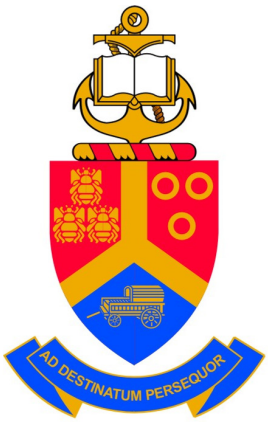


University of Pretoria

Digital Image Correlation: Applications in Vehicle Dynamics

by

Theunis Richard Botha

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor of Engineering
(Mechanical Engineering)

in the Faculty of

Engineering, Built Environment and Information Technology (EBIT)

at the

University of Pretoria,
Pretoria

February 2015

Summary

Title: Digital Image Correlation: Applications in Vehicle Dynamics
Author: Theunis Richard Botha
Study Leader: Prof. P.S. Els
Department: Mechanical and Aeronautical Engineering, University of Pretoria
Degree: Doctor of Philosophy in Mechanical Engineering

Except for aerodynamics forces, all vehicle excitation forces are generated at the tyre-road interface. Considering low speed applications, such as terramechanics where the aerodynamics force are negligible, the road-tyre interaction is of extreme importance. Crucial variables which govern the forces generated at the tyre-road interface are tyre side-slip angle, tyre longitudinal slip ratio and terrain profile. Solutions to measure these variables exist for smooth hard roads but the solutions experience challenges on rough and off-road terrain.

Digital Image Correlation is concerned with tracking the changes of a scene in a sequence of images or in images obtained from multiple viewpoints. These methods are frequently used in micro and nano-scale mechanical testing due to its ease of implementation and use as well as its non-contact approach. As a result these techniques are being implemented in many fields from material testing, physics, film animations and engineering. The aim of this thesis is to investigate the applications of Digital Image Correlation in vehicle dynamics using cost efficient off-the-shelf digital cameras and lenses. The following three vehicle dynamic problems are investigated: The measurement of the vehicle side-slip angle, longitudinal slip-ratio of a pneumatic tyre and high fidelity terrain profiling.

The vehicle side-slip angle can be used as a measure of the vehicle stability and therefore be used to improve the effectiveness of vehicle stability controllers. The tyre side-slip angle is also a vital measurement in characterising the lateral force characteristics of pneumatic tyres. A planar measuring method using Digital Image Correlation is shown to accurately measure the side-slip angle. The method is expanded by developing two additional algorithms which can measure all translational and rotational velocities. These methods are validated on both smooth surfaces and rough off-road terrain.

A method is also implemented whereby the longitudinal slip ratio of a tyre can be measured using a single camera. Therefore, doing away with the conventional method of using three independent

measurement systems. Features in the contact patch, encompassing both the tyre and the road, are tracked in a sequence of images. The features are classified into features lying on the tyre, road and outliers using a clustering algorithm. This enables the system to determine the tyre and road velocities from which the slip ratio is determined.

High fidelity terrain profiling is performed using a calibrated stereographic rig to obtain a three dimensional point cloud of the scene which is being viewed. The point cloud generated at one sample contains a grid of points encompassing a large area with points spaced both laterally and longitudinally. Overlapping point clouds are generated and joined using various registration techniques. The joined point clouds are sub-sampled to obtain a regularised grid of point containing a single point cloud of non overlapping points.

The proposed techniques create new possibilities in the field of vehicle dynamics. Enabling the side-slip angle to be measured in rough of road conditions while providing additional measurements. The longitudinal slip ratio which is measured directly at the contact patch could pave the way for better understanding the mechanism of the longitudinal tyre force generation. The inexpensive road profiling systems enables multiple sensors to be used in terramechanics tests to determine the impact of a vehicle on the environment. The thesis presents the mere tip of the ice berg concerning digital image correlation used in vehicle dynamics with many more possibilities waiting to be discovered.

Acknowledgements

I would like to show my gratitude to

- My parents which have provided me with a roof over my head and food in my stomach which allowed me to further my studies
- My supervisor Professor Schalk Els who has provided me with guidance over the past several years. Your mentor-ship and support has allowed me to successfully complete my PhD
- My fellow PhD students Herman, Joachim and Anria with whom I have had many conversations pertaining to our studies
- Carl Becker who was always eager to help with testing or helping with other small things. Also for the swimming which always gave me time to reflect on a problem
- The rest of the vehicle dynamics group, Glenn, Wietsche, Brett, Michael, Hans, Torsten and Jaco for all the interesting times
- My brothers Philip and Johann who always kept me occupied with other activities
- Antoinette who gave me time to relax when I needed it and for listening to all the rumblings of a mad scientist
- The financial assistance of the National Research Foundation (DAAD-NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the DAAD-NRF.

Contents

List of Figures

List of Tables

List of Symbols

Roman Symbols

Symbol	Description	Units
CoC	Circle of confusion	*mm*
c_x	Lateral pixel coordinate of camera optical axis projected onto the sensor	*pixel*
c_y	Vertical pixel coordinate of camera optical axis projected onto the sensor	*pixel*
d	Binocular disparity between left and right views	*pixel*
$\boldsymbol{D}$	Diagonal singular values obtained from Singular Value Decomposition	
D_F	Furthest distance from lens in focus	*mm*
D_N	Nearest distance from lens in focus	*mm*
$\boldsymbol{e}$	Epipolar point	*pixel*
$\boldsymbol{E}$	Essential matrix	
f	Focal length of lens	*mm*
f_{number}	Aperture opening size	
$f_{overlap}$	Fraction of overlap	
f_x	Pixel focal length of lens in lateral axis	*pixel*
f_y	Pixel focal length of lens in vertical axis	*pixel*
$\boldsymbol{F}$	Fundamental matrix	
FPS	Frames per second	*Hz*
h_i	ith value of the Homography matrix H	
$\boldsymbol{H}(I)$	Hessian matrix of pixel intensity I	
I	Pixel intensity	*byte*
$\boldsymbol{I_t}$	Image at time t	
I_t	Pixel intensity gradient with respect to time	
I_x	Pixel intensity gradient in x-direction	
I_y	Pixel intensity gradient in y-direction	
k_i	i^{th} Radial distortion correction parameter	
$\boldsymbol{K}$	Intrinsic camera parameters	
$\boldsymbol{K_i}$	Intrinsic parameters of i^{th} camera	
KP_i	Key Point i	*pixel*
$KP_i(V_i)$	Velocity of Key Point i	
L_{frame}	Length of the area capture in a frame	*mm*

Symbol	Description	Units
L_{pixel}	Length of area captured per pixel	*mm*
L_{sensor}	Length of sensor	*mm*
$\boldsymbol{H}$	Homography matrix	
$\boldsymbol{M}$	Average Hessian matrix in a window	
$\boldsymbol{p}$	Pixel coordinates $\begin{bmatrix} x & y \end{bmatrix}^T$ of point P projected onto image plane	
p_i	i^{th} Tangential distortion correction parameter	
p_{ij}	Distance between clusters i and j	
$\boldsymbol{P}$	Real world coordinates of a point in space	
$\boldsymbol{q}$	Normalised image coordinates	
q_i^j	ith value from the point coordinate $\boldsymbol{q}$ of the jth correspondence	
r	Radial pixel distance from optical centre	*pixel*
$\boldsymbol{R}$	Rotational transformation matrix	
R_e	Effective tyre rolling radius	*m*
$\boldsymbol{R_x}$	Rotational transformation matrix about x-axis	
$\boldsymbol{R_{stereo}}$	Relative rotational transformation matrix between stereographic rigs	
$R_{y,t}$	Rolling radius from camera at time t	*pixel*
s	Scalar value	
S	Longitudinal slip ratio	*%*
S_x	Lateral pixel size	*pixel/mm*
S_y	Vertical pixel size	*pixel/mm*
$\boldsymbol{t}$	Translation vector in x,y and z axis	*mm*
t	Time	*s*
t_x	Translation in x-axis	*mm*
$[\boldsymbol{t}]_{\boldsymbol{x}}$	Skew symmetric translational matrix	
$\boldsymbol{t_{stereo}}$	Relative translation displacement vector between stereographic rigs	
$\boldsymbol{T}$	Transformation matrix containing rotational and translational matrix	*mm*
$\boldsymbol{U}$	Left singular vector matrix obtained from Singular Value Decomposition (SVD)	
$\boldsymbol{V}$	Right singular vector matrix obtained from Singular Value Decomposition (SVD)	
V	Longitudinal vehicle velocity	*m/s*
V_a	Actual tyre velocity at contact patch	*m/s*
V_{max}	Maximum operating vehicle speed of system	*m/s*
$V_{max,overlap}$	Maximum operating vehicle speed of system as limited by the overlap of images	*m/s*
$V_{max,blur}$	Maximum operating vehicle speed of system based on the sharpness of an image	*m/s*
V_{road}	Velocity of road from camera	*pixel*
V_t	Theoretical tyre velocity at contact patch	*m/s*
V_{tyre}	Velocity of tyre from camera	*pixel*
V_x	Velocity in x-direction	*pixel/s*
V_y	Velocity in y-direction	*pixel/s*
w	Scalar value	
w_i	Inverse to a power interpolation weighting value	
x	Height of object projected on image plane (2D) or lateral pixel coordinate (3D)	*mm*
$\bar{\boldsymbol{x}}$	x and y coordinates of point being interpolated	
$\boldsymbol{x}$	x and y coordinates of point being used to interpolate point $\bar{\boldsymbol{x}}$	
x_c	Lateral distortion corrected pixel coordinate	*pixel*
$x_{i,j}$	Longitudinal x value of point cloud at indices row$i$ and column j	*mm*

Symbol	Description	Units
x_{max}	Maximum x value of point in point cloud	*mm*
x_{min}	Minimum x value of point in point cloud	*mm*
x_{step_i}	Increase in longitudinal x value with an increase in row (i) index	*mm*
x_{step_j}	Increase in longitudinal x value with an increase in column (j) index	*mm*
X	Height of object in real world	*mm*
$\boldsymbol{X}_{ik}$	kth component of cluster i's centroid	
y	Vertical pixel coordinate (3D)	*mm*
y_c	Vertical distortion corrected pixel coordinate	*pixel*
$y_{i,j}$	Lateral y value of point cloud at indices row$i$ and column j	*mm*
y_{max}	Maximum y value of point in point cloud	*mm*
y_{min}	Minimum y value of point in point cloud	*mm*
Y	Lateral length of object in real world	*mm*
x_{step_i}	Increase in lateral y value with an increase in row (i) index	*mm*
x_{step_j}	Increase in lateral y value with an increase in column (j) index	*mm*
z_i	Vertical height of point $\boldsymbol{x}$ used in interpolation	*mm*
$\bar{z}$	Interpolated Vertical height of point $\boldsymbol{x}$	*mm*
Z	Distance between lens and object	*mm*

Greek Symbols

Symbol	Description	Units
δx_r	Lateral radial distortion correction distance	*pixel*
δx_t	Lateral tangential distortion correction distance	*pixel*
δy_r	Vertical distortion correction distance	*pixel*
δy_t	vertical tangential distortion correction distance	*pixel*
Δi	Difference in row (i) index	
Δj	Difference in column (j) index	
Δx	Difference in longitudinal position (x)	*mm*
Δy	Difference in lateral position (y)	*mm*
ϵ	Error value	
$\prod$	Image plane	
ψ	Angle of rotation about x-axis	*radians*
ϕ	Angle of rotation about x-axis	*radians*
θ	Angle of rotation about z-axis	*radians*
ω	Wheel angular velocity	*radians/s*

List of Abbreviations

Abbreviation	Description
2D	Two Dimensional
3D	Three Dimensional
ABS	Anti-Lock Braking System
ADC	Analogue to Digital Converter
CCD	Charge Coupled Device
CMOS	Complimentary Metal Oxide Semiconductor
CoC	Circle of Confusion
CG	Centre of Gravity
DIC	Digital Image Correlation
DoG	Difference of Gaussian
FPGA	Field Programmable Gate Array
EPS	Electronic Stability Program
GPS	Global Position System
ICP	Iterative Closest Point
IMU	Inertial Measurement System
INS	Inertial Navigation System
ISO	International Standards Organisation
MEMS	MicroElectroMechanical Systems
OpenCV	Open Computer Vision
OpenPCL	Open Point Cloud Library
RANSAC	RANdom SAmple Consensus
RMS	Root Mean Square
ROI	Region of Interest
RTK	Real Time Kinematic
SAD	Sum of Absolute Differences
SIFT	Scale-Invariant Feature Transform
STD	Standard Deviation
SURF	Speeded Up Robust Features
SUV	Sports Utility Vehicle
SVD	Single Value Decomposition

Chapter 1

Introduction

Except for aerodynamics forces, all vehicle excitation forces are generated at the tyre-road interface. Considering low speed applications such as terramechanics and in-situ tyre testing, where the aerodynamics force are negligible, the road-tyre interaction is of extreme importance. Crucial variables which govern the forces generated at the tyre-road interface are tyre side-slip angle, tyre longitudinal slip ratio and terrain profile. These variables capture the mechanisms which create the lateral, longitudinal and vertical tyre forces respectively. Solutions to measure these variables exist for smooth hard roads. These solutions not only have their drawbacks on the hard surfaces but also experience challenges on rough and off-road terrain. This study addresses the shortcoming of these solutions by means of digital image correlation, creating solutions with improved performance on hard surfaces with the ability to measure the variables on rough and off-road terrain.

Digital Image Correlation (DIC) is a mathematical technique whereby naturally occurring or artificially created surface contrast patterns are tracked using algorithms from one image to another using algorithms. The advent of inexpensive, high frame rate cameras as well as higher throughput computers have opened up the way for DIC techniques to be used in various fields. These DIC techniques use non-contact optical techniques to measure the changes in image sequences. The DIC algorithms and implementations can be tailored depending on the desired measurement. This diversity and customisability has allowed DIC to be applied in various fields to obtain different measurements. Certain algorithms are used in material testing to characterise multi axial behaviour of rubber-like materials [Chevalier et al., 2001] and to investigate crack growth [Abanto-Bueno and Lambros, 2002]. DIC is

also used in digital particle image velocimetry to measure the direction of velocity and displacement fields of particles. High contrast particles are embedded in various mediums to measure the velocity pattern of the medium. These mediums can include liquids such as water and oil, particle based substances such as sand and mud or even solid materials. These techniques are also used in human motion tracking [Cai and Aggarwal, 1999] not just for the purpose of studying the motion of animals and humans but also to capture the motion of actors for animated films.

Stereography is the ability to determine the depth of a scene using DIC algorithms and multiple cameras [Grimson, 1981]. The ability to measure the depth of a scene has been extensively used in robotics in order to perform both path planning and navigation [Murray and Little, 2000], [Kriegman et al., 1989]. DIC provides the capability to both track and determine the three dimensional coordinates of particles. This has enabled a range of new possible measurements. Applications include strain based measurements [Pan et al., 2009] and 3D displacement and velocity measurements [Jin et al., 2010].

The maturity of the technology has reached a point where the algorithms are robust and relatively easy to implement especially when using open source algorithms. Due to the maturity of the techniques as well as the reduced cost and size of digital cameras, the techniques have been employed to a greater extent than previously. However, not much work has been performed in the domain of vehicle dynamics. Moser *et al* [Moser et al., 2010] used DIC to generate 3D measurements in tyre testing. However, the applications focus mainly on laboratory testing, such as determining the properties of the material used in the construction of the tyre as well as the outside deformation of the side-wall of a tyre when being loaded. Stereography is also used in vehicle crash tests [Rentschler and Uffenkamp, 1999] to perform automated deformation analysis of the vehicle. There has therefore been a limited amount of vehicle dynamics applications in which DIC is used. This study attempts to fill this gap in the research by applying DIC techniques in vehicle dynamics in order to improve and provide additional tools which can be used to improve the safety of vehicles. The systems also aim to provide additional information to researchers which have not been available previously, to enable new insights into vehicle engineering and especially into the mechanisms of force generation at the tyre-road interface.

1.1 Problem Statement

The purpose of this study is to investigate the usefulness of DIC techniques in vehicle dynamics. In order to fully understand DIC a basic understanding of cameras and camera models is provided as well as describing the DIC algorithms which are used in this study. The study attempts to address the question by solving three problems often experienced in vehicle dynamics. These problems are:

1. Measuring of side-slip angle at low vehicle speeds and over rough surfaces
2. Measuring longitudinal tyre slip ratio using a single measurement system capable of measuring the slip ratio at the contact patch
3. Obtain high fidelity terrain profiles in the presence of disturbances to the measurement system while at the same time reducing the reliance on an inertial navigation system

Each problem is addressed in a separate chapter, discussing the shortcoming of current methods used to solve the problems and describing the application of DIC in detail in order to solve the problem. Each solution is validated using experimental testing and compared to the current methods used to demonstrate the advantages of the proposed techniques.

1.2 Layout of the Study

The layout of the study is as follows:

The purpose of Chapter 2 is to describe the basics of camera systems, how digital images are obtained and the effects of camera settings on the captured images. The understanding of camera systems is vital to capture images which are sharp and are exposed properly in order for DIC techniques to be applied successfully.

In Chapter 3 the foundation is laid for the DIC techniques by examining the mathematical camera models used in calibration and lens distortion removal. In this chapter the calibration between two fixed cameras, to create a stereographic rig from which depth can be obtained, is also described.

Detail descriptions of the DIC algorithms which are implemented in this study to enable the tracking of features and to generate a 3D point cloud from a stereographic rig are provided in Chapter 4

Chapters 2 - 4 provide an introductory understanding of cameras and of the DIC techniques used in this study. The purpose of these chapters are not to serve as an comprehensive study of these techniques but merely to provide the reader with a basic understanding of the concepts and notations used in later chapters. The principal focus of this study is the implementation and usage of these concept and techniques to address problems in vehicle dynamics.

The investigation of the plausibility of using DIC in solving vehicle dynamic problems is performed in Chapters 5 - 7. In Chapter 5 the measurement of side-slip angle using three different DIC-based methods is described. The first is a planar method which measures the side-slip angle only. Two additional methods are introduced which also measures all rotational and translational velocities in addition to the side-slip angle. These methods are validated on both smooth surfaces and rough off-road terrain. The methods show excellent capability of measuring the side-slip angle, especially at low vehicle speeds and over rough terrain, where existing commercial solutions fail.

A method whereby the longitudinal slip ratio of a pneumatic tyre can be measured using a single camera is described in Chapter 6. This method replaces the traditional practice of obtaining the slip ratio by measuring the vehicle speed, wheel rotational speed and rolling radius using three separate sensors. The method tracks features in the contact patch region encompassing both the tyre and road. The features are classified into three groups, features on the road, tyre and outliers. From the classified groups the velocity of the tyre and road are determined from which the slip ratio can be determined. The technique allows the slip to be measured directly at the contact patch and negates the assumption of a constant rolling radius. This technique could bring new insights into the tyre longitudinal force generation mechanism.

A terrain profiling technique which uses DIC to obtain point clouds of which one sample contains points spaced both laterally and longitudinally is described in Chapter 7. Point clouds are taken such that they overlap. These overlapping point clouds are stitched together to produce a continuous point cloud made up of many individual clouds. A sub-sampling algorithm is used to create a single point cloud which contains a regularised grid of points. The technique negates the use of an expensive fast sampling Inertial Navigation System (INS) which is often used with line scan sensors to create a 3D road profiling system. The method is experimentally verified over artificial terrains with known profiles. This allows a known disturbance to be applied to the system while profiling a known road surface.

Chapter 2

Camera Basics

Digital Image Correlation techniques apply software algorithms to digital images. Therefore, the acquisition of images is a fundamentally important part of the process. The focus of this chapter is on the basics of camera systems, how they are modelled in physics, the effect of lens and camera settings and lighting. The aim is to ensure that captured images are sharp with enough exposure to successfully apply DIC algorithms to. The aim of this chapter is not to serve as a comprehensive summary of all camera systems and sensor architectures but merely to provide a basic guideline of the parameters which may be altered and their effect.

2.1 Digital Cameras

In order to implement DIC techniques it is required to have the images in digital form. The simplest method to obtain digital images is by means of a digital camera or digital video camera. Although other forms of images can be digitized using scanners this is beyond the scope of this study.

2.1.1 Digital Camera Sensors

Digital image sensors are used to convert light into electronic signals. Two different types of sensors, which are often used to convert light into electronic signals, are the Charged Coupling Device (CCD) and the Complimentary Metal Oxide Semiconductor (CMOS) sensors. Both sensors use photo diodes which releases electrons when light photons hit the diode. The amount of charge built up from the released electrons are proportional to the intensity of the light hitting the diode. The differences

between the CCD and CMOS sensors are how they are manufactured and how the charge from each diode is converted into a digital value.

The most basic method to create a camera sensor would be to have an array of photo diodes with each photo diode connected using an array of wires to an Analogue to Digital Converter (ADC). The ADC converts the analogue value obtained from the sensor into a digital value which can be used in computer algorithms. As each diode accumulates a charge, the charge can be sent via wires to an ADC to get a digital value proportional to the charge and therefore light hitting the diode. The problem with such a systems is that the wires would result in capacitive coupling. This would cause a charge to be leaked from one diode to the next resulting in distortion of the image. The difference between the CCD and CMOS sensors is the method used to solve this problem.

Charged Coupling Device (CCD)

In a CCD sensor each pixel contains a diode and a storage cell or buffer bin. As the sensor shutter opens, and is therefore exposed to light, each diode accumulates a charge proportional to the light intensity. Upon the closure of the shutter the light is shifted from the diode to the diode's storage cell. The sensor exposes each pixel at the same instance and duration. The sensor then starts the process of digitising the stored charge. The special nature of CCD sensors is the ability to shift the charge across to neighbouring storage cells. This is in fact the origin of the CCD sensor which is a device with the ability to act as a shift register for charges. Its application to imaging came after its invention.

Each column of the sensor is coupled allowing the charges to be shifted along the column (vertically). The charges are shifted vertically until the first pixel row is in the serial (horizontal) shift registry. This registry shifts the charge horizontally into a capacitor which generates a voltage dependent on the capacitance and the charge in the capacitor. The voltage is then amplified and digitised using an ADC which creates a digital value dependent on the charge and therefore the light intensity. Often the digital value of a pixel ranges between a value 0 for no illumination to 255, 8bit/1byte value, for a completely saturated pixel. However, higher order values are possible with higher resolution ADC's but this is however seldom required. The charges are shifted until the complete row has been digitised after which another vertical shift occurs which shifts the 2nd pixel row into the serial shit register. This process continues until all pixel values have been digitised after which the process

starts all over with a new frame.

This process allows the charges of each diode to be transferred to a single ADC while reducing capacitive coupling which causes leakage. However, this requires a special manufacturing process to allow charges to be shifted across. This makes the sensor relatively expensive to produce.

Complimentary Metal Oxide Semiconductor (CMOS)

In a CMOS sensor the problem is solved by simply having a capacitor and amplifier on each pixel. Therefore, there is no need to shift the charges across the device to one capacitor as in the CCD sensor. The number of ADCs per sensor ranges from a single ADC per several pixel columns to an ADC for each pixel. The voltage is simply transferred using normal semiconductor wires. This sensor contains a more complex pixel with a diode, capacitor and amplifier, possibly ADC, but the manufacturing process is very simplistic following standard semi-conductor fabrication methods.

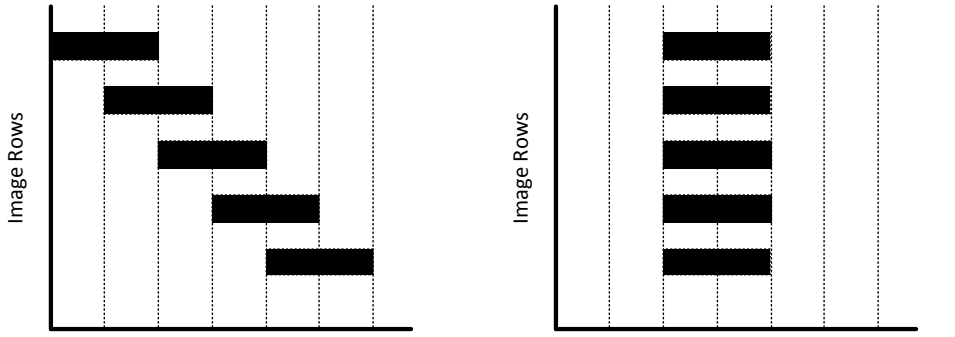
2.1.2 CCD vs CMOS

The question of which sensor type to use is application dependent as each have their own advantages and disadvantages. The CCD sensor has a larger diode per pixel as compared to the CMOS sensor, which contains many parts per pixel, which results in better light sensitivity and a wider dynamic range for the CCD sensor. CCD sensors also have less noise because of a quieter sensor substrate (reduced on-chip circuitry) and the fact that it uses a single amplifier and ADC as opposed to CMOS sensors. CMOS sensors on the other hand are much faster than CCD sensors due to multiple amplifiers and ADC's. CMOS sensors can also easily employ windowing which allows only a portion of the sensor values to be read out. Since CMOS sensors employ a standard manufacturing process they are also easier and cheaper to manufacture than CCD sensors. CMOS sensors also consume less power than CCD's giving them an advantage in low power applications.

One of the most significant differences is in shuttering. Most digital sensors have electronic shutters as opposed to mechanical shutters with CCD sensors offering superior electronic shutters by default. The ability to start and stop exposure is a standard feature on CCD sensors. Typically all pixel's exposure starts and stops almost simultaneously resulting in what is known as a global shutter. In contrast CMOS sensors are exposed one row at a time resulting in what is known as a rolling shutter. The exposure of the image sensor pixels is shown graphically in Figure 2.1. The

difference between a rolling and global shutter has significant effects when imaging objects in motion as illustrated in Figure 2.2. The figure shows that the motion of the fan blade becomes distorted due to the exposure mechanism of a rolling shutter. This effect is dependent on the speed of the motion as well as the exposure time of the sensor. If the sensor's exposure time is set very short the effect may be negligible. For this reason CCD sensors with global shutters are preferred when applying DIC to images. A study of the modelling of shutter mechanism and their effects on motion blur is provided by Schoberl *et al* [Schoberl et al., 2009].

In general CMOS sensors are used in many daily devices such as camcorders, cellphones, computers etc. for their cost and power advantages. On the other hand CCD sensors are still used in many scientific applications requiring global shutters and better quality images. However, CMOS sensors are making huge strides in sensitivity and global shutters. Some high-end CMOS sensors now have global shutters and improved dynamic range [SONY, 2014]. Durini [Durini, 2014] provides a detailed description of the internal workings of most imaging sensors and gives a detailed comparison between CCD and CMOS sensors.



(a) Rolling shutter exposure time for imaging sensor rows

(b) Global shutter exposure time for imaging sensor rows

Figure 2.1: Global vs rolling shutter exposure time

2.2 Lenses and Camera Settings

When capturing images there are several different camera and lens settings than can be set to obtain a more desired effect. On a digital camera these would include mainly sensitivity and shutter exposure time. On the lens the settings are the aperture size and focal length. These settings and their effects are briefly discussed in the following section.

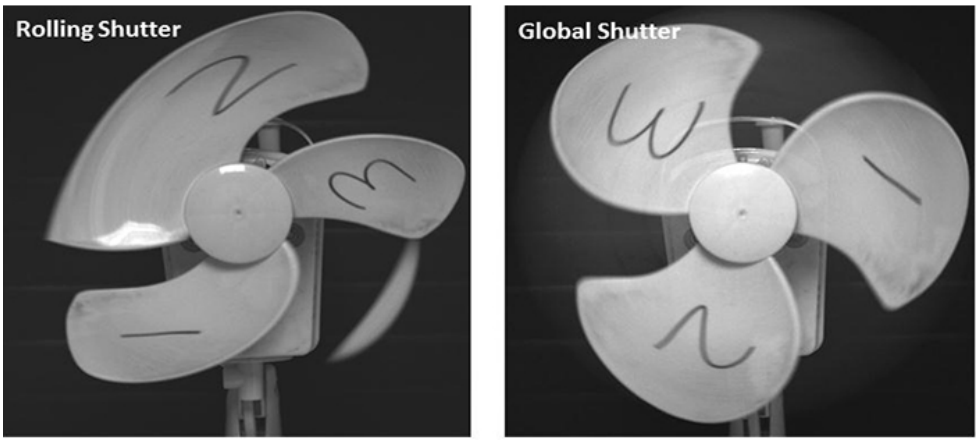


Figure 2.2: Distortion effect of rolling shutter compared to global shutter [Andor, 2014]

2.2.1 Camera: Sensor Sensitivity (ISO)

The imaging sensitivity more commonly known as the ISO, is simply a measure of the amplification used prior to the ADC. The higher the ISO value the higher the amplifier gain and the more sensitive the sensor [Jacobson et al., 2000]. The result is that with a higher sensitivity a smaller amount of light is needed to obtain the same effect at a lower sensitivity. This means that either a shorter exposure time or a smaller aperture is required. The side-effect of a higher sensitivity is that the noise levels are also amplified which could result in a more noisy (grainy) image. The effect of the ISO setting is shown in Figure 2.3. The exposure of the image increases as the ISO value increases, unfortunately the increase in noise is not as evident as it would be with much higher ISO values $ISO > 1600$.

(a) ISO=100, f_{number}=5.6, $t_{exposure}$=1s, f=50mm

(b) ISO=400, f_{number}=5.6, $t_{exposure}$=1s, f=50mm

(c) ISO=1600, f_{number}=5.6, $t_{exposure}$=1s, f=50mm

Figure 2.3: The effect of sensor sensitivity (ISO), f_{number} is the aperture size, $t_{exposure}$ is the exposure time and f is the focal length

2.2.2 Camera: Shutter Exposure Time

The shutter exposure time is simply the time duration the photo diode is exposed to light. With a longer exposure time the photo diode has a longer period of time to build up a charge representing the light intensity of that pixel. Therefore, with a longer exposure a higher charge is built up. This is typically required if the scene is very dark with insufficient light. If the exposure is set too high the result is an over exposed image with all detail washed away. The other side effect of a longer exposure time is that it causes motion blur should there be motion in the scene. Since this study considers capturing images of dynamic scenes it is required to have the exposure time as short as possible to retain clear features in the scene. The effect on increasing the exposure time is shown in Figure 2.4, illustrating that the image becomes more exposed as the exposure time increases. The effect of motion blur is illustrated in Figure 2.5 where the image with a longer exposure time shows considerable more blur. The exposure between the images were kept similar by increasing the image sensitivity on the shorter exposure time.

(a) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=0.5s, f=50mm (b) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=1.5s, f=50mm (c) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=3s, f=50mm

Figure 2.4: Effect of exposure time on image, ISO is the sensor sensitivity, f_{number} is the aperture size and f is the focal length

2.2.3 Lens: Aperture

The lens aperture is the opening through which light travels to hit the sensor. The larger the opening, the more light photons are allowed into the sensor, the faster the diode releases charge. The aperture opening is usually specified by an f-number. The f-number is the ratio of lens focal length to effective aperture opening diameter. Typical f-numbers range from $f/1.4$ to $f/22$, with a lower f-number denoting a larger aperture opening. The effect of the aperture opening is shown in Figure 2.6, the

(a) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=2s, f=50mm

(b) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=0.3s, f=50mm

Figure 2.5: effect on exposure time on motion blur, ISO is the sensor sensitivity, f_{number} is the aperture size and f is the focal length

figure shows that with an increase in the f-number, smaller aperture, the exposure of the image decreases. The aperture also affects the depth of field which is explained further in section 2.2.5.

(a) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=1s, f=50mm

(b) ISO=400, f_{number}=$f/8$, $t_{exposure}$=1s, f=50mm

(c) ISO=400, f_{number}=$f/11$, $t_{exposure}$=1s, f=50mm

Figure 2.6: Aperture opening and the effect on the exposure of the image, ISO is the sensor sensitivity, $t_{exposure}$ is the exposure time and f is the focal length

2.2.4 Lens: Focal Length

The focal length of a lens is the optical distance of the plane at which the light rays converge from the lens to form a sharp clear image of the scene. The focal length is determined when focusing on an object placed at infinity. The focal length also presents a measure of how strongly a lens converges light. A lens with a smaller focal length has a stronger optical power and therefore converges light more. This therefore also gives an indication of the magnification of a lens. The effect of the focal length is shown in Figure 2.7. The figure shows that the focal length has no effect on the exposure of the image. The effect observed is that the image region is increased as the focal length decreases. The magnification of the image thus increases as the focal length increases.

(a) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=2s, f=25mm (b) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=2s, f=37mm (c) ISO=400, f_{number}=$f/5.6$, $t_{exposure}$=2s, f=50mm

Figure 2.7: Focal length on image, ISO is the sensor sensitivity, f_{number} is the aperture size, $t_{exposure}$ is the exposure time and f is the focal length

2.2.5 Compromises between Camera and Lens Settings

All the camera and lens settings discussed above need to be considered together as they effect each other and often a compromise between settings is made. The three settings which affect the light intensity are sensitivity, shutter exposure and aperture. As stated earlier it is desired to have a sensitivity low enough as not to create excessive noise on the image. The shutter exposure time is desired to be short enough that no motion blur is introduced onto the image. The decrease in both the sensitivity and exposure time increases the amount of light required to pass through the lens to produce an image which is sufficiently illuminated. This thus requires the aperture of the lens to be set as large as possible to let in more light. This however has a detrimental effect on the depth of field of the scene.

The depth of field of a scene is the distance between the nearest and furthest object in a scene which is deemed sufficiently sharp or in focus. A lens can only perfectly focus an object at one distance. Any object further away or closer than the point of focus becomes blurred. The further away the object the more out of focus or blurry the object becomes An object is deemed in focus if the light from the object is spread over a circle smaller than the Circle of Confusion (CoC). The CoC is the circle size of which the area is smaller than the perceptible resolution of the human eye. Therefore, any circle smaller CoC may not have a perceptible change in clarity to the unaided human eye [Jacobson et al., 2000]. The standard value of the CoC is dependent on image format since it is a factor of the amount of enlargement between the image format and the final image. The larger the image format the larger the CoC due to a smaller enlargement factor between image format and final image.

The distance of the nearest object D_N and the farthest object D_F in focus is given in equations 2.1 and 2.2 [Jacobson et al., 2000].

$$D_N = \frac{Zf^2}{f^2 + f_{number}CoC(Z-f)} \tag{2.1}$$

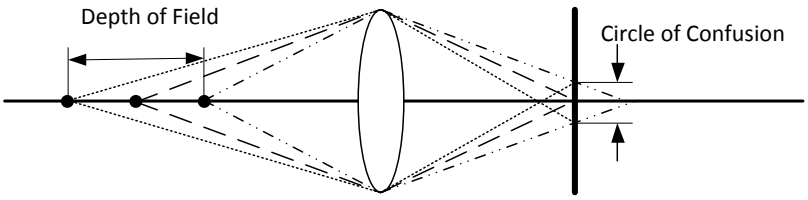
$$D_F = \frac{Zf^2}{f^2 - f_{number}CoC(Z-f)} \tag{2.2}$$

where Z is the distance of the object in precise focus, f is the focal length of the lens, f_{number} is the aperture opening. Since the CoC is fixed for the imaging format the three values that can be altered to change the depth of field are the.

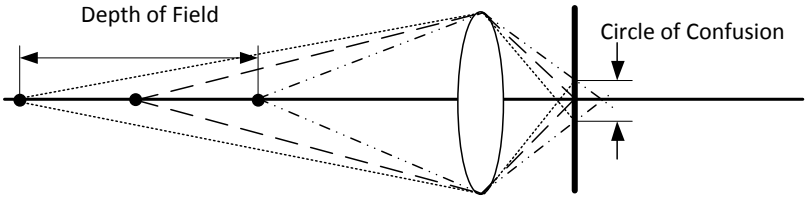
1. Aperture opening f_{number}
2. Distance of the object Z
3. Focal length of lens f

The effect the focal length has on the depth of field is shown graphically in Figure 2.8. A smaller focal length increases the depth of field. However, changing the focal length of the lens has the effect that should the depth of field be increased the area in view also increases. This effectively means that the resolution of the desired object is reduced since a larger region is imaged. The effect the aperture and distance to the object has on the depth of field is shown graphically in Figure 2.9. With a further placed object the depth of field is increased. However, it has the same effect as the focal length that the effective resolution of the object deteriorates. Thus, the only parameter which can be adjusted to increase the focal point without affecting the region of interest is the aperture.

It is thus often required to have a small aperture opening to increase the depth of field which improves the ability to adequately focus on the object as well as remove sensitivity to motion in the direction of the lens. Therefore, it is often required that the three parameters, sensor sensitivity, exposure time and aperture opening, to be as low as possible to reduce noise, reduce blur due to motion and improve depth of field respectively. However, decreasing all of these values effects the brightness of the image. Therefore, a compromise between the parameters must be reached, giving preference to the parameters which are more critical in the desired application. For instance if the object motion is large the exposure time is desired to be as short as possible. If the object moves



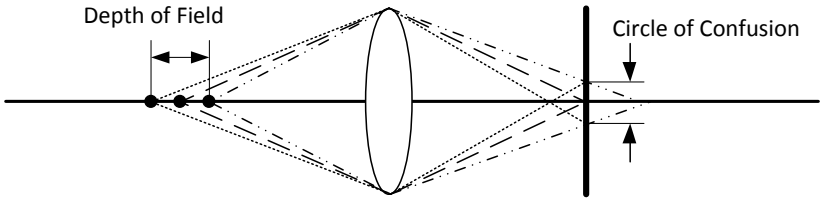
(a) Depth of field with large focal length



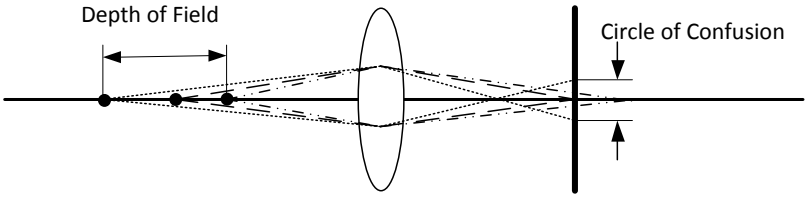
(b) Effect of smaller focal length on depth of field

Figure 2.8: Effect of focal length on depth of field

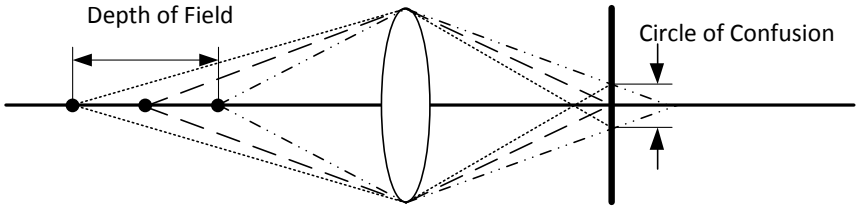
away or towards the camera the aperture size needs to be as small as possible to ensure the object remains in focus. Another alternative is to use artificial lighting either in the form of strobe lighting or simply a constant artificial light source. If the object is flooded with light it enables all three parameters to be set as low as possible to obtain a clearer image of the scene which is still adequately exposed. Adequate exposure and clear images are required if DIC algorithms are to be successfully applied to images.



(a) Depth of field determined from circle of confusion



(b) Effect of aperture on depth of field



(c) Effect of distance to focus point on depth of field

Figure 2.9: Effect of lens settings on depth of field

2.3 Conclusion

The basics of imaging in terms of the hardware aspect as well as the effect of the camera and lens settings have been discussed in this chapter. Having knowledge of the effects of these settings is vital in capturing images which are not blurred or underexposed. It is desired to have images which are sharp, with proper exposure which retains the clarity of the original scene to increase the effectiveness of the DIC techniques. It is shown that often a compromise between settings are needed in order to obtain the desired effect. In Chapter 3 the modelling of cameras in order to remove distortion and allow for depth to be determined using multiple cameras is discussed.

Chapter 3

Mathematical Camera Models and Calibration

In Chapter 2 the focus was on how to obtain clear digital images of a scene on which are suitable to use in DIC algorithms. The digital representation of an image is obtained in the form of an array of pixels. Where each pixel contains 3 byte values representing red, green and blue light colour spectrum or a single byte value for grey scale images. However, often these images may contain lens distortion which must be removed prior to DIC correlation. Or alternatively multiple views, usually from two cameras or more, need to be calibrated against one another to allow depth to be determined. In order to achieve this, a mathematical camera representation and calibration techniques for cameras are needed. In this chapter the mathematical representation of a camera as well as calibration techniques used to remove lens distortion and calibrate between multiple camera views are discussed. This study does not attempt to develop or analyse any of the calibration algorithms. The study makes use of open source implementations of these algorithms. The aim of this chapter is therefore to merely give an underlying understanding of camera models and the requirements for calibration.

3.1 Pinhole Camera Model

The mathematical model often used to represent a camera is the simplistic pinhole camera [Prince, 2012]. In the simple pinhole camera model, a single light ray enters from a particular point in a scene or object and is projected onto the imaging plane. The image is assumed to always be in focus and the

relative size of the image and object are related by their distances to and from the camera as well as the focal length of the lens. This relationship can easily be determined from the geometry shown in Figure 3.1 as:

$$x = -f\frac{X}{Z} \tag{3.1}$$

where x is the height of the object in the image plane, X is the object height and Z is the distance from the camera along the optical axis. The negative sign indicates the inversion effect of the lens.

