LOW COST IMPLEMENTATION OF SINGLE-CAMERA STEREO VISION CALIBRATION

ATHUKORALAGE SUPUN GAYANTHA DISSANAYAKA

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN ROBOTICS AND COMPUTATIONAL INTELLIGENCE SYSTEMS
SCHOOL OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2024

KMITL-2024-EN-M- 407-311

Thesis	Single Camera Stereo Vision Camera Calibration
Student	Mr. Athukoralage Supun Gayantha Dissanayaka
Student ID.	61601098
Degree	Master of Engineering
Program	Robotics and AI Engineering
Year	2024
Thesis Advisor	Prof. Dr. Pitikate Sooraksa
Co-Thesis Advisor	Assoc. Prof. Dr. Somyot Kaitwanidvilai

ABSTACT IN THAI

ระบบสเตอริโอที่เรียบง่ายสามารถสร้างขึ้นจากกล้องเดี่ยวโดยใช้ปริซึมในเส้นทางแสงเพื่อให้ได้ภา สองมุมมองของระบบ ความเรียบง่ายของระบบเหล่านี้มีข้อดีหลายประการ โดยเฉพาะอย่างยิ่งหากเป้าหมายคือหุ่นยนต์ใต้น้ำ ที่ขนาดกะทัดรัดและความสามารถในการปิดผนึกชิ้นส่วนออปติคอลเป็นปัจจัยสำคัญ อย่างไรก็ตาม การกระจายแสงโดยปริซึม รวมถึงความบิดเบือนของเลนส์ ทำให้การสอบเทียบเป็นเรื่องท้าทาย โดยการใช้วิธีการที่ไม่ต้องใช้แบบจำลอง เราสามารถสอบเทียบระบบสเตอริโอที่ใช้ปริซึมได้อย่างมีประสิทธิภาพ นอกจากนี้ เรายังมุ่งหมายที่จะใช้ปริซึม 45° ที่หาได้ง่าย ซึ่งมีการกระจายแสงในระบบมาก แต่ยังคงความเรียบง่ายและลดต้นทุนเมื่อเทียบกับปริซึมมุมต่ำที่สั่งทำพิเศษ ไฟ LED สมัยใหม่ให้แหล่งกำเนิดแสงที่มีความเข้มสูงและแบนด์วิดท์ต่ำ เราใช้แหล่งกำเนิดแสงสามตัวที่จัดวางโดยประมาณบนช่องสัญญาณ RGB ของกล้องเชิงพาณิชย์ที่หาได้ทั่วไป ระบบของเราใช้รูปแบบเป้าหมายวงกลมที่ครอบคลุมบริเวณที่มองเห็นได้สองตาในฉาก และรวบรวมชุดภาพที่ระยะทางที่ทราบ โดยใช้แหล่งกำเนิดแสงสามตัวที่แยกกัน จากภาพเหล่านี้ เราสร้างตารางการค้นหาสองตาราง หนึ่งสำหรับแต่ละพิกเซลในภาพและการเบี่ยงเบนที่ได้จากการจับคู่จุดที่สอดคล้องกัน Cp(u,v,du) ซึ่งมีสามมิติ และตารางการค้นหาอีกหนึ่งตาราง ซึ่งมีมิติเดียว Cz(z) ดังนั้นจึงไม่ใหญ่เกินไป และไม่เกินความสามารถของหน่วยความจำของระบบกล้องสมัยใหม่ขนาดเล็ก แต่ให้เวลาการค้นหาที่รวดเร็ว O(1) เหมาะสำหรับระบบเรียลไทม์ กลยุทธ์การสอบเทียบของเราช่วยให้ระบบสเตอริโอที่สร้างขึ้นจากกล้องเดี่ยวสามารถวัดความลึกในฉากได้:กล้องเดี่ยวไม่ต้องการการซิงโครไนซ์อิเล็กทรอนิกส์และสร้างจากชิ้นส่วนออปติคอลที่หาง่ายและราคาถูกเพียงชิ้นเดียว - ปริซึมมุมฉาก

Thesis	Single Camera Stereo Vision Camera Calibration
Student	Mr. Athukoralage Supun Gayantha Dissanayaka
Student ID.	61601098
Degree	Master of Engineering
Program	Robotics and AI Engineering
Year	2024
Thesis Advisor	Prof. Dr. Pitikate Sooraksa
Co-Thesis Advisor	Assoc. Prof. Dr. Somyot Kaitwanidvilai

ABSTRACT IN ENGLISH

A simple stereo system can be constructed from a single camera using a prism in the optical path to provide the required two views of a system. The simplicity of these systems has several advantages, particularly if the target is an underwater robot, where compact size and ability to seal the optical components are key factors. However, dispersion by the prism, in addition to the lens distortion, makes calibration challenging. By using a model-free approach, we were able to calibrate a prism-based stereo system effectively. We also aimed to use readily available 45° prisms, which present significant dispersion in the system, but retain simplicity and reduce costs, compared to custom low angle prisms. Modern LEDs provide high intensity, low bandwidth light sources and we used a set of three sources, roughly centered on the RGB channels of a readily available commercial camera. Our system used a circular target pattern covering the binocularly visible region in the scene and collected sets of images at known distances, using three separate light sources. From these images, we generated two look-up tables, one for each pixel in the image and a disparity derived by matching corresponding points, $Cp(u,v,du)$, which has three dimensions, and another look-up table, which has a single dimension, $Cz(z)$, so are not quite large, and not beyond the memory capability of even small modern camera systems, but provide fast, $O(1)$, lookup times, suitable for real-time systems. Our calibration strategy enables a simple stereo system built from a single camera to

measure depths in a scene: the single camera requires no electronic synchronization and is built from a single, inexpensive, and readily available optical component – a right-angle prism.

ACKNOWLEDGEMENT

I am deeply grateful to everyone who has supported me throughout my journey to complete my master's degree in Robotics and Computational Intelligence. This work would not have been possible without the guidance, encouragement, and assistance of many individuals. First and foremost, I would like to express my sincere gratitude to my advisors, Assoc. Prof Dr. John Morris, Prof. Dr. Pitikhate Sooraksa, Assoc. Prof Dr. Somyot Kaitvanidvilai for their continuous support, insightful feedback, and invaluable guidance. Their expertise and advice have been instrumental in shaping this research and helping me navigate through the challenges encountered. A heartfelt thank you to my family for their unwavering support and encouragement throughout my academic journey. To my mom, Rohinie Dissanayaka, thank you for always believing in me and for providing me with the foundation to pursue my dreams. To my siblings, Dasun Dissanayaka and Lakmali Dissanayaka for your encouragement and love have been a source of strength for me. I am also grateful to my friends and colleagues who have provided moral support, inspiration, and encouragement during the thesis writing process. Special thanks to Ashan Eranga for their patience, understanding, and for providing a listening ear when needed. Finally, I would like to acknowledge the support of King Mongkut's Institute of Technology, Robotics and AI department for providing the necessary resources and a conducive environment for research. This journey has been challenging yet rewarding, and I am thankful to everyone who has been a part of it.

Thank you all

Athukoralage Supun G Dissanayaka

Contents

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS AND SYMBOLS

$\mathcal{C}_p$	Look up table of Corresponding points
$\mathcal{C}_z$	Look up table of distances
$\mathcal{O}(1)$	Look up time

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND

This project was motivated by a desire to equip an underwater robot with a single camera with the ability to navigate accurately underwater – detecting obstacles and possible threats in its path. Traditional stereo rig setups face significant challenges in cameras alignment and calibration due to the complex setup needed in an underwater environment. Also handling the refraction of the two cameras and thermal gradients, leading to varying refractive indices, is more difficult underwater. The potential to use a single prism, easily added and sealed into the robot, but providing the two images needed to accurately measure distance, was the driving force and the design and calibration of such a system was studied in detail. This work focused on one important aspect: camera calibration - a process to obtain intrinsic focal length and resolution and extrinsic position and orientation of the virtual cameras formed by prism refraction.

1.2 INTRODUCTION

Computer stereo vision modeled human vision that produces a pair of images by projection of light rays onto the retina of the eyes. From the resultant two-dimensional images, we can compute 3D depths enabling us to interact with the 3D-surroundings. Various algorithms were developed to replicate the ability of human stereo vision. 3D reconstruction obtains the shape and appearance of an arbitrary object or scene at a given depth using various computer techniques for computing depths. These methods fall into three classes:

1) triangulation,
2) time of flight and
3) modulation.

Miquel and Gabriel (2015) showed that each of these falls into multiple categories and 3D depths may be obtained by a single method or a combination of them.

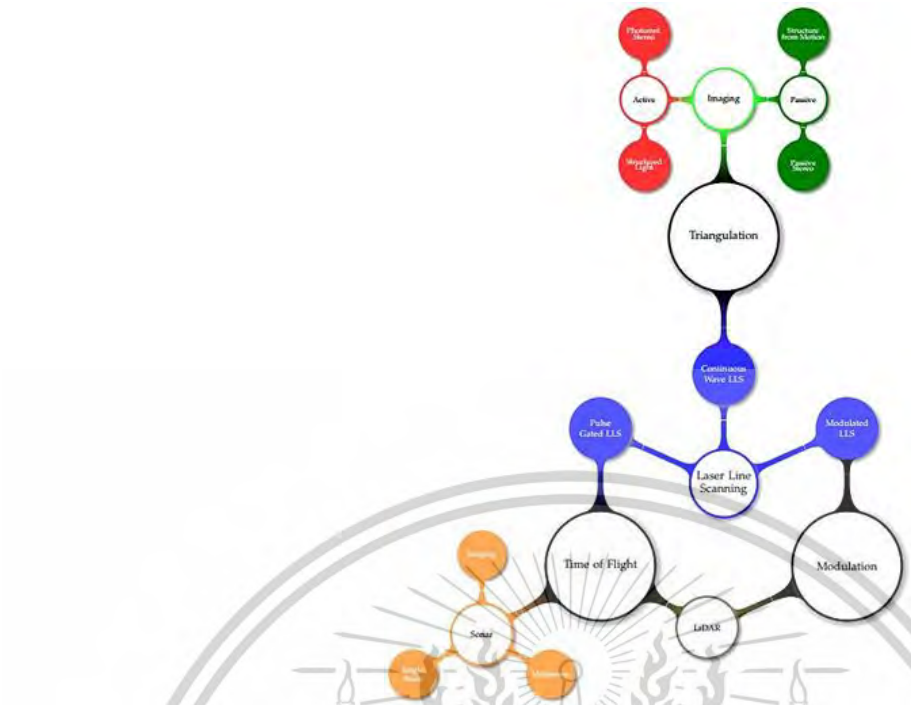


Figure 1.1 3D reconstruction methods (Massot-Campos & Oliver-Codina, 2015)

1.2.1 Passive methods

Passive methods, obtain the 3D depth without interfering with the scene. Commonly used methods are

1) binocular stereo vision and
2) monocular cues.

1.2.2 Active methods

Active methods contact the object or project energy to acquire 3D data. Numerical approximations are used to construct the 3D profile and create a model. These sensors, interact with the scene using various probe sources:

1) ultrasound, i.e. sonar,
2) microwave, i.e. radar,
3) light, i.e. LIDAR,
4) moving light sources or even
5) mechanical range finders, e.g. depth gauges,

to measure distance to an object in a scene.

1.2.3 Passive methods

Passive methods obtain the 3D depth without interfering with the scene. Commonly used methods are

1) binocular stereo vision and
2) monocular cues.

1.2.3.1 Triangulation

Triangulation methods target common features, using two or more depth acquisition devices. Commonly, two cameras or a 'stereo rig' will be used to find the matched scene points. However, the remaining features can be back projected onto the world as light rays going out from two cameras. The triangle formed between the feature in the space and the two cameras is the basis for triangulation (Massot-Campos & Oliver-Codina, 2015)

1.2.3.2 Time of flight

Time discrimination methods use depth sensors consisting of an infrared (IR) emitter and a receiver. A reflection from IR signal projected to a scene can be received by the receiver. Common types of signals are high-speed pulse and continuous wave. Distance to the object is computed from half of the traveled length, based on the travel time of the pulse or the phase shift between the emitted and received waves. Generally, the speed of the probe in the medium needs to be known and considered: usually for light in air, this can be assumed constant, but, in a water medium, temperature, salinity and pressure changes affect the speed of sound (Massot-Campos & Oliver-Codina, 2015) (Nguyen et al., 2018).

1.2.3.3 Modulation

Modulation uses frequency domain differences, e.g. amplitude and phase of a modulated signal. In some of these systems, components, in a relatively narrow band, are lost due to the scattering of the pulses. The receivers are usually photo multiplier tubes or photon counters. These sensors are triggered during a time window, and the incoming light is integrated. After demodulation, 3D information can be obtained from the phase difference.

CHAPTER 2
LITERATURE REVIEW

2.1 DEPTH ACQUISITION METHODS

Recent depth sensors allowed better quality in preserving 3D models. Different methods are selected for specific applications, based on some or all the following criteria (Gomes et al., 2014) :

1) technical reliability - precise sensors will generate more accurate 3D models.
2) price - low-cost solutions are preferred.
3) portability and flexibility - lightweight and mobile solutions may be needed.
4) acquisition time, when dynamic scenes must be followed.
5) resolution - required by the application.
6) repeatability or precision - drifts of the measurements may not be acceptable.
7) accuracy - allowed statistical variations of repeated measurements.
8) environmental sensitivity - robustness in varying light, temperature wind and pressure conditions etc.

This section summarizes the 3D acquisition methods - referring to the techniques in Figure 1.1.

2.1.1 Sonar, Radar, LIDAR and Time of Flight

There are a large variety of probe pulses: systems can emit sound or EM radiation - at a wide range of frequencies, ranging from the relatively low RADAR frequencies to optical frequencies, e.g. LIDAR - generally as pulses of sounds, phonons or photons, and receive reflections from scene objects.

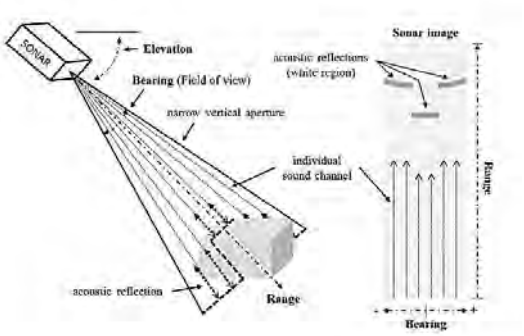


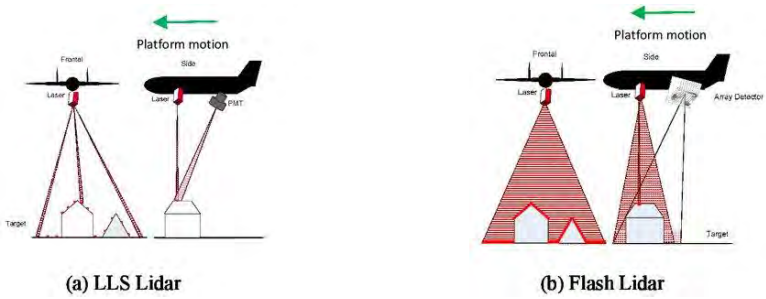
Figure 2.1 Sonar image acquisition - taken from Y. Lee et al. (2016)

2.1.2 Light detection and ranging (LIDAR)

A laser is scanned over a wide area to sense a large scene volume. Weather monitoring systems often use LIDAR to measure clouds over long distances.

