, while the other corresponds to the moving object in front of the camera. As estimation of error in disparity from the object car moving in front also remains constant for a longer period

of time (similar to static objects). Unlike other dynamic objects where the estimated error in disparity also constant for each object but as they exist in the field of view for shorter period of time they do not generate a peak in the histogram. The histogram for such a scenario can be seen in figure 6.6.

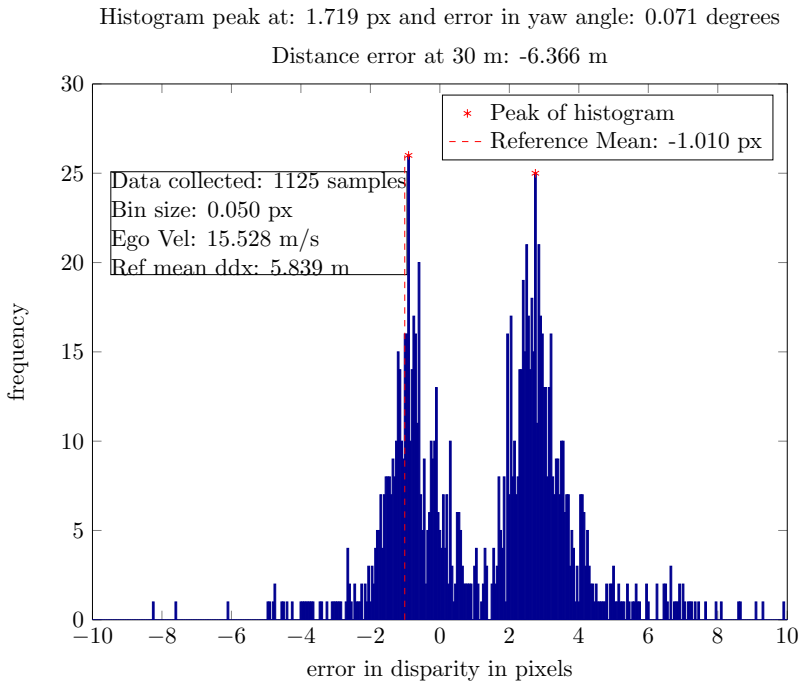

FIGURE 6.6: A bimodal histogram generated when there is a moving object present in the view of the stereo camera for a long time

It can be seen from figure 6.6 that there are two significant peaks of the histogram. If just the maximum of the frequencies is considered for the estimation of error in disparity then either of the peaks can contribute to the estimation. It cannot be known for certain what is the correct de-calibration of the camera.

6.7 Classification of static and non-static objects

A solution for a scenario mentioned above is needed which is independent of the assumption that the scene is static object dominant. That implies a classification of static and non-static objects is required. So that the algorithm can be only processed for static objects neglecting the non-static objects. This way the assumption that all the objects in the scene are static is no longer required.

6.7.1 Threshold on measured absolute velocity of the object

An idea for classification is to use a threshold on the measured absolute velocity. This involves checking the measured absolute velocity of each object. If it is greater than the threshold considered then it belongs to non-static objects. Otherwise it belongs to the group of static objects. i.e.

$$v_{obj}^{abs} > \theta \text{then that object belongs to non-static objects} \tag{6.3}$$

Consider an ego vehicle is traveling at a velocity of 2 m/s and a object is measured to have -8 m/s (moving towards the ego vehicle) but it is actually static. Then the resultant yaw error would be -0.22 degrees. This is determined as explained before from the equations 4.27 and 4.28. So if a threshold on absolute velocity of the object is made at ±8 m/s then any calibration error in yaw angle greater 0.22 degrees cannot be detected. Also, if an object is having an absolute velocity less than 8 m/s but moving in the field of the view of the stereo camera for a longer time then the problem stated above still persists.

Hence it is not a good option to use a threshold on the absolute velocity of the detected object.

6.7.2 Using the object type classifier information

Another approach is by using an image based classifier the object determined to be belonging to one of the training classes. Each class of the training class is given a type label. In this context these labels are car (OBJECT_CAR), truck (OBJECT_TRUCK), pedestrian (OBJECT_PEDESTRIAN) or no type (OBJECT_NO_TYPE). The object is then determined by the classifier to be belonging to one of the class labels. [12]

Once the type of the object is known the classification for static and non-static objects can be approximated by using prior knowledge about the scene. In general traffic scenario objects that belong to car, truck or pedestrian move. The objects that belong to no class are generally road side objects like traffic signals, guard railings etc. These are static. Using this inference, the static and non-static objects can be classified. This classification applied to the error in disparity histogram can be seen in the figure 6.7. The frequencies in green belong to objects that are of no type and the frequencies that are in blue color belong to objects that are of type car, truck and pedestrian.

So, if all the objects that belong to OBJECT_NO_TYPE are considered then the resulting error in disparity histogram would be shown in the figure 6.8

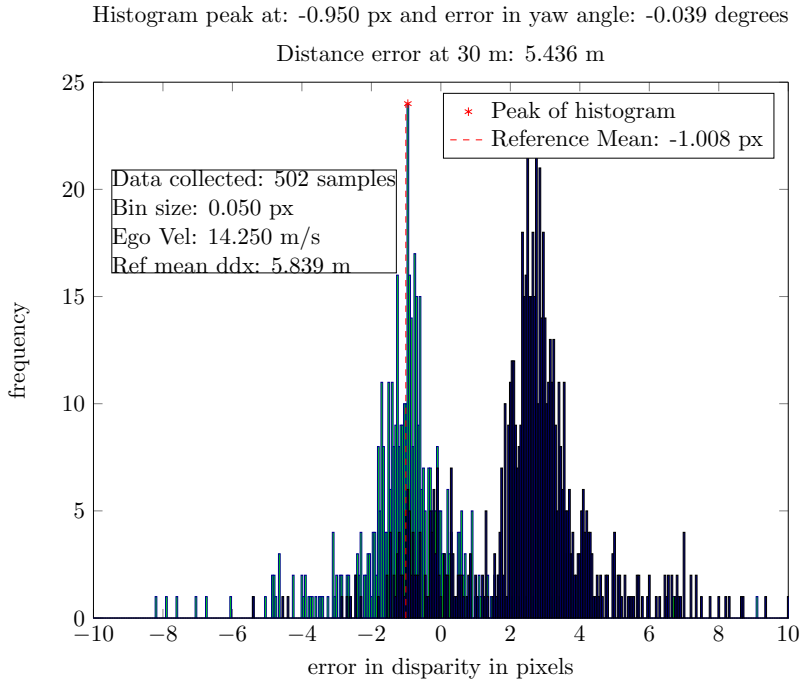

FIGURE 6.7: Bimodal histogram generated after classification based on error in disparity data from static and non-static objects

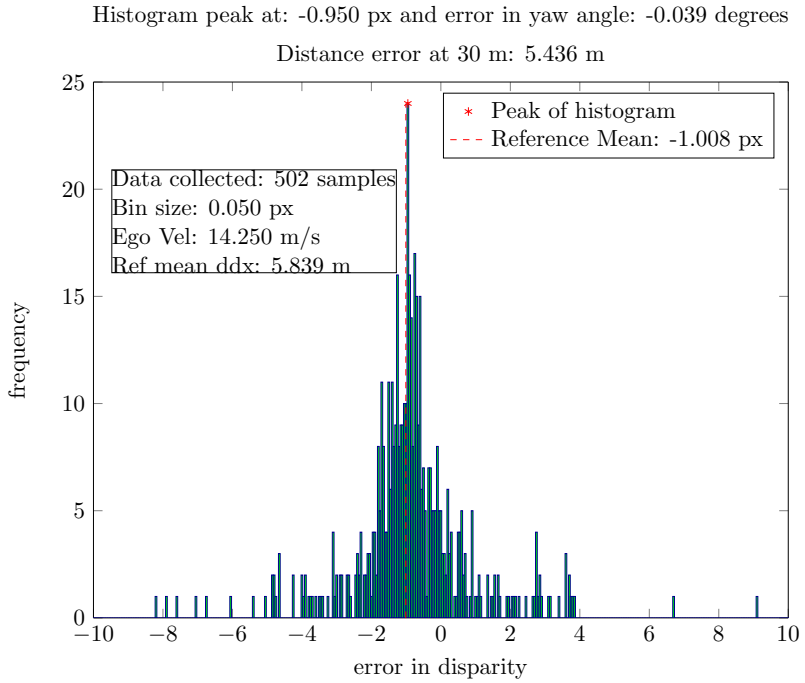

FIGURE 6.8: Bimodal histogram generated after filtered for error in disparity data from static objects

It can be observed that now the histogram has only one significant peak that corresponds to the data mainly from static objects. It can also be observed that the overall amount of data is decreased considerably. This is a disadvantage by the method of classification

as the time required to collect for the required amount data increases.

6.8 Determining the optimal window size

The window size plays a crucial role in the estimation of the error in disparity. The size of the window must be in such a way that the time taken for the first output is less, but the data in window must be sufficient enough to provide a proper peak of the histogram.

In order to determine the optimal size of the window, a test is conducted with different window sizes on the data from a same scenario. From each instance of sliding window a peak of histogram for error in disparity is obtained. An optimal size for a window can be defined as the size for which the standard deviation of the output from all the windows over time is low. The result for the test conducted is shown below. A mean standard deviations in error in disparity is the mean of the standard deviations for about 30 sequences.