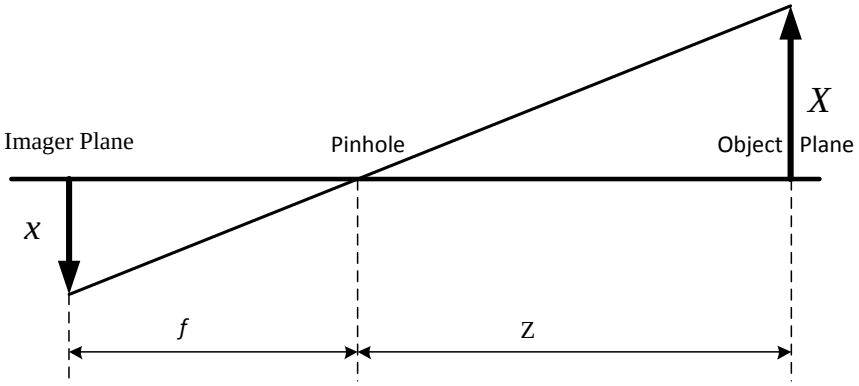


Figure 3.1: Basic 2 dimensional camera model

In order to simplify the analysis, without effecting the overall nature of the problem, the imaging plane is often placed in front of the camera on the same side as the object. This creates an equivalent problem which only removes the inversion effect and simplifies the mathematical analysis. In a more general analysis the geometry of the analysis is represented in Figure 3.2. From this geometry it can be shown that:

$$x = f\frac{X}{Z} \tag{3.2}$$

$$y = f\frac{Y}{Z} \tag{3.3}$$

where x and y are the coordinates of the projected point on the image plane. It is however more desirable to relate the projection not onto the image plane but rather the image sensor itself. One of

the differences is that the origin of the image sensor does not necessary intersect with the optical axis. This is due to manufacturing tolerances as well the selection of the sensor origin which is generally selected to be in one of the corners of the sensor rather than the centre. Another difference is that the focal length of the sensor and lens combination is represented in terms of *pixels* rather than physical distance in *mm* for just the lens. This focal length is also not necessarily the same in the x and y direction due to the pixels often having rectangular as opposed to square shapes. The sensor's focal length is given as the product of the actual lens focal length in *mm* and the pixel size S_x and S_y in *pixel/mm*. In this study the lens focal length f is used in all geometric relationships, with the pixel size S_x and S_y used to convert the geometric distance to pixel distance. The sensor focal length is therefore given as:

$$f_x = S_x f \tag{3.4}$$

$$f_y = S_y f \tag{3.5}$$

from which the pixel coordinates of a point in the real world projected onto the image sensor are obtained as:

$$x = f_x \frac{X}{Z} + c_x \tag{3.6}$$

$$y = f_y \frac{Y}{Z} + c_y \tag{3.7}$$

where c_x and c_y are the pixel distances from the sensor origin to the intersection of the optical axis on to the sensor.

3.2 Homogeneous Coordinates

The homogeneous coordinates system is a coordinate system used to represent projective geometry. The coordinate system not only simplifies the problem of projecting world coordinates onto a plane, such as the image plane, but also introduces a method of representing a point at infinity using finite coordinates. It allows representation of affine and projective transformations using a single matrix. For this reason it is often used in computer graphics [Riesenfeld, 1981] and camera systems [Hartley et al., 1992]. Homogeneous coordinates can be used to represent any number of dimensions. It is however often used to present both 2D and 3D Euclidean coordinates. The representation

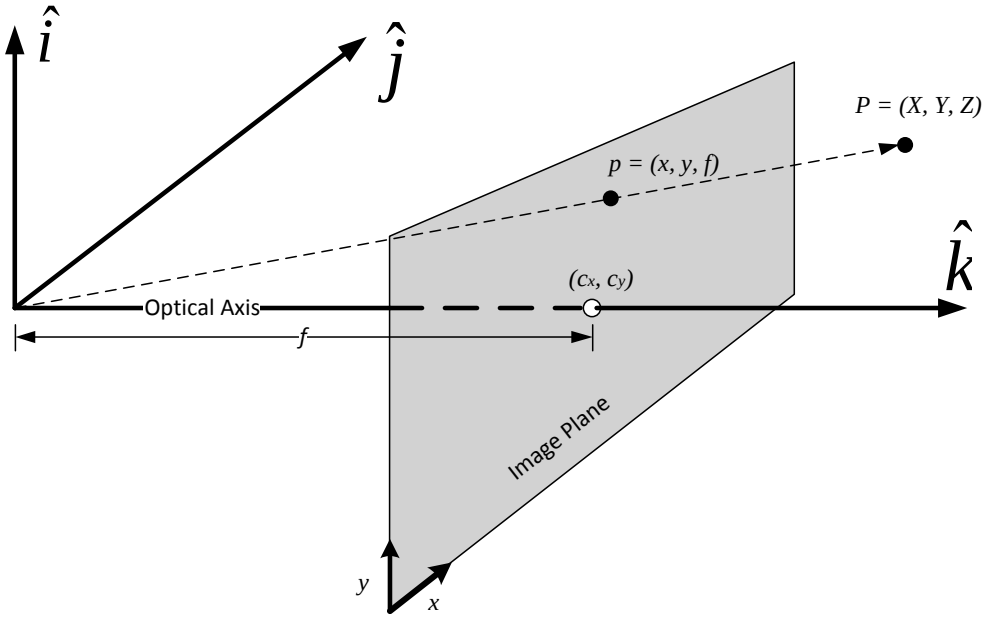


Figure 3.2: General 3 dimensional pinhole camera model

of 2D Euclidean space in homogeneous coordinates is described since it is easier to visualise and illustrate therefore making the concept of homogeneous coordinates easier to grasp. In homogeneous coordinates an extra dimension is added, thus 2D points are represented by three values and 3D is represented by four and so forth. This can be viewed as representing each point by means of a line which intersects the origin of the coordinate system. This is illustrated in Figure 3.3. The mapping of points from Euclidean to homogeneous coordinates is easily achieved by amending the point with an additional value, as:

$$\boldsymbol{p} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} xw \\ yw \\ w \end{bmatrix}, w \neq 0 \tag{3.8}$$

where w is any non-zero scale value. The concept of homogeneous coordinates is that the 2D points can be represented by any combination of triple $\begin{bmatrix} xw & yw & w \end{bmatrix}^T$ and that the 2D Euclidean plane is mapped in homogeneous coordinates when $w = 1$. Thus, we can recover the original 2D point

simply by dividing by the last value in homogeneous coordinates, as:

$$\boldsymbol{p} = \begin{bmatrix} xw \\ yw \\ w \end{bmatrix} = \frac{1}{w} \begin{bmatrix} xw \\ yw \\ w \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}, w \neq 0 \tag{3.9}$$

This concept in camera projections is useful as it represents a light ray being projected onto a plane. As in camera systems the depth of the point mapped on the the image plane cannot be determined and simply represents a single light ray emanating from the centre of projection. This makes homogeneous coordinates ideal to represent projective geometry. In homogeneous coordinates a point at infinity is represented when $w = 0$, thus infinity can be represented by finite values.

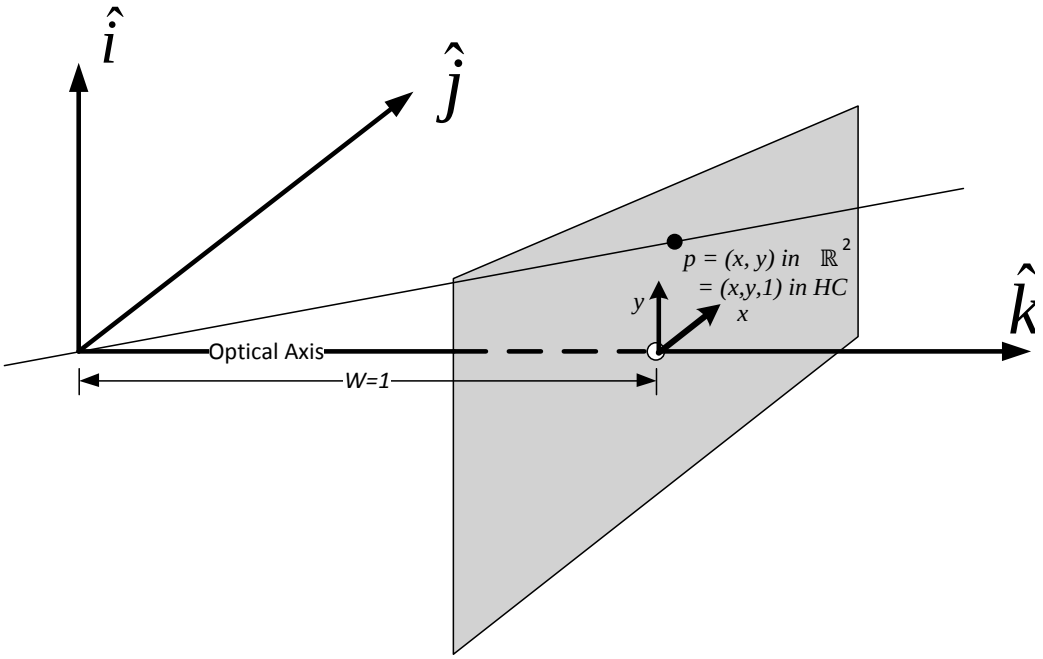


Figure 3.3: Representation of 2D Euclidean coordinates in homogeneous coordinates

Using homogeneous coordinates the mapping of the world coordinates onto the image plane can be presented by the matrix transform (derived [Heikkila and Silvén, 1997]):

$$\boldsymbol{p} = \boldsymbol{KP}, \text{ where } \boldsymbol{p} = \begin{bmatrix} xw \\ yw \\ w \end{bmatrix}, \boldsymbol{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{P} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.10}$$

multiplying out equation 3.10 it is found that $w = Z$ which gives equations 3.6 and 3.7. This enables the projection of world points onto the image plane to be presented using a single matrix $\boldsymbol{K}$. The matrix $\boldsymbol{K}$ is known as the camera intrinsic matrix as it contains parameters internal to the camera and lens. Due to homogeneous coordinates simplifying the method of mapping coordinates from the real world onto the image plane it is often used to simplify the calibration process in cameras as explained in section 3.4.1.

3.3 Lens Distortion

Many factors in the camera system can cause distortions on the projected image. The two forms of distortions which are most prevalent are that of radial and tangential distortion.

3.3.1 Radial Distortion

As a result of inferior or cheap manufacturing processes, lenses will distort the scene. The most common form of distortion is that of radial distortion. Radial distortion is a lens which fails to be rectilinear, that is to say that straight lines in a scene a warped when projected onto the image sensor. This effect is especially pronounced near the edges of the image. This due to the lens refracting the rays further away from center of the lens more than the rays close to the center. This phenomenon is most noticeable on cheap lenses as well as wide angle lenses. In wide angle lenses the distortion takes the form of so called barrel distortion, named so due to the distortion shape representing a barrel. In low-end telephoto lenses the distortion has a different appearance and resembles a pincushion of which the corners flare outward. Due to its appearance it is often referred to as pincushion distortion. Barrel distortion is graphically depicted in Figure 3.4.

In order to correct for lens distortion a model is fitted which approximates the distortion of the lens. The radial lens distortion is given as [Slama et al., 1980]:

$$x_c = x + \delta x_r \tag{3.11}$$

$$y_c = y + \delta y_r \tag{3.12}$$

$$\delta x_r = x(k_1 r^2 + k_2 r^4 + ...) \tag{3.13}$$

$$\delta y_r = y(k_1 r^2 + k_2 r^4 + ...) \tag{3.14}$$

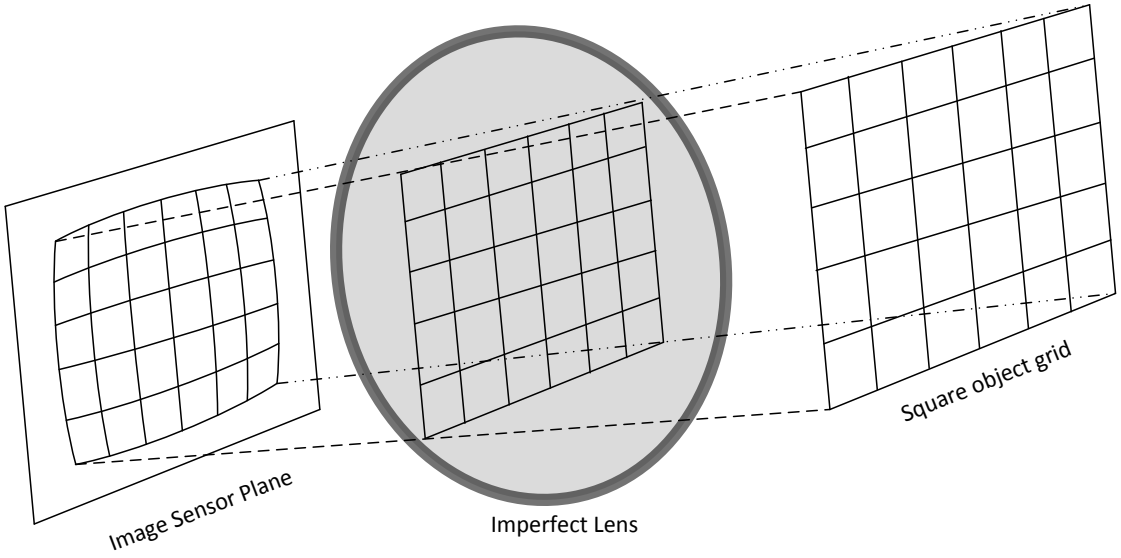


Figure 3.4: Imperfect lens causing barrel radial distortion of square grid on image sensor plane

where x_c and y_c are the corrected coordinates by adding the radial distortion corrections δx_r and δx_r. The coefficients k_i are the radial lens distortion parameters of the Taylor expansion and $r^2 = x^2 + y^2$ is the radial distance from the optical centre. For most lenses the distortion is low and can be approximated with a second order expansion having two parameters. With very wide angle (fish eye) lenses an extra order is sometimes needed to better approximate the distortion, this gives upto three parameters needed to model radial distortion.

3.3.2 Tangential Distortion

Another form of distortion is tangential distortion. This distortion arises when the lens is not parallel to the imaging plane due to poor manufacturing or assembly. This form of distortion causes the projected image to be slanted on the image plane as shown in Figure 3.5

Tangential distortion correction model contains two parameters, p_1 and p_2, and is modelled using the following [Slama et al., 1980]:

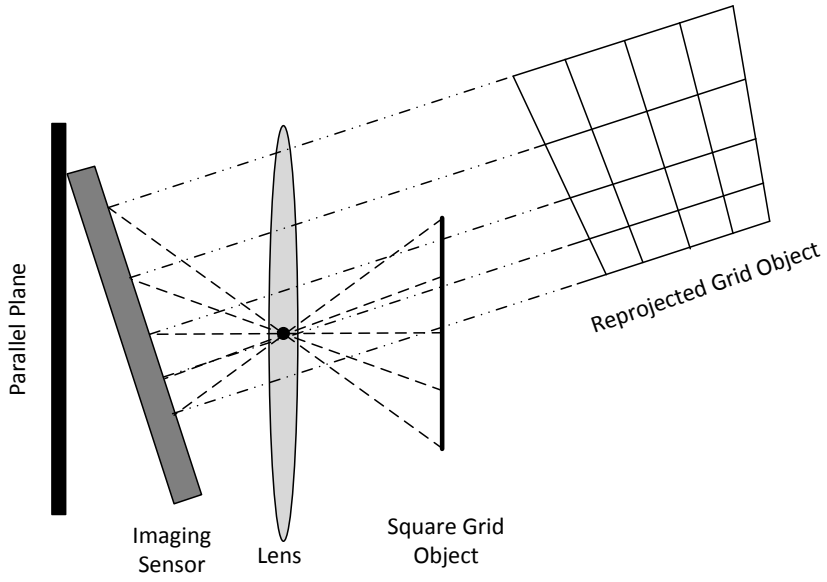


Figure 3.5: Skew lens and sensor causing tangential distortion of a square grid

$$x_c = x + \delta x_t \tag{3.15}$$

$$y_c = y + \delta y_t \tag{3.16}$$

$$\delta x_t = \left[2p_1 y + p_2(r^2 + 2x^2)\right] \tag{3.17}$$

$$\delta y_t = \left[p_1(r^2 + 2y^2) + 2p_2 x\right] \tag{3.18}$$

3.3.3 Combined Radial and Tangential Distortion

Correcting for both radial and tangential distortion we add the distortion corrections as:

$$x_c = x + \delta x_r + \delta x_t \tag{3.19}$$

$$y_c = y + \delta y_r + \delta y_t \tag{3.20}$$

$$x_c = x + x(k_1 r^2 + k_2 r^4 + ...) + \left[2p_1 y + p_2(r^2 + 2x^2)\right] \tag{3.21}$$

$$y_c = y + y(k_1 r^2 + k_2 r^4 + ...) + \left[p_1(r^2 + 2y^2) + 2p_2 x\right] \tag{3.22}$$

Therefore, to correct for the most common lens distortions requires solving for at least 4 parameters, two for radial and two for tangential distortion. With lenses that give more radial distortions a higher order radial model is used increasing the number of distortion parameters.

3.4 Single View Calibration

In order to successfully calibrate a camera it is required to solve for the camera intrinsic parameters in section 3.2 and the distortion coefficients in section 3.3. One way to solve for the intrinsic parameters is to solve directly using equation 3.10. This method entails using several real world points projected onto the imaging plane. This however requires knowledge of the exact world coordinates of the points $\boldsymbol{P}$ projected onto the image plane to obtain points $\boldsymbol{p}$. From the known real world locations and projected locations the intrinsic parameters can be solved for. Once the intrinsic parameters are known a pinhole camera model can be used to determine where the real world coordinates should have been projected onto the image plane without lens distortion effecting the projection. These pinhole projections are equivalent to the "corrected" points (x_c and y_c) after lens distortion is removed. The difference between the actual projected locations (x and y) and the theoretical pin hole projections (x_c and y_c) along with the distortion model equations (3.21 and 3.22) are then used to solve for the distortion coefficients. The intrinsic parameters can then be re-estimated using the projected point corrected using the distortion models.

While the methodology of using known real world points is not impossible it is very difficult and tedious to perform and requires accurate measurements of the real world coordinates. Therefore, an

alternative method is used which utilises points in the real world which lie on a planar surface. In order to understand this process requires knowledge of a special transform known as planar homography.

3.4.1 Planar Homography

Planar homography refers to the projective mapping of a planar surface onto another planar surface. In camera systems we are interested in the mapping of a 3D plane onto the 2D plane of the camera sensor. In homogeneous coordinates the mapping is obtained by means of a single matrix multiplication. With both the point on the plane in the real world and the projected point on the sensor in homogeneous coordinates the mapping is found as:

$$\boldsymbol{p} = s\boldsymbol{H}\boldsymbol{P}, \text{ where } \boldsymbol{P} = \begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T, \boldsymbol{p} = \begin{bmatrix} x & y & 1 \end{bmatrix}^T \tag{3.23}$$

where s is an arbitrary scalar value. As in section 3.2 it indicates that homography transformation $\boldsymbol{H}$ is only defined up to a scale factor (due to the universal scale ambiguity). The homography transformation consists of two parts. Firstly a physical transformation consisting of rotational and translational motion which locates the plane in 3D Euclidean space. The second part is the projection onto the imaging plane, therefore the homography matrix can be described as:

$$\boldsymbol{H} = \boldsymbol{K}\boldsymbol{T} \tag{3.24}$$

where $\boldsymbol{K}$ is the intrinsic camera matrix of equation 3.10, $\boldsymbol{T}$ is the transformation matrix containing the rotational and translational transformation. The rotation is encoded into a 3×3 rotational matrix which incorporates all three Euler angle rotations. The Euler angle rotations are written in matrix form as:

$$\boldsymbol{R} = \boldsymbol{R}_z(\theta)\boldsymbol{R}_y(\phi)\boldsymbol{R}_x(\psi) \tag{3.25}$$

$$\boldsymbol{R}_x(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \tag{3.26}$$

$$\boldsymbol{R}_y(\phi) = \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \tag{3.27}$$

$$\boldsymbol{R}_z(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.28}$$

The translation is simply presented by a single column vector with three values representing the translational in the x, y and z axis as:

$$\boldsymbol{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{3.29}$$

These two matrices are written into one 3×4 matrix to represent the physical rotational and translational transformation of the plane:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \end{bmatrix} \tag{3.30}$$

This transformation matrix $\boldsymbol{T}$ is also known as the extrinsic parameters as they describe the location and orientation of an object relative to the camera. The use of the homography matrix will be shown to be very useful when performing camera calibration as discussed section 3.4.2.

Therefore the homograhic transform in equation 3.23 can be rewritten as:

$$\boldsymbol{p} = s\boldsymbol{H}\boldsymbol{P} \tag{3.31}$$

$$\boldsymbol{p} = s\boldsymbol{K}\begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \end{bmatrix}\boldsymbol{P} \tag{3.32}$$

The homography matrix is a 3×3 matrix consisting of nine unknowns. However the last value is simply a normalisation value which leaves eight unknowns to be solved. Therefore, four non-collinear points, of which each point yields two constraints one in x and another in y, are required in order to estimate the homography matrix. It is easy to visualize this by imagining the four points representing the corners of the planar surface which enables the mapping of a square into any quadrilateral. While more points can be used to estimate the homography matrix in a least squares sense, this is only useful in noisy environments since any more additional points will not provide additional unique constraints.

3.4.2 Calibration Using Planar Surface

As mentioned in section 3.4 using real world coordinates to obtain all the calibrated parameters is a difficult process. Instead it is possible to utilise points on a planar surface. The 3D coordinates of the points on the planar surface is known relative to the planar surface's axis system. The origin of the axis system is typically located at the first point. The position and rotation, i.e. extrinsic parameters, of the planar surface's axis system relative to the camera is however unknown. This is shown in Figure 3.6. A planar surface with N points ($N >= 4$) provides $2N$ constraints. It is required that at least ten constraints be provided to solve for the four intrinsic parameters and the six extrinsic parameters of the planar surface. It therefore seems that a single view from a planar surface with at least five points is sufficient to solve for all unknowns. However, as stated in section 3.4.1 a planar surface will at most yield eight unique constraints from four points. Therefore, multiple views of the planar surface is needed. Each view gives at most eight constraints but introduces six new unknowns since the extrinsic parameters change between views while the intrinsic parameters remain constant. Therefore, the minimal number of views K needed are determined as:

$$N_{constraints} \geq N_{Unknowns} \tag{3.33}$$

$$8K \geq 6K + 4 \tag{3.34}$$

$$K \geq 2 \tag{3.35}$$

Therefore, a minimal of two views are needed to solve for both the extrinsic and intrinsic parameters. However, in practice more than two views are used to improve estimation of the intrinsic parameters with each view containing more than four points to improve estimation of the homography matrix in the presence of noise. Initially the distortion parameters are set to zero, the intrinsic

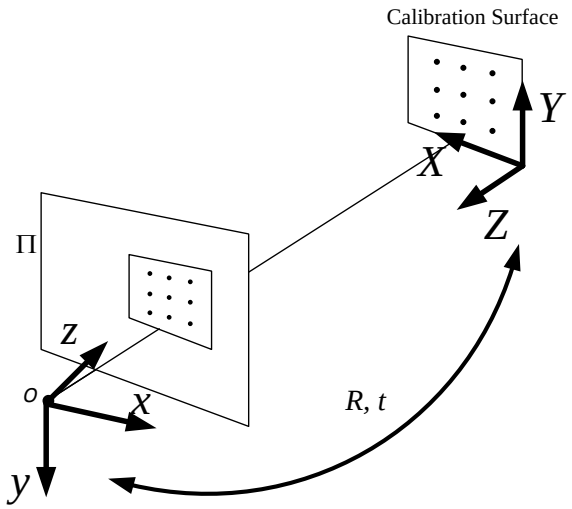


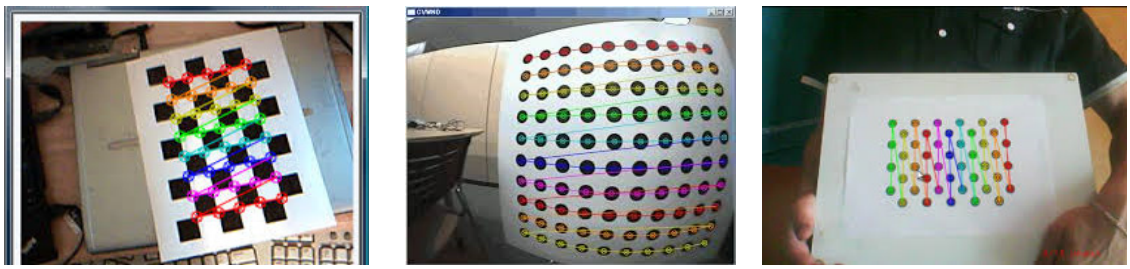
Figure 3.6: Calibration of camera

parameters are then estimated. The intrinsic parameters are used with a pinhole camera model in order to determine the distortion parameters. It is assumed that the pinhole camera model is perfect and is therefore used to project the planar surface points onto the imaging plane. These projected point are assumed to be the undistorted points (x_c and y_c) and the actual measured points on the sensor is the distorted points. Multiple points are then used to estimate the distortion coefficients. The distortion model is then used to correct the measured point, which are in-turn used to re-estimate the intrinsic camera parameters. The detail of how the parameters are estimated are not part of the scope of this study but can be found in [Zhang, 2000].

3.4.3 Calibration Using OpenCV

The calibration process in this study is performed using the open source library Open Compute Vision (OpenCV) [Bradski, 2000]. The library incorporates a calibration algorithm which requires the coordinates of the planar points relative to the planar surface's axis system as well as the corresponding projected image sensor coordinates. This is accomplished by searching for the planar points in the imaging sensor. There are several functions which can be used to perform this. All the functions rely on the points firstly having a specific layout, i.e. square grid, asymmetric grid and be of known dimensions. The dimensions are simply the number of points per row and column. The second

requirement is that the points can easily be found in the image by having a high contrast. This can be achieved by creating a black dot on a white surface. Two algorithms are mainly used namely an algorithm which uses a chessboard pattern to identify corner locations in the chessboard and a second algorithm that identifies circular points in either a square or asymmetrical grid pattern. The identification of the planar points which make up these grid patterns are shown in Figure 3.7. The axis system of the planar surface is often selected such that the origin is at the first point. The directions of the axes are chosen that the X and Y axis are orientated column and row wise respectively. The Z is away from the planar surface with all planar point lying on the $X-Y$ plane i.e. all for all points $Z=0$. The planar axis system therefore moves with the surface but is solved for in each view. With enough points obtained from multiple views it is possible to estimate the camera intrinsic and lens distortion parameters as discussed in section 3.4.2.



(a) Chessboard pattern (b) Symmetric circle grid (c) Asymmetric circle grid

Figure 3.7: Planar surface points pattern used in calibraton

3.5 Multiple View Calibration

Section 3.4 dealt with the calibration and removal of distortion of a single camera. This section deals with the calibration over multiple views. This form of calibration is used to obtain the relative position and orientation between the two views. The views can be obtained from either two cameras imaging the same scene at the same time, often referred to as stereographic vision, or of two views taken from the same camera imaging a scene from different locations. In the former the views are taken at the same time instance while in the latter there is a finite time difference between the two views allowing for the camera to undergo motion between the two views.

When the two views are obtained from two cameras, the system is referred to as a stereographic rig. The cameras are fixed relative to one another such that their position and orientation does not

change in time. This allows for the calibration procedure to be performed once creating a calibrated stereographic rig. This rig allows the calculation of the depth of the scene being viewed. When the two views are from the same camera the calibration procedure is performed between each new pair of the scene since the relative orientation and position is not consistent.

3.5.1 Epipolar Geometry

The geometry governing multiple views is called epipolar geometry. This geometry relates the scene in one image to the scene in another. The epipolar geometry is displayed in Figure 3.8. Point $\boldsymbol{P}$ is projected onto the left image plane $\prod$ and right image plane $\prod'$ to obtain points $\boldsymbol{p}$ and $\boldsymbol{p'}$ respectively. The epipolar points e is the projection of the center of projection $\boldsymbol{O'}$ of plane $\prod'$ onto the plane $\prod$. The other epipolar point $\boldsymbol{e'}$ is the projection of the other plane's $\prod$ center of projection $\boldsymbol{O}$ onto the plane $\prod'$. The lines $\boldsymbol{ep}$ and $\boldsymbol{e'p'}$ represent the epipolar lines.

From section 3.4 it is noted that using a single view, a point projected onto the image plane can only be determined as a ray joining the centre of projection and the projected point, as no depth can be extracted. The significance of the epipolar lines is that the ray $\boldsymbol{OP}$ in plane $\prod$ (the left image) is projected onto the plane $\prod'$ (the right image) as the epipolar line $\boldsymbol{e'p'}$. Therefore, the epipolar line $\boldsymbol{e'p'}$ contains the projection of all points along the line $\boldsymbol{OP}$. This means that in a stereographic rig a feature in one image must lie in the corresponding epipolar line, also known as the epipolar constraint. The significance of this will be taken advantage of in a stereographic rig in section 4.2.

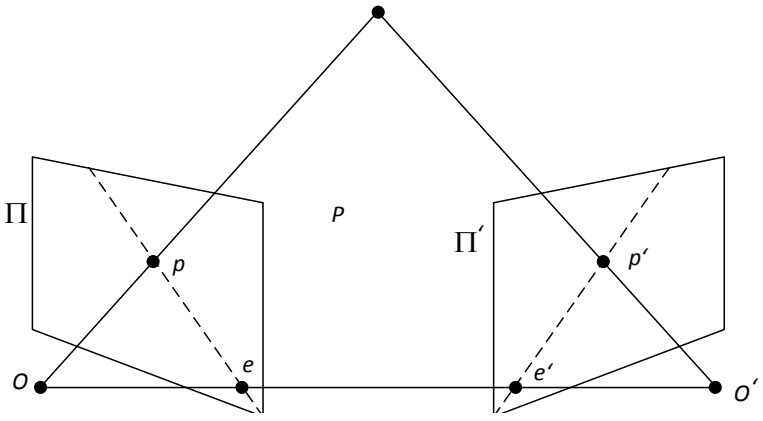


Figure 3.8: Epipolar geometry of multiple view system

3.5.2 Stereography to Determine Depth

It is possible to determine the depth of a feature by using two views. This is accomplished by back projecting the ray of the same feature from the two views. The point at which these rays intersect determines the location of this feature in space. If the relative orientation and translation of the two cameras are known then the exact 3D location relative to the camera can be determined. This is similar to one method the human eye uses to determine depth [Qian, 1997]. An important parameter in stereographic vision is binocular disparity. Binocular disparity is the positional difference between the two images' projection of a certain point in space. The positional difference in these views are a result of the views having a translational and rotational separation between them. This concept is illustrated in Figure 3.9. The figure shows three features which are projected onto two views with a lateral separation between them. It can be observed that the point closest to the cameras has a high degree of positional separation between the two views and that this separation becomes less, the further a point is away from the cameras. This phenomenon is easily demonstrated by placing a finger in front of the eyes. If one eye is opened at a time it is possible to estimate the positional separation between the two eyes. If the finger is moved further away the positional separation decreases and vice versa if the finger is moved closer.

It was stated that it is possible to determine the 3D position of a feature by determining its disparity between two views and having knowledge of the geometric relationship between the two views. The method for determining the disparity between two views is discussed in section 4.2. The remainder of the chapter discusses how the geometric position between two cameras can be obtained to create a calibrated stereographic rig.

3.5.3 Calibration

The calibration process for a stereo camera rig involves determining the geometric relationship between two cameras. This can easily be accomplished by making use of the single camera calibration discussed in section 3.4. The calibration process involves both cameras taking views of the same planar calibration surface at the same time. From the calibration process the intrinsic and distortion parameters are determined along with the extrinsic parameters (positional relationship between camera and calibration surface). For each view of the calibration surface a combination of rotational matrices ($\boldsymbol{R}$ and $\boldsymbol{R'}$) and translational vectors ($\boldsymbol{t}$ and $\boldsymbol{t'}$), as illustrated in Figure 3.10, are obtained

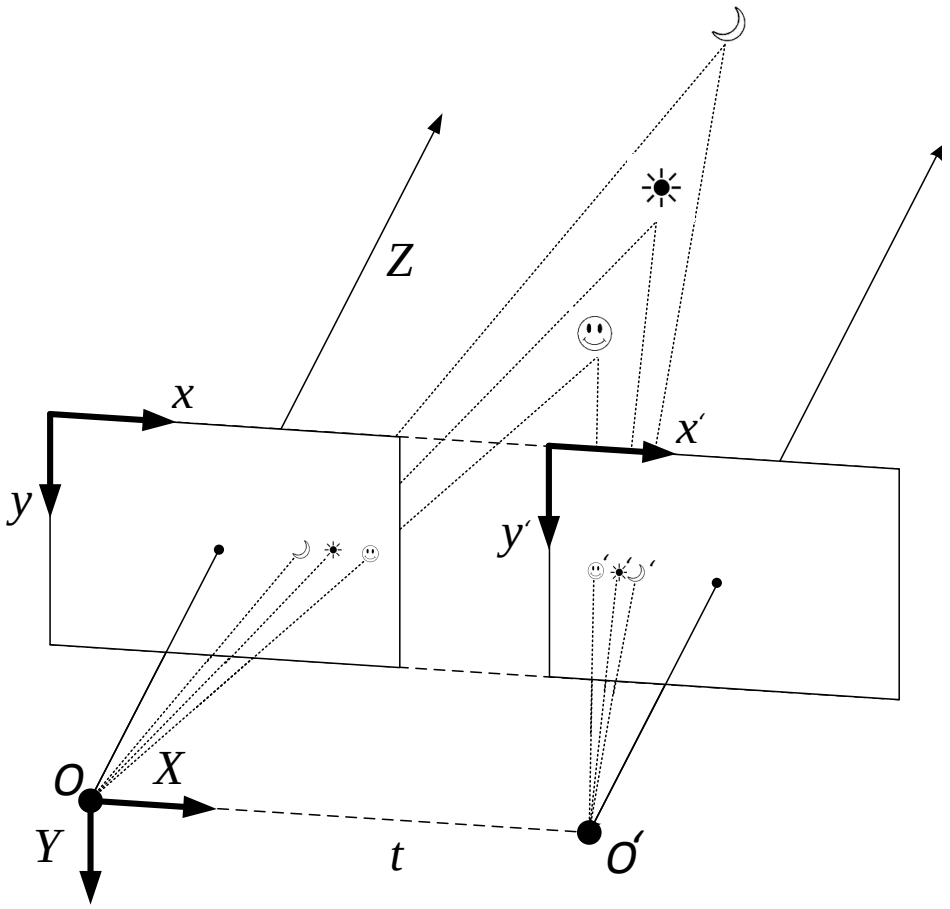


Figure 3.9: Stereovision geometry

for the left and right views respectively. Using geometry it can be found that the relative orientation $\boldsymbol{R_{stereo}}$ and translation $\boldsymbol{t_{stereo}}$ between the two camera is obtained from:

$$\boldsymbol{R_{stereo}} = \boldsymbol{R'R}^T \tag{3.36}$$

$$\boldsymbol{t_{stereo}} = \boldsymbol{t'} - \boldsymbol{R_{stereo}t} \tag{3.37}$$

Thus, it is possible to calibrate the stereo rig by applying the single view calibration to both cameras at the same time. Multiple of these views are used to obtain the average of the relative position and orientation between the cameras.

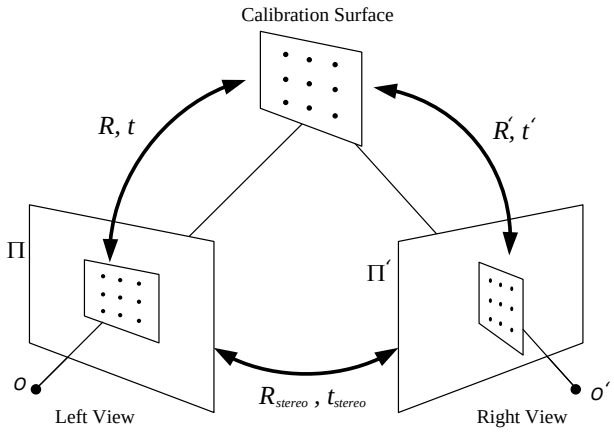


Figure 3.10: Relative orientation and translational position between two views

3.6 Conclusion

In this chapter the calibration process of a camera was discussed. A method for determining the intrinsic camera parameters and distortion coefficients was shown. The calibration process allows the images to be rectified such that a straight line in the real world corresponds to a straight line in the projected image plane. In the multiple view calibration process the relative orientation and translation between two fixed cameras can be obtained using the single camera calibration process. The calibration process was shown to simplify obtaining the disparity of a feature in section 4.2. The disparity and relative orientation and position of the cameras allows the depth of a feature to be determined in real world coordinates. Once the cameras are calibrated and the images rectified DIC techniques can be applied to the images in order to track particles and determine the 3D position of particles. The DIC techniques used in this study are discussed in Chapter 4.

Chapter 4

Digital Image Correlation Techniques

In Chapters 2 and 3 the basic understanding of cameras and how they are modelled was presented. In this chapter the basic principle of most algorithm used in this study is presented. The chapter does not go into detail of the numerical methods used in the implementation of each algorithm, the study also makes use of open source implementations of these algorithms. The aim of this chapter is therefore to provide a basic understanding of the algorithms needed for subsequent chapters.

Digital Image Correlation (DIC) is a method whereby optical methods are used to track changes between images. In order to track changes between images all motion trackers require that the surface contain some arbitrary features which create brightness variations across the image. The brightness variations can be used to differentiate regions from one another, thus monotone surfaces may be very difficult to apply DIC techniques to. However, most terrains have arbitrary features such as grooves, stones and other discolourations which can be used to identify unique regions. In certain measurements an arbitrary pattern is created on the surface. Often this pattern is created by using spray paint to generate a speckle pattern on the surface which has a different contrast to the surface.

DIC methods can be used to obtain a variety of parameters from displacements and velocities of particles to the strain of an object in an image [Xavier et al., 2012]. There are numerous methods which can be used to perform DIC, however the underlying methodology remains the same. The techniques try to match regions in one image to regions in another, effectively tracking these regions across images. The images can be from different views each obtained with a different camera, or make up a time series of images, $\boldsymbol{I_t}$, $[t = 0, 1, \ldots, n]$, taken with a finite time difference apart.

The image correlation can be performed on the whole image by tracking all regions in the image in another image. This often involves an exhaustive search across the image which can be a very time consuming process and can yield many false correspondences. Two techniques can be used to speed up and improve correspondence matching. The first technique is to perform the matching on a smaller subset of points which are unique and will provide less false matched points. Another technique is to make use of special geometry to reduce the exhaustive search across the whole image to a much smaller search region. The geometry technique is often used in calibrated stereo rigs to perform fast full image correlation.

4.1 Sparse Image Correlation

In sparse image correlation a smaller subset of features, which are unique in the image and therefore easy to track, are obtained and tracked from one image to another.

4.1.1 Corner Features

In order to track a subset of regions in an image first requires the identification of regions in the image which are easily identifiable and therefore easy to track. A simplistic feature to track in images are corner or edge features. These are features which have a change in texture (change in pixel intensity/contrast) in at least two directions. This is generally achieved by taking the second derivative of the image intensity in various directions to obtain the hessian matrix $\boldsymbol{H}(I)$ as:

$$\boldsymbol{H}(I) = \begin{bmatrix} \frac{\delta^2 I}{\delta x^2} & \frac{\delta^2 I}{\delta x \delta y} \\ \frac{\delta^2 I}{\delta x \delta y} & \frac{\delta^2 I}{\delta y^2} \end{bmatrix} \tag{4.1}$$

A region is classified as a good feature to track if the smallest eigenvalue of $\boldsymbol{H}(I)$ is above a certain threshold, as well as the relative ratio of the eigenvalues are below another threshold [Shi and Tomasi, 1994]. This results not only in pixels with good contrast but gives preference to corner points instead of edge points. This method is also applied over a small window around the pixel to find the average gradients over a region instead of at a single pixel. This is accomplished as:

$$\boldsymbol{M} = \int\int_{Window} \boldsymbol{H}(I) w \delta x \delta y \tag{4.2}$$

$$\boldsymbol{M} = \begin{bmatrix} \sum_i^n \sum_j^m \frac{\delta^2 I}{\delta x^2} w & \sum_i^n \sum_j^m \frac{\delta^2 I}{\delta x \delta y} w \\ \sum_i^n \sum_j^m \frac{\delta^2 I}{\delta x \delta y} w & \sum_i^n \sum_j^m \frac{\delta^2 I}{\delta y^2} w \end{bmatrix} \tag{4.3}$$

where the $Window$ is of size $n \times m$ and w is a weighting function which can be either uniform with all points having equal weighting or Gaussian giving points closer to the centre more weight.