2.1.3 Laser line scanning (LLS)

Laser line scanning systems deploy lines of laser pulses, which can use less energy than area-based emitters, e.g. those for LIDAR, and thus may be suitable for longer range imaging. Range to an object can be also measured by pulsed lasers and range measuring by the phase difference between the transmitted and retrieved signal Phase difference methods use lasers that continuously emit light- or continuous wave CW lasers (Wehr & Lohr, 1999).



(a) LLS Lidar (b) Flash Lidar

Figure 2.2 Lidar image acquisition - taken from Ouyang et al.,(2014)

2.1.4 Structured Lighting Systems

Structured lighting systems project known patterns on to the scene and capture images with a single camera (Bell et al., 1999)

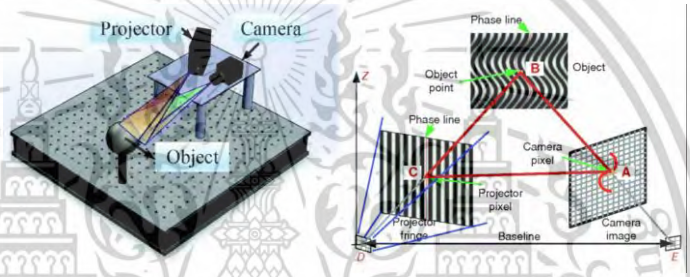


Figure 2.3 Structured lighting system - taken from Bell et al.,(1999)

In these systems, the pattern projector takes the place of the second camera: the position of sections of the pattern in the image vary with the depth and depth can be calculated from these 'disparities' as with two camera systems. They work well in scenes with low texture: the pattern adds the texture needed to assist matching. This strategy has become well known with the advent of a commercial system, Kinect (J. M. Chen et al., 2013), but the strategy has been known since the early 1990 (Forbes, 2019) These systems are fundamentally similar to two camera systems - the optics of the pattern projector must have similar capabilities or precision - and suffer from the same problems with occlusions, so they are not discussed further here.

2.1.5 Structure from motion

Structure from Motion is a triangulation technique, which takes multiple images of a scene using a single camera: either the camera or the scene can move, e.g. by placing a target object on a rotating table. Having obtained multiple images, SfM can use any available matching algorithms, for example, Schonberger & Frahm, (2016) started with feature extraction and matching, followed by geometric verification - see

Figure 1.3. The resulting scene graph was the initial step towards a model by selected two view reconstructions, before incrementally registering new images, triangulating scene points, filtering outliers and refining the reconstruction using bundle adjustment Schonberger & Frahm, (2016)– see Figure 1.5.

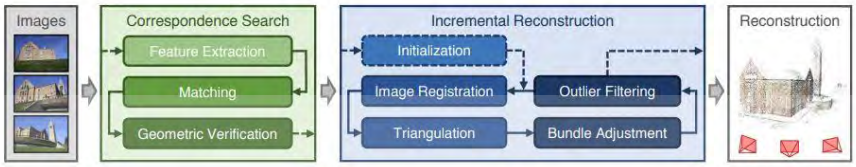


Figure 2.4 SfM process flow - taken from Schonberger & Frahm (2016)

Next, in this chapter, I discuss available stereo vision techniques, covering conventional stereo calibration and correspondence algorithms, depth recover, and single camera stereo vision techniques.

2.2 STEREO VISION FUNDAMENTALS

Computer based vision systems enable us to measure the positions and shapes of objects in any environment. Generally, they use the same techniques, as animals with two eyes, to assess the distance to objects confronting them: Therefore, stereovision requires two or more cameras to capture the same scene to calculate the disparities for depth recovery. It has been used for many years in robotics, medicine, industry or pattern recognition, artificial intelligence and many other fields as briefly:

Table 2.1 Traditional stereo vision applications

Field	Stereovision Application	Reference
Robotics	Autonomous Navigation (Mobile Robots)	(P. Cui & Yue, 2007), (Wooden et al., 2010), (Sanchez-Rodriguez & Aceves-Lopez, 2018)
	Outdoor Mapping	
	Collision Avoidance	
	Stereoscopic particle image velocimetry	

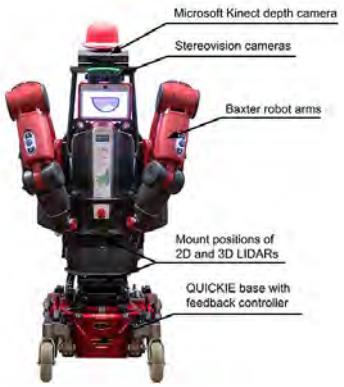


Figure 2.5 Stereo vision equipped Robot De Niro (Falck et al., 2020)

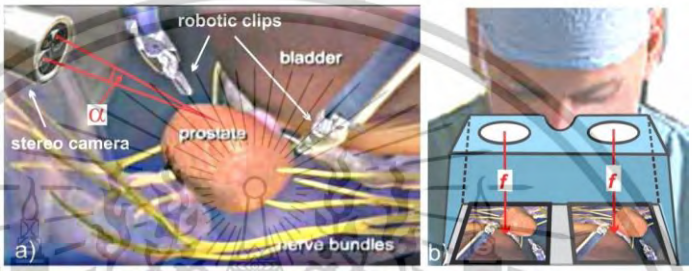


Figure 2.6 Stereo vision assisted surgery – taken from Rosas (2011)

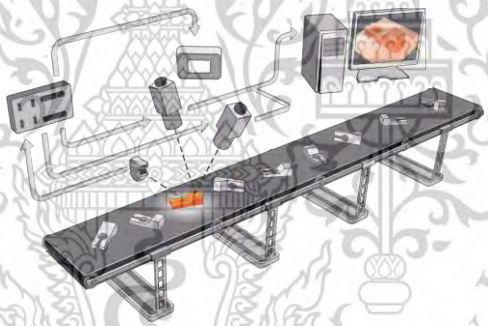


Figure 2.7 Stereo vision use in a production line (3D Vision: MVTec software)

In industrial processes, stereo vision inspection systems monitor both visual and physical characteristics of a product to maintain quality and consistency of the production. Physical measurement of manufactured products, e.g. surface roughness, and complex shape contour geometry are being replaced by non-contact stereo vision systems. As a result, industrial product physical characteristics of length, height, outside diameter and other dimensions, as well as the visual appearance, can be measured without distortion created by contact measuring methods. Robotics applications use stereovision to assist navigation. All these applications use dual camera stereovision to achieve 3D perceptions, similar to human vision, where the same scene point is viewed

from two or more viewing angles. However, it is possible to use a single camera to obtain depths, with additional support features, for example mirrors, prisms mounted on the lens and additional optical components, as long as separate views are obtained.

2.3 HISTORY OF STEREOVISION

Stereoscopy was discovered by Sir Charles Wheatstone in 1838. Later it was improved by Sir David Brewster in 1849, whose stereograph had a pair of images, generated by a camera with two lenses placed 60 mm apart, in same horizontal line to simulate the position of human eyes, and then mounted the prints side by side laterally. The final stereo image could be viewed; by the stereoscope, which had two eye pieces, which could view the laterally mounted images, placed in a holder in front of the lenses. Then, the two images were brought together and matched by a human brain to create a 3D illusion. In the past, stereographs were created for a variety of scenes. Mostly landscapes and monuments and composed narrative scenes of a humorous or slightly suggestive nature. A simple hand-held stereoscope was introduced by Holmes (Pietrobruno, 2011). Stereographs became famous after Queen Victoria expressed an interest in them, when they were exhibited at the 1851 Crystal Palace exhibition (Pietrobruno, 2011). Later, stereographs became educational and a recreational device, with considerable impact on public knowledge and taste during the 19th century.

Figure 2.8 Pocket Stereoscope with a test image, taken from, Brewster, (1856)

2.4 CONVENTIONAL STEREO VISION SYSTEMS

Conventional stereo camera systems generate accurate 3D models of real objects by capturing an image of the scene from different orientations. These viewing

angles can be obtained by using more than one camera, moving a single camera around the object or rotating the object within the field of view of a single camera. The separate viewing angles of the scene generate a ‘disparity’ (or separation) between objects appearing in the images. Stereo vision is a passive technique, i.e. it acquires high resolution images, and no energy emission, whereas other depth measuring devices, e.g. laser scanners, sonar or radar, are active techniques that emit a probe beam. Stereo vision is discussed in detail in many textbooks and review papers: a selection of these appears in Table 2.2

Table 2.2 Stereo vision sources

Author	Notes	Reference
Conventional stereo techniques		
Okutomi and Kanade	Multiple baseline stereo	(Okutomi & Kanade, 1993)
Faugeras	Three-dimensional computer vision: A geometric viewpoint	(Faugeras, 1993)
Hartley and Zisserman	Multiple view geometry	(Andrew, 2001)
Delaunoy et al.	Full 3D Helmholtz stereovision algorithms	(Delaunoy et al., 2010)
Reviews		
Barnard and Fischler	Computational stereo	(Barnard & Fischler, 1982)
Dhond and Aggarwal	Structure from a stereo review	(Dhond & Aggarwal, 1989)
Scharstein and Szeliski	Evaluation of stereo vision algorithms	(Scharstein & Szeliski, 2002)
Lazaros et. al	Review of stereo vision algorithms	(Lazaros et al., 2008)
Hamzah and Ibrahim	Survey on disparity map algorithms	(Hamzah & Ibrahim, 2016)
Kumari and Kaur	Stereo matching techniques for 3D vision	(Kumari & Kaur, 2016)
Perez et. al	Robotic vision techniques in industrial environments	(Pérez et al., 2016)

Machine vision algorithms have generally assumed several characteristics of objects in the scene and used several computation steps. These assumptions involved are (Z. Zhang, 2000), (Andrew, 2001)

- objects have Lambertian surfaces (i.e. their appearance will not vary with view angle),

- the scene consists of smooth surfaces,
- the cameras have been calibrated and
- aligned in epipolar geometry.

Common algorithm steps are:

- model the camera noise, differences in electronic gains,
- identify matching image points in the two images,
- depth calculation and
- interpolation

See figure 1.1.

2.5 CORRESPONDENCE PROBLEM

A key problem for automated stereo vision is that you need to find or match corresponding points in each image, from which the disparity and thus the depth is calculated. It determines three-dimensional coordinates X, Y, Z of an object using two-dimensional image coordinates x, y Thus, an object point must appear in two or more images. Two or more views often lead to occlusions or points which are visible in only one image or occluded by an object in front. Figure 1.2 shows a classical stereo vision system and the coordinates of points in the scene, in which the optical axes of the cameras are parallel to each other and perpendicular to the baseline, are calculated from:

- X_w, Y_w, Z_w - World coordinate system
- X_R, Y_R, Z_R - Optical center of the right camera coordinate system
- X_L, Y_L, Z_L - Optical center of left camera coordinate system
- x_{il}, y_{il}, z_{il} - Left camera image coordinate system
- x_{ir}, y_{ir}, z_{ir} – Right camera image coordinate system
- LR represents the epipolar line
- Q is an arbitrary scene point in the world coordinate system

where L is the baseline distance connecting the optical centers of the cameras in the X_w direction and f is the focal length (assumed to be identical for both cameras). Coordinates of scene points are calculated from:

$$X_w = \frac{\frac{L(x_l - x_r)}{2}}{x_l - x_r} \tag{1.1}$$

$$Y_w = \frac{\frac{L(y_l - y_r)}{2}}{y_l - y_r} \tag{1.2}$$

$$Z_w = \frac{\frac{L(z_l - z_r)}{2}}{z_l - z_r} \tag{1.3}$$

$x_l - x_r$ is the **disparity**, which is the relative position of a corresponding scene point, in the two image planes. A detailed model can be found in many texts, including Faugeras (1993).

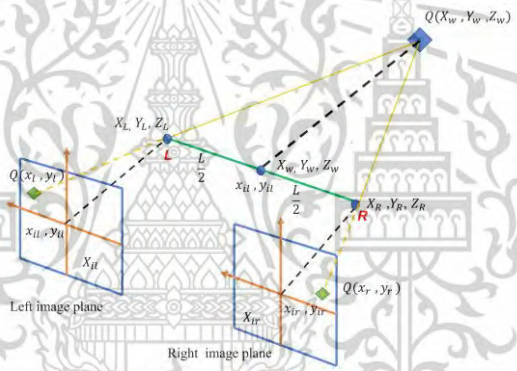


Figure 2.9 Two camera stereo vision system, showing a general scene, Q (X_w Y_w, Z_w) the optical centers of the two cameras at (X_{LR}, YLR, Z_{LR})

Solving the correspondence problem, i.e. identifying pairs of corresponding points, has occupied the attention of many researchers for several decades now. It is reviewed extensively by Scharstein & Szeliski (2002) whose review provided a taxonomy and metrics for the performance of correspondence algorithms They grouped correspondence algorithms into two groups - local ones, which used local information around an image point to determine the corresponding point in the other image - and global ones, which used a large neighborhood of each image point to find its corresponding pair. Figure 2.10 shows the steps of development of stereo vision algorithms.

Figure 2.10 Development steps of stereo vision algorithms

2.6 LOCAL METHODS

Local methods trade accuracy for speed and depend only on intensity values, within a finite support window; they compute the disparity at a given point using the intensity values in that window. Various cost functions were used to determine whether points match or not: one of the simplest uses sums of squared differences of intensities in a small window to find the disparity and selects the minimum at each pixel, for quite simple code. Other functions are based on rank transformation and normalized cross-correlation methods, for cost computation and matching to determine disparity - see figure 2.11.

The disparity map, using these algorithms, is commonly achieved by a winner take all optimization (WTA): the disparity having the minimum cost is assigned to each pixel of the image.

