A METHOD FOR EXTRACTING TREE HEIGHT, DIAMETER AT BREAST HEIGHT, AND CROWN DIAMETER FROM HORIZONTALLY DISPLACED STEREOSCOPIC IMAGES

Kiplimo Cedric

A Thesis Submitted in Partial Fulfillment for the Award of the degree of Master of Science in Telecommunication Engineering in the School of Engineering, Dedan Kimathi University of Technology (DeKUT)

MARCH, 2024

DECLARATION

This thesis is my original work and has not been submitted for the award of any degree in any other university

Signature..... Date.....

Cedric Kiplimo

E224-01-0258/2021

This thesis has been submitted with our approval as the University Supervisors:

FIRST SUPERVISOR

Signature..... Date.....

Prof Ciira wa Maina, PhD

Department of Electrical and Electronic Engineering

Dedan Kimathi University of Technology

SECOND SUPERVISOR

Signature..... Date.....

Mr Billy Okal

NVIDIA Corporation, Santa Clara, California

ACKNOWLEDGEMENT

The **LORD God** *"in whom are hid all the treasures of wisdom and knowledge,"* is the undisputed source of every gleam of sublime thought and every flash of intellect in these pages.

I register utmost gratitude and profound indebtedness to **Prof Ciira wa Maina** and **Mr Billy Okal** for their guidance and supervision. The generous knowledge and expertise provided, patience shown, and the critical review and feedback given have contributed to an insightful research journey.

My studies were made possible by the generous scholarship I was awarded by the **Dedan Kimathi University of Technology** through the graduate assistantship program. The institution's support has been a cornerstone in achieving the research goals set forth in this study.

All of my research work was undertaken at the **Centre for Data Science and Artificial Intelligence (DSAIL)**, which afforded me extensive facilities and an exceptionally supportive atmosphere. I am profoundly thankful for the invaluable support provided by the centre.

Much gratitude to **Dr Edwell Tafara** (Chair – Department of Electrical Engineering) for his impeccable leadership and **Dr Waweru Njeri** whose lectures laid a good foundation for my research.

I thank **NVIDIA Corporation** for a hardware grant to the Centre for Data Science and Artificial Intelligence (DSAIL).

Many thanks to **my classmates and colleagues** for their encouragement and support while I carried out this research. My good friend **Miriam Njoroge** gave me a helping hand when performing multiple experiments. I owe her a great debt of gratitude. Special thanks to **Daniel Nyakundi** and **Eddy Kerosi** for the time they sacrificed to help collect the hundreds of images used when undertaking this research.

Finally, I am eternally grateful to **my dear parents** who, throughout my research journey, bore me up on the wings of their unconditional love and support. Their presence has been as enriching as it has been a beacon of hope and strength.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

ALS	Aerial Laser Scanning
CHM	Canopy Height Model
DBH	Diameter at Breast Height
FIA	Forest Inventory and Analysis
FRA	Forest Resources Assessment
ITS	Intelligent Transportation System
LiDAR	Light Detection and Ranging
MLS	Mobile Laser Scanning
NFI	National Forest Inventory
NGO	Non-Governmental Organisations
RMSE	Root Mean Square Error
RS	Remote Sensing
SfM	Structure from Motion
SLAM	Semantic Localisation and Mapping
SLOAM	Semantic Lidar Odometry and Mapping
TH	Tree Height
TLS	Terrestrial Laser Scanning
UAV	Unmanned Aerial Vehicle

ABSTRACT

Forests are very important because they provide fuel, medicine , and water for food production to many populations in the world. They also aid in climate change mitigation since they are the largest terrestrial carbon sink. Because of the many benefits drawn from forests, it is essential to collect and store measurement data to help monitor them. This systematic collection of forest data for use in assessment and analysis is called forest inventory. The most measured tree biophysical parameters in forest inventory are the diameter at breast height (DBH) and tree height (TH). Manual techniques used for forest inventory are slow and labour-intensive because tools often have to come into contact with trees, making it necessary to deploy more personnel to speed up the process. Many non-contact techniques proposed by researchers are expensive, computationally heavy, require extensive training or exhibit low accuracy. This study presents computational geometric algorithms for estimating the breast height (BH) location and extracting the DBH, TH and crown diameter (CD) from disparity images derived from stereoscopic images. It took 30 minutes to capture 105 image pairs of trunks belonging to 20 trees and 10 image pairs of full trees. Images for locating the BH were taken at 5 m, 6 m, 7 m, 8 m and 9 m, those for estimating the DBH were taken at 5 m and 8 m, and those for estimating TH and CD were recorded at arbitrary distances (typically less than 10 m) while ensuring that the whole vertical extents of the trees were visible in the images. The results indicated RMSEs of 1.02 cm ($R^2 = 0.9913$) and 0.94 cm ($R^2 = 0.9927$) in DBH estimation at 5 m and 8 m respectively. The biases in DBH estimation at these distances were 0.3 cm and 0.42 cm respectively. An MAE of 3.15 cm was achieved in the BH location, and RMSEs of 19.32 cm ($R^2 = 0.9540$) and 31.93 cm ($R^2 = 0.9209$) and biases of 3.5 cm and -0.3 cm in TH and CD estimation respectively. This implies that the technique can be used for DBH and TH estimation with very good accuracy. CD and TH estimation can be done only for short trees ($\leq$ 6 m in height) at a camera baseline of 12.9 cm, albeit with a slightly lower accuracy for CD.

CHAPTER 1

INTRODUCTION

1.1 Background

Forests are important to the world because they provide both utilitarian and ecosystem services. To more than one billion people, they provide medicines, fuel, and supply water for food production. Apart from these utilitarian services, forests also protect the world's watersheds, maintain soil structure, and function as a carbon store [1]–[3]. They cover nearly one-third of the earth's land surface and are the dominant source of plant biomass. Since 50% of plant biomass comprises carbon, it is clear that forests play a critical role in the carbon cycle [4]. Tree biomass is usually estimated using equations where tree attributes such as diameters and tree heights are key input variables [5]. These attributes are measured through a process known as forest inventory. Forest inventories are the primary information sources used in forest management and forestry policy formulation [6]. Some countries have National Forestry Inventory (NFI) programs for the collection and keeping of national forest resources data [7].

The world's first NFI was formed in Norway in 1917, although there is evidence that interest in national forest statistics there dates back to 1737. Sweden and Finland followed shortly afterwards in the early 1920s [8]. Raphael Zon and William Sparhawk published the first comprehensive report on global forest resources, setting the stage for later assessments which were continued by FAO [9]. Not long after this report was published, the United States Congress created the Forest Inventory and

Analysis (FIA) program in 1928 and mandated it to be the custodian of forest resources data in the US. FIA's task was to provide annual state forest inventory data of the highest quality, quinquennial (every five years) national forest statistics as well as define national standards to be followed in forest data collection [7].

The United Nations Food and Agriculture Organisation (FAO) has been conducting global assessments of forest resources since 1948 when it published the first assessment. Since 1990, reports of the assessment have been published every five years. The latest report, known as FRA 2020, was published in 2020 [1]. In the early years, these evaluations focused on collecting information on the available timber resources needed for construction following damage caused by the Second World War [9], [10]. Over the years, however, the Global Forest Resources Assessment (FRA) has grown in scope to include all the elements that comprise sustainable forest management. As a result, there has been a shift from a utilitarian evaluation of forests to one that is focused on entire ecosystems and their ecology [10].

Currently, FAO relies heavily on forest resource data received from countries when compiling their quinquennial reports [10]. FAO has been using a three-tier system for classifying data since FRA 2015. Tier 3 data are those sourced through field data collection and remote sensing (RS) in the last 10 years. When such data gets older than 10 years, it becomes tier 2 data. Tier 1 data are of the lowest quality and is sourced mainly through expert assessments [11]. In the early years of FRA, there was very limited country involvement and most of the data collection was done by FAO. At present, the burden of reporting is on individual countries and pressure has mounted on these nations to develop efficient and faster methods of data collection [10]. It is worth

noting that since the scope of reporting has grown substantially over the last three decades, the demand for information has increased to levels that developing nations have not been able to match. This inability is in large part because both the data and the funds required to collect it are often not available [9].

Most developing nations do not have comprehensive NFIs. This has made it difficult for them to provide up-to-date data to FAO and, consequently, concerns have been raised about the accuracy of forest statistics published in FRAs. Keenan *et al.* observed a strong relationship between the quality of data reported by countries and the amount of funding targeted at forest inventory [10]. They point out that between FRA 2005 and FRA 2015, there was a 14% increase in the proportion of tropical forests (which are located in developing countries) whose reported data quality was the highest. Since tier 3 data is expensive to acquire, targeted funding to developing nations improves reported data quality. FRA 2020 shows that African countries which received targeted funding showed the greatest improvement in their ability to produce their own NFIs. This detail shows that there will be growth in the proportion of tier 3 (NFI and RS) data if targeted funding is sustained [1], [11].

Policy and investment decisions by government agencies, NGOs and scientists are informed by reliable information on the status and change in forest area in regions of interest [11]. As already noted, investment in tropical countries has led to an increase in the data quality [10]. High-quality data is not only useful for reporting to international agencies but also for forest management within those countries. Nations can make informed policy changes such as increasing forest cover, diversifying tree species, banning forest harvesting in certain areas, etc. when forest inventory data is

available and accurate [8]. As an example, policy and decision-making in China rely heavily on the availability of forest resource data at different scales [12]. Norway has stringent rules forbidding the harvesting of trees below a certain diameter, while forest stands in certain areas have been marked for conservation [8]. Knowledge of tree attributes is essential when making decisions like these. The ultimate goal of all forest research is to produce accurate information that facilitates decision-making [6].

Another important objective of forest inventory is to monitor forest health. Forest health evaluation aims to draw links between observations made and the causes or drivers of change, whether human or non-human. Trumbore *et al.* [13] divided the indicators of forest health into utilitarian and ecosystem indicators. Utilitarian indicators included factors like wood yield and forest growth while ecosystem indicators focus on ecological aspects such as soil fertility and habitat quality. The main tools recommended for assessment included remote sensing and inventory plots, which are both NFI techniques. Forest recovery trajectories help planners to anticipate any disruptions to supply chains of forest products and plan to mitigate against the disruptors [13].

The kind of data required for decision and policymaking should exhibit statistical rigour. While analyzing the history of the FIA program, Smith points out that it has been useful for supplying data that is both statistically grounded and scientifically authentic [7]. The role of providing detailed and exhaustive data cuts across most NFIs in industrialised nations. In China, the NFI's role is to provide accurate forest data in time to facilitate decision-making [12], implying that the process should be both accurate and fast. Gadow *et al.* suggested that, since forest research data facilitates

decision-making, it ought to be accurate [6]. This kind of quality requires acquisition techniques whose accuracy compares to manual methods [7]. In developing nations where the proportion of tier 3 data reported is still low, such methods should be made more accessible [11].

The level of detail and accuracy in available data depends on the data collection techniques chosen. It has been stated before that the highest quality of data (tier 3) is provided by NFIs and RS [11]. While these two methods may provide detailed forest statistics, RS is more suitable where scale is of interest such as in estimating forest cover [12]. The assumption that RS provides accurate information about how forests have changed over time has been described as tenuous by experts [9]. For example, this technique is known to be inept at differentiating between forest and woody horticultural trees and has very low accuracy at determining the proportion of land under tree cover, where tree cover densities are below 30% [9], [10]. Its accuracy at measuring under canopy attributes such as basal area and diameter at breast height is also low [12]. Put differently, RS is not capable of detecting other less visible ways forests change [9]. Despite its misgivings, RS remains an indispensable technique for performing forest inventory.

The consensus among researchers is that the type and quality of data needed inform the inventory techniques deployed for the collection of forest attribute data. This is evident in FIA's three-phase approach to forest inventory. In phase 1, remotely sensed data is collected using satellite images and aerial photographs, and then it is interpreted. The interpretation comprises the classification of photo points as either forest or non-forest. This process provides a high-level view of the forest and does not

provide much detail. The second and third phases involve field data collection where tree species, heights, diameters at breast height (DBH), vigour measurements etc. are the variables of interest [7]. In China, experts recognise that most forest attributes cannot be estimated with adequate accuracy using RS. As such, field measurements continue to be necessary for the foreseeable future [12]. The case is similar in Norway where tree-level measurements such as growing stock are done using small sample plots [8].

Since forest inventory exercises often involve covering vast areas during surveys, it is desirable that the techniques deployed be fast, accurate, reliable, practical and of low cost [14]. Low costs makes the collection of tier-3 data affordable for many developing countries [11]. The constraints of accuracy, cost, and speed should be optimised after which the best method is chosen [6]. Another point to note is that the manual techniques in widespread use currently require a lot of training and experience to ensure the accuracy of data [8]. Many researchers have developed non-contact methods for measuring important tree attributes taking advantage of recent advancements in image processing, photogrammetry and computer vision [14]–[18]. In analysing these methods, the desirable features of speed, accuracy, reliability, practicality, and cost must be taken into consideration [14].

1.2 Problem Statement

Many developing countries lack reliable and high-quality forest resource data. FAO requires UN member countries to report their national forest statistics every five years [10]. However, developing nations face challenges in collecting and reporting accurate

data [11]. FRA 2020 showed that only 69% of growing stock status data from Africa were based on tier-3 data, the highest quality level [1]. This low proportion of high-quality data is attributed to inadequate funding for forest inventory activities and, indeed, targeted funding has been shown to lead to positive results [11]. This gap is due to insufficient funding for forest inventory activities, which are costly and involve the use of advanced techniques such as laser scanning.

Manual forest inventory methods are slow and costly to apply over large areas. Foresters use tools like diameter tapes and callipers for measuring the tree DBH, and hypsometers for tree heights. The use of these tools requires significant effort, expertise, yields subjective estimates from field staff, and are impractical for extensive forest stands [8]. Furthermore, some foresters encounter dangerous animals when measuring trees using manual methods that require coming into contact with them. While non-contact techniques such as Structure from Motion (SfM) and Simultaneous Localisation and Mapping (SLAM) can assess multiple trees simultaneously, they report higher inaccuracies for distant trees [5], [18]. These methods demand significant computation time, and the software required for processing acquired images is expensive [19]–[22]. Thus, the gains in efficient data acquisition are offset by computing costs.

1.3 Objectives

1.3.1 Main Objective

The main goal of this thesis is to develop a method for measuring tree height, diameter at breast height, and crown diameter using horizontally displaced stereoscopic image

pairs.

1.3.2 Specific Objectives

The specific objectives are:

(i) To develop the geometric framework for calculating diameter at breast height, tree height, and crown diameter.

(ii) To develop an algorithm for extracting tree height, diameter at breast height, and crown diameter from horizontally displaced stereoscopic image pairs.

(iii) To validate the technique in a sparse park setting.

1.4 Significance of this Research

Developing nations have continued to lag in reporting tier-3 forest resources data. Although targeted funding has led to the increase in the proportion of this data, funding alone is not a sufficient solution for this problem. The main reason why the targeted funding has led to the evident increase in the proportion of high-quality data reported by developing countries is because the cost of data collection using techniques in widespread use is very high. Developing inventory techniques that are fast, low-cost, practical, and as accurate as current methods will significantly reduce the cost of forest inventory activities. Yet another benefit accruing from this study is the reduced acquisition time for tree attributes. Foresters usually have to deploy different tools to measure DBH, tree heights, and crown diameters. To measure all these attributes with a single setup would result in much shorter measurement times.

All these factors motivate the desire for a fast, accurate, practical, and low-cost method for measuring tree attributes.

The main contributions of this study are:

(a) A method for leveraging stereoscopic vision to estimate tree height and crown diameter.

(b) Algorithms for estimating tree height, diameter at breast height, and crown diameter using disparity maps derived from ground-based stereoscopic images.

(c) A technique for preventing errors in distance estimation caused by the presence of anomalies in disparity maps.

1.5 Scope of the Study

This research study is limited to the extraction of attributes of a single tree from a pair of stereoscopic images. The attributes of interest are the tree height, diameter at breast height and crown diameter only. Once the algorithm had been developed, it was tested and validated using data acquired from a park setting, where the trees are at least 5 m apart to ensure that the tree crowns are clearly delineated.

1.6 Thesis Organisation

The rest of this thesis is arranged as follows. Chapter 2 presents the main concepts behind stereoscopic vision and scene reconstruction, and a critical review of the various non-contact techniques that have recently been developed by researchers. The gaps in existing knowledge identified during the literature survey have also been

highlighted. Chapter 3 describes the proposed approach, the experiments performed and the metrics used for evaluating the performance of the proposed method. Chapter 4 presents the results of the study and discusses their meaning and significance to forest inventory. Chapter 5 presents the conclusions arrived at after the investigation, highlights some recommendations to improve the technique, and points out directions for further research.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents a critical review of what has been achieved in the past on the development of non-contact techniques of forest inventory. The theory of stereoscopic vision as well as the research gaps identified are also discussed.

2.2 Main Concepts

The theoretical framework for stereoscopic vision is developed and discussed in great detail by Hartley and Zisserman [23] and by Szeliski [24] in their textbooks. The main concepts presented and discussed in this section as well as section 2.3 are mature and well-established in the literature, and the methodology developed in this thesis builds on top of them.

2.2.1 Perspective Projection

A camera captures an image by projecting points on a real-world scene ($\mathbb{R}^3$) onto a 2D image plane ($\mathbb{R}^2$). Figure 2.1 is adapted from Hartley and Zisserman's textbook [23] and is an illustration of how a point P_o in $\mathbb{R}^3$ is mapped onto the point P_i in $\mathbb{R}^2$ (image plane). P_o is represented by the vector $\overline{r_o} = (x_o, y_o, z_o)$ while Pi by the vector $\overline{r_i} = (x_i, y_i, f)$, where f is the focal length of the camera.

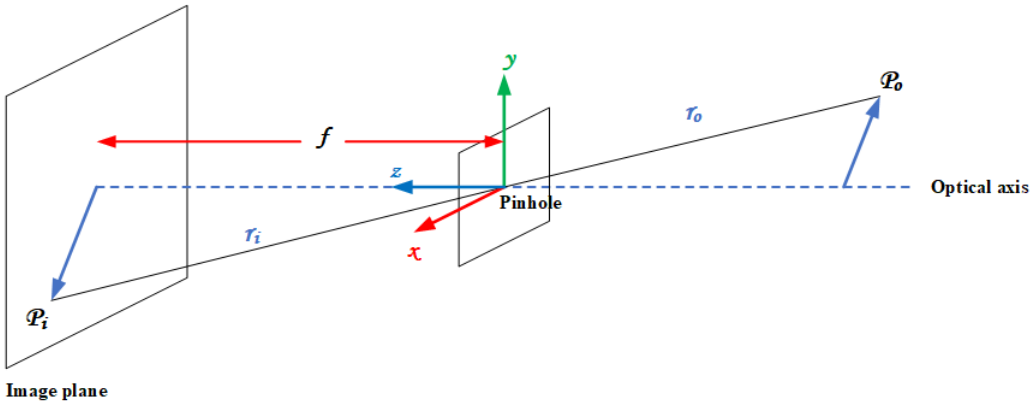


Figure 2.1: Perspective projection.

By applying the concept of similar triangles to figure 2.1, the result is the set of perspective projection equations captured in equation 2.1.

$$\frac{\overline{r_o}}{f} = \frac{\overline{r_i}}{z_o} \Rightarrow \frac{x_i}{f} = \frac{x_o}{z_o} \; and \; \frac{y_i}{f} = \frac{y_o}{z_o} \tag{2.1}$$

These perspective projection equations are useful for developing the camera's forward imaging model [23].

2.2.2 Forward Imaging Model

The objective of the forward imaging model is to obtain the description of a point P_o in $\mathbb{R}^3$ (real-world scene) with respect to the camera coordinate frame C. This model is illustrated by figure 2.2, also adapted from [23]. The point P_o has 3D coordinates with respect to both C and the world coordinate reference W.

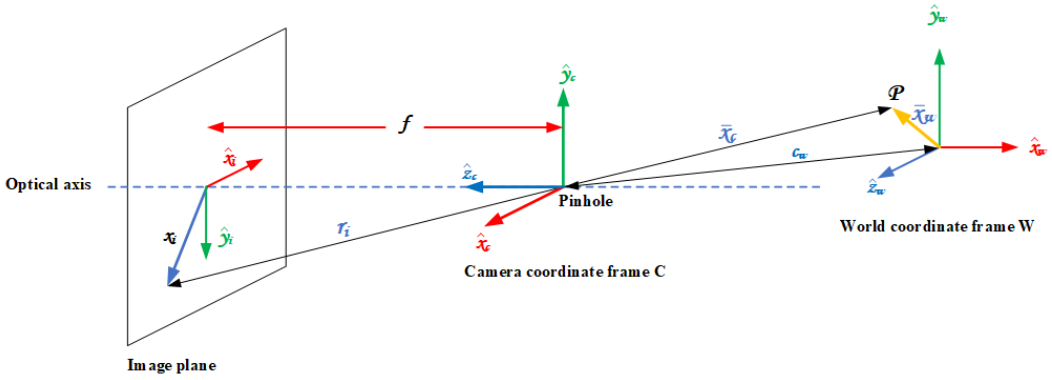


Figure 2.2: Forward imaging model.

Hartley and Zisserman [23] describe the mapping from $\mathbb{R}^3$ to $\mathbb{R}^2$ in three steps. First of all, the location of P_o with respect to C and on the image plane are highlighted. P_o's location in C is the vector $\overline{x_c}$ and it is mapped onto a 2D point on the image plane described by the vector $\overline{x_i}$ where:

$$\overline{x_i} = \begin{pmatrix} x_i \\ y_i \end{pmatrix}; \quad \overline{x_c} = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} \tag{2.2}$$

The transformation that maps vector $\overline{x_c}$ to $\overline{x_i}$ is the perspective projection.

Second, the perspective projection equations in equation 2.1 are substituted into the $\mathbb{R}^2$ location of P_o provided in equation 2.2. The camera image plane has the pixel densities (in pixels/mm) m_x and m_y in the x and y directions respectively. By combining these two quantities with equation 2.2 Then the corresponding 2D location of P on the image plane (u, v) can be found using equation 2.3.

$$u = m_x x_i = m_x f \frac{x_c}{z_c}, \quad u = m_y y_i = m_y f \frac{y_c}{z_c} \tag{2.3}$$

Finally, equation 2.3 is generalised to the case where the principal point of the image, denoted by coordinates (o_x, o_y), does not correspond to the origin of C. In such a case, equation 2.3 is rewritten as:

$$u = m_x f \frac{x_c}{z_c} + o_x, \quad u = m_y f \frac{y_c}{z_c} + o_y \tag{2.4}$$

and finally

$$u = f_x \frac{x_c}{z_c} + o_x, \quad u = f_y \frac{y_c}{z_c} + o_y \tag{2.5}$$

where f_x and f_y are the focal lengths in pixels along the x and y directions. The parameters f_x, f_y, o_x, and o_y constitute the intrinsic parameters of the camera. It is important to note that the two projection equations for u and v are non-linear. The coordinate frame W is mapped onto C by a linear transformation.