Window size	Mean standard deviation in disparity error	distance error at 30m
500	0.3930 px	1.7 m
1000	0.1578 px	0.72m
1500	0.1220 px	0.56m
2000	0.1056 px	0.49m
2500	0.0901 px	0.42m
3000	0.0778 px	0.36m

It can be observed that the standard deviation decreases as the size of the window increases. It can also be observed that there is a convergence in the mean standard deviation at window size of 2000 bins. An optimal size for the window depends on the accuracy and the time for first output that is required. It depends on the application. For this thesis and its application on a active system a window size of 2000 bins is chosen.

6.9 Determining the optimal histogram bin size

Choosing a bin size for the histogram is a balance between the accuracy needed and how good the histogram is the portraying the data. If the histogram bin size is high the accuracy of the algorithm is significantly low. As the true error in disparity from the histogram can be anywhere in the maximum frequency bin not necessarily in the center

of the bin. If the histogram bin size is very less then the histogram will result in many local maximum peaks instead of one significant peak.

One way of determning the optimal bin size is to check with different bin sizes on the data. The result of such test is show in the figures 6.9,6.10 and 6.11:

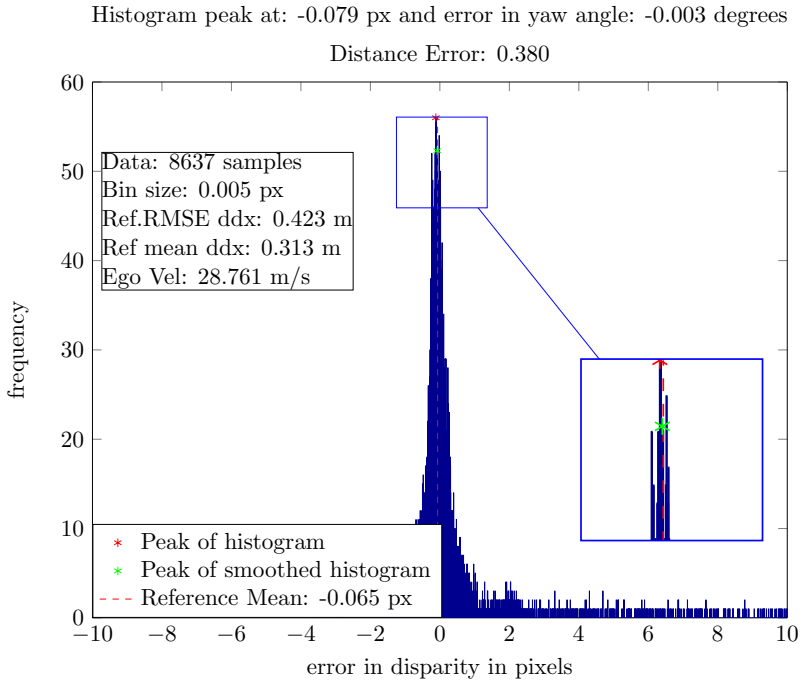

FIGURE 6.9: Histogram when a bin size of 0.005 pixel is chosen

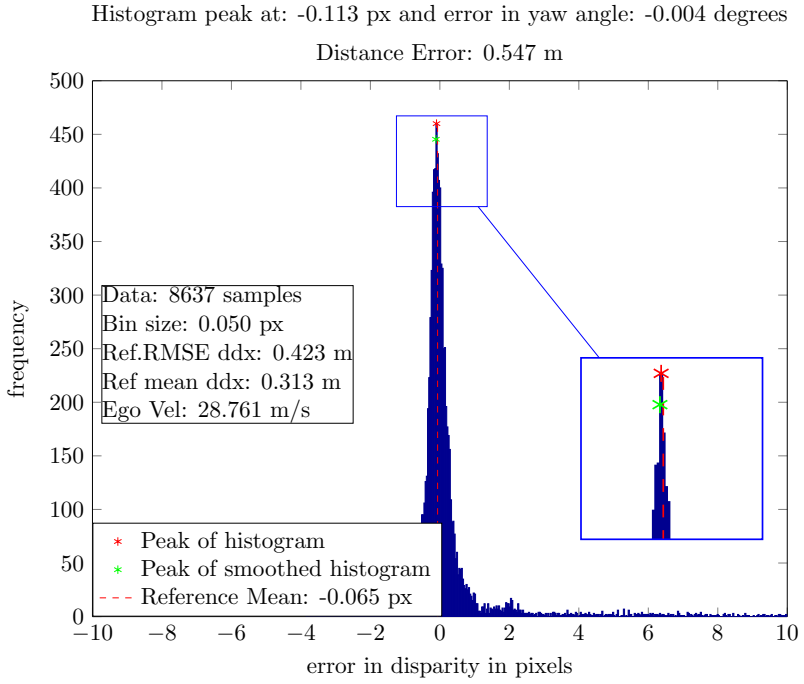

FIGURE 6.10: Histogram when a bin size of 0.05 pixel is chosen

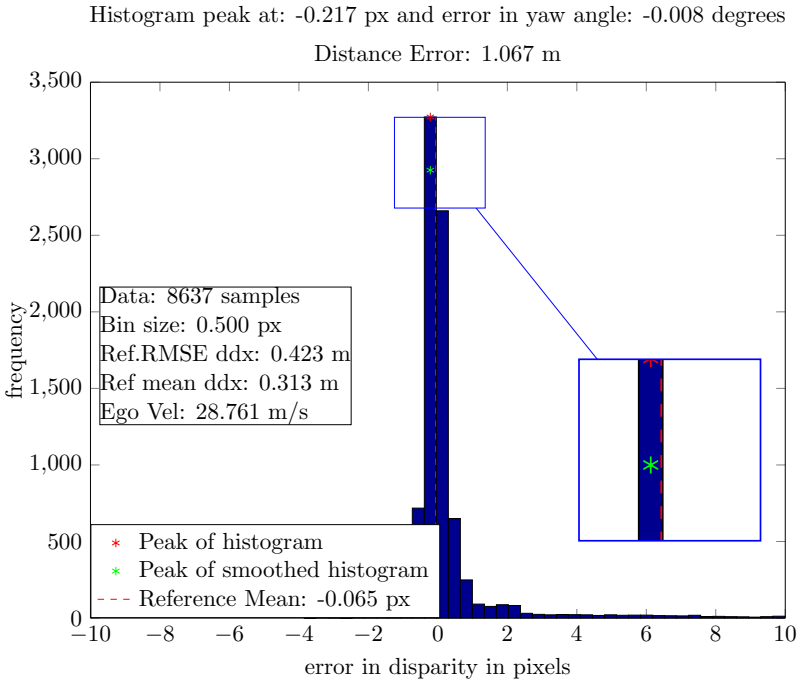

FIGURE 6.11: Histogram when a bin size of 0.5 pixel is chosen

It can be observed that in the figure 6.9 the number of bins is too high and the bin size is very low and hence it gives the many local peaks in the histogram. In figure 6.10 the bin size is too high and hence the accuracy of the algorithm is affected. In figure 6.11 the number of bins is not too few or too high and is a balance between the two factors mentioned. For this thesis a bin size of 0.05 pixel is used.

6.10 Summary

The algorithm is adapted for a real world scenario and the problems that are present during the adaptation are discussed and solution is presented. Optimal window size and the histogram bin size have been discussed. In the next chapter the results of the algorithm under different scenarios will be presented along with the comparison of results with reference data.

Chapter 7

Results and Discussion

This chapter will present the results of the algorithm presented in chapter 6 applied to different situations like in city scenario, night time and highway scenarios.

7.1 Validation

The results from the algorithm have to be validated against some reference data. This data is generated from a laser range finder. A laser range finder uses a laser beam to find the distance of the object. It uses the time of flight principle to find out the distance of the object. Based on the time taken by the laser beam from its emission to being detected again after hitting a reflecting surface, the distance is determined. The precision of the laser range finder that uses very sharp laser pulses and has a very fast detector can range an object to within a few millimeters. [13]

This distance data is compared to measured distance by the stereo camera and a reference distance error is determined. This reference distance error is used to validate the results from the algorithm presented in chapter 6.

7.2 Reference Data

The reference distance error mentioned in the section 7.1 is the mean of the distance error obtained from all objects at various depths interpolated (as distance error varies over the reference distance) to a reference distance of 30 m. That is the objects that are present any distance other than 30m then the distance errors for those objects are converted to disparity errors . Then the mean disparity error is found out and then it is further converted back to disparity error at 30m by using the equation 6.1. For

example, if there are two objects at 20m and 40m with distance errors -1.5m and -4m then a distance error at 30m can be found as:

- Convert distance errors into disparity errors using the equation 6.1, i.e. -1.5m at 20m and -4m at 30m will be equal to 0.68 px and 0.46 px error in disparity (if $f = 1400$ and $b = 0.12$ m).
- Find the mean of the disparity error i.e. $\bar{d}_{\text{error}} = 0.57$ px.
- Convert the mean disparity error in to distance error at 30m using the formula 6.1. So, $D_{\text{error at 30m}} = -2.78$ m.

This mean disparity value will be affected by the wrong corresponding object establishment between reference data and the measured data. That is the comparison of measurement data and reference data of two different objects. Hence, instead of comparing the result of the algorithm to such single reference mean value, in this thesis the result is compared to all the individual reference distance error data from every object at every time instance.