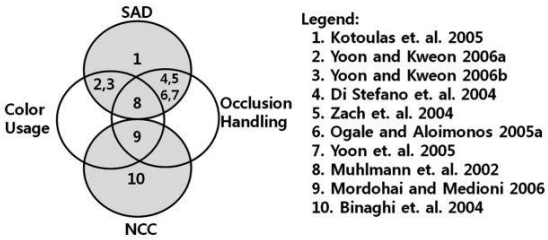


Figure 2.11 Local methods categorization - taken from Lazaros et al. (2008)

2.7 GLOBAL METHODS

Global methods treat the disparity as a minimization problem using a large area of the image. These functions have two terms: the data term, which penalizes solutions that do not match the target, and a smoothness term, which implements a piece wise smoothing to the neighboring pixels. Although global methods produce better results, they are computationally expensive. Hence, they are not ideal for real time stereo vision applications. (Scharstein & Szeliski, 2002), (Hamzah & Ibrahim, 2016); (Kalarot & Morris, Feb 4-6)

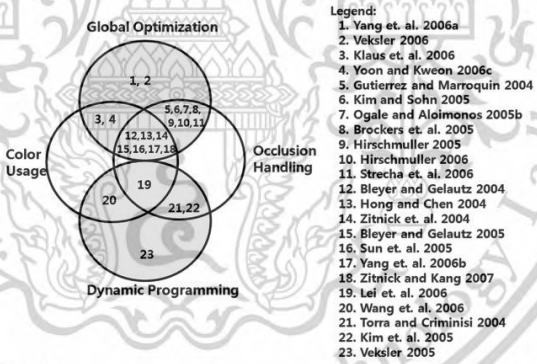


Figure 2.12 Global methods categorization - taken from Lazaros et al. (2008)

2.8 DISPARITY MATCHING COST AND AGGREGATION

Generally, stereo vision algorithms compare intensities of two pixels in the generated images to an associate them with a single scene point. Therefore, they require a cost criterion to match the pixels. A scene point, Q, is viewed from the optical centers, L and R of the two cameras. It produces one image in the left plane, $x_{il,}$ and another in the corresponding right image on the x_{ir} plane - see figure 2.9. The points, $\boldsymbol{Q}(\boldsymbol{x_l}, \boldsymbol{y_l})$ and $\boldsymbol{P}(\boldsymbol{x_r}, \boldsymbol{y_r})$, represent the matching pixel positions in the left and right

image planes. In a correctly aligned system, left and right matching image points must be located on a horizontal line, $\boldsymbol{LR}$, an epipolar line. This is commonly referred to as the epipolar constraint (Gong et al., 2007). The difference or disparity between the corresponding pixels in the two images can be associated with a depth value via 3D projection. This alignment reduces one factor in the computational complexity from $\boldsymbol{\mathcal{O}(w^2h^2)}$ to $\boldsymbol{\mathcal{O}(w^2h)}$ – see figure 1.6, because the search for matches can be constrained to a straight line of length, $\boldsymbol{w}$, the epipolar line. Early disparity computation algorithms used pixel-based metrics. These methods include

- absolute differences,
- squared differences,
- sampling-insensitive absolute differences (modification of above methods) and
- truncated absolute differences (modification of above methods) that can be applied to grayscale or color images. Area or window-based methods use more matching data than the pixel-based methods. These techniques are generally more precise, because they consider many pixels covering an image region of interest (ROI). Some algorithms based on areas, include
- sum of absolute differences
- sum of squared differences
- normalized cross correlation
- rank transformation and
- census transformation.

which calculates the cost of matching over a ROI. A common disadvantage of these methods is that they assume that all pixels, within the ROI, have the same disparity value. However, this is not true for depth discontinuities or edges in the image region, therefore, inappropriate selection of the image region results in poor depth estimations.

2.9 DISPARITY COMPUTATION

In local methods, the disparity for each pixel is usually selected by a winner-take-all strategy, where the chosen disparity, d_Q is determined from the minimum cost value,

$$d_Q = \ minC(Q, d) \text{ for } d \in D \qquad 1.4$$

where, D represents the set of all allowed disparities and $C\ (Q, d)$ is the aggregated cost obtained from the matching cost calculation. This technique has been used extensively in early papers, see references in Scharstein & Szeliski (2002).

There are several limitations of this method:

- a point in one image can match several points in the other image,
- occluded regions lead to incorrectly matched pixels,
- due to summing or averaging over regions
- accuracy of local methods, depend on the cost computation and aggregation stages and
- occlusions are not recognized correctly.

Disparity calculation, using global methods, assumes the scene is smooth, except at boundaries. Therefore, neighboring pixels will have similar disparities: these techniques solve the energy minimization problem. Hence, the optimal energy disparity function is

$$E(d) = E_{data}(d) + \alpha E_{smooth}(d) \qquad 1.5$$

where $E_{data}(d)$ is an energy function for the matching cost at point, (x, y).
The smoothing function $E_{smooth}(d)$ assumes neighboring pixels have similar disparities and α is a weighting factor.
Global methods include:

- dynamic programming
- belief propagation,
- graph cut and
- multiresolution energy functions (Gudis et al., 2012).

2.10 DISPARITY REFINEMENT

Disparity refinement is often applied to reduce noise and smooth the disparity map. Common steps in this task are

- regularization, reducing variations of inconsistent pixels,
- occlusion filling and
- interpolating to `fill in` when disparities are unclear or
- aggregating areas of the image, e.g. texture less areas and edges (Scharstein & Szeliski, 2002).

2.11 ABSOLUTE DIFFERENCES (AD)

This pixel-based function calculates the differences of intensities between the pixels in x_{il} and x_{ir}, given as:

$$AD(x, y, d) = | x_{il}(x, y) - x_{ir}(x - d, y)| \quad 1.6$$

is probably the simplest matching function and runs fast compared to other methods. An AD function suits low texture scenes. However, it does not provide correct disparity maps for highly textured scenes. A truncated absolute differences function has been implemented to overcome this: it uses colors and gradients to match pixels in varying illumination scenes (L. Wang et al., 2006).

2.12 SQUARED DIFFERENCES (SD)

As an alternative to absolute differences, the larger differences are magnified by squared differences between corresponding pixels in x_{il} and x_{ir} can be computed:

$$SD(x, y, d) = | x_{il}(x, y) - x_{ir}(x - d, y)|^2 \quad 1.7$$

2.13 FEATURE BASED STEREO MATCHING

Features, e.g. edges, shapes, segmentation, textures or gradient peaks, can be used to compute the matching cost. Also, statistical characteristics, e.g. histograms, medians or minima, and transformation features, e.g. Hough, wavelet, Gabor transformations, are used to calculate the matching cost. For example, Sharma & Moon (2013) used a feature-based algorithm, obtained from scale invariant feature transform (SIFT), for autonomous navigation. Although the computational cost was low, due to the

correlation between selected features, the SIFT algorithm was often unable to produce the complete disparity map (Hamzah & Ibrahim, 2016). Further limitations were low accuracy in texture-less and occluded areas.

2.14 SUM OF ABSOLUTE DIFFERENCES (SAD)

Sum of Squar Sum of absolute differences is an area based matching cost computation function; it sums absolute differences between intensity of each individual pixels in a selected region.

$$S_{AD}(x, y, d) = \sum_{(x,y)\,\epsilon\, r}^{n} | x_{il}(x, y) - x_{ir}(x - d, y)| \quad 1.8$$

where the pixel intensity differences are computed over a region - typically, a small rectangular region. Tippetts et al., (2011) analyzed the performance of SAD using a real time human pose tracking, while Lee and Shrama (2011) applied SAD to calculate matching costs in real time, using a Graphical Processing Unit (GPU), because SAD costs for multiple pixels can be computed simultaneously. Both concluded that SAD had a low computational complexity, while maintaining reasonable accuracy of the disparity map.

2.15 SUM OF SQUARED DIFFERENCES (SSD)

SSD is a variant of SAD, which computes the squared differences in pixel intensities, rather than absolute differences.

$$SSD(x, y, d) = \sum_{(x,y)\,\epsilon\, r}^{n} | x_{il}(x, y) - x_{ir}(x - d, y)| \quad 1.9$$

2.16 DYNAMIC PROGRAMMING (DP)

Dynamic programming is a common problem-solving approach: it assumes that it can break a problem into subproblems and that the optimal solution to the larger problem depends only on the solutions to the subproblems, i.e. it does not need to back-track. Gimel'farb's Symmetric Dynamic Programming Stereo (SDPS) computes the minimum cost using a matrix of already computed pixel pair matching costs between corresponding scan lines of two images (G. L. Gimel'farb, 2002). As with all matching

algorithms, image areas affected by occlusions, i.e. the matching pixel is occluded in one of the two images, are handled by bi-linear interpolation technique. to a single pixel in the other corresponding image. However, it is theoretically impossible to determine a unique correct disparity for occluded pixels, there are always multiple possible solutions. One contribution of SPDS was that it provided a simple and fast way to assign a disparity, which was a possible solution, for occluded pixels. It assumed a virtual Cyclopean eye in the center between the left and the right cameras, classed pixels as either binocular (seen by both cameras) or monocular (seen by only the left or the right camera). Binocular points led to known disparities, but monocular points, occurring between two binocular points, were assigned to disparities by a linear interpolation. Another limitation of DP is that it is required to have the relative ordering of the pixels on a particular scan line remain unchanged between the corresponding image regions, i.e. it can be confused by a narrow pole in front of a background object - perhaps not unlike human vision, which may sometimes be confused by such illusions. However, DP is faster in reconstructing 3D surfaces. Therefore, it has been used in real time stereo vision applications (Khan et al., 2009) NCSM (Noise driven concurrent stereo matching) was introduced to estimate the image noise and obtain more stable disparity maps by using SPDS (G. Gimel'farb et al., 2006).

2.17 GRAPH CUT (GC)

Graph cut algorithms find a globally optimal segmentation solution. A graph is partitioned into two disjoint sets, then by simply removing edges connecting the two sub-graphs, we can compute the degree of dissimilarity between the two sub graphs as the total weight of the edges that have been removed. This method performed well in texture less areas and near discontinuities (Scharstein & Szeliski, 2002) .However, the computational costs remained high. As an enhancement, Wang et al. (2006) used a hierarchical bilateral disparity structure (HBDS) graph cut algorithm, to speed up computation and to improve accuracy of the disparity maps (L. Wang et al., 2006).

2.18 BELIEF PROPAGATION (BP)

Belief Propagation is another global energy optimization method. It works by passing messages around a graph, defined by a four-connected image grid. This method is computationally intensive, even for low resolution images and small numbers of

disparity levels. Let hw be the image size, L be the number of disparity levels and T be the iterations, then, computational complexity is originally $O(ThwL^2)$. It is also memory intensive, providing a challenge for high resolution images at video rates. A standard four-connected belief propagation algorithm requires at least $4NL$ variables to store all the messages and NL for the data cost (Q. Yang et al., 2010). However, a Bumblebee XB3 camera could capture stereo pairs with 1280 × 960 resolution and more than 400 disparities at 16 frames per second.

2.19 NEURAL NETS (NN)

The neural net community believes that neural nets can solve every problem and have great promise. Neural nets have demonstrated to be useful when classifying objects. For example, in image classification, Hamid et al.,(2020) implemented a convolutional neural net (CNN) for vehicle counting. They were able to classify different types of vehicles, using their CNN based counting system (Hamid et al., 2020). However, calculating accurate depths increased the computational complexity. classes Zbontar et al.(2016). used CNN to produce matching costs as in figure 2.13. Their net had eight layers, fed by 9×9 gray image patches. The first layer was a 5×5 convolution layer with 32 kernels, followed by seven fully connected layers. The output vectors of the convolutional layers (Layer 1) passed through Layer 2 and 3 (fully connected layer with two hundred neurons each). The vectors processed from left and right image patches were concatenated together from two channels of the 200-dimensional vector (left and right) into a single 400-dimension vector. This was passed to Layer 4 of their net. Layers 4 through 7 were single layers with three hundred neurons each. The final Layer produced a distribution of good and bad match

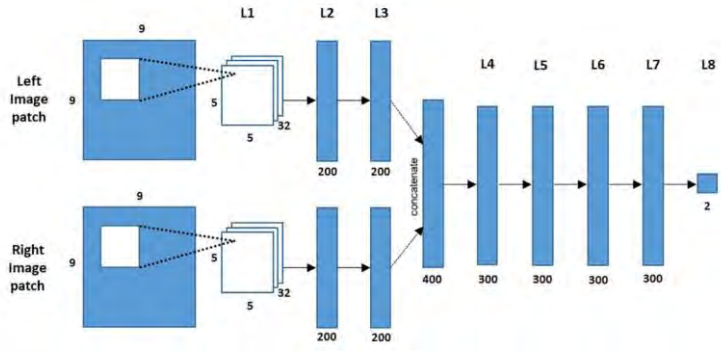


Figure 2.13 CNN for stereo matching - taken from Hamid et al. (2020)

2.20 OPTICAL ARRANGEMENTS

This section discusses optical arrangements used to capture stereo images. It starts with `traditional' two camera systems, which uses two cameras to capture two different images. It then discusses single camera systems.

2.20.1 Traditional multiple camera systems

Multiple camera systems take many forms, from simple pairs of cameras, mounted on a rigid base, to various multiple camera arrangements and moving camera systems, which simply take multiple images from a moving camera, effectively mimicking multiple cameras. These systems are described in texts (Kaczmarek, 2017) or reviews (Holzmann & Hochgatterer, 2012). Some examples are mentioned in the introduction, see figures 1.6 to 1.8 and are not further described here, because my work focused on a single camera system.

2.21 COMPUTER BASED SINGLE CAMERA STEREO

With the rapid development of the mobile stereo vision application in recent years, minimal hardware, faster response and low-cost solutions were required. The use of single camera stereo vision has demonstrated better prospects for mobile systems, as it allows compact and low-cost designs. Additional optics allow the single camera to capture two different views of the scene. In mobile robot applications, the need for multiple cameras or complex optics generally leads to physically large systems. Particularly in an underwater environment, where these key components must be mounted in a sealed casing, small and compact optical systems are desirable. Many single camera systems have been designed, with varying complexity additional optics, and they are reviewed in this section. Table 3 summarizes single camera stereo

vision techniques reported in the literature. Previous work indicates that compact optical systems are desirable due to the mobility of the stereo vision application.
In this dissertation, I introduce a single camera stereo vision system that uses a bi-prism and a single camera. The prism divides the scene into two halves to obtain the disparity for further three-dimensional construction.