2.2.3 Homogeneous Coordinates

To convert the non-linear projection equations to their linear equivalents, the concept of homogeneous coordinates is used (figure 2.3). A 2D point (u, v) on the image plane is described with respect to the homogeneous coordinate frame $(\tilde{u}, \tilde{v}, \tilde{w})$ with the uv plane located at $\tilde{w} = 1$.

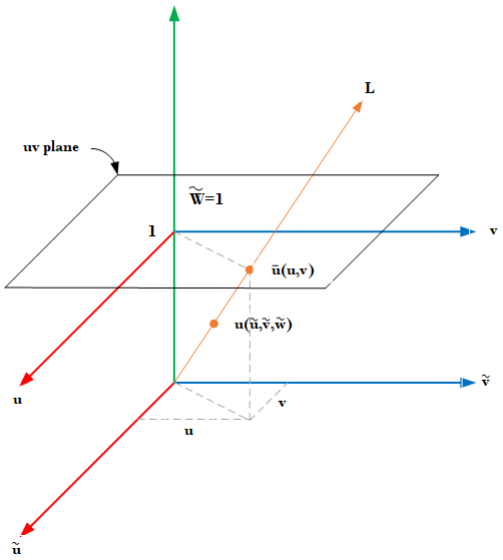


Figure 2.3: Homogeneous Coordinates

The transformation between the two frames is:

$$u = \frac{\tilde{u}}{\tilde{w}}, \qquad v = \frac{\tilde{v}}{\tilde{w}} \tag{2.6}$$

and consequently

$$u = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \equiv \begin{pmatrix} \tilde{w}u \\ \tilde{w}v \\ \tilde{w} \end{pmatrix} \equiv \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{w} \end{pmatrix} = \bar{u} \tag{2.7}$$

Equation 2.7 maps the 2D point $\bar{u} = (u, v)$ onto the 3D point $u = (\tilde{u}, \tilde{v}, \tilde{w})$ where the third coordinate $\tilde{w} \neq 0$ is fictitious and is only used for scaling and normalization. All the points (except the origin) along the line L crossing the origin and the point $\bar{u} = (\tilde{u}, \tilde{v}, \tilde{w})$ represent the homogeneous coordinates of $\bar{u}$ (u, v). Given any point along L, $\bar{u}$ can be found. The line L (excluding the origin) also represents the set of all possible 3D points in the real-world space that get mapped onto the 2D point $\bar{u}$ (u, v) on the image plane. Put differently, for any point $\bar{u}$ (u, v) on the image plane, it is not possible to compute its depth in the scene. Instead, all that is known about the

location of its corresponding 3D location lies along an outgoing ray [23], [24]. With the equations now linearised, they represent a linear transformation from the camera 3D space to the image plane that can be captured in a matrix known as the intrinsic matrix M_{int} which is written as in equation 2.8.

$$M_{int} = \begin{bmatrix} f_x & 0 & o_x & 0 \\ 0 & f_x & o_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.8}$$

For the complete mapping of a point in $\mathbb{R}^3$ with respect to W on to the image plane, a second matrix is needed to express its coordinates relative to W to their equivalents relative to C. This matrix is known as the extrinsic matrix and is written as in equation 2.9

$$M_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.9}$$

The product of matrices M_{int} and M_{ext} yields a 4×4 projection matrix which completely maps a point in $\mathbb{R}^3$ to one in $\mathbb{R}^2$ (the image plane).

2.3 Reconstructing Scene Geometry

Stereoscopic vision provides a way to reconstruct a scene from at least two images taken at different points of view. The reconstructed image, known as a disparity image, is one in which a pixel's greyscale intensity is an index to its depth in the scene. Darker pixels are further away while brighter ones are close to the camera. The theory of stereoscopic vision and the process of deriving the disparity map is presented in sections 2.3.1 to

2.3.4.

2.3.1 Simple Stereo

To recover the 3D structure of a scene, two images horizontally displaced with respect to each other are used to triangulate the depth of points on the scene recorded on both images. Figure 2.4, adapted from [23] and [24], is an illustration of this two-view geometry. The horizontal separation distance between the two camera centres is known as the baseline.

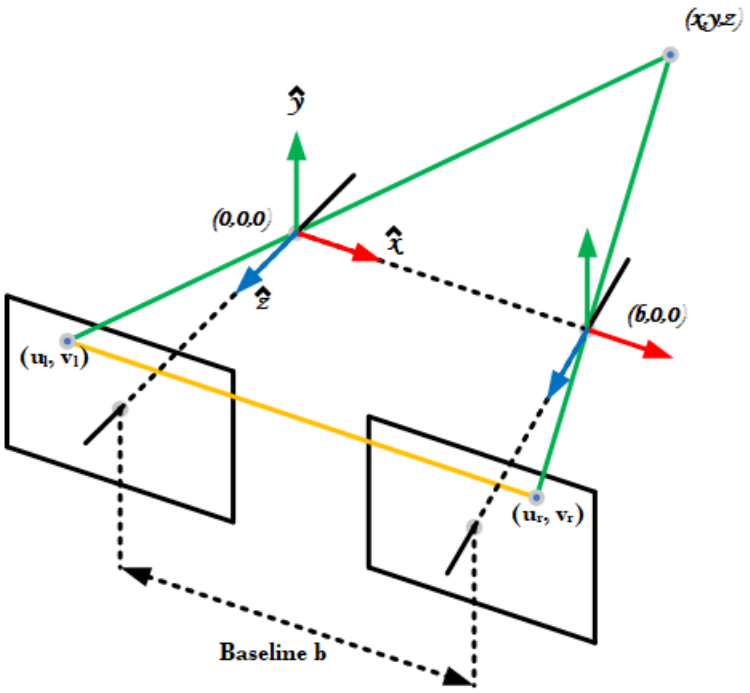


Figure 2.4: Simple Stereo.

The two images contain overlapping projections of the same 3D scene captured simultaneously by two identical cameras. The point with coordinates (x, y, z) in the scene is projected onto (u_l, v_l) and (u_r, v_r) in the left and right images respectively. Relative to each image, this point lies along an outgoing ray (the set of all homogeneous coordinates) such that the rays from the corresponding points on each

image intersect at the exact location of P in $\mathbb{R}^3$. In his textbook, Szeliski [24] provides the derivation of the depth z of the point from the projection equations and geometry of figure 2.4. This derivation is captured in equation 2.10.

$$\begin{gathered} u_l = f_x \frac{x}{z} + o_x; \quad u_r = f_x \frac{x-b}{z} + o_x \\ v_l = f_y \frac{y}{z} + o_y; \quad v_r = f_y \frac{y}{z} + o_y \\ u_l - u_r = f_x \frac{b}{z} \\ \Rightarrow \quad z = \frac{b f_x}{u_l - u_r}; \;\; x = \frac{b(u_l - o_x)}{u_l - u_r}; \;\; y = \frac{b(v_l - o_y)}{f_y(u_l - u_r)} \end{gathered} \tag{2.10}$$

The value b is the stereo baseline and $u_l - u_r$ is the horizontal disparity in pixels between the corresponding pixels on both images. Beginning wit equation 2.1, the location of a point on the scene relative to the camera coordinate frame C has been derived. Equation 2.10 contains all the information needed to reconstruct the 3D structure of the scene. More precise depth estimates are achieved when the disparity is larger. The relative motion between the two cameras is captured by the fundamental matrix F which has the form in equation 2.11 [23], [24].

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \tag{2.11}$$

2.3.2 Epipolar Geometry and Image Rectification

When an image is captured by a digital camera, information about the depths of points in the scene is lost. To recover the depth of a point in the image plane, at least two views of the scene are needed, and these views can be acquired by using more than one camera. The projective geometry that relates these views is known as epipolar

geometry [23], [24]. Human beings perceive their environment in three dimensions in a similar fashion using two eyes to achieve stereoscopic vision.

Figure 2.5 is Bradski and Kaehler's [25] illustration of epipolar geometry. Similar illustrations have also been done by Hartley and Zisserman [23] and by Szeliski [24] in their textbooks. This figure shows a point in $\mathbb{R}^3$ mapped onto points x_l and x_r in the left and right planes respectively. It can be seen that x_l, x_r, and the point in $\mathbb{R}^3$ are coplanar constituting a plane known as the epipolar plane. The camera centres C_l and C_r also form a part of this plane. The projection of each camera centre onto the other image plane is known as an epipole, and the figure shows the two epipoles e_l and e_r in the left and right image planes [23]. Figure 2.5 shows that the epipolar plane intersects the two image planes along two lines, l_l and l_r. These lines are referred to as epipolar lines (epiline in short) and they all pass through the epipoles. For every point X in $\mathbb{R}^3$, there is one and only one corresponding epiline in each image plane. This is known as the epipolar constraint.

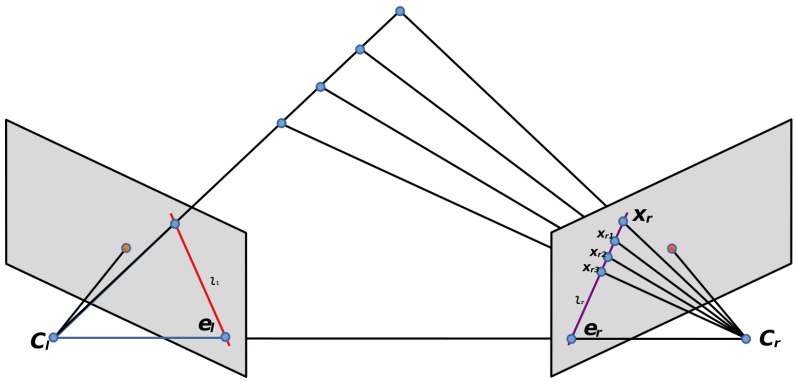


Figure 2.5: Epipolar geometry.

2.3.2.1 Epipolar Constraint

When only x_l is known and x_r is unknown, it is not possible to recover the depth of X because all points on the line $x_l X$ project onto x_l. However, the location of x_r in this case, will be constrained such that it falls along the epiline l_r. Therefore, the stereo correspondence algorithm will restrict its search to a one-dimensional search along this line, and this fact is pointed out by both Hartley and Zisserman [23] and Szeliski [24].

Figure 2.6, adapted from Bradski and Kaehler's textbook [25], describes the relative position and orientation of the two cameras which is captured by two matrices t and R where:

- t is the position of the right camera in the left camera's coordinate frame.
- R is the rotation of the left camera in the right camera's coordinate frame.

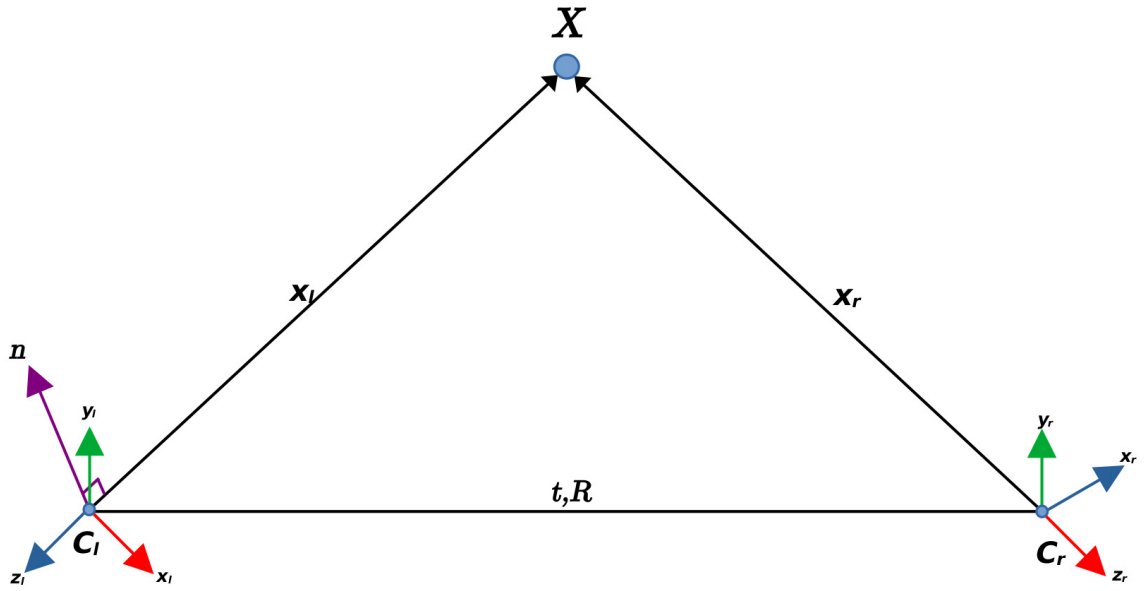


Figure 2.6: Relative position and orientation of left and right cameras.

The goal of epipolar geometry is to establish the relationship between the image planes

of the two cameras. Hartley and Zisserman [23] discuss in great detail the process of deriving this relationship. A summary of that derivation is provided here. There is a vector $\hat{n}$ normal to the epipolar plane and its equation is:

$$\hat{\mathbf{n}} \;=\; \mathbf{t} \times \bar{\mathbf{x}}_l \tag{2.12}$$

Vector x_l is the location of point X in the scene with respect to the left camera's coordinate frame. The dot product of $\hat{n}$ and x_l must be zero since they are perpendicular. This requirement is known as the epipolar constraint [23] and is expressed as in equation 2.13.

$$\bar{\mathbf{x}}_l \cdot (\mathbf{t} \times \bar{\mathbf{x}}_l) \;=\; 0 \tag{2.13}$$

Equation 2.14 is the matrix form of equation 2.13.

$$\bar{\mathbf{x}}_l^T \; T_X \; \bar{\mathbf{x}}_l \;=\; 0 \tag{2.14}$$

Where T_X is the translation matrix that transforms the vector x_l in the same way as the cross product $t \times \bar{x_l}$ does. Given R, t, and $\bar{x_l}$, $\bar{x_r}$ can be expressed as a rotation of $\bar{x_r}$ followed by a translation as shown by equation 2.15 [23].

$$\bar{\mathbf{x}}_l = \mathbf{R} \; \bar{\mathbf{x}}_r \;\; + \mathbf{t} \tag{2.15}$$

Combining equations 2.14 and 2.15 yields equation 2.16

$$\bar{\mathbf{x}}_l^T (T_X \mathbf{R} \bar{\mathbf{x}}_r + T_X \mathbf{t}) = 0 \tag{2.16}$$

The last term in the parentheses is equal to zero because it is a cross-product of two similar vectors. Equation 2.16 reduces to:

$$\bar{\mathbf{x}}_l^T T_X \mathbf{R} \bar{\mathbf{x}}_r = \bar{\mathbf{x}}_l^T E \bar{\mathbf{x}}_r = 0 \tag{2.17}$$

The product of the translation matrix T_X and the rotation matrix R forms a new 3×3 matrix known as the essential matrix E. Matrices T_X and R can be decoupled from E using singular value decomposition. The objective of stereo camera calibration is to compute the essential matrix E, from which T_X and R can be obtained.

Equation 2.17 cannot be used to compute E since both $\bar{x_l}$ and $\bar{x_r}$ are unknown. However, by incorporating the homogeneous coordinates of point X in the image planes and the camera matrices, the result is a more useful representation of equation 2.17. Equations for $\bar{x_l}$ and $\bar{x_r}$ are rewritten to arrive at equation 2.18 [23].

$$\bar{x_l} \;\; x_{lH} \; z_l \; K_l^{-1^T}; \qquad\qquad K_r^{-1} \;\; \bar{x_r} \; z_r \; x_{rH} \tag{2.18}$$

In equation 2.18, x_{lH} is x_l expressed as homogenous coordinates and the same is true for x_rH. z_l and z_r are the real-world depths of X. Equation 2.17 is now rewritten as:

$$\begin{gathered} x_{lH} z_l K_l^{-1^T} E K_r^{-1} \bar{x_r} z_r x_{rH} = 0 \\ z_l \neq 0; \qquad z_r \neq 0 \\ x_{lH} K_l^{-1^T} E K_r^{-1} x_{rH} = 0 \end{gathered} \tag{2.19}$$

The product $K_l^{-1^T} E K_r^{-1}$ also a new 3×3 matrix known as the fundamental matrix F.

Equation 2.19 now simplifies to:

$$x_{lH} F x_{rH} = 0 \tag{2.20}$$

With both x_{lH} and x_{rH} being known, F can be computed from equation 2.20. This new equation represents a mapping between a point x_{lH} in one image plane and its corresponding epiline Fx_{rH} in the other image plane [23]. The fundamental matrix F is the algebraic description of epipolar geometry. Being a singular matrix, it may be conveniently expressed as a product $F = [e]_X M$ where M is a non-singular matrix and $[e]_X$ a skew-symmetric matrix where e is the first image's epipole [23].

2.3.2.2 Image Rectification

The epipolar constraint provides a way for constraining the search for pixel correspondence between the left and right images by reducing the search space to only one dimension. According to [23], a more efficient approach is to first warp the images so that the epilines run parallel to the x-axis and correspond between the two views. This process is known as image rectification and it results in disparities between the images being in the x direction only and y disparities being zero [24].

Hartley and Zisserman [23] describe the process of making the epilines parallel in their textbook. A projective transformation H is needed which maps the image's epipole to infinity. Given two images J_l and J_r with the fundamental matrix $F = [e]_X M$, the goal is to apply two projective transformations H_l and H_r respectively on them such that an epipolar line in J_l is matched to its corresponding epiline in J_r. H_l and H_r form a

matched pair of transformations which satisfy the condition for some vector **a**[23].

$$H_l(I + H_r e_r \mathbf{a}^T) H_r M \tag{2.21}$$

Since H_r maps e_r to a point in infinity $(1,0,0)^T$, $I + H_r e_r \mathbf{a}^T = I + (1,0,0)^T \mathbf{a}^T$ will be an affine transformation of the form in equation 2.22 [23].

$$H_A = \begin{bmatrix} a & b & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.22}$$

Where there are n point correspondences between the two images, a transformation that minimizes the least squares distance in equation 2.23 [23] is sought.

$$\sum_{i}^{n} d(H_l x_{l_i}, H_r x_{r_i})^2 \tag{2.23}$$

The rectified images will be the result of applying H_l on J and H_r on J_r. Once the images are rectified, the disparity map is the computed by applying the stereo matching algorithm on them.

2.3.3 Stereo Matching

Correspondence matching between the two images must be done to determine where the corresponding points lie before calculating their disparities. Scharstein and Szeliski [26] provide a taxonomy of dense correspondence algorithms applied in computing the disparity map. Their taxonomy was based on the observation that the stereo correspondence algorithms generally implement a subset of four steps which

include matching cost computation, cost aggregation, disparity map computation and optimization, and disparity refinement [24]. One commonly used correspondence algorithm today is the the Semi-Global Matching (SGM) algorithm proposed by Heiko Hirschmuller [27]. A modified implementation of SGM, known as Semi-Global Block Matching (SGBM), is available in the OpenCV library [25] as well as in MATLAB [28]. Figure 2.7 is an illustration by MathWorks of the block matching step of the SGBM algorithm as implemented in MATLAB [28].

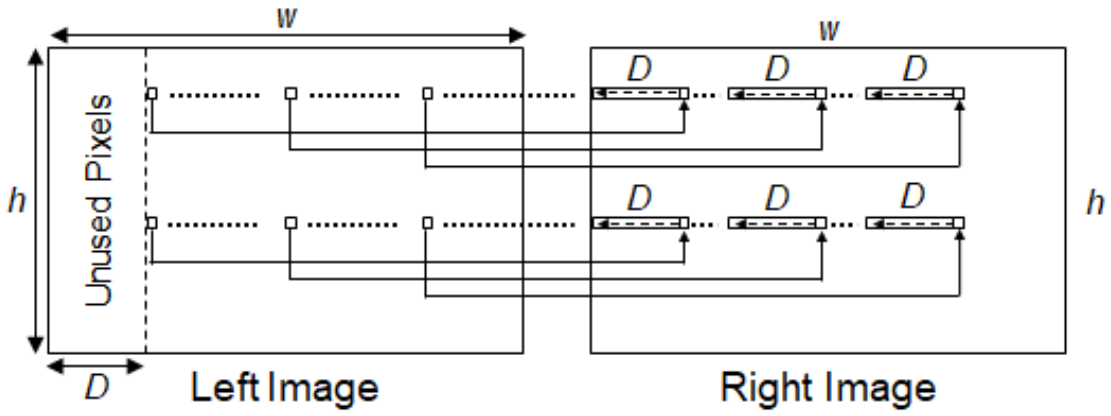


Figure 2.7: Disparity between left and right images.

Heiko Hirschmuller's SGM algorithm is very popular, implementations of it are widely available, and is representative of the algorithms in Scharstein and Szeliski's taxonomy [26]. For these reasons, the working of its modified version (SGBM) has been discussed in further detail here. The algorithm is implemented in three major steps: matching cost calculation, directional cost aggregation and post-processing.