An example of such reference data comparison is shown as reference distance error vs reference distance graph as in the figure 7.3. Here the blue data corresponds to the distance error of all the objects in the scene at various distances at various time instances. The data marked with red circle corresponds to the data of the distance error of objects that are labelled as OBJECT_NO_TYPE. The mean distance error is plotted as the yellow line in the figure. The green line is the estimated distance error plot for various reference distances. It is calculated as follows:

The estimated error in disparity value is converted to the distance error for various reference distances by using the formula 6.1.

Here the reference distance D_{ref} is varied so that, for a given error in disparity the distance error $D_{error} = D_m - \tilde{D}$ where D_m is the measured depth and $\tilde{D}$ is the true depth, is obtained for every reference distance.

Behavior of reference data

It can be observed from the figure 7.3, that the reference has following a quadratic behavior. This can be explained by a simulation of reference distance error for various reference distances, given error in disparity is positive and when it is negative. The result of such simulation is shown in the figure 7.1. It shows that when a negative error in disparity is present the distance error starts from 0 and continuously increases as the reference distance increases. When the error in disparity is positive the distance error starts from 0 and continuously decreases as the reference distance increases.

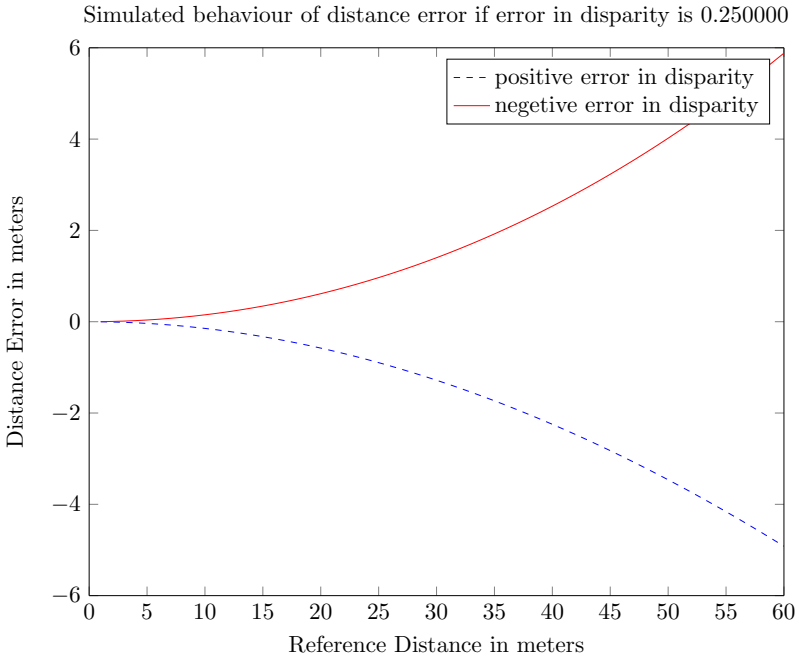

FIGURE 7.1: Behaviour of distance error when error in disparity is positive and negative

7.3 Results

As stated in the section 6.8, the size of the window effects the estimation of the algorithm. The size of the window depends on the requirement of the application and is user specific. A too large window size delays the first output of the algorithm and also takes a long time for the algorithm to adapt to the changes in the de-calibration of the camera. A too small window size may not produce a significant peak in the histogram. The results for this thesis are presented for all the data that is available for about a length of 5 minutes of a scenario. It is taken for granted that there is no change in de-calibration of the camera with in a 5 minutes time span. Hence the results presented will be for all the data that is available instead of using sliding window.

A comparison of estimation of error in disparity when windowing is done using a window size of 2000 samples against when all data is used is shown for a scenario 7.3.1.

7.3.1 Scenario 1

A scenario where ego car is moving on a highway during day time and when the error in disparity is positive.

Estimation of error in disparity when a window size of 2000 samples is considered Plot of histogram for the estimation of error in disparity is shown in the figure

7.2 and comparison of estimated error in disparity from algorithm and reference data is shown in the figure 7.3

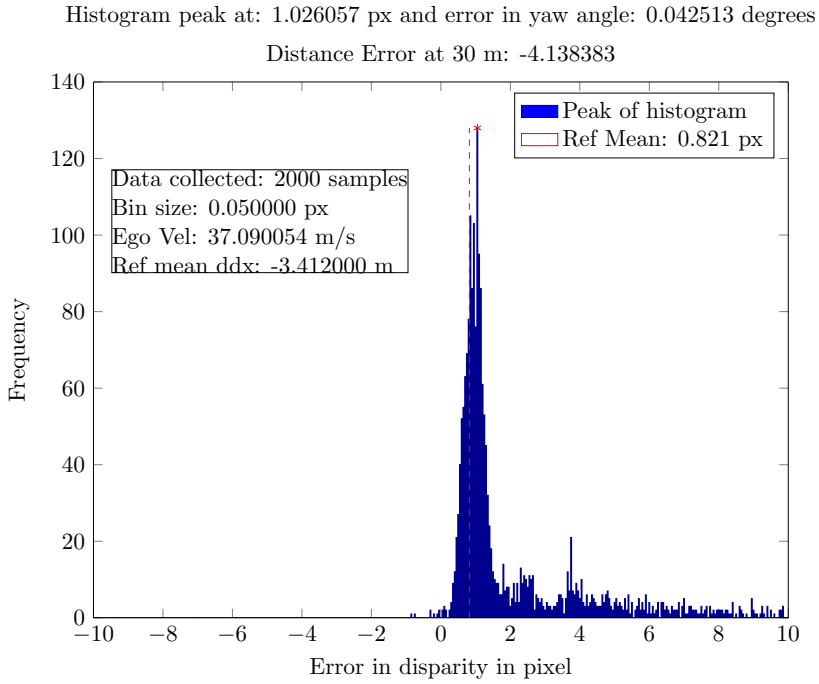

FIGURE 7.2: Estimation of error in disparity for a scenario of ego car travelling in highway and during day time

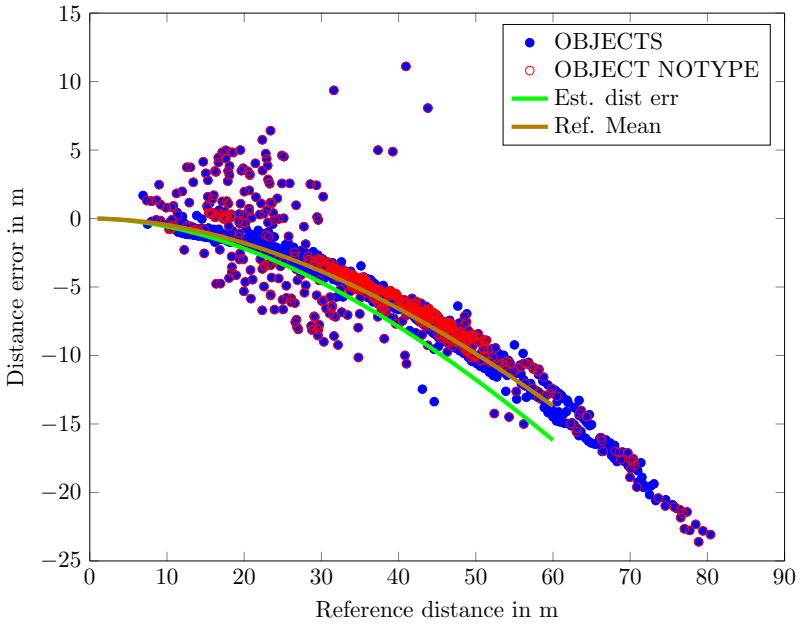

FIGURE 7.3: Comparison of estimated error in disparity and reference data for a scenario of ego car travelling in highway and during day time

Estimation of error in disparity when all the data is used

Plot of histogram for the estimation of error in disparity is shown in the figure 7.4 and comparison of estimated error in disparity from algorithm and reference data is shown in the figure 7.5

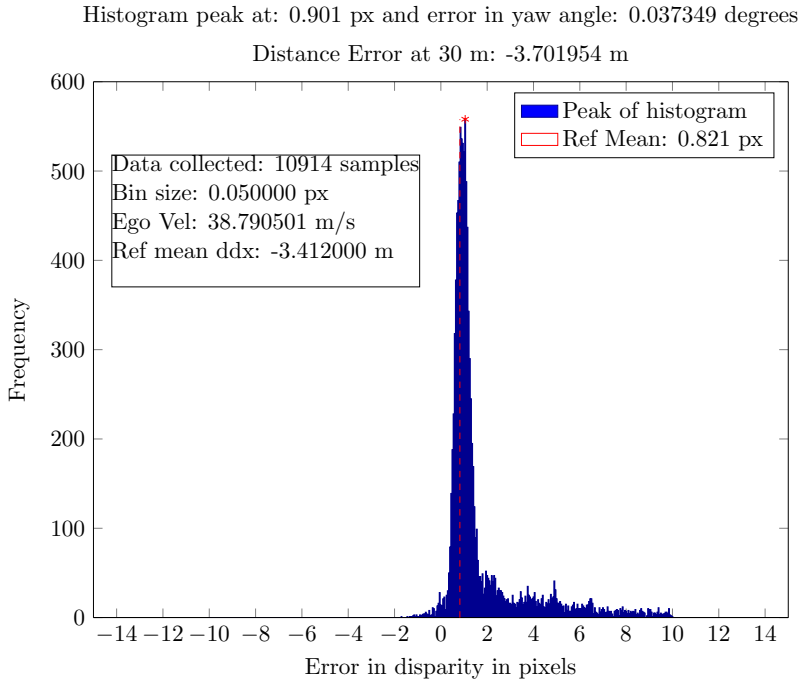

FIGURE 7.4: Estimation of error in disparity for a scenario of ego car travelling in highway and during day time