Table 2.3 Previous work on single camera stereo vision systems

Author	Notes	Reference
Mirror techniques		
Goshtasby and Gruver	Single camera and two mirrors	(Goshtasby & Gruver, 1993)
Inaba et al.	Single camera and four mirrors	(Inaba et al., 1993)
Zhang and Tsui	Single fixed focus camera with a planar mirror	(Z. Y. Zhang & Tsui, 1998)
Glukman and Nayar	Single camera with a fisheye lens and a convex mirror	(Gluckman & Nayar, 1999)
Ramsgaard et al.	Single camera and two orthogonal planar mirrors	(Ramsgaard et al., 2000)
Kim et al.	Single camera and a rotating mirror	(Kim et al., 2005)
Jangb et al.	Single camera with conic mirrors	(Jang et al., 2005)
Mouaddib et al.	Catadioptric omnidirectional camera and multiple mirrors	(Mouaddib et al., 2005)
Sooyeong et al.	Concave lens and a convex mirror	(Yi & Ahuja, 2006)
Duvieubourg et al.	Single camera, two lateral mirrors and a rotating prism	(Duvieubourg et al., 2006)
Li et al.	Single camera with a fisheye lens and a convex mirror	(W. Li & Li, 2011)

Feng and Pan	Single auto focus camera with a planar Mirror	(Feng & Pan, 2015)
Zhou et al.	Single camera and two pyramid mirrors	(Zhou et al., 2016)
Yu and Pan	Single camera and four-mirror adapter	(Yu & Pan, 2016)
S Barone et al.	Single low frame rate camera and two Planar mirrors	(Barone et al., 2018)
Zhong et al.	Single color camera and two planar mirrors	(Zhong & Quan, 2018)
Diffraction gratings		
Trivi and Rabel	Single camera, monochromatic light and diffraction grater	(Trivi & Rabal, 1988)
Henao et al.	Adding a diffraction grafting to the Burch experiment	(Henao et al., 1993)
Rabal et al.	Speckle pattern shearing interferometry	(Rabal et al., 1996)
Xia et al.	Transmission diffraction grater and a single camera	(Xia et al., 2013)
Pan and Wang	Transmission diffraction grater, single camera and a monochromatic light	(Pan & Wang, 2013)
Xia et al.	diffraction graft assisted fluorescent microscope	(Xia et al., 2014)
Prism Based		
Lee and Kwoen	Single Camera and a Biprism	(D. Lee & Kweon, 2000)
Lim and Xiao	Biprism single camera system	(Lim & Xiao, 2005)
Xiao and Lim	Trinocular vision system using a multi faced prism	(Xiao & Lim, 2007)
Cheng et al.	Micro prism array plate replacing the single prism	(C.-Y. Chen et al., 2008)

Cui et al.	Muti ocular prism stereo using geometric approach	(X. Cui et al., 2012)
Kee et al.	Virtual epipolar line construction on single camera stereo vision	(Kee et al., 2012)
Zhao et al.	Geometry based algorithm to solve the stereo correspondence in single camera multi faced prism system	(Zhao et al., 2012)
Lim et al.	Geometric ray approach to calibrate the extrinsic parameters of the generated virtual cameras	(Lim et al., 2013)
Wang et al.	Geometry based image rectification for trinocular single camera system	(D. Wang et al., 2013)
Wu et al.	Modified virtual points approach by ray tracing for single camera stereo vision	(L. Wu et al., 2014)
Cui et al.	Perspective model for prism based stereo vision	(X. Cui et al., 2015)
Kee et al.	Parameter and error analysis of single lens prism based	(Kee et al., 2015)
Qian and Lim	Image distortion correction for thick prisms and lens	(Qian & Lim, 2016)
Yang et al.	Micro prism array based compact stereo endoscopic camera having a single image sensor	(S.-P. Yang et al., 2016)
Wu et al.	Bi- prism based stereo vision for long working distance	(L. Wu et al., 2018)

Chen et al.	Biprism based stereo vision for underwater environment	(Yuan et al., 2019)
Telecentric Lens, Optical filters, Beam splitters		
Wu et al.	Bilateral telecentric lens with a bi prism	(L. Wu et al., 2015)
Wu et al.	Telecentric lens bi prism system error analysis and correction	(L. Wu et al., 2016)
Li et al.	Monocular stereo vision 3CCD color camera, Zoom lens, beam splitter, optical band pass filters and plane mirrors	(C. Li et al., 2019)
Reviews		
Pan et al.	Review of Single Camera Stereo	(Pan et al., 2018)

2.22 GLASS PLATE ARRANGEMENTS

Nishimoto and Shirai (1988) arranged a movable glass plate in front of the camera, that created a stereo pair because the effective optical axis of the camera was deviated as the glass plate moved – see Figure 2.14.

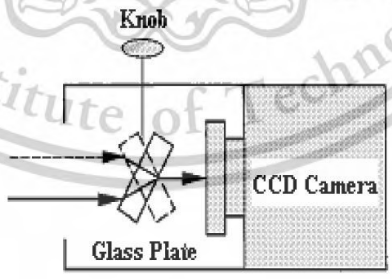


Figure 2.14 Single camera stereo vision using a glass plate with position changed by the control knob - taken from Nishimoto & Shirai (1988).

Gao & Ahuja (2004).improved this model to capture multiple images as the plate rotated: it provided substantial number of image pairs, with larger disparities, in a wider field of view.

2.23 MIRROR BASED ARRANGEMENTS

Several systems added mirrors in front of the camera - see Table 3. Teoh and Zhang (1984) placed two mirrors at fixed angles at the top and bottom and a third mirror rotated freely in the middle. Stereo pairs were obtained when the third mirror was aligned parallel to one of the fixed mirrors - see figure 2.15.

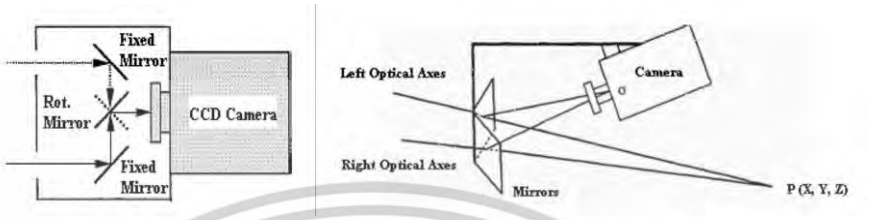


Figure 2.15 Single camera stereo vision systems using multiple mirrors – taken from Teoh & Zhang (1984)

Another simple mirror arrangement put two differently angled mirrors in the camera field of view, so the left and right halves of the captured image were obtained from different angles - see figure 2.15 (left). These systems use planar mirrors and thus do not add additional distortions to the images; however, the additional mirrors require space and precise, rigid alignments and constrain the effective binocular vision region.

2.24 DIFFRACTION GRATINGS

The earliest grating based stereo vision systems used a vertically placed diffraction grating between the lens and the test object, with a perpendicular alignment to the optical axis. Two monochromatic light sources were used to illuminate the scene, so that the stereo pairs were not affected by dispersion, i.e. chromatic aberrations. By analyzing the optical model, see figure 2.16, the distance between the images can be derived by ray tracing.

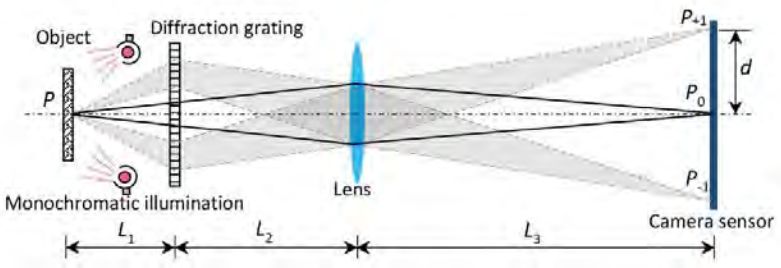


Figure 2.16 Single camera stereo vision system using a diffraction grating – taken from Pan & Wang (2013)

2.25 OPTICAL FILTERS AND BEAM SPLITTERS

A narrow band pass filter selectively transmits a narrow band of the optical spectrum. Pan et al. (2018) used an optical arrangement, consisting of a single camera, a beam splitter placed in front of the lens and two bandpass filters, for different wavelengths, mounted on left and right sides of the beam splitter and three planar mirrors. Figure 2.17 shows their optical arrangement.

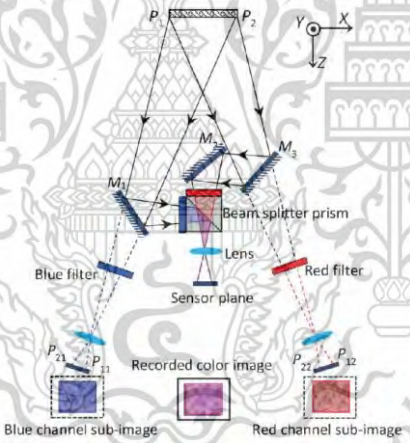


Figure 2.17 Single camera stereo vision using beam splitters and optical filters – taken from Pan et al. (2018).

2.26 SIMPLE PRISM BASED SYSTEM

The system shown in figure 2.18 has a biprism placed on the optical axis of a camera. Rays through the top and bottom half of the prism are deviated in opposite directions, effectively creating two 'virtual' left and right cameras. Although potentially simple - requiring a single, cheap and easily obtained prism - the images are influenced by the prism dispersion, in addition to the lens distortion inherent in almost every camera.

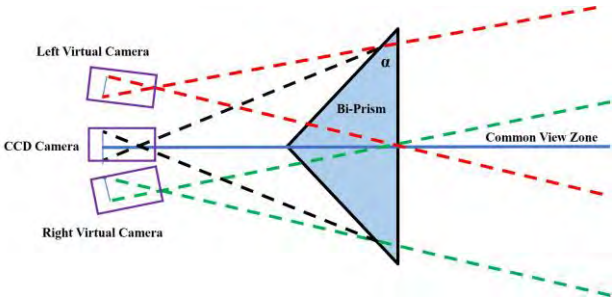


Figure 2.18 Single camera stereo vision system

Various techniques have been used to alleviate the prism dispersion problem.

2.26.1 Telecentric lens

Conventional lenses have angular fields of view. When the distance between lens and the scene increases, magnification decreases. Similarly, human vision perceives depth. This field creates parallax or perspective error, which decreases depth accuracy. Telecentric lenses eliminate the parallax error by having a constant, non-angular FOV; at any distance from the lens: a telecentric lens will always have the same FOV (Batchelor, 2012); (L. Wu et al., 2015).

2.26.2 Monochromatic light

Monochromatic light sources produce and reflect light of a single wavelength. Therefore, the 'color' of the scene remains consistent across the image pair. Since there is no prism dispersion with single wavelength light, this improves the edge detection capabilities of the image by providing sharper, clearer edges in the scene. Monochromatic light sources can also help reduce noise in the images, which can improve the accuracy of depth estimation (Pan et al., 2018).

CHAPTER 3
CALIBRATION

3.1 OVERVIEW

Camera calibration is necessary in computer vision systems to ensure accurate and reliable measurements of the scene. Calibration involves finding the intrinsic and extrinsic parameters of the camera. Intrinsic parameters are specific to the camera, whereas extrinsic parameters describe the position and orientation of the camera with respect to the world coordinate system. The required parameters are focal length, principal point, lens distortion coefficients and camera orientation in space. Systems are typically calibrated by capturing images of a calibration target with known geometry, e.g. a checkerboard or a set of dots arranged in a grid pattern. Since dealing with the specific problems of calibrating a general, i.e. not limited to narrow angle prisms or a monochromatic light, biprism system, in this chapter, I discuss, first, calibration of 'conventional' dual camera stereo systems, followed by previously reported biprism systems. The following section describes the design of biprism systems and a chapter on the challenges found calibrating my system.

Rapidity, robustness and accuracy are the major challenges faced by camera calibration algorithms. 3D depth evaluated by a stereo algorithm critically depends on varying camera positions. Two accurately known camera positions lead to a proper calibration which defines the accuracy of 3D depths. Camera calibrations fall in to linear and non-linear techniques. Linear models use a pinhole camera model to find the relevant parameters by solving a set of linear equations describing image, camera and world coordinate systems. Linear models compute rapidly but accuracy suffers when the real lens distortion is neglected. Non-linear models consider lens imperfections and distortions to improve accuracy (Slama, 1980; Heikkila & Silvén, 1997).

Zhang (2000) placed a planar pattern in front of a camera in more than two different orientations to obtain a robust less complex calibration, followed by a set of linear equations to find the camera parameters. Also, he used a non- linear approach to model the radial distortion of the lens. A planar calibration pattern approach was modified to handle out of focus cameras by not considering the extrinsic parameters

of the camera, reduce reprojection errors and used a target using of imperfect, not measured shapes, to calibrate the camera, eliminate the effects movements of the system that introduces changes in the perpendicular (orthogonal) relationship between the camera optical axes, e.g. due to slight tilts or rotations of the cameras relative to each other, or due to vibrations or misalignments, that affects camera calibration, enhance feature point extraction under varying light conditions etc. (Albarelli et al., 2009; Huang et al., 2013; Jia et al., 2015; Liu et al., 2016).

J. Zhang et al. (2018). also tested a camera calibration method using a 2.5D coding target that requires a single captured image capable of calibrating all camera parameters including a distortion model

The majority of these calibrations have been applied to two camera stereo vision systems. The cost and accuracy of these systems were affected by the physical setup of the camera system.

3.2 TWO CAMERA STEREO CALIBRATION

Since the relative orientation and separation of the two cameras must stay constant, it is usually essential to capture the scene with two (or more) cameras simultaneously. Movement of the object will cause the two views to vary, making matching of individual scene points harder. Varying illumination also causes the intensity of individual scene points to vary and not accurately match. Synchronization ensures that individual ‘shots’ of the scene use the same illumination. Thus, identical image conditions, allowing only the view to vary, are needed to detect the matching features, in both left and right images, in the binocular region, from which depths can be calculated. The ability to match features in both images is the stereo correspondence problem - it includes having a 'binocular field of view' (a field in which matching features in both images may be seen), as well as ensuring, as much as possible, identical conditions to simplify feature matching (Kala, 2016). My system eliminated part of the hardware problem, by using a single camera, so that images are captured synchronously, 'freeze' motion to the limit allowed by the aperture time, use identical lighting, a single, common lens and identical electronic gain settings. However, the biprism introduced an extra distortion, added to the lens distortion, which must be eliminated or a method to neglect the effect of prism added distortion from any

real camera system must be introduced. Much of this thesis discusses a method of avoiding the effect of prism distortion.

3.3 SINGLE STEREO CAMERA CALIBRATION

Several single camera stereo systems have been described to avoid synchronization and complex optical hardware. Several single camera techniques use **multiple mirrors** to generate the required two separate 'views' of the scene. Mouaddib et al. (2005) used a number of convex mirrors to provide the baseline from which two views can be seen. Takahashi et al. (2012) used mirrors to observe three known 3D points of an object in three poses to calibrate the extrinsic parameters. Another single camera system used **polarizers.** Polarization was used to separate light from different angles and capture two views without losing resolution (Y. Li et al., 2018). However, polarization led to a significant loss of light, creating a problem in calibration and subsequent image matching. These single stereo vision systems required complex installations with multiple mirrors or additional polarizers and special mirrors or filters.