These locations can be further enhanced by obtaining the location in a sub-pixel manner. A simple method to achieve this would be to fit a curve, for example a parabola, to the pixel values surrounding an edge. The exact corner or edge location can then be found where the gradient is a maximum. Newer techniques exist which achieve the same result [Lucchese and Mitra, 2002], [Yimin et al., 2010] using more complex methods. These features are then tracked from one image to the next using optical flow techniques. One of the most widely used techniques for sparse optical flow (small subsets of regions are tracked instead of the whole image) is the Lucas-Kanade optical flow algorithm [Lucas et al., 1981]. The algorithm tracks the feature by making use of the brightness consistency i.e. a feature's brightness does not change in time. Therefore, the brightness intensity of a pixel can be described by:

$$I(x, y, t) = I(x + dx, y + dy, t + 1) \tag{4.4}$$

$$I(x, y, t) = I(x, y, t) + \frac{\delta I}{\delta x} dx + \frac{\delta I}{\delta y} dy + \frac{\delta I}{\delta t} dt + higher\, order\, terms \tag{4.5}$$

$$I(x, y, t) \approx I(x, y, t) + \frac{\delta I}{\delta x} dx + \frac{\delta I}{\delta y} dy + \frac{\delta I}{\delta t} dt \tag{4.6}$$

where $\frac{\delta I}{\delta x}$ and $\frac{\delta I}{\delta y}$ are the brightness gradients in the x and y direction respectively and $\frac{\delta I}{\delta t}$ is the time derivative of the brightness. From equation 4.6 it can be shown that:

$$\frac{\delta I}{\delta x}dx + \frac{\delta I}{\delta y}dy + \frac{\delta I}{\delta t}dt = 0 \tag{4.7}$$

$$\frac{\delta I}{\delta x}\frac{dx}{dt} + \frac{\delta I}{\delta y}\frac{dy}{dt} + \frac{\delta I}{\delta t}\frac{dt}{dt} = 0 \tag{4.8}$$

$$\frac{\delta I}{\delta x}V_x + \frac{\delta I}{\delta y}V_y + \frac{\delta I}{\delta t} = 0 \tag{4.9}$$

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \begin{bmatrix} V_x \\ V_y \end{bmatrix} = -I_t \tag{4.10}$$

From equation 4.10 the velocity V_x in the x direction and V_y in the y direction can be determined if the brightness gradient at the pixel, as well as the change in brightness in time, is known. The time derivative is determined by determining the change in brightness of the same location over two images which are taken a finite time difference apart.

However, this equation has two unknowns and contains only one equation. The assumption that the Lucas-Kanade optical flow algorithm makes to increase the number of equations is to assume that surrounding pixels exhibit the same motion. Therefore, a number of these equations can be used to solve for the velocity of the motion. These velocities are often solved for by means of a least squares approach over all pixels in a window surrounding the feature. The algorithm then steps in this direction and performs another iteration until the exact location is found. The exact location of the feature is determined by minimising the residual:

$$\epsilon = \sum_{i}^{m}\sum_{j}^{m}[\boldsymbol{I_t}(x, y, t) - \boldsymbol{I_{t+1}}(x + dx, y + dy)] \tag{4.11}$$

The algorithm makes three assumptions on the motion in order to track corners. The first assumption is that of brightness constancy. This implies that the brightness of the feature or region being tracked, is consistent from image to image. In other words the brightness time derivative $\frac{\delta I}{\delta t}$ is caused by the motion of the feature and not due to the changing lighting conditions. This assumption, even if invalid when shadows are generated, can be forced to be valid by adjusting lighting, adjusting sensor gain, adjusting aperture or, in post processing, by image contrast adjustments.

The second assumption is that the feature's motion is relatively small. This implies that the feature does not move much from image to image. This assumption would be the first to be found invalid, as the time steps between images are often not as small as desired. The Lucas-Kanade algorithm

tries to alleviate this problem by performing an iterative approach. This improves convergence as long as the initial step is close to the solution. However, often much larger motions are observed. To circumvent this problem an additional modification is introduced which performs the tracking over layers of scaled images. Each layer contains an increasingly smaller image from the bottom to the top. Thus, each layer effectively reduces the feature motion in increasing magnitude. The algorithm is performed on the top (smallest image) and the solution is then used on the next layer down until the algorithm is applied to the bottom level which corresponds to the original image. In this way much larger motions between images can be tracked.

The third and final assumption is that a small windowed region around the corner exhibits the same motion as that of the feature. This assumption is valid most of the time when lateral motion is observed from a stationary scene, but may be invalid when multiple objects in the scene move considerably or when the feature undergoes mostly rotational motion.

4.1.2 SIFT Feature Descriptor

In 4.1.1 a simplistic feature and tracking algorithm was presented based on tracking corner features. This algorithm sometimes fails to consistently track a sufficient number of features needed to perform an analysis. This is due to the assumptions made by the algorithm being violated in certain cases. The most common assumption which is violated is the assumption of small displacements. Often much larger displacements occur than which the optical flow method can deal with. Most features are also scale dependent, therefore if the scale changes the feature either disappears or cannot be tracked satisfactorily.

As a result, a more complex feature descriptor technique is used if the simpler methods fail to track features adequately. Feature descriptors use a different means to obtain unique key points. From these key points, features which uniquely identify these regions are then extracted, effectively providing each key point with a fingerprint. Feature descriptor methods differ in approach by performing the key point and feature extraction step on both images rather than just the first image. In [Shi and Tomasi, 1994] features are identified on only the first image and tracked on the second image using optical flow algorithms. The feature descriptor methods perform the key point search and feature extraction on both images. The methods then try to match features across images. This approach allows much larger displacements to be tracked.

The method of obtaining key points and the type of features extracted often makes up the main difference between these DIC algorithms. Two robust feature descriptors often used are the Scale-Invariant Feature Transform (SIFT) [Lowe, 2004] and Speeded Up Robust Features (SURF) [Bay et al., 2008].

In this study the SIFT feature descriptor will be used. SIFT key points are determined by local minima and maxima from the result of applying Difference of Gaussians (DoG) to the image. The DoG is applied by blurring the image by means of convolving the image with a Gaussian kernel. The standard deviation σ of the Gaussian kernel is increased to increase the amount of blurring. The blurred images are then subtracted from one another to obtain the DoG images. By subtracting the images from one another only spatial frequencies between a certain range are preserved. Thus, DoG is a form of band-pass filter applied to the images. With larger standard deviations the detector enables larger minima or maxima to be detected. This enables key points to be detected at different scales. Thus, key points are not only a function of image locations but scale as well (x, y, σ).

This process is performed on numerous pyramidal scaled images (smaller resolutions images). This is in order to enlarge the scale space. Performing DoG on smaller images is more efficient than increasing the standard deviation. Each list of DoG images for a given image size is called an octave. In most implementations the scale space consists of 4 octaves, each containing 5 DoG scale levels. The scale space is shown in Figure 4.1 and the result of performing the DoG process is illustrated in Figure 4.2.

Local minima and maxima are obtained by looking at the 26 neighbouring pixels of the pixel under consideration. These neighbouring pixels contain nine pixels in the considered scale space image and eight features in the next larger and smaller scale space images. This is shown graphically in Figure 4.3a. If an extrema is found the pixel location and scale is stored. This generates a large list of possible key points. The list is reduced by removing key points of which the contrast is below a certain threshold. Also key points which describe edges rather than corners are removed. Edge key points are classified by taking the hessian matrix of the key point (equation 4.1.1) and discarding points of which the ratio between eigenvalues are above a threshold.

The process above generates a list of key point locations and scales. Next an orientation is also assigned to each key point based on the direction of the local image gradient to obtain invariance to rotation.

After key points have been assigned, features are extracted around this region. The features are

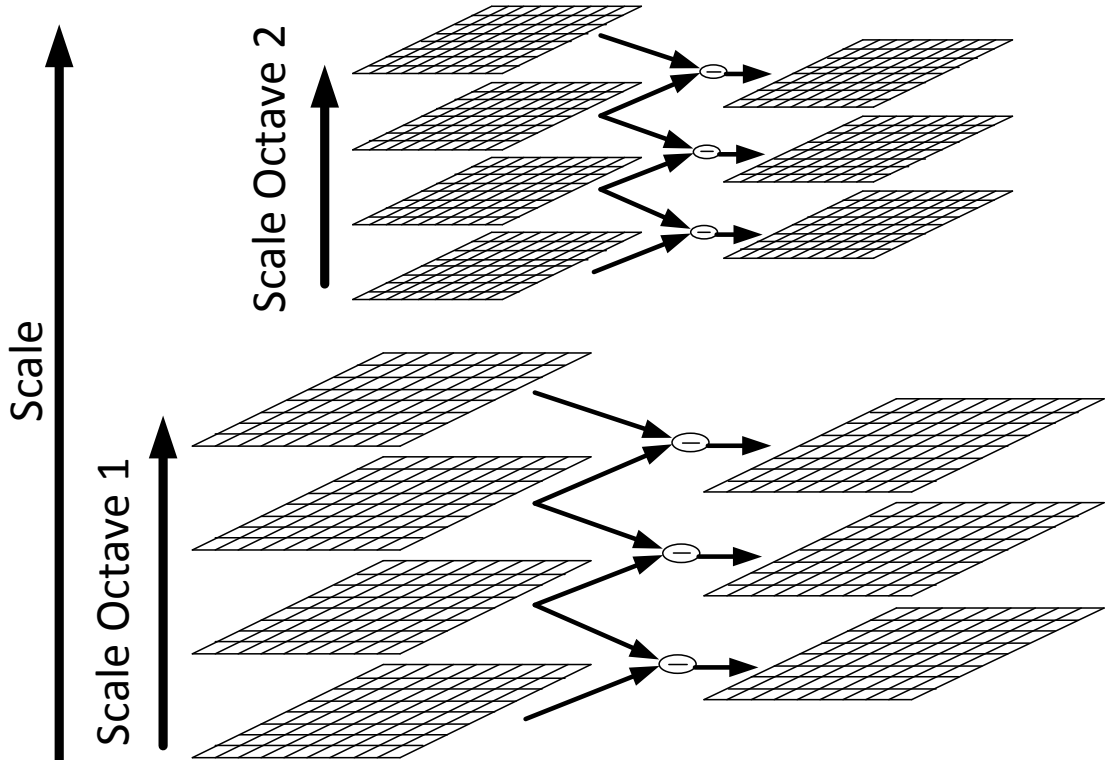
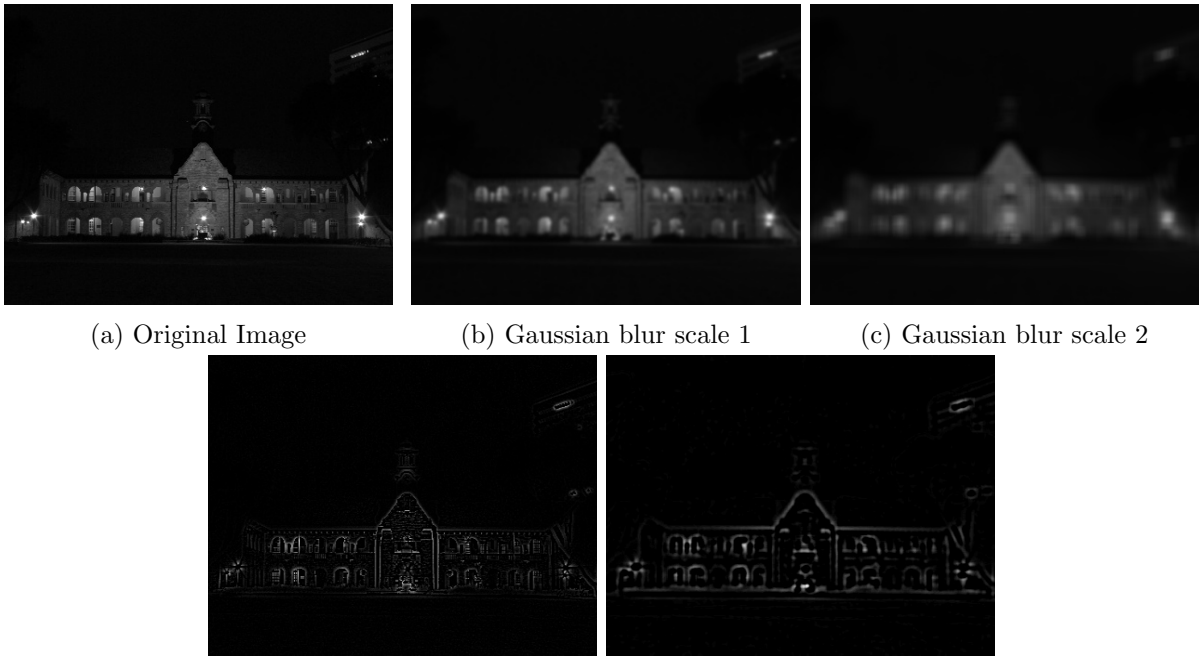


Figure 4.1: Rectification of stereo vision

the magnitude and direction of the image gradient in a window of 4×4, where each window grid point describes the magnitude of the gradient in eight directions. This is done by taking a window of size 16×16 around the pixel. Subregions of 4×4 are created within the larger region creating 4×4 larger subregions around the pixel, each of which has a size of 4×4 pixels. In one of these larger subregions the gradient magnitude and direction of each pixel is determined. The magnitudes are weighted using a Gaussian kernel centred at the the key point. In each subregion a histogram is created of 8 bins specifying a direction (45 *deg* increments). The histograms are filled using the weighted magnitude and direction of the pixel gradients. This generates 4×4 histograms of eight bins producing a final feature vector of 128 bins. This process is shown in Figure 4.3b. The feature vector is normalized to obtain invariance to illumination.

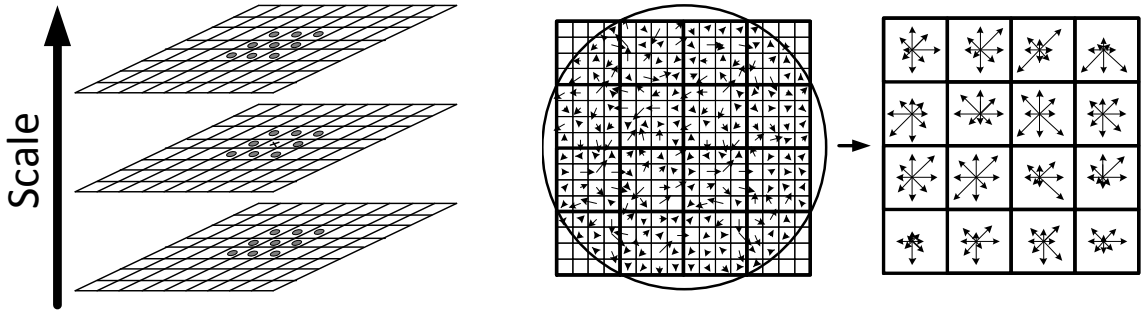
An attempt is then made to match the key points from the two separate images therefore tracking the key points from one image to another. This is performed by using a suitable distance metric between features, usually the Euclidean metric. The corresponding key point to image $\boldsymbol{I_t}$ is obtained by finding the key point in image $\boldsymbol{I_{t+1}}$ which has the shortest distance between feature histograms.



(a) Original Image (b) Gaussian blur scale 1 (c) Gaussian blur scale 2

(d) Difference of Gaussian (DoG) scale 1 (e) Difference of Gaussian (DoG) scale 2

Figure 4.2: SIFT feautre descripor method obtaining difference of Gaussians



(a) Key point detection (b) Features descriptors

Figure 4.3: SIFT key point and feature descriptors

This is performed either using an exhaustive brute force method or a random sampling method. The key point with the smallest distance is taken as the matching key point.

This algorithm, while much more robust than the combination of methods discussed in section 4.1.1, it is also more computationally expensive. Therefore, it is advised if possible to use the method discussed in section 4.1.1 unless this method proves to be insufficient for the task at hand.

4.2 Full Image Correlation

In section 4.1 the DIC techniques used to track a subset of points across images were described. In this section the DIC techniques are applied across the full image. This is useful in a stereographic rig to determine the depth of the whole scene. If the disparity of each pixel in an image can be determined then the 3D coordinates of each pixel can be determined as well. Considering that the techniques in section 4.1 cannot be applied to the full image, the simplest method to match a region is to map a region in one to another using an exhaustive search and suitable correlation metric. The simplest metric is minimising the Sum of Absolute Differences (SAD) given as:

$$SAD = \sum_{i=-N/2}^{N/2} \sum_{i=-N/2}^{N/2} |I(x+i, y+j) - I'(x+i+dx, y+dy+j)| \tag{4.12}$$

where I and I' are the brightness of the left and right image respectively. This metric maps a widow of size $N \times N$ from one image to another. In order to find the corresponding region requires that the window be moved across the whole second image, for each pixel in the first image. The pixel location in the second image with the smallest SAD metric value, corresponds to the matching region from the first region. This is a very computationally expensive process.

To speed up the method, the searching region is reduced using the geometry relating the two images. The epipolar geometry discussed in section 3.5.1 provides this ability. The epipolar geometry allows a feature in one image to be mapped onto a line in the other image, this allows the exhaustive search to be reduced to a much simpler line search. The process is simplified by first re-projecting both images such that the epioplar lines are horizontal in both images and correspond with the same row. Therefore, a feature in row m in the left image is located in the same row m in the right image.

This reduces the complexity of solving for the image disparity. The disparity is computed relative to the left image. The disparity of a pixel in the left image is obtained by starting at the same pixel location on the right image and then moving a window to the left of this position. This is done since in most circumstances the feature in the right image would be found left of the features' corresponding location in the left image, as shown in Figure 4.5. The algorithm makes provision for geometry in which the feature in the right image is located to the right of the position of the left image.

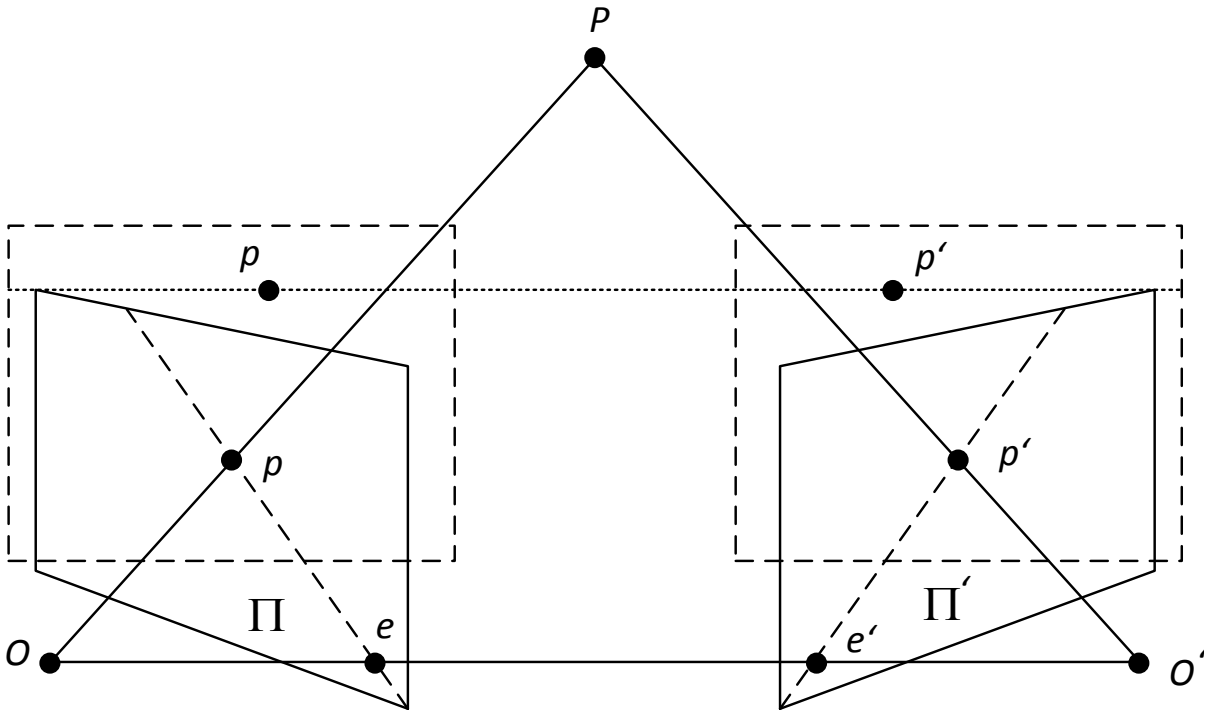


Figure 4.4: Rectification of stereo vision

There are other methods which determine the disparity of an image but the method using the SAD metric, also known as the block matching method [Konolige, 1998], is used in this dtufy for its simplicity and low computational cost. The method employs other heuristic techniques such as pre and post filtering to improve disparity estimation and reduce false matching.

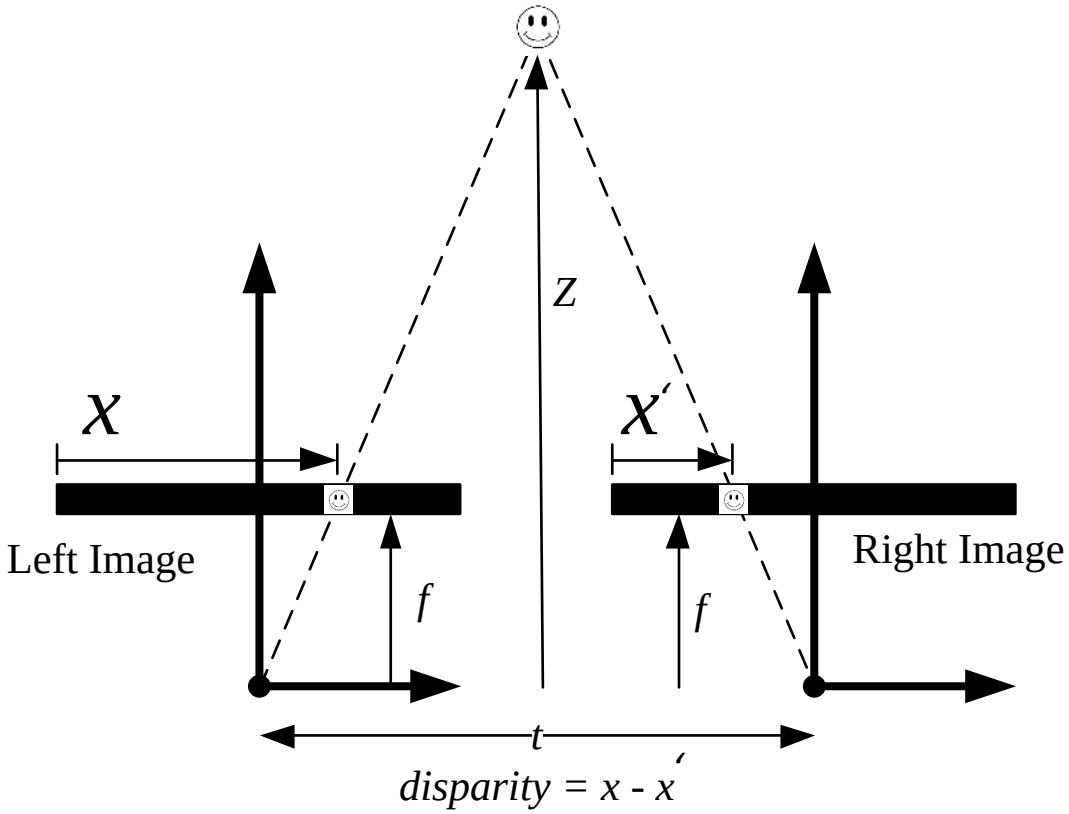


Figure 4.5: Disparity mapping between images

4.2.1 Depth Resolution

The depth resolution is an important consideration in stereographic rigs. It determines how accurate the depth can be determined from the geometry of a stereographic rig. Considering the canonical geometry where the two optical axes of the cameras are parallel as in Figure 4.5, the depth of the feature Z can be derived as:

$$\frac{t-(x-x')}{Z-f} = \frac{t}{Z} \tag{4.13}$$

$$\frac{t-d}{Z-f} = \frac{t}{Z} \tag{4.14}$$

$$Z = \frac{ft}{d} \tag{4.15}$$

It can also be derived from equation 4.15 that the depth resolution ΔZ in terms of the focal length f and translation displacement t (distance between cameras) as well as the minimum measurable disparity Δd is [Grewe and Kak, 1994]:

$$\Delta Z = \frac{-\Delta d Z^2}{tf + \Delta Z} \tag{4.16}$$

It is evident from equation 4.16 that the depth resolution is dependent on the distance away from the camera. This relationship is shown graphically in Figure 4.6. The figure shows that the depth resolution deteriorates exponentially as the object moves further away from the cameras. It is therefore desired to have the object as close as possible to the stereographic rig as practically possible, taking into consideration the focussing and viewing area, in order to improve the depth resolution. Alternatively the depth resolution can be improved by means of increasing either the focal length f or the translational separation t. The implication of increasing the focal length is that the angle of view (i.e. area of view), the depth of field and minimum focus distance decreases. This often leads to the object being placed further away which negates the effect of increasing the focal length. The alternative is to increase the translational separation between the cameras t. The only side effect to increasing the translation separation is that objects close to the camera may not be in view of both cameras.

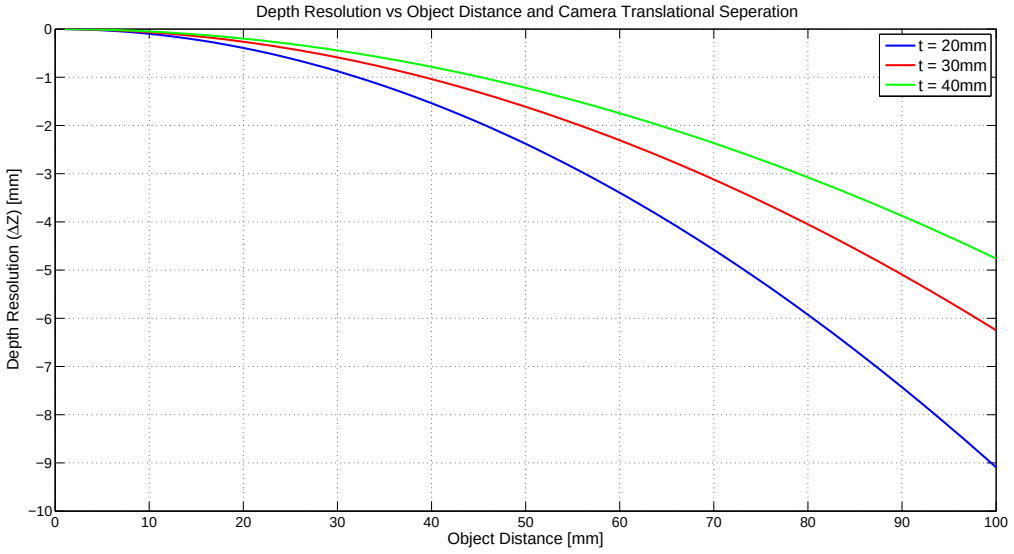


Figure 4.6: Depth resolution as a function of object distance and camera translational separation

A method to circumvent this is to angle one of the cameras inward, this allows the object to be visible in both cameras even when very close. The depth resolution for a camera configuration which is not parallel (also called canonical) but convergent is derived as [Dhond and Aggarwal, 1991]:

$$\Delta Z = \Delta d \cos^2(\phi - \beta)[A - B] \tag{4.17}$$

$$\beta = \tan^{-1}\left[\frac{b - (Zx/f)}{z}\right], A = \frac{Z^2}{tf}, B = \frac{t}{f}\left[1 - \frac{Zx}{tf}\right]^2 \tag{4.18}$$

Equation 4.17 shows that the depth resolution is not only a function of depth Z, focal length f, translational separation t and the pan angle ϕ but also the horizontal location on the image x. Therefore, given a convergent configuration the depth resolution changes across the image and is not consistent as in the canonical configuration. The depth resolution for various pan angles are shown in Figure 4.7. The figure shows the constant depth resolution for a pan angle $\phi = 0deg$, canonical configuration, across the image plane. Whereas the plots for pan angles of $\phi = 10deg$ and $\phi = 20deg$ show how the resolution changes with the horizontal position of the object in the image plane. It should also be noted that all the distances used in this chapter are specified in world units [*mm*], however the image sensor coordinates are specified in *pixels*.

The relationship between these measurement are given by the pixel size S_x introduced in section 3.1. Given that the smallest accuracy of the disparity mapping is one pixel, sub pixel accuracy of $1/16^{th}$ pixel is attainable with block matching, the corresponding minimum measurable disparity d in *mm* is given as $1/S_x$. The pixel size can roughly be determined as the number of horizontal pixels divided by the horizontal length of the image sensor. This means that the higher the number of pixels and the smaller the image sensor the better the depth resolution.

Similar equations can be obtained for the horizontal and vertical resolution of the system, however the depth resolution is the parameter which is often of most concern.

The OpenCV library is used to employ both the motion tracking algorithms as well as the disparity mapping.

4.3 Conclusion

The focus of this chapter was the description of the DIC techniques used in this study, in order to track regions in one image in another image. Two sparse tracking algorithms were described

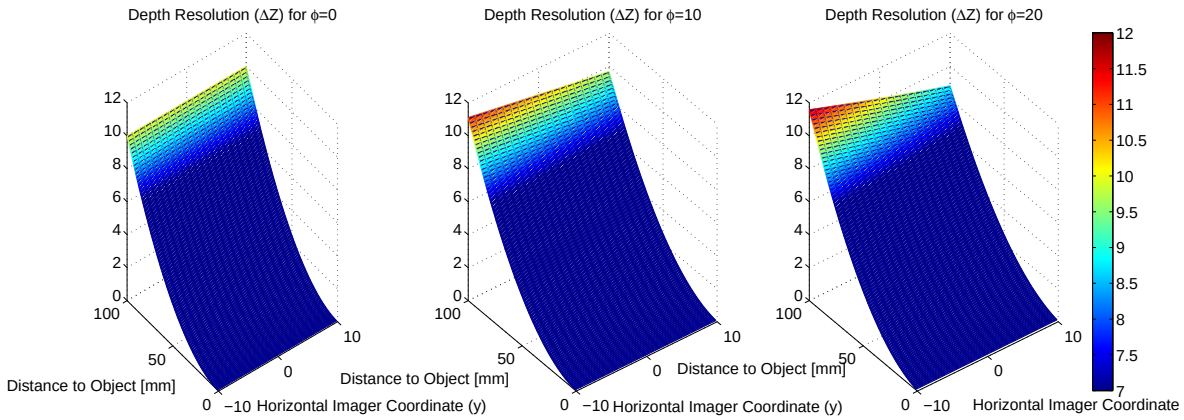


Figure 4.7: Depth resolution as a function of object distance and camera translational separation

which is used to track a smaller number of features in the image. The first algorithm, using corner features and optical flow, provides a fast method which is used to track small feature motions. The second algorithm, the SIFT feature descriptor, is more robust but is computationally more expensive allowing larger displacements to be tracked. A block matching algorithm described how the problem of disparity mapping can be solved in stereography to obtain the 3D coordinates of almost every pixel in a stereographic rig. Equations show how depth resolution of a stereographic rig can be improved to obtain the best resolution from the system. These DIC techniques are applied either separately or in a combination to solve problems in vehicle dynamics in Chapters 5, 6 and 7.

Chapter 5

Side-slip Angle Measurement on Rough Terrain

In this chapter, methods whereby the vehicle side-slip angle can be measured accurately at low speeds on any terrain using inexpensive cameras are presented. Most commercial side-slip angle sensor systems and slip angle estimation techniques rely on smooth terrain and high vehicle speeds, typically above $20 km/h$, to provide accurate measurements. The systems are also limited to small slip angles, below $15 deg$, while very large slip angles can be encountered on slippery surfaces However, during certain in-situ tyre and vehicle testing under severe off-road conditions, the vehicle may be travelling significantly slower than the speeds at which current sensors and estimation techniques can provide sufficiently accurate results. Terramechanics tests are typical case in point. Three algorithms capable of determining the side-slip angle from overlapping images are presented. The first is a simple fast planar method. The second is a more complex algorithm which can extract not only the vehicle slip angle but also its rotational velocities and scaled translational velocities. The third algorithm uses a calibrated stereo rig to obtain all rotational and translational movements in real world coordinates, i.e the 3D motion in *mm* and *deg*. The second and third methods are aimed more at rough terrain applications, where the terrain induces motion components other than typical predominant yaw-plane motion at which side-slip angle is traditionally measured. These systems allow not only for accurate measurement at low speeds but also at higher speeds depending on the camera speed and lighting.

5.1 Introduction

Vehicle side-slip angle is an important measurement for the handling and stability of vehicles. The side-slip angle can describe the over steer and under steer characteristics of a vehicle when measured on the front and rear axle of the vehicle. It can also be a measure of vehicle directional stability. Measurement of the side-slip angle is therefore important when performing high speed dynamic manoeuvres with vehicles.

The tyre side-slip angle is also of great importance when performing tyre parametrization and evaluation. The lateral tyre force generation is a complex phenomenon in which the deformation of the rubber during cornering is non-uniform across the contact patch as illustrated by Fiala's mathematical formulation [Fiala, 1954] of tyre force generation. It can be shown that the side-force generation of a tyre can be sufficiently described using the average side-slip angle of a tyre [Fiala, 1954], [Bakker et al., 1987]. Accurate measurement of the side-slip angle is therefore crucial in validating and characterizing tyre models. In many terramechanics studies the tyre tests are performed in a laboratory with a controlled environment [Sandu et al., 2008] [Krick, 1973]. In these laboratory tests the side-slip angle can be controlled accurately, however, these test rigs are expensive and consist of only a tyre and not a complete vehicle. The tests are also conducted at low speeds over artificially created terrain. Numerous tests are conducted in-situ with vehicles travelling across real terrains. In these situations the side-slip angle needs to be accurately measured. Since these tests are typically conducted at low speeds, the estimation methods are generally not applicable.

Stability and traction control algorithms can also be improved if a cost effective on-line measurement of the side-slip angle is available. The side-slip angle and its derivative offer more practical information about the vehicle stability compared to the yaw rate for vehicle stability [Inagaki et al., 1994]. Chung and Yi [Chung and Yi, 2006] developed a side-slip angle based stability control scheme which shows overall improvement in vehicle performance.

In literature many studies concentrate on estimating the side-slip angle through other vehicle measurements. The side-slip angle is often estimated using sensors such as accelerometers, rate gyroscopes and GPS [Botha and Els, 2012], [Bevly et al., 2006]. The main disadvantages of the estimation methods are that they rely on high dynamic situations such as high speed manoeuvres where sensor excitations are large in comparison to sensor noise. Additionally, when considering off-road terrain the additional ground excitation may become too large rendering these methods

unsuitable. While in off-road driving scenarios especially on deformable or very low friction terrains, such as mud, sand, loose gravel, ice and snow, the side-slip angle may become large when fully sliding and can also occur at much slower speeds as compared to hard terrains.

One of the most popular and common commercial side-slip angle measurement sensor is the Kistler Correvit S-HR [Kistler, 2014]. This sensor uses both Doppler Effect and an absolute measuring method. However, the sensor generally provides low accuracy below speeds of $15 km/h$ and has a maximum side-slip angle range of $\pm 20 deg$. The sensor is mainly developed for use on on-road vehicles where there is very little vertical body motion and the vehicle mainly undergoes planar motion, i.e. high speed driving on smooth roads. It will however be shown that vertical motion, as well as roll, has a negative effect on the accuracy of the side-slip angle measured with the sensor.

In this chapter three algorithms are described, whereby the side-slip angle can be measured using either a single camera facing the ground or a calibrated stereographic rig containing two cameras. The three algorithms make use of the DIC techniques described in Chapter 4 as well as other imaging techniques, to determine the side-slip angle. The first method proposed is a simplistic technique which is very computationally inexpensive but only measures longitudinal and lateral velocities from which the slip angle can be calculated. The second method is more computationally expensive but can measure all rotational and scaled translational velocities. The third technique makes use of a calibrated stereographic rig to determine all rotations and translations in real world units. All methods can measure any degree of slip angle and can accurately measure slip angles up to very low speeds.

5.2 Slip Angle Measurement Using Efficient 2D Planar Method

The DIC technique described in section 4.1.1 is used to track features in successive images in the image sequence $[\boldsymbol{I_t}, \boldsymbol{I_{t+1}}, ..., \boldsymbol{I_n}]$. The image sequence is obtained using a calibrated and undistorted single camera mounted perpendicular to the road. The camera captures images at a frame rate which guarantees that a large portion of the scene overlaps in successive images. An analysis of the maximum permissible vehicle speed for which this is possible is given in section 5.6.

After features have been tracked between two frames, these tracked features can be used to estimate the relative motion between the frames. This can be achieved provided the features are stationary and that any lens distortion is removed.

The tracked features describe the re-projected relative motion of the camera over the road. Since

the lateral and longitudinal displacement of the features are measured simultaneously, the angle of the feature's motion can be considered as the camera direction of travel. Thus, if the camera is mounted such that the features move horizontally, or laterally, when the side-slip angle is zero, then the angle the features move in is effectively the side-slip angle. Offset correction can be performed by measuring the slip angle with the cameras when the slip angle should be zero i.e. driving in a straight line on a level flat road. The assumption that the feature's angle of motion represents the slip angle holds true if the feature's motions contain purely lateral and longitudinal motion, thus, all features will travel in the same direction. On smooth roads this assumption is generally valid as the vehicle undergoes mostly planar motion. Figure 5.1 depicts the tracking of features from one frame to another over a level concrete surface with the line trace showing the relative motion between the previous frame and the current frame. From this image it can be observed that most features seem to be travelling in the same direction with the same magnitude, however at least one tracked feature can be observed to be tracked incorrectly. The incorrect tracking of features presents a problem to the measuring of the direction of travel. While Figure 5.1 shows the majority of the features moving in the same direction, it could occur that numerous features are tracked incorrectly causing an error in the measurement. It is therefore necessary to employ a form of outlier rejection to remove any incorrectly tracked features. The first phase of outlier rejection is provided by the Lucas-Kanade algorithm (see section 4.1.1) implemented in OpenCV. The rejection is based on the difference between windows, of which the feature is at the centre, on the first image and the tracked feature in the second image. If this error is larger than a certain prescribed threshold the feature is rejected. While this simple method does remove some outliers it fails to remove all outliers, as shown with one feature in Figure 5.1, therefore another outlier rejection namely the RANdom SAmple Consesus (RANSAC) algorithm [Fischler and Bolles, 1981] is employed as well.

5.2.1 RANSAC

The RANSAC algorithm is an iterative procedure to estimate the parameters of a model from a set of data which is corrupted by outliers. In this simple case the model is simply the direction of travel. The synopsis of the method is to randomly sample a minimum set of data points from which to develop the model, in this case only one sample is necessary to estimate the direction of travel. The model is derived from this minimal set and each remaining data point is tested against the model

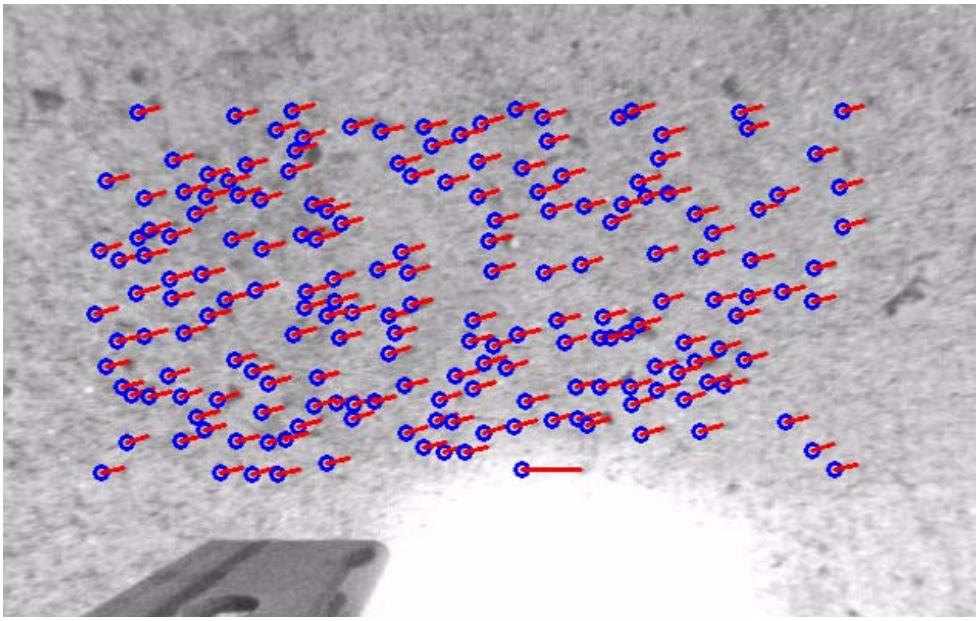


Figure 5.1: Tracking of corner features with the trace showing the direction of motion

using a suitable metric. In this case the difference in angle. If the error of the metric is below a certain threshold, the data point is considered to agree with the model and counted as an inlier. The process of sampling a minimum set of data points and developing a model continues for a suitable number of iterations. Upon completion the model with the most inliers is returned as the correct model. In this case rather than use the best model as the correct model, the inliers of the best model is used to average the model over multiple samples. This algorithm thus allows the model to be based on the average value of inliers only. The result of the outlier rejection process is shown in Figure 5.2 with certain tracked features classified as outliers, shown in green. The RANSAC algorithm is also used to determine the average velocity magnitude between the two images $\boldsymbol{I_t}$ and $\boldsymbol{I_{t+1}}$. If the velocity magnitude is smaller than a threshold then the image $\boldsymbol{I_{t+1}}$ is iterated to image $\boldsymbol{I_{t+2}}$. The velocity magnitude is recalculated and check is performed again unitl the magnitude is above the threshold. This scheme creates a variable frequency sampling method which provides better results at low speeds. A flow chart of the complete method is shown in Figure 5.3.

The simple 2D planar algorithm, which uses only one camera, is computationally inexpensive and can easily be implemented on an embedded system to create a system that can output an analogue slip angle value in real time. The embedded system need not require much to be capable of providing real time outputs at $500Hz$.