2.3.3.1 Matching Cost Calculation

In block matching, the goal is to find the block of pixels in the second image that most closely matches a block of pixels in the base image based on a matching cost. This matching cost is calculated based on the Birchfield – Tomasi sub-pixel metric [29]. This metric measures the dissimilarity between two pixels using the linearly interpolated intensity function around the corresponding pixel, thus making this approach robust to image sampling. Given a left and right scanline in the left and right images respectively, the aim is to determine the dissimilarity between a pixel x_L in the left scan line and another at x_R in the right scanline. This is achieved by first defining two symmetric functions as shown in equation 2.24.

$$\begin{aligned} \bar{d}(x_L, x_R, I_L, I_R) &= \min_{\{x_R - \frac{1}{2} \leq x \leq x_R + \frac{1}{2}\}} |I_L(x_L) - \hat{I}_R(x)| \\ \bar{d}(x_R, x_L, I_R, I_L) &= \min_{\{x_L - \frac{1}{2} \leq x \leq x_L + \frac{1}{2}\}} |\hat{I}_L(x_L) - I_R(x)| \end{aligned} \tag{2.24}$$

Where $\hat{I}_L$ and $\hat{I}_R$ are the linear interpolating functions between the sample points in the left and the right scanlines respectively. The Birchfield – Tomasi sub-pixel dissimilarity metric is the minimum of the two quantities as calculated in equation 2.24 [29].

$$d(x_L, x_R) = \min\left(\bar{d}(x_L, x_R, I_L, I_R), \bar{d}(x_R, x_L, I_R, I_L)\right) \tag{2.25}$$

In the next step, the intensity at the point half-way between x_R and the pixel to its

immediate left, and the analogous quantity are calculated as follows:

$$\begin{aligned} I_R^- &\equiv \hat{I}_R\left(x_R - \frac{1}{2}\right) = \frac{1}{2}\Big(I_R(x_R) + I_R(x_R - 1)\Big) \\ I_R^+ &\equiv \hat{I}_R\left(x_R + \frac{1}{2}\right) = \frac{1}{2}\Big(I_R(x_R) + I_R(x_R + 1)\Big) \end{aligned} \tag{2.26}$$

From these two equations, the values I_min and I_max are defined by equation 2.27.

$$\begin{aligned} I_{min} &= \min\big(I_R^-, I_R^+, I_R(x_R)\big) \\ I_{max} &= \max\big(I_R^-, I_R^+, I_R(x_R)\big) \end{aligned} \tag{2.27}$$

From these two values defined in equation 2.27,

$$\bar{d}(x_L, x_R, I_L, I_R) = \max\big(0, I_L(x_L) - I_{max}, I_{min} - I_L(x_L)\big) \tag{2.28}$$

The corresponding equation for $\bar{d}(x_L, x_R, I_L, I_R)$ can be arrived at in a similar manner, thus completing the symmetry and giving the expression for the Birchfield – Tomasi metric $d(x_L, x_R)$. These computations add only marginal time to that taken by the absolute intensity difference, usually no more than 10%.

2.3.3.2 Directional Cost Aggregation

Heiko Hirschmuller [29] developed an equation for aggregating the directional cost, which is also adopted in the SGBM algorithm. The cost $S(p, d)$, captured in equation

2.29, is computed by aggregating the cost from r different directions at each pixel p.

$$S(\mathbf{p}, d) = \sum_{r} L_r(\mathbf{p}, d) \tag{2.29}$$

The 1D minimum cost path $L_r(\mathbf{p}, d)$ of the pixel p at disparity d for any given direction r is calculated recursively using equation 2.30.

$$\begin{aligned} L_r(\mathbf{p}, d) = C(\mathbf{p}, d) \\ + \min(L_r(\mathbf{p\text{-}r}, d), L_r(\mathbf{p\text{-}r}, d-1) + P_1, L_r(\mathbf{p\text{-}r}, d+1) \\ + P_1, \min_i L_r(\mathbf{p\text{-}r}, i) + P_2) - \min_k L_r(\mathbf{p\text{-}r}, k) \end{aligned} \tag{2.30}$$

Where $L_r(\mathbf{p}, d)$ is the current cost of pixel p and disparity d in direction r, $C(\mathbf{p}, d)$ is the matching cost at pixel p and disparity d, $L_r(\mathbf{p\text{-}r}, d-1)$ is the previous cost pixel in direction r at disparity $d-1$, $L_r(\mathbf{p\text{-}r}, d+1)$ is the previous cost pixel in direction r at disparity $d+1$, $\min_i L_r(\mathbf{p\text{-}r}, i)$ is the minimum cost of pixel in direction r for the previous computation and P_1, P_2 the penalty for discontinuity.

The cost along each path is independent, and the total cost is the aggregation of the costs in all directions. Figure 2.8, adapted from [27], is an illustration of the directional cost aggregation. The block that minimises this aggregate cost is considered the correct match and is the one used to compute the disparity, in pixel space, of the scene captured in its pixels. This process is repeated until all the blocks in the images have been matched and their disparities calculated. The end result is a disparity map ready for post-processing.

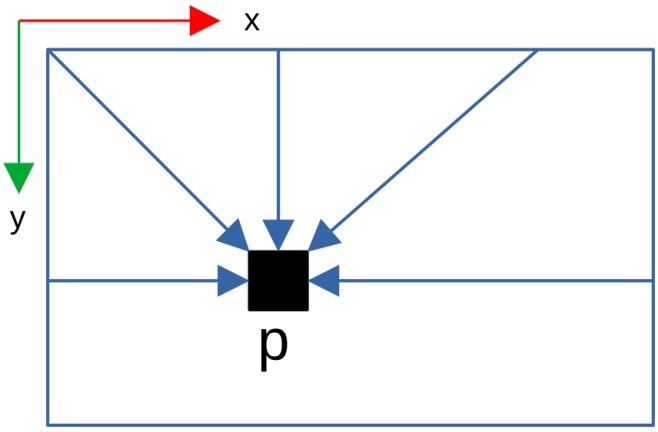


Figure 2.8: Cost aggregation in 5 directions

2.3.3.3 Post-Processing

During this step, the corresponding pixels with minimum cost are identified and their indices recorded, quadratic interpolation is performed at the sub-pixel level to fill occluded pixels, and uniqueness constraints are applied to ensure consistency and reliability of the obtained disparity image [27]. The resulting disparity map exhibits much smoother transitions between dark and bright regions compared to the one before post-processing.

2.3.4 Similarity Metrics

In simple stereo, the corresponding points in the two images are not vertically displaced relative to one another. The searching algorithm performs template matching along a horizontal scan line rather than the entire image. Szeliski discusses the common similarity metrics applied in stereo matching at great length in his textbook [24]. Other than the Birchfield-Tomasi metric discussed in section 2.3.3, the other three commonly applied ones highlighted by Szeliski are:

(i) Sum of absolute differences (SAD): For a pixel $(k, l) \in L$:

$$SAD_{min(k,l)} = \sum_{i,j \in T} |E_l(i, j) - E_r(i + k, j + l)| \tag{2.31}$$

(ii) Sum of Squared Differences (SSD):

$$SSD_{min(k,l)} = \sum_{i,j \in T} |E_l(i, j) - E_r(i + k, j + l)|^2 \tag{2.32}$$

(iii) Normalized Cross Correlation (NCC):

$$NCC_{max(k,l)} = \frac{\sum_{i,j \in T} E_l(i, j) \sum_{i,j \in T} E_r(i + k, j + l)}{\sum_{i,j \in T} E_l(i, j)^2 \sum_{i,j \in T} E_r(i + k, j + l)^2} \tag{2.33}$$

2.3.5 Extracting Measurements from Reconstructed Scenes

In order to extract real world measurements from reconstructed scenes, it is necessary to develop mathematical equations to analyse these scenes. This is true for all techniques of scene reconstruction and is not limited to stereoscopic vision. Other techniques of scene reconstruction include Structure from Motion (SfM), which involves generating 3D point clouds of a scene from multiple overlapping images, and laser scanning, which involves generating similar point clouds using a laser scanner.

For a single approach to parameter estimation, e.g., SfM, there are different procedures followed in estimating the value of the DBH. Where point clouds are involved, it is common to find researchers using cylinder-fitting for estimation [30]. Liu *et al.* criticised this approach noting that cylinders would not fit perfectly to the point cloud because tree trunks have a degree of taper that makes their diameters

decrease with height [31]. In place of it, they proposed a multi-height diameter estimation approach combined with an outlier detection algorithm. The diameter at each height was estimated by fitting a circle to the point cloud using the well-known RANSAC algorithm. The least squares circle fitting algorithm has also been used by other researchers for DBH estimation [32]. Trairattanapa *et al.* provide an insightful comparison between least squares and the Random Hough Transform circle fitting approaches for estimating the DBH and found the latter to be superior in performance with a precision of 92.67% compared to the former's 87.81%. Although circle fitting seems to perform well for point clouds, it is not suitable for use with disparity images because the density of points on these images is too low. Instead, for techniques implementing stereoscopic vision, the use of multiple-view geometry is dominant [17], [18], [33]. Furthermore, most studies implementing stereoscopic vision have used spherical images for tree attribute estimation. The use of ordinary perspective images is comparatively rare [17].

2.4 Non-Contact Methods of Forest Inventory

2.4.1 Tree Attributes of Interest

TH and DBH are the two most measured tree attributes during forest inventory exercises [19]. Their high significance derives from the fact they have the strongest correlation with other variables of interest such as biomass, timber volume, and basal area [8], [9], [19], [34]. These parameters are also relevant in applications outside forestry. For example, it is known that crop height and crown diameter are useful for inferring the progress of growth, and hence the yield of coffee bushes [22]. Biomass,

which is an index of the carbon storage capability of a forest stand, is computed using allometric equations which have DBH and TH as the two most important inputs [4], [35], [36]. While estimating the amount of biomass contained in mangrove trees in Kenya, Kairo *et al.* [37] found that the DBH and tree height were the most important predictors of tree biomass. A similar conclusion was reached by Jones *et al.* [34] while conducting a comparable study in Australia. Adhikari *et al.* [38], who also carried out their research in Kenya, used DBH to model the value of above-ground biomass. Needless to say, TH and DBH are indispensable to forest health assessment [13].

Both DBH and TH can be measured at the individual tree level or for multiple trees at once. Stand-level statistics are then derived from the data obtained using those two approaches. Individual tree-level measurements are those where tree attributes in a sample plot are taken tree by tree [17], [39], [40]. Manual measurement methods are the most common techniques used in this approach. They include using callipers and diameter tapes to measure diameters and hypsometers for tree heights. Relascopes, which are more sophisticated pieces of equipment, are also used to measure tree diameters, heights and stand basal area [17]. In the last two decades, researchers have developed non-contact methods to measure TH and DBH based on the principles of photogrammetry and laser scanning (LS) [14], [15], [40]. LS, which makes use of light detection and ranging (LiDAR), is used for obtaining attributes of multiple trees simultaneously [41]. Photogrammetry, on the other hand, can be used in both multiple-tree and single-tree inventory. These methods will be discussed in further detail in the sections that follow.

2.4.2 Comparison Between Multiple and Single Tree Inventory

Many benefits accrue from estimating tree attributes at the individual tree level. One prominent advantage is accuracy. Data obtained using manual techniques are used as a reference for gauging the performance of other methods [17]. The accuracy levels of figures derived from single-tree photogrammetry are comparatively higher than in the case of multiple-tree inventory [15], [39], [41], [42]. In photogrammetry, experiments performed by numerous researchers show that accuracy levels of tree attribute data decrease with distance from the camera [15], [19]. Thus, attributes of trees closer to the camera are measured more accurately than those further away. A major concern with single tree measurements is the speed of data acquisition because forest inventories require collecting attributes of hundreds, perhaps thousands, of trees [5], [17], [41], [43]. While this is true, recent research shows that photogrammetric systems can be designed which take as short as less than ten seconds to acquire tree attributes. Eliopoulos *et al.* reported an acquisition time of 3 seconds when measuring DBH using stereoscopic photogrammetry [17]. If durations like these become typical, this issue will not be as concerning.

Multiple tree inventory also offers diverse advantages. One very well-known gain of this approach is the significant reduction in data acquisition time [5]. There are three main approaches to achieving this: panoramic imagery, taking multiple overlapping images or using TLS [5], [44], [45]. The approach of taking multiple photographs aims to generate a point cloud and is known as structure from motion (SfM). Although these approaches exhibit decreased accuracy with distance, researchers have discovered that this method generates fairly accurate results when smaller sample plots are taken.

Accuracies are also high for forests with low tree density [5]. Common devices used here are LiDAR cameras (also called laser scanners), hemispherical and spherical cameras and normal digital cameras. Because LiDAR cameras are very expensive, they can be substituted with consumer-grade spherical cameras, which are readily available and cheaper than the former by two to three orders of magnitude [5], [18].

Tree occlusion and overlapping stems are one of the major demerits of multiple-tree inventory using spherical imagery. This is especially true in dense forests [5]. For any single location of the camera, some trees will be either fully occluded or partially occluded (overlapping stems) [18]. For obvious reasons, high measurement errors are reported in the case of partially occluded stems. Besides occlusion, current spherical cameras generate images by stitching together two hemispherical images, a process that is not error-free [5]. A work-around that has been implemented by some researchers is that of taking images from three points with angular displacements of 120° relative to the plot centre [5], [18]. While this reduces the number of occlusions, it does not eliminate the problem.

Another demerit of acquiring attributes of multiple trees at once is the computational complexity involved. When laser scanning or SfM is used for acquisition, the process begins with the generation of point clouds [5]. Point cloud generation from images requires huge processing resources and is often achieved using expensive proprietary software such as AgiSoft PhotoScan [20]–[22] and Pix4D [34]. Depending on the number of images to be processed, the interval between acquisition and data availability is hours at best and could even stretch to days [22]. All this is compounded by the fact that when the attributes are finally obtained, accuracy levels are rather low for trees further

away from the camera [39], [45]. Although accuracy levels for LiDAR point clouds are very high, laser scanners (LiDAR cameras) are extremely expensive [46]. For forest agencies and private forest owners, the cost involved in this approach far outweighs the benefits.

2.4.3 Working with 3D Point Clouds

Closely related to stereoscopic vision is another type of photogrammetry known as structure from motion (SfM). It is a range imaging method for generating the 3D structure of a scene from a sequence of 2D images. The process of 3D reconstruction using this technique applies the principles of multiple-view geometry. Unlike binocular vision where the depth information is obtained from only two images, SfM reconstructs a scene from numerous overlapping images [19], [20], [39]. The number of images used varies depending on the application and can number even in the hundreds. The SfM pipeline, as outlined in [23] and [24], happens in the following steps:

(i) Points of interest (features) are identified in all the images and described using computer vision algorithms.

(ii) The features are matched across images following the rules of epipolar geometry (the geometry of stereo vision).

(iii) The set of all corresponding point pairs is used to reconstruct a 3D structure in the form of a point cloud.

Correspondence matching between images in the sequence is only possible due to the overlap imposed during acquisition. A large overlap ($\approx$ 80%) is needed between

photographs to generate accurate point clouds [22]. Image acquisition can be done by humans using a stop-and-go approach or by using Unmanned Aerial Vehicles (UAVs) with pre-planned flight paths, speed, and imaging intervals [39].

Many photogrammetric studies using SfM point out that this approach has varied weaknesses. Foremost of these is the computational complexity involved in finding correspondences and generating the point clouds [19]. While stereoscopic vision also involves finding matching image features, this step is magnified many times over in SfM because matching has to be done over a long sequence of images [19], [39]. The processing time depends on the resolution and number of images, computing hardware, the algorithm used, and the parameters used in the software. To retain high levels of accuracy, one needs more images of high resolution, which explains why computational complexity is not a minor concern [39]. Additionally, correspondence matching and point cloud generation are often achieved using very expensive software such as AgiSoft PhotoScan [20]–[22]. Compounded with this issue is the problem of reducing accuracy with distance from the camera. Some researchers have even challenged the wisdom of using point clouds when measurements can be done directly from image pairs [18]. These demerits are glaring.

Structure from Motion has been applied in numerous studies to collect forest inventory data such as tree heights, trunk diameters, canopy area, basal area, and tree density [19]–[21], [34], [39], [45], [47]. Liang *et al.* employed this technique to measure the DBH values of 25 trees and achieved a root mean square error (RMSE) of 2.39 cm. Using a Samsung NX 300 hand-held camera of focal length 16 mm, they captured 973 images and then generated a point cloud using AgiSoft PhotoScan, a process which

took 32 hours [39]. A similar experiment was performed by Bayati *et al.* to create a 3D reconstruction of an uneven-aged forest and reported an RMSE of 2.17 cm. The latter study successfully reconstructed the forest with a remarkable 42 images only, which explains their much shorter processing time of 8.85 hours [19]. This might be an indication that a larger number of images is not necessarily advantageous. More recently, Jones *et al.* used UAV imagery and SfM with an overlap of 80% to measure tree heights and reported an R^2 value of 0.98 for tree heights measured from point clouds [34]. In a Malaysian mangrove reserve, the authors of [20] created a canopy height model (CHM) from a point cloud generated from drone imagery and used it to estimate tree heights. The reported results were p values of 0.375 and $\leq$0.0001 for the median tree height in two different zones in the reserve. They attributed the poor accuracies in the latter zone to large disparities in tree. One SfM study focused on tree parameters of forest stands under regeneration in Norway [21]. The variables of interest for young trees in this study were the mean tree height and the tree density. Their results were an RMSE of 0.56 – 0.97 m and RMSE% (RMSE as a percentage of the mean) of 23.6 - 32.6% in mean tree height, and an RMSE of 185 – 413 trees per hectare and RMSE% of 13% - 28.4% in tree density. The accuracy values were measured relative to data collected using ALS. When comparing the performance of ALS and SfM-derived data against field-measured data for measuring individual tree, Guerra-Hernandez *et al.* found Pearson correlation coefficients of 0.66 (ALS) and 0.61(SfM) [45]. They concluded that SfM-generated point clouds are as good as ALS-derived ones.

Closely related, yet in contrast to the structure from motion (SfM) is the Simultaneous

Localization and Mapping (SLAM) algorithm. Whereas SfM recovers the depth of the scene from a set of 2D images, SLAM calculates the position and orientation of the camera relative to its surroundings while mapping the environment at the same time. To successfully achieve this, successive frames should contain an adequate number of features to be tracked. Fan *et al.* employed SLAM to estimate the tree positions, DBH, and tree height running on a special purpose-built mobile phone [40]. Their study reported RMSEs of 1.26 cm (6.39%), and 1.11 m (7.43%) in DBH and tree height. Chen *et al.* implemented a form of SLAM using Lidar Odometry and Mapping (LOAM) for forest mapping. They reported a mean error of 1.67 cm (0.67 in) in DBH measurements [16]. Although SLAM has great potential, it requires the use of expensive equipment and is also computationally expensive.

LS is known to generate very accurate point clouds. Because of this advantage, this technique is sometimes used to generate reference data for evaluating the performance of other methods [45]. Measurement with LS is done based on the time-of-flight and phase difference of the light pulses generated by the scanner. Each point on the scene is described by x, y, and z coordinates, colour and reflectance values assigned by a specialist software when the point cloud is generated [48], [49]. LS can be deployed at the ground level (terrestrial, or TLS, and mobile, or MLS) or in the air (aerial, or ALS). For TLS, the scanner is mounted on a tripod stand while ALS scanners are mounted on UAVs or manned aircrafts. Guerra-Hernandez *et al.* used an aircraft-mounted laser scanner to collect ALS data for comparison with SfM data. They found comparable performances for the two methods except for individual tree crown identification. In this case, inaccuracies resulted from the fact that tree crowns often overlap making

morphological separation impossible [45]. Cabo *et al.* developed an algorithm for automatically extracting tree heights and DBH from ALS point clouds. They reported RMSEs of 0.8 – 1.3 cm and 0.3 – 0.7 m in tree height and DBH respectively [41]. The authors of [46] used a UAV-mounted LiDAR to generate point cloud data, which was then analysed using various machine learning models to predict tree attributes. Their results were a relative RMSE of $\leq$9% for DBH measurements. Although ALS and TLS studies usually report highly accurate results, the cost of laser scanners ranges from \$15,000 to upwards of \$100,000. Additional equipment required during data acquisition includes an inertial measurement unit (IMU) for registering the locations of the laser scanner in all cases, and an aircraft or UAV for ALS [21], [45]. These further inflate the cost of laser scanning, thus making the astronomical cost involved a major impediment to wider deployments.

In working with 3D point clouds (TLS or SfM), there are various approaches used in estimating the value of DBH. Marzulli *et al.* [30] developed an automated cylinder-fitting approach for approximating the diameter of 30 cm sections along the length of the trunks. They reported an RMSE and bias of 1.9 cm and -1.99% respectively. Liu *et al.* used the RANSAC circle fitting algorithm to estimate the trunk diameter at different heights of the tree and, where necessary, interpolation to find the DBH [31]. They used outlier detection to eliminate diameters that did not belong to the trunk and reported RMSE and MAE of 1.97 cm and 1.65 cm (6.31%) respectively. McGlade *et al.* [32] also used a circle-fitting approach applied on 20 cm sections of stems between 1.2 m and 1.4 m above the ground from point clouds generated by the Microsoft Azure Kinect depth sensor. They achieved an RMSE of 8.43 cm and a bias

of 2.05 cm. The work of Trairattanapa *et al.* [50] provides insights into the comparison between least squares circle fitting and Random Hough transform (RHT) circle fitting in DBH extraction. In their study, the RHT approach outperformed the least squares approach by achieving 92.67% precision compared to the latter's 87.81%.

2.4.4 Use of Reference Objects

To reduce the complexity of algorithms used for parameter estimation in photogrammetry and laser scanning (LS), other researchers have resorted to the use of reference objects or benchmarks in images. In this method, a reference object of known dimensions is captured together with the trees of interest and then used as a reference for calculating tree parameters. Collazos *et al.* [51] developed a photogrammetric system for estimating the DBH, TH and other parameters from SfM point clouds based on a scale factor. Good performance was obtained for DBH measurement while the other parameters showed poorer performance. Piermattei *et al.* [47] used multiple targets in a field plot to help estimate the scale of their SfM point clouds. As already stated, they reported a minimum RMSE of 1.21 cm, a minimum bias of -0.71 cm in DBH estimation, and a tree detection rate of 91%. Byrne *et al.* [52] used a laser rangefinder to measure the distance between the camera and the tree of interest. They reported absolute errors in DBH estimation of $\leq$2.54 cm for all trees measured. Han and Wang [53] used markers placed along the tree trunk to extract tree heights from images with a mean relative error of 3.62%.

2.4.5 Towards Techniques Requiring Less Training and Expertise

Many researchers argue that the development of forest inventory techniques that will require less training and expertise to use will be beneficial. In their study, Piermattei *et al.* [47] used commercial software to generate SfM point clouds and point out how the use of such software automates the point cloud generation step, thus making it require less expertise. A similar approach was also implemented by Marzulli *et al.* [30] who developed an automated approach for extracting tree stems and point clouds and fitting them with cylinders to estimate DBH values. St-Onge *et al.* [54] used proprietary software to process stereophotogrammetric satellite imagery and extract Canopy Height Models with RMSE of between 2.53 m and 2.95 m. Such automation in techniques reduces the effort and training required to extract parameters post-acquisition. One of the main reasons why SfM and Laser Scanning techniques require extensive training and expertise to use is that point cloud processing involves significant human interaction. The interactive steps in the processing pipeline demand a high level of mastery of skills for working with point clouds [19], [31], [47], [54]. Since speed is one of the desirable traits of a forest inventory technique, human interaction should be kept at a minimum. These two methods inherently consume a lot of time during data acquisition and processing making real-time estimation difficult to achieve [55]. Two key features that must be achieved to make real-time estimation possible are reduced computation time and zero or minimal human interaction [56]–[58]. Research studies that have reported either real-time parameter estimation or potential for real-time estimation include [17], [40], and [55]. The first two of these studies used stereoscopic vision to estimate tree attributes. This method

is fast because it involves finding correspondences between only two images, thus making it attractive for research in tree parameter estimation.