3.4 PRISM STEREO CALIBRATION

Introducing a biprism to single camera system would solve the hardware and software complexity. D. Lee & Kweon (2000) used a bi-prism, with a single lens that captured a pair of stereo images, due to differing refraction through two sections of the prism - using an optical arrangement similar to that shown on the bottom of Figure 3.1. They introduced a simple mathematical model which provided the disparity between left and right half of the image when the prism angle was small This concept was extended by Lim et al. (2013b), who used a geometrical approach to find extrinsic parameters of the virtual cameras created using a biprism placed in front of the camera Genovese et al. (2013) noted that distortion induced by the biprism could not be handled by existing lens distortion correction functions. Hence, they chose a model-free approach by calculating a 'piece wise' function that described the image deformation over the measurement area. Pan et al. (2012) placed a high quality telecentric lens in line with a biprism to maintain the system optical axis perpendicular to the upper surface of the prism to ensure sharp captured images. Distance between the lens and prism could be manually adjusted to generate two virtual images on the

camera sensor. For a successful configuration, the apex of the prism should divide the FOV in half thus creating two virtual cameras. Further, Pan et al., (2018). used a monochromatic (blue) light source to avoid the prism dispersion problems. Attaching a bandpass filter to the lens helped eliminate effects of chromatic aberration caused by the refraction through the different media (Pan et al., 2012; Pan et al., 2018). Since these methods needed special setups, e.g. a custom fabricated biprism, monochromatic light or bandpass filters and a telecentric lens, I explored the possibility of achieving similar results, using readily available prisms and a simple camera setup. Using similarly readily available and inexpensive lenses and cameras, and thus some not precisely known configuration parameters, e.g. position of the prism relative to the system optical center, I was able to introduce a simple prism distortion correction method.

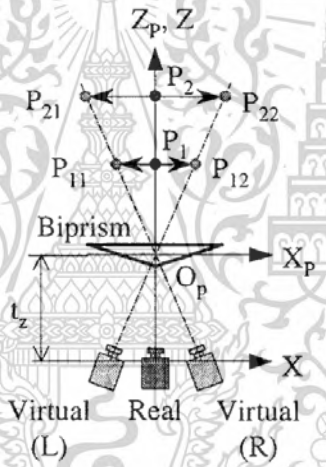


Figure 3.1 Equivalent stereo camera system to a biprism-stereo system taken from D. Lee & Kweon (2000)

3.5 PRISM STEREO CAMERA CALIBRATION PREVIOUS WORK

In prism-based single-camera stereo-DIC systems, the imaging lens and bi-prism are crucial components, typically assumed to be ideal in the imaging model. However, in reality, both the imaging lens and biprism have imperfections due to manufacturing errors. Additionally, any misalignment between the camera and biprism can further affect the accuracy of the imaging model. These deficiencies and misalignments deviate from the ideal imaging model, inevitably introducing errors into the measurement results. To minimize these errors and achieve high-precision 3D shape and deformation measurements, careful calibration of the bi-prism-based single-

camera stereo-DIC system is essential. Table 3.1 summarizes the previous calibration techniques of biprism stereo vision camera calibration.

Table 3.1 Previous Work

Author	Notes	Reference
Genovese et al	Model free camera calibration approach	(Genovese et al., 2013)
L.Wu et al	Calibration-free single-lens 3D-DIC system based on a bilateral telecentric lens and a bi-prism.	(L. Wu et al., 2015)
L. F. Wu et al	Refractions within the bi-prism were analyzed in 3D space, and errors introduced by the imperfect imaging model were minimized by accounting for the misalignment between the bi-prism and the camera, as well as lens distortion.	(L. F. Wu et al., 2016)
X. Li et al	one-step calibration method is employed to obtain the intrinsic parameters of biprism bi-telecentric single camera	(X. Li et al., 2024)

3.6 GEOMETRIC SETUP

Several systems were used in the preliminary work in this study. These systems used cameras in a commercial underwater robot, modified with a 3D printed camera mount to which a prism was mounted as a ‘window’ between the water and the internal hardware. Some of the images captured in them appear in the images reported here. However, the final calibration experiments used the system described in Table 3.2.

Table 3.2 System Specification

Parameter	Value	Note
Camera	Raspberry Pi camera V2	
Sensor Resolution	3280 × 2464 pixels	8 Megapixels
Pixel Size	1.12μ × 1.12μ	Sony IMX219
FOV	62.2°	Horizontal
Focus	1m to infinity	Fixed focus
Light Source	RGB	
Prism		
Material	N-BK7	Right angle prism
Refractive index	1.515	
Angle	450	
Length of legs	15mm	
Surface flatness	$\lambda/8$	

A mount for the prism, in front of the camera lens, to which the camera PCB could be bolted, was 3D printed. See it in Figure 3.2 in front of the camera lens. The base of the prism was placed as close as possible to the camera lens. The prism apex was roughly aligned with the optical axis of the camera, so that the biprism divided the camera sensor in half. An important aspect of this system was that the alignment of the prism with camera optical axis did not need to be perfect – see later images of the captured in Figure 3.8. Note that the prism exchanges the sides of the captured scene image:

A multicolor strip of LEDs, with red, green, and blue LEDs, was mounted on a large board. The prism and camera were inserted into an aperture in this board - see Figure 3.2.1. No alignment in this system was particularly critical, although as noted in the conclusion, the overall calibration curves would be smoother if the prism was better aligned with the camera.

The target was imaged with each color in turn.

A diagram of the setup is in Figure 3.3. The left half shows the transit of a ray through the prism, labelling angles relative to the edges of the prism. The right half shows - viewed from the top and from left to right - the camera image plane, the lens,

the prism and three points in the scene. X_q is a scene point, imaged onto two points - following two paths through the prism, P_l and P_r, on the image plane. A point, **XL**, in the lower half of the scene, is imaged to P_l, which is on the left side of the image points, i.e. the images are reversed. Thus, we have two virtual cameras, viewing a 'binocular region', similar to the conventional stereo vision, required for 3D depth calculation.

Figure 3.2 Prism-Camera mount

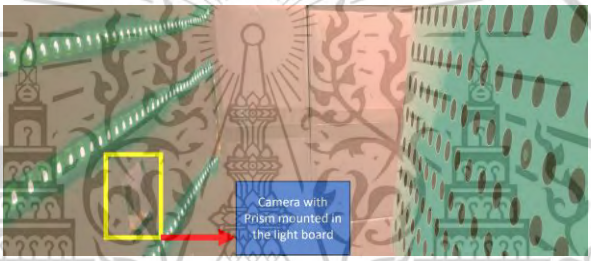


Figure 3.3 Light board with stereo rig inserted in center

3.7 DESIGN CALCULATIONS

To calculate the depth of a scene object, we must have two views of it from different viewpoints (or virtual viewpoints in a biprism system). Thus, a key step in designing the system is calculating the binocular region, in which scene points are seen by both cameras (or virtual cameras in our system, since the prism makes it appear as if there are two separate cameras). We back projected rays from the image plane to the scene point: Figure 3.5 shows the biprism stereo vision system and an associated co-ordinate system. The points, **(X, Y, Z)**, represent a scene point, **Xq**, whereas **(u, v)** represents image coordinates. The prism angle was **α**, **d** is the distance from the optical center, **C** to the biprism and **f** is the focal length and **δ** is the deviation of a ray as it passed through the prism

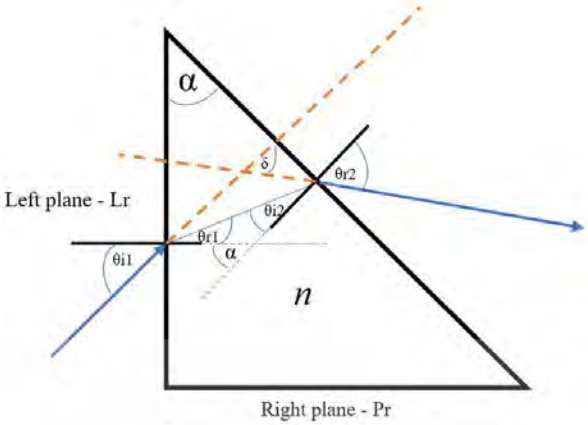


Figure 3.4 Model for incident ray from the left, transiting the prism and exit refracted ray on the right

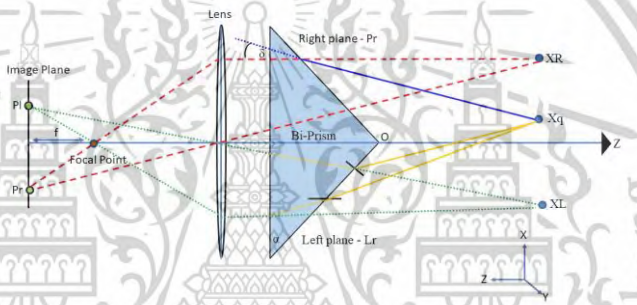


Figure 3.5 Model of bi-prism vision system.

Back projection simplified the calculations (Lim & Qian, 2016). From the left of Figure 3.5, we obtain,

$$\alpha = \theta_{r1} + \theta_{i2} \tag{1}$$

where Θ_{ix} and Θ_{rx} are the incident and refracted angles at surface, $x \in \{1, 2\}$. The total deviation of a back projection ray from the image plane, with incident angle, i_1, on the 'front' (base) surface of the prism, is

$$\delta = \theta_{i1} + \theta_{r2} - \alpha \tag{2}$$

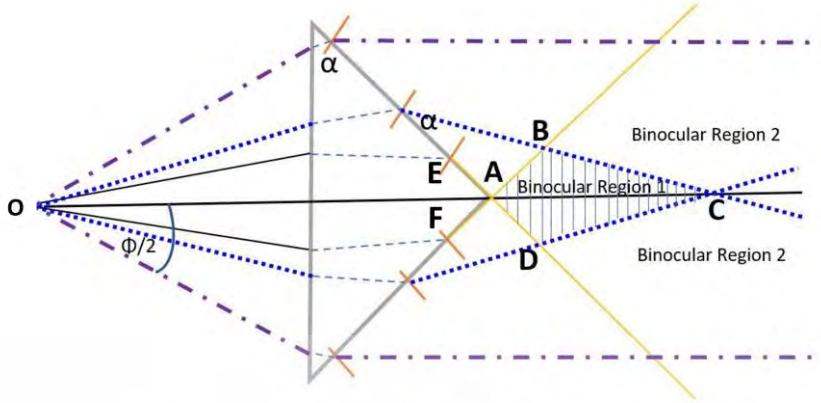


Figure 3.6 Definition of the binocular region

To determine the binocular region, we first consider the rays parallel to the optical axis. Note that if the prism angle α is larger than the critical angle,

$$\theta_c = \sin^{-1}(n1/n2),$$

there will be a small angular field, close to the optical axis, in which rays are totally internally reflected. For a 45° BK-7 prism,

$$\theta_c = \sin^{-1}(1/1.5) = 41.8°,$$

from BK-7 to air, so that a narrow angular field of 3.2°, between OE and OF, does not appear in images. The largest emitted angle comes from rays emitted at grazing incidence to the prism edge, i.e. at α = 45° from the optical axis, rays AB and AD, marked in yellow in Figure 5. The other edges of the binocular region are determined by the field of view of the camera: examples are the rays BC and DC. Note that there is a point, C, at which the disparity is zero, as usual for a verging axis stereo camera system. The resulting binocular region is labelled in Figure 3.6.

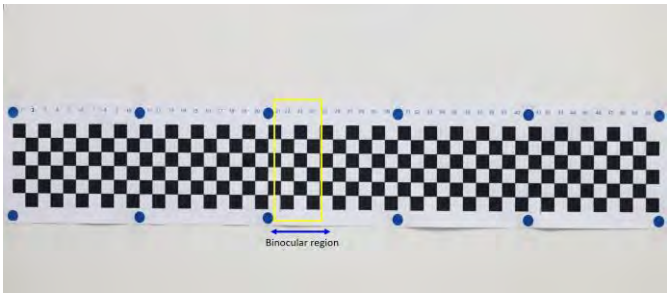


Figure 3.7 Wide target viewed without the prism in optical path

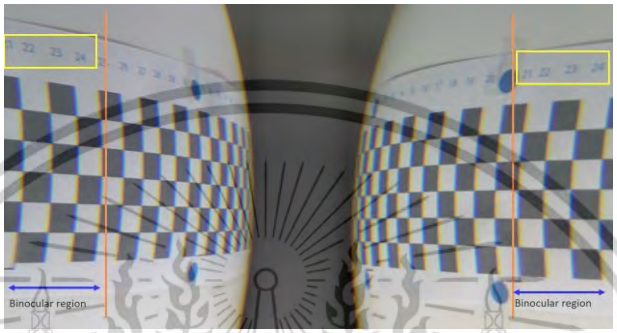


Figure 3.8 Wide target used to illustrate scene setup and prism distortion.

Figure 3.7 showing a wide target seen with the prism in the path: Note that the binocular region marked here is a relatively narrow region of the field of view. The yellow boxes outline regions which appear in both images.

After passing through the prism, rays are deviated by **δ**, see Equation 2. So, the other edge of the binocular region is determined by the extreme rays in the camera field of view. Two examples are shown in Figure 3.6, for a narrow lens (inner pair of dotted blue rays) and a wider lens (outer pair of dot-dash purple rays). After transiting the prism, rays from the wider lens extend to ∞ and define the binocular region, labeled 'Region 2'. Rays from the narrow lens intersect at C and delimit the striped binocular region. Thus, designers can choose a lens that leads to a needed binocular region, i.e. areas of interest in which depths can be measured.

A further example appears in the test pattern in Figure 3.8, where the binocular regions, shown by numbered columns, which appear in both left and right images, are marked in red in the lower image.

3.8 DISPERSION

Introducing a prism into the optical path of a camera ensured that color images 'spread' due to the variation of the refractive index, and rays of different colors

diverged as they passed through the prism. Thus, if I wish to image colored scenes, I must allow for these divergences. Several previous studies (Pan et al., 2012); (Yu & Pan, 2016); (L. Wu et al., 2016) avoided this problem by using monochromatic light sources or filters. However, to compute depths in arbitrary scenes, where we cannot control the imaging wavelength or objects in the scene that might not have, conveniently, narrow bands of colors, that assist image matching, allowances must be made for these divergences.

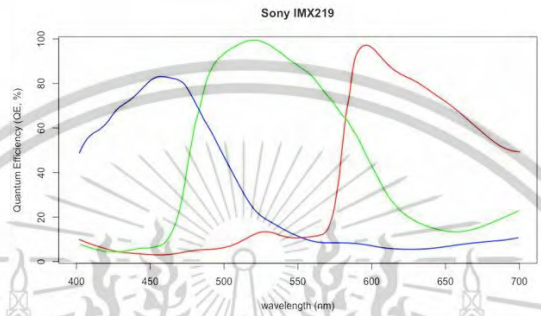


Figure 3.9 Spectral response of the Sony IMX219 chip

3.9 CALIBRATION OF A GENERAL BI-PRISM SYSTEM

In this section, I describe the calibration model and the challenges in calibrating the bi-prism system. It describes some design trials, not all successful, but reported here for the benefit of following researchers, and the final scheme used to calibrate the system. It discusses the targets, their shapes and colors.

3.10 CAMERA CALIBRATION PROCEDURE

The basic calibration steps follow traditional patterns using known targets, e.g. checkerboards or patterns of circles, that cover the useful region in the scene, so that optical aberrations could be removed, and simple projective geometry based on pinhole camera models used for depth estimation, e.g. Lee and Kweon(D. Lee & Kweon, 2000). However, prism generated aberrations need to be removed first. These aberrations can be seen in Figure 3.10, where it was notable that, even though the target was held perpendicular to the optical axis, so that all squares would appear with similar widths, as seen in Figure 3.7, with the prism removed.