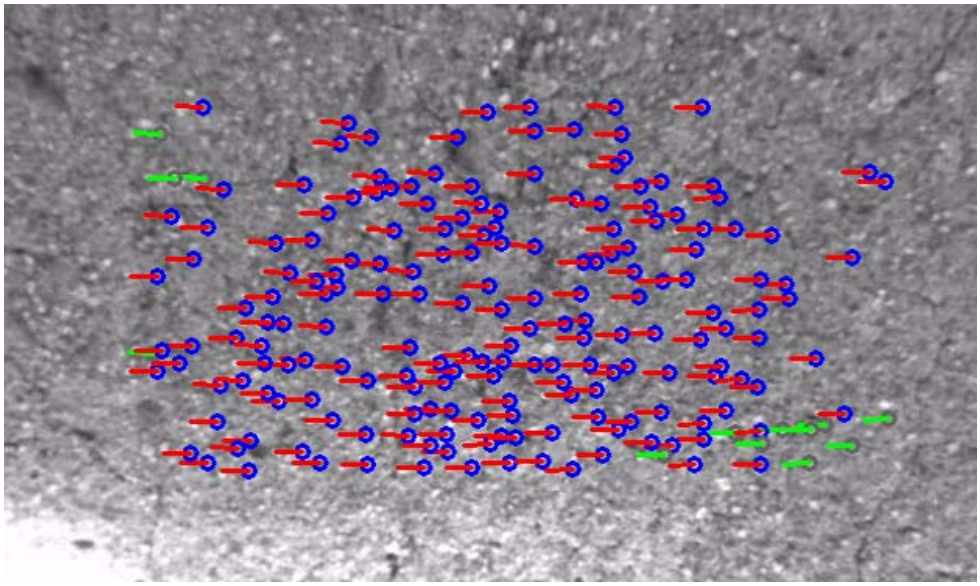


Figure 5.2: Feature tracking and outlier rejection (red particles are inliers and green are outliers) for the planar method

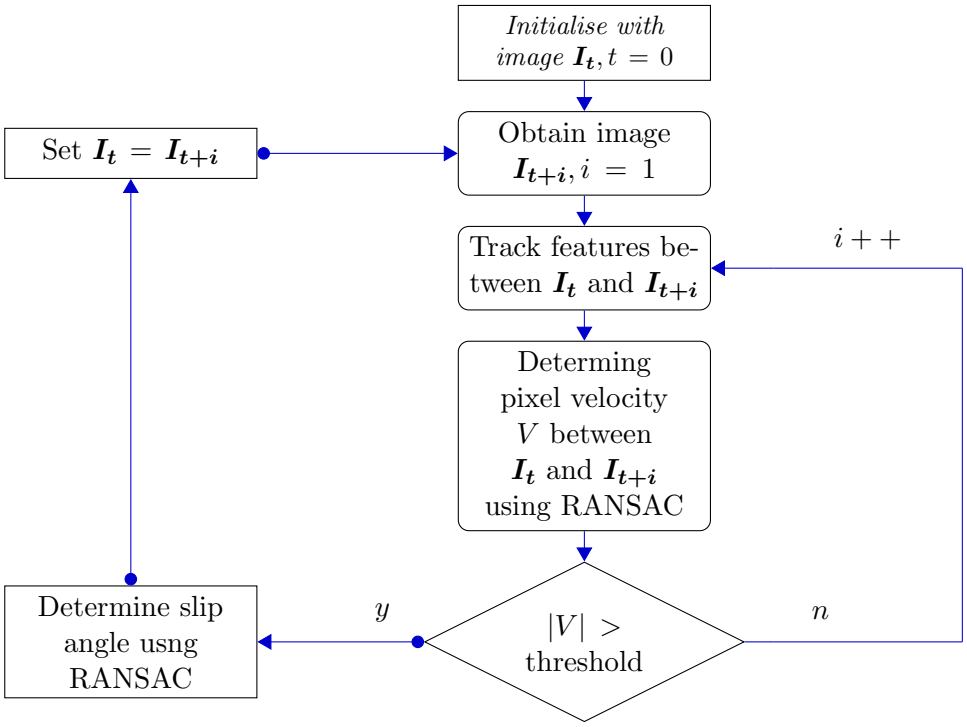


Figure 5.3: Flow chart of 2D planar method

5.3 Slip Angle Measurement Using Pose Estimation from Image Correspondence

The algorithm described in section 5.2 is very basic and can easily be implemented in real time. However, the largest problem with this model is that it assumes only planar motion. While this assumption is valid under traditional paved road conditions, the assumption breaks down when considering off-road terrain, which may induce vertical motion as well as rotational motions. The effect of these motions on the feature's motion which are re-projected onto the camera image is illustrated in Figure 5.4. This figure shows that motions other than planar may create an ambiguity in measuring the lateral and longitudinal motion. The figure shows the induced motion if it is assumed that the camera centre is the centre of rotation, this may not necessarily be the case especially if the camera is not mounted close to the Centre of Gravity (CG) of the vehicle. If the camera is mounted away from the CG certain motions may have a lesser effect, such as the yaw rate. However, mounting the sensor further away from the CG will result in the other rotations, pitch and roll, inducing longitudinal, lateral and vertical motion of the sensor. It may be very difficult to differentiate between the translational motions of the sensor, which are induced by the rotational motions, and the overall lateral vehicle motion. It is therefore recommended to mount the sensor close to the CG to reduce these effects when traversing over rough roads. This however still requires the compensation of the other motions in order to determine the direction of travel in the horizontal plane.

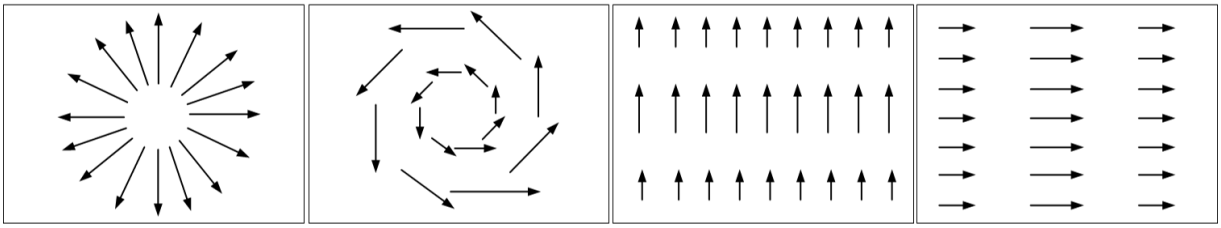


Figure 5.4: Projected motion onto camera sensor for (from left to right): a) vertical motion b) yaw motion c) roll motion d) pitch motion of a planar object

To take all of these motions into account requires the implementation of pose estimation algorithms using image correspondences. The problem of estimating relative pose from multiple viewpoints of scene features has been studied for a long time, initially for photogrammetric purposes and more recently for computer vision. The estimation of the pose is the same problem as in the

stereographic rig where it is required to know the relative position and orientation between the two cameras. In this case the difference lies in the fact that the two views of the scene are captured with one camera in motion. Also in the stereographic rig a known calibration surface is used which is impossible to use in the single camera case. The motion is also not constant which requires that the calibration procedure be performed at each new frame. Instead another approach is used to estimate the epipolar geometry between the two views which captures the relative motion between the two views.

From a mathematical viewpoint the Fundamental matrix $\boldsymbol{F}$ describes the relative projection between two un-calibrated viewpoints.

$$\boldsymbol{F} = \boldsymbol{K}_2^{-1T}[\boldsymbol{t}]_{\boldsymbol{x}}\boldsymbol{R}\boldsymbol{K}_1^{-1} \tag{5.1}$$

Where $\boldsymbol{F}$ is the 3×3 fundamental matrix, $\boldsymbol{K}_1$ and $\boldsymbol{K}_2$ are the calibration matrix from viewpoint 1 and 2 respectively, $\boldsymbol{R}$ is the relative orientation between the two views and $[\boldsymbol{t}]_{\boldsymbol{x}}$ is the skew symmetric matrix containing the relative translations:

$$[\boldsymbol{t}]_{\boldsymbol{x}} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \tag{5.2}$$

The fundamental matrix also encodes the epipolar constraints between the projections of a real world coordinate onto the two camera image coordinates. The epipolar constraint describes the relation of these images to one another viewing the same scene. The constraint is described as:

$$\boldsymbol{p}'^{T}\boldsymbol{F}\boldsymbol{p} = 0 \tag{5.3}$$

Where $\boldsymbol{p}$ and $\boldsymbol{p}'$ are point correspondences described in homogeneous coordinates of an object in the first and second images respectively. The homogeneous coordinates are a 3 vector coordinate obtained from the image as $\boldsymbol{p} = \begin{bmatrix} x & y & 1 \end{bmatrix}^T$, where x and y are the image coordinates. In most cases the calibration matrices of the cameras are performed beforehand using a known planar surface (see section 3.4). Since the calibration matrix is constant for one camera over multiple views $\boldsymbol{K}_1 = \boldsymbol{K}_2$. The epipolar constraint of equation 5.3 can be simplified to:

$$\boldsymbol{q}'^{T}\boldsymbol{E}\boldsymbol{q} = 0 \tag{5.4}$$

Where $\boldsymbol{E}$ is known as the essential matrix and maps the relative translational and rotational motion between the views as:

$$\boldsymbol{E} = [\boldsymbol{t}]_x \boldsymbol{R} \tag{5.5}$$

The homogeneous image coordinates $\boldsymbol{q}$ and $\boldsymbol{q'}$ are now presented in terms of the normalized coordinates

$$\boldsymbol{q} = \boldsymbol{K}_1^{-1}\boldsymbol{p} \text{ and } \boldsymbol{q}' = \boldsymbol{K}_2^{-1}\boldsymbol{p}' \tag{5.6}$$

The remainder of this section focuses on the estimation of the essential matrix from point correspondences. These point correspondences are the same features which are tracked between the two views in the planar method. The essential matrix is estimated using multiple correspondences from the two viewpoints. The first method required 8 point correspondences and is therefore termed the 8-point linear algorithm introduced by Longuet-Higgins [Longuet-Higgins, 1987]. While the 8-point algorithm is simplistic in nature it is unfortunately very sensitive to noise in the correspondences. Consequently many alternative algorithms have been developed to estimate the essential matrix and while these algorithms are exceedingly more complex they generally provide superior robustness in the presence of noise [Rodehorst et al., 2008]. These algorithms also require less correspondences, however, generally produce more than one candidate solution to the essential matrix. Especially of interest is the 5-point algorithm developed by Stewenius [Stewenius et al., 2006]. This algorithm is not only robust in the presence of noise, it has also been observed that it is stable if the correspondences are obtained from coplanar objects. Since the terrain which the vehicle is driven across is most likely coplanar or near coplanar, especially when the road is flat, the 5-point algorithm is chosen in this study for estimating the essential matrix. The first step in the algorithm is to use the epipolar constraint 5.4, the constraint is employed as follows:

$$\boldsymbol{q}'^T \boldsymbol{E} \boldsymbol{q} = 0 \tag{5.7}$$

$$\tilde{\boldsymbol{q}}^T \tilde{\boldsymbol{E}} = 0 \tag{5.8}$$

where

$$\tilde{\boldsymbol{q}} = \begin{bmatrix} q'_1 & q_2 q'_1 & q_3 q'_1 & q_1 q'_2 & q_2 q'_2 & q_3 q'_2 & q_1 q'_3 & q_2 q'_3 & q_3 q'_3 \end{bmatrix}^T \tag{5.9}$$

$$\tilde{\boldsymbol{E}} = \begin{bmatrix} E_{12} & E_{13} & E_{21} & E_{22} & E_{23} & E_{31} & E_{32} & E_{33} \end{bmatrix}^T \tag{5.10}$$

The vector $\tilde{\boldsymbol{q}}$ is shown for one point correspondence. By using five correspondences a 5×9 matrix, build-up of different $\tilde{\boldsymbol{q}}$ vectors, is obtained. The Singular Value Decomposition (SVD) algorithm can then be used to obtain the 4 null vectors of the matrix [Khare et al., 2010]. The 4 vectors of size 1×9 can then be written in a 3×3 matrix form, obtaining matrices E_1, E_2, E_3, E_4. The essential matrix is then given by:

$$\boldsymbol{E} = x\boldsymbol{E}_1 + y\boldsymbol{E}_2 + z\boldsymbol{E}_3 + w\boldsymbol{E}_4 \tag{5.11}$$

where x, y, z, w are scalar values. It is generally assumed that $w = 1$ since the scalar values are only defined up to a common scale factor. This then requires the solution of the remaining scalar values using the trace and determinate constraints. The solution is obtained from solving a multivariate polynomial equation, generally obtained using the Grobner basis method. The method is quite arduous and thus the detail of the solution is omitted here, for detailed descriptions see Stewenius et al [Stewenius et al., 2006] or the alternative method by Nistér [Nistér, 2004].

The 5-point algorithm can return up to 10 real solutions all of which are candidate solutions to the essential matrix. To obtain the best essential matrix the RANSAC algorithm is employed. A random subset of the minimum number of correspondences is drawn and the 5-point algorithm is used to solve for up to 10 plausible solutions. Each solution is then checked for inliers among the remaining correspondences. The metric used to determine whether a sample is an inlier is the first order geometric distance (Sampson-distance) given as:

$$d_{sampson} = \frac{\boldsymbol{q'^T E q})^2}{(\boldsymbol{Eq})_1^2 + (\boldsymbol{Eq})_2^2 + (\boldsymbol{E}^T\boldsymbol{q'})_1^2 + (\boldsymbol{E}^T\boldsymbol{q'})_2^2} \tag{5.12}$$

where $(\boldsymbol{Eq})_i$ is the ith entry in the vector $\boldsymbol{Eq}$. Another test conducted to determine if the essential matrix is indeed the true essential matrix is that of the cheirality test. This test first requires extracting the relative orientation and translation between the frames from the essential matrix. This is accomplished from the thereom given by Tsai and Huang [Tsai and Huang, 1984].

The solution is obtained by using the singular value decomposition on the essential matrix. It should be noted that four possible solutions are obtained from this procedure:

$$\boldsymbol{E} = \boldsymbol{U} diag(1,1,0)\boldsymbol{V}^T \tag{5.13}$$

$$\boldsymbol{D} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.14}$$

$$\boldsymbol{R}_a = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T \tag{5.15}$$

$$\boldsymbol{R}_b = \boldsymbol{U}\boldsymbol{D}^T\boldsymbol{V}^T \tag{5.16}$$

$$\boldsymbol{t}_a = \begin{bmatrix} u_{13} & u_{23} & u_{33} \end{bmatrix}^T \tag{5.17}$$

$$\boldsymbol{t}_b = -\boldsymbol{t}_a \tag{5.18}$$

Where matrices $\boldsymbol{U}$ and $\boldsymbol{V}^T$ are the matrices obtained from the singular value decomposition. Two possible rotation matrices $\boldsymbol{R}_a$ and $\boldsymbol{R}_b$ are obtained along with the two possible translation matrices $\boldsymbol{t}_a$ and $\boldsymbol{t}_b$. Any combination of these rotation and translation matrices satisfies the epipolar constraints and is thus a possible solution, yielding four possible solutions of which only one is correct. One of the other solutions corresponds to the twisted pair which is obtained by rotating one of the views by 180*deg*. The remaining two solutions correspond to reflections of the true configuration and its twisted pair.

To remove the ambiguity of the motion solutions the cheirality of all four possible solution are tested. This is accomplished by triangulating [Hartley and Sturm, 1997] the five correspondences used to obtain the essential matrix from which the rotation and translational motion is obtained. The triangulation performs a back projection from the two camera images to obtain the scaled world coordinates of each correspondence. The world coordinate is then tested to see whether it lies in front of both cameras or behind any of the cameras. The true configuration should require all five correspondences to be in front of both cameras. Therefore, the solution which contains all five point correspondences in front of the camera is selected as the true motion.

Rodenhorst [Rodehorst et al., 2008] advises that, for efficiency considerations the cheirality and Sampson-distance inlier test not be conducted on each possible essential matrix. Thus, the cheirality of each real solution from equation 5.11 is tested first. Only candidate essential matrices which have passed the cheirality test, i.e. those with all five correspondences in front of both views, are passed on

for further analysis where the number of inliers are tested. This methodology, provided that enough RANSAC iterations and enough correct correspondences are present, should result in the correct solution to the essential matrix. From the essential matrix the relative rotation and translation can be extracted from equations 5.13 - 5.18.

To improve accuracy and allow for triangulation requires that the camera has undergone some translational motion between the two views under consideration. In order to guarantee this, the magnitude of translation, in pixels, is calculated directly from the point correspondences using a RANSAC approach. If the translation between frames is less than a certain threshold, no computation is performed and the frame $\boldsymbol{I_{t+1}}$ is iterated to $\boldsymbol{I_{t+2}}$, without advancing frame $\boldsymbol{I_t}$. This effectively means that the side-slip angle is computed between frames $\boldsymbol{I_t}$ and $\boldsymbol{I_{t+2}}$, producing a variable frequency sampling strategy to ensure that the motion remains above a certain threshold. This ensures that enough motion is present before estimating the essential matrix.

The outlier rejection of the pose method is shown in Figure 5.5, the figure shows that the system classifies more correspondences as outliers compared to the planar method. This is partly due to the enforcement of a stricter inlier threshold to improve the final model.

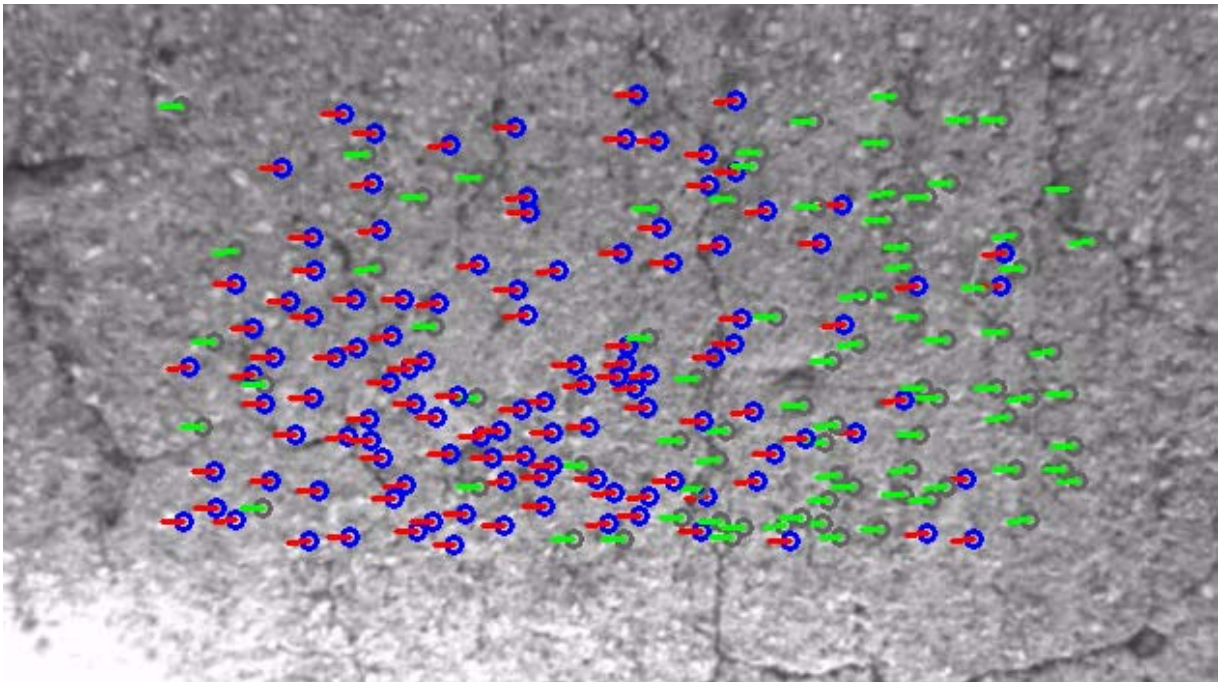


Figure 5.5: Feature tracking and outlier rejection (red particles are inliers and green are outliers) for the pose method

This algorithm thus not only provides the slip angle which can be extracted from the longitudinal and lateral velocities, but also the vertical velocity as well as all three rotation velocities from a single camera. These additional measurements can be used to evaluate the calculated slip angle to ensure that the measured slip angle is a result of the vehicle motion rather than the sensor motion. The algorithm however requires that a computationally expensive algorithm be performed numerous times per measurement. This makes the methodology difficult to implement in real-time. A flow diagram of the 2D Pose method is shown in Figure 5.6.

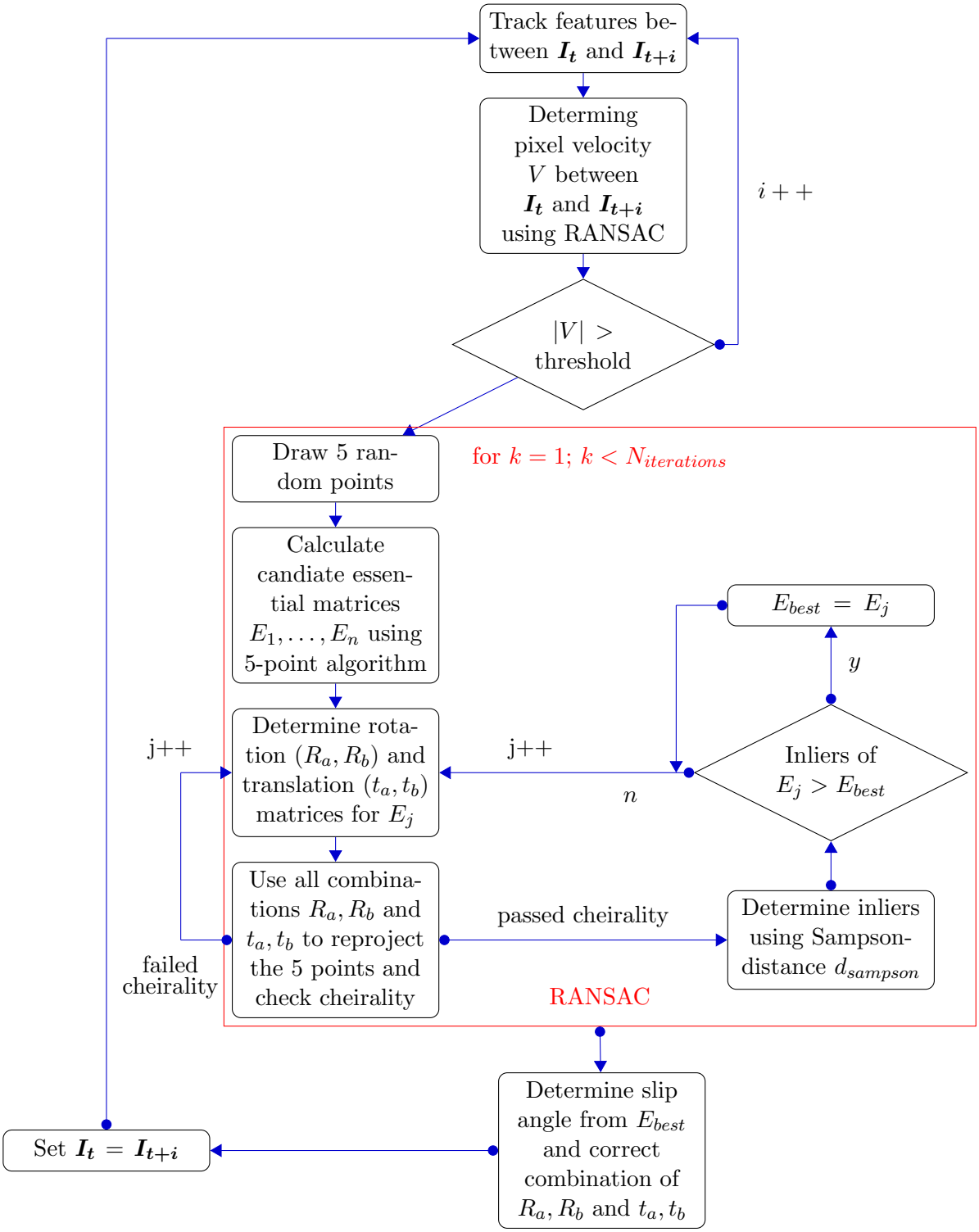


Figure 5.6: Flow chart of 2D Pose method

5.4 Slip Angle Measurement Using Stereograhical System

The pose based system described above has the advantage that it can obtain all translational motions in scale as well as all rotational motions. However, the algorithm is both computationally expensive and the translation motion is obtained also in scale. The scale is dependent on the relative motion and is thus not consistent. Therefore, another method is proposed which obtains all measurement in real world unscaled coordinates. The algorithm is also capable of real time performance with special hardware. The method uses two cameras to view the same scene at the same time. It therefore allows for 3D measurements of the scene.

5.4.1 Slip From 3D Points

Once a 3D map has been generated from the two cameras, features can be tracked in the left image using the same method as the planar system. Since the disparity and therefore 3D coordinates are performed relative to the left image the 2D pixel coordinates from the tracking process can easily be transformed into 3D real world coordinates. This yields several 3D points in space which have undergone rigid motion between two views of which the position is known in both views. The relative translation and rotation can then be determined provided at least three point correspondences have been obtained [Arun et al., 1987]. The problem being solved is the rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$ in:

$$\boldsymbol{P}_2 = \boldsymbol{R}\boldsymbol{P}_1 + \boldsymbol{t} \tag{5.19}$$

Where the N 3D point sets $\boldsymbol{P}_1$ undergoes rigid motion to yield the points set $\boldsymbol{P}_2$. The method starts of by finding the centres of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$:

$$\bar{\boldsymbol{P}}_1 = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{P}_1^i \tag{5.20}$$

where $\boldsymbol{P}_1^i$ is the ith 3D point correspondence. Next the covariance matrix between points set $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are determined as:

$$\boldsymbol{\Sigma} = \sum_{i=1}^{N} (\boldsymbol{P}_1^i - \bar{\boldsymbol{P}}_1)(\boldsymbol{P}_2^i - \bar{\boldsymbol{P}}_2)^T \tag{5.21}$$

The rotation matrix $\boldsymbol{R}$ is obtained from the Singular Value Decomposition (SVD) of the covariance matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \tag{5.22}$$

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{U}^T \tag{5.23}$$

If the determinant of $\boldsymbol{R}$ is negative the solution corresponds to an inflection, the rotation matrix is then recalculated as:

$$\boldsymbol{R} = \boldsymbol{V}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\boldsymbol{U}^T \tag{5.24}$$

The translation vector is obtained from:

$$\boldsymbol{t} = \bar{\boldsymbol{P}}_2 - \boldsymbol{R}\bar{\boldsymbol{P}}_1 \tag{5.25}$$

While this method requires at least 3 point correspondences, the method will calculate the rotation and translation in a least mean square sense for any point sets with N correspondences. Thus, several point correspondences ($\approx$ 50 points) are used to obtain the average motion between frames. However, to prevent any outliers from influencing the calculation of the rotation and translation the RANSAC algorithm is used to remove outliers. The model is derived from 3 randomly drawn point correspondences. The remaining data set is tested against the derived model for inliers using the following error metric, $\epsilon = \left|\left|\boldsymbol{P}_2^i - (\boldsymbol{R}\boldsymbol{P}_1^i + \boldsymbol{t})\right|\right|$. Upon completion the model with the most inliers is returned along with the set of inliers. The rotation and translation is then determined in a least squares sense using all of the inliers of the best model. Figure 5.7 shows the tracked features on a colour depth map of the surface, as well as the corresponding original left camera image. The algorithm also incorporates a variable frequency sampling algorithm to improve the accuracy at low speeds. If the translation between two frames is below a threshold, image $\boldsymbol{I_{t+1}}$ is incremented to image $\boldsymbol{I_{t+2}}$ while image $\boldsymbol{I_t}$ remains fixed.

This yields a system whereby all translations and rotations can be obtained unscaled in real world coordinates using a two camera system. The most computationally expensive part of the

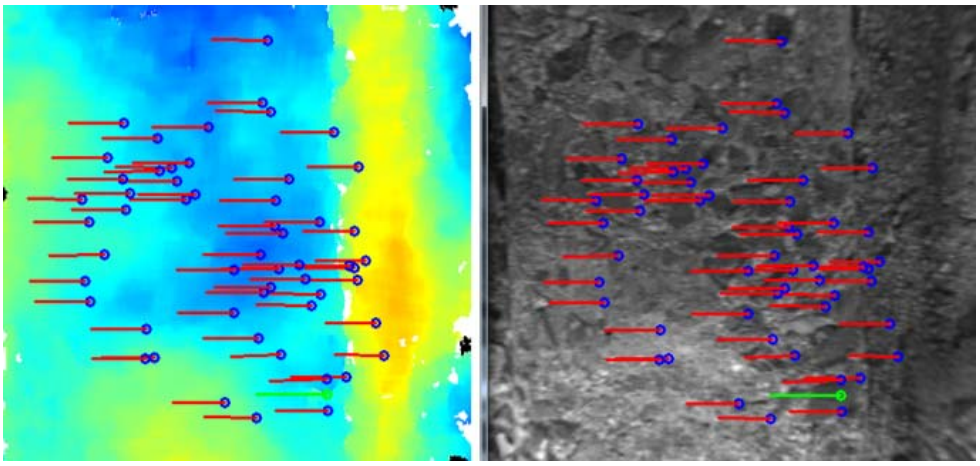


Figure 5.7: Feature tracking and outlier rejection (red particles are inliers and green are outliers) for the stereographic rig in the depth colour (left) map and right on the original image (right)

algorithm is the disparity mapping. However, using Field Programmable Gate Array (FPGA) has yielded systems capable of performing the disparity mapping using more advanced algorithms at over 500*fps* [Georgoulas and Andreadis, 2009]. Thus, this system is capable of yielding a system capable of real time performance at over 500*fps*.

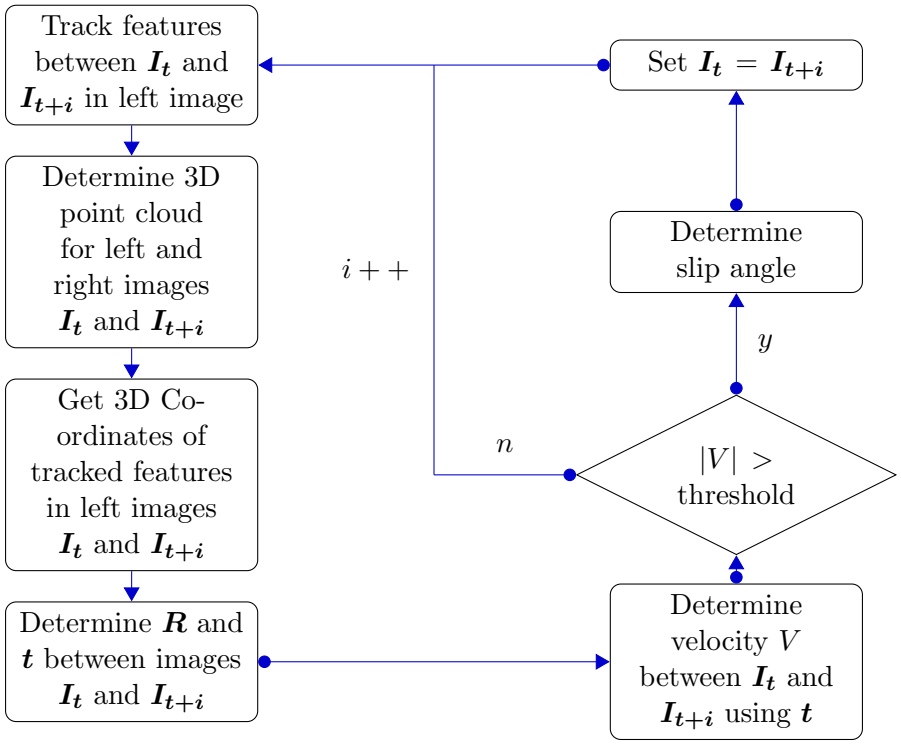


Figure 5.8: Flow chart of 3D method

5.5 Experimental Test

In order to validate the results of the algorithms, tests were performed on both smooth roads as well as off-road terrain. The setup consisted of a calibrated stereographic camera rig. The cameras are Point Grey Flea3 USB 3.0 [Point Grey, 2014] cameras with maximum resolution of 1280×1024 and a global shutter, a detailed specification is given in Appendix A. The lens is a Kowa 4*mm* focal length c-mount lens [Kowa, 2014], the details of the lens are attached in Appendix A. The lens is a wide angle lens providing a large image area even when mounted close to the ground. The cameras are set to resolution 640×480 at which a maximum of 600*fps* can be obtained per camera. The cameras are mounted next to a Corrsys-Datron Correvit S-HR slip angle sensor viewing the same region of road and sampled at 250*fps*. The sensors were synchronised using a synchronising pulse, generated by the data acquisition device.

To validate the measurements both sensors were mounted on a gimbal which allowed the sensors to be rotated in one degree increments relative to the vehicle as shown in Figure 5.9. This allows for the side-slip angle to be set at a constant angle while the vehicle moves in a straight line. The gimbal system is mounted underneath the vehicle as close as possible to the CG of the vehicle to reduce the effect of the roll and pitch motions. The slip angle was set to 0, 2, 5 and 10 *deg* whilst driving over both a flat surface and a rough Belgian paving surface at Gerotek Test Facilities (Gerotek Test Facilities, 2014).

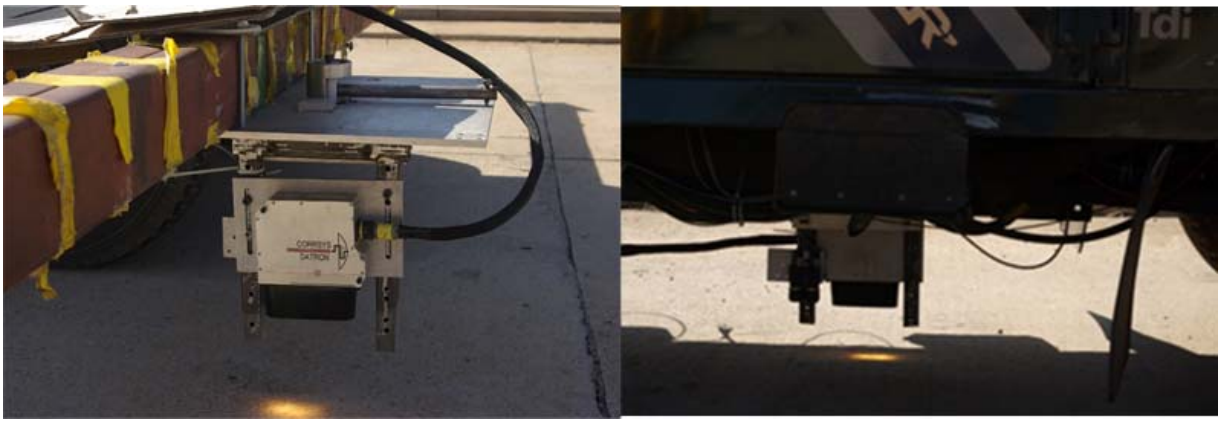


Figure 5.9: Test setup showing two cameras and Correvit S-HR slip angle sensor

The vehicle speed and side-slip angle from all sensors and algorithms driving over a flat surface with the gimbal set at a side-slip angle of $0deg$ is shown in Figure 5.10. Each algorithm is individually compared to the Correvit S-HR slip angle sensor. It can be noted that the Correvit S-HR sensor has a larger noise band than the other measurements especially at lower speeds. The Correvit S-HR only provides decent accuracy at speeds above 10-15km/h, even then the noise band is not consistent. At best the noise levels are comparable to the measurements obtained from the camera based measurements and at worst almost an order of magnitude higher. The camera based systems have better noise bands at lower speeds with the 3D system having the smallest noise band at very low speeds. All measurements also increase in accuracy as vehicle speeds increase. The vehicle speed obtained from the 3D method also compares exceptionally well to that of the velocity of the Correvit S-HR sensor. The vehicle speed can also be obtained from the 2D planar algorithms however the measurement is scaled, therefore if the vehicle height remains constant the scale will be the same. However, since the height of the sensor may change, the measurement of the method is not reliable. The vehicle speed obtained from the 3D algorithm is obtained using real world coordinates and is therefore independent of sensor height. The noise band of the velocity from the 3D algorithm is also smaller compared to the Correvit S-HR sensor. No noticeable difference can be ascertained between the different camera based measurements. Figure 5.11 shows some of the other motions measured by the pose and 3D algorithms. These measurements show no discernible motions as expected as the vehicle is travelling over a smooth surface.

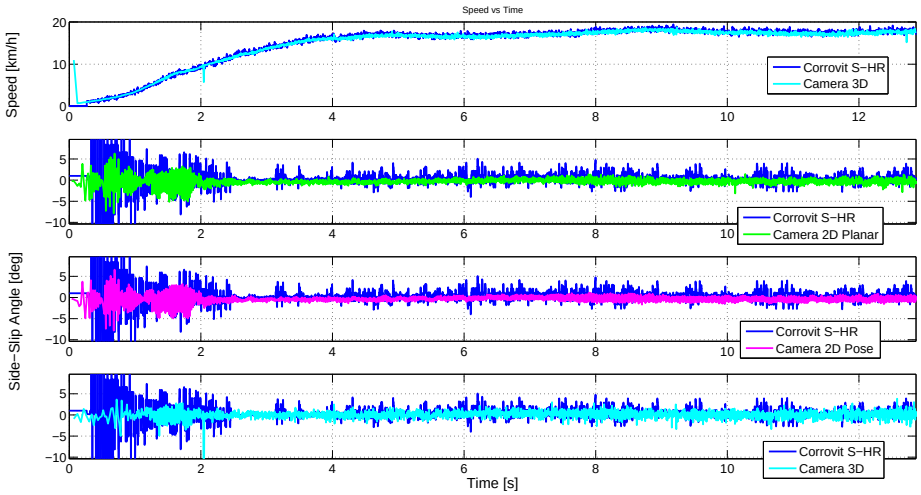


Figure 5.10: Vehicle speed and side-slip angle comparison between camera based techniques and Correvit S-HR side-slip angle sensor on flat surface for zero degrees slip angle.

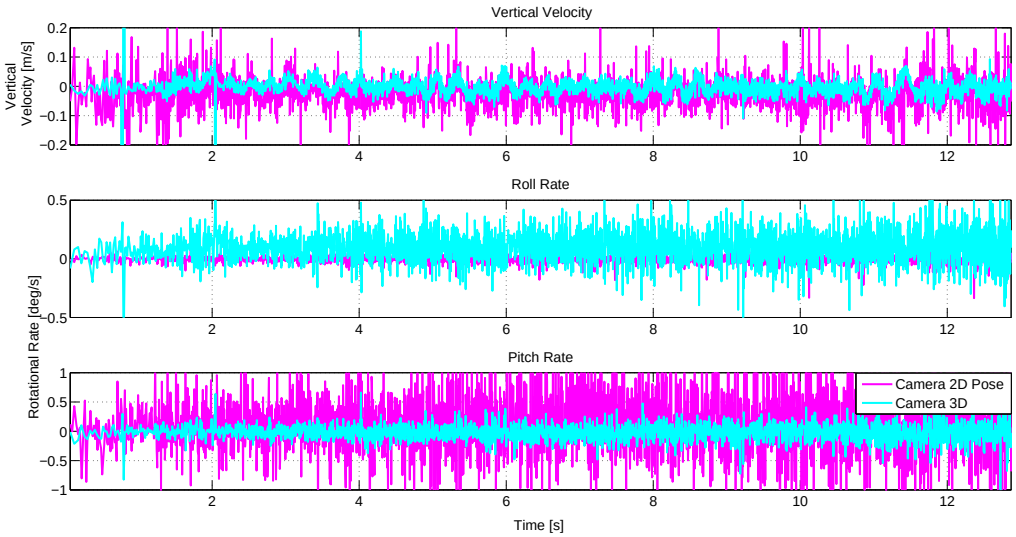


Figure 5.11: Vertical velocity, roll and pitch rate from 2D pose and 3D camera based methods, on flat surface for zero slip angle.

Figure 5.12 depicts the speed and side-slip angle from all sensors and algorithms driving over a Belgian paving surface with the side-slip angle set at $10deg$. The Belgian paving measurement provide similar results to that observed on the flat surface. These tendencies include a smaller noise band at lower speeds as well as a smaller noise band overall for the camera based systems as compared to the Correvit S-HR sensor. Figure 5.13 shows the vertical velocity, roll rate and pitch rate from the 2D pose method and 3D method. The vertical velocity from the 2D pose method is scaled such that the magnitudes appear to be the same. From these plots it is clear that there is a change in vertical velocity which appears oscillatory, however, the roll and pitch rate show no discernible trend or oscillations. The vertical velocity of the 2D pose method appears to contain considerably more noise than the 3D method. This is mostly due to the inconsistent scale of the 2D pose method which makes it seem noisy. The 3D method however contains a much cleaner vertical velocity measurement. A zoomed in section is displayed in Figure 5.14 comparing all methods against one another. The figure shows that all camera methods appear to provide the same measurements irrespective of whether they take other motions into consideration. The side-slip angle however still appears to be oscillatory even though the vehicle maintains a straight driving line. The oscillatory nature is still likely caused by other vehicle motions. These motions most likely cause the slip angle sensor to move relative to the ground as the vehicle body moves. This results in a side-slip angle to be generated at the sensor

which is not necessarily caused by the side-slip angle of the vehicle. To further investigate this, a Power Spectral Density of both the side-slip angle and the vertical velocity is shown in Figure 5.14. The plot indicates that both measurements have high energy content at the same low frequencies. This indicates that the vertical velocity could be the effect which causes the oscillatory nature in the side-slip angle.