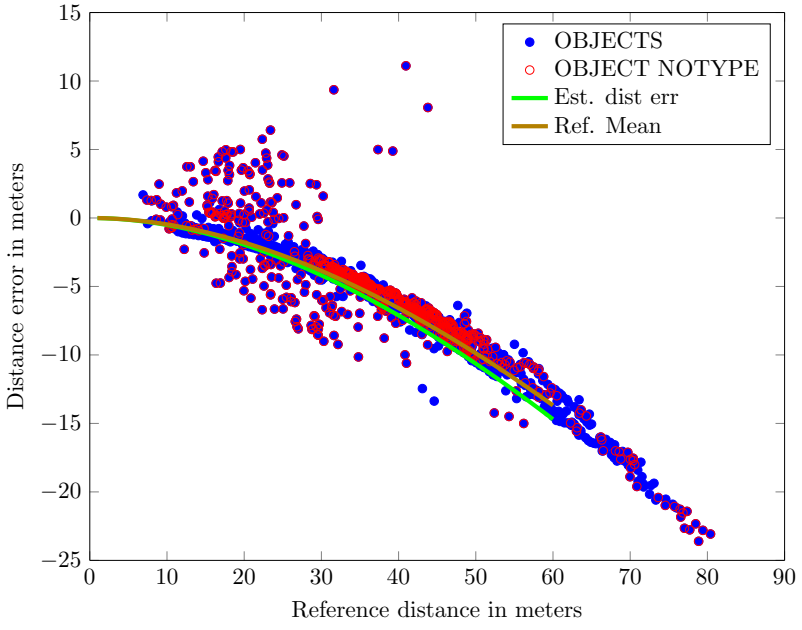

FIGURE 7.5: Comparison of estimated error in disparity and reference data for a scenario of ego car travelling in highway and during day time

It can be observed that the difference between is less, but when all the data is used the estimation of error in disparity is closer to the reference error in disparity and is better

than the case where only 2000 samples are used.

7.3.2 Scenario 2

A scenario where ego car is moving in city outskirts during night time and when the error in disparity is negative.

Plot of histogram for the estimation of error in disparity is shown in the figure 7.6 and comparison of estimated error in disparity from algorithm and reference data is shown in the figure 7.7

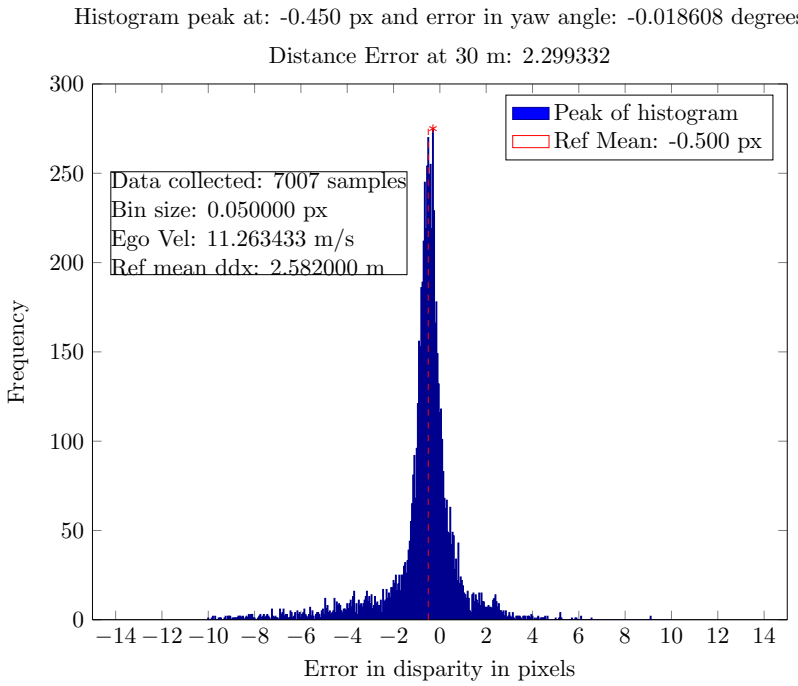

FIGURE 7.6: Estimation of error in disparity for a scenario of ego car travelling in city outskirts and during night time

7.3.3 Scenario 3

A scenario where ego car is moving in city outskirts during night time and when the error in disparity is positive and when the there is some noise in the data.

Plot of histogram for the estimation of error in disparity from histogram is shown in the figure 7.8

Comparison of estimated error in disparity from algorithm and reference data is shown in the figure 7.9

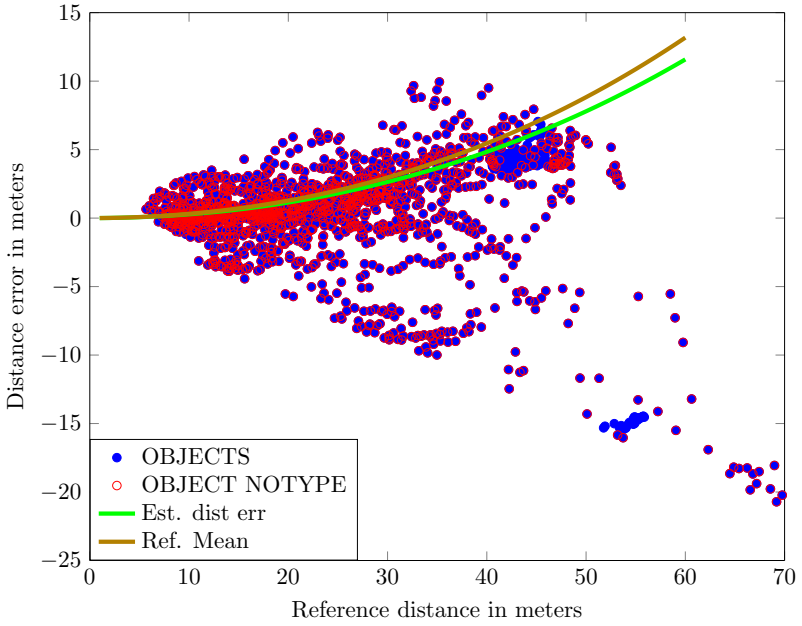

FIGURE 7.7: Comparison of estimated error in disparity and reference data for a scenario of ego car travelling in city outskirts and during night time

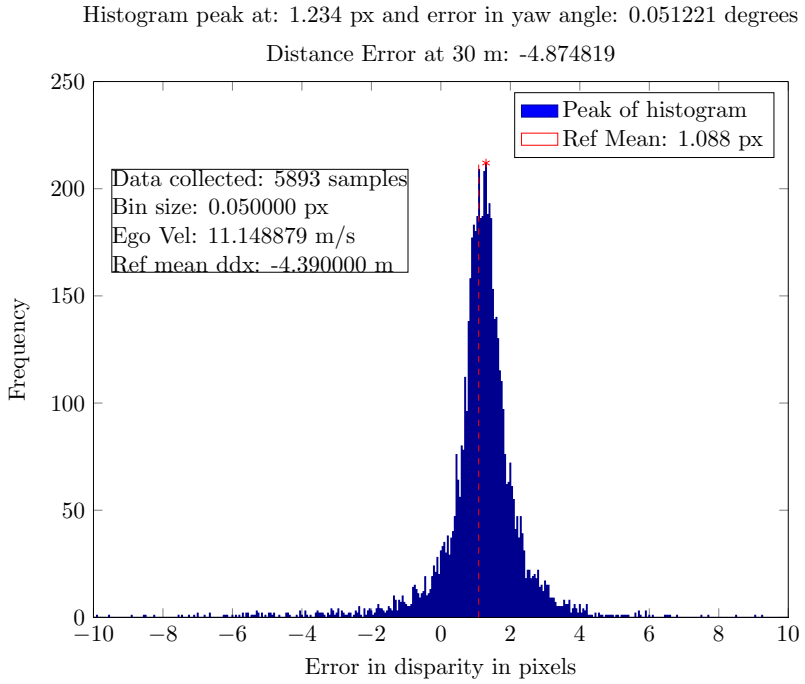

FIGURE 7.8: Estimation of error in disparity for a scenario of ego car travelling in city and during night time

7.3.4 Importance of comparing the estimated error in disparity with entire reference data

If a wrong corresponding object is established for comparison of reference data and the measured data, then the mean of the reference data will no longer be useful for comparing