In the lower image, although the squares had similar widths over a small region, e.g. the binocular region marked, significant variations remained. However, this variation

was close enough to linear to avoid complex inverse tan corrections, and a need to know the actual angle at which a square corner appeared. However, this approach had limitations due to the edge distortion induced by lens and chromatic aberration. Therefore, the OpenCV corner detector was not able to detect the distorted edges.

3.10.1 Checkerboard target

The traditional black and white checkerboard used by Z. Zhang (2000) - see the Figure 3.7 was rejected for several reasons:

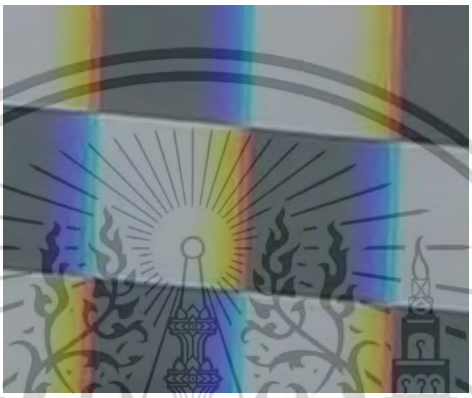


Figure 3.10 Chromatic aberration induced by prism

- Dispersion also caused significant chromatic aberration.
- Consequently, the 'corners' were not precise - see later Figures 3.12 and 3.13, the OpenCV Harris corner detector often failed to find them.
- With a white light source, the vertical edges were dispersed over many pixels, although the horizontal edges (perpendicular to the prism apex) were relatively unaffected - see Figure 3.10.

3.10.2 Circle target approach

Circular patterns have also been used by many - who believed that detecting circle centers was easier, for example, Hui et al. (2009) used a solid circle pattern with different diameters to calibrate their stereo vision system. Simple trials with colored circles, matching the RGB channels of the camera, on a white background in white light - see Figure 3.11, so that the R, G and B channels of the camera would select monochromatic images, but fuzzy images resulted from the white background - see Figures 3.8 and 3.11. Fundamentally, all trials with various color combinations and shapes of targets were defeated when using white light imaging sources with the prism

in place. Selecting color bands using the inherent filters of the camera or choosing targets with narrow bands of color failed to produce desired sharp images needed for accurate calibration: either the target edges were blurred or, equivalently, the colored background edges were blurred. The use of multiple bi-colored targets was rejected as being impractical due to the number of targets needed and the difficulty of finding suitable colors with minimal spectral overlaps in the camera spectral responses.

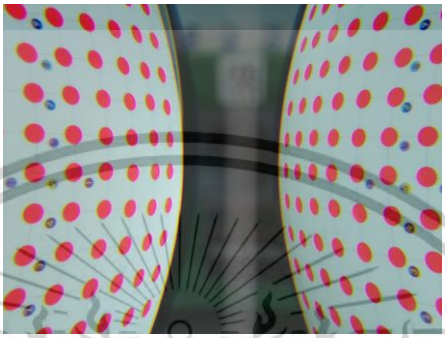


Figure 3.11 Red circles under white light

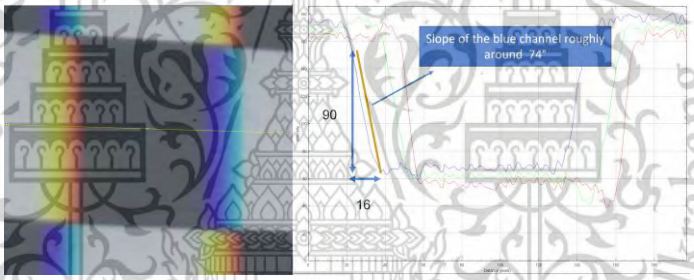


Figure 3.12 Square target vertical edge distortion

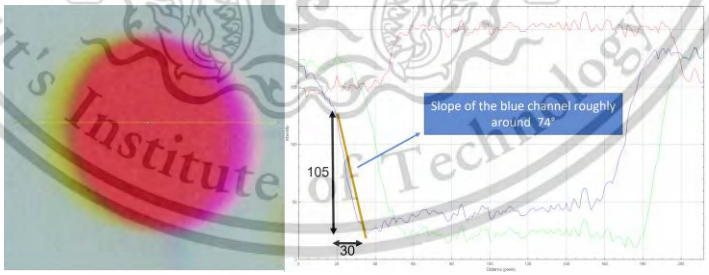


Figure 3.13 Circle target contour distortion

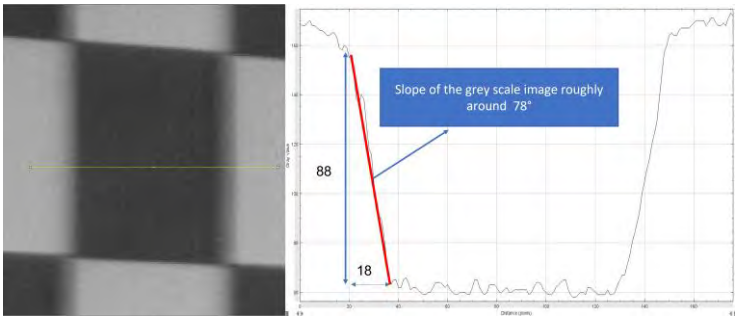


Figure 3.14 Visible edge dispersion of the target image in grayscale

3.10.3 Monochromatic illumination

The ready availability of strips of LED lights of various colors prompted a trial using monochromatic illumination. I tested a black and white circle target under monochromatic light - see Figure 3.15. Black circles on a white background were selected, over red circles on a white background since red circles did not appear in the camera red channel. The target was printed onto a 2×1 m white plastic background. One dot was chosen as the 'reference' and marked with a white circle in its center - see Figure 3.15. While not critical, it made it easy to recognize corresponding points in the left and right images when the prism was inserted into the optical path.

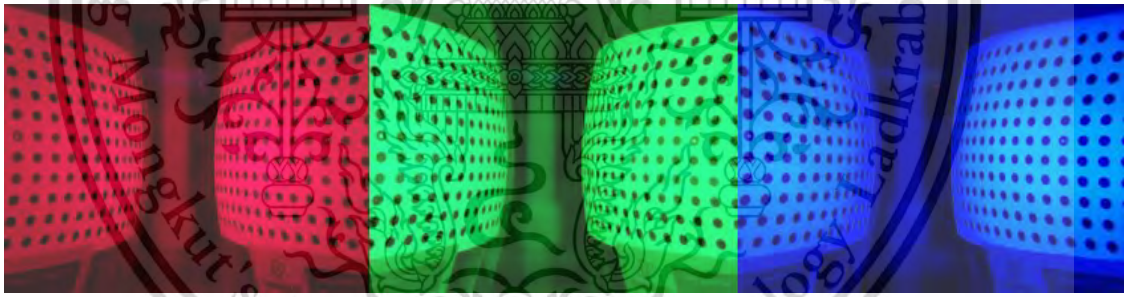


Figure 3.15 Circle features under red, green, blue illumination

Circular targets under monochromatic illumination showed clear circle contour shapes considerably similar to the calibration target under white light without the prism - see Figures 3.16 and 3.17.

3.11 CALIBRATION STEPS

Several images were captured for each illumination (R, G and B) at each target distance (z = 600, 700,....,1200 mm).

Processing steps:

1. Images at depths from z = 600 mm to z = 1200 mm were individually processed.
2. Regions of interest (ROIs) were selected.
3. RGB images were separated into color bands: only the color band corresponding to the illumination was used.
4. Images were inverted and thresholded to form binary images.
5. Centers of the images were estimated manually and with an OpenCV program.
6. Images were separated into left and right regions relative to the center.
7. Targets were identified in the left and right half images.
8. Targets were linked into a rectangular grid- horizontally and vertically - see Figure 3.19.
9. Grid points were assigned indices to label corresponding targets in left and right halves - see Figure 3.18.
10. Two calibration look-up tables were generated with sparse set of points for the points and
11. Matching disparities, $\mathcal{C}_p$(u, v, du), and sets of distances $\mathcal{C}_z$(z).

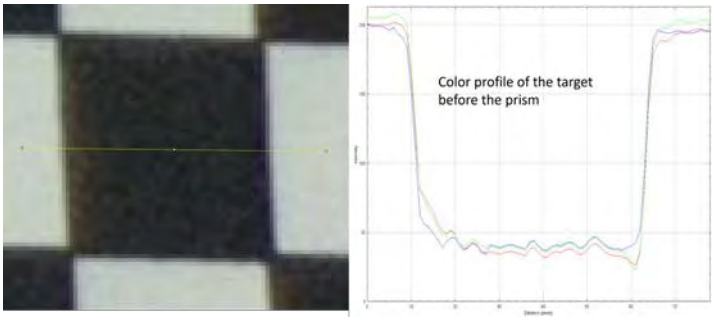


Figure 3.16 Target image square - single channel taken from RGB image without the prism

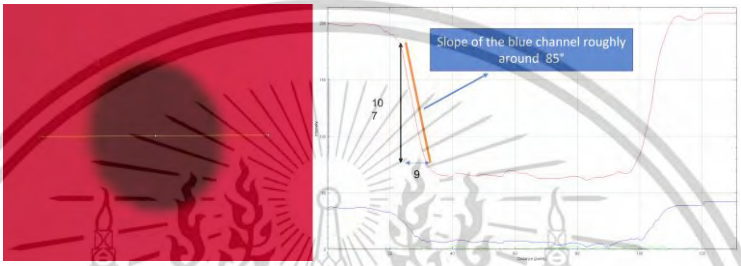


Figure 3.17 Circle target under red light with prism in the optical path

3.12 CREATION OF DEPTH MAPS

Using the two look-up tables, Cp, containing (u, v, d) points and the corresponding look-up table, Cz of distances, z, to create a depth map for a new, unknown image, $\mathcal{I}$, the steps were:

1. Using the two look-up tables, Cp, containing u, v, d points and the corresponding look-up table, Cz of distances, z.
2. Identify the left and right images, $\mathcal{I} \longrightarrow (I_L, I_R)$.
3. Form a depth map large enough to encompass the binocular region in the left image, $\mathcal{I}_L$,
4. For each recognizable point in the left image at u_L, v_L, using any of the several available matching strategies, e.g. SDPS [21], find the corresponding point (u_R,v_R) in the right image.
5. Calculate the disparity, d from the $(u_L,v_L) \longrightarrow (u_R,v_R)$ vector.
6. Locate the eight points surrounding the point (u_L,v_L,d) in the point look-up table, $\mathcal{C}_p$.
7. Using the offsets from the points in $\mathcal{C}_p$, computed by linear interpolation, compute the required z, using the same offset points in the $\mathcal{C}_z$ table.

8. These z's form the (sparse) depth map.

The sparsity of the depth map will depend on the original image, the number of recognizable matching points in it and the matching algorithm, i.e. how dense are the matching points. As noted in the further improvement section, the accuracy of the depth map may be increased by using more circles in the calibration targets.

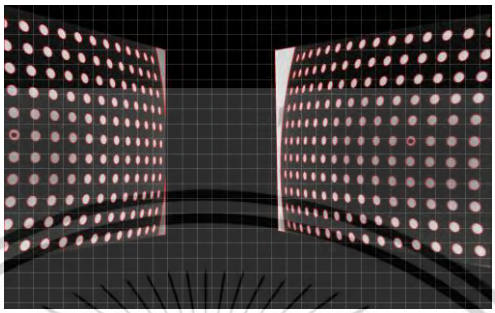


Figure 3.18 Contours detected after inversion, thresholding and binarization; Note: (1) Large contours resulting from the target board edges were rejected by size, (2) Small targets inside centers of a large target were used to mark corresponding left and right targets - detected as 'inside' larger targets and rejected in the next step

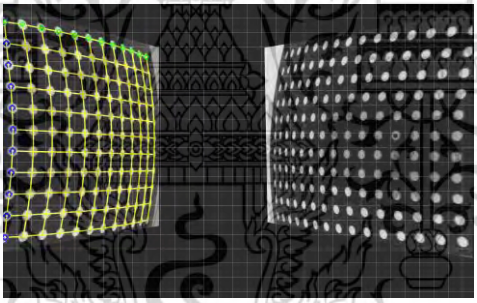


Figure 3.19 Left Image: Individual target circles linked to rectangular grid

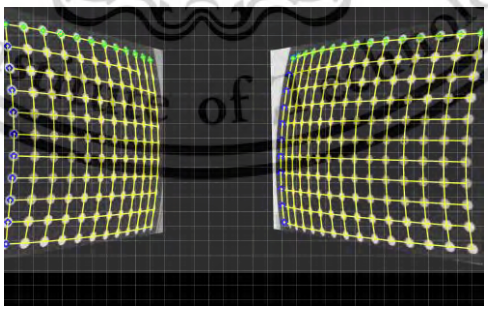


Figure 3.20 Left and Right: Individual target circles linked to rectangular grid to match the corresponding points

3.14 PSEUDOCODE

A more formal discussion of the several steps in the calibration follows as pseudocode.

Calibration

for each z ∈ (600, 700, . . . , 1200) do

for each band ∈ (red, green, blue) do

findCorrespondences() (Alg 2)

end

end

Algorithm 1: Overall process

findCorrespondences()

Data: image list (I_j ∈ 0..n−1)

Result: Calibration map

for each I_j in image_list **do**

A ⟵ $I^c j$ Select c ∈ R, G, B channel from RGB image

B ⟵ A Working image - Choose region interest: mask out background

C ⟵ B Invert image

D ⟵ C Threshold image

locateCircleCentres()

irj ⟵ Rotate ij using corresponding points.

end

Algorithm 2: Calibration Steps

CHAPTER 4

RESULTS

I used a stereo vision method, that was able to find the depth of a scene point, z, using two look-up tables. The observable range of depths varied from 600- 1200 mm due to the limitation of the camera and the prism parameters - see Table 3.2. The first look-up table was produced using the image pixel space $\mathcal{C}_p$, it is sparse and has elements (u, v, du). The second look-up table has the known distances, z, in $\mathcal{C}_z$, matching elements in Cp. Examples of curves showing elements in $\mathcal{C}_p$ appear in Figure 19, which show that smooth curves for disparity versus distances were extracted from the images for known individual distances. To improve the accuracy of the detection of (u, v) target points, we were able to make three calibration curves - one for each RGB color band, so that Figure 4.1 shows a set of three curves.