2.4.6 Stereoscopic Photogrammetry

2.4.6.1 An Overview

Stereoscopic photogrammetry has drawn a lot of interest from researchers because of its simplicity, speed, and accuracy [17]. This method operates on the principle of stereopsis, which gives human beings the visual ability to perceive the world in three dimensions. Human beings can tell whether an object is close to them or further away based on visual information derived from their eyes [24]. Similarly, two cameras horizontally displaced relative to each other generate a stereo image pair, which is then used to compute the depth of points in the scene [59]. This kind of computer vision is known as binocular vision. The technique, in varied forms, has been applied to generate depth maps [59], measure tree attributes [14], [17], [60]–[62], estimate object distances [63], and aid traffic safety [64]. The theoretical foundations of stereopsis will be discussed in detail in section 2.3.

2.4.6.2 Use Cases in Forest Inventory

Studies abound in which scientists have used stereoscopic vision in varied forms to estimate forest attributes such as TH, DBH and tree position. Perng *et al.* [18] began by deriving the specific geometry of a vertically displaced pair of spherical images and used their results to compute DBH and tree distance from the camera. Their study reported RMSEs in DBH of 3.74 cm – 22.2 cm which correspond to an RMSPE (root mean squared percentage error) of 10.23% - 36.81% for tree distances of between 0 and

15 m. The lowest MEs were observed for closer trees and vice versa. The values of RMSE and RMSPE were very low when DBH was $\leq$20 cm and grew very large for DBH $\geq$ 20 cm [18]. A similar experiment was conducted by Wang *et al.* who found an RMSE of 2.4$\pm$1.6 cm [5]. Like other similar studies, these two experienced the problem of decreasing accuracy with distance, which is a limitation of all forms of light sensors [5], [18]. Perng *et al.* attributed part of their larger errors encountered to weaknesses in their derivation which included several sequential approximations that they thought to have propagated the errors [18]. Another source of error in DBH computation was thought to be the image stitching process which is not completely error-free.

It is possible to obtain stereo image pairs using spaceborne systems. One interesting study used advanced stereo-radargrammetric data and image processing using SAR satellite data in X-band. By taking a triplet of images from space and using them to compute canopy height models of terrestrial forests, the authors reported a canopy height estimation of over 20%. Despite the worrying underestimation, they achieved a forest classification accuracy of 90% using a maximum likelihood classifier [60]. Elsewhere, using stereoscopic systems with very high resolutions mounted on separate space satellites, the authors of [65] extracted the canopy heights of forests on mountainous landscapes with RMSE of 1.15 m and R^2 of 0.92, and those on flat landscapes with RMSE of 0.99 m and R^2 of 0.96. These two studies provide insights into the capabilities of spaceborne stereoscopic systems.

Research on terrestrial stereoscopic systems with narrow fields of view is limited compared to those based on panoramas. Few recent studies have opened up this subspace for future work to be done. A photogrammetric system developed by

Forsman *et al.* comprised a rig holding five cameras such that no subset of three cameras was collinear. Its baseline varied between 57 cm and 113 cm so that objects within the range of 1 – 20 m could be measured accurately. The design with five cameras was chosen to allow tolerance for occlusions. This study reported RMSEs of 2.8 – 9.5 cm in the DBH estimates [14]. Although the five cameras were all of the expensive professional types, the cost of the setup could be reduced significantly by choosing consumer-grade cameras. The work done by Eliopoulos *et al.* features the Intel RealSense D435 purpose-built stereo camera purchased at $179. The acquisition time for 40 trees (imaged three times each at varying distances from the trunk) was 30 minutes. The reported RMSEs were 1.28 cm (at 1 m), 1.47 cm (at 3 m) and 2.57 cm (at 5 m). Aside from the good accuracies obtained, this study achieved real-time extraction of DBH values. This is the first study to have focused on imaging a single tree and obtaining the DBH using stereoscopic vision.

Available literature shows that the use of self-made stereo systems dominates research in the area in comparison to purpose-built stereo cameras. This may be attributed to the fact that ready-made sets do not allow for adjusting the camera baseline, and thus limit the range within which depths can be measured accurately. In [18], the authors reported that after testing the performance of their system using baseline values of 30 cm and 60 cm. The 30 cm baseline provided the lowest RMSE values for measuring DBH. Wang *et al.* [5] used a spherical stereo camera with a baseline of 1 m to acquire stereo image pairs. A similar baseline value was used by Olaveri-Monreal *et al.* for building a real-time distance measurement system for a traffic safety system [64]. No purpose-built stereo with such long baselines is currently available in the market. The

setup used by Forsman *et al.* comprised five cameras mounted on a rig with three baseline values depending on the camera pair permutation [14]. They reported that their design provided robustness against occlusions. One advantage of self-built stereo sets is that they are cheaper to create. An example is an affordable system designed by Strotov *et al.* using two USB Logitech C270 HD web cameras known to cost about $70 each [63].

2.4.6.3 Other Use Cases

Stereoscopic photogrammetry use cases stretch beyond measuring forest attributes. One such use case is in an Intelligent Transportation System (ITS) where Olaveri-Monreal *et al.* [64] used a stereo system to measure the distance to a trailing vehicle and provide unobtrusive warning whenever tailgating behaviour was observed. Strotov *et al.* built a similar system to measure object distances in real-time with edge processing [63]. The algorithm used for the latter application provided accurate distance measurements when the target object was far enough from the camera and the size of the object of interest in the image was not too large to be stored for processing. Marino *et al.* applied stereo vision to perform an analysis of forest crowns to analyse continuity in canopy fuels and generate data useful for wildfire prevention [43]. These studies provide insights into the potential of stereoscopic photogrammetry.

2.4.7 Challenges Presented by Real Forest Setups

When studying the performance of non-contact methods in real forest setups, researchers encounter diverse challenges that should be addressed to make these methods useable in real forests. Some of these challenges are unique to the technique

being studied. One such issue is the natural taper in tree trunks that makes the use of cylinder fitting algorithms inappropriate in point clouds for diameter estimation [31]. Although not as common in planted forests, tree trunk bifurcation (tree forks) presents a difficulty in DBH estimation since diameters for both trunks must be measured [51]. This becomes especially difficult when the individual trunks are not delineated at the breast height. Non-straight tree trunks also present a unique challenge to extraction algorithms [51]. In photogrammetric methods which involve heavy use of digital image processing techniques, image segmentation is notoriously challenging because colour and texture information is not enough to distinguish object types [66]–[68]. Object occlusion is a ubiquitous problem in forest settings for all types of optical sensors such as cameras and laser scanners. Occlusion happens when tree crowns or trunks overlap or when tree branches and trunks are sheltered by leaves, thus making edge finding very difficult or impossible [52], [69], [70]. In urban parks, the background is further complicated by the presence of cars, poles, and people [57]. As much as possible, the parameter extraction algorithms should be robust to these challenges [52].

2.5 Research Gaps

Considering past works in developing non-contact methods for measuring individual tree attributes, a vast majority of studies have focused on laser scanning and SfM. From the literature review already presented, both techniques are computationally heavy and time-consuming with laser scanning also requiring very expensive equipment. By contrast, stereoscopic vision is low cost, can be optimized for fast computation, and can yield highly accurate distance measurements. From the foregoing literature survey,

five studies have leveraged stereoscopic vision in various forms. These studies are those carried out by Wang *et al.* [5], [44], Eliopoulos *et al.* [17], Perng *et al* [18], and Malekabadi *et al.* [55]. Given the limited research on stereoscopic vision systems for forest inventory, this study undertakes to explore how this technology can be used to estimate tree attributes. It is worth devoting some time to explaining how the research presented in the subsequent sections improves on or differs from past works and fills some existing research gaps in non-contact forest inventory.

The following gaps have been identified:

(i) There seems to be no non-contact forest inventory study based on stereoscopic vision that has presented algorithms that may be applied in tree attribute estimation. This thesis presents algorithms to be used in estimating three tree attributes from disparity maps derived from stereoscopic images.

(ii) The bulk of research work done on non-contact forest inventory techniques has focused on SfM and laser scanning. Limited research has been done on stereoscopic vision despite the advantages it offers in cost, speed, accuracy, and computational simplicity. Of the five studies mentioned, only two studies derived measurements from normal perspective images while the other two relied on spherical images. This study seeks to build on the previous studies by working with perspective images.

(iii) Based on the numerous literature surveyed, only the study conducted by Malekabadi *et al.* has implemented the use of disparity maps to extract individual tree parameters but only tested their technique on only two artificial trees. In addition, the geometry for calculating those parameters was not

presented in their study or any other study. Further, neither this study nor any other seems to have proposed a method for addressing the effect of the presence of anomalies (large sudden changes in pixel intensities) in disparity maps on distance estimation. This research study seeks to develop algorithms that can extract these parameters from actual trees. These algorithms rely on the geometric framework for deriving the location of the breast height and calculating the DBH, TH, and CD based on information from disparity maps. The geometry may be generalized to apply to any other application involving measurement. This thesis also builds on the work of Malekabadi *et al.* by applying a median filter on regions of interest to overcome the effect of these anomalies.

(iv) The work of Eliopoulos *et al.* proposes a method for computing the DBH of a single tree from image pairs taken by a stereo camera equipped with an infrared (IR) projector. The IR projector aids the correspondence algorithm by adding texture to the images. This study explores the use of only RGB images with no IR information overlay to extract disparity maps. It also builds on the work of Eliopoulos *et al.* by adding two parameters i.e., CD and TH, to be measured.

(v) The studies carried out by Wang *et al.*, and Perng *et al.* all make use of spherical image pairs. To compute the parameters, correspondences between images are identified manually and then spherical geometry is applied. Further, their calculations required prior knowledge of the camera's height above the ground (i.e., relative to the trunk base). The method proposed in this study explores the use of perspective images and then extracts tree parameters from

the computed disparity image. It also enables the estimation of these parameters without prior knowledge of the camera's location.

Based on the gaps identified above, a method for extracting the diameter at breast height, crown diameter, and the height of a tree from a pair of horizontally displaced perspective images has been formulated, tested, and validated. Images are segmented using the well-known grab cut algorithm and results are saved as segmentation masks. With the masks in place, the proposed technique applies geometric computational algorithms to extract the tree attributes of interest from the disparity maps without human interaction.

CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This chapter presents the proposed method for extracting individual tree biophysical parameters from stereoscopic image pairs. The contents include information about the study area, materials and software used, proposed technique, and performance evaluation.

3.2 Study Area

The trees whose measurements were taken during this study are found within the recreation parks of the Dedan Kimathi University of Technology located 6km to the north of Nyeri town in central Kenya. The parks are found within reasonable walking distance from coordinates (0°23'47.8"S, 36°57'38.7"E). This location is chosen because it has desirable characteristics such as sparse tree distribution with no overlapping tree crowns. From observation, the trees in these parks vary greatly in height and diameter. They are also at least 5 m apart and their crowns are all clearly delineated, which make it suitable for validating the proposed method. Additionally, parts of the parks have a flat terrain while others slope gently with no steep terrains present at all. The most common tree species spotted in the parks are Silky Oak (*Grevillea robusta*), Mexican Cypress (*Cupressus lusitanica*), and Willow-leaf Podocarp (*Podocarpus salignus*). Some photos captured at the proposed study location are provided in Appendix I.

3.3 Materials and Data Acquisition

3.3.1 Materials

3.3.1.1 Nvidia Jetson Nano 2GB Developer Kit

This is a small but powerful single-board computer with the ability to run multiple artificial intelligence applications such as image classification and segmentation, and speech processing. For this study, this kit was used to provide the computing power to perform image processing for all the images of trees captured in the parks. Some of its technical specifications of interest are [71]:

- GPU: 128-core NVIDIA Maxwell™
- CPU: Quad-core ARM® A57 @ 1.43 GHz
- Memory: 2 GB 64-bit LPDDR4 25.6 GB/s
- Connectivity: Gigabit Ethernet, 802.11ac wireless
- Camera: 1x MIPI CSI-2 connector
- Display: HDMI
- USB: 1x USB 3.0 Type A,2x USB 2.0 Type A, USB 2.0 Micro-B

A photo of the kit can be found in Appendix I.

3.3.1.2 Cameras

For this study, a pair of Logitech C270 HD 720p USB cameras were used to build a stereo camera. This type of camera was chosen because it is low-cost and easy to control programmatically. The specifications are as follows [72]:

- Maximum resolution: 720 pixels
- Frame rate: 30 fps
- Diagonal field of view (dFoV): 55°
- Mega pixels: 0.9

A photo of the camera can be found in Appendix I.

3.3.1.3 Camera Rig

To hold the two cameras together, a rig comprising two interlocking halves was designed on Autodesk® AutoCAD and 3D-printed. Each half contained a hollow section where a part of the camera would rest while the other half completes the assembly to secure the camera firmly in place. The two cameras were separated by a horizontal distance (known as baseline) of approximately 12.9 cm. This rig made it possible to move the setup from place to place without needing to recalibrate the stereo camera because it could tolerate mechanical disturbance caused by shaking during transportation. Refer to Appendix I for a photo of the rig.

3.3.1.4 Bosch GLM20 Laser Rangefinder

The ground truth tree heights were measured using the Bosch GLM20 laser rangefinder. This device has a range of 0.15 m to 20 m and an accuracy of ±3 mm making it suitable for measuring trees with heights of less than 6 m. It consists of a laser beam generator and a receiver and estimates the distance as the product of the time of flight of the beam and the speed of light in air. Other specifications of the device are [73]:

- Dimensions: $100 \times 36 \times 23mm$
- Weight: 130g
- Operating temperature: -10°C to 40°C
- Laser beam colour: Red
- Power source: $(2 \times 1.5V)$ AAA batteries

3.3.2 Camera Calibration

Once the stereo camera assembly had been prepared, the setup was calibrated to learn the internal parameters of each camera as well as their relative positions and orientations. The calibration takes place in two stages:

3.3.2.1 Single camera calibration

Each camera was calibrated to establish the relationship between the positions of objects in the real-world coordinate frame and their mapping onto the 2D image plane. This step yields the intrinsic and extrinsic camera matrices presented in equations 2.8 and 2.9 respectively. The product of these two matrices yields the projection matrix P, a single linear transformation that maps a point in the scene to one on the image plane. To determine these matrices, OpenCV's implementation of the method developed by Zhang [74] was followed. The procedure is as follows:

(i) At least 20 images of a checkerboard pattern (figure 3.1) of dimensions 7×10 boxes were captured by the camera at different angles and distances.

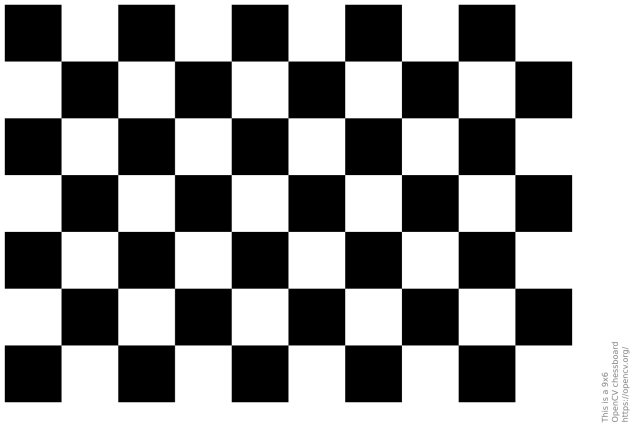


Figure 3.1: Checkerboard pattern for camera calibration

(ii) The correspondences between the points of interest in the actual 3D scene (object points) and their images (image points) were found using OpenCV.

(iii) For each correspondence point, the mapping between object and image points was expressed as a matrix equation $\bar{u} = P\bar{x}$ where $\bar{u}$ is a 3×1 vector of the image point homogeneous coordinates, $\bar{x}$ a 4×1 vector of the object point homogeneous coordinates, and P the $4{\times}3$ projection matrix to be found.

(iv) The matrix equation in (iii) above was expanded into two linear equations relating the x and y ordinates of the image point to the product on the right-hand side. This was done for every corresponding point.

(v) All the equations in (iv) above were rearranged to form a second matrix equation of the form $A\mathbf{P} = 0$ where P is the projection matrix rearranged into a 12×1 vector and A is an $n \times 12$ matrix where n is the number of corresponding points.

(vi) The Marquardt – Levenberg nonlinear optimization algorithm [75] was applied in solving for P in the equation $A\mathbf{P} = 0$.

(vii) The intrinsic and extrinsic matrices were decoupled from the projection matrix P using the QR decomposition approach.

In step (vi) above, the Marquardt – Levenberg algorithm [75] was chosen for non-linear optimization in camera calibration because it offers the advantages of efficiency, robustness to noise, and fast convergence [76]. Other optimisation techniques which can be used al alternatives here include genetic algorithms, particle swarm optimization, and simulated annealing [77]–[79].

3.3.2.2 Stereo Camera Calibration

This second stage of calibration is performed to establish the geometric relationship between the two cameras in the real-world coordinate frame. The most important outputs of this process are the fundamental (F) and the essential (E) matrices, from which the translation (T_X) and rotation (R) matrices can be obtained, and the rectification transforms H_l and H_r. These two matrices establish the association between the stereo-image pairs. The theory for calculating the E and F matrices is presented in sections 2.3.3.1 and 2.3.3.2 The following procedure was followed:

(i) At least 20 image pairs of the checkerboard pattern boxes were captured by the stereo camera at different angles and distances.

(ii) The correspondences between the points of interest in the left image x_l and the right image x_r were found using OpenCV.

(iii) Equation 2.20 was solved to find the fundamental matrix F from which the essential matrix E was obtained.

(iv) E was then decomposed into T_X and R using singular value decomposition.

(v) A pair of matching projective transformations H_l and H_r satisfying equations 2.21 and 2.23 were computed.

3.3.3 Ground Truth Data Acquisition

The ground truth data acquired during field exercises were used as the benchmark against which the values predicted by our method were compared. These measurements were obtained as follows:

3.3.3.1 Diameter at Breast Height

The common practice in DBH measurement is to use measuring tapes wrapped around the tree at the breast height level (1.3 m from the trunk base), or callipers to measure two perpendicular diameters and find their average. In this study, a measuring tape was wrapped around the tree trunk at the breast height (BH) and the circumference read off to the nearest millimetre. This circumference was then divided by π to find the reference DBH value. Figure 3.2 illustrates how the measurement was carried out.

Figure 3.2: Reference DBH measurement using a measuring tape

3.3.3.2 Crown Diameter

The CD was measured by projecting the crown edges to the ground and measuring the length between the edges along the axis of the tree crown to the nearest centimetre. For trees with smaller crowns, a rope was held tightly so that it spanned the whole diameter of the crown along one axis. The crown edges were then projected onto the rope and the distance between the projections was measured using a measuring tape. This latter case is illustrated by the image in figure 3.3.

Figure 3.3: Measuring the CD using a rope and measuring tape

3.3.3.3 Tree Height

The process of measuring the TH was comparatively more complicated and required the use of a laser rangefinder held at a fixed position. The Bosch GLM20 laser rangefinder was used to measure three lengths: length to tree base (d_b), length to the treetop (d_t), and horizontal distance to the tree trunk (d_h). Figure 3.4 illustrates how the laser rangefinder is used to measure tree heights. From these three lengths, the value of TH was calculated using equation 3.1.

$$TH = \sqrt{d_b^2 - d_h^2} + \sqrt{d_t^2 - d_h^2} \tag{3.1}$$

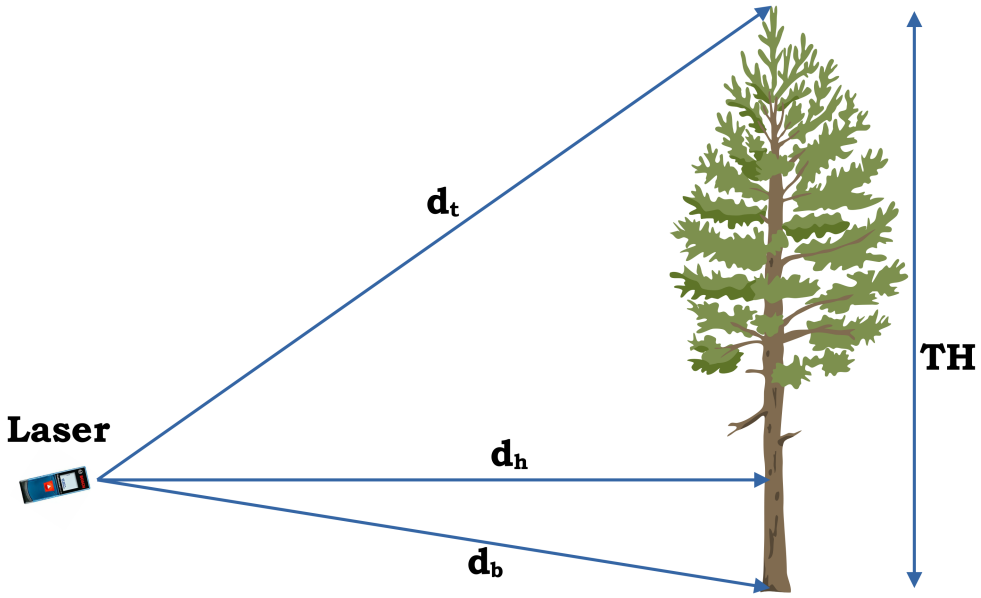


Figure 3.4: Using the laser rangefinder to measure TH

3.3.4 Terrestrial Stereo Photogrammetric Survey

Images were all taken in late August and early September 2022 using the stereo camera built for this research by assembling two Logitech web cameras. The weather conditions on those days were generally calm and cloudy. In all cases, the stereo camera was held steadily and its height above the ground was not measured as it is not required by the system developed in this study.