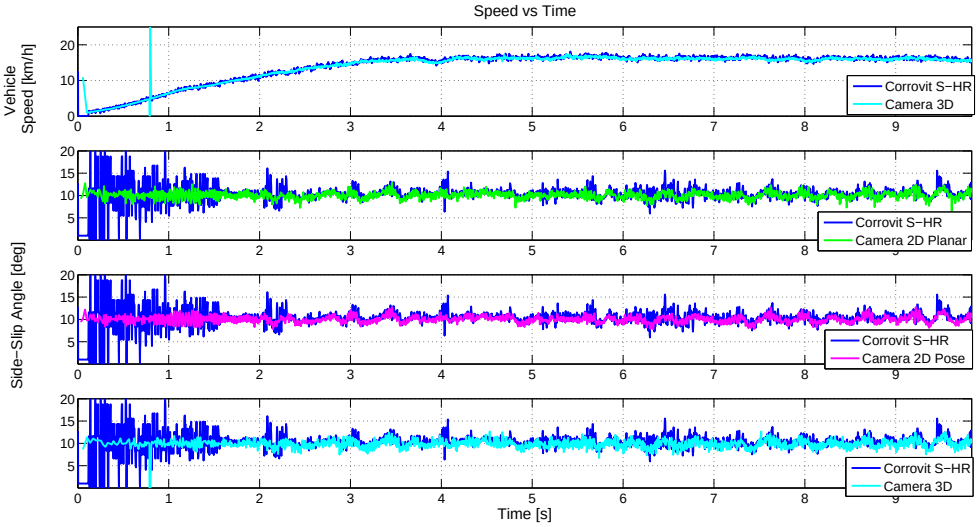


Figure 5.12: Vehicle speed and side-slip angle comparison between camera based techniques and Correvit S-HR side-slip angle sensor on Belgian paving surface for 10 degrees slip angle

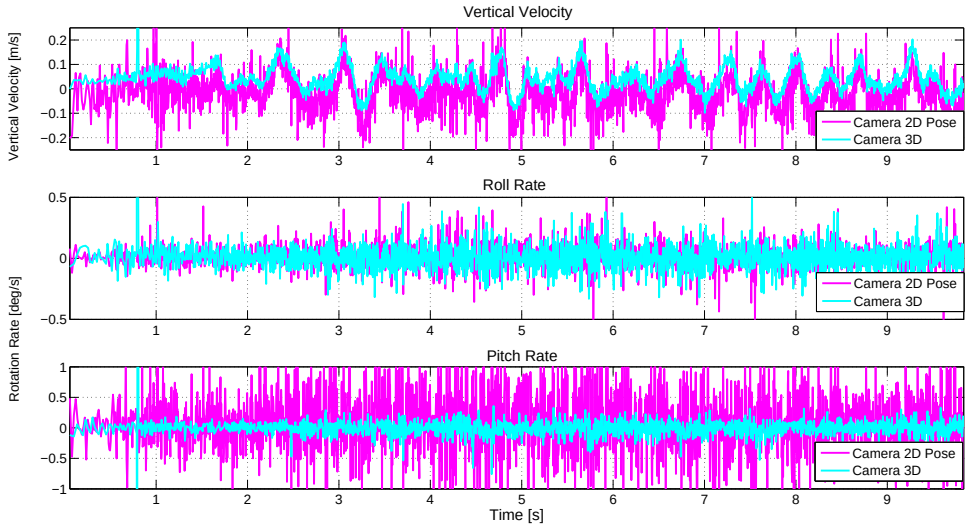


Figure 5.13: Vertical velocity, roll and pitch rate from 2D pose and 3D camera based methods on flat surface for zero slip angle

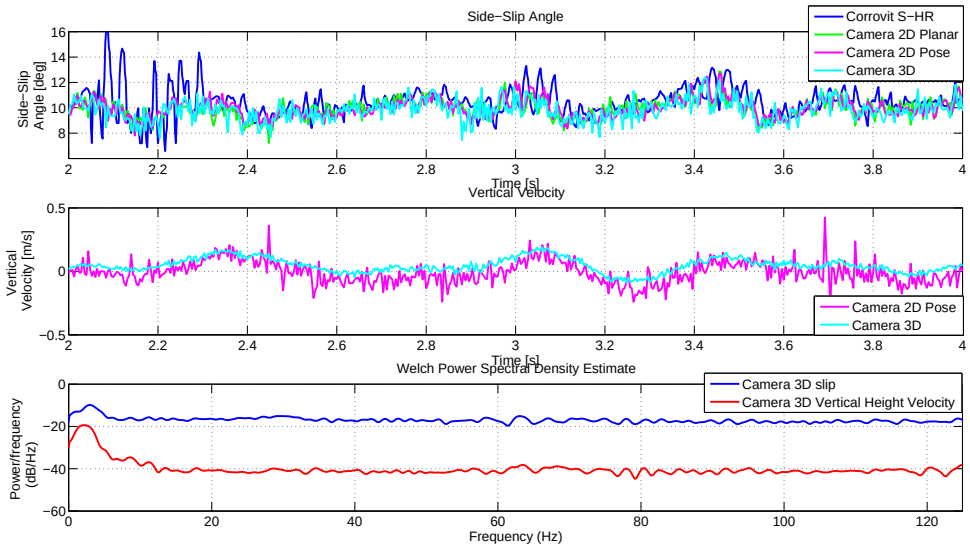


Figure 5.14: Enlarged region comparing different side-slip angle techniques, vertical velocity and power spectral density of 3D camera method side-slip angle and vertical velocity

Another experiment conducted was to move the mounting point of the sensors further away from the CG to near one of the wheels. The reason for this mounting position is that it is often easier to mount the sensors at this location as compared to the CG where space and ground clearance may be an issue. The side-slip angle of the tyres is also often measured to conduct tyre lateral force characterisation. The vehicle was driven in a straight line over the Belgian paving as in the previous test. Figure 5.15 shows a comparison of the side-slip angle at the different mounting positions with the slip angle set at $5deg$, the measurements are obtained using the 2D pose algorithm as the 3D method was not implemented at the wheel. The figure clearly shows that the measurements close to the wheel have much larger oscillations compared to the measurements at the CG. The vertical velocity also shows much larger values, however, it should be noted that the scale may not be exactly identical. This test indicates that the reduction of the other vehicle motions improves the accuracy of the measured side-slip angle. The effect these motions have on the sensor should also be reduced by mounting the sensor as close to the vehicle CG as possible.

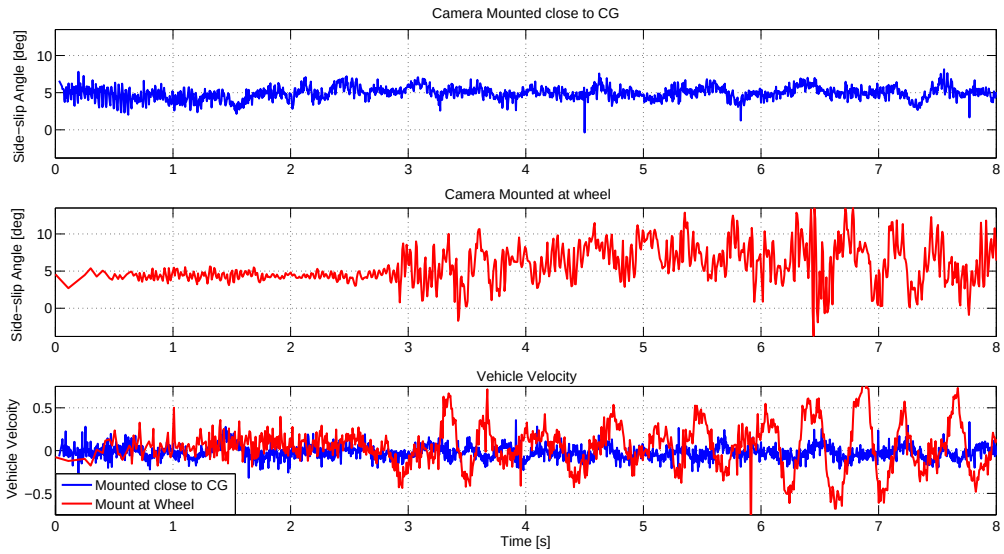


Figure 5.15: Comparison between side-slip angle and vertical velocity at the CG and wheel mounting locations

Table 5.1: RMS, mean and standard deviation for all techniques at various slip angles over a flat and Belgian paving surface

	Correvit S-HR			**2D Planar**			**2D Pose**			**3D**		
Flat	**RMS**	**Mean**	**STD**	**RMS**	**Mean**	**STD**	**RMS**	**Mean**	**STD**	**RMS**	**Mean**	**STD**
0	2.30	0.42	2.26	0.91	-0.36	0.84	0.91	-0.31	0.85	0.83	0.06	0.83
2	2.76	2.45	1.28	1.87	1.75	0.66	1.89	1.76	0.69	1.81	1.7	0.6
5	5.86	5.48	2.09	4.98	4.90	0.89	5.02	4.92	0.98	4.93	4.85	0.87
10	10.39	10.10	2.45	9.96	9.95	0.53	9.97	9.95	0.60	9.81	9.77	0.87
Belg	**RMS**	**Mean**	**STD**	**RMS**	**Mean**	**STD**	**RMS**	**Mean**	**STD**	**RMS**	**Mean**	**STD**
0	1.71	0.65	1.58	0.91	-0.09	0.91	0.94	-0.02	0.93	0.92	-0.09	0.92
2	2.98	2.52	1.59	2.10	1.85	0.99	2.18	1.92	1.04	2.06	1.83	0.95
5	5.60	5.42	1.42	4.92	4.85	0.83	4.97	4.90	0.86	4.89	4.79	0.96
10	10.56	10.35	2.12	10.03	10.00	0.75	10.07	10.04	0.78	9.9	9.87	0.87

The remainder of the tests at the other side-slip angles are tabulated in Table 5.1. The results are presented as RMS values, mean and Standard Deviation (STD). These values indicate how consistently the measurement measures the correct side-slip angle. The results are also graphically depicted in Figures 5.16 and 5.17. These figures however show the RMS and mean error. The results show that the Correvit S-HR has a consistently larger standard deviation and that the mean and RMS values are further away from the correct set point. The results between the camera based measurements are very similar showing almost no discernible difference between the three algorithms.

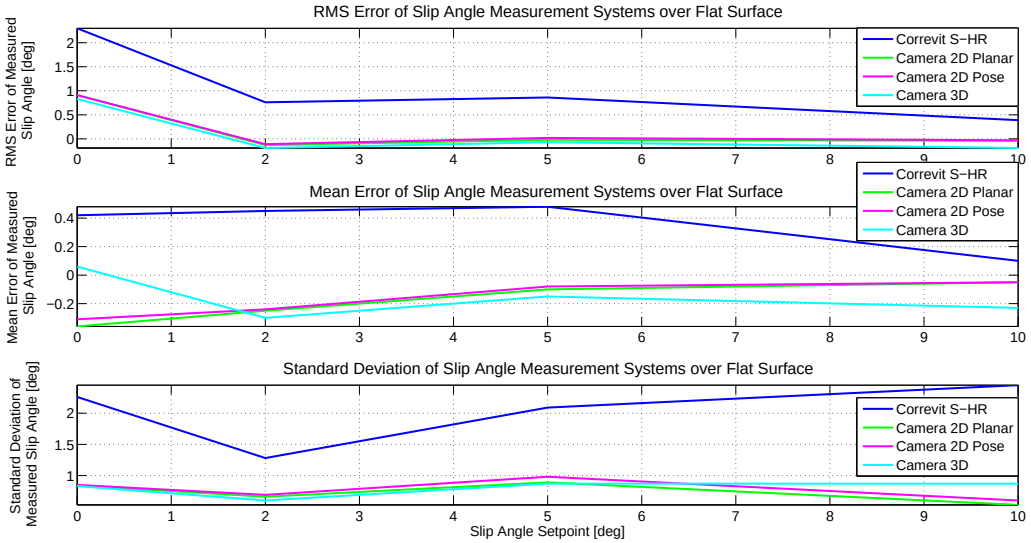


Figure 5.16: Comparison of RMS error, mean error and standard deviation between slip angle measurement systems over flat surface

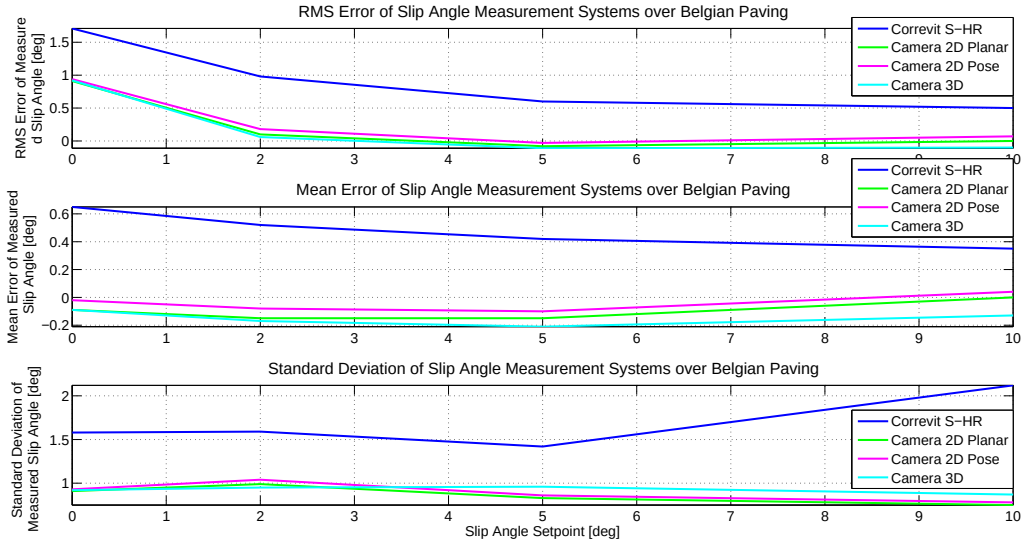


Figure 5.17: Comparison of RMS error, mean error and standard deviation between slip angle measurement systems over Belgian paving

5.6 Maximum Operating Speed

The maximum operating vehicle speed at which successive overlapping images can be obtained is limited by the size of the area captured and the sampling frequency. The size of the area captured is a function of the distance between the surface and the camera as well as the focal length of the lens as described in Chapter 2. The shutter opening time should also be as short as possible to reduce motion blur. A simplistic estimate of the maximum speed can therefore be based on firstly the required distance of overlap between images. The second limitation is the motion blur of the image. This is conservatively set at the magnitude of motion which would move the value from one pixel to the next. This is determined by first the spatial resolution per pixel and the shortest exposure time achievable by the camera. An estimate of the theoretical maximum operating speed is therefore obtained from the following calculations:

$$V_{max} = min(V_{max,overlap}, V_{max,blur}) \tag{5.26}$$

$$V_{max,overlap} = f_{overlap} L_{frame} FPS \tag{5.27}$$

$$= f_{overlap} \frac{L_{sensor} Z}{f} FPS \tag{5.28}$$

$$V_{max,blur} = \frac{L_{pixel}}{t_{exposure}} \tag{5.29}$$

$$= \frac{L_{frame}}{N_{pixel} t_{exposure}} \tag{5.30}$$

$$= \frac{\frac{L_{sensor} Z}{f}}{N_{pixel} t_{exposure}} \tag{5.31}$$

where V_{max}, $V_{max,overlap}$ and $V_{max,blur}$ are the absolute maximum speed, the maximum speed as limited by the overlap and the maximum speed limited by motion blur respectively. $f_{overlap}$ is the fraction of overlap required, L_{frame}, L_{sensor} and L_{pixel} are the length captured by a frame, the length of the sensor and the length captured by a single pixel respectively. FPS is the frame rate and N_{pixel} is the number of pixels along the image sensor length. Considering the most common smallest sensor size $1/3"$, diagonally, set at a resolution containing 640*pixels* in a length of 4.8*mm*. With the camera system used capable of 500*fps*, at 640×480 resolution, with the shortest exposure time of 0.006*ms* and using a 4*mm* focal length lens mounted 300*mm* from the terrain. The required overlap is set at 50%, therefore the maximum possible operating speed of:

$$V_{max,overlap} = f_{overlap}\frac{L_{sensor}Z}{f}FPS \quad (5.32)$$
$$= \frac{1}{2}\frac{(4.8)(300)}{4}500 \quad (5.33)$$
$$= 90m/s \quad (5.34)$$
$$V_{max,blur} = \frac{\frac{L_{sensor}Z}{f}}{N_{pixel}t_{exposure}} \quad (5.35)$$
$$= \frac{\frac{(4.8)(300)}{4}}{(640)(0.006e-3)} \quad (5.36)$$
$$= 93.75m/s \quad (5.37)$$

The theoretical speed is limited by the ability of the camera to capture images which have sufficient overlap. The maximum speed is however over $300km/h$ and sufficient for any vehicle testing. The maximum speed can be improved by employing different cameras, changing the geometry or lenses. If the lens focal length is reduced or the camera is moved further away then the maximum speed due to the overlap will increase, while reducing the accuracy. The accuracy can be compensated for by increasing camera resolution, this would however reduce the maximum frame rate possible unless higher throughput cameras are used. The motion blur can also be reduced by using cameras with faster shutters or by employing strobe lighting with a very short exposure time. This would increase the maximum speed due to motion blur.

5.7 Future Work

It is noted in this chapter that the Pose and the 3D algorithms not only measures the side-slip angle but also enables the measurement of the rotational velocities and translational velocities, scaled in the Pose method. A future study could investigate the possibility of using these two algorithms to perform the function of rate gyroscopes. Currently there is also no easy means to measure the vertical velocity of the vehicle. The vertical velocity is a useful measure used in the sky hook [Hrovat et al., 1988] and ground hook control [Novak and Valasek, 1996] suspension control algorithm. In both sky hook and ground hook the direction of the body vertical velocity is used to improve the ride and the handling respectively.

Another study that could be of great value is whether the 3D method can be used to determine the pitch and yaw angles as well as the vertical height of the vehicle body. This could be achieved

not only by integrating the velocities but also by using all the 3D points of the surface. If the road is flat a plane could be fitted through the 3D points of the surface. From this plane, the roll and pitch angles as well as the vertical height of the plane can be obtained and therefore of the vehicle as well.

5.8 Conclusion

Three techniques, capable of measuring the side-slip angle based on camera images were presented in this chapter. Two of the methods use a single inexpensive camera to measure the side-slip angle. The first method measures the slip angle using the projected features on the camera image. This method relies mostly on the camera motion being planar. The second method uses a 5 point pose estimation algorithm to estimate the scaled translational and rotational velocities. This pose technique corrects for certain errors induced by other vehicle motion and provides additional information compared to the planar technique. The third method relies on two inexpensive cameras configured in a calibrated stereographic rig to measure 3D world coordinates. This allows for all translations and rotations to be measured in unscaled real world coordinates. The algorithms are compared to a Correvit S-HR commercial side-slip angle sensor. Tests were conducted over both a flat surface as well as a Belgian paving surface.

The tests show that the camera based techniques have a lower noise band especially at low speeds compared to the Correvit S-HR sensor. No discernible differences between the three camera based techniques were observed even over the rough terrain.

However, the pose and 3D algorithms provide more information compared to the Correvit S-HR and 2D planar methods. The 2D planar method can easily be converted into a real time method. The 3D algorithm can also potentially be performed in real time using FPGAs, as previous studies have shown. The 2D pose method on the other hand is the most computationally intensive algorithm of all three and is therefore currently limited to off-line analysis. It however does not require any additional requirement than the 2D planar method to be performed. If off-line analysis are considered the raw image data can easily be passed through both single camera techniques. The 3D camera technique, while providing better information, has the added complexity of using two cameras. The cameras must be calibrated and have a rigid connection between one another. The images of the cameras should also have similar illumination and should have similar focus points.

If planar motion or small rotational excitation is experienced the planar motion is recommended.

However, the 3D or 2D pose methods are recommended if the vehicle experiences larger excitations, especially if these excitations are large in comparison to the vehicle velocity.

Chapter 6

Tyre Longitudinal Slip-Ratio Measurement

In Chapter 5 the side-slip angle of the vehicle or tyre is measured using DIC. Another important tyre measurement is the tyre longitudinal slip ratio. Measurement of the tyre longitudinal slip ratio is often estimated from three independent measurement devices namely wheel rotation speed, vehicle speed and tyre rolling radius. This produces an expensive measurement system which indirectly determine the slip-ratio. This chapter presents a method by which the slip-ratio is determined from a camera using DIC techniques. The camera, mounted in such a way that the contact interface between the tyre and the road is captured, enables the system to measure the tyre tread speed and ground speed at the contact patch. The slip ratio is then determined from these two measurements obtained from the same set of images.

6.1 Introduction

The mechanism for generating tyre longitudinal forces is an exceedingly complex process which occurs as a result of the relative sliding motion between a pneumatic tyre and the road. While the sliding motion is non-uniform across the contact patch [Deur et al., 2005], with regions where the contact patch adheres to the surface and regions where it slides, it can be shown that the generated longitudinal force can be sufficiently described as the ratio of the average relative speed between the tyre and the ground and the vehicle speed. This parameter is termed the wheel slip-ratio [Sakai, 1981]. This

slip-ratio is essential when constructing tyre longitudinal models and can be a useful measurement in vehicle stability programs such as traction control, ABS, ESP, and other. The slip ratio definition often used in many studies is given as [Wong, 2001]:

$$S = -\frac{V_t - V_a}{V_t} \tag{6.1}$$

where S is the slip ratio as a fraction, V_t is the theoretical speed (at zero slip) at the center of the wheel, V_a is the actual tyre speed. The theoretical tyre speed V_t is often merely represented as the vehicle longitudinal speed V. The actual tyre speed is determined as:

$$V_a = \omega R_e \tag{6.2}$$

where ω is the tyre angular velocity and R_e is the effective tyre radius.

In experimental laboratory tests aimed at charactering longitudinal tyre models, both the tyre and the surface are controlled accurately to move at known speeds. Thus, by measuring the tyre radius the wheel slip-ratio can be accurately determined. The tyre contact surface of these test setups typically consists of a steel rotating drum [Hüsemann and Wöhrmann, 2010]. Drum tests however have the disadvantage that the surface is usually smooth and not representative of asphalt and the tyre makes contact with a circular surface as opposed to a flat surface. While a flat track testing machine [Cabrera et al., 2003], alleviates the problem of a rounded surface, the surface however usually still consists of a steel belt which is sometimes coated with an artificial surface to simulate real road surfaces. Additionally, these testing machines are exceedingly expensive and are quite difficult to scale for larger tyres used in mining and agriculture.

An alternative to laboratory testing is in-situ testing, on real road surfaces, conducted using a tyre test machine or the actual vehicle the tyres are used on. This methodology alleviates the problems associated with laboratory testing. The most difficult task however is the accurate estimation of the wheel slip-ratio. Traditionally the wheel slip-ratio is computed from equation 6.1 by measuring the three parameters separately using three measurement devices. While it is fairly inexpensive to accurately measure the wheel rotation speed, the rotation speed is measured at the rim and not at the contact patch. Additional filtering often needs to be applied to the wheel speed signal. Accurate measurements of the vehicle speed and effective tyre roll radius are more expensive to obtain. Miller et al [Miller et al., 2001] shows how the slip-ratio can be determined using GPS. To obtain accurate GPS measurements at a high sampling frequency (50-100Hz) however requires an

expensive differential GPS receiver. It should also be noted that GPS signals can have a latency of up to 150ms [Bevly et al., 2006] affecting real time results under dynamic test conditions. The tyre rolling radius is most often not measured but rather assumed constant. While Miller et al determined the rolling radius using the GPS vehicle speed and wheel speed, this can only be accomplished under no-slip, i.e. when no braking is performed. Thus, for the entirety of the braking tests the rolling radius is assumed constant, however load transferred due to braking has an effect on the tyre rolling radius. It is therefore important to continuously measure the rolling radius.

Since wheel speed is measured at the rim, and not the contact point, the estimated slip-ratio is assumed to be produced immediately at the contact patch. The generated longitudinal force however has a small phase lag compared to the slip-ratio, [Rill, 2006], [Clover and Bernard, 1998]. The phase lag is partially attributed to the transient transfer of the longitudinal forces generated at the contact patch to the wheel hub due to the tyre inertia. The transient formation of the forces at the contact patch is often neglected. The cause of this transient build-up of forces at the contact patch is not well understood. The transient build-up is partially associated with the viscoelastic nature of the tread and the generation of the slip condition at the contact patch [Mavros et al., 2005], pertaining to the deformation of the tread and side wall [Van Zanten et al., 1990].

In this chapter a method whereby the tyre slip-ratio is measured at the contact patch is presented. The method comprises of a single camera, mounted in such a way to a tyre test machine or vehicle so that it captures both the tyre and the road at the tyre-road interface. The slip ratio is measured by performing DIC techniques on consecutive frames of the contact patch encompassing both the tyre and road. This method is applicable to hard terrains in which the tyre-road interface is visible i.e. not soft soil such as sand wher the interface may be covered in sand.

6.2 Longitudinal Slip Ratio from Digital Image Correlation

The slip ratio is determined by post processing a sequence of images obtained from a single calibrated and undistorted camera. Features on both the tyre and the road, close to the contact patch, are tracked using the DIC method described in 4.1.2. The SIFT feature tracker is used due to the poor result obtained using the Lucas-Kanade (section 4.1.1) optical flow algorithm. This could be due to larger motions being experienced or one of the other assumptions of the algorithm being invalid. A small window is selected in image $\boldsymbol{I_t}$ which encapsulates both the tyre and road close to the contact

interface. The key points in this image is tracked across the whole of image $\boldsymbol{I_{t+1}}$ (see Figure 6.1).

The DIC method used can at times produce mismatched key points. In order to reduce the mismatched key points, an initial outlier rejection is performed. Since the speed of the regions under consideration should be mostly horizontal, matches that result in large deviation in the y-direction are rejected, i.e. key points which move too much in the vertical direction.

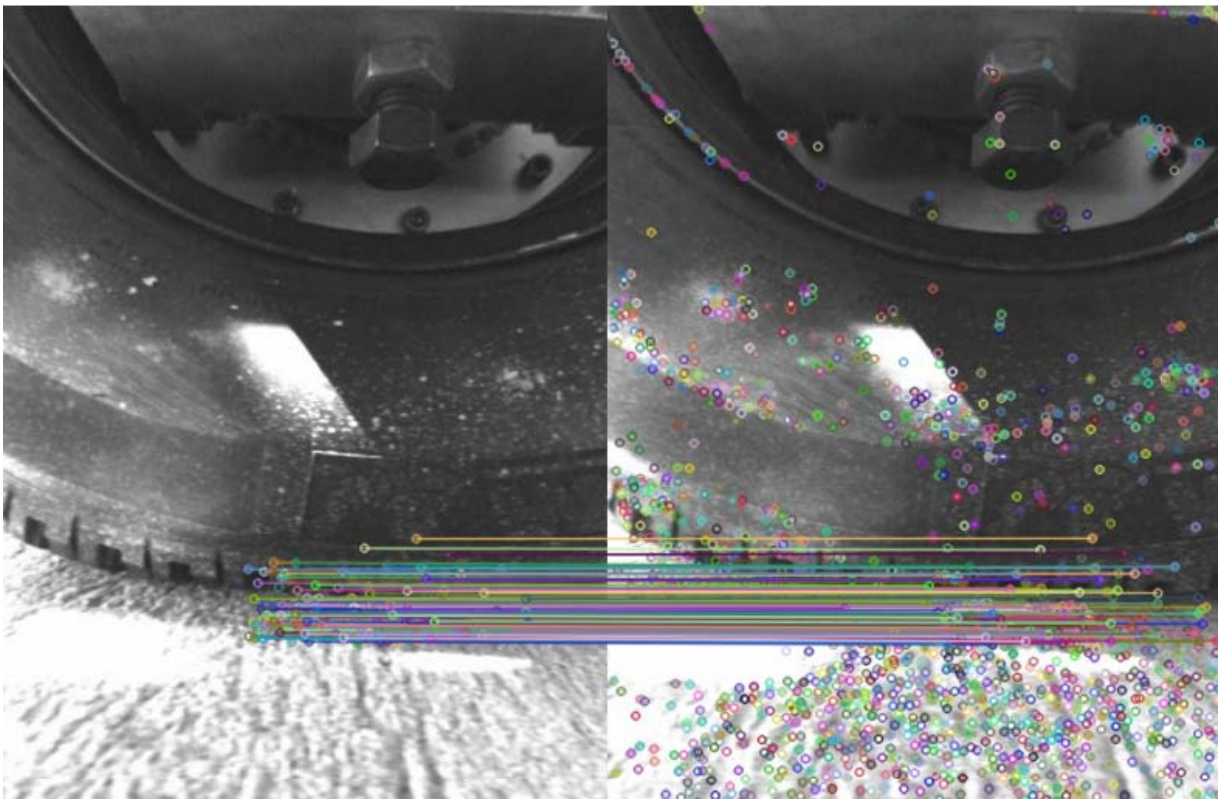


Figure 6.1: Tracking of key points from contact patch in image $\boldsymbol{I_t}$ (left), to image $\boldsymbol{I_{t+1}}$ (right), and matched key points are joined with a line across the two images

Once the key points have been matched they have to be separated into three categories, points lying on the tyre, points lying on the road and any additional outliers. At first a simplistic method was proposed whereby a horizontal line is drawn at the contact point. Any points above this line is considered to be on the tyre and any points below are on the road. This method, while being very efficient and successful at the start of the test, causes some points to be categorized incorrectly due to the contact point moving vertically during the test. The method also fails to remove any additional outliers which are present. The vertical movement of the contact patch is due to the roll radius changing as a result of load being transferred during braking as well as the test rig bouncing. A more robust method for categorizing the points is therefore required.

In this study a clustering technique is implemented to characterize the key points into the three categories. Two clustering methods generally used are the K-means and agglomerative clustering technique [Gupta et al., 2011]. In the K-means approach, K cluster centroids are chosen and all points are joined to these centroids. The centroids are updated after all points have been joined to the clusters. The process then restarts by joining all the points to the new updated centroids. This process continues until the centroid's position change very little. The disadvantage of this method is that the process continues until steady state of the centroids are reached. In the agglomerative clustering technique each point is considered as a cluster. Iteratively the two clusters which are the closest are joined and the cluster centroid is recalculated. The process is repeated until the desired numbers of clusters are obtained. The disadvantage of this method is that prior knowledge of the number of clusters is required. In this application while we require three groups (tyre, road and outliers), the situation could arise when there is no outlier cluster, resulting in two clusters containing either tyre or road points. An alternate case is where there may be more than one outlier cluster, with one outlier cluster closer in distance to either the tyre cluster or the road cluster rather than another outlier cluster. This would result in outliers being clustered together with either the tyre or road clusters.

The clustering algorithm implemented in this study includes aspects from both clustering techniques. Two main clusters are initially created, a tyre and a road cluster, containing initial points in each cluster. The remaining key points are considered to be an individual cluster themselves. The distance from each cluster to one of the two main clusters is calculated. The cluster with the smallest distance to one of the main clusters is then included into the corresponding main cluster. The cluster centroid is recalculated along with the distance to each of the remaining individual clusters. This process is repeated until the smallest distance between clusters exceeds some threshold. The remaining clusters, not included into one of the main clusters, are considered outliers. This process is graphically illustrated in Figure 6.2. The initialisation creates the initial tyre (cluster 1) and road (cluster 2) clusters. Iteratively the point closest to any of these two clusters is joined into that cluster. In the first iteration a point is joined into cluster 2. The centre of this cluster is recalculated and the distance to each point is recalculated. This process is continued for N iterations until the distance of a point to a cluster is larger than a threshold. The final clusters are then obtained with the remaining points classified as outliers.

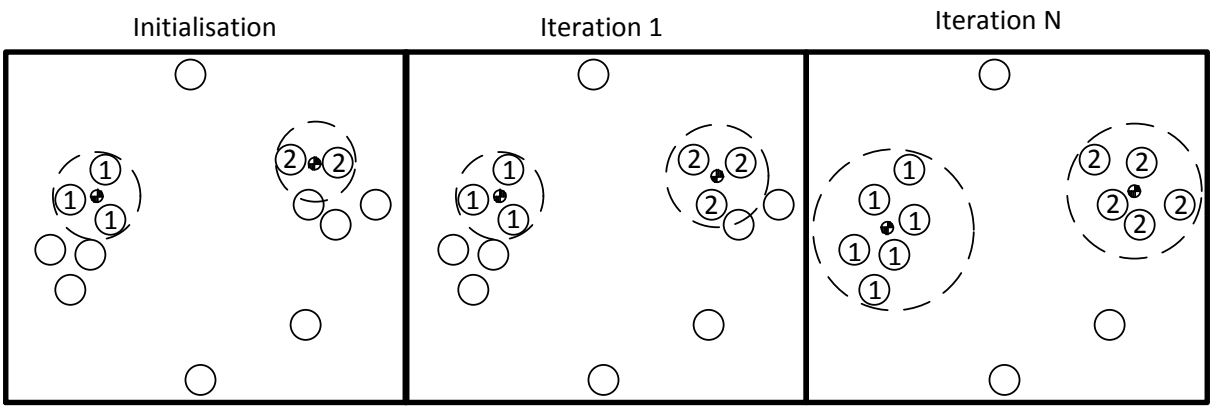


Figure 6.2: Clustering algorithm. Initial tyre and road clusters are created (left). Points are then iteratively merged into these clusters based on distance to cluster (middle). Process continues until the distance to a cluster exceeds some threshold (right).

The distance metric used is the Euclidean metric given as:

$$p_{ij} = \sqrt{\sum_{k=1}^{d} (\boldsymbol{X}_{ik} - \boldsymbol{X}_{jk})^2} \tag{6.3}$$

where p_{ij} is the Euclidean distance between clusters i and j in d-dimensional space. $\boldsymbol{X}_{ik}$ is the kth component of cluster i's centroid. In this study a three dimensional space is used to describe each cluster. The first component is the speed of the cluster. This component is used since the tyre and road should have different speeds, especially during slip. The second component is the x-pixel coordinate from the image where the cluster is situated. This component is included because the tyre and road are situated on different vertical locations on the image. The last component is the homography error of the cluster. Since there is a possible difference in magnitude in the three components, the components are normalized such that each component vector has a mean of zero and a standard deviation of one. The cluster center is simply taken as the mean of the cluster data points. The following section discusses why a homography error is included as a component of the clusters.

As discussed in section 3.4.1 planar homography is the projective mapping of a planar surface to another planar surface. The homography is represented by a 3×3 matrix $\boldsymbol{H}$, with the mapping of image correspondences on a plane mapped as:

$$\boldsymbol{H}\boldsymbol{q} = \boldsymbol{q}' \tag{6.4}$$

Where $\boldsymbol{q}$ is the homogeneous coordinates of a correspondence in image $\boldsymbol{I_t}$, and $\boldsymbol{q'}$ is the associated correspondence in image $\boldsymbol{I_{t+1}}$. As stated in section 3.4.1 the homographic matrix contains 9 elements, but is ambiguous up to scale and therefore has only 8 degrees of freedom. Thus, only 8 independent elements need to be solved for. The scale can be arbitrarily chosen or set such that $||\boldsymbol{H}|| = 1$. The homographic matrix is estimated from the fact that the cross product of $\boldsymbol{Hq}$ and $\boldsymbol{q'}$ is zero:

$$\boldsymbol{q'} \times \boldsymbol{Hq} = \begin{bmatrix} y'h_3\boldsymbol{q} - z'h_2\boldsymbol{q} \\ z'h_1\boldsymbol{q} - x'h_3\boldsymbol{q} \\ x'h_2\boldsymbol{q} - y'h_1\boldsymbol{q} \end{bmatrix} = 0 \tag{6.5}$$

Where h_1 is the first row of the homohraphy matrix $\boldsymbol{H}$. The third equation is a linear combination of the first two, therefore each correspondence provides two linearly independent constraints on the elements of $\boldsymbol{H}$. Thus, in order to estimate the homographic matrix $\boldsymbol{H}$ at least four correspondences are required which lie in a plane. Equation 6.5 can be rewritten as:

$$\boldsymbol{q'} \times \boldsymbol{Hq} = \boldsymbol{A} \begin{bmatrix} h_1 \\ \vdots \\ h_8 \end{bmatrix} \tag{6.6}$$

$$= \begin{bmatrix} -q_1^1 & -q_2^1 & -1 & 0 & 0 & 0 & q_1'^1 q_1^1 & q_1'^1 q_2^1 & q_1'^1 \\ 0 & 0 & 0 & -q_1^1 & -q_2^1 & -1 & q_2'^1 q_1^1 & q_2'^1 q_2^1 & q_2'^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -q_1^4 & -q_2^4 & -1 & 0 & 0 & 0 & q_1'^4 q_1^4 & q_1'^4 q_2^4 & q_1'^4 \\ 0 & 0 & 0 & -q_1^4 & -q_2^4 & -1 & q_2'^4 q_1^4 & q_2'^4 q_2^4 & q_2'^4 \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_8 \end{bmatrix} = 0 \tag{6.7}$$

where q_i^j is the i^{th} value of the j^{th} correspondence and h_k is the k^{th} value of the homographic matrix. The homograhic vector in equation 6.7 is solved using Singular Value Decomposition (SVD) and the solutions are obtained from the right singular vector corresponding to the smallest singular value. An overview of this method is provided in Hartley and Zisserman [Hartley and Zisserman, 2003].

The estimation is achieved by means of the RANSAC algorithm provided in section 5.2.1. The inlier metric used in the RANSAC algorithm is the error $\epsilon = ||\boldsymbol{q'} - \boldsymbol{Hq}||$. The correct model is taken as the model with the most inliers

The homography is important as it can be used to model the road surface the tyre is rolling over under the assumption that the road can be represented as a plane. A key point which strongly agrees with the model is considered to be on the road and a point which does not is considered to lie off the plane or a mismatched outlier. To generate the homography, correspondences which are guaranteed to be part of the road plane are chosen. This is achieved be performing DIC on a region below the contact patch which only contains the surface the tyre is rolling over. The homography is also used to seed the road cluster with initial points. Key points with a homographic error below a certain threshold are considered to be on the road and included into this cluster, with the threshold being set very low to ensure only road points are selected. This process is repeated at each image $\boldsymbol{I_t}$ to seed the initial road cluster with points.

The initial points on the tyre are selected from the first image $\boldsymbol{I_0}$ based on their vertical pixel height being within a prescribed range above the initial road points, as well as their velocity magnitudes having similar values to the road points thus assuming an initial no slip condition. This process is only performed on the first image after which the cluster is seeded with points from the previous image $\boldsymbol{I_{t-1}}$. The points from the previous image are selected to be the points with the highest vertical pixel coordinate. This is done to prevent the tyre cluster from migrating towards the road cluster. The points selected are also not allowed to be chosen to seed the tyre cluster in the next iteration. If no new points have been included into the cluster, then in the following iteration the tyre cluster is initialized using the same method as initializing the cluster using the first image $\boldsymbol{I_0}$. The tyre-road contact point's vertical pixel coordinate is also determined at each iteration. This is accomplished by taking the distance halfway between the centres of the tyre and road clusters vertical pixel coordinate. However, instead of completely replacing the previous iteration's estimate of the contact point height with the new estimate, the estimate is updated by a fraction of the error between new and previous estimate as:

$$R_{y,t+1} = R_{y,t} + c\,(R_{y,t+1} - R_{y,t}) \tag{6.8}$$

where $R_{y,t+1}$ is the updated position of the contact point in the y-direction, $R_{y,t+1}$ is the new estimate of contact point and c is a constant describing the rate of update, $c \leq 1$. The tyre -road contact point is used to position the window from which the key points are generated to allow an

equal amount of tyre and road region being included. This parameter also effectively means that the change in effective tyre roll radius is captured from the camera.

Figure 6.3 depicts one iteration showing images $\boldsymbol{I_t}$ and $\boldsymbol{I_{t+1}}$ on the left and right respectively. The left image shows the trace of each key point depicting the velocities of the key point as well as the windows from which key points are generated. The image on the right depicts the clustering result with the red points classified as tyre points, the blue points on the road and green points classified as outliers. The red horizontal line also depicts the tyre-road contact height. The figure shows excellent classification of both tyre and road points with no outliers contained within these groups. There are however some points which should be classified as either being a tyre or road points that are classified as outliers. This is of no concern as enough points are classified correctly with no inclusion of outliers. This method was found to be very robust at classifying the key points correctly while removing all outliers.