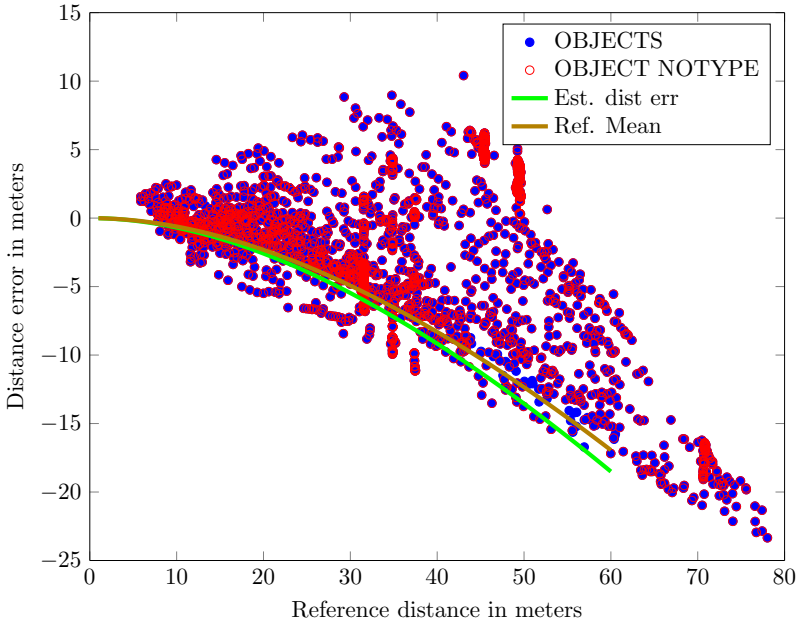

FIGURE 7.9: Comparison of estimated error in disparity and reference data for a scenario of ego car travelling in city and during night time

the estimated error in disparity. This is the reason for comparing the estimated error in disparity with the each reference distance error data from each object at every time instance. Some examples for such scenario can be seen in the figure 7.11 and 7.13

Plot of histogram for the estimation of error in disparity from histogram is shown in the figure 7.10

Comparison of estimated error in disparity from algorithm and reference data is shown in the figure 7.11

7.4 Open Issues

For few scenarios the estimation from the algorithm is very close or the above the tolerance mentioned in the section 6.2. These results are shown in figures 7.14,7.15,7.16,7.17. It can be observed that for these cases the estimation of the error in disparity is bad and is very close the tolerance of 1.28 m error between estimation and reference value.

7.5 Possible causes for the open issues

Error in measurement of ego velocity

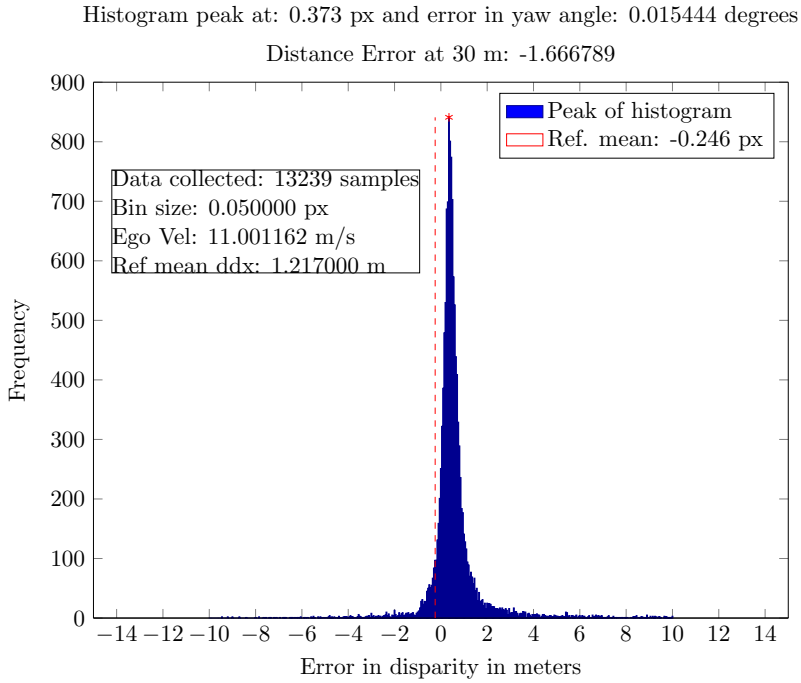

FIGURE 7.10: Estimation of error in disparity for a scenario of ego car travelling in city outskirts and during day time

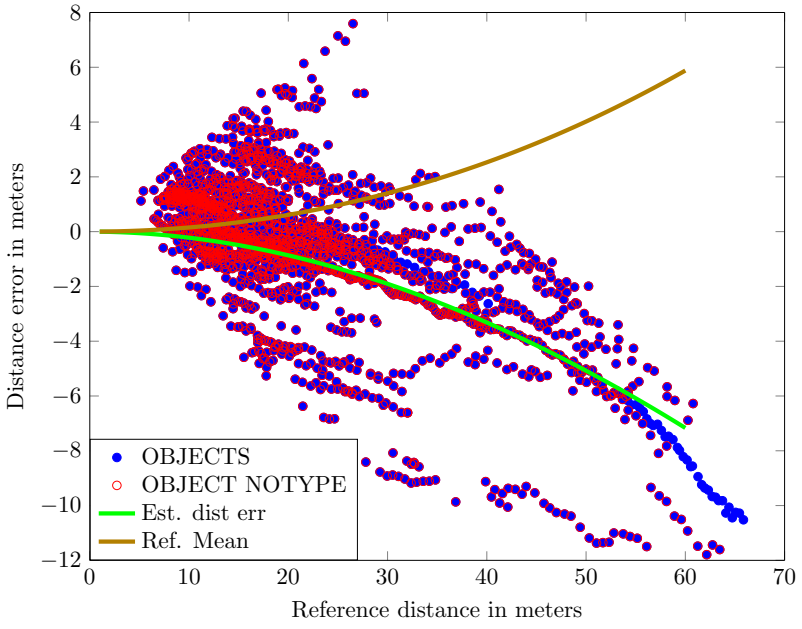

FIGURE 7.11: Comparison of estimated error in disparity and reference data for a scenario of ego car travelling in city outskirts and during day time

For the algorithm in this thesis it is assumed that the there is no error in ego velocity as mentioned in the section 4.2. When this assumption is violated, this will affect the estimation of the error in disparity by the algorithm. This is studied by a simulation. In the simulation an error in disparity of 0.25 pixels is introduced with different errors in ego velocity and the estimation of the algorithm under the simulation conditions

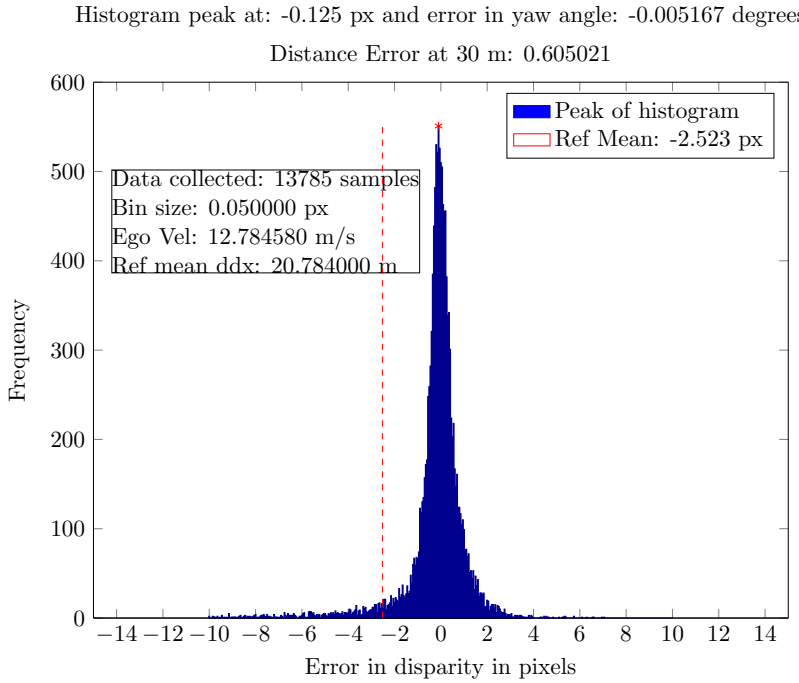

FIGURE 7.12: Estimation of error in disparity for a scenario of city outskirts and during night time

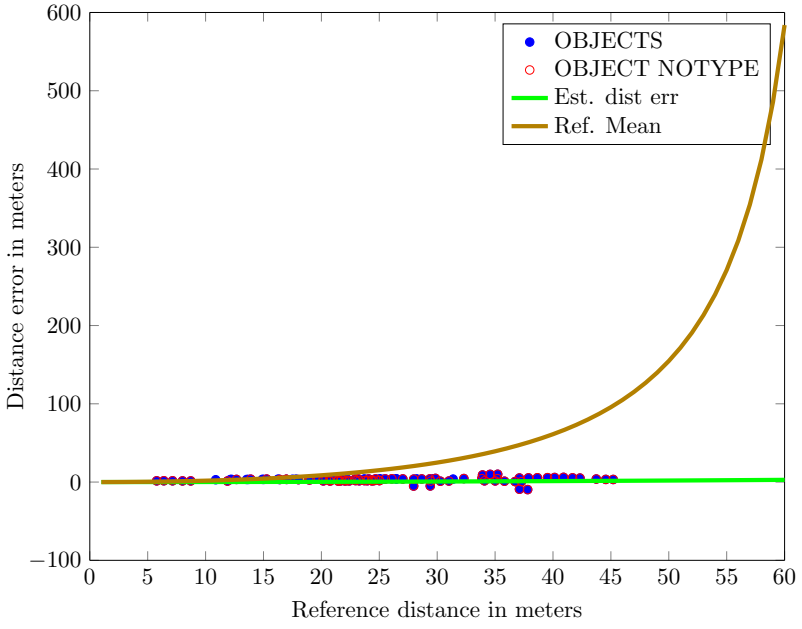

FIGURE 7.13: Comparison of estimated error in disparity and reference data for a scenario of city outskirts and during night time

mentioned in the chapter 5 is calculated. The behavior of error in disparity estimation against ego velocity is plotted in the figure 7.18.