3.3.4.1 Trunk Images for BH Location and DBH Estimation

For DBH extraction, the stereo camera was positioned at 5 m and 8 m away from the tree trunk and two perpendicular views of the trunk were acquired for each distance. The capturing of these perpendicular views was informed by the industry best

practice applied in DBH measurement which usually requires finding an average of two perpendicular diameters. Therefore, 4 images were taken for each tree making a total of 80 image pairs for the 20 trees studied. In addition to this, 5 trees were selected to study the performance of the method developed in this research in estimating the position of the breast height in an image. Each of these trees was wrapped with white tape at the breast height for later comparison to the estimated location. One image was taken for each tree at intervals of 1 m from 5 m to 9 m from the trunk making a total of 25 images for all trees.

3.3.4.2 Full Tree Images for TH and CD Estimation

Given the small baseline of 12.9 cm for the stereo camera and the low resolution of the camerasused in this study, its ability to estimate distances accurately was limited to a range of approximately 12 m. The maximum tree height that could be captured within this distance turned out to be approximately 6 m. For TH and CD extraction, the camera was held at arbitrary distances from the tree and a total of 10 images of 10 trees were captured.

3.4 The Proposed Methodology

The flowchart for the proposed approach is shown in figure 3.5. It captures the full pipeline for extracting tree attributes from stereoscopic images beginning with image segmentation. The images captured in the field are segmented and the edges of the resulting mask are smoothened using mathematical morphology. The dimensions of the structuring elements were 40 $\times$ 1 (elongated) for tree trunk masks and square 5 $\times$

5 for full tree masks. With the segmentation complete, a full disparity map was derived from the left and right images using a stereo correspondence algorithm. The camera parameters obtained during calibration are necessary to compute this disparity map. To obtain a segmented disparity containing only the foreground object (tree trunk or full tree), the mask was applied on the full disparity map. This resulting disparity map was then fed into the proposed parameter extraction algorithm to estimate the tree attributes. Finally, the extracted parameters were compared to ground truth values to assess the performance of the proposed technique. Each of these steps is described in detail in the subsequent sections.

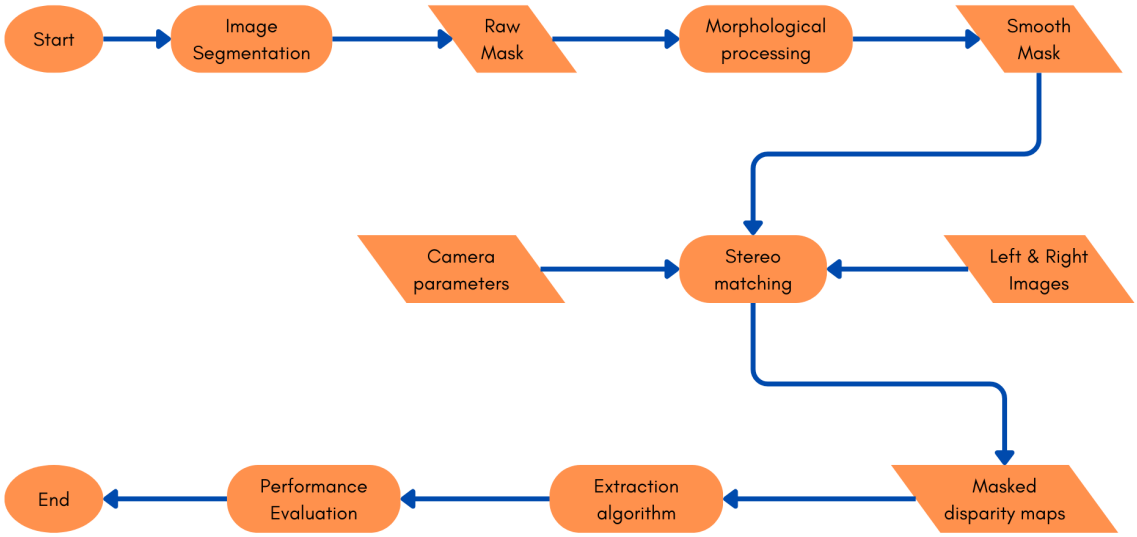


Figure 3.5: Flowchart for the proposed methodology

3.4.1 Full Tree and Trunk Segmentation

Before performing any segmentation on the images, contrast enhancement was done on images to make segmentation more accurate. This was achieved using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method [80]. The

enhanced images were then segmented using the Grab Cut method. This algorithm begins by drawing a bounding box around the region (tree or trunk) of interest. Every pixel outside of this bounding box becomes a background while those inside the box are used to define a colour distribution model using a Gaussian Mixture Model (GMM). In this GMM, each pixel is labelled as either background, foreground or unknown. After this, the algorithm creates a graph initialized with weights. The vertices V are the pixels in the box while the weights W represent the similarity between those pixels. Two nodes, the sink and source nodes are added to the graph to represent the background and foreground respectively [81].

A Min-Cut algorithm is then run on this graph to perform an initial segmentation into two subsets A and B representing the foreground and background respectively. This algorithm defines the minimum cut as the smallest sum of weights whose removal would sever the sink from the source. The user aids the process by continuously labelling more pixels as background and foreground and the Min-Cut (equation 3.2) is performed iteratively until the segmentation is complete [81].

$$mincut(A, B) = min\left(\sum_{u \in A, v \in B} w(u, v)\right) \tag{3.2}$$

For every image pair, only the left image was segmented as it was used as the base image for disparity map computation. The segmentation results obtained in this step were saved as binary masks for later use in segmenting the disparity maps.

3.4.2 Image Rectification

The image pairs captured by the stereo camera in the field do not have the corresponding points aligned horizontally. To correct this and facilitate efficient computation of the disparity map, the projective transformations H_l and H_r obtained during stereo calibration were used to warp the left and the right images respectively.

3.4.3 Disparity Map Computation

Disparity maps were generated using OpenCV's implementation of the SGBM algorithm. The inputs to the algorithm were the rectified left and right images and the camera calibration parameters. For this study, the left image was used as the base image during disparity computation. In implementing this algorithm, the following were the major model parameters used:

- Block size: 11 pixels
- Minimum number of disparities: 0 pixels
- Number of disparities: 128 pixels
- Uniqueness ratio: 10
- Left-Right Consistency check: 5

Two important parameters of the SGBM algorithm are usually tweaked in operation. These are the disparity levels and the number of directions for matching cost aggregation. The number of disparity levels defines the domain for matching. Given D disparity levels, then for a pixel in the left image, a search is performed from a

block of D pixels in the right image [28]. As a result, the first D columns of the left image are usually not included in the search because they are missing in the right image. Increasing the value of D decreases the minimum distance whose depth can be detected and vice versa. The cost function (equation 2.30) was used to calculate the cost of matching a block of pixels in the left image to a block in the right image [27]. The number of directions determines how the total cost is aggregated making the process one of directional cost calculation. By default, the algorithm implements 5 directions of cost aggregation (figure 2.8), although up to 8 directions can be implemented albeit with longer computation times [28]. The default setting was used in this research.

Post-processing steps applied to the resulting disparity image include peak removal by invalidating small segments, image normalization to fit the intensities within the range of 0 to 255, and median filtering to remove other irregularities. The disparity map generated at this point is a 3D map of the whole scene captured by the image. Since the rectified left image was warped relative to the segmentation mask created by the method described in section 3.4.1, applying it directly on the full disparity image would not produce the desired results. To retain only the foreground (tree trunk or full tree), this segmentation mask was first warped using the transformation H_l and then applied to the full disparity map to yield a segmented one.

3.4.4 Disparity – Distance Relationship

In performing accurate distance estimation using the disparity map, a consistent relationship between distance and disparity was established. Disparity maps for a single tree were computed from image pairs taken in the range of 3.0 m to 12.6 m

from the trunk at intervals of 30 cm. The disparities of the trunk's mid-point at the middle row of the disparity map were recorded. A distance-disparity scatter plot was then obtained followed by curve fitting using MATLAB's implementation of the Trust-Region-Reflective Least Squares algorithm [82].

From two-view geometry, the depth of a point on the scene varies inversely to its disparity. An inverse relationship can be modelled using an asymptotic curve such that of a rational polynomial. To find the polynomial that best fit the data, various equations where the order of the denominator was greater than that of the numerator by one were fit using MATLAB. The polynomial that resulted in the lowest least squares error was taken as the best fit.

Given a disparity image D_m with the left image as the base, the real-world cartesian coordinates of any point b on the image in the camera coordinate frame is the vector:

$$\begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix} = \frac{b}{u_l - u_r} \left(\begin{bmatrix} u_l \\ v_l \\ f_x \end{bmatrix} - \begin{bmatrix} o_x \\ o_y \\ 0 \end{bmatrix} \right) \tag{3.3}$$

where z_b is the real-world depth of the point in the disparity image as calculated using the equation of the fitted curve, u_l and v_l are its x and y ordinates, (o_x, o_y) is the location of the principal point of the image, and $(u_l - u_r)$ is the disparity of that point, which is related to the grayscale intensity value from the disparity image. Its value can be obtained from z_b which is already known. Once these coordinates are known, they are used in the algorithms presented in section 3.4.7.

3.4.5 Geometry Derivation

The algorithms for extracting the three parameters of interest rely heavily on the scene geometry and camera specifications. In the figures presented in the following sections, the geometry for calculating the tree attributes of interest in this study is presented. The real-world distances between the stereo camera and various points on the tree are obtained from the segmented disparity maps. The greyscale values in the disparity images are resolved into distances using equation 4.1.

3.4.5.1 Working with Camera Fields of View

The vertical and horizontal fields of view (vFoV and hFoV) of the camera are first derived before presenting the geometry for calculating each tree parameter. The diagonal field of view (dFoV) of the camera used in this study is 55°. The camera sensor has dimensions of 1280 × 720 pixels. Hypothetically, this sensor would have $\sqrt{1280^2 + 720^2} = 1468.60$ pixels along its diagonal. Beginning with the diagonal field of view, the horizontal (dFoV) and vertical fields of view (vFoV) can be calculated by applying the sine rule to the corresponding sides using equations 3.4 and 3.5.

$$\frac{1468.60px}{tan(\frac{dFoV}{2})} = \frac{1280px}{tan(\frac{hFoV}{2})} = \frac{720px}{tan(\frac{vFoV}{2})}$$

$$\frac{1468.60}{tan(27.5)} = \frac{1280}{tan(\frac{hFoV}{2})} = \frac{720}{tan(\frac{vFoV}{2})}$$

$$hFoV = 2 \times tan^{-1}\left(\frac{1280\ tan(27.5)}{1468.60}\right) = 48.81° \tag{3.4}$$

$$vFoV = 2 \times tan^{-1}\left(\frac{720\ tan(27.5)}{1468.60}\right) = 28.63° \tag{3.5}$$

3.4.5.2 Estimating the Location of the Breast Height

The geometry for locating the breast height in a disparity image is described in figure 3.6. Point C is the position of the stereo camera when the image is taken and h_f is the breast height i.e., $h_f = 1.3m$. The breast height spans s_h pixels vertically along the height of the disparity image and the goal in estimating the location of the breast height is to find this value. From equation 4.1 and the disparity (greyscale intensity) of the trunk base (point B in figure 3.7) read off from the segmented disparity image, the coordinates of point B in the camera's coordinate frame are found as (x_b, y_b, z_b) using equation 3.3.

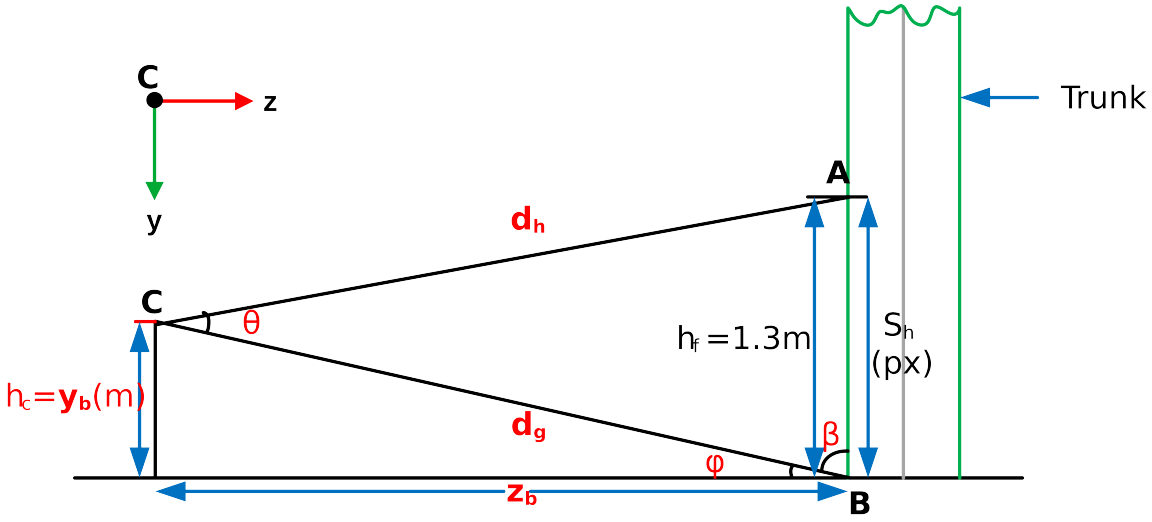


Figure 3.6: Geometry for Estimating the BH Location

The angle β is calculated using trigonometry as:

$$\beta = tan^{-1}\left(\frac{z_b}{y_b}\right) \tag{3.6}$$

The two sides (AB and BC) of ΔABC and the angle (β) between them are now known. Applying the cosine rule to ΔABC:

$$d_h^2 = h_f^2 + d_g^2 - 2h_f d_g cos\beta \tag{3.7}$$

But $h_f = 1.3m$ as already stated. The value of d_g is found from the coordinates of B.

$$\begin{aligned} d_g &= \sqrt{x_b^2 + y_b^2 + z_b^2} \\ d_h^2 &= 1.3^2 + d_g^2 - 2.6 d_g cos\beta \\ \therefore d_h &= \sqrt{1.69 + d_g^2 - 2.6 d_g cos\beta} \end{aligned} \tag{3.8}$$

To find angle θ, the sine rule is applied to ΔABC:

$$\frac{sin\beta}{d_h} = \frac{sin\theta}{1.3} \quad \Rightarrow \quad \theta = sin^{-1}\left(\frac{1.3\ sin\beta}{d_h}\right) \tag{3.9}$$

The camera has a vertical field of view $vFoV = 28.63°$ (equation 3.5). The next step is to find the angle subtended by the breast height, which also spans s_h pixels vertically in the disparity map. Applying proportionality:

$$\frac{720}{tan\left(\frac{vFoV}{2}\right)} = \frac{s_h}{tan\left(\frac{\theta}{2}\right)} \quad \Rightarrow \quad \frac{720}{tan\ 14.31} = \frac{s_h}{tan\left(\frac{\theta}{2}\right)} \tag{3.10}$$

The only unknown in the above equation is s_h, which is easily found by making it the subject as shown by equation 3.11.

$$s_h(px) = \frac{720 \times tan\left(\frac{\theta}{2}\right)}{tan\left(\frac{vFoV}{2}\right)} = \frac{720 \times tan\left(\frac{\theta}{2}\right)}{tan\ 14.31} = 2822.61 \times\ tan\left(\frac{\theta}{2}\right)\ px \tag{3.11}$$

Now the span in pixels of the trunk between its base and breast height is known, which

means that the location of the breast height has been found.

3.4.5.3 Estimating the DBH

The geometry for estimating the DBH is obtained from figure 3.7, which is adapted from Eliopoulos *et al.*'s work [17]. The horizontal field of view of the camera is hFoV=48.81°(equation 3.4).

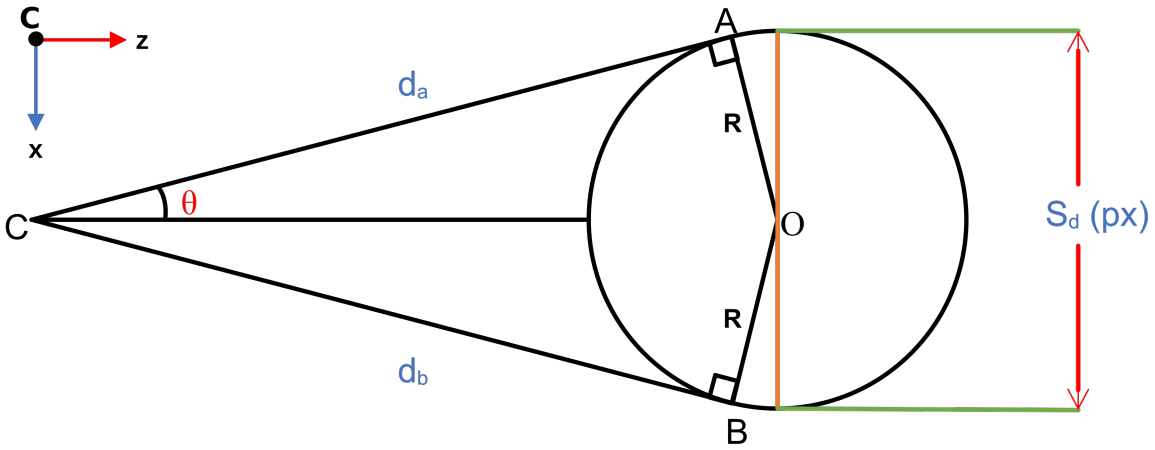


Figure 3.7: Geometry for Estimating the DBH.

The DBH spans s_d pixels in the disparity image. Therefore:

$$\frac{1280\ px}{tan\left(\frac{hFoV}{2}\right)} = \frac{s_h\ px}{tan\ \theta} \quad \Rightarrow \quad \frac{1280}{tan\ 24.41} = \frac{s_d}{tan\ \theta}$$

$$1280 \times tan\ \theta = s_d \times tan\ 24.41$$

$$\theta = tan^{-1}\left(\frac{s_d \times tan\ 24.41}{1280}\right) = tan^{-1}(s_d \times 3.546 \times 10^{-4})$$

$$R = d_a \times tan\ \theta \tag{3.12}$$

The distance d_a is obtained by resolving the disparity of point A into distance using

equation 4.1. The DBH is twice the value of R.

$$DBH = 2 \times R \tag{3.13}$$

3.4.5.4 TH Estimation

To estimate the TH, figure 3.8 is used. The lengths d_T and d_B are determined by finding the coordinates of points T and B respectively using equation 3.3 and calculating their respective distances from the camera.

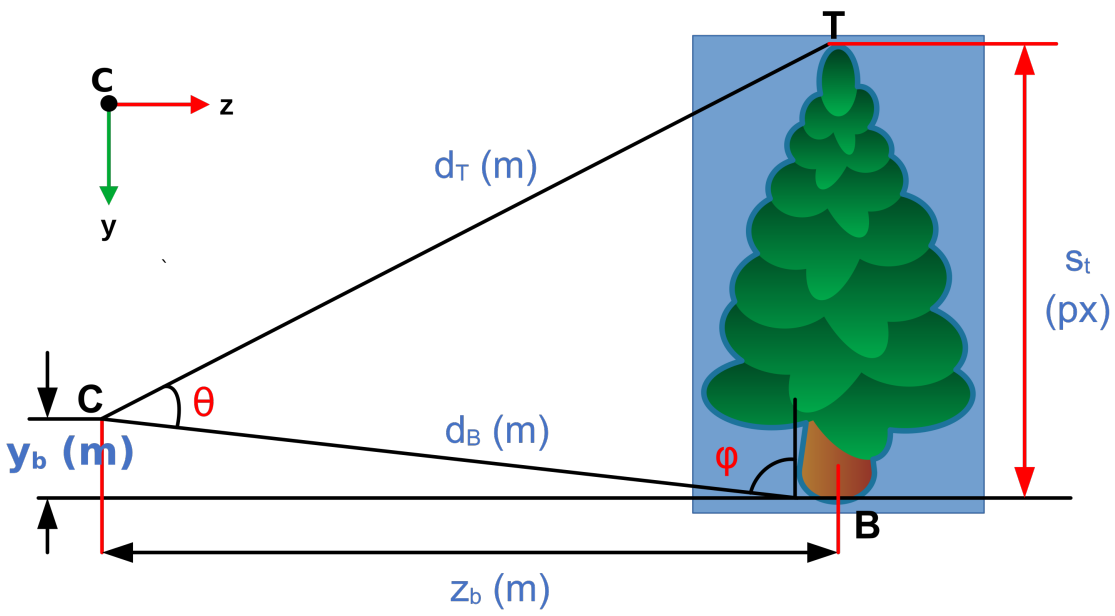


Figure 3.8: Geometry for Estimating the TH

It follows that:

$$\frac{720\ px}{tan\left(\frac{vFoV}{2}\right)} = \frac{s_t\ px}{tan\ \frac{\theta}{2}} \quad \Rightarrow \quad \frac{720}{tan\ 14.31} = \frac{s_t}{tan\ \frac{\theta}{2}}$$

$$\frac{\theta}{2} = tan^{-1}\left(\frac{s_t \times tan\ 14.31}{720}\right) \quad \Rightarrow \quad \theta = 2 \times tan^{-1}(s_t \times 3.543 \times 10^{-4})$$

$$\phi = tan^{-1}(\frac{z_B}{y_B})$$

$$\frac{\sin\theta}{TB} = \frac{\sin\phi}{d_T} \tag{3.14}$$

The value of tree height, which is equal to the length TB in figure 3.8, is found as in equation 3.15.

$$TB = TH = \frac{d_T \sin\theta}{\sin\phi} \tag{3.15}$$

3.4.5.5 CD Estimation

To compute the crown diameter, figure 3.9 is used. The geometry is like that of the DBH.