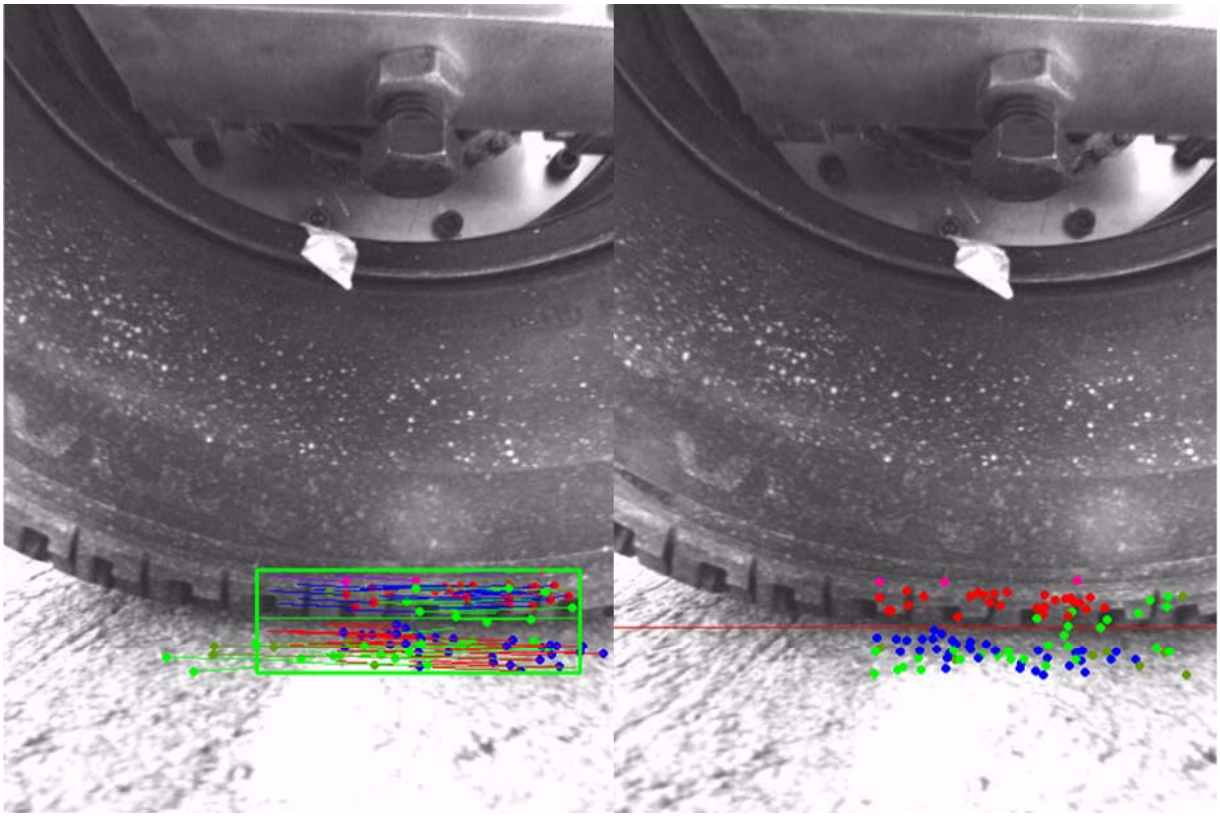


Figure 6.3: Tracking and classification of key points from contact patch in image $\boldsymbol{I_t}$ (left), to image $\boldsymbol{I_{t+1}}$ (right). Points classified as tyre points have red dots and those on the road have blue dots. Outliers are coloured green

After classification, the data points inside the clusters are used to obtain the velocity of the road and of the tyre at the contact point. One aspect which needs to be taken into consideration is the parallax effect caused by the difference in distance from the camera (depth) to the key points. This effect is illustrated in Figure 6.4, where key point KP_i is closer to the camera than key point KP_j. As illustrated this causes the velocity measured at KP_i (V_i) to be much larger than the velocity measured at KP_j (V_j).

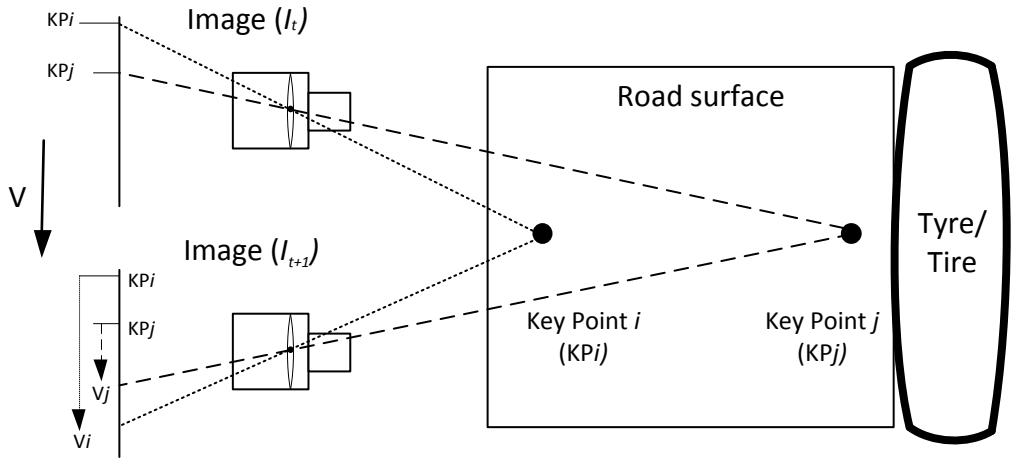


Figure 6.4: Illustration of the parallax effect caused by points having different depths from the camera

To compensate for this effect all velocities must be transferred to the velocity measured at the contact point. The change in depth for the road can be captured by the vertical pixel coordinate, Figure 6.5 shows a plot of the key point velocity against the vertical pixel coordinate of the key point. From this figure it is evident that there is a substantial change in velocity with respect to a change in vertical pixel location for the key points located on the road. Comparatively there is no noticeable change in velocity for key points located on the tyre. This is mostly due to all point lying on the tread which seem to experience the same speed irrespective of distance from the point of rotation. The accuracy of the system however could be improved by focusing on a smaller region. This may lead to small variations in the tread speed both vertically and longitudinally to be detected. With the current setup and obtained relationship the tyre velocity is simply taken as the average of all key points on the tyre. However, for the road the parallax effect must be compensated for. It is also noticeable from the figure that the relationship between the vertical pixel coordinate and the velocity

of the road key points seems to be linear in nature. Thus, a straight line is fitted through these points to obtain a model for this relationship. This is achieved by making use of the RANSAC algorithm. However, in this RANSAC algorithm all the inlier points for the best model are used to fit a linear Least Means Square fit between all the inlier points. This effectively bases the fit on inlier points only, removing any outliers which could distort the relationship. This process is performed at each iteration since the model changes as a result of a change in the vertical displacement of the tyre i.e. the bouncing of the tyre. Additionally the fit from the previous iteration is used as the first model in the current RANSAC algorithm. Once this model has been obtained the model is simply used to determine the velocity at the contact point $R_{y,t+1}$.

From this process the velocity of the tyre and road has been determined at the contact point from which the slip ratio is determined using equation 6.9. A flow chart of the complete algorithm is shown in Figure 6.6.

$$s = -\frac{V_{road} - V_{tyre}}{V_{road}} \tag{6.9}$$

A flow diagram of the complete algorithm is shown in Figure 6.6.

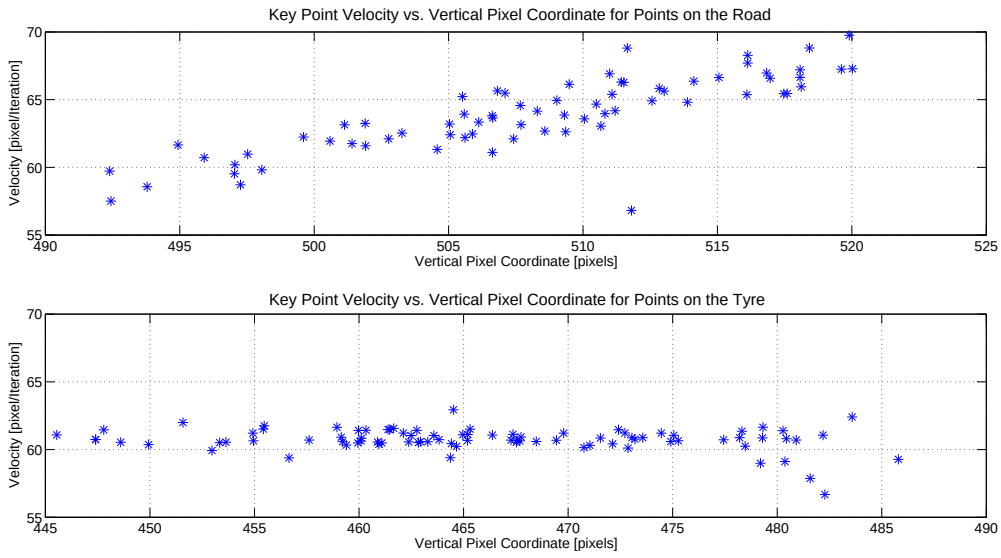


Figure 6.5: Velocity of key points versus vertical pixel coordinate for points on the road (top) and points on the tyre (bottom)

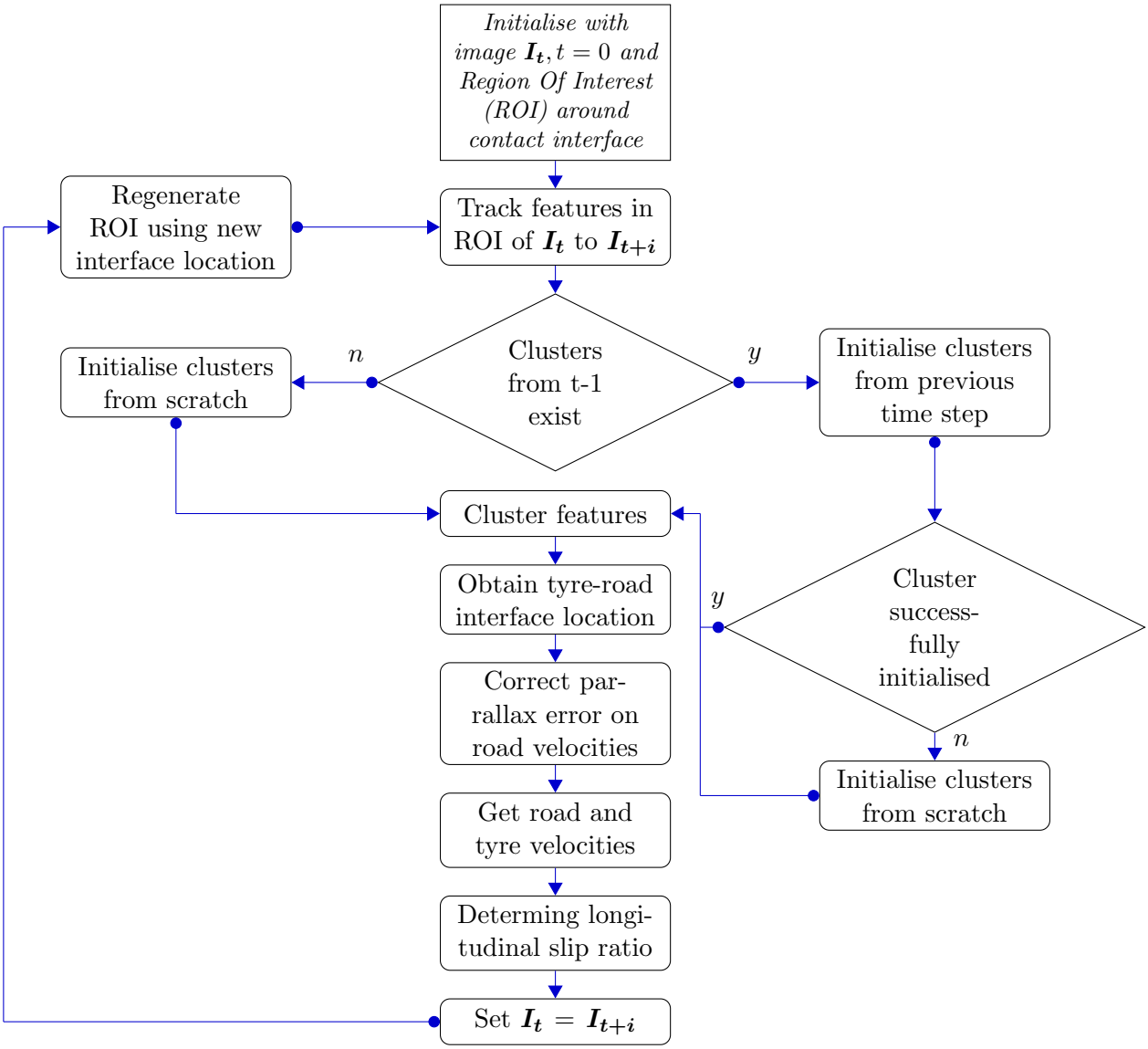


Figure 6.6: Flow chart of longitudinal slip ratio measurement algorithm

6.3 Experimental Setup And Results

The method is validated by comparing it to the traditional method of estimating the slip-ratio from equation 6.1. The method is tested on a tyre test rig as shown in Figure 6.7. The tyre test rig is equipped with a wheel force transducer, developed by the University of Pretoria, which can measure all six force/moment components as well as the rotational angle of the wheel. A road profiling laser is used to measure the effective tyre rolling radius and a differential GPS is used to measure the test

rig speed.

The tyre test rig is pulled using a large truck with the vehicle speed kept as constant as possible. The test rig is equipped with an electronically controlled pneumatic braking system which is programmed to provide a gradual increase in braking, thus generating a gradually increasing slip-ratio. A Point Grey Flea3 (USB 3.0) (see Appendix A for specifications) camera is mounted facing the contact patch of the tyre, as shown encircled in Figure 6.7. A $4mm$ wide angle lens is used (see Appendix A for specifications). The camera is sampled and synchronized with the data by using a $100Hz$ trigger. During the braking test a fairly low resolution grey-scale video of 640×480 pixels is captured of the contact patch region. The slip-ratio is then determined in post processing from the recorded video.

Figure 6.7: Tyre test rig, showing encircled camera mounting position

In order to compare the measurements obtained using the camera, with units in pixels, to the measurement obtained with traditional means required the scaling of the camera measurements. The comparison of effective roll radius and vehicle speed along with the hydraulic brake pressure is presented in Figure 6.8. The data obtained from the camera starts at $16s$, and is synchronised using the trigger signal. The effective roll radius plot, in the middle, illustrates the change in roll radius as a result of the load transferred, depicted by the gradual increase, as well as the bouncing of the tyre test rig depicted by the oscillatory nature of the graph. The plot also shows that initially the camera roll radius estimation has a slight error, however, as the test continues better correlation is obtained with the camera capturing both the gradual increase as well as oscillatory change in the tyre effective roll radius. The bottom plot depicts the vehicle speed obtained from the GPS as well as the camera. It can be observed that the camera shows good correlation overall but contains a gradual

decrease and is more oscillatory. These effects are caused by the change in camera position relative to the road, and can be noted to be the inverse of the effective roll radius measurement from the camera. This is because the change in camera position creates a slight change in parallax, however this phenomena occurs on the tyre velocity measurement as well, depicted in Figure 6.9, effectively negating this effect when determining the slip ratio. It should also be stated that the dip in GPS obtained vehicle velocity between $20s$ and $22s$ is due to a loss of a satellite and is therefore not a true representation of the actual vehicle speed.

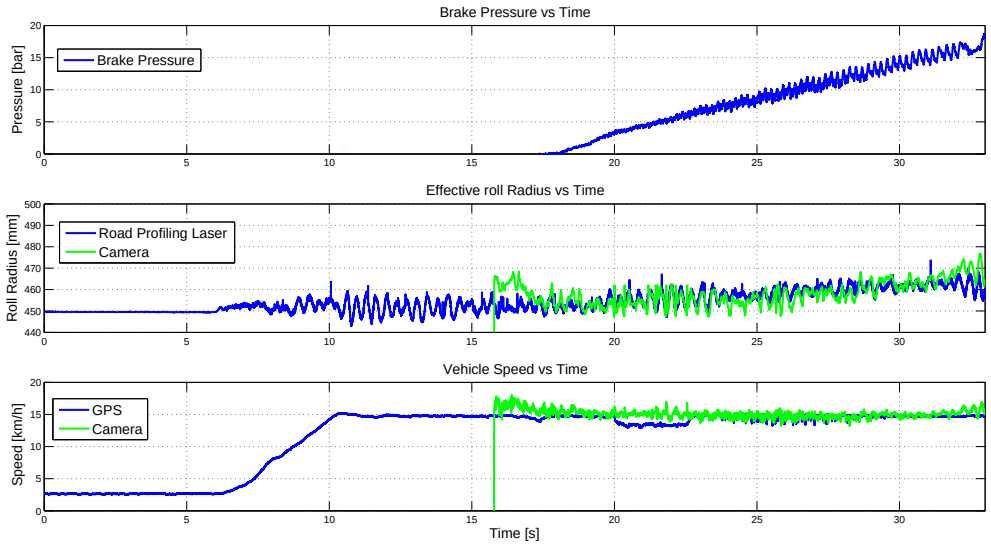


Figure 6.8: Hydraulic brake pressure, comparison between effective roll radius of a road profiling laser and the camera system, vehicle speed comparison between GPS and camera system

Figure 6.9 also shows the wheel torque measured at the wheel (middle plot). It is evident from this plot that the torque has a high degree of oscillation. The same oscillations are evident from the brake pressure. These oscillations were caused by misalignment in the brake system as evident from the brake pressure. These oscillations also caused the arm which creates a fixed reference for the angle sensor in the Wheel Force Transducer to vibrate at the same frequency. As a result of this an artificial signal is generated on the angle measurement, which after differentiating to obtain the wheel rotational speed is amplified. The result is that the slip measurement during the large oscillations is largely erroneous. This is better illustrated in Figure 6.10 showing an enlarged region containing the oscillations, and the region with less oscillatory behaviour. While it is evident that the oscillatory brake torque should result in an oscillatory tyre speed, the excessive oscillations from the

wheel rotational speed sensor is largely noise. This noise generated by the movement of the fixed reference is a problem associated with this technique to measure wheel speed. The fixed reference will in general not be a rigid connection with not only the arm being flexible but the mounting point may experience motion which could induce noise on the measured angle. The motion of the mounting point could be due to the motion of the vehicle body or a change in camber or caster of the wheel itself. A better alternative is to use a proximity based measurement often used in ABS systems to measure wheel speed. However, even these measurements can be corrupted with noise due to manufacturing tolerances, corrosion, dirt, low-frequency disturbances due to varying air gap between sensor head and the sensor wheel and disturbances on the transmission path [Schwarz et al., 1997]. The generated noise would require filtering which could lead to a phase lag in the measurement therefore skewing the results.

The camera based tyre speed shows the same trend however at a reduced magnitude. In the regions where the oscillatory wheel torque is much less pronounced, which also resulted in less induced noise on the angle sensor, the data shows good correlation between the contact patch tyre speed obtained from the rotational wheel speed sensor and the camera based measurement.

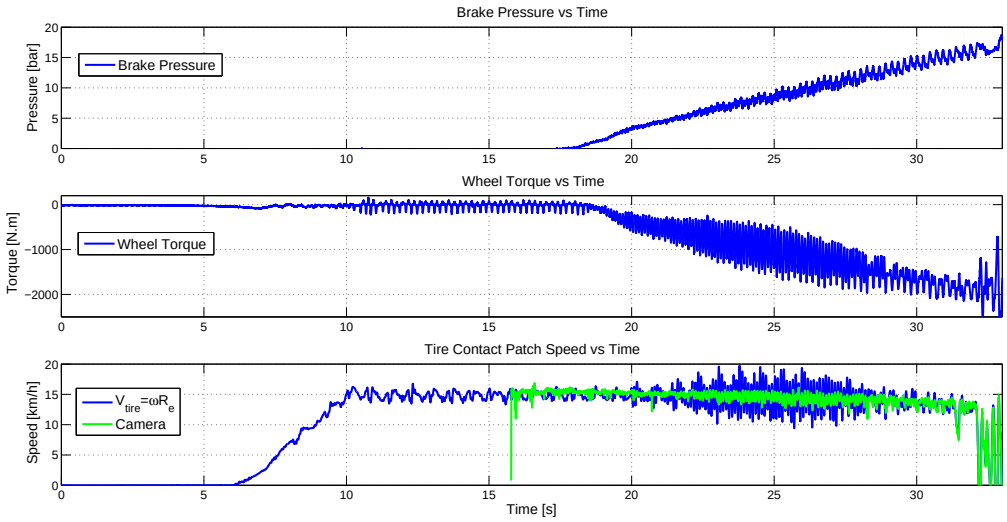


Figure 6.9: Hydraulic brake pressure, wheel torque and estimated tyre contact patch speed comparison between camera based method and using the effective roll radius and wheel rotation speed

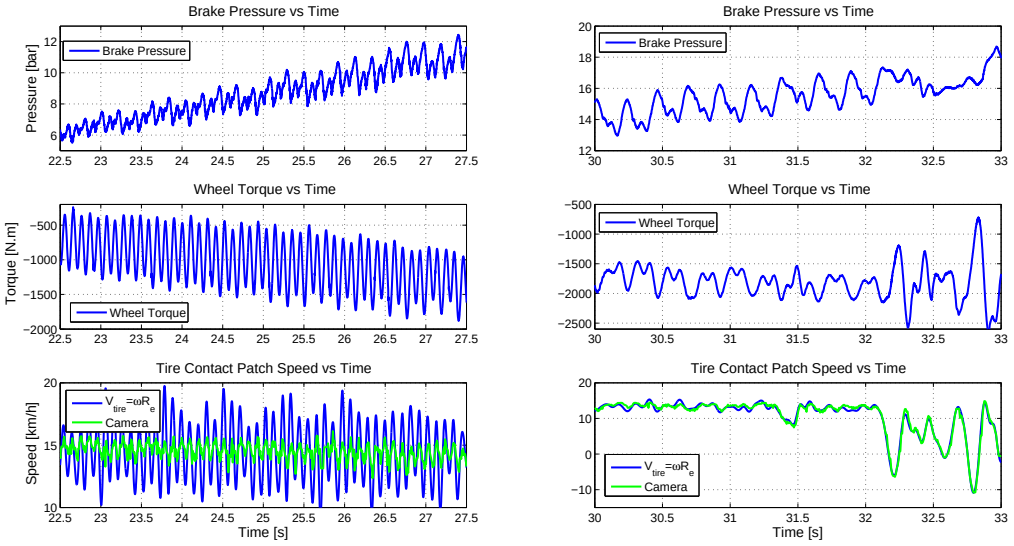


Figure 6.10: Enlarged regions of the hydraulic brake pressure, wheel torque and estimated tyre contact patch speed comparison between camera based method and using the effective roll radius and wheel rotation speed

Figure 6.11 depicts the comparison of the slip-ratio determined from the camera system to that of the traditional method using equation 6.1. Again the oscillatory nature of the method from equation 6.1 is evident from the top plot. If the region from 30-33s is compared, as shown in the bottom plot, excellent comparison of the slip-ratio is shown. It should be noted that the slip-ratio values exceeding 100% should be inverted to be negative, however, the graph is better displayed by keeping these values positive. It is also noted that there is a slight phase lag of 10ms between the signals, this value however falls within the sampling frequency of the camera. It is therefore impossible to say with the current data whether the slip-ratio generated at the tyre is indeed out of phase of that determined using the rim rotational speed. For an analysis into the phase difference a higher frequency is required, as well as, good control over the phase shifts caused by any filtering process during sampling. The longitudinal force as a function of tyre longitudinal slip-ratio is displayed in Figure 6.12. This figure clearly illustrates that the relationship obtained from the camera presents typical results expected from such a test [Van Zanten et al., 1990] but the conventional method gives unsatisfactory results. It can be concluded that the proposed methodology represents an inexpensive alternative method to measuring the slip-ratio of a tyre.

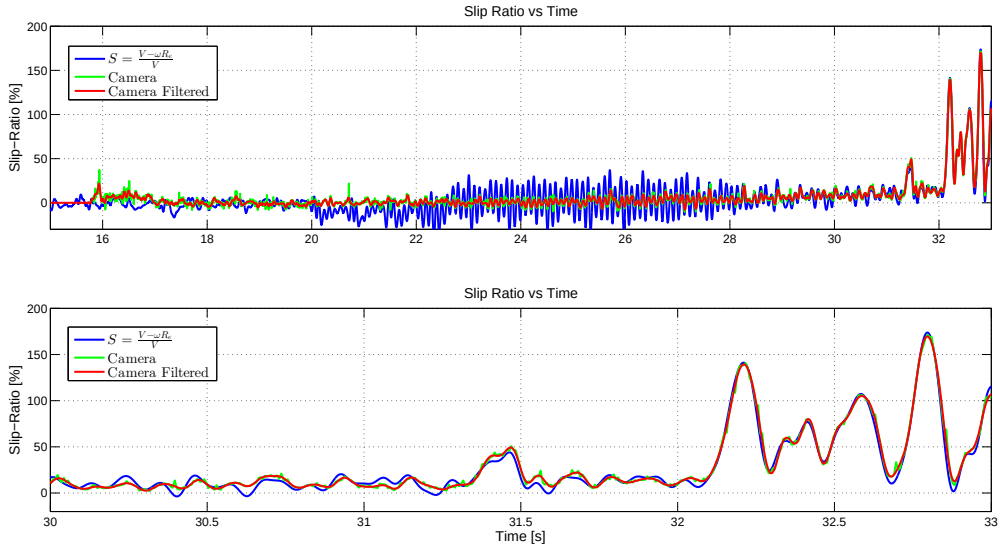


Figure 6.11: Slip-ratio comparison between the camera based method and the traditional method using equation 6.1

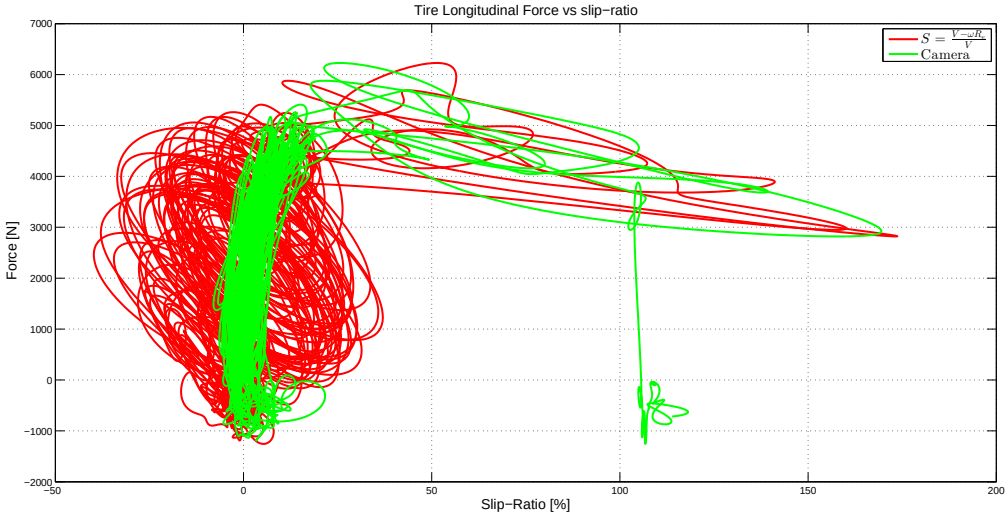


Figure 6.12: Longitudinal tyre force vs. slip-ratio comparison using slip-ratio based from equation 6.1 and measured using a camera

6.4 Future Work

Measuring of the rolling radius, tyre and road velocities has been shown to be possible on smooth land terrain. The following challenge is to test the system on rough terrain. The challenges on rough road are more complex since the tyre deformation is more complex. The tyre contact patch may have varying rolling radii depending on the tyre deformation as well as wheel camber. This can partly be solved by using cameras on either side of the tyre or even by placing a camera inside of the tyre. This would provide more and better information than the traditional method used to measure the slip ratio. Additionally radial lines can be made on the tyre to observe the deformation of the tyre while rolling. A stereographic system can also be used to determine the velocities in real world coordinates. This would remove the need to correct the parallax error caused in the single camera system. The method would also allow the effective rolling radius to be calculated using the translational contact patch velocity and wheel rotational velocity. This method could therefore be used to aid in the understanding the generation of longitudinal slip at the contact patch as well as other additional measurements. The drawback of the system is that the contact patch be visible to the camera. The system is therefore not applicable to soft soil terrain.

The method however obtains all measurements in units of *pixels*. For the purposes of measuring the slip-ratio this has been proved to be of no concern. However, if the measurements can be obtained in real world coordinates then much more useful information can be obtained. As stated this can be achieved using a stereographic rig. Additionally if a stereographic rig could be installed on the inside of the tyre which always faces the contact patch then the velocity profile of the whole contact patch both longitudinally and laterally can be measured. If the road velocity is determined at the same instance and it is assumed that the velocity of the inside of the tyre is the same as the outside of the tyre then the slip across the contact patch can be determined. While this method would neglect the tyre tread, the tread could perhaps be included if modelled correctly, it would allow measurements to be obtained which could never been obtained before.

6.5 Conclusion

In this chapter DIC was used to develop a method whereby the longitudinal slip ratio of a tyre can be measured. The method consists of an inexpensive camera mounted such that it captures images of the tyre -road interface. DIC techniques are used to track points on both the tyre and road at the contact interface. A clustering methodology is used to differentiate between points on the tyre and points on the road, as well as, remove outliers. The method also tracks the vertical height of the contact patch, thus effectively measuring the effective roll radius. This allows the method to compensate for parallax errors which are introduced as the tyre deflection changes. The method is shown to accurately estimate the slip-ratio when compared to traditional methods of estimating the longitudinal slip-ratio of a pneumatic tyre.

Chapter 7

Rough Terrain Profiling Using Digital Image Correlation

In Chapters 5 and 6 DIC techniques are used to measure certain operating conditions of a tyre such as the side-slip angle and longitudinal slip ratio. In this chapter the focus is on profiling the terrain over which the tyre travels.

Profiling of terrains is an important aspect of vehicle dynamics simulations especially over rough terrains. The accurate measurement of rough off-road terrains allows for more accurate multi body simulations. Many techniques exist for the measuring of road profiles inertial, mechanical or optical. The 3D profiling of terrains is often performed by utilising a line scan sensor which measures several points in the lateral direction of the terrain. The type of sensor used ranges from simple road following wheels, laser sensor to more complex LiDAR sensors. The sensor is moved longitudinally, often with a vehicle, while measuring the sensor's orientation and position. The individual measured points are then stitched together using the orientation and position of the sensor to obtain a full 3D point cloud terrain profile. The sensor's position and orientation needs to be accurately determined in order to combine the measured points to obtain an accurate representation of the measured profile. The position and orientation of the sensor is normally measured using an expensive Inertial Measurement Unit (IMU) or Inertial Navigation System (INS) with high sensitivity, low noise and low drift.

In this chapter a terrain profiling technique which utilises stereography, based on two inexpensive digital cameras, to obtain three-dimensional measurements of the terrain is described. The advantage

of this technique is that each measurement results in a surface containing lateral and longitudinal position as well as height, rather than a traditional line scan as in most systems. Corresponding camera images are taken so that they have an area which overlaps. The relative position and orientation between the individual surfaces can be obtained using either 2D or 3D features in the corresponding surfaces. This creates a system which negates the use of an expensive IMU system to determine orientation and position of each measurement. LiDAR sensors, while accurate, generate huge quantities of points which require sub sampling to reduce the data sets to usable data sets. The sub sampling is computationally expensive performing exhaustive searches to find neighbouring points. The camera technique presented here also requires sub-sampling. A sub sampling routine is presented which makes use of the structure of the data sets to significantly speed up sub sampling.

7.1 Introduction

The vertical dynamics of a vehicle is important for analysing the ride comfort and fatigue life of a vehicle. Computer aided simulation is extensively used to speed up and reduce the costs of vehicle development. The terrain input is the main source of vertical excitation of a vehicle [Andren, 2006] and therefore requires a realistic representation of terrain profiles. While some terrain profiles can easily be represented using standard road profile models [ISO 8608, 1995], extremely rough terrain surfaces common in in agricultural or off-road terrain cannot. The accurate measurement of real terrain surfaces not only provides a realistic representation of the desired terrain on which the vehicle will operate, but can also serve as validation between the measurement and simulation.

In terramechanics soil deformation is used not only to evaluate the performance of wheels in deformable soil [Shibly et al., 2005] but also used to improve soil models for simulation and prediction [Iagnemma et al., 2003]. Accurate measurement of the deformed soil can improve the development for off-road tyres. Improved off-road tyres would lead not only to improved tractive performance of the wheels but also decrease the rutting of soils. The rutting of soils generally increases the rolling resistance of tyres [Kurjenluoma et al., 2009] leading to a decrease in efficiency. The soil rut depth is also a useful measure, due to its simplicity, of the amount of soil compaction [Soane et al., 1981]. The soil compaction can lead to soil degradation by enhancing harmful physical, chemical and biological processes, which may affect the energy used in crop production [Soane and Van Ouwerkerk, 1995]. Anderson [Anderson et al., 2005] also uses the rutting of terrains to assess the level of vehicle impact

on the environment. Therefore, the development of a sensor to automatically and accurately measure soil deformation would lead to better energy efficiency not only in the vehicles themselves but also in the production of crops.

Numerous methods have been developed to measure road profiles. General Motors used a displacement transducer to measure the relative distance between the road and vehicle and an IMU using an accelerometer to measure the body displacement [Spangler and Kelly, 1966]. This system was used to measure the low wavelengths of a single track and highlights the need of the profiler to determine the absolute position of the vehicle body. Kern and Ferris [Kern and Ferris, 2007] developed a 3D version using a line scan laser measuring a lateral line across the road. The lateral lines are then pieced together to form a 3D surface. In order to piece the line scans together requires accurate measurement of the laser sensors orientation and position. This results in an expensive system using an expensive displacement transducer and a highly accurate INS. More cost effective systems replace the displacement transducer with cheaper alternatives such as cameras [Kertész et al., 2008] and or a road following wheel [Becker and Els, 2014]. However, all these systems still require the use of an expensive INS system to obtain orientation and position of the sensor.

This chapter discusses the development an inexpensive road profiling system where a 3D road profile is measured using photogrammetric methods based on two cameras. The cameras not only provide a full 3D surface with lateral, longitudinal and vertical coordinates with every measurement taken but also obtain a colour bitmap of the surface. The images are captured such that consecutive images overlap. The overlap allows the joining of 3D surfaces using both the colour bitmap and 3D surfaces. This method of joining negates the use of an accurate and expensive INS system, however, over longer distances accumulated errors may be large and the use of a slow sampling and inexpensive INS system is advised to remove drift. The low cost of the system may also enable profile measurement from any platform and may eventually be feasible for mass-production. This could allow the system to be installed on commercial off-road equipment.

7.2 Stereography

A stereographic rig, as described in section 3.5.2, is used along with the DIC correlation technique described in section 4.2 to determine the depth of the scene being viewed. For every pixel feature which is in view of both cameras an associated 3D coordinate can be obtained. Using two 640×480

resolution cameras a resolution of up to 640×480 can be obtained. In the set-up used, an average region contains 400×400 points that are in view of both cameras and for which therefore 3D points can be obtained. As noted in section 4.2.1, the spatial resolution between points depends on the focal lengths, resolution of the cameras as well as the distance from the surface to the cameras. The larger the area viewed the larger the spatial distance. The increase in viewing area however comes at a general loss of accuracy. This however can be compensated for by using higher resolution cameras. Higher resolution cameras come not only at a slightly higher monetary cost but a substantially higher computational cost as well. Thus, when measuring depth with stereographic cameras, compromises between area, resolution and computational cost need to be considered. The cameras were set-up so that the region captured was $250 \times 250mm$ with a spatial distance between points of about $0.6mm$. This region is large enough to encompass the contact patch of most SUVs and off-road vehicles.

7.3 Compilation of Surfaces

In most 3D profilometers a line scan of the surface is obtained and joined using an INS system. However, with the method presented here every sample taken with the cameras creates a 3D surface instead of a 2D line. Thus, these surfaces can be combined by simply having an overlap between consecutively sampled surfaces. The process of joining overlapping 3D surfaces is called registration. Three different registration techniques are applied to the individual surfaces to create a larger 3D surface.

The first registration technique uses the DIC correlation described in section 4.1.1 to track features in the left image from frame to frame. Once a set of 2D image correspondences between two images are obtained, with a relative change in motion between them, the 3D coordinates of these correspondences can be obtained as well. This yields several 3D points in space which have undergone rigid motion (rotation and translation) between two views of which the position is known in both views. The relative translation and rotation can then be determined provided at least three point correspondences have been obtained [Arun et al., 1987]. The problem being solved is the rotation matrix $\boldsymbol{R}$ and translation vector $\boldsymbol{t}$ in:

$$\boldsymbol{P}_2 = \boldsymbol{R}\boldsymbol{P}_1 + \boldsymbol{t} \tag{7.1}$$

This problem can be solved using the method described in section 5.4.1. This yields the relative orientation and translation between two consecutive frames. The rotation and translation is calculated at each sample with the point correspondences relative to the camera. Thus, each calculated rotation and translation presents the motion which occurred from sample one to sample two, relative to sample one. The rotation and translation matrices are combined into one transformation matrix in homogeneous coordinates as:

$$\boldsymbol{T}_{local} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{7.2}$$

where R_{ij} represents the component from the i^{th} row and j^{th} column in the rotation matrix $\boldsymbol{R}$ and t_x is the translation in the x-direction. The transformation of a point from sample one to sample two is then performed as:

$$\begin{bmatrix} P^i_{2x} \\ P^i_{2y} \\ P^i_{2z} \\ 1 \end{bmatrix} = \boldsymbol{T} \begin{bmatrix} P^i_{1x} \\ P^i_{1y} \\ P^i_{1z} \\ 1 \end{bmatrix} \tag{7.3}$$

where the point correspondence P^i_1 is appended by 1 to create a homogeneous coordinate. This multiplication performs both the rotation and translation in one single matrix multiplication. It also allows the easy accumulation of rotations and translations between two relative samples to determine the global position and orientation $\boldsymbol{T}_{total}$ of each sample's surface (i.e. relative to the very first sample) as:

$$\boldsymbol{T}_{total} = \boldsymbol{T}_{total}\boldsymbol{T}_{local} \tag{7.4}$$

This method is computationally fast compared to the other two registration techniques used later, which work directly on the 3D point clouds produced at each sample. As such this technique is also used to determine when the other registration techniques are to be used. This is accomplished by determining the distance travelled since the last time a 3D point cloud registration technique was

used. Once a certain distance has been travelled which would assure a certain percentage in overlap, the system triggers the use of an additional registration technique.

One of the additional registration techniques used is that of Iterative Closest Point (ICP). The ICP algorithm works on two point clouds, the previous point cloud n and the latest generated point cloud $n+1$. The ICP algorithm first determines the closest matching points between the two clouds for all points. This step is by far the most time consuming and modifications to the ICP algorithm are generally directed at speeding up the correspondence matching. The ICP algorithm proposed by Besl and McKay [Besl and McKay, 1992] calculates the distance from each point in the original cloud to all points in the transformed cloud to obtain the closest point. This process is time consuming as it performs an exhaustive search. More efficient variations are also available [Rusinkiewicz and Levoy, 2001]. After correspondences have been obtained the algorithm solves for the rigid transformation which reduces the distances between the matches. The process finds a local minimum and thus requires several iterations before convergence is reached. The ICP algorithm used is implemented in the Open Point Cloud Library (OpenPCL) [OpenPCL, 2014]. The transformation matrix returned by the ICP algorithm is then used to correct the total transformation matrix as:

$$\boldsymbol{T}_{total} = \boldsymbol{T}_{ICP}\boldsymbol{T}_{total} \tag{7.5}$$

The ICP algorithm works best if the point clouds are initially aligned to some degree. This initial alignment is performed by the first registration algorithm, however, if the point cloud cannot find matches a third registration technique is used to perform the initial alignment. This additional registration technique is very similar in approach to the first, with the exceptions that both key points and features are obtained from the spatial domain, i.e. from the 3D point cloud, rather than from the colour map of the raw images. Tombari *et al* [Tombari et al., 2013] evaluated the performance of various 3D key point detectors and found that Intrinsic Shape Signatures (ISS) key points [Zhong, 2009] provide good performance while being particularly efficient. The algorithm works by determining the covariance matrix of points surrounding a proposed key point. The eigenvalues of the covariance matrix is used to determine the saliency or uniqueness of the key point. If the ratios of the successive eigenvalues are below a certain threshold then the point is kept. The rationale is to avoid detecting key points at points which have a similar spread about the principal directions where a repeatable canonical reference cannot be established. The saliencies of the remaining points

are determined by the value of the smallest eigenvalue. This produces a method which is efficient in finding reasonable key points. After key points have been identified, features describing the region need to be determined. For this Point Feature Histograms are often used [Rusu et al., 2007]. Wahl *et al* [Wahl et al., 2003] describes four features that can be created from the distance vector between two points and their surface normal. The process is then to obtain these features between all points within a certain radius of the key point. Each feature is then divided into a histogram with a certain number of bins. Each bin contains the number of points within the region that falls within the bin range. The feature vector thus contains 4 histograms of n bins. The feature vectors can then be mapped using the Sample Consensus Initial Alignment (SAC-IA) [Rusu et al., 2009] method. In this method a subset of samples are drawn from the first image and a list of possible matches are obtained in the second image using a suitable metric, such as the Mahalanobis [Mahalanobis, 1936] or the Euclidean distance. A correspondence is chosen at random from the list and the rigid transformation is computed from the correspondences. The quality of the transform is then determined, the process is repeated and the best transformation returned. This transformation is then applied to the point cloud after which another iteration of ICP is used to refine the registration. This algorithm is outlined in a flow chart in Figure 7.1

With the registration techniques applied, multiple point cloud surfaces are obtained which overlap by about 50%. These point clouds require interpolation or sub sampling to obtain a final point cloud which averages over all point cloud surfaces. Detwieler and Ferris [Detweiler and Ferris, 2010] conducted a study on interpolation methods for 3D terrain surfaces. They concluded that the Inverse Distance to a Power method was the recommended method when fine resolution is required to describe localized roughness and that Kriging is ideal for identifying global trends of the surface.