It can be observed that as the % of error in measured ego velocity is increasing the estimation of the error in disparity is further away from the introduced error in disparity.

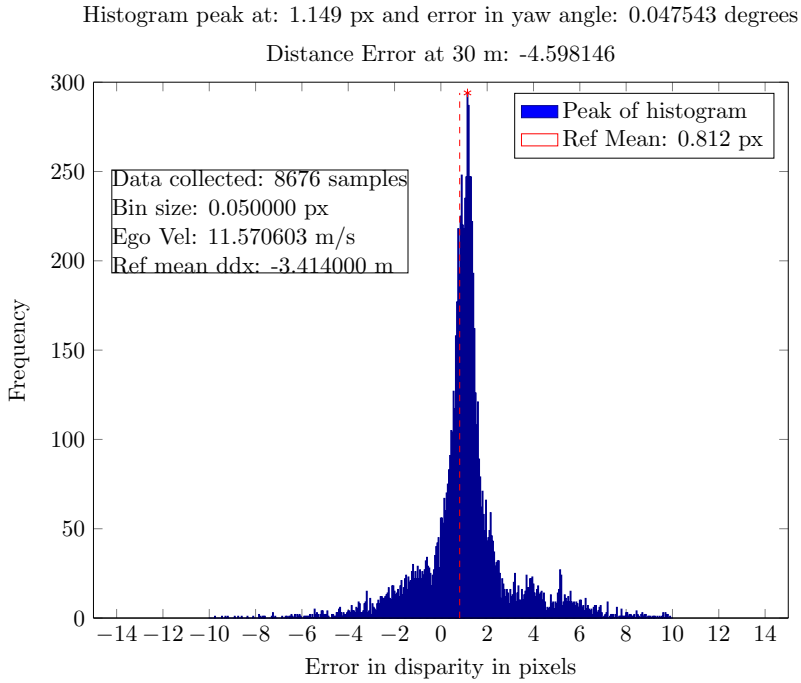

FIGURE 7.14: Estimation of error in disparity for a scenario of ego car travelling in city outskirts and during night time

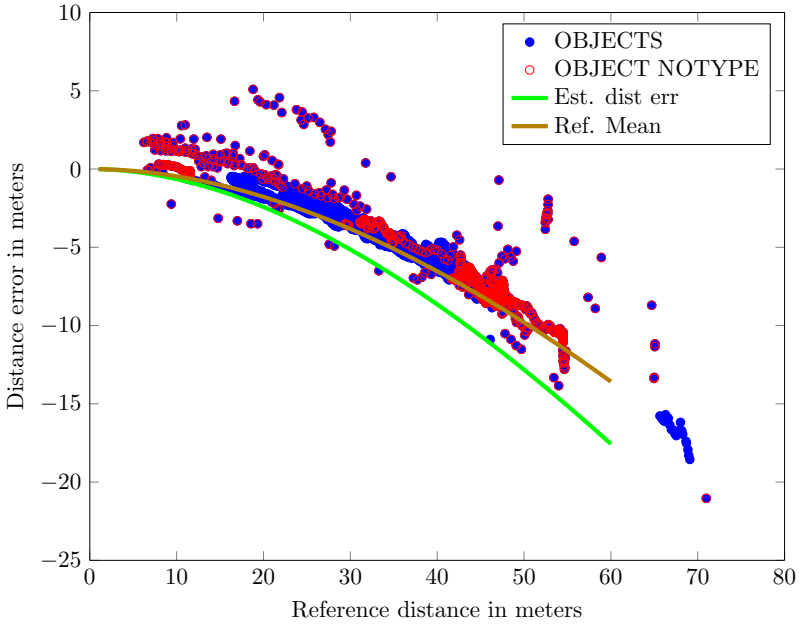

FIGURE 7.15: Comparison of estimated error in disparity and reference data for a scenario of ego car travelling in city outskirts and during night time

Also, as the ego velocity is increasing this effect is amplified. So, if there in an error in the ego velocity considered for the estimation of error in disparity by the method explained by this thesis, the estimated distance error is expected to be deviate from the reference distance error. This deviation depends on the measured ego velocity and the % of error present in it.

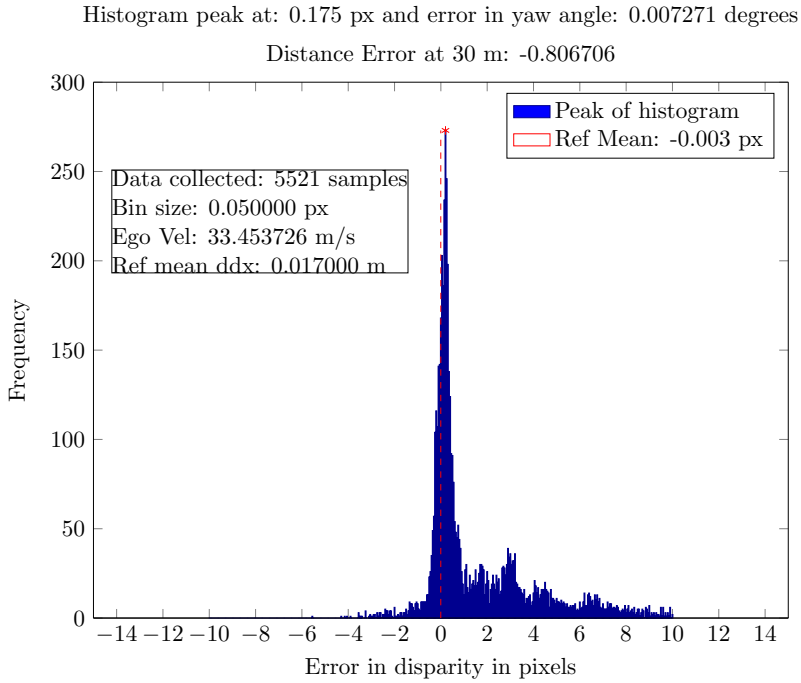

FIGURE 7.16: Estimation of error in disparity for a scenario of ego car travelling in high way during day time

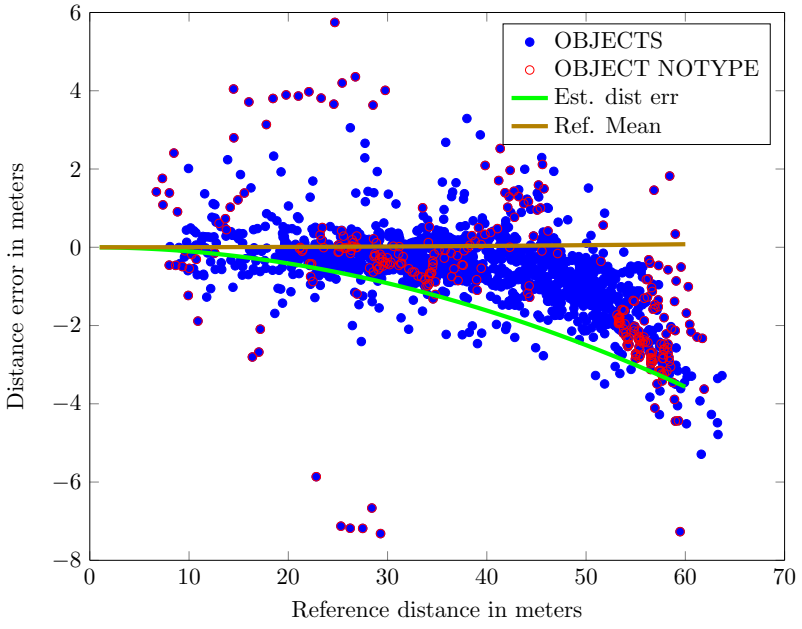

FIGURE 7.17: Comparison of estimated error in disparity and reference data for a scenario of ego car travelling in high way during day time

7.6 Conclusion

The algorithm is tested on about 30 scenarios and the average difference between the reference distance error and the estimated distance error is found out to be 0.56m at 30m. That is equal to 0.11 pixel disparity error. This value is measure of approximate

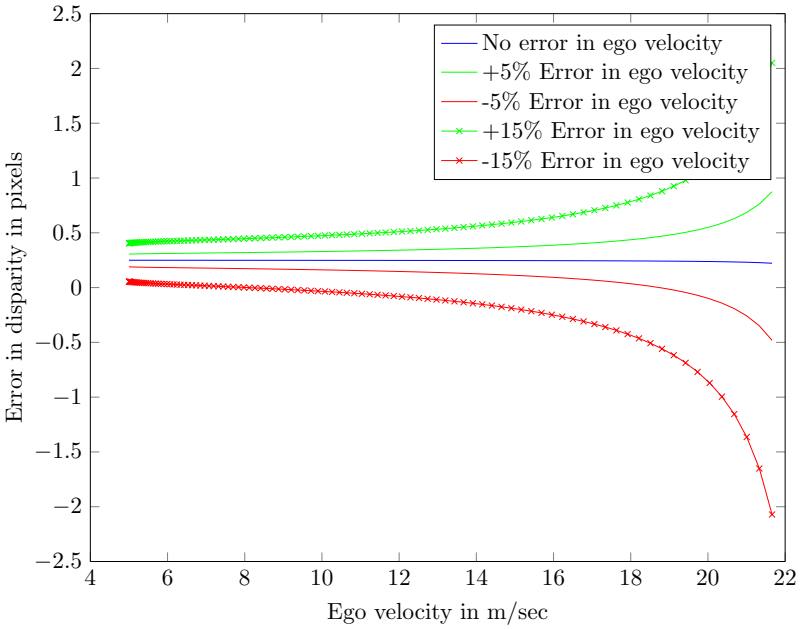

FIGURE 7.18: Simulation of the variation of error in disparity estimation when there is error in ego velocity

accuracy of the algorithm. This value is considerably lower than the tolerance of 1.28 m distance error at 30m that has been taken. This result of 0.56 m error at 30m is very less than the required tolerance for the application of this concept to a real world safety system. Hence the algorithm proposed is very useful not only for de-calibration detection but also for the estimation of the error in disparity present due to the yaw angle misalignment. Although some open issues are presented in the section 7.4, the results of these cases are also below the required tolerance stated above. Hence it is safe to say that the proposed algorithm can be used in a real world driver assistance systems where driver safety is crucial.