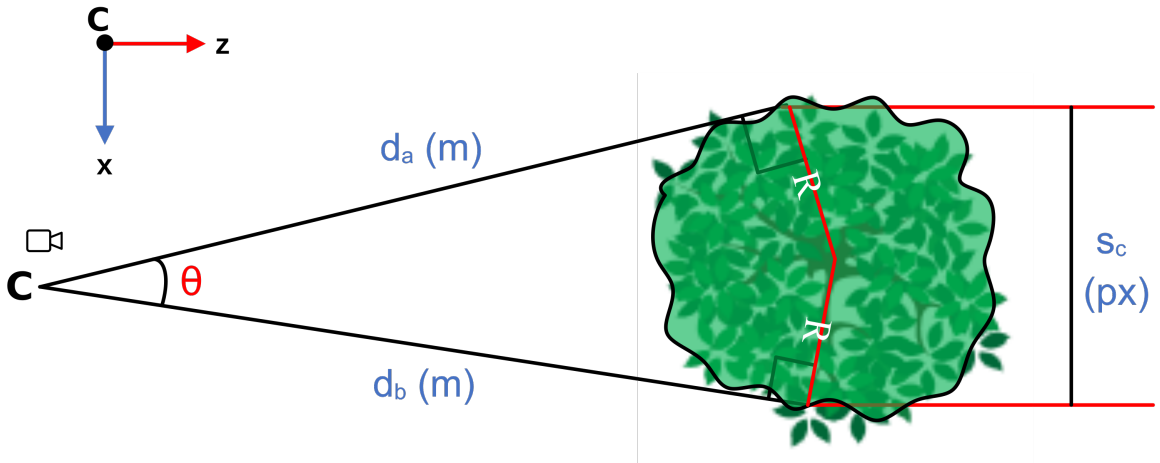


Figure 3.9: Geometry for Estimating the CD

Beginning with the $hFoV = 48.81°$:

$$\frac{1280\ px}{\tan\left(\frac{hFoV}{2}\right)} = \frac{s_c\ px}{\tan\left(\frac{\theta}{2}\right)}$$
$$\frac{1280}{\tan 24.41} = \frac{s_c}{\tan\left(\frac{\theta}{2}\right)} \quad \Rightarrow \quad \theta = 2 \times tan^{-1}(s_c \times 3.546 \times 10^{-4})$$
$$R = d_a \times \tan\left(\frac{\theta}{2}\right) \tag{3.16}$$

The length d_a is obtained from the coordinates of the crown edge pixel found using equations 3.3 and 4.1.

$$CD = 2R = 2 \times d_a \times \tan\left(\frac{\theta}{2}\right) \tag{3.17}$$

3.4.6 Finding the Pixels of Interest

The pixels of interest are the edge pixels at the base, top, crown extremes, and breast height of the tree. The pixels of interest for TH and CD estimation are the easiest to locate since they're simply the top and base pixels, and crown edge pixels respectively. In the case of the DBH, the method presented for locating the breast height (section 3.3.5.2) will be applied and the edge pixels at that height subsequently be identified. In a good disparity image, the pixels of interest have greyscale intensity values equal to or very close to those of pixels in their neighbourhood. However, anomalies (pixels with very different greyscale intensities) may be present in some cases. Since the real-world distance to a pixel is based on its greyscale intensity, these anomalies can cause errors in distance estimation and consequently in parameter extraction.

To eliminate the effect of these anomalies, regions of interest are formed as sets of

pixels surrounding the pixels of interest (figure 3.10). In the case of base and top pixels, the region of interest includes all pixels in the bottom 20 rows and the top 20 rows of the object pixels in the disparity map. In the case of the DBH, all non-zero pixels 5 rows above and 5 rows below the breast height location constitute the region of interest. When deciding on the size of these regions of interest, consideration was made for those pixels whose real-world distances from the camera were approximately equal. The disparity of the pixel of interest was taken to be the median of pixel intensities in the region of interest.

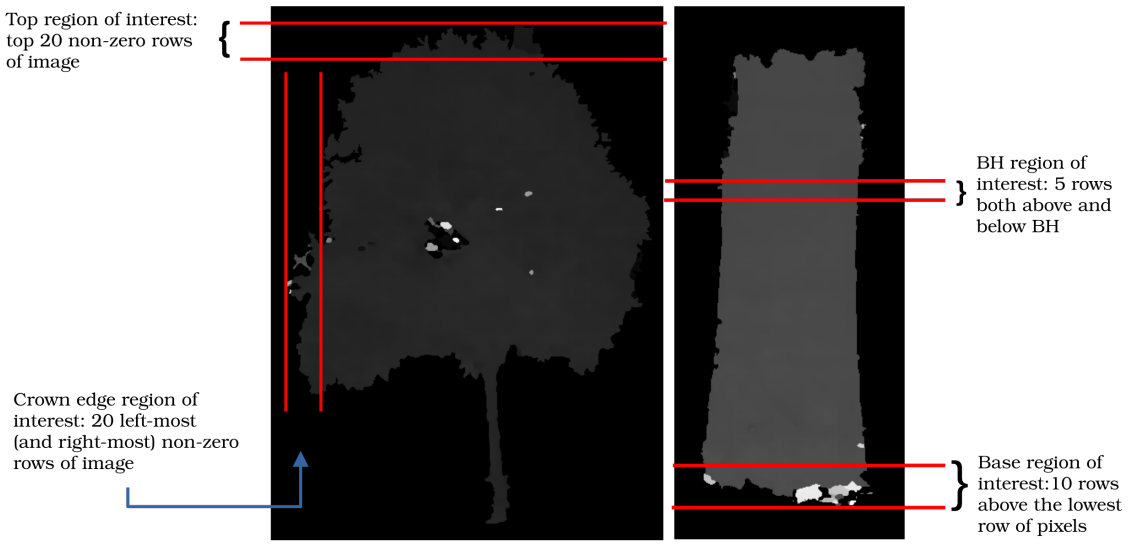


Figure 3.10: Forming regions of interest

3.4.7 Parameter Extraction Algorithms

Computational geometric algorithms were written in Python to extract the DBH, CD and TH using the approach presented in section 3.4.5.

3.4.7.1 Algorithm for Pixel of Interest Identification

Input:	Left and right images, mask, calibration parameters
Output:	Vector of intensities of pixels of interest
Step 1:	Compute disparity image
Step 2:	Mask the disparity image
Step 3:	Find the nonzero pixels in the masked disparity image
Step 4:	Calculate the row of the breast height
Step 5:	Identify the base, breast height, top, and crown regions of interest
Step 6:	Find median intensities in the base, breast height, top, and crown regions of interest
Step 7:	Save intensities in a vector

3.4.7.2 Algorithm for DBH Extraction

Input:	Left & right images, mask, calibration parameters, reference DBH
Output:	DBH value, Error in estimation
Step 1:	Compute disparity image
Step 2:	Mask the disparity image
Step 3:	Find the nonzero pixels in the masked disparity image
Step 4:	Calculate the row of the breast height
Step 5:	Identify the base and breast height regions of interest
Step 6:	Find median intensities in the base (b) and breast height (h) regions of interest
Step 7:	Find coordinates of b as (x_h, y_h, z_h) and distance to b as $d_b = \sqrt{(x_b^2 + y_b^2 + z_b^2)}$
Step 8:	Find coordinates of h as (x_b, y_b, z_b) and distance to h as $d_a = \sqrt{(x_h^2 + y_h^2 + z_h^2)}$
Step 9:	Find angle θ_1 subtended by breast height at the camera
Step 10:	Find number of nonzero pixels at the breast height
Step 11:	Find angle θ_2 subtended by breast height at the camera
Step 12:	Calculate DBH from θ_2 and d_a, and compute error in estimation

3.4.7.3 Algorithm for CD Extraction

Input:	Left & right images, mask, calibration parameters, reference CD
Output:	CD value, Error in estimation
Step 1:	Compute disparity image
Step 2:	Mask the disparity image
Step 3:	Find the nonzero pixels in the masked disparity image
Step 4:	Identify the crown region of interest
Step 5:	Find median intensities in the crown (c) regions of interest
Step 6:	Find coordinates of c as (x_c, y_c, z_c) and distance to c as $d_c = \sqrt{(x_c^2 + y_c^2 + z_c^2)}$
Step 7:	Find angle θ subtended by the crown at the camera
Step 8:	Calculate CD from θ and d_c, and compute error in estimation

3.4.7.4 Algorithm for TH Extraction

Input:	Left & right images, mask, calibration parameters, reference TH
Output:	TH value, Error in estimation
Step 1:	Compute disparity image
Step 2:	Mask the disparity image
Step 3:	Find the nonzero pixels in the masked disparity image
Step 4:	Identify the base and top regions of interest
Step 5:	Find median intensities in the base (b) and top (t) regions of interest
Step 6:	Find coordinates of b as (x_h, y_h, z_h) and distance to b as $d_b = \sqrt{(x_b^2 + y_b^2 + z_b^2)}$
Step 7:	Find coordinates of t as (x_t, y_t, z_t) and distance to t as $d_t = \sqrt{(x_t^2 + y_t^2 + z_t^2)}$
Step 8:	Find complementary of angle ϕ subtended at tree base by camera height
Step 9:	Find angle θ subtended by tree height at the camera
Step 10:	Calculate TH from θ, ϕ and d_t, and compute error in estimation

3.4.8 Performance Evaluation

The values of the DBH, tree height and crown diameter measured using manual methods were used as the reference data. The performance metrics used to evaluate the performance of the technique developed in this study include the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Bias, Root Mean Square Error (RMSE), and Coefficient of Determination (R^2). Yin and Wang [83] recommend that for assessing the accuracy in tree parameter estimation, these metrics are the most important for performance evaluation. The Wilcoxon Signed Ranked Test was also used to find the similarity between the DBH estimates at 5 m and 8 m from the tree.

3.4.8.1 Mean Absolute Error (MAE)

This value is the arithmetic mean of the absolute errors in the predicted measurements. It is calculated using equation 3.18.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x_{ir}| \tag{3.18}$$

where x_i is the measured value, x_{ir} the reference value, and n the number of trees measured.

3.4.8.2 Mean Absolute Percentage Error (MAPE)

This is the arithmetic mean of the absolute errors in prediction expressed as percentages of the reference values. It is calculated as shown by equation 3.19.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - x_{ir}|}{x_{ir}} \times 100\% \tag{3.19}$$

3.4.8.3 Bias

This is the mean difference between the attributes' ground truth values and those estimated by the proposed approach. It is calculated as shown by equation 3.20.

$$bias = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{ir}) \tag{3.20}$$

3.4.8.4 Root Mean Square Error (RMSE)

This metric represents the root of the second moment of the differences between reference values and the predicted values (predicted by the proposed method) [83]. It is calculated using equation 3.21.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_{ir})^2}{n}} \tag{3.21}$$

3.4.8.5 Coefficient of Determination(R^2)

This refers to the proportion of variation in the reference values that can be explained by variations in the estimated measurements [83]. It is given by equation 3.22.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - x_{ir})^2}{\sum_{i=1}^{n}(x_i - \bar{x_i})^2} \tag{3.22}$$

where $\bar{x_i}$ is the mean of the reference values.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the results obtained after applying the proposed approach to extract individual tree attributes from stereoscopic image pairs. The measurements obtained are compared to the ground truth values and the overall performance is juxtaposed with those achieved in other studies.

4.2 Results

4.2.1 Disparity – Distance Relationship

The result of fitting a curve to the distance-disparity points was a rational polynomial with a numerator of order 2 and a denominator of order 3 as shown in equation 4.1. In this equation, x is the pixel intensity and y is the distance in meters. Its plot is shown in figure 4.1.

$$y = \frac{346x^2 - 116.7x - 1.961}{x^3 - 5.863x^2 + 47.24x + 487.6} \tag{4.1}$$

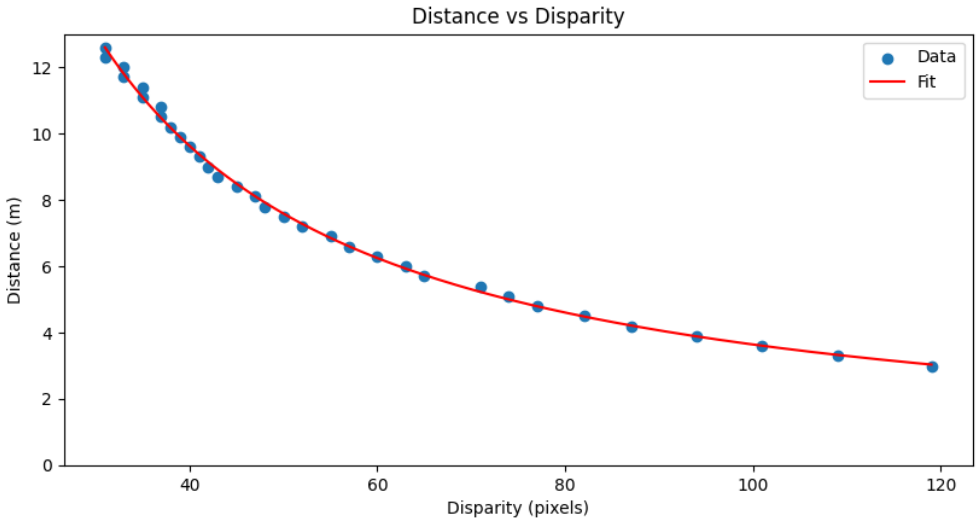


Figure 4.1: Relationship between disparity and distance

4.2.2 Image Acquisition and Processing Time

Having already written Python code for acquiring the images, each image was captured instantly. Most of the time was spent moving between trees. A significant amount of time was spent ensuring that the two views at each distance from the tree were perpendicular and in movement between trees to change scenes. On average, it took approximately a minute and 15 seconds to capture the 4 image pairs required for estimating the DBH of each tree, adding up to 25 minutes for the 20 trees (80 image pairs) studied. When acquiring the images used for estimating the location of the breast height, it took about a minute to capture 5 images (at different distances) of every tree, which makes a total of 5 minutes for the 5 trees (25 image pairs) considered in this aspect of the study. Thus, it took 30 minutes in all to acquire 105 image pairs of tree trunks for studying the performance of the proposed technique in estimating the location of the breast height as well as the value of the DBH. The 10 image pairs of full trees needed for TH and CD extraction were taken in approximately 3 minutes.

The process of deriving a disparity image from every image pair took 3.1 seconds on average. Applying the segmentation algorithm described in section 3.4.1 took approximately one minute for every image. For every image pair acquired by the stereo camera, only the left image was segmented to obtain a mask that was later applied to the disparity map. With the segmented disparity map in place, extracting each of the parameters took less than a second. Table 4.1 shows a summarized breakdown of the time taken for image acquisition and processing.

Table 4.1: Time taken to acquire and process images

Activity	**Time Taken**
Trunk image acquisition (105 image pairs): distance measurement, perpendicularity checking, image capturing	30 minutes
Full tree image acquisition (10 image pairs)	3 minutes
Disparity image computation (1 image pair): enhancement, correspondence matching, post-processing, masking	3.1 seconds
Image segmentation (left image only)	1 minute
DBH extraction	$<$ 1 second
TH extraction	$<$ 1 second
CD extraction	$<$ 1 second

4.2.3 Image Segmentation

The accuracy of parameter estimation, especially the DBH, in our proposed method is very dependent on the successful separation of the target tree from the background.

Applying the Grab Cut algorithm on the image gave impressive results as it removed most of the background pixels (figures 4.2b and 4.3b). For some trees, the trunks were so different from the background pixels that very minimal interaction was needed during segmentation. However, for most trees, some trunk sections matched the backgrounds, and a few iterations were still not sufficient to perform a good segmentation. The morphological operations discussed in section 4.1.3 were used to correct the results of the segmentation. Figures 4.2 and 4.3 show the images of a full tree and a tree trunk respectively through the segmentation pipeline beginning with applying the grab cut algorithm on the original image to obtain the mask. The segmentation masks in figures 4.2b and 4.3b was applied to the full disparity map in figures 4.2c and 4.3c to obtain the masked disparity maps in part figures 4.2d and 4.3d.

Figure 4.2: Image of a full tree through the segmentation pipeline.

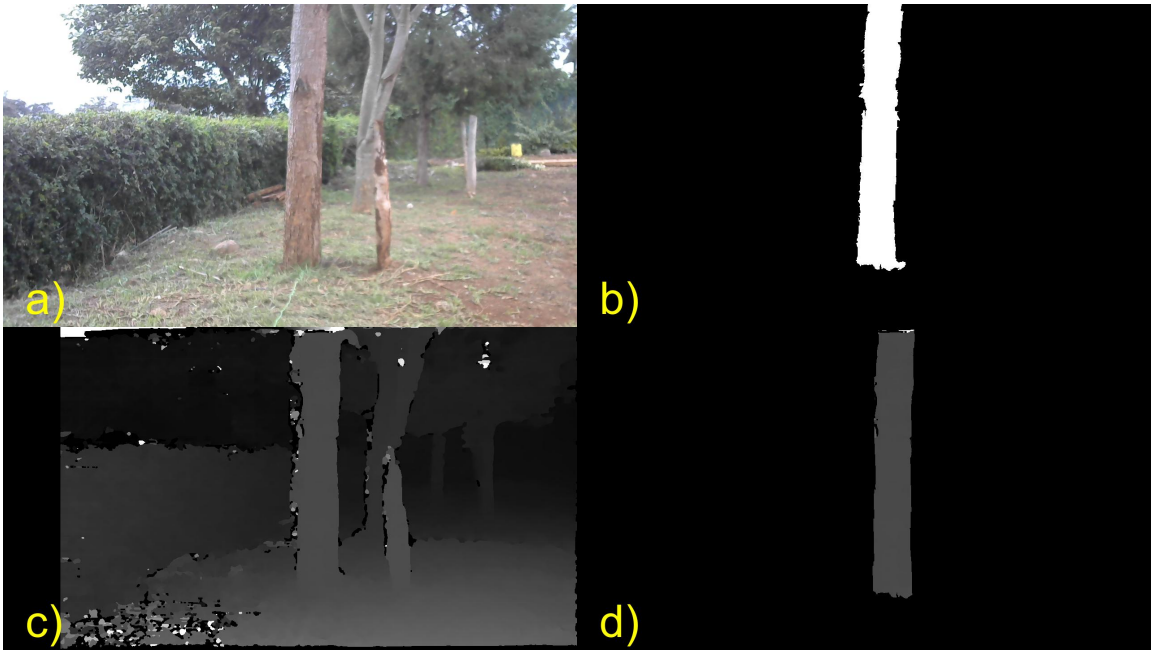


Figure 4.3: Image of a tree trunk through the segmentation pipeline.

4.2.4 Morphological operations

In applying morphological opening and closing, the vertical nature of tree trunks was taken into consideration. Most segmentation results contained small blobs (white pixels) extending out of the trunk and holes (black pixels) within the trunk. The effect of blobs located at the breast height is to increase the number of pixels spanned by the breast height while that of holes is to reduce this number below the actual value. Since the extraction algorithm uses the number of pixels for DBH extraction, increasing this number above the actual leads to an overestimation of the DBH while reducing it does the opposite. These blobs and holes were removed by applying morphological closing followed by opening. The results were smooth edges on the segmentation masks as seen in Figure 4.4. The image on the left has tiny blobs along the tree trunk edges while that on the right has smooth edges.

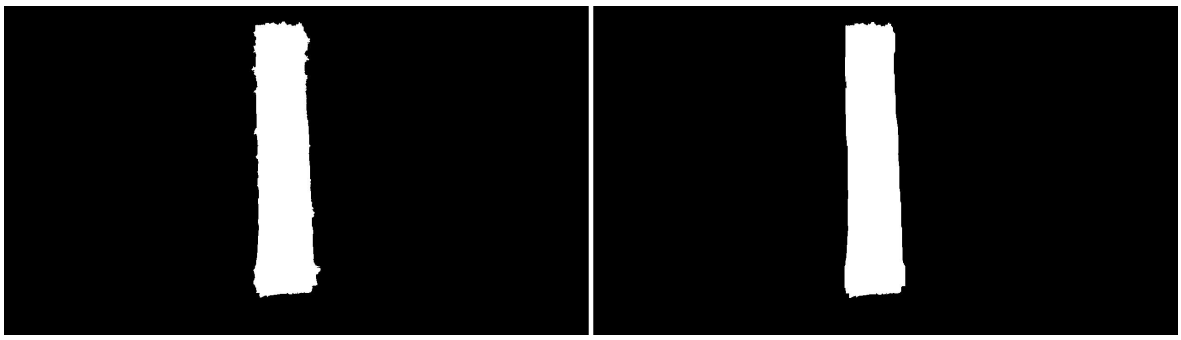


Figure 4.4: Morphological operations on the mask

4.2.5 DBH Estimation

Table 4.2 and Table 4.3 below show the mean DBH values extracted at 5 m and 8 m respectively from the tree trunk. At each distance, 2 DBH estimations (one for each of the two perpendicular views) were performed, and their mean was recorded. From these tables, the percentage absolute error was consistently below 10% except in one case. For the DBH estimated at 5 m, the technique reported a mean absolute error (MAE) of 0.77 cm (2.22%), RMSE of 1.02 cm and relative RMSE of 0.09 cm. For estimates made at 8 m, these values are an MAE of 0.76 cm (2.20%), RMSE of 0.94 cm and relative RMSE of 0.09 cm. The Coefficients of Determination were $R^2 = 0.9913$ and $R^2 = 0.9927$ for the estimates at 5 m and 8 m respectively. The performance of the technique at the two distances is comparable with a marginal improvement observed at 8 m. The corresponding regression plots are shown in Figure 4.5.