7.3.1 Inverse Distance to a Power Method

The Inverse Distance to a Power method is a weighted average interpolator. With this method data points are weighted according to their distances from the node under consideration, with points closer receiving a higher weighting than those further away. The weighting function can be also be altered using a weighting power. This power controls how quickly the weighting changes with distance. The higher the power the less impact the points further away from the node has on the value of the node.

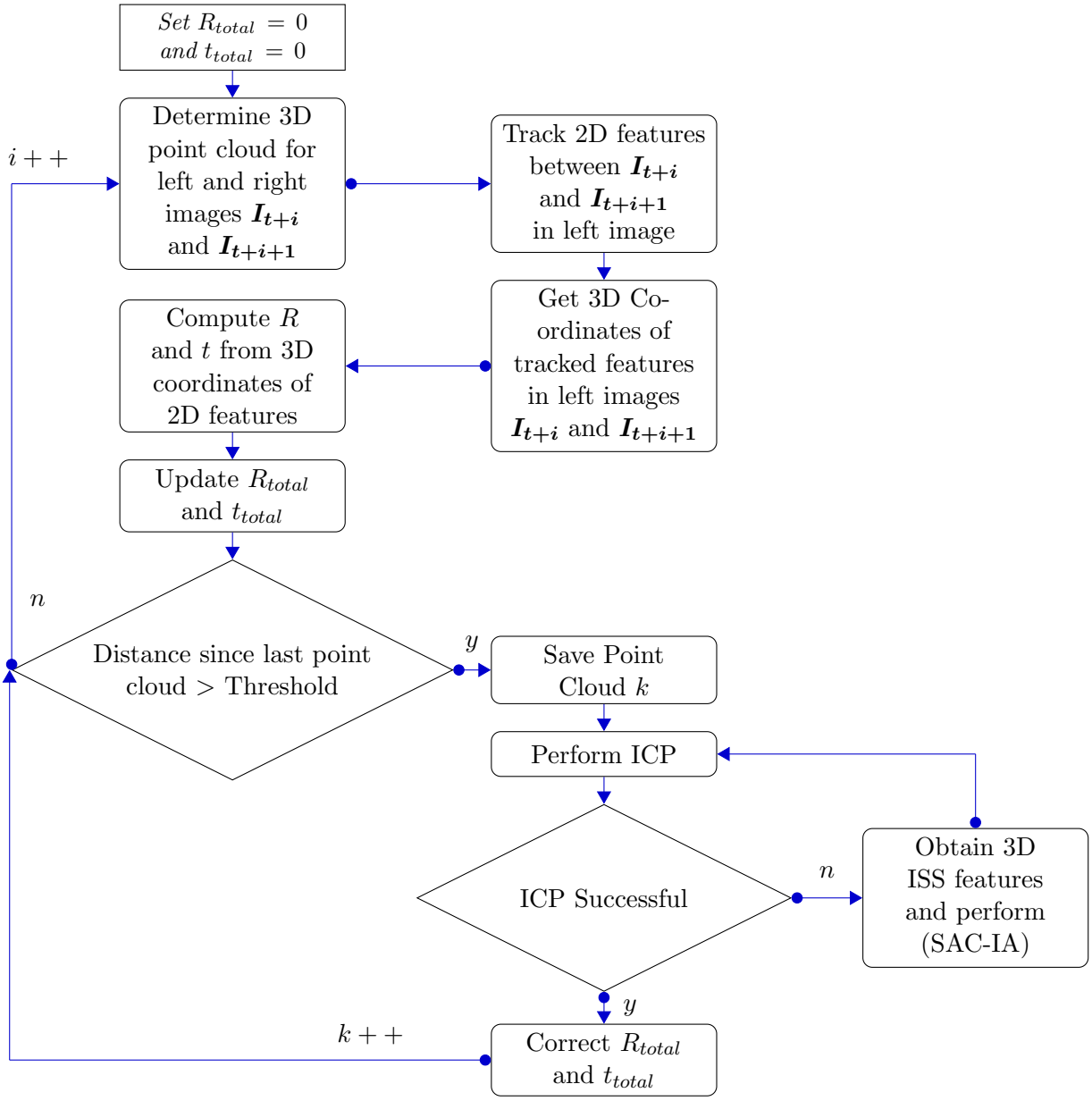


Figure 7.1: Flow chart for 3D terrain profiling algorithm

7.3.2 Kriging

Kriging is a geostatistical estimator method. It follows in principle the same methodology as the Inverse Distance to a Power method whereby the predicted value is the weighted average of surrounding points. It differs in how weights are assigned to surrounding points. While the Inverse Distance to a Power method is a deterministic interpolation method where weights are based directly

on the values of surrounding points, in Kriging a statistical model is fitted to surrounding points. Kriging thus assumes that the distance and direction of points reflects a spatial correlation that can describe variations in the surface. Kriging is therefore a useful interpolation tool which can capture global trends in data such as long wavelength sinusoidal undulations and other repeated patterns. Kriging is however a very computationally expensive algorithm which takes an order of magnitude longer than the Inverse Distance to a Power method to solve.

7.3.3 Sub-sampling Algorithm

Since most off-road terrains can be considered random, with the exception of washboard roads and bricked roads (Belgian paving), the Inverse Distance to a Power Method is used to perform the interpolation. The algorithm is also much less computationally expensive than Kriging thus reducing the interpolation time. The Inverse Distance to a Power Method equation is given as:

$$\bar{z}(\bar{\boldsymbol{x}}) = \sum_{i=1}^{N} \frac{w_i z_i(\boldsymbol{x})}{\sum_{i=1}^{N} w_i} \tag{7.6}$$

where w_i is the weighting given as:

$$w_i = \frac{1}{d(\bar{\boldsymbol{x}}, \boldsymbol{x})^P} \tag{7.7}$$

where $d(\bar{\boldsymbol{x}}, \boldsymbol{x})$ is the Euclidean distance between the point being interpolated $\bar{\boldsymbol{x}}$ and the point from the original point cloud $\boldsymbol{x}$. The domain of N points is taken as all points which lie within a certain radius from the interpolation point. Most often the points are obtained from an exhaustive search over all points to obtain those which lie within this radius. However, since the 3D surface obtained from the stereographic system has a certain structure, the domain of points used for interpolation can be obtained in a much faster way. The underlying structure of the surface, which is made up of a 3 matrices $\begin{bmatrix} X & Y & Z \end{bmatrix}$, is that the longitudinal (X) and lateral distances (Y) when mapped according to the row (i) and column (j) indices form a plane, as shown in Figure 7.2.

This results in a fairly consistent increase in longitudinal and lateral distances as the row and column indices are incremented. As a result of this a much faster interpolation strategy is proposed which does not perform an exhaustive search over all points in the point cloud. The process starts by determining the bounds of each surface $\begin{bmatrix} x_{min} & y_{min} & x_{max} & y_{max} \end{bmatrix}$. Next the incremental increase

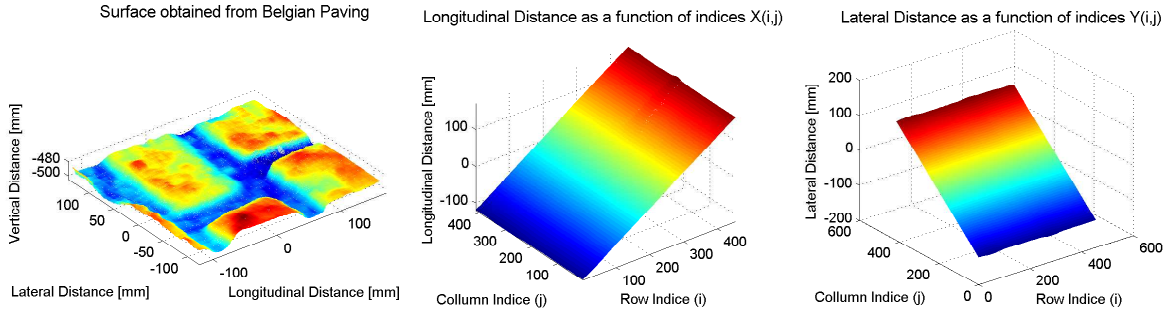


Figure 7.2: Measured surface. Longitudinal distance value as a function of index. Lateral distance value as a function of index

in longitudinal and lateral distance, as the row (i) and column (j) indices are incremented, are determined. These values are given as $\begin{bmatrix} x_{step_i} & x_{step_j} & y_{step_i} & y_{step_j} \end{bmatrix}$, where x_{step_i} is the increase in longitudinal distance as an increase in the row index (i). Therefore, each longitudinal and lateral value as a function of the matrix indices can be described as:

$$x_{i,j} \approx x_{0,0} + ix_{step_i} + jx_{step_j} \tag{7.8}$$

$$y_{i,j} \approx y_{0,0} + iy_{step_i} + jy_{step_j} \tag{7.9}$$

Where $x_{i,j}$ is the longitudinal position of point (i,j). If the equations are solved for in terms of indices (i,j), one obtains:

$$j \approx \frac{y_{0,0} - \frac{x_{0,0}}{x_{step_i}} y_{step_i}}{y_{step_j} - \frac{x_{step_j}}{x_{step_i}} y_{step_i}} \tag{7.10}$$

$$i \approx \frac{x_{0,0} - jx_{step_j}}{x_{step_i}} \tag{7.11}$$

This allows one to obtain the indices (i,j) given the longitudinal and lateral position (x,y) and knowing the start position $(x_{0,0}, y_{0,0})$ of a point cloud. Since this is only an approximation, the process can be repeated a few times (2-3 times) by simply changing equations 7.10 and 7.11 as:

$$\Delta j \approx \frac{\Delta y - \frac{\Delta x}{x_{step_i}} y_{step_i}}{y_{step_j} - \frac{x_{step_j}}{x_{step_i}} y_{step_i}} \tag{7.12}$$

$$\Delta i \approx \frac{\Delta x - j x_{step_j}}{x_{step_i}} \tag{7.13}$$

This should then converge to the region in the surface where the point being interpolated should lie. The interpolation is then performed with radii in the Δi and Δj direction obtained from equations 7.12 and 7.13 by substituting Δx and Δy by the search radius R. While the interpolation region is not strictly circular but rather rectangular this has a very small effect on the interpolated value especially with a large search radius R_i and R_j. Thus, the interpolation process can be summarized as:

- Find all surfaces in which the point (x, y) lie using the ranges $\begin{bmatrix} x_{min} & y_{min} & x_{max} & y_{max} \end{bmatrix}$ of the surfaces .
- Find the (i, j) indices where the point (x, y) lies within a surface using equations 7.10 to 7.13
- Find the search region in terms of indices using equations 7.12 and 7.13
- Interpolate over all surfaces using the Inverse Distance to a Power Method

This process works as long as the point (x, y) lies within at least one surface. If the point is found to lie outside all surfaces an extrapolation procedure is used. The extrapolation procedure assigns the value to the point of its nearest neighbour. To find the nearest neighbour the surfaces are ranked in terms of the distance from the point to the start of the point cloud surface $(x_{0,0}, y_{0,0})$. Starting at the closest surface. A boundary search is performed in that surface and the closest point retained, a boundary search is then computed for the next closest surface. This process is repeated until the point retained from the next surface is further away than the previous surface thus halting the process. While the interpolation process can be considered to be fairly fast the extrapolation process is still regarded as slow in comparison.

7.4 Experimental Results

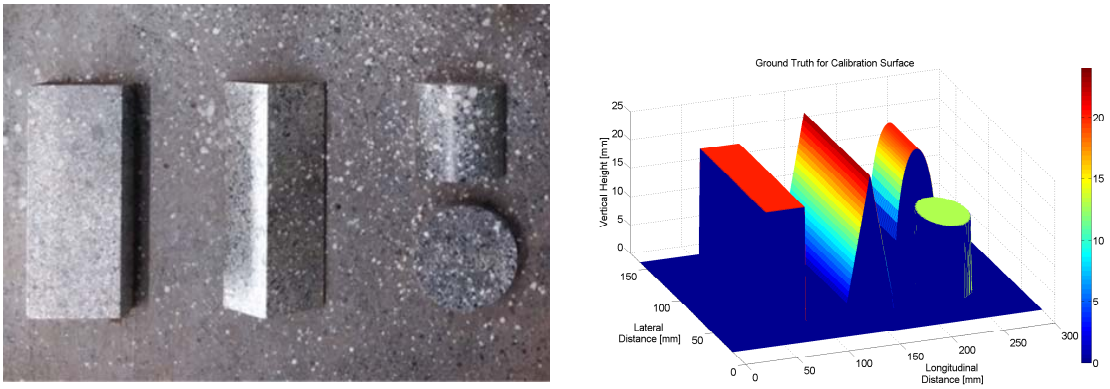
The initial experimental tests were calibration tests conducted in a laboratory to determine the accuracy in a controlled manner. Further tests where conducted over artificial terrains to determine the ability to profile roads with known disturbances. Finally real rough terrain profiles are used to evaluate the method.

7.4.1 Laboratory Calibration Tests

The accuracy of the stereovision, registration and interpolation procedures are evaluated using a small calibration surface. Smith and Ferris [Smith III and Ferris, 2010] suggest that a calibration surface should contain contours with sharp-edged high-bandwidth events. The surface thus consists not only of round contours but also contours with sudden changes in height or sharp edges. The surface is sprayed with a speckle pattern to aid the stereovison system in determining the depth. The calibration surface along with the ground truth of the surface is shown in Figure 7.3.

The cameras are Point Grey Chameleon USB 2.0 (see Appendix A for specifications) cameras with maximum resolution of 1296×964 and a global shutter. The lens is a Fujinon 6*mm* focal length c-mount lens (see Appendix A for specifications). The camera resolution is set fairly low at 640×480 to speed up computation. The surface is measured by moving the stereovision system across the system by hand. The final surface is a 1*mm* spaced interpolated surface obtained from 15 point cloud surfaces which are joined together. The final measured surface is shown in the first plot of Figure 7.4. The second plot shows the vertical distance error surface. The error is therefore merely the difference in vertical distance between the measured and ground truth. This error metric is fairly conservative as compared to an error distance which measures the distance to the nearest ground truth point. This plot indicates that large errors exists at objects which have sharp edges. This is partly due to the metric used as well as that these regions may never have been seen by both cameras as the cameras were still pointed down vertically. The problem of hidden regions is a problem with most measurement systems especially optically based systems. It can generally be corrected by viewing the regions at different angles which would enable the system to then measure these regions. While the system easily allows this to be performed, this method is however not considered in this study. In this chapter most measurements are made vertical to the surface. Thus, at these sharp

edges the surface is not measured and therefore these regions are interpolated therefore causing large errors. The RMS error over the whole error surface is 2.4*mm*, however this is mostly dominated by the larger errors at sharp edges. The last plot shows a sectional view of the profiles as well as the error surface. The plot clearly shows the error at the sharp edges. However, at the smoother areas the errors are much smaller staying mostly below 1*mm*. This indicates that the system slightly smooths sharp edges. In vehicle dynamics applications sharp edges are not of significant concern since the tyre itself smooths sharp edges to a greater extent than the stereographic system. The measurement system accuracy can still be improved with higher resolution settings as well as changing the camera geometry, this is however not performed in this study. The system is deemed to have satisfactory result for the purpose of terrain measurements.



(a) Calibration surface (b) Ground truth of calibration surface

Figure 7.3: Calibration surface

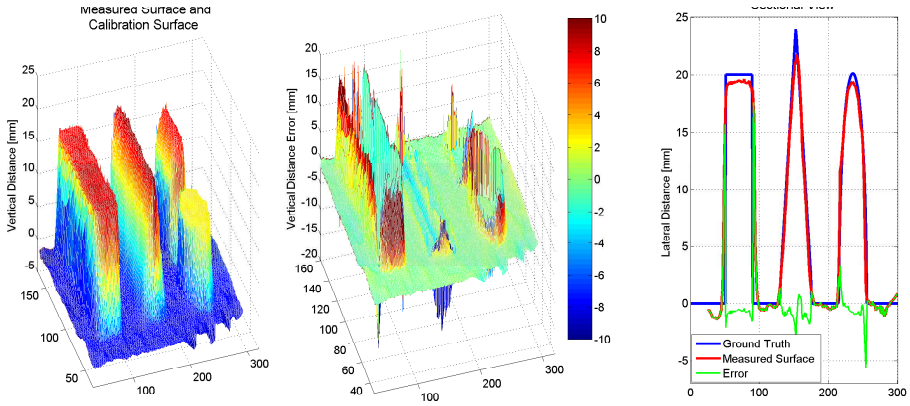


Figure 7.4: Final interpolated surface. Error surface of measured surface. Sectional view

7.4.2 Real Road Tests

The stereographic system was experimentally tested on real road surfaces on the suspension track of the Gerotek Test Facilities [Gerotek Test Facilities, 2014] The system was attached to another road profiler as described by [Becker and Els, 2014]. The system described in [Becker and Els, 2014] consists of a beam with three wheels and multiple road following arms attached to potentiometers (The road profiler is shown in Figure 7.5). As the arms move over the surface the change in angle is measured which is related to the height of the arm. This allows the height to be calculated at each arm relative to the beam. The system is also equipped with a Novatel SPAN-CPT GNSS+INS receiver [Novatel, 2014] which allows the measured motion relative to the beam to be transformed into motion relative to a global reference. The system therefore allows the mapping of a terrain in global coordinates with a resolution of 40mm lateral and 6mm longitudinal.

Figure 7.5: CanCan machine using road wheel followers and INS system

The profiler with the camera system was tested over multiple surfaces which introduce various forms of excitations to the system. Due to the design of the arm based system, it is referred to as the CanCan machine in this study. The experimental set-up along with some of the road profiles used, such as sinusoidal profiles, is shown in Figure 7.6. The camera was mounted at a height of $500mm$ this allowed the camera to measure an individual surface of dimensions $260 \times 260mm$ containing a grid of 400×400 points, with a spacing of about $0.65mm$ between points.

Figure 7.6: Sinusoidal profile. Experimental setup

Flat Surface

A flat $4m$ long section was profiled first, the final cloud is interpolated from 38 individual surfaces. It was found that due to insufficient calibration that the measured surface contained drift. The insufficient calibration causes a curvature error in the individual surfaces captured. The combined surfaces thus gradually cause the final surface to drift. The drift is also not linear as the curvature error exponentially increases the vertical height drift. As a result the drift is removed by means of detrending. The detrending is achieved by fitting a 2nd degree polynomial through the surface along the longitudinal distance. This polynomial, which captures the mean of the surface, is subtracted from the original surface. This yields a surface from which low frequency drift is removed without affecting the higher frequencies. Figure 7.7 shows the surface profile as well as the detrented surface. The detrented surface removes almost all drift from the surface. In order to evaluate the camera based profiler the camera measurements such as roll, pitch and vertical height are detrented as well using a single degree polynomial.

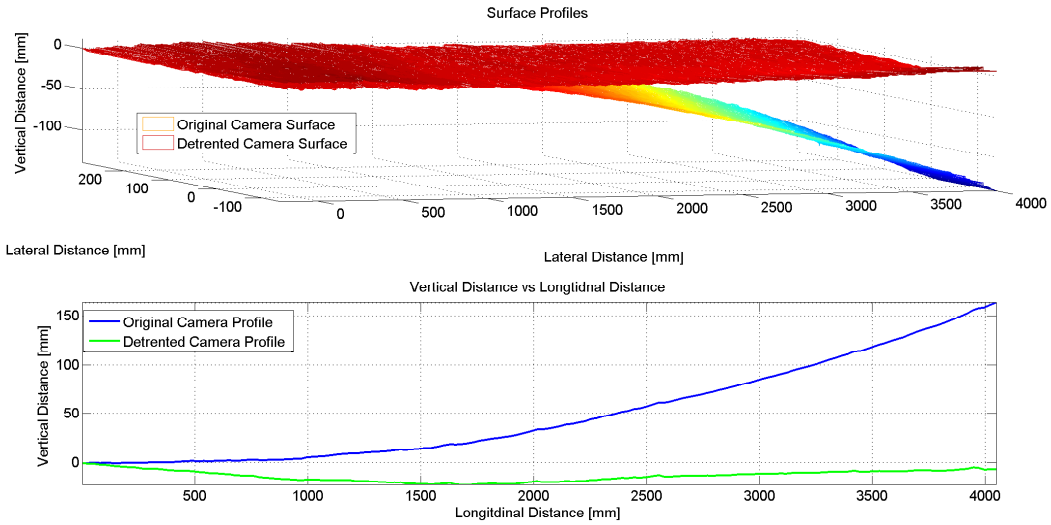


Figure 7.7: Measured surface and vertical displacement measuring a flat surface

Roll Excitation

The profilometer was excited in the roll direction by letting the left wheel follow a sinusoidal vertical motion while the two right wheels remained at a constant vertical height. The vertical profile followed by the left wheel is a $60mm$ amplitude sinusoid with a $1800mm$ wavelength cast in concrete, as shown in Figure 7.6 (left). The stereographic system thus experienced a roll motion while still profiling a flat surface. The roll and pitch angles experienced by the profilometer are used to evaluate the ability of the stereographic profiling system to remove the roll excitation from the profile.

The baseline measurements are obtained from the geometry of the problem, knowing the dimensions of the CanCan machine as well as the profile which each wheel followed. This allows for exact determination of the roll and pitch angles as well as vertical height of the sensors. In the case of the roll excitation the roll angle is additionally determined by road profiling lasers. This is achieved by placing two road profiling lasers a certain distance apart on the rear wheels. From the known and measured displacements the roll angle can be determined from basic trigonometry. This method is not used in the other excitations since the lasers could not measure a flat reference surface. The baselines are used to compare the measured angles from firstly the Novatel-INS which uses high accuracy fibre optic gyroscopes to determine the roll, pitch and yaw angles and a GPS system combined with accelerations to determine the lateral, longitudinal and vertical position. Secondly the detrented

measurements from the camera based system are used for comparison. The camera measurements are obtained by integrating the relative velocity measurements obtained between frames. Two data sets are used in this comparison. The difference in the data sets are mostly the accuracy obtained from the Novatel-INS. This is done in order to demonstrate the varying accuracy which can be expected from the GPS components of an INS system.

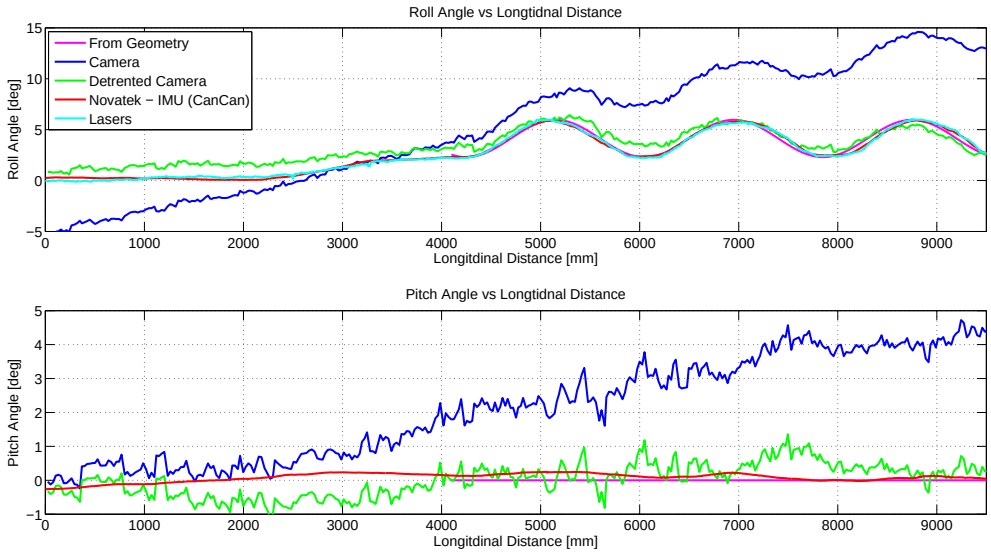


Figure 7.8: Measured roll and pitch angle with roll excitation, best INS measurement

The first plot in Figure 7.8 shows the measurements obtained from the best data set. The figure shows that the Novatel-INS measures the roll and pitch angles perfectly, following the ideal and laser measurements very closely. The detrented camera measurements also compare well to that of the baseline measurements. The camera measurements have the correct amplitude and frequency, however showing slightly more noise than the Novatel-INS. The noise amplitude is comparable to most MicroElectroMechanical Systems (MEMS) based INS systems. This demonstrates that the system successfully measures the roll excitation.

Figure 7.9 shows the same plots for the worst INS measurement. The plots indicate that while the INS system obtains the correct amplitudes of the angles the longitudinal displacement is incorrect. This is to be expected since the angles are obtained using high accuracy fibre optic gyroscopes. However, the longitudinal position is obtained primarily from GPS, with some filtering using accelerometers. GPS measurement accuracy is highly dependent on the satellite constellation and

varies during the day. A jump in position can also occur when a satellite is lost or a new one is acquired. This results in GPS accuracy varying during testing causing repeatability problems. Even though the angles are measured accurately, most road profilers obtain the positions from the INS system and an error in the positions will drastically affect the final surface obtained.

It is possible to improve the INS system accuracy significantly by utilising a GPS base station or Real Time Kinematic (RTK) GPS. The base station uses a stationary measurement point to improve and remove unnecessary jumps in GPS measurements. Use of such a system is expected to greatly improve the accuracy as well as consistency of the GPS. In comparison the results from the camera profiler are unaffected and remain consistent. This indicates that this system is very repeatable with few external factors affecting the accuracy. In Figure 7.10 the profiled surface with roll excitation is shown, the profile is, as required, fairly flat with at most 15mm deviation from zero over a distance of 11m, indicating almost complete removal of the roll excitation.

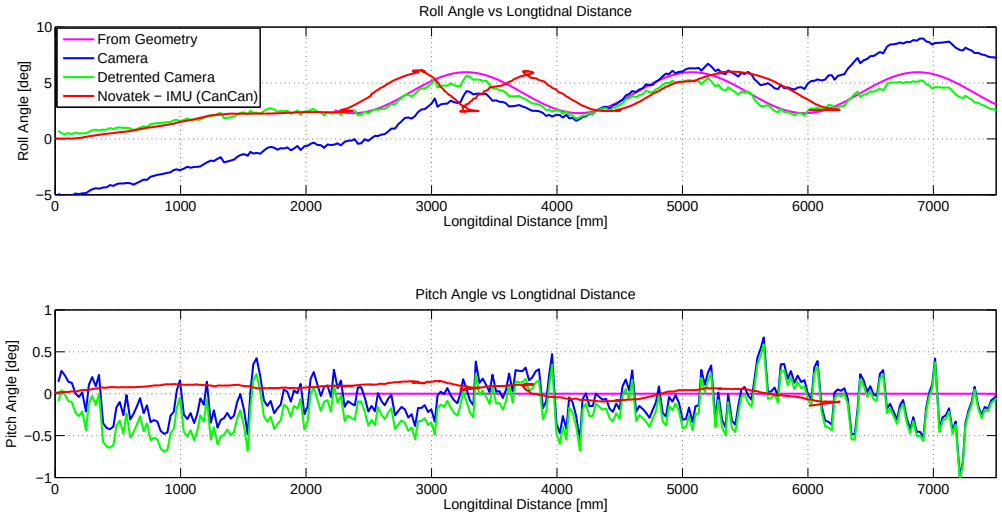


Figure 7.9: Measured roll and pitch angle with roll excitation, worst INS measurement

Pitch Excitation

The system was next excited in the pitch direction. This was achieved by having all the wheels of the profiler move across a sinusoidal profile. This not only induces a pitch excitation but also a vertical motion of the measuring systems as well. The profiler also traverses the road while profiling the

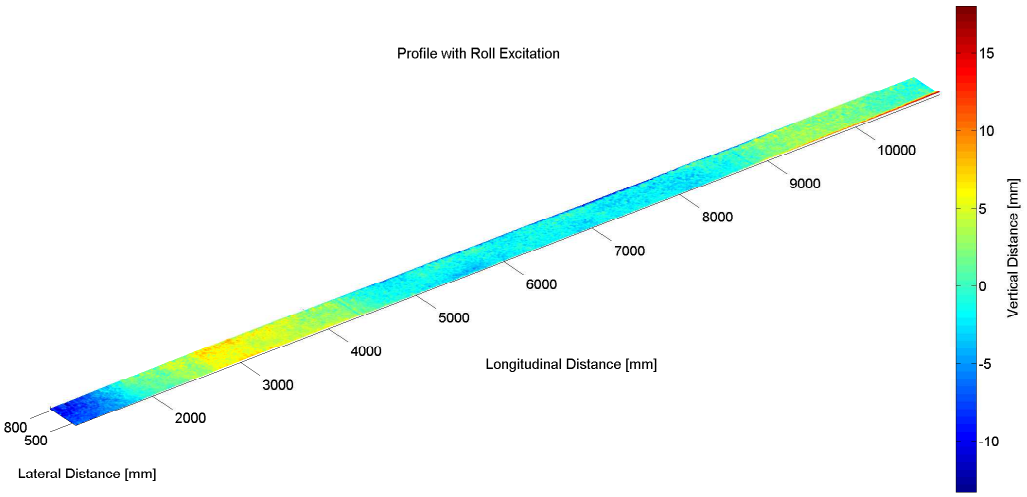


Figure 7.10: Measured flat surface with roll excitation

same sinusoidal profile. The plots in Figure 7.11 show that the pitch angles of both the camera and INS compare excellently to the baseline measurements. It should be noted that the baseline profile only starts around the 5000*mm* mark, the section in front of this is an unknown profile. However, the camera and INS system still provide similar responses in this unknown region. The INS angle however has a slightly longer wavelength which results in the profiles being out of phase at the end of the test. This again shows the effect which can be induced by the small errors of the GPS system. While the camera measurements again contain more noise the measurements compare well to the baseline profile with similar amplitudes and the correct wavelength. The vertical height plot shows that the INS system also contains considerable drift, this is especially true in the vertical direction due to the GPS having unsatisfactory vertical height accuracy. This can adversely affect the profiled terrain which relies heavily on all measurements from the INS system. Both systems measure similar amplitudes with the camera system measuring the correct wavelength compared to the INS system. Overall the camera vertical height measurement provides better results compared to the INS system which is noisier and contains random jumps.

Figure 7.12 shows the final profiled road. The 3D sinusoidal profile is shown in the first plot. The edges of the profile are affected by the extrapolation process and should therefore be disregarded. A sectional plot of the profile from the camera based system is compared to the baseline profile and the

CanCan machine is given in the bottom plot. This shows that the camera system provided better results than the CanCan system which was highly affected by the vertical height drift in the INS measurements. The amplitude obtained from the camera system is around $100mm$ which is a 16% error on the $120mm$ profile. However, it should be noted that this is still better than the INS system. The accuracy of the camera system could be improved by increasing the size of an individual surface from $300 \times 300mm$ to a size which is comparable to the wavelength being measured. This however will result in a loss of measurement resolution if the camera resolution is not changed accordingly.

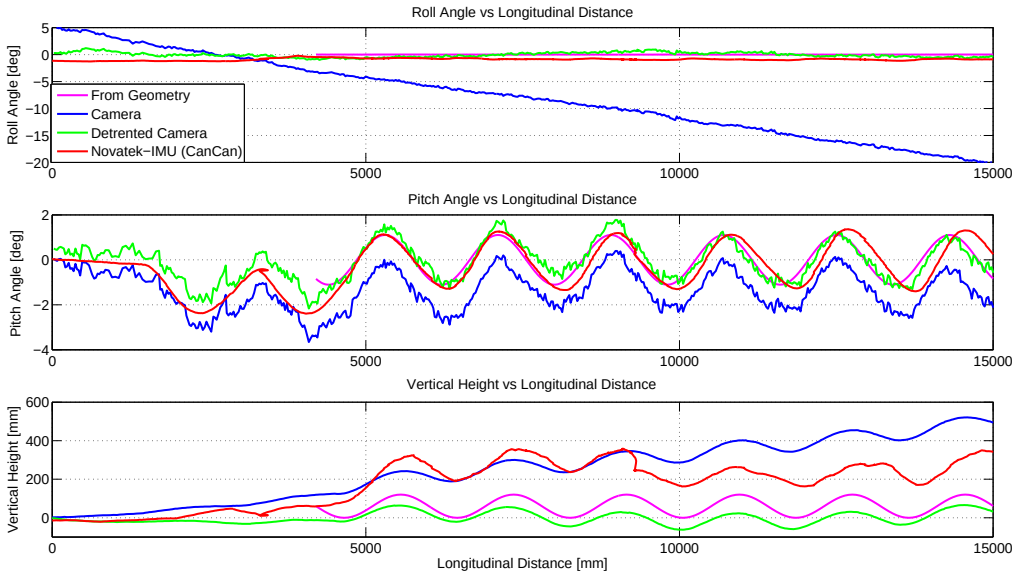


Figure 7.11: Measured roll, pitch and vertical height with pitch and vertical excitation

Roll and Pitch Excitation

In this test the road profiler underwent a combined roll and pitch excitation by having the left wheel follow a sinusoidal profile which is $180deg$ out of phase with the right wheels. This induces both a roll and pitch excitation at the same time while the camera system profiled a flat section of the track. Figure 7.13 shows the angles obtained from the camera as well as the INS. The angles from both systems follow the baseline profile well with the INS obtaining a slightly longer wavelength of the sinusoidal profile. The profiled road in Figure 7.14 shows that the profiled road remains flat, deviating only up to a maximum of $20mm$. The camera system therefore successfully removes the roll and pitch excitations.

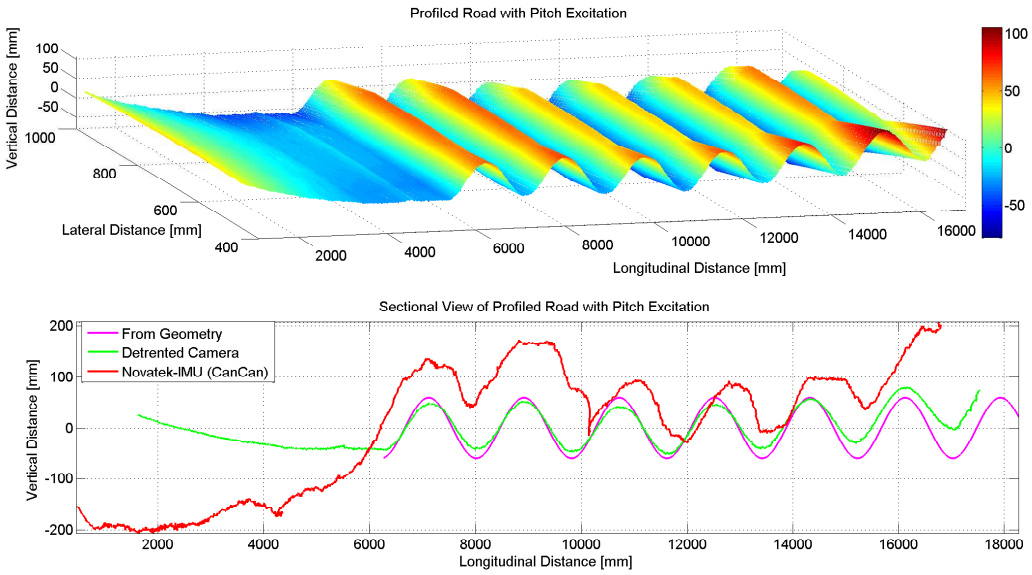


Figure 7.12: Measured 3D surface with sectional view of sinusoidal profile with pitch and vertical excitation

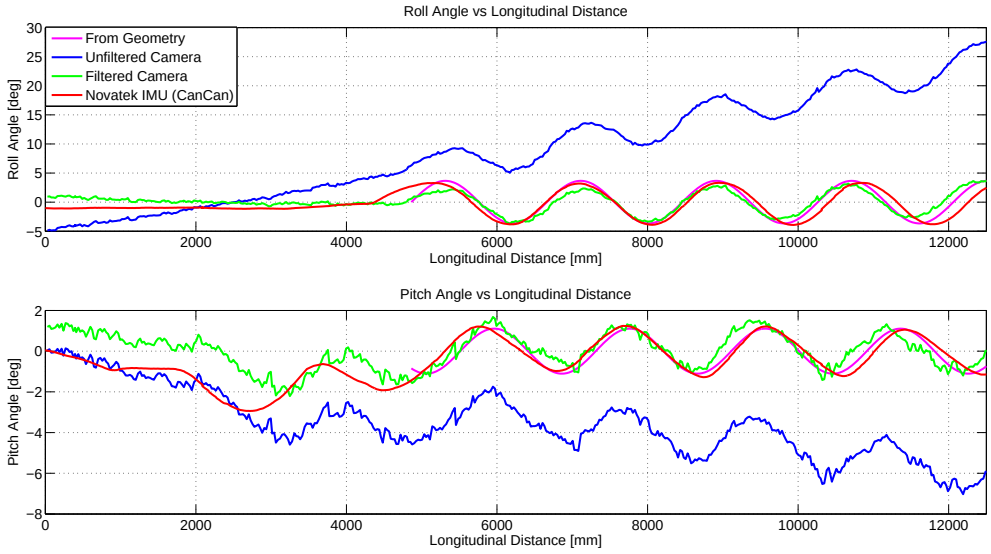


Figure 7.13: Measured roll and pitch angles with roll, pitch and vertical excitation

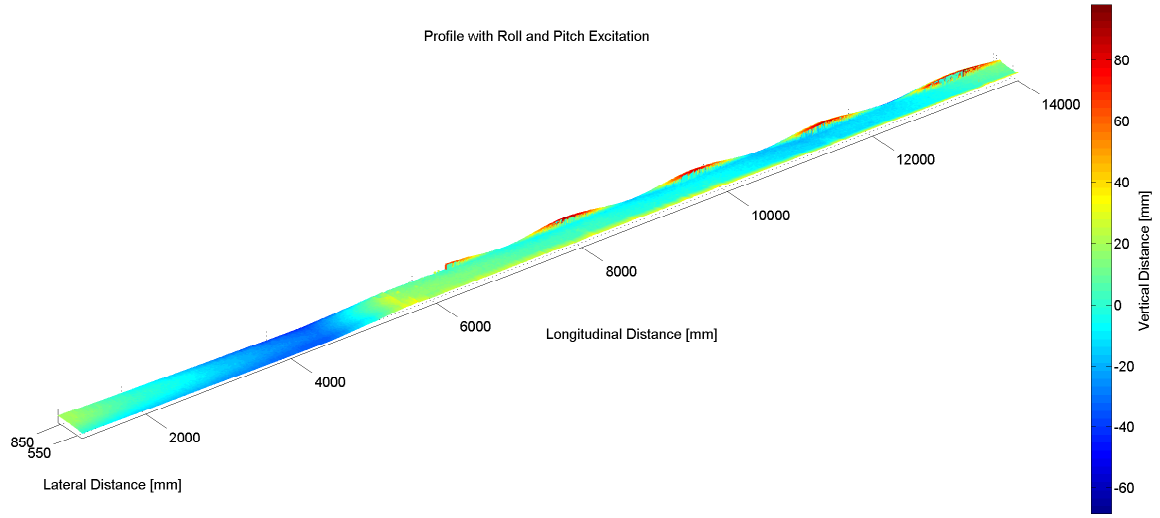


Figure 7.14: Measured flat surface with roll, pitch and vertical excitation

Belgian Paving

A section of Belgian paving was profiled which contains much higher frequencies as compared to the previous terrains which were either flat or low frequency sinusoidal profiles. A image of the Belgian paving is shown in Figure 7.15, this figure gives an indication of the roughness and high frequency nature of the terrain. The profile obtained using the cameras is shown in Figure 7.16. The figure shows a $10m$ section of Belgian paving superimposed on a low frequency profile. A zoomed in section of the profile shows the resolution that can be obtained from the camera system, showing the bricks as well as spaces between the bricks very clearly. This method obtains better results than the CanCan machine which uses a road following wheel (stylus) with circular tip to measure the profile. The spatial resolution of the system tested can be easily improved by increasing the camera resolution without affecting the area measured. The diameter of the road following wheel effectively filters the profile depending on the radius of the stylus tip. A Power Spectral Density of the profile is given in the bottom plot. This plot shows that the camera method effectively captures the two wavelengths at $7.32 cycles/m$ as well as $14.65 cycles/m$ corresponding to the average brick and spacing dimensions of $136mm$ and $68mm$. While the CanCan method does indicate that there is similar energies the peaks are much less pronounced and at lower spatial frequencies. The reason for the smaller peaks is partly

due to the filtering effect but also due to the short section recorded at with a higher sampling spacing than the camera. The peaks are found at lower spatial frequencies since the stylus radius affects the measured profile at lower spatial frequencies indicating that the profiles are larger than actual. The study in [Becker and Els, 2014] found that the system can successfully capture the spatial frequencies corresponding to the brick and spacing frequencies if the stylus radius is taken into consideration.