Chapter 8

Conclusion

In this thesis an algorithm to detect the de-calibration in the stereo camera due to misalignment in the yaw angle between the cameras, has been proposed. It is tested on a simulation system and found out that is able to estimate the error in yaw angle well under ideal conditions. The algorithm is then adapted to detect the de-calibration for the real world data from the stereo setup used at Robert Bosch and the results are presented. The algorithm is further tested on several scenarios like day time, in city scenarios and also on challenging scenarios like night time. It is found out that the algorithm is not only able to detect the de-calibration but also provide the estimate of de-calibration present which is much lower than the tolerance required for a real world application.

From the above results, it is shown that by using both the image information and the motion information from the objects detected and that of the camera rig, the challenges present due to the scene characteristics, like weather conditions, day or night can be confronted. This algorithm requires less processing power, as does not require pixel level processing. It can also be easily added as an additional system for de-calibration detection to an existing one for cross validation.

Appendix A

Appendix

A.1 Results of Simulation

This section provides the additional results from the simulation system that are not presented in the chapter 5.

Estimation of error in disparity in a case when ego car is moving with constant velocity

When the ego car is moving with constant velocity towards a static object , the depth vs time and disparity vs time and the estimation of error in disparity vs time graph when β is neglected is shown in the figures A.1, A.2 and A.3. The estimation of error in disparity when β is approximated is shown in figure A.4.

Estimation of error in disparity in a case when ego car is moving with varying acceleration

When the ego car is moving with constant velocity towards a static object , the depth vs time and disparity vs time and the estimation of error in disparity vs time graph when β is neglected is shown in the figures A.5, A.6 and A.7. The estimation of error in disparity when β is approximated is shown in figure A.8.

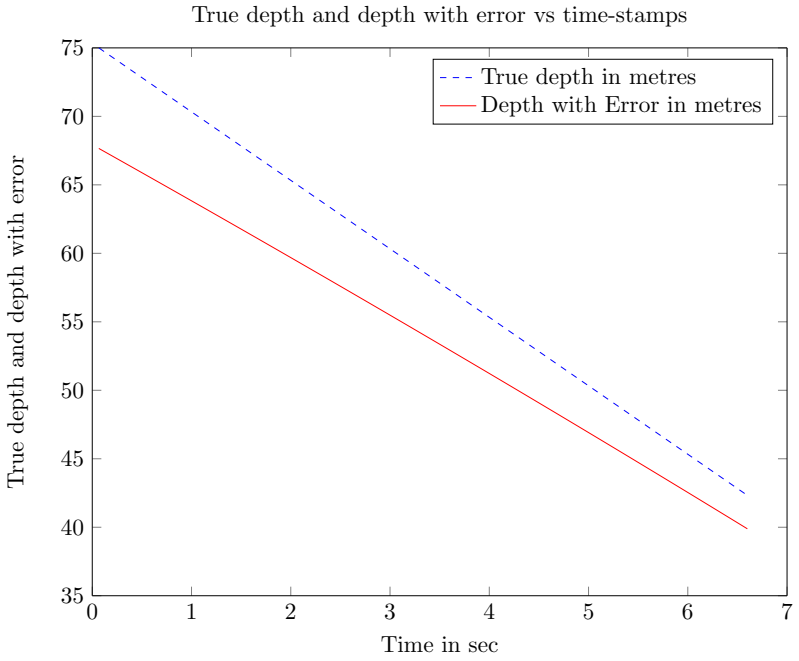

FIGURE A.1: Variation of true depth and measured depth when ego vehicle is moving with constant velocity

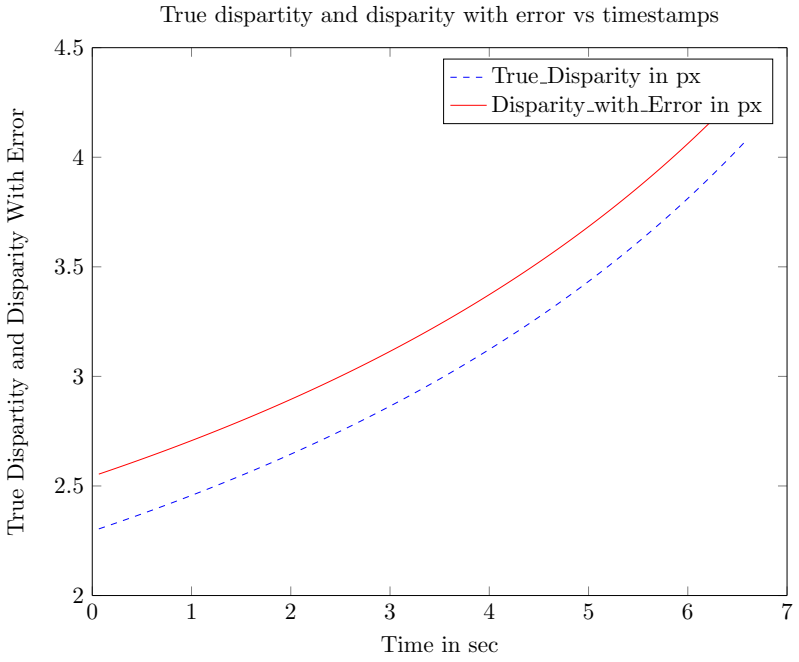

FIGURE A.2: Variation of true disparity and measured disparity when ego vehicle is moving with constant velocity

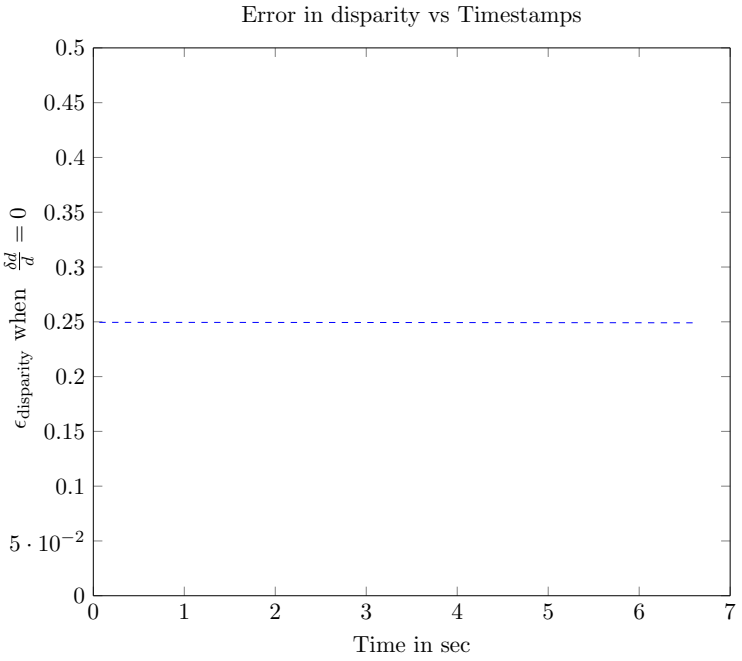

FIGURE A.3: Behaviour of estimated error in disparity when the ego vehicle is moving with constant velocity and β is neglected

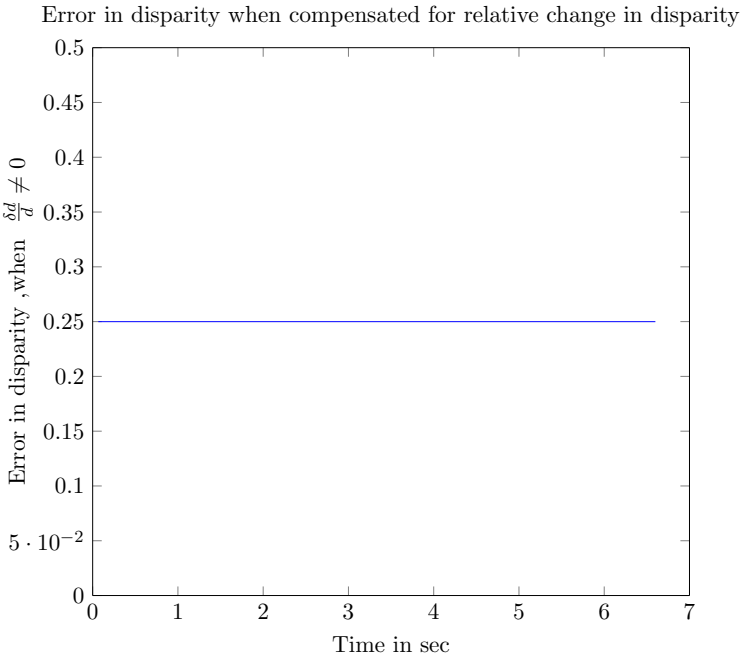

FIGURE A.4: Behaviour of estimated error in disparity when the ego vehicle is moving with constant velocity and β is approximated from measured disparity