Table 4.2: DBH Values Extracted at 5 m from the camera

Tree	Ground Truth DBH (cm)	Extracted DBH (cm)	AE (cm)	APE (%)
1	28.49	29.18	0.69	2.40
2	28.52	27.58	0.95	3.31
3	13.81	14.25	0.44	3.19
4	46.76	45.81	0.95	2.03
5	23.91	24.31	0.40	1.65
6	20.31	20.42	0.11	0.52
7	33.68	32.98	0.70	2.08
8	24.83	25.80	0.97	3.89
9	28.01	27.69	0.32	1.14
10	46.79	48.40	1.61	3.44
11	53.95	57.00	3.05	5.64
12	26.10	27.53	1.43	5.48
13	49.18	48.39	0.79	1.61
14	25.46	25.77	0.31	1.22
15	33.80	34.77	0.97	2.86
16	43.13	43.31	0.17	0.41
17	44.44	44.43	0.01	0.02
18	39.22	39.65	0.43	1.10
19	47.62	48.55	0.93	1.95
20	41.22	41.01	0.21	0.52

Table 4.3: DBH Values Extracted at 8 m from the camera

Tree	Ground Truth DBH (cm)	Extracted DBH (cm)	AE (cm)	APE (%)
1	28.49	28.28	0.21	0.74
2	28.52	29.38	0.86	3.00
3	13.81	13.52	0.30	2.14
4	46.76	47.73	0.97	2.07
5	23.91	23.59	0.32	1.34
6	20.31	21.04	0.73	3.59
7	33.68	32.28	1.40	4.16
8	24.83	25.29	0.46	1.83
9	28.01	28.50	0.49	1.75
10	46.79	47.47	0.68	1.44
11	53.95	53.19	0.76	1.41
12	26.10	26.99	0.89	3.41
13	49.18	48.93	0.25	0.51
14	25.46	25.30	0.16	0.63
15	33.80	34.11	0.31	0.90
16	43.13	43.52	0.38	0.89
17	44.44	44.93	0.48	1.09
18	39.22	40.92	1.70	4.32
19	47.62	49.42	1.80	3.77
20	41.22	43.27	2.05	4.97

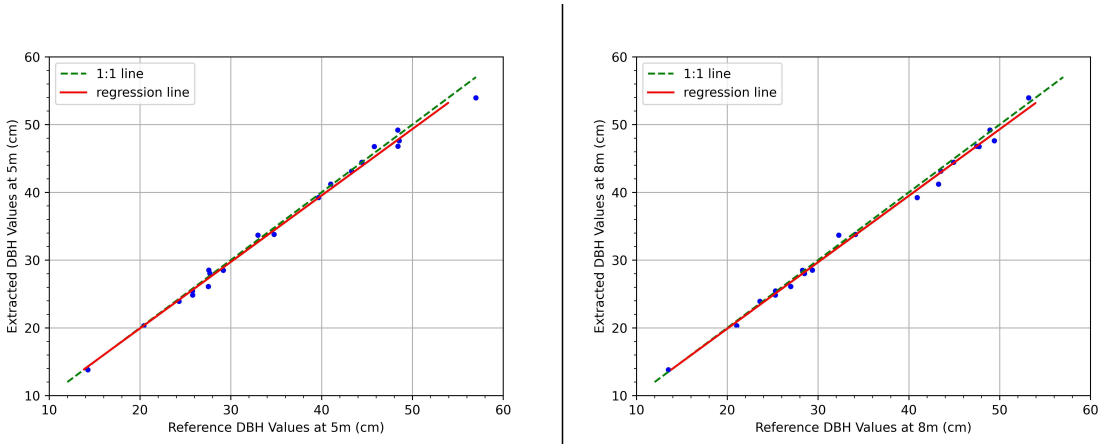


Figure 4.5: Regression plots for DBH extraction at 5 m and 8 m

4.2.6 DBH Errors vs Distance

The MAE and MAPE in DBH estimation were found to grow as the distance from the camera was increased. For 25 image pairs of 5 trees taken between 5 m and 9 m from the tree, the variation in MAPE with distance is shown in figure 4.6. The large MAPE of 9.72% obtained at 9 m from the tree was largely contributed by a single measurement which gave a MAPE of 20.45%. This large error was a statistical outlier when compared to the rest of the DBH estimates summarised in table 4.2 and table 4.3. For example, the results in table 4.3 for DBH estimations at 5 m indicate that only two tree measurements had an absolute error above 5%.

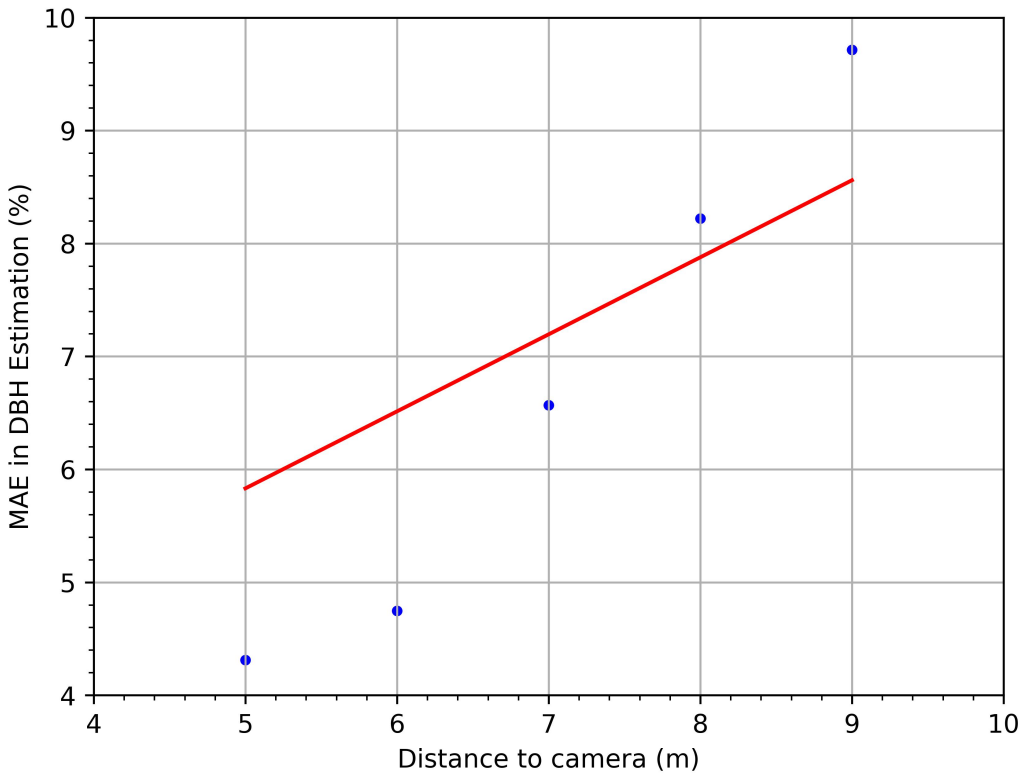


Figure 4.6: Error in DBH Estimation with distance from the tree

4.2.7 Breast Height Location

The difference in pixels and centimetres between the actual and the estimated BH location was recorded in table 4.4. The MAE was found to be 6.9 pixels (2.35%) which translates to a real-world space MAE of 3.15 cm (2.43%). Figure 4.7 shows a sample image comparing the actual and estimated BH locations. The black arrow points to the actual BH indicated by a white tape wrapped around the tree while the red line below it is the estimated BH. It can be seen in this case that the difference between the two is very small. As such, the result is an insignificant error in DBH estimation since the degree of tree taper for most of the trees studied was very small.

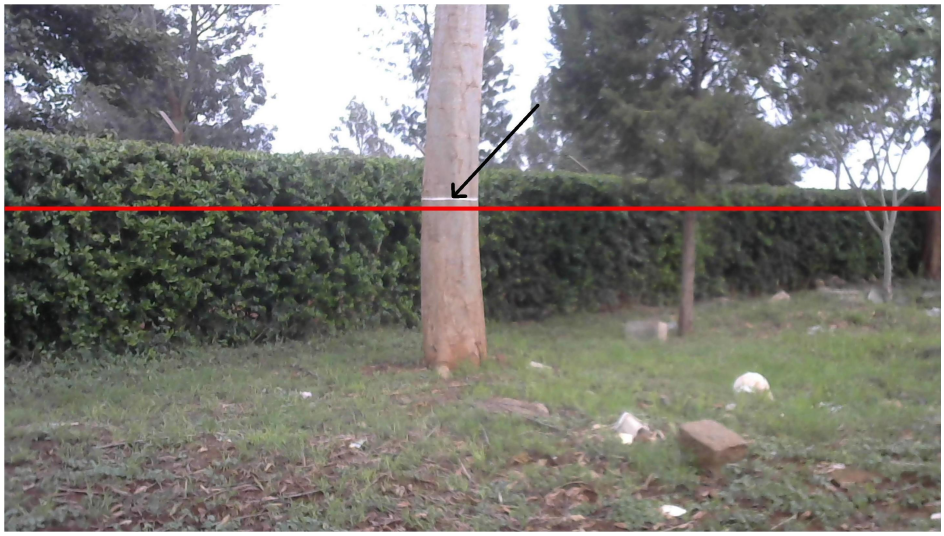


Figure 4.7: Actual vs Estimated BH Location

Table 4.4: Breast height estimation errors

Image	Ground Truth BH (px)	Extracted BH (px)	AE (px)	AE (cm)	APE (%)
1	179	179	0	0.00	0.00
2	285	276	9	4.09	3.15
3	248	248	0	0.00	0.00
4	285	275	10	5.25	4.04
5	287	272	15	6.44	4.95
6	313	304	9	3.54	2.72
7	266	256	10	5.25	4.04
8	325	311	14	6.01	4.63
9	312	312	0	0.00	0.00
10	280	278	2	0.97	0.74

4.2.8 Crown Diameter Estimation

The values of CD predicted by our algorithm are recorded in table 4.5. These values were extracted from images taken from arbitrary distances. For the 10 trees studied, an MAE of 25.3 cm (7.03%) was achieved. The values of RMSE and relative RMSE

were found to be 31.93 cm and 0.28 cm respectively. A good proportion of the variance in the ground truth values could be explained by the measurements obtained from the proposed technique as shown by the R^2 value of 9209. The regression plot is shown in figure 4.8.

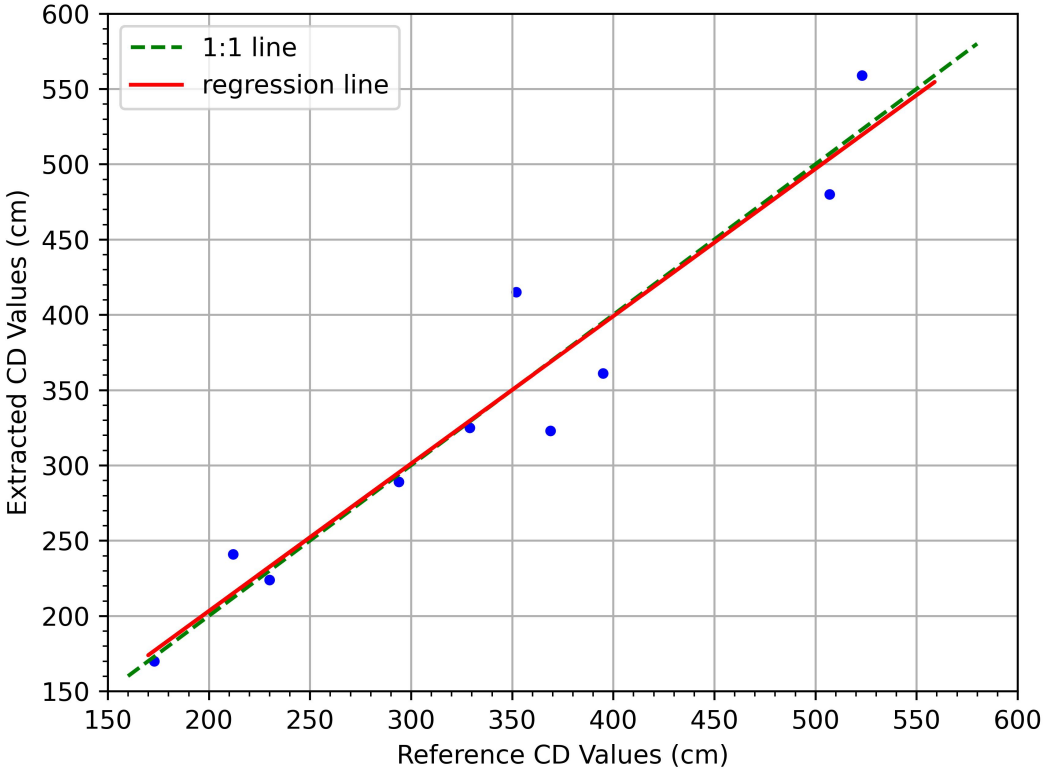


Figure 4.8: Regression plots comparing reference and extracted CD values

Table 4.5: Extracted CD Values

Tree	Ground Truth CD (cm)	Extracted CD (cm)	AE (cm)	APE (%)
1	415	352	63	15.18
2	559	523	36	6.44
3	480	507	27	5.63
4	170	173	3	1.76
5	289	294	5	1.73
6	224	230	6	2.68
7	241	212	29	12.03
8	361	395	34	9.42
9	325	329	4	1.23
10	323	369	46	14.24

4.2.9 Tree Height Estimation

The performance of our technique showed higher performance in predicting the values of TH in comparison to CD (table 4.6). The values of MAE and RMSE, and were found to be 17.9 cm (MAPE = 4.57%), and 19.32 cm respectively. The R^2 value was 0.9540, which shows that the measurements by the propose technique can explain most of the variance in the ground truth values. The regression plot is shown in figure 4.9.

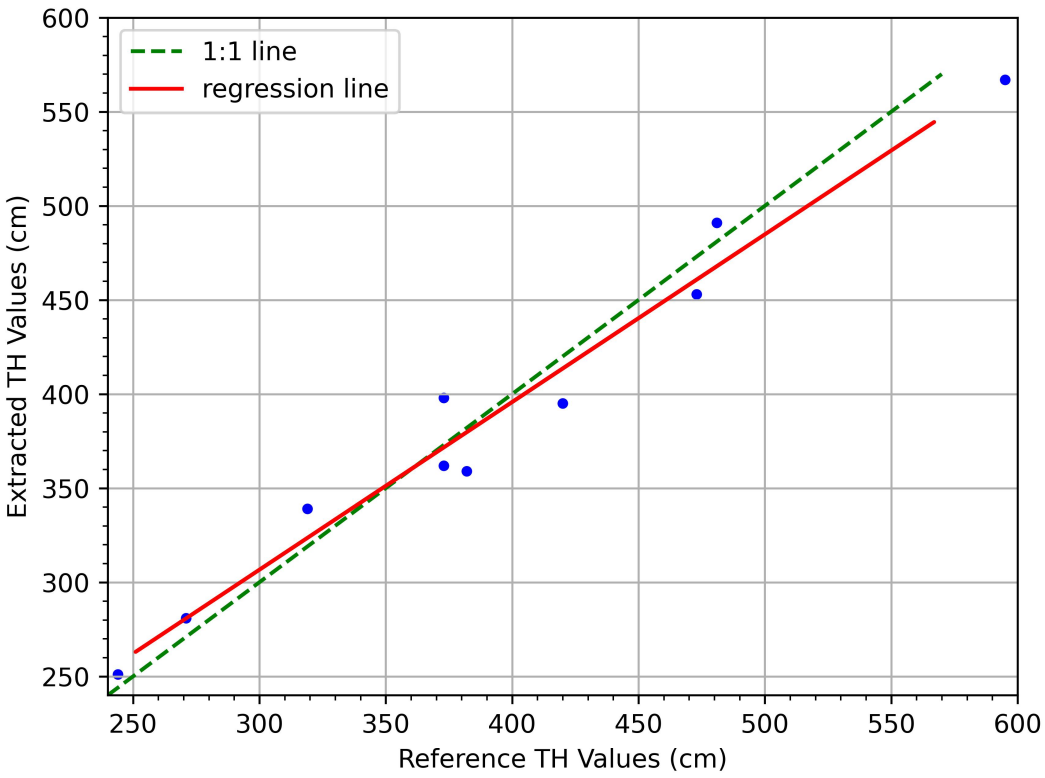


Figure 4.9: Regression plots comparing reference and extracted TH values

Table 4.6: Extracted TH Values

Tree	Ground Truth TH (cm)	Extracted TH (cm)	AE (cm)	APE (%)
1	339	319	20	5.90
2	453	473	20	4.42
3	567	595	28	4.94
4	251	244	7	2.79
5	281	271	10	3.56
6	362	373	11	3.04
7	359	382	23	6.41
8	398	373	25	6.28
9	491	481	10	2.04
10	395	420	25	6.33

4.3 Discussion

In this work, a technique for automatically extracting three important tree biophysical parameters has been proposed. The proposed technique involves extracting tree biophysical parameters from disparity maps extracted from RGB images only. So far, based on the literature surveyed in this thesis, only Malekabadi *et al.* [55] have implemented this approach, albeit without extracting parameters from real trees. It is also worth mentioning that the performance of the proposed technique based on the reported results ranks above those of other researchers using stereoscopic vision based on RMSE and MAE. Gonzalez *et al.* [33], who extracted tree parameters from hemispherical image pairs and achieved an RMSE of 5.69 cm with a tree distance of up to 15 m while Perng *et al.* [18] reported an RMSE of 6.84 cm with for tree distances of 5 – 10 m. Perng *et al.* also reported a mean error (ME) of -0.48 cm, a value that is always less than the MAE for any dataset. At 5 m from the tree, Eliopoulos *et al.* achieved an RMSE of 2.57 cm and MAE of 2.25 cm. In contrast, the results presented in this thesis indicate lower RMSE values of 1.02 cm and 0.94 cm at 5 m and 8 m respectively, and MAE of 0.77 cm and 0.76 cm at the same distances respectively. RMSE is a very important metric because it shows how well a model predicts a variable. The smaller it is the less the uncertainty in the prediction [30], which indicates a lower uncertainty of prediction reported by our study. The small RMSE values reported in this thesis indicate a lower uncertainty in the measurements extracted by the proposed approach.

Overall, for all three tree parameters, the proportion of the variance in the field-measured data that could be explained by the measurements extracted by the

proposed technique was high. The R^2 values for all measurements were very close to 1, with DBH values at 5 m from the tree registering the highest performance. The slopes of the regression lines were also close to 1 in all instances indicating that the extracted measurements were very close to the ground truth values. To find the similarity between the DBH estimates at 5 m and 8 m, the Wilcoxon Signed Ranked Test was applied. The result was a p-value of 0.81, which indicated that these measurements were very similar.

The bias reported for DBH estimation was 0.3 cm (1.08%) and 0.42 cm (1.2%) at 5 m and 8 m respectively from the camera. This indicates an overestimation of the DBH in both cases with a larger overestimation for the latter case. Results for breast height estimation showed that the algorithm consistently identified a point below the 1.3 m mark as the breast height location. This error in breast height estimation is the most likely cause of DBH overestimation since points below the BH were wider than those at the BH for all thee trees studied. The bias values in TH and CD were 3.5 cm (0.9%) and -0.3 cm (-0.09%), which are much smaller relative to the ground truth values. Therefore, despite the lower accuracies of CD and TH estimates in comparison to those of DBH, a lower bias was still achieved for the studied sample. Since the bias errors in estimating the three attributes are quite small in proportion to the ground truth values, the result is that the error in the estimation of tree parameters over large areas is also expected to be small relative to the actual values.

In comparison to the SfM study by Marzulli *et al.* who reported a bias of between -0.58 cm and -2.04 cm in DBH estimation [30], the bias achieved in this study was smaller. Sanchez-Gonzalez *et al.* [33] also achieved a higher overestimation with a

bias of between 1.4 cm and 2.88 cm. When comparing the performance of SfM and TLS, Piermattei *et al.* [47] reported a minimum bias of -0.71 cm and -0.38 cm respectively relative to field measurements. The lower amount of bias reported in this thesis indicates a higher performance than that reported in these three studies.

Given that the image acquisition takes place instantly and disparity image generation and parameter extraction took less than 4 seconds (table 4.1), this technique has the potential for use in real-time parameter estimation. This is very encouraging as very few researchers who have developed techniques based on stereoscopic vision have either achieved this ability or reported the potential for real-time estimation. These include the work done by Malekabadi *et al.* [55] and Eliopoulos *et al.* [17].

The DBH estimates reported here are reasonably accurate. Of the more than 80 measurements extracted by the algorithm, only two trees registered an absolute error greater than 10%. The possible sources of these errors include shaking hands during image acquisition, inaccurate image segmentation, or errors stemming from image normalization. Since the stereo camera was held by hand during image acquisition and there exists a short lag between the time the left and right images are captured, shaking can lead to errors in disparity image computation. In the case of the 20.45% error, this is the most likely culprit, since a different image pair of the same tree taken from the same distance yielded an absolute error of less than 5%.

The analysis of the performance of this technique in estimating DBH as the distance from the camera increases is not surprising at all. The problem of decreasing accuracy with distance is inherent to all types of light sensors [15], [18], [19]. This is one of the major issues facing all attempts to perform measurements of multiple trees at

once since trees further away from the sensor tend to exhibit larger errors in recorded values [5]. Additionally, stereoscopic vision systems rely on disparity values to estimate distances correctly. For any stereo camera baseline, no observable change in disparity is seen as objects move further away [23], [24]beyod a certain distance from the camera. Therefore, as the camera is moved further from the tree, the accuracy in DBH estimates reduces.

It was important to show how accurately the proposed technique could estimate the location of the breast height at 1.3 m above the ground. It is reasonable to assume that a failure to locate this position correctly will lead to significant errors in DBH measurement due to tree taper. The diameters at lower heights are larger while those higher are smaller. An MAE of 3.15 cm in breast height location is reported in this work. There seems to be no other study, other than that done by Eliopoulos *et al*, [17] provides some information on the performance in this area, which makes performance comparison difficult to do. Even then, the study includes one image showing the computed and actual breast height. This information does not lend itself to any meaningful comparison with the results reported in our study. The good performance of the proposed method at estimating the location of the breast height provided the confidence to proceed and apply it in DBH extraction.

Studies focusing on TH measurement usually use RS techniques to create Canopy Height Models (CHMs) rather than measuring a single tree at a time [54]. There are hardly any studies where terrestrial photogrammetry has been used for this purpose because forest trees are very tall and have a lot of obstructions. The study done by Malekabadi *et al.* [55], while similar to this study in scope, only tested performance on

artefact trees with comparatively clean backgrounds. No parameters were extracted from actual trees with complex backgrounds. They reported an error of 2.2% - 6.6% in TH estimation which is comparable to the absolute error values of 2.79% - 6.33% reported in this thesis. Another unique aspect of this study is the focus on small trees less than 6 m tall. The accuracies of the proposed method at estimating both DBH and TH are very similar, and this is most likely because the distance to the camera in both cases was less than 10 m. As already pointed out earlier, the degradation in the accuracy of distance estimation becomes more severe once the camera is more than 12 m away.

The task of CD estimation is usually difficult because individual crown delineation is a rather complex task. Even interactive segmentation algorithms such as that used in this study do not perform well when the background and foreground pixels are similar and the transition between them is seamless. The larger errors reported in CD estimation compared to DBH and TH can be largely attributed to this issue. Failure to correctly identify the crown edge pixels during segmentation can yield highly inaccurate results in estimation, a process that requires good segmentation. Based on the scrutiny of many other studies for comparative analysis, the work of Malekabadi *et al.* [55] most closely aligns with the proposed method at CD estimation and their results can be easily compared with those reported in this thesis. Their study reported impressive errors of between -0.8% and 3.1% in CD estimation while the proposed method achieved errors of between -15.18% and 14.24%. Their superior performance can still be attributed to working with artefact trees in an ideal environment.

The major limitations of the proposed technique include inaccurate DBH extraction

for trees with bifurcation, occluded trunks, trees with slanting trunks and those with bulges at breast height. The standard practice for trees that bulge at the breast height is taking the diameter above the bulge as the DBH, something that the proposed algorithms do not take into consideration. The proposed system was designed to operate within a range not exceeding 12.6 m. Outside this regime, the accuracy in distance estimation reduces significantly. It was observed that accuracies began to plummet at 9 m. These limitations apply in a similar way to both CD and TH. Another limitation of the proposed method is the use of an interactive segmentation algorithm. While this has been the case, the segmentation step can be viewed as an independent component that can be automated later and improved without sacrificing performance. Automating the segmentation step is one of the objectives for future research. The work done by Jodas *et al.* [57] provides one of the attractive ways to achieve this.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This thesis has presented a technique for extracting tree DBH, CD, and TH from stereoscopic image pairs using the concept of disparity images. The depth information available from the disparity maps was used as the basis for deriving the geometry for calculating the tree attributes. Computational geometric algorithms were then written to extract the tree parameters with good accuracy and minimal uncertainty in model predictions indicated by small RMSE values.