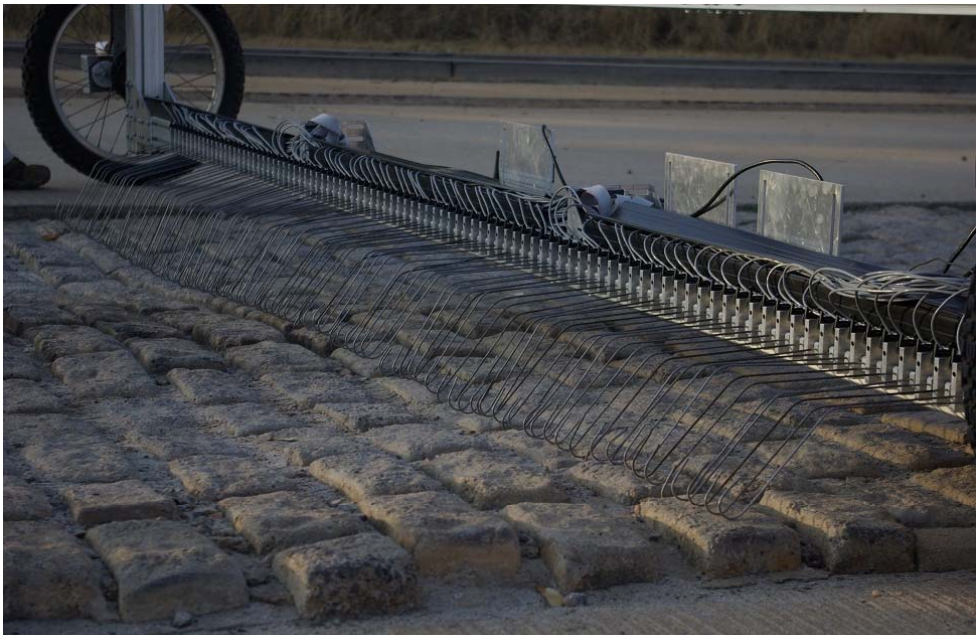


Figure 7.15: Belgian paving at Gerotek Test Facilities

An additional advantage of the system is that it uses cameras to profile the terrain. Therefore the colour information of the terrain is obtained as well. This is useful in simulation environment to create a terrain profiles which not only represents the actual terrain but have the correct visual appearance as well. This is useful in marketing videos or for driver simulators. While it is possible to create point clouds which have the correct colour, the process requires an additional registration technique [Mastin et al., 2009]. With this system the real colour of each 3D point is obtained with very little effort, requiring no additional computational power. An illustration of a 3D terrain profile where the colour is based on the height of the pixel as well as the Grayscale colour of the pixel, as obtained from the camera, is shown in Figure 7.17.

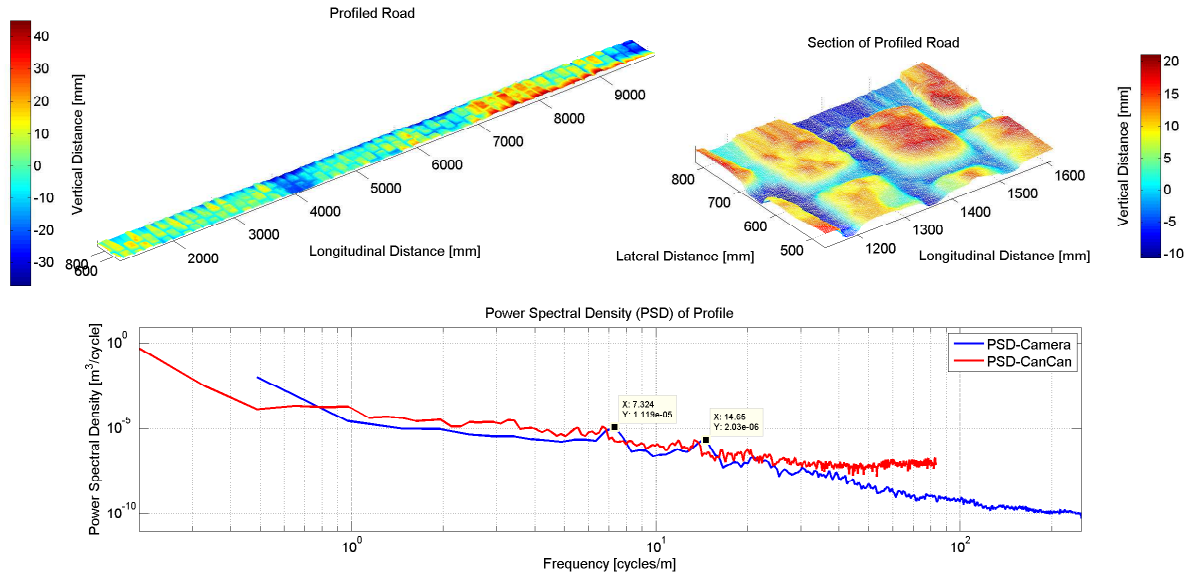
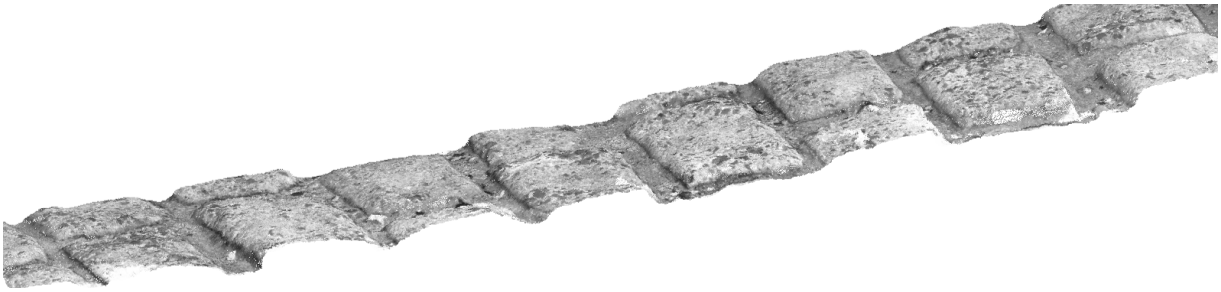
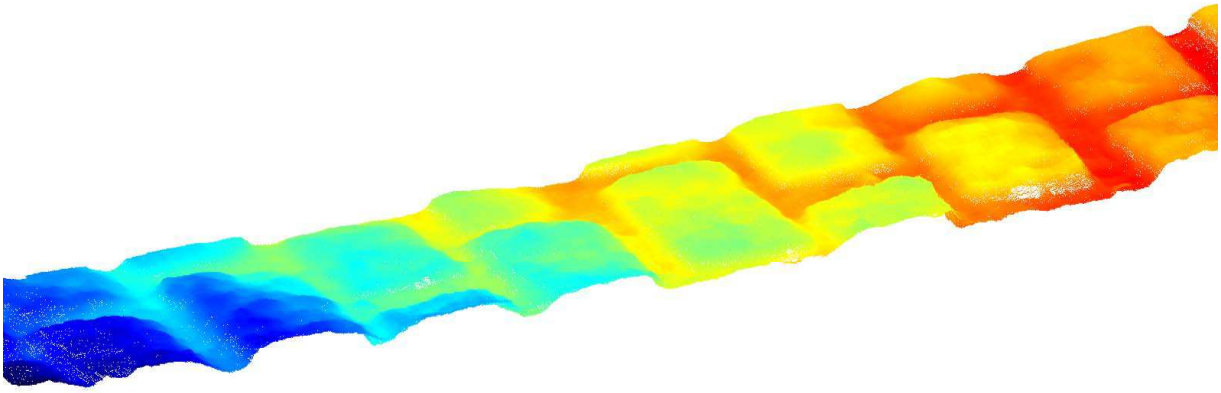


Figure 7.16: Measured 3D surface of Belgian paving. Zoomed in section of Belgian paving. Power spectral density of Belgian paving with proposed method and CanCan machine



(a) Grayscale point cloud



(b) Height coloured point cloud

Figure 7.17: Point clouds with camera obtained colour spectrum or height based colour

7.5 Drift Removal

Since the registration methods use frames relative to one another to determine the relative motion it is impossible to obtain drift free estimates. While the drift can be reduced by proper calibration of the cameras a small amount of drift from noise and errors in calibration will lead to drift in the measurement over time. Possible solutions to this would be to use another set of cameras which always views fixed markers at some distance away. This would allow all frame motions to be corrected by means of motion determined from the same fixed points. The accuracy of this method is the largest concern as stereographic measurement accuracy deteriorates at greater distances as illustrated in section 4.2.1. Another alternative which is frequently employed is to use an IMU/INS system to remove drift. Multiple algorithms can be employed to fuse the camera measurement with that of the IMU/INS [Julier and Uhlmann, 2005], [Zhang et al., 2005], [Li and Mourikis, 2013]. These algorithms make use of Kalman filters to fuse the different INS and camera measurements together. While an IMU/INS is required to remove drift the requirement of the system is much lower than required for most road profilers. The system accuracy need not be very high, neither does it require a high sampling frequency. Infrequent measurements would be enough to reduce the overall drift of the system. However if the system is equipped with GPS it would lend itself to the same problems that a conventional profiler has with dropped satellites, jumps in position etc. This may necessitate the development of another means of removing drift for road profiling.

7.6 Future Work

The developed system opens up possibilities in many areas of vehicle dynamics and terramechanics. The system provides an inexpensive technique to measure the 3D profile of a terrain. As a result of the inexpensive nature and customisability of these cameras, multiple cameras can be added to vehicles. This will allow for example, in terramechanics not only the measurement of the rut depth before and after every wheel but also to obtain full 3D surface of the whole terrain before and after every wheel. This will allow for better development of off-road vehicle wheels with better traction and efficiency while providing a smaller impact on the environment.

The system also has the potential of being applied in real time by firstly utilising faster cameras to obtain frames at higher speed. The 3D depth and initial alignment phases can be performed in

real time. Using Field Programmable Gate Array (FPGA) has yielded systems capable of performing the 3D depth mapping using more advanced algorithms, than presented here, at over 500*fps* [Georgoulas and Andreadis, 2009]. Thus, this system is capable of measuring rotational and translational rates at over 500*fps*. This can not only be used to measure terrain at higher vehicle speeds but also be used in other vehicle dynamics applications. The translational and rotational rates as well as positions and angles can be used to complement or provide a cheaper alternative to INS systems.

7.7 Conclusion

A road profiler system which provides automatic continuous measurement of road profiles by making use of two cameras in a stereographic system was presented in this chapter. The method of determining 3D profiles from a set of cameras is described. The camera motion is tracked using 2D image features along with their 3D coordinates yielding a fast camera pose technique. Once a certain distance of camera motion is reached a point cloud is captured which is aligned with the previous point cloud using 3D registration techniques. This yields multiple overlapping 3D point clouds. A novel interpolation strategy is presented which uses characteristics of the captured point clouds to increase the speed of obtaining the interpolation region of an Inverse Distance to a Power method. The presented method thus yields a system of which a regularised grid of the surface can be obtained at any resolution at a high computational efficiency.

The accuracy of the system is excellent, at the most basic setting, and can be improved by means of better cameras and lenses as well as changing the geometry of the two cameras with respect to each other as well as with respect to the road surface. The measurements obtained from the system, although containing drift, provides better accuracy and consistency than the GPS system employed by the INS system. While the GPS measurements can be improved by means of a RTK GPS, the overall positional accuracy is still questionable in providing millimetre accuracy required by road profilers. The camera system is shown to be able to remove all three translational and rotational disturbances from the system while allowing these disturbances to be measured as well. The developed system does not require an IMU/INS to combine multiple sensor readings as opposed to most methods. However, the inclusion of a low cost IMU/INS will help reduce some of the drift in the system.

The system also has the potential to vastly improve the field of terramechanics by providing an inexpensive sensor which can measure full 3D surfaces of the undeformed and deformed soil being

worked by a vehicle. The system also has the potential to be implemented in real time to not only profile terrain at higher speeds but also to measure translation and rotational rates, positions and orientations.

Conclusions and Recommendations

Conclusions

The application of Digital Image Correlation (DIC) towards solving vehicle dynamic related problems has been proven successful. DIC is successfully implemented to solve three problems often faced in vehicle dynamics. The solutions are compared against existing methods used to address the same problem. The DIC techniques proved to provide better results than the current methods employed especially when cost is taken into account. The extent to which DIC applications can be used in vehicle dynamics is unknown with a multitude of other possible application waiting to be discovered.

This study discussed the basics of digital imaging, covering the basic knowledge needed to obtain clear images which can be used in DIC. The effect of the aperture, exposure time and sensor sensitivity on the image exposure is discussed. A compromise between these settings needs to be attained since it is desirable to have all of these parameters as low as possible to increase the depth of field, minimise the effect of motion blur and reduce the noise on the image respectively. However, without artificial or sufficient passive lighting a compromise between these settings is desired to obtain images which are still sufficiently illuminated.

The models used to mathematically model cameras are also discussed. These models are used in the calibration of camera systems to obtain the intrinsic camera parameters. The intrinsic camera parameters model the projection of real world points onto points on the imaging plane. The modelling of lens distortion is used to obtain undistorted images of a scene which can be used to accurately track motion in the images. The calibration of multiple camera systems which involve solving for the extrinsic camera parameters, relative orientation and translation between cameras, is also discussed. The intrinsic, extrinsic and distortion coefficients are solved using multiple views of a planar surface

with known points. The process is performed using the OpenCV library.

A detailed discussion of the DIC algorithms used in this study a provided in Chapter 4. The chapter discusses the algorithms used to track discrete features using two different algorithms. The first algorithm finds corner features in an initial image, which are tracked to the following image using an optical flow technique. The optical flow technique makes certain assumptions and solves for the motion direction using image intensity gradients as well as image temporal gradients. The exact feature location is obtained using a correlation function on a window surrounding the feature. This tracking algorithm is computationally inexpensive and is the initial tracking method used. The second tracking algorithm is often more robust but computationally more expensive. The algorithm detects key points in both images and extract features from these locations. The features are then mapped to the corresponding features in the opposite image using a suitable metric. A DIC technique used to obtain the full image disparity in a stereographic rig in order to map pixels to a 3D coordinate is also discussed. This creates a system of which the 3D coordinates of the full scene can be obtained, enabling 3D coordinates of features to be tracked in real world coordinates.

DIC is used to measure the side-slip angle of a vehicle or tyre. Three different algorithms are developed. The first algorithm is a fast simple planar algorithm which measures the side-slip angle only. The second algorithm uses a single camera and a pose algorithm to measure all scaled translational velocities and rotational velocities. The Pose method is computationally expensive to use and only obtains the translational velocities in scale. The technique has however no additional requirements compared to the planar technique. The computational cost and scaled translational velocities of the Pose method resulted in the use of a stereographic rig, with two cameras, to obtain all translational and rotational velocities in real world coordinates. This algorithm, with the aid of special hardware, can be realised in a real time system capable of a few hundred frames per second. The last two algorithms were developed to reduce the effect of other motions on the measured slip angle when travelling over rough terrain. All algorithms provided a smaller noise band compared to a commercial side-slip angle sensor, especially in low speed applications. A comparative study between two different mounting locations, close to the CG and at the wheel, is also performed. This study indicated that the sensor should be mounted as close as possible to the CG to reduce the effect of other vehicle motions on the measured side-slip angle.

The application of DIC was also used to measure the longitudinal slip ratio of a pneumatic tyre. The application involves mounting a single camera observing the contact interface encapsulating both

the tyre and road. Features which are in close proximity to the contact interface are tracked and classified into three groups (point on the tyre, road and outliers) using clustering. From these groups the velocities of the tyre and road are determined. A method is used to compensate for the parallax of the camera thereby accurately obtaining the velocity of the road. The method is compared to the traditional means used to measure the slip ratio and shows excellent ability to measure the slip ratio. The method measures the slip ratio directly at the contact patch and therefore is not susceptible to any slip between the tyre and rim or the vertical and rotational deformation of the tyre. The system could provide better insight in the longitudinal force generation mechanism of a pneumatic tyre since the slip is measured directly at the tyre-road interface.

A calibrated stereographic rig is used to create a road profiling device. The road profiling device differs from conventional road profiling devices in that a single measurement results in a grid of points that are spaced both laterally and longitudinal as opposed to just laterally. This means that a single sample consists of a full surface rather than a line scan. In traditional systems the line scans are joined using an expensive inertial navigation system. The accuracy of the inertial navigation system is shown to be vital to the measurements system. In the proposed method, overlapping surfaces are sampled which are joined by tracking features in the overlapping surfaces. This initially aligns the point clouds which are corrected using the iterative closest point algorithm. The proposed system contained drift due to insufficient calibration of the cameras. The low frequency drift is removed using detrending. This removes the low frequency drift without affecting the higher frequencies which enables the system to be compared to a baseline measurement and a inertial navigation system. The system showed remarkable consistency and accuracy in measuring the disturbances to the sensor while accurately profiling the high frequency profiles in the terrain. The system reduced the requirement of an absolute positioning system enabling a low cost and slow sampling GPS system to be used to simply remove drift.

This study has unequivocally shown that DIC can successfully be used in vehicle dynamics to solve certain problems. These technique also has tremendous possibilities to bring to resolution certain questions not only in vehicle dynamics but other fields as well. As the cost, size and performance of cameras as well as computers improve, the range of applications will grow. Since these techniques have the possibility to provide a wide range of measurements using the same system, the potential applications of these techniques are endless.

Recommendations

Throughout this study, several shortcomings and additional application for the proposed DIC applications have been identified. In the side-slip angle measurement system both the Pose and 3D methods have potential to be used as rate gyroscopes. The 3D method could also be used to determine the roll and pitch angles as well as vertical height of the vehicle. This would enable the system to provide numerous additional measurements enabling a range of new applications such as suspension control, roll over detection and vehicle performance testing.

The longitudinal slip ratio measurement system also measures the tyre velocity, terrain velocity and slip ratio. The system however only measures in units of *pixels*. Use of a 3D stereographic rig would enable the measurements to be obtained in real world coordinates. This would enable the measurements to contain more meaning since no scale is needed and there would be no need to correct for the parallax error. The system should also use higher focal length lenses to reduce the area of interest. This would enable better results and should enable the system to determine the velocity variations along the longitudinal distance of the contact patch. Additionally a stereographic rig could be used on the inside of the tyre to measure the full velocity profile of the contact patch both longitudinally and laterally. This could possibly enable the slip across the contact patch to be determined which would be of great value in the understanding of the longitudinal tyre force generation mechanism.

The calibration of the stereographic rig should be improved to reduce the drift of the terrain profiling system. If the drift of the terrain profiler is reduced it would greatly improve its capability. However, while the drift can be reduced it would be impossible to completely remove the drift. It is therefore recommended to implement an absolute measuring system as well. Often GPS is used as an absolute measuring system however this study has shown that even GPS systems have their limitations. While the addition of a base station would greatly improve results the vertical accuracy would still not be enough for the purposes of terrain profiling. The proposed method however does not require a high accuracy and fast sampling rate GPS system. Therefore, the minimal errors of the GPS system should have a smaller effect. Additionally stereography can possibly be used to reduce drift by using it as an absolute measurement system by constantly measuring fixed points a certain distance away. This would remove the need of GPS and would simply require an additional set of inexpensive cameras.

Appendices

Appendix A

Camera Specifications

Point Grey Chameleon CMLN-13S2C-CS

Specification	Specification Value
Resolution	1296 x 964
Frame Rate	18 FPS
Megapixels	1.3 MP
Chroma	Color
Sensor Name	Sony ICX445
Sensor Type	CCD
Readout Method	Global shutter
Sensor Format	1/3"
Pixel Size	3.75 μm
Lens Mount	CS-mount
ADC	12-bit
Quantum Efficiency Blue (% at 470 nm)	46
Quantum Efficiency Green (% at 525 nm)	50
Quantum Efficiency Red (% at 640 nm)	34.8
Temporal Dark Noise (e-)	7.65
Saturation Capacity (e-)	6365
Dynamic Range (dB)	57.85
Gain Range	0 dB to 24 dB

Specification	Specification Value
Exposure Range	0.01 ms to 10 seconds
Trigger Modes	Standard, bulb, skip frames, overlapped
Partial Image Modes	Pixel binning, ROI
Image Processing	Color/Greyscale conversion, gamma, lookup table, white balance
User Sets	3 memory channels for custom camera settings
Flash Memory	256 KB non-volatile memory
Non-isolated I/O Ports	2 bi-directional
Serial Port	1 (over non-isolated I/O)
Auxiliary Output	3.3 V, 150 mA maximum
Interface	USB 2.0
Power Requirements	4.745 to 5.25 V via Mini-B USB 2.0 or JST 7-pin GPIO
Power Consumption (Maximum)	2 W
Dimensions	25.5 mm x 44 mm x 41 mm
Mass	37 grams
Machine Vision Standard	IIDC v1.31
Compliance	CE, FCC, RoHS
Temperature (Operating)	0°C to 45°C
Temperature (Storage)	-30°C to 60°C
Humidity (Operating)	20 to 80
Humidity (Storage)	20 to 95
Warranty	1 year

Point Grey Flea3 FL3-U3-13Y3M-C

Specification	Specification Value
Resolution	1280 x 1024
Frame Rate	150 FPS
Megapixels	1.3 MP
Chroma	Mono
Sensor Name	On Semi VITA1300

Specification	**Specification Value**
Sensor Type	CMOS
Readout Method	Global shutter
Sensor Format	1/2"
Pixel Size	4.8 μm
Lens Mount	C-mount
ADC	10-bit
Quantum Efficiency (% at 525 nm)	61
Temporal Dark Noise (e-)	26.26
Saturation Capacity (e-)	10226.41
Dynamic Range (dB)	51.64
Gain Range	0 dB to 18 dB
Exposure Range	0.006 ms to 1 second
Trigger Modes	Standard, bulb, multi shot
Partial Image Modes	Pixel binning, ROI
Image Processing	Gamma, lookup table, hue, saturation, and sharpness
Image Buffer	32 MB
User Sets	2 memory channels for custom camera settings
Flash Memory	1 MB non-volatile memory
Opto-isolated I/O Ports	1 input, 1 output
Non-isolated I/O Ports	2 bi-directional
Serial Port	1 (over non-isolated I/O)
Auxiliary Output	3.3 V, 150 mA maximum
Interface	USB 3.0
Power Requirements	5-24 V via GPIO or 5 V via USB 3.0
Power Consumption (Maximum)	¡3 W
Dimensions	29 mm x 29 mm x 30 mm
Mass	41 g
Machine Vision Standard	IIDC v1.32
Compliance	CE, FCC, RoHS
Temperature (Operating)	0°C to 45°C
Temperature (Storage)	-30°C to 60°C
Humidity (Operating)	20 to 80% (no condensation)

Specification	Specification Value
Humidity (Storage)	20 to 95% (no condensation)
Warranty	3 years

Lens Specifications

Fujinon DF6HA-1B

Specification	Specification Value
Mount Type	C-Mount
Imager Size	1/2" CCD
Focal Length	6mm
Aperture	f/1.2 to f/16
Iris	Manual
Focus	Manual
Maximum Horizontal Angle-of-View	56°C 09'
Maximum Vertical Angle-of-View	43°C 36'
Minimum Object Distance (M.O.D.)	4" (0.1 m)
Macro	None
Dimensions (WxHxD)	1.1 x 1.1 x 1.4" (29.5 x 29.5 x 36.7 mm)
	27 mm Filter Thread
Weight	1.9 oz (55 g)

Kowa LM4NCL

Specification	Specification Value
Main Sensor size	1/2
Focal length (mm)	3.5 mm

Specification	Specification Value
Iris range (F-stop)	f/1.40 - f/16
Focusing range (m)	0.20 m - ∞
Iris Control	Manual
Focus Control	Manual
Shooting range at MOD (m)	
horizontal:	396.00 m
vertical:	247.00 m
Resolution	
Center:	100.00 lp/mm
Corner:	60.00 lp/mm
Pixel-Size:	5.00 m
Distortion (TV)	-28.00%
Back focus in air (mm)	8.89 mm
Size	ø31 x 30.5 mm
Mount	C-mount
Weight	73 g
Temperature range	-10°C - +45°C
Angle of view	
Sensor size 1/3 inch	
horizontal	77.00°
vertical	57.70°
Sensor size 1/2 inch	
horizontal	103.60°
vertical	76.60°

Bibliography

[Abanto-Bueno and Lambros, 2002] Abanto-Bueno, J. and Lambros, J. (2002). Investigation of crack growth in functionally graded materials using digital image correlation. *Engineering Fracture Mechanics*, 69(14):1695–1711.

[Anderson et al., 2005] Anderson, A. B., Palazzo, A. J., Ayers, P. D., Fehmi, J. S., Shoop, S., and Sullivan, P. (2005). Assessing the impacts of military vehicle traffic on natural areas. introduction to the special issue and review of the relevant military vehicle impact literature. *Journal of Terramechanics*, 42(3):143–158.

[Andor, 2014] Andor (2014). Rolling and global shutter. http://www.andor.com/learning-academy/rolling-and-global-shutter-exposure-flexibility. [Online; accessed 2014-11-07].

[Andren, 2006] Andren, P. (2006). Power spectral density approximations of longitudinal road profiles. *International Journal of Vehicle Design*, 40(1):2–14.

[Arun et al., 1987] Arun, K. S., Huang, T. S., and Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (5):698–700.

[Bakker et al., 1987] Bakker, E., Nyborg, L., and Pacejka, H. B. (1987). Tyre modelling for use in vehicle dynamics studies. Technical report, SAE Technical Paper.

[Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.

[Becker and Els, 2014] Becker, C. M. and Els, P. S. (2014). Profiling of rough terrain. *International Journal of Vehicle Design*, 64(2):240–261.

[Besl and McKay, 1992] Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics.

[Bevly et al., 2006] Bevly, D. M., Ryu, J., and Gerdes, J. C. (2006). Integrating ins sensors with gps measurements for continuous estimation of vehicle sideslip, roll, and tire cornering stiffness. *Intelligent Transportation Systems, IEEE Transactions on*, 7(4):483–493.

[Botha and Els, 2012] Botha, T. R. and Els, P. S. (2012). Vehicle sideslip estimation using unscented kalman filter, ahrs and gps. In *ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 651–659. American Society of Mechanical Engineers.

[Bradski, 2000] Bradski, G. (2000). Opencv. *Dr. Dobb's Journal of Software Tools*.

[Cabrera et al., 2003] Cabrera, J., Ortiz, A., Simon, A., García, F., and La Blanca, A. P. (2003). A versatile flat track tire testing machine. *Vehicle System Dynamics*, 40(4):271–284.

[Cai and Aggarwal, 1999] Cai, Q. and Aggarwal, J. K. (1999). Tracking human motion in structured environments using a distributed-camera system. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(11):1241–1247.

[Chevalier et al., 2001] Chevalier, L., Calloch, S., Hild, F., and Marco, Y. (2001). Digital image correlation used to analyze the multiaxial behavior of rubber-like materials. *European Journal of Mechanics-A/Solids*, 20(2):169–187.

[Chung and Yi, 2006] Chung, T. and Yi, K. (2006). Design and evaluation of side slip angle-based vehicle stability control scheme on a virtual test track. *Control Systems Technology, IEEE Transactions on*, 14(2):224–234.

[Clover and Bernard, 1998] Clover, C. and Bernard, J. (1998). Longitudinal tire dynamics. *Vehicle System Dynamics*, 29(4):231–260.

[Detweiler and Ferris, 2010] Detweiler, Z. R. and Ferris, J. B. (2010). Interpolation methods for high-fidelity three-dimensional terrain surfaces. *Journal of Terramechanics*, 47(4):209–217.

[Deur et al., 2005] Deur, J., Ivanović, V., Troulis, M., Miano, C., Hrovat, D., and Asgari, J. (2005). Extensions of the lugre tyre friction model related to variable slip speed along the contact patch length. *Vehicle System Dynamics*, 43(sup1):508–524.

[Dhond and Aggarwal, 1991] Dhond, U. R. and Aggarwal, J. K. (1991). A cost-benefit analysis of a third camera for stereo correspondence. *International Journal of Computer Vision*, 6(1):39–58.

[Durini, 2014] Durini, D. (2014). *High performance silicon imaging : fundamentals and applications CMOS and CCD sensors*. Woodhead Publishin.

[Fiala, 1954] Fiala, E. (1954). Seitenkrafte am rollenden luftreifen. *ZVD I*, 96:Nr–29.

[Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

[Georgoulas and Andreadis, 2009] Georgoulas, C. and Andreadis, I. (2009). A real-time occlusion aware hardware structure for disparity map computation. In *Image Analysis and Processing–ICIAP 2009*, pages 721–730. Springer.

[Gerotek Test Facilities, 2014] Gerotek Test Facilities (2014). Gerotek Test Facilities. www.gerotek.co.za. [Online; accessed 2014-09-08].

[Grewe and Kak, 1994] Grewe, L. L. and Kak, A. C. (1994). Stereo vision. In *Handbook of pattern recognition and image processing (vol. 2)*, pages 239–317. Academic Press, Inc.

[Grimson, 1981] Grimson, W. E. L. (1981). A computer implementation of a theory of human stereo vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, pages 217–253.

[Gupta et al., 2011] Gupta, E. A., Gupta, E. A., and Mishra, E. A. (2011). Research paper on cluster techniques of data variations. *International Journal of Advance Technology & Engineering Research*, 1(1):39–47.

[Hartley et al., 1992] Hartley, R., Gupta, R., and Chang, T. (1992). Stereo from uncalibrated cameras. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 761–764. IEEE.

[Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

[Hartley and Sturm, 1997] Hartley, R. I. and Sturm, P. (1997). Triangulation. *Computer vision and image understanding*, 68(2):146–157.

[Heikkila and Silvén, 1997] Heikkila, J. and Silvén, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 1106–1112. IEEE.

[Hrovat et al., 1988] Hrovat, D., Margolis, D., and Hubbard, M. (1988). An approach toward the optimal semi-active suspension. *Journal of Dynamic Systems, Measurement, and Control*, 110(3):288–296.

[Hüsemann and Wöhrmann, 2010] Hüsemann, T. and Wöhrmann, M. (2010). The impact of tire measurement data on tire modeling and vehicle dynamics analysis. *Tire Science and Technology*, 38(2):155–180.

[Iagnemma et al., 2003] Iagnemma, K., Kang, S., Brooks, C., and Dubowsky, S. (2003). Multi-sensor terrain estimation for planetary rovers. In *Proceedings of the 8th international symposium on artificial intelligence, robotics, and automation in space. IEEE Press, New York*.

[Inagaki et al., 1994] Inagaki, S., Kshiro, I., and Yamamoto, M. (1994). Analysis on vehicle stability in critical cornering using phase-plane method. In *International Symposium on Advanced Vehicle Control (1994: Tsukuba-shi, Japan). Proceedings of the International Symposium on Advanced Vehicle Control 1994*.

[ISO 8608, 1995] ISO 8608 (1995). Mechanical vibration - road surface profiles - reporting of measured data. Technical report, International Organization for Standardization.

[Jacobson et al., 2000] Jacobson, R., Ray, S., Attridge, G. G., and Axford, N. (2000). *Manual of Photography*. Taylor & Francis.

[Jin et al., 2010] Jin, T., Jia, H., Hou, W., Yamamoto, R., Nagai, N., Fujii, Y., Maru, K., Ohta, N., and Shimada, K. (2010). Evaluating 3d position and velocity of subject in parabolic flight

experiment by use of the binocular stereo vision measurement. *Chinese Optics Letters*, 8(6):601–605.

[Julier and Uhlmann, 2005] Julier, S. J. and Uhlmann, J. K. (2005). Fusion of time delayed measurements with uncertain time delays. Technical report, DTIC Document.

[Kern and Ferris, 2007] Kern, J. V. and Ferris, J. B. (2007). Development of a 3d vehicle-terrain measurement system part i: Equipment setup. In *Proceedings of the Joint North America, Asia-Pacific ISTVS Conference, Fairbanks, AK*.

[Kertész et al., 2008] Kertész, I., Lovas, T., and Barsi, A. (2008). Photogrammetric pavement detection system. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37.

[Khare et al., 2010] Khare, S., Pillai, H., and Belur, M. (2010). Algorithm to compute minimal nullspace basis of a polynomial matrix. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems–MTNS*, volume 5.

[Kistler, 2014] Kistler (2014). Correvit s-hr.

[Konolige, 1998] Konolige, K. (1998). Small vision systems: Hardware and implementation. In *Robotics Research*, pages 203–212. Springer.

[Kowa, 2014] Kowa (2014). LM4NCL C-mount lens. http://www.kowa.eu/lenses/en/LM4NCL.187.php. [Online; accessed 2014-11-20].

[Krick, 1973] Krick, G. (1973). Behaviour of tyres driven in soft ground with side slip. *Journal of Terramechanics*, 9(4):9–30.

[Kriegman et al., 1989] Kriegman, D., Triendl, E., and Binford, T. O. (1989). Stereo vision and navigation in buildings for mobile robots. *Robotics and Automation, IEEE Transactions on*, 5(6):792–803.

[Kurjenluoma et al., 2009] Kurjenluoma, J., Alakukku, L., and Ahokas, J. (2009). Rolling resistance and rut formation by implement tyres on tilled clay soil. *Journal of Terramechanics*, 46(6):267–275.

[Li and Mourikis, 2013] Li, M. and Mourikis, A. I. (2013). 3-d motion estimation and online temporal calibration for camera-imu systems. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 5709–5716. IEEE.

[Longuet-Higgins, 1987] Longuet-Higgins, H. C. (1987). A computer algorithm for reconstructing a scene from two projections. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, MA Fischler and O. Firschein, eds*, pages 61–62.

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

[Lucas et al., 1981] Lucas, B. D., Kanade, T., et al. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679.

[Lucchese and Mitra, 2002] Lucchese, L. and Mitra, S. K. (2002). Using saddle points for subpixel feature detection in camera calibration targets. In *Circuits and Systems, 2002. APCCAS'02. 2002 Asia-Pacific Conference on*, volume 2, pages 191–195. IEEE.

[Mahalanobis, 1936] Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55.

[Mastin et al., 2009] Mastin, A., Kepner, J., and Fisher, J. (2009). Automatic registration of lidar and optical images of urban scenes. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2639–2646. IEEE.

[Mavros et al., 2005] Mavros, G., Rahnejat, H., and King, P. (2005). Transient analysis of tyre friction generation using a brush model with interconnected viscoelastic bristles. *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, 219(3):275–283.

[Miller et al., 2001] Miller, S. L., Youngberg, B., Millie, A., Schweizer, P., and Gerdes, J. C. (2001). Calculating longitudinal wheel slip and tire parameters using gps velocity. In *American Control Conference, 2001. Proceedings of the 2001*, volume 3, pages 1800–1805. IEEE.

[Moser et al., 2010] Moser, R. A., Sube, H. J., Turner, J. L., and Zakelj, P. (2010). 3d digital imaging correlation: Applications to tire testing. *Tire Science and Technology*, 38(2):100–118.

[Murray and Little, 2000] Murray, D. and Little, J. J. (2000). Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 8(2):161–171.

[Nistér, 2004] Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(6):756–770.

[Novak and Valasek, 1996] Novak, M. and Valasek, M. (1996). A new concept of semi-active control of trucks suspension. In *Proc. of AVEC*, volume 96, pages 141–151.

[Novatel, 2014] Novatel (2014). Span-cpt. http://www.novatel.com/products/span-gnss-inertial-systems/span-combined-systems/span-cpt/#overview. [Online; accessed 2014-09-08].

[OpenPCL, 2014] OpenPCL (2014). Open point cloud library. http://www.pointclouds.org. [Online; 2014-09-08].

[Pan et al., 2009] Pan, B., Qian, K., Xie, H., and Asundi, A. (2009). Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. *Measurement science and technology*, 20(6):062001.

[Point Grey, 2014] Point Grey (2014). Flea3 USB 3.0 camera. http://www.ptgrey.com/flea3-13-mp-mono-usb3-vision-vita-1300-camera. [Online; accessed 2014-11-20].

[Prince, 2012] Prince, S. J. (2012). *Computer vision: models, learning, and inference*. Cambridge University Press.

[Qian, 1997] Qian, N. (1997). Binocular disparity and the perception of depth. *Neuron*, 18(3):359–368.

[Rentschler and Uffenkamp, 1999] Rentschler, W. and Uffenkamp, V. (1999). Digital photogrammetry in analysis of crash tests. Technical report, SAE Technical Paper.

[Riesenfeld, 1981] Riesenfeld, R. F. (1981). Homogeneous coordinates and projective planes in computer graphics. *IEEE Computer Graphics and Applications*, 1(1):50–55.

[Rill, 2006] Rill, G. (2006). First order tire dynamics. In *Proceedings of the III European Conference on Computational Mechanics Solids, Structures and Coupled Problems in Engineering, Lisbon, Portugal*, volume 58.

[Rodehorst et al., 2008] Rodehorst, V., Heinrichs, M., and Hellwich, O. (2008). Evaluation of relative pose estimation methods for multi-camera setups. *International Archives of Photogrammetry and Remote Sensing (ISPRS08)*, pages 135–140.

[Rusinkiewicz and Levoy, 2001] Rusinkiewicz, S. and Levoy, M. (2001). Efficient variants of the icp algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 145–152. IEEE.

[Rusu et al., 2009] Rusu, R. B., Blodow, N., and Beetz, M. (2009). Fast point feature histograms (fpfh) for 3d registration. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3212–3217. IEEE.

[Rusu et al., 2007] Rusu, R. B., Blodow, N., Marton, Z., Soos, A., and Beetz, M. (2007). Towards 3d object maps for autonomous household robots. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 3191–3198. IEEE.

[Sakai, 1981] Sakai, H. (1981). Theoretical and experimental studies on the dynamic properties of tyres. part 1: review of theories of rubber friction. *International Journal of Vehicle Design*, 2(1).

[Sandu et al., 2008] Sandu, C., Taylor, B., Biggans, J., and Ahmadian, M. (2008). Building an infrastructure for indoor terramechanics studies: the development of a terramechanics rig at virginia tech. In *Proceedings of 16th ISTVS international conference, Turin, Italy*, pages 177–85.

[Schoberl et al., 2009] Schoberl, M., Fobel, S., Bloss, H., and Kaup, A. (2009). Modeling of image shutters and motion blur in analog and digital camera systems. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3457–3460. IEEE.

[Schwarz et al., 1997] Schwarz, R., Nelles, O., Scheerer, P., and Isermann, R. (1997). Increasing signal accuracy of automotive wheel-speed sensors by online learning. In *American Control Conference, 1997. Proceedings of the 1997*, volume 2, pages 1131–1135. IEEE.

[Shi and Tomasi, 1994] Shi, J. and Tomasi, C. (1994). Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE.

[Shibly et al., 2005] Shibly, H., Iagnemma, K., and Dubowsky, S. (2005). An equivalent soil mechanics formulation for rigid wheels in deformable terrain, with application to planetary exploration rovers. *Journal of Terramechanics*, 42(1):1–13.

[Slama et al., 1980] Slama, C. C., Theurer, C., and Henriksen, S. W. (1980). *Manual of photogrammetry.* Number Ed. 4. American Society of photogrammetry.

[Smith III and Ferris, 2010] Smith III, H. and Ferris, J. B. (2010). Calibration surface design and validation for terrain measurement systems. *Journal of Testing and Evaluation*, 38(4):431–438.

[Soane et al., 1981] Soane, B., Blackwell, P., Dickson, J., and Painter, D. (1981). Compaction by agricultural vehicles: a review i. soil and wheel characteristics. *Soil and Tillage Research*, 1:207–237.

[Soane and Van Ouwerkerk, 1995] Soane, B. and Van Ouwerkerk, C. (1995). Implications of soil compaction in crop production for the quality of the environment. *Soil and Tillage Research*, 35(1):5–22.

[SONY, 2014] SONY (2014). Sony pregius global shutte sensor. http://www.sony.net/Products/SC-HP/sensor/technology/pregius.html. [Online; accessed 2015-02-11].

[Spangler and Kelly, 1966] Spangler, E. B. and Kelly, W. J. (1966). Gmr road profilometer-a method for measuring road profile. *Highway Research Record*.

[Stewenius et al., 2006] Stewenius, H., Engels, C., and Nistér, D. (2006). Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294.

[Tombari et al., 2013] Tombari, F., Salti, S., and Di Stefano, L. (2013). Performance evaluation of 3d keypoint detectors. *International Journal of Computer Vision*, 102(1-3):198–220.

[Tsai and Huang, 1984] Tsai, R. Y. and Huang, T. S. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (1):13–27.

[Van Zanten et al., 1990] Van Zanten, A., Erhardt, R., and Lutz, A. (1990). Measurement and simulation of transients in longitudinal and lateral tire forces. Technical report, SAE Technical Paper.

[Wahl et al., 2003] Wahl, E., Hillenbrand, U., and Hirzinger, G. (2003). Surflet-pair-relation histograms: a statistical 3d-shape representation for rapid classification. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 474–481. IEEE.

[Wong, 2001] Wong, J. Y. (2001). *Theory of ground vehicles*. John Wiley & Sons.

[Xavier et al., 2012] Xavier, J., Sousa, A. M., Morais, J. J., Filipe, V. M., and Vaz, M. (2012). Measuring displacement fields by cross-correlation and a differential technique: experimental validation. *Optical Engineering*, 51(4):043602–1.

[Yimin et al., 2010] Yimin, L., Naiguang, L., Xiaoping, L., and Peng, S. (2010). A novel approach to sub-pixel corner detection of the grid in camera calibration. In *Computer Application and System Modeling (ICCASM), 2010 International Conference on*, volume 5, pages V5–18. IEEE.

[Zhang et al., 2005] Zhang, K., Li, X. R., and Zhu, Y. (2005). Optimal update with out-of-sequence measurements. *Signal Processing, IEEE Transactions on*, 53(6):1992–2004.

[Zhang, 2000] Zhang, Z. (2000). A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334.

[Zhong, 2009] Zhong, Y. (2009). Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 689–696. IEEE.