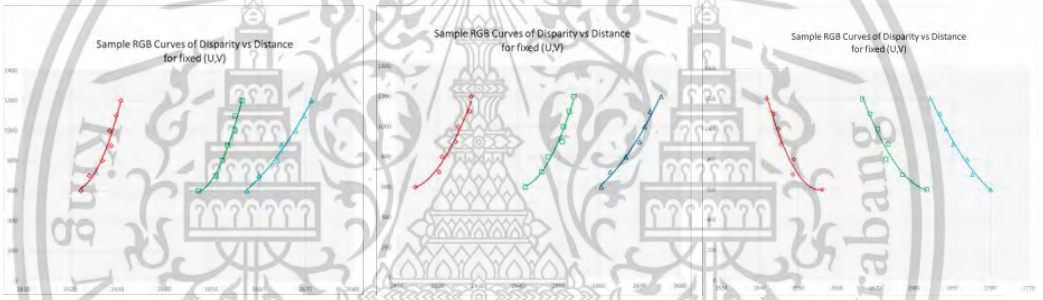


Figure 4.1 Sample disparity vs distance curves for fixed reference points in a row of targets. From left to right, the chosen targets were in the center, below the center and above it.

These curves showed that the calibration points derived from individual targets, set at distances from 600 to 1200 mm, lay on smooth curves. To form the full calibration look-up tables, the measured disparities from all actual target points (~230 at different (u, v) image points in the binocular region) were combined to form sparse look-up tables. Although searching for the eight points neighboring any desired point to generate new distance maps may be formally computationally expensive, the small total size of the lookup tables makes using a more sophisticated and complex algorithm unjustified.

CHAPTER 5

CONCLUSION

A simple calibration technique for the flexible prism stereo system was described. This system used readily available and inexpensive components - a small camera and a commonly available 45° prism. A simply constructed circle target was used to generate calibration tables. The calibration system collected a set of images at known distances in the scene. Processing the images identified the targets using an OpenCV contour detector and then located the centers. Corresponding points were matched, and the sparse look-up tables built.

5.1 IMPROVING ACCURACY

The partial curves in Figure 4.1 show that the data from the three RGB channels show different disparity values, but similar patterns. This enabled accuracy improvement by averaging from the three monochrome sources to generate final depth maps. If unknown targets are imaged in three different bands, reflections from the monochrome sources will be, in the absence of fluorescence or phosphorescence contributions, themselves monochrome and not severely affected by the prism deviations: typical readily available monochrome LEDs are quite narrow sources (usually 20-50 nm optical bandwidths). Thus, depth maps derived from each band can be reliably averaged to increase depth accuracy. This potential will be explored in a further paper.

5.2 FURTHER IMPROVEMENTS

After examining the results from these trials, it was found that improvements might be made by:

1. using a larger number of smaller targets to increase the density of points in the calibration look-up tables: targets only need to be visible in the cameras and could be ~15 mm in diameter or even smaller, for high resolution cameras, than the 25 mm targets used here and
2. labelling more targets with, e.g. squares or other simple shapes, to speed up recognition of matching targets in the calibration stage.

3. Introducing equilateral prism with two reflective mirrors, creates virtual viewpoints with accurate angular offset (Somanath et al., 2010). Thus, enabling stereo matching without misalignments. Figure 5.1 shows a ray diagram of the formation of the image pairs, where OC is optical center, scene point Q, Q_L and Q_R are virtual points respectively.

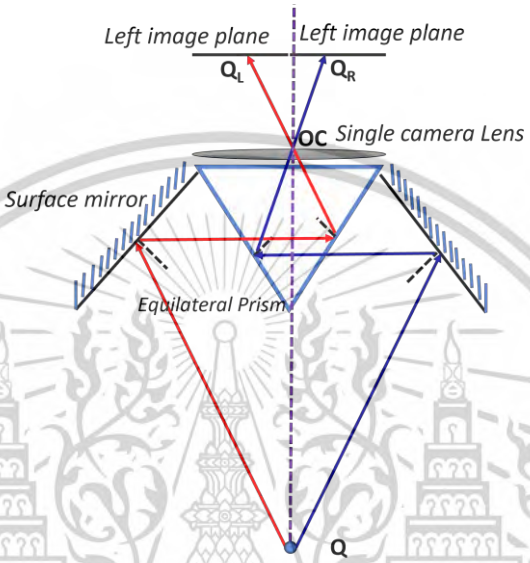


Figure 5.1 Ray diagram of the mirrors and prism arrangement

4. Using a customized low angle prism,
   - provides smaller angular separation, creating consistent disparity between virtual points. Thus, it helps to secure accurate depths at close or moderate distance,
   - reduces the complexity of correcting distortions as low angle prisms introduce, comparatively less image distortion compared to high angle prisms,
   - maintain significant overlap of the field of view for dense stereo matching,
   - minimize the error in epipolar geometry.
   - Therefore, formal models can be used to solve the calibration problem.

Figure 5.2 shows the Binocular vision (BV) region of the single-lens prism-based stereovision system with the increasing value of prism angle θ. There is no observable BV region when θ is equal to 0 (Figure 5.2 (a)); when θ is slightly larger, BV region of the two half image planes is increasing. However, BV region in Figure 5.2(b) is too small to be useful for stereovision. Thus, the BV region of the system should fall in the range between Figure 5.2(c) and Figure 5.2(f). Note that, when prism angle θ is roughly larger than 20° maximize the chromatic aberration effect, making the depth measurement system ineffective (X. Cui et al., 2015).

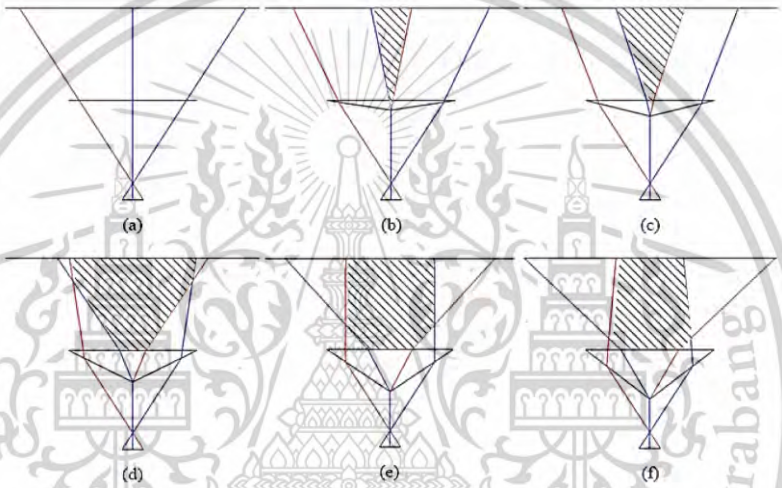


Figure 5.2 FOV of the system with the increasing value of Prism angle

A completely different single camera stereovision to use quadruple mirrors. Basic concept is to use mirrors to split the image plane simulating a stereovision setup Figure 5.3 shows a quadruple mirror setup. However, there are several considerations as follows.

- Precision – precise placement and orientation is required to avoid distortion
- Optical quality – High quality mirrors are required to maintain the image clarity and avoid image distortion.
- Image overlap – Ensure the reflected views cover the target scene

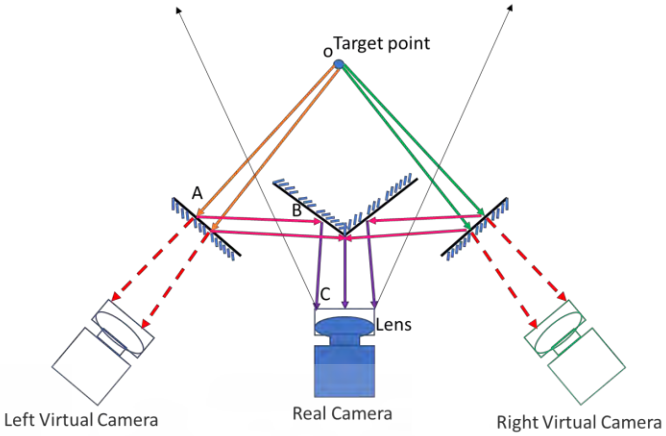


Figure 5.3 Alternative quadruple mirror system

5.3 CONTRIBUTIONS

This work contributed to the following:

1. a stereo system using a single camera built from inexpensive, readily available right-angle prisms,
2. custom (and therefore expensive) low angle prisms, used in previous work, were not required,
3. no mirrors, filters or expensive optics, e.g. telecentric lenses, were required,
4. simple optical arrangements, without precise alignments, were effective,
5. readily available multicolor LED sources provided the monochromes sources and
6. the potential to improve accuracy by averaging results from the individual measurements using different optical bands requiring, at most, three image frames, as the bands were switched.

Further, the potential to use the simple optical requirements in underwater systems is also noted: the prism can be the 'window' of a sealed underwater housing, simplifying the optical paths and the numbers of differing refractive index paths.

CHAPTER 6

BIBLIOGRAPHY

Albarelli, A., Rodolà, E., & Torsello, A. (2009). Robust camera calibration using inaccurate targets. *IEEE Trans. Pattern Anal. Mach. Intell.*, *31*, 376–383.

Andrew, A. M. (2001). Multiple view geometry in computer vision. *Kybernetes*.

Barnard, S. T., & Fischler, M. A. (1982). Computational stereo. *ACM Computing Surveys (CSUR)*, *14*(4), 553–572.

Barone, S., Neri, P., Paoli, A., & Razionale, A. V. (2018). 3D vibration measurements by a virtual-stereo-camera system based on a single low frame rate camera. *Procedia Structural Integrity*, *12*, 122–129.

Batchelor, B. G. (2012). *Machine vision handbook*. Springer.

Bell, T., Li, B., & Zhang, S. (1999). Structured light techniques and applications. *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1–24.

Brewster, D. (1856). *The Stereoscope; Its History, Theory, and Construction: With Its Application to the Fine and Useful Arts and to Education*. John Murray.

Chen, C.-Y., Yang, T.-T., & Sun, W.-S. (2008). Optics system design applying a micro-prism array of a single lens stereo image pair. *Optics Express*, *16*(20), 15495–15505.

Chen, J. M., Liu, X., Zhao, Z. X., & Zhong, W. T. (2013). Research on landscape environment with 3D-reconstruction and volume measurement of fruit tree canopy based on Kinect. *Advanced Materials Research*, *788*, 480–485.

Cui, P., & Yue, F. (2007). Stereo vision-based autonomous navigation for lunar rovers. *Aircraft Engineering and Aerospace Technology*.

Cui, X., Lim, K. B., Guo, Q., & Wang, D. (2012). Accurate geometrical optics model for single-lens stereovision system using a prism. *JOSA A*, *29*(9), 1828–1837.

Cui, X., Zhao, Y., Lim, K., & Wu, T. (2015). Perspective projection model for prism-based stereovision. *Optics Express*, *23*(21), 27542–27557.

Delaunoy, A., Prados, E., & Belhumeur, P. N. (2010). Towards full 3D Helmholtz stereovision algorithms. *Asian Conference on Computer Vision*, 39–52.

Dhond, U. R., & Aggarwal, J. K. (1989). Structure from stereo-a review. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(6), 1489–1510.

Duvieubourg, L., Ambellouis, S., & Cabestaing, F. (2006). SINGLE-CAMERA STEREOVISION SETUP. In *Computer Vision and Graphics* (pp. 173–178). Springer.

Falck, F., Doshi, S., Tormento, M., Nersisyan, G., Smuts, N., Lingi, J., Rants, K., Saputra, R. P., Wang, K., & Kormushev, P. (2020). Robot DE NIRO: a human-centered, autonomous, mobile research platform for cognitively-enhanced manipulation. *Frontiers in Robotics and AI*, *7*, 66.

Faugeras, O. (1993). *Three-dimensional computer vision: A geometric viewpoint*. MIT press.

Feng, X., & Pan, D. (2015). Research on the application of single camera stereo vision sensor in three-dimensional point measurement. *Journal of Modern Optics*, *62*(15), 1204–1210.

Forbes, A. (2019). Structured light from lasers. *Laser & Photonics Reviews*, *13*(11), 1900140.

Gao, C., & Ahuja, N. (2004). Single camera stereo using planar parallel plate. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, *4*, 108–111.

Genovese, K., Casaletto, L., Rayas, J. A., Flores, V. H., & Martinez, A. (2013). Stereo-digital image correlation (DIC) measurements with a single camera using a biprism. *Optics and Lasers in Engineering*, *51*(3), 278–285.

Gimel'farb, G. L. (2002). Probabilistic regularisation and symmetry in binocular dynamic programming stereo. *Pattern Recognition Lett*, *23*(4), 431–442.

Gimel'farb, G., Liu, J., Morris, J., & Delmas, P. (2006). Concurrent Stereo under Photometric Image Distortions. *International Conference on Pattern Recognition*, I: 111-114. http://doi.ieeecomputersociety.org/10.1109/ICPR.2006.401

Gluckman, J., & Nayar, S. (1999). Planar catadioptric stereo: Geometry and calibration. *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, *1*, 22-28 Vol. 1.

Gomes, L., Bellon, O. R. P., & Silva, L. (2014). 3D reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters*, *50*, 3–14.

Gong, M., Yang, R., Wang, L., & Gong, M. (2007). A performance study on different cost aggregation approaches used in real-time stereo matching. *International Journal of Computer Vision*, *75*, 283–296.

Goshtasby, A., & Gruver, W. A. (1993). Design of a single-lens stereo camera system. *Pattern Recognition*, *26*(6), 923–937. https://doi.org/10.1016/0031-3203(93)90058-5

Gudis, E., van der Wal, G., Kuthirummal, S., & Chai, S. (2012). Multi-resolution real-time dense stereo vision processing in FPGA. *2012 IEEE 20th International Symposium on Field-Programmable Custom Computing Machines*, 29–32.

Hamid, M. S., Abd Manap, N., Hamzah, R. A., & Kadmin, A. F. (2020). Stereo Matching Algorithm based on Deep Learning: A Survey. *Journal of King Saud University-Computer and Information Sciences*.

Hamzah, R. A., & Ibrahim, H. (2016). Literature survey on stereo vision disparity map algorithms. *Journal of Sensors*, *2016*.

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

Henao, R., Medina, F., Rabal, H. J., & Trivi, M. (1993). Three-dimensional speckle measurements with a diffraction grating. *Applied Optics*, *32*(5), 726–729.

Holzmann, C., & Hochgatterer, M. (2012). Measuring distance with mobile phones using single-camera stereo vision. *2012 32nd International Conference on Distributed Computing Systems Workshops*, 88–93.

Huang, L., Zhang, Q., & Asundi, A. (2013). Flexible camera calibration using not-measured imperfect target. *Applied Optics*, *52*(25), 6278–6286.

Hui, Z., Liyan, Z., Hongtao, W., & Jianfu, C. (2009). Surface measurement based on instantaneous random illumination. *Chinese Journal of Aeronautics*, *22*(3), 316–324.

Inaba, M., Hara, T., & Inoue, H. (1993). A stereo viewer based on a single camera with view-control mechanisms. *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '93)*, *3*, 1857–1865 vol.3.

Jang, G., Kim, S., & Kweon, I. (2005). Single camera catadioptric stereo system. *Proc. of Workshop on Omnidirectional Vision, Camera Networks and Nonclassical Cameras (OMNIVIS2005)*.