5.2 Conclusions

The geometry for calculating the three individual tree attributes was first derived. By combining the geometry presented and digital image processing techniques, an algorithm was written that could automatically to extract a tree attribute from a stereoscopic image pair. Overall, this technique registered good accuracy and minimal uncertainty in parameter estimation as indicated by small RMSE and MAE values.

The algorithms developed in this study for estimating tree parameters were validated by testing them on actual trees located in a park with low tree density ($\geq$ 5 m apart e.g., recreation parks). The extracted measurements were compared to ground truth values and the performance of the proposed method was also compared to that achieved in previous studies. Having surveyed numerous studies to find similar ones for comparative performance, the MAE $\leq$ 1 cm in DBH estimation reported in this

thesis turned out to be superior. Comparable performance was achieved for TH estimation while no meaningful comparison was done for CD estimation because a suitable study for comparison was not found.

This proposed approach is suitable for use by forest management agencies and individual forest business owners for rapid acquisition of tree parameters. The heights and crowns of taller trees can be measured by adjusting the stereo camera baseline.

5.3 Recommendations

Based on the results reported in this study, the application of stereoscopic vision for extraction of individual tree parameters from disparity images in real forest setups can be explored.

Further research should be done to improve on the limitations of this study. For example, the extraction of tree attributes from slanting trees, trees with bifurcation, as well as from multiple trees simultaneously are areas that should be explored. In addition, studies focusing on improving the accuracy of attribute estimation by the proposed technique are also encouraged. The method presented in this study is also not suited for use in extracting the DBH of leaning, curved, and crooked tree trunks as well as forked trees below the breast height. Another limitation of this study is the estimation of tree attributes for trees planted on a steep slope. Suitable improvements can be made to the algorithms to incorporate these scenarios. These limitations are topics that should be explored in future research.

REFERENCES

[1] FAO. *Global Forest Resources Assessment 2020*. FAO, 2020.

[2] Ministry of Environment and Forestry. *National Strategy for Achieving and Maintaining Over 10% Tree Cover By 2022 (2019-2022)*. May 2019.

[3] Nathaniel Eisen. "Restoring the Commons: Joint Reforestation Governance in Kenya". In: (Mar. 2019).

[4] Tianyu Hu, Yanjun Su, Baolin Xue, et al. "Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data". In: *Remote Sensing* 8.7 (2016), p. 565.

[5] Haozhou Wang, Ting-Ru Yang, Joni Waldy, et al. "Estimating Individual Tree Heights and DBHs from Vertically Displaced Spherical Image Pairs". In: *Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS)* 13.1 (2021), pp. 1–14.

[6] Klaus von Gadow, Chun Yu Zhang, and Xiu Hai Zhao. "Science-based forest design". In: *Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS)* 1.1 (2009), pp. 14–25.

[7] W Brad Smith. "Forest inventory and analysis: a national inventory and monitoring program". In: *Environmental pollution* 116 (2002), S233–S242.

[8] Johannes Breidenbach, Aksel Granhus, Gro Hylen, et al. "A century of National Forest Inventory in Norway–informing past, present, and future decisions". In: *Forest Ecosystems* 7.1 (2020), pp. 1–19.

[9] Kenneth G MacDicken. “Global forest resources assessment 2015: what, why and how?” In: *Forest Ecology and Management* 352 (2015), pp. 3–8.

[10] Rodney J Keenan, Gregory A Reams, Frédéric Achard, et al. “Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015”. In: *Forest Ecology and Management* 352 (2015), pp. 9–20.

[11] Mst Karimon Nesha, Martin Herold, Veronique De Sy, et al. “An assessment of data sources, data quality and changes in national forest monitoring capacities in the Global Forest Resources Assessment 2005–2020”. In: *Environmental Research Letters* 16.5 (2021), p. 054029.

[12] XD Lei, MP Tang, YC Lu, et al. “Forest inventory in China: status and challenges”. In: *International Forestry Review* 11.1 (2009), pp. 52–63.

[13] Susan Trumbore, Paulo Brando, and Henrik Hartmann. “Forest health and global change”. In: *Science* 349.6250 (2015), pp. 814–818.

[14] Mona Forsman, Niclas Börlin, and Johan Holmgren. “Estimation of tree stem attributes using terrestrial photogrammetry with a camera rig”. In: *Forests* 7.3 (2016), p. 61.

[15] Carlos Henrique Souza Celes, Raquel Fernandes de Araujo, Fabiano Emmert, et al. “Digital Approach for Measuring Tree Diameters in the Amazon Forest”. In: *Floresta e Ambiente* 26 (2019).

[16] Steven W Chen, Guilherme V Nardari, Elijah S Lee, et al. “Sloam: Semantic lidar odometry and mapping for forest inventory”. In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 612–619.

[17] Nicholas J Eliopoulos, Yezhi Shen, Minh Luong Nguyen, et al. “Rapid tree diameter computation with terrestrial stereoscopic photogrammetry”. In: *Journal of Forestry* 118.4 (2020), pp. 355–361.

[18] Bo-Hao Perng, Tzeng Yih Lam, and Mei-Kuei Lu. “Stereoscopic imaging with spherical panoramas for measuring tree distance and diameter under forest canopies”. In: *Forestry: An International Journal of Forest Research* 91.5 (2018), pp. 662–673.

[19] Hadi Bayati, Akbar Najafi, Javad Vahidi, et al. “3D reconstruction of uneven-aged forest in single tree scale using digital camera and SfM-MVS technique”. In: *Scandinavian Journal of Forest Research* 36.2-3 (2021), pp. 210–220.

[20] Viviana Otero, Ruben Van De Kerchove, Behara Satyanarayana, et al. “Managing mangrove forests from the sky: Forest inventory using field data and Unmanned Aerial Vehicle (UAV) imagery in the Matang Mangrove Forest Reserve, peninsular Malaysia”. In: *Forest ecology and management* 411 (2018), pp. 35–45.

[21] Stefano Puliti, Svein Solberg, and Aksel Granhus. “Use of UAV photogrammetric data for estimation of biophysical properties in forest stands under regeneration”. In: *Remote Sensing* 11.3 (2019), p. 233.

[22] Luana Mendes dos Santos, Brenon Diennevan de Souza Barbosa, Adriano Valentim Diotto, et al. “Biophysical parameters of coffee crop estimated by UAV RGB images”. In: *Precision Agriculture* 21.6 (2020), pp. 1227–1241.

[23] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Second. Cambridge University Press, 2003.

[24] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer, Sept. 2010.

[25] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[26] D. Scharstein, R. Szeliski, and R. Zabih. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms". In: *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*. 2001, pp. 131–140.

[27] Heiko Hirschmuller. "Stereo processing by semiglobal matching and mutual information". In: *IEEE Transactions on pattern analysis and machine intelligence* 30.2 (2007), pp. 328–341.

[28] MathWorks. *Stereo Disparity Using Semi-Global Block Matching*. URL: `https://www.mathworks.com/help/visionhdl/ug/stereoscopic-disparity.html` (visited on 11/01/2022).

[29] S. Birchfield and C. Tomasi. "A pixel dissimilarity measure that is insensitive to image sampling". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.4 (1998), pp. 401–406.

[30] Maria Immacolata Marzulli, Pasi Raumonen, Roberto Greco, et al. "Estimating tree stem diameters and volume from smartphone photogrammetric point clouds". In: *Forestry: An International Journal of Forest Research* 93.3 (2020), pp. 411–429.

[31] Lulu Liu, Aiwu Zhang, Shen Xiao, et al. “Single tree segmentation and diameter at breast height estimation with mobile LiDAR”. In: *Ieee Access* 9 (2021), pp. 24314–24325.

[32] James McGlade, Luke Wallace, Bryan Hally, et al. “An early exploration of the use of the Microsoft Azure Kinect for estimation of urban tree Diameter at Breast Height”. In: *Remote Sensing Letters* 11.11 (2020), pp. 963–972.

[33] Mariola Sanchez-Gonzalez, Miguel Cabrera, Pedro Javier Herrera, et al. “Basal area and diameter distribution estimation using stereoscopic hemispherical images”. In: *Photogrammetric Engineering & Remote Sensing* 82.8 (2016), pp. 605–616.

[34] Alice R Jones, Ramesh Raja Segaran, Kenneth D Clarke, et al. “Estimating mangrove tree biomass and carbon content: a comparison of forest inventory techniques and drone imagery”. In: *Frontiers in Marine Science* 6 (2020), p. 784.

[35] Charles N Kuria, Balozi B Kirongo, and Wilson Kipkore. “Growth and Yield Models for Plantation-Grown Cupressus lusitanica for Central Kenya”. In: *African Journal of Education, Science and Technology* 5.2 (2019), pp. 34–58.

[36] Shu Gideo Neba. “Assessment And Prediction Of Above-Ground Biomass In Selectivelylogged Forest Concessions Using Field Measurements And Remotesensing Data: Case Study In South East Cameroon”. MA thesis. University of Helsinki, May 2013.

[37] J. G. Kairo, J. Bosire, J. Langat, et al. “Allometry and biomass distribution in replanted mangrove plantations at Gazi Bay, Kenya”. In: *Aquatic Conservation: Marine and Freshwater Ecosystems* 19.S1 (May 2009), S63–S69.

[38] Hari Adhikari, Janne Heiskanen, Mika Siljander, et al. “Determinants of Aboveground Biomass across an Afromontane Landscape Mosaic in Kenya”. In: *Remote Sensing* 9.8 (Aug. 2017), p. 827.

[39] Xinlian Liang, Anttoni Jaakkola, Yunsheng Wang, et al. “The use of a hand-held camera for individual tree 3D mapping in forest sample plots”. In: *Remote Sensing* 6.7 (2014), pp. 6587–6603.

[40] Yongxiang Fan, Zhongke Feng, Abdul Mannan, et al. “Estimating tree position, diameter at breast height, and tree height in real-time using a mobile phone with RGB-D SLAM”. In: *Remote Sensing* 10.11 (2018), p. 1845.

[41] Carlos Cabo, Celestino Ordóñez, Carlos A López-Sánchez, et al. “Automatic dendrometry: Tree detection, tree height and diameter estimation using terrestrial laser scanning”. In: *International journal of applied earth observation and geoinformation* 69 (2018), pp. 164–174.

[42] Guangpeng Fan, Feixiang Chen, Yan Li, et al. “Development and testing of a new ground measurement tool to assist in forest GIS surveys”. In: *Forests* 10.8 (2019), p. 643.

[43] Eva Marino, Fernando Montes, José Luis Tomé, et al. “Vertical forest structure analysis for wildfire prevention: comparing airborne laser scanning data and stereoscopic hemispherical images”. In: *International journal of applied earth observation and geoinformation* 73 (2018), pp. 438–449.

[44] Haozhou Wang, John A Kershaw, Ting-Ru Yang, et al. “An integrated system for estimating forest basal area from spherical images”. In: *Math. Comput. For. Nat. Resour. Sci.* 12 (2020), pp. 1–15.

[45] Juan Guerra-Hernández, Diogo N Cosenza, Luiz Carlos Estraviz Rodriguez, et al. “Comparison of ALS-and UAV (SfM)-derived high-density point clouds for individual tree detection in Eucalyptus plantations”. In: *International Journal of Remote Sensing* 39.15-16 (2018), pp. 5211–5235.

[46] Ana Paula Dalla Corte, Deivison Venicio Souza, Franciel Eduardo Rex, et al. “Forest inventory with high-density UAV-Lidar: Machine learning approaches for predicting individual tree attributes”. In: *Computers and Electronics in Agriculture* 179 (2020), p. 105815.

[47] Livia Piermattei, Wilfried Karel, Di Wang, et al. “Terrestrial structure from motion photogrammetry for deriving forest inventory data”. In: *Remote Sensing* 11.8 (2019), p. 950.

[48] Chau-Chang Wang, ed. *Laser Scanning, Theory and Applications*. InTech, Apr. 2011.

[49] Mathias Lemmens. “Terrestrial Laser Scanning”. In: *Geo-information*. Springer Netherlands, 2011, pp. 101–121.

[50] Visarut Trairattanapa, Ankit A Ravankar, and Takanori Emaru. “Estimation of Tree Diameter at Breast Height using Stereo Camera by Drone Surveying and Mobile Scanning Methods”. In: *2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*. IEEE. 2020, pp. 946–951.

[51] Daniel Tobón Collazos, Victor Romero Cano, Juan Carlos Perafan Villota, et al. “A photogrammetric system for dendrometric feature estimation of individual trees”. In: *2018 IEEE 2nd Colombian Conference on Robotics and Automation (CCRA)*. IEEE. 2018, pp. 1–6.

[52] Jeffrey Byrne and Sanjiv Singh. *Precise Image Segmentation for Forest Inventory*. Tech. rep. CMU-RI-TR-98-14. Pittsburgh, PA: Carnegie Mellon University, 1998.

[53] Dianyuan Han and Chengduan Wang. “Tree height measurement based on image processing embedded in smart mobile phone”. In: *2011 International Conference on Multimedia Technology*. IEEE. 2011, pp. 3293–3296.

[54] B St-Onge, C Vega, RA Fournier, et al. “Mapping canopy height using a combination of digital stereo-photogrammetry and lidar”. In: *International Journal of Remote Sensing* 29.11 (2008), pp. 3343–3364.

[55] Ayoub Jafari Malekabadi, Mehdi Khojastehpour, and Bagher Emadi. “Disparity map computation of tree using stereo vision system and effects of canopy shapes and foliage density”. In: *Computers and electronics in agriculture* 156 (2019), pp. 627–644.

[56] Wajid Ali, Fredrik Georgsson, and Thomas Hellstrom. “Visual tree detection for autonomous navigation in forest environment”. In: *2008 IEEE Intelligent Vehicles Symposium*. 2008, pp. 560–565.

[57] Danilo Samuel Jodas, Sergio Brazolin, Takashi Yojo, et al. “A Deep Learning-based Approach for Tree Trunk Segmentation”. In: *2021 34th*

SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). 2021, pp. 370–377.

[58] Liying Zheng, Jingtao Zhang, and Qianyu Wang. "Mean-shift-based color segmentation of images containing green vegetation". In: *Computers and Electronics in Agriculture* 65.1 (2009), pp. 93–98.

[59] Asra Aslam, Mohd Ansari, et al. "Depth-map generation using pixel matching in stereoscopic pair of images". In: *arXiv preprint arXiv:1902.03471* (2019).

[60] Roland Perko, Hannes Raggam, Janik Deutscher, et al. "Forest Assessment Using High Resolution SAR Data in X-Band". In: *Remote Sensing* 3.4 (Apr. 2011), pp. 792–815.

[61] Yueting Wang, Xiaoli Zhang, and Zhengqi Guo. "Estimation of tree height and aboveground biomass of coniferous forests in North China using stereo ZY-3, multispectral Sentinel-2, and DEM data". In: *Ecological Indicators* 126 (2021), p. 107645.

[62] Changming Sun, Ronald Jones, Hugues Talbot, et al. "Measuring the distance of vegetation from powerlines using stereo vision". In: *ISPRS journal of photogrammetry and remote sensing* 60.4 (2006), pp. 269–283.

[63] Simon E. Korepanov, Sergey A. Smirnov, Valery V. Strotov, et al. "Object distance estimation algorithm for real-time FPGA-based stereoscopic vision system". In: *High-Performance Computing in Geoscience and Remote Sensing VIII*. Ed. by Bormin Huang, Sebastián López, and Zhensen Wu. SPIE, Oct. 2018.

[64] Cristina Olaverri-Monreal, Gerd Ch. Krizek, Florian Michaeler, et al. "Collaborative approach for a safe driving distance using stereoscopic image processing". In: *Future Generation Computer Systems* 95 (June 2019), pp. 880–889.

[65] Wenjian Ni, Zhiyu Zhang, Guoqing Sun, et al. "Modeling the Stereoscopic Features of Mountainous Forest Landscapes for the Extraction of Forest Heights from Stereo Imagery". In: *Remote Sensing* 11.10 (May 2019), p. 1222.

[66] Thomas Hellström, Ahmad Ostovar, T Hellström, et al. "Detection of trees based on quality guided image segmentation". In: *Proceedings of the Second International RHEA Conference, Madrid, Spain*. 2014, pp. 21–23.

[67] Yining Deng, B.S. Manjunath, and H. Shin. "Color image segmentation". In: *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*. Vol. 2. 1999, 446–451 Vol. 2.

[68] Calvin Hung, Juan Nieto, Zachary Taylor, et al. "Orchard fruit segmentation using multi-spectral feature learning". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2013, pp. 5314–5320.

[69] Xiao-song Wang, Xin-yuan Huang, and Hui Fu. "The Study of Color Tree Image Segmentation". In: *2009 Second International Workshop on Computer Science and Engineering*. Vol. 2. 2009, pp. 303–307.

[70] Lei Cheng and Tieying Song. "An efficient approach for tree digital image segmentation". In: *Forestry Studies in China* 6.3 (2004), pp. 43–49.

[71] Nvidia. *Jetson Nano 2GB Developer Kit*. URL: `https://developer.nvidia.com/embedded/jetson-nano-2gb-developer-kit` (visited on 11/25/2021).

[72] Logitech. *C270 HD Webcam*. URL: https://www.logitech.com/en-us/products/webcams/c270-hd-webcam.960-000694.html (visited on 11/25/2021).

[73] Robert Bosch Power Tools GmbH. *GLM 20*. URL: https://www.boschtools.com/us/en/boschtools-ocs/laser-measures-glm-20-143533-p (visited on 11/01/2022).

[74] Z. Zhang. "A flexible new technique for camera calibration". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (2000), pp. 1330–1334.

[75] K. Levenberg. "A method for the solution of certain nonlinear problems in least squares". In: *Quarterly of Applied Mathematics* 2 (1944), pp. 164–168.

[76] Guoliang Hu, Zuofeng Zhou, Jianzhong Cao, et al. "Non-linear calibration optimisation based on the Levenberg-Marquardt algorithm". en. In: *IET Image Process.* 14.7 (May 2020), pp. 1402–1414.

[77] Y Zhang and Q Ji. "Camera calibration with genetic algorithms". In: *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*. Seoul, South Korea: IEEE, 2002.

[78] Xueqin Lü, Lingzheng Meng, Liyuan Long, et al. "Comprehensive improvement of camera calibration based on mutation particle swarm optimization". en. In: *Measurement (Lond.)* 187.110303 (Jan. 2022), p. 110303.

[79] Kashif Bilal and Junaid Qureshi. "Nature inspired optimization techniques for Camera calibration". In: *2008 4th International Conference on Emerging Technologies*. Rawalpindi, Pakistan: IEEE, Oct. 2008.

[80] Karel Zuiderveld. “Contrast Limited Adaptive Histogram Equalization”. In: *Graphics Gems*. Elsevier, 1994, pp. 474–485.

[81] Rafael C Gonzalez and Richard E Woods. *Digital Image Processing*. 4th ed. Upper Saddle River, NJ: Pearson, 2017.

[82] Richard H. Byrd, Robert B. Schnabel, and Gerald A. Shultz. “Approximate solution of the trust region problem by minimization over two-dimensional subspaces”. In: *Mathematical Programming* 40-40.1-3 (Jan. 1988), pp. 247–263.

[83] Dameng Yin and Le Wang. “How to assess the accuracy of the individual tree-based forest inventory derived from remotely sensed data: a review”. In: *International Journal of Remote Sensing* 37.19 (Aug. 2016), pp. 4521–4553.

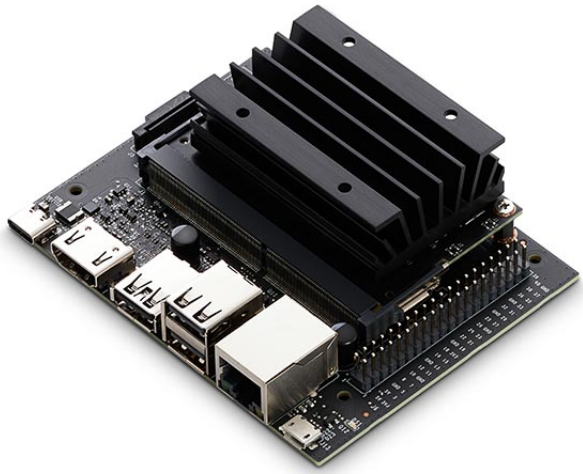
APPENDICES

Appendix I: Images of the Materials and the Study Area

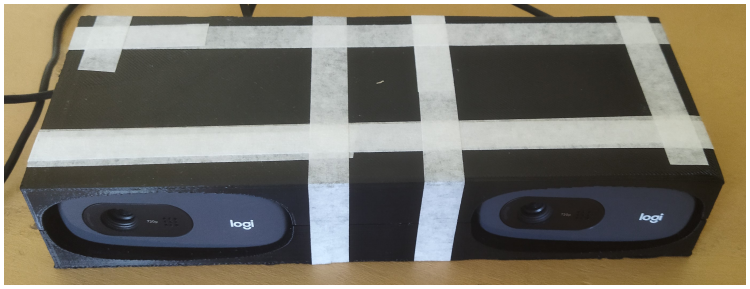
Images of the Materials

Camera

Jetson Nano 2GB Developer Kit



Camera Rig



Bosch GLM20 Laser Rangefinder

Photographs of the Study Area





Appendix II: Software and Data

The tree images used in arriving at the findings of this study, as well as the software code for extracting the tree biophysical parameters algorithmically, are publicly available via the following links:

- Images: `https://doi.org/10.17632/nx3ggv7pxf.1`
- Code: `https://github.com/DeKUT-DSAIL/forest-monitoring`

Appendix III: Note on Publication

The work done in this thesis was packaged in a piece of software and published in the journal *SoftwareX* by Elsevier. The publication's citation is:

C. Kiplimo, C. E. Epege, C. wa Maina, and B. Okal, "DSAIL-TreeVision: A software tool for extracting tree biophysical parameters from stereoscopic images," *SoftwareX*, vol. 26, p. 101661, May 2024.

The article is available online at https://www.doi.org/10.1016/j.softx.2024.101661