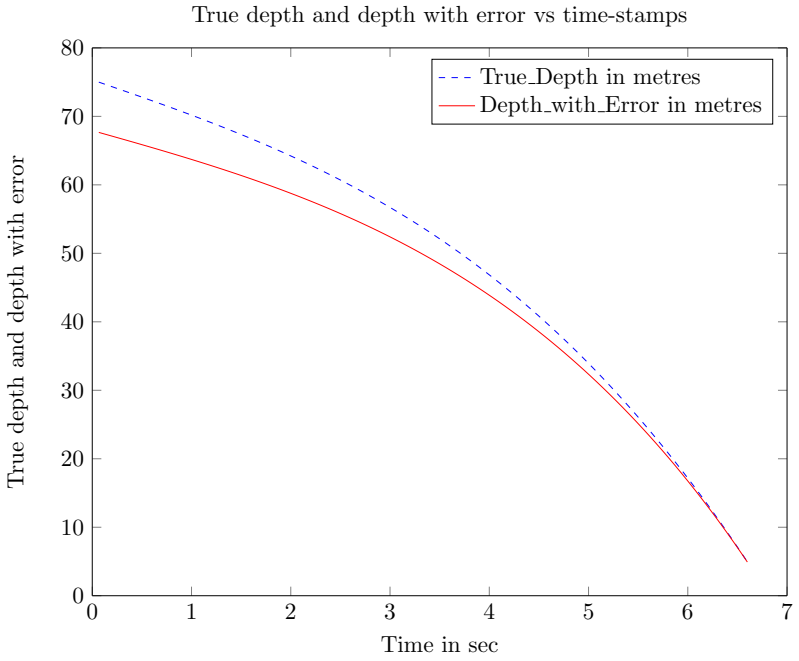

FIGURE A.5: Variation of true depth and measured depth when ego vehicle is moving with varying acceleration

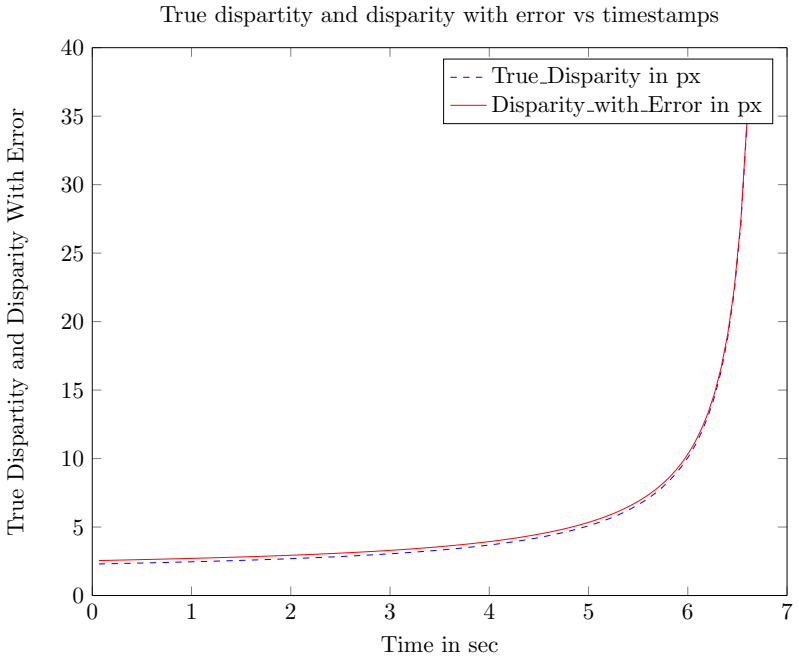

FIGURE A.6: Variation of true disparity and measured disparity when ego vehicle is moving with varying acceleration

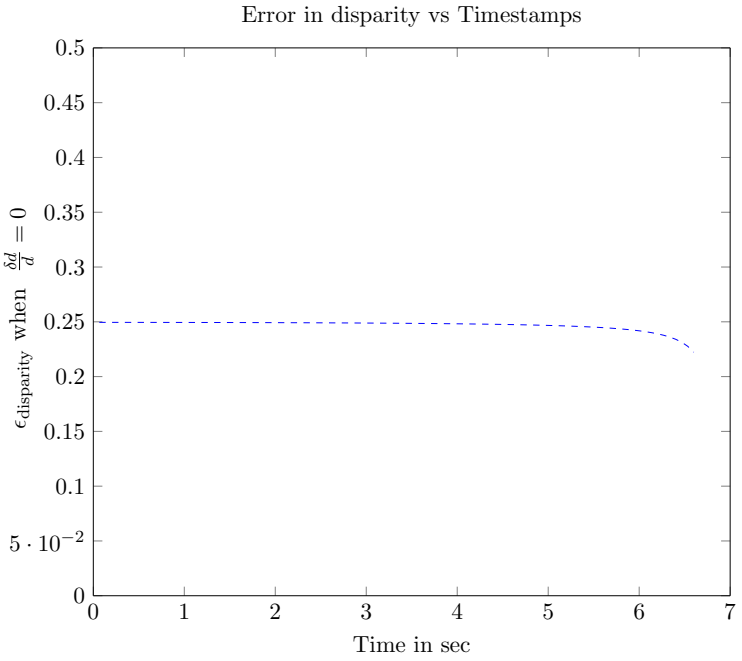

FIGURE A.7: Behaviour of estimated error in disparity when the ego vehicle is moving with varying acceleration and β is neglected

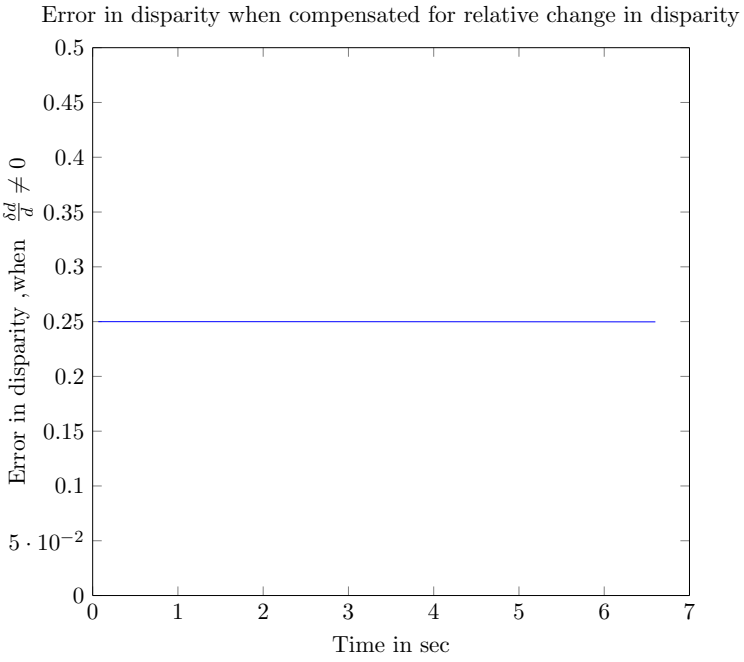

FIGURE A.8: Behaviour of estimated error in disparity when the ego vehicle is moving with varying acceleration and β is approximated from measured disparity

Bibliography

[1] Christian Whler. 3d computer vision, efficient methods and applications. 2013.

[2] F.Kuwar F.Mahmood, Syed.M.B.Haider. Investigating the performance of correspondence algorithms in vision based driver-assistance in indoor environment. *International Journal of Computer Applications (0975 8887)*, December 2012.

[3] Wenyi Zhao and N. Nandhakumar. Effects of camera alignment errors on stereoscopic depth estimates. April 1996.

[4] T. Daboczi A. Bodis. Optimization methods to calibrate a stereo rig with increased accuracy for vehicular applications.

[5] Gang Xu and Zhengyou Zhang. Epipolar geometry in stereo, motion and object recognition. 1996.

[6] Sebastian Jansson. On vergence calibration of a stereo camera. 2012.

[7] Berthold Klaus Paul Horn. Robot vision. 1986.

[8] Hartley and Zisserman. Multiple view geometry in computer vision (2nd edition). 2003.

[9] G. Csurka R. Horaud and D. Demirdijian. Stereo calibration from rigid motions, pattern analysis and machine intelligence. 2000.

[10] Simon H. K. Schauwecker, S. Morales and R. Klette. A comparative study of stereo-matching algorithms for road-modeling in the presence of windscreen wipers.

[11] Vladimir Igorevi Arnold. Mathematical methods of classical mechanics (2nd edition). 1989.

[12] S. Moscatelli Y. Kodratoff. Machine learning for object recognition and scene analysis. 1994.

[13] Edmund G.R. Kraal Christopher Lavers. Advanced electrotechnology for marine engineers. 2014.