Jia, Z., Yang, J., Liu, W., Wang, F., Liu, Y., Wang, L., Fan, C., & Zhao, K. (2015). Improved camera calibration method based on perpendicularity compensation for binocular stereo vision measurement system. *Optics Express*, *23*(12), 15205–15223.

Kaczmarek, A. L. (2017). Stereo vision with Equal Baseline Multiple Camera Set (EBMCS) for obtaining depth maps of plants. *Computers and Electronics in Agriculture*, *135*, 23–37.

Kalarot, R., & Morris, J. (Feb 4-6). Implementation of Symmetric Dynamic Programming Stereo Matching Algorithm Using CUDA. *16th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, 141–146.

Kee, W. L., Bai, Y., & Lim, K. B. (2015). Parameter error analysis of single-lens prism-based stereovision system. *JOSA A*, *32*(3), 367–373.

Kee, W. L., Lim, K. B., & Wang, D.-L. (2012). Virtual epipolar line construction of single-lens bi-prism stereovision system. *J. Electron. Sci. Technol.*, *10*(2), 97–101.

Khan, T., Morris, J., Javed, K., & Gimelfarb, G. (2009). Salmon: Precise 3D Contours in Real Time. *Proc 8^th IEEE Intl Conf on Dependable, Autonomic and Secure Computing*, 424–429. http://dx.doi.org/10.1109/DASC.2009.113

Kim, H., Lin, C.-S., Song, J., & Chae, H. (2005). Distance measurement using a single camera with a rotating mirror. *International Journal of Control, Automation, and Systems*, *3*(4), 542–551.

Kumari, D., & Kaur, K. (2016). A survey on stereo matching techniques for 3D vision in image processing. *Int. J. Eng. Manuf*, *4*, 40–49.

Lazaros, N., Sirakoulis, G. C., & Gasteratos, A. (2008). Review of stereo vision algorithms: From software to hardware. *International Journal of Optomechatronics*, *2*(4), 435–462.

Lee, D., & Kweon, I. (2000). A novel stereo camera system by a biprism. *IEEE Transactions on Robotics and Automation*, *16*(5), 528–541.

Lee, S. H., & Sharma, S. (2011). Real-time disparity estimation algorithm for stereo camera systems. *IEEE Transactions on Consumer Electronics*, *57*(3), 1018–1026. IEEE Transactions on Consumer Electronics. https://doi.org/10.1109/TCE.2011.6018850

Lee, Y., Kim, T., Choi, J., Jung, J., & Choi, H.-T. (2016). Preliminary Result on Underwater 3-Dimensional Reconstruction Using Imaging Sonar. *International Conference on Advanced Engineering Theory and Applications*, 252–258.

Li, C., Dong, B., & Pan, B. (2019). A flexible and easy-to-implement single-camera microscopic 3D digital image correlation technique. *Measurement Science and Technology*, *30*(8), 085002.

Li, W., & Li, Y. F. (2011). Single-camera panoramic stereo imaging system with a fisheye lens and a convex mirror. *Optics Express*, *19*(7), 5855–5867.

Li, X., Zhao, J., He, S., & Zhang, Z. (2024). Micro bi-prism bi-telecentric single lens Stereo-DIC system using one-step calibration method. *Optics & Laser Technology*, *171*, 110303. https://doi.org/10.1016/j.optlastec.2023.110303

Li, Y., Zhang, Y., Li, H., Zhang, W., & Zhang, Q. (2018). Epipolar geometry and stereo matching algorithm for underwater fish-eye images. *International Journal of Advanced Robotic Systems*, *March-April 2018*. https://doi.org/10.1177/1729881418764715

Lim, K. B., Kee, W. L., & Wang, D. (2013). Virtual camera calibration and stereo correspondence of single-lens bi-prism stereovision system using geometrical approach. *Signal Processing: Image Communication*, *28*(9), 1059–1071.

Lim, K. B., & Qian, B. (2016). Biprism distortion modeling and calibration for a single-lens stereovision system. *J. Opt. Soc. Am. A*, *33*(11), 2213–2224. https://doi.org/10.1364/JOSAA.33.002213

Lim, K. B., & Xiao, Y. (2005). Virtual stereovision system: New understanding on single-lens stereovision using a biprism. *Journal of Electronic Imaging*, *14*(4), 043020.

Liu, Z., Yin, Y., Liu, S., & Chen, X. (2016). Extrinsic parameter calibration of stereo vision sensors using spot laser projector. *Applied Optics*, *55*(25), 7098–7105.

Massot-Campos, M., & Oliver-Codina, G. (2015). Optical sensors and methods for underwater 3D reconstruction. *Sensors*, *15*(12), 31525–31557.

Mouaddib, E. M., Sagawa, R., Echigo, T., & Yagi, Y. (2005). Stereovision with a Single Camera and Multiple Mirrors. *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, 800–805. https://doi.org/10.1109/ROBOT.2005.1570215

Nguyen, T.-N., Huynh, H.-H., & Meunier, J. (2018). 3D reconstruction with time-of-flight depth camera and multiple mirrors. *IEEE Access*, *6*, 38106–38114.

Nishimoto, Y., & Shirai, Y. (1988). A feature-based stereo model using disparity histograms of multi-resolution channels. *Advanced Robotics*, *3*(1), 17–33.

Okutomi, M., & Kanade, T. (1993). A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(4), 353–363.

Ouyang, B., Caimi, F. M., Dalgleish, F. R., Nootz, G., & Vuorenkoski, A. K. (2014). 3D imaging using compressive line sensing serial imaging system. *Compressive Sensing III*, *9109*, 91090V.

Pan, B., & Wang, Q. (2013). Single-camera microscopic stereo digital image correlation using a diffraction grating. *Optics Express*, *21*(21), 25056–25068.

Pan, B., Wu, D., & Xia, Y. (2012). An active imaging digital image correlation method for deformation measurement insensitive to ambient light. *Optics & Laser Technology*, *44*(1), 204–209.

Pan, B., Yu, L., & Zhang, Q. (2018). Review of single-camera stereo-digital image correlation techniques for full-field 3D shape and deformation measurement. *Science China Technological Sciences*, *61*(1), 2–20.

Pérez, L., Rodríguez, Í., Rodríguez, N., Usamentiaga, R., & García, D. F. (2016). Robot guidance using machine vision techniques in industrial environments: A comparative review. *Sensors*, *16*(3), 335.

Pietrobruno, S. (2011). The stereoscope and the miniature. *Early Popular Visual Culture*, *9*(3), 171–190. https://doi.org/10.1080/17460654.2011.601159

Qian, B., & Lim, K. B. (2016). Image distortion correction for single-lens stereo vision system employing a biprism. *Journal of Electronic Imaging*, *25*(4), 043024.

Rabal, H., Henao, R., & Torroba, R. (1996). Digital speckle pattern shearing interferometry using diffraction gratings. *Optics Communications*, *126*(4–6), 191–196.

Ramsgaard, B. K., Balslev, I., & Arnspang, J. (2000). Mirror-based trinocular systems in robot-vision. *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, *4*, 499–502 vol.4. https://doi.org/10.1109/ICPR.2000.902966

Rosas, H. (2011). Perception and Reality in Stereo Vision: Technological Applications. In *Advances in Stereo Vision*. IntechOpen.

Sanchez-Rodriguez, J.-P., & Aceves-Lopez, A. (2018). A survey on stereo vision-based autonomous navigation for multi-rotor MUAVs. *Robotica*, *36*(8), 1225–1243.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, *47*(1), 7–42.

Schonberger, J. L., & Frahm, J.-M. (2016). Structure-from-motion revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4104–4113.

Sharma, K., & Moon, I. (2013). Improved scale-invariant feature transform feature-matching technique-based object tracking in video sequences via a neural network and Kinect sensor. *Journal of Electronic Imaging*, *22*(3), 033017. https://doi.org/10.1117/1.JEI.22.3.033017

Slama, C. C. (1980). *Manual of Photogrammetry.* America Society of Photogrammetry,.

Somanath, G., Rohith, M. V., & Kambhamettu, C. (2010). Single Camera Stereo System Using Prism and Mirrors. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, R. Chung, R. Hammound, M. Hussain, T. Kar-Han, R. Crawfis, D. Thalmann, D. Kao, & L. Avila (Eds.), *Advances in Visual Computing* (Vol. 6454, pp. 170–181). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17274-8_17

Takahashi, K., Nobuhara, S., & Matsuyama, T. (2012). A new mirror-based extrinsic camera calibration using an orthogonality constraint. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 1051–1058.

Teoh, W., & Zhang, X. (1984). An inexpensive stereoscopic vision system for robots. *Proceedings. 1984 IEEE International Conference on Robotics and Automation*, *1*, 186–189.

Tippetts, B. J., Lee, D.-J., Archibald, J. K., & Lillywhite, K. D. (2011). Dense Disparity Real-Time Stereo Vision Algorithm for Resource-Limited Systems. *IEEE Transactions on Circuits and Systems for Video Technology*, *21*(10), 1547–1555. IEEE Transactions on Circuits and Systems for Video Technology. https://doi.org/10.1109/TCSVT.2011.2163444

Trivi, M., & Rabal, H. J. (1988). Stereoscopic uses of diffraction gratings. *Applied Optics*, *27*(6), 1007–1009.

Wang, D., Lim, K. B., & Kee, W. L. (2013). Geometrical approach for rectification of single-lens stereovision system with a triprism. *Machine Vision and Applications*, *24*(4), 821–833.

Wang, L., Gong, M., Gong, M., & Yang, R. (2006). How far can we go with local optimization in real-time stereo matching. *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, 129–136.

Wehr, A., & Lohr, U. (1999). Airborne laser scanning—An introduction and overview. *ISPRS Journal of Photogrammetry and Remote Sensing*, *54*(2–3), 68–82.

Wooden, D., Malchano, M., Blankespoor, K., Howardy, A., Rizzi, A. A., & Raibert, M. (2010). Autonomous navigation for BigDog. *2010 IEEE International Conference on Robotics and Automation*, 4736–4741.

Wu, L. F., Zhu, J. G., Xie, H. M., & Zhang, Q. (2016). An Accurate Method for Shape Retrieval and Displacement Measurement Using Bi-Prism-Based Single Lens 3D Digital

Image Correlation. *Experimental Mechanics*, *56*(9), 1611–1624. https://doi.org/10.1007/s11340-016-0193-7

Wu, L., Yin, Y., Zhang, Q., Fang, D., Zhang, R., Zhu, J., & Xie, H. (2018). Bi-prism-based single-lens three dimensional digital image correlation system with a long working distance: Methodology and application in extreme high temperature deformation test. *Science China Technological Sciences*, *61*(1), 37–50.

Wu, L., Zhu, J., & Xie, H. (2014). A modified virtual point model of the 3D DIC technique using a single camera and a bi-prism. *Measurement Science and Technology*, *25*(11), 115008.

Wu, L., Zhu, J., & Xie, H. (2015). Single-lens 3D digital image correlation system based on a bilateral telecentric lens and a bi-prism: Validation and application. *Applied Optics*, *54*(26), 7842–7850.

Wu, L., Zhu, J., Xie, H., & Zhou, M. (2016). Single-lens 3D digital image correlation system based on a bilateral telecentric lens and a bi-prism: Systematic error analysis and correction. *Optics and Lasers in Engineering*, *87*, 129–138.

Xia, S., Gdoutou, A., & Ravichandran, G. (2013). Diffraction assisted image correlation: A novel method for measuring three-dimensional deformation using two-dimensional digital image correlation. *Experimental Mechanics*, *53*(5), 755–765.

Xia, S., Pan, Z., & Zhang, J. (2014). Optical microscope for three-dimensional surface displacement and shape measurements at the microscale. *Optics Letters*, *39*(14), 4267–4270.

Xiao, Y., & Lim, K. B. (2007). A prism-based single-lens stereovision system: From trinocular to multi-ocular. *Image and Vision Computing*, *25*(11), 1725–1736.

Yang, Q., Wang, L., & Ahuja, N. (2010). A Constant-Space Belief Propagation Algorithm for Stereo Matching. *CVPR*, 1458–1465.

Yang, S.-P., Kim, J.-J., Jang, K.-W., Song, W.-K., & Jeong, K.-H. (2016). Compact stereo endoscopic camera using microprism arrays. *Optics Letters*, *41*(6), 1285–1288.

Yi, S., & Ahuja, N. (2006). An Omnidirectional Stereo Vision System Using a Single Camera. *18th International Conference on Pattern Recognition (ICPR'06)*, *4*, 861–865. https://doi.org/10.1109/ICPR.2006.263

Yu, L., & Pan, B. (2016). Single-camera stereo-digital image correlation with a four-mirror adapter: Optimized design and validation. *Optics and Lasers in Engineering*, *87*, 120–128.

Yuan, T., Dai, X., Shao, X., Zu, Z., Cheng, X., Yang, F., & He, X. (2019). Dual-biprism-based digital image correlation for defect detection of pipelines. *Optical Engineering*, *58*(1), 1–13. https://doi.org/10.1117/1.OE.58.1.014107

Zbontar, J., LeCun, Y., & others. (2016). Stereo matching by training a convolutional neural network to compare image patches. *J. Mach. Learn. Res.*, *17*(1), 2287–2318.

Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *22*(11), 1330–1334.

Zhang, Z. Y., & Tsui, H. T. (1998). 3D reconstruction from a single view of an object and its image in a plane mirror. *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, *2*, 1174–1176 vol.2. https://doi.org/10.1109/ICPR.1998.711905

Zhao, M., Lim, K. B., & Kee, W. L. (2012). Geometrical-analysis-based algorithm for stereo matching of single-lens binocular and multi-ocular stereovision system. *J. Electron. Sci. Technol*, *10*(2), 107–112.

Zhong, F., & Quan, C. (2018). A Single Color Camera Stereo Vision System. *IEEE Sensors Journal*, *18*(4), 1474–1482. https://doi.org/10.1109/JSEN.2017.2786747

Zhou, F., Chai, X., Chen, X., & Song, Y. (2016). Omnidirectional stereo vision sensor based on single camera and catoptric system. *Appl. Opt.*, *55*(25), 6813–6820. https://doi.org/10.1364/AO.55.006813

AUTHOR BIOGRAPHY

Name	Mr. Athukoralage Supun Gayantha Dissanayaka
Date of Birth	July 10, 1990, in Colombo
Address:	213, Saman mawatha, Siddhamulla, Piliyandala, Colombo, 10530

Educational Background:

2014:	Bachelor of Science in Mechatronics Asian Institute of Technology Thailand.
2024:	Master of Engineering in Robotics and Computational Intelligence, King Mongkut's Institute of Technology Ladkrabang.

Work Experience and Research Achievements:

2015 – 2017	Quality Assurance Engineer and Production Engineer at Global Thaixon Precision Industry Co.Ltd.
2018-2024	Laboratory Engineer and Instructor at King Mongkut's Institute of Technology Ladkrabang.