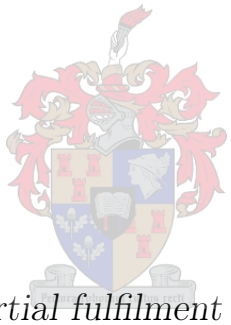
Feasibility Study of Stereo Vision as a Stand-Alone Sensor for Obstacle Detection

by

Armand Nolte



Thesis presented in partial fulfilment of the requirements for the degree of Master of Engineering (Mechanical) in the Faculty of Engineering at Stellenbosch University

Supervisor: Dr. W. Smit

March 2017

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2017

Copyright © 2017 Stellenbosch University
All rights reserved.

Stellenbosch University https://scholar.sun.ac.za

Abstract

Feasibility Study of Stereo Vision as a Stand-Alone Sensor for Obstacle Detection

A. Nolte

Department of Mechanical and Mechatronic Engineering,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MEng (Mech)

March 2017

The need for obstacle detection and the estimation of the relative distance of the objects is increasing in the field of robotics. The feasibility of using stereo vision as a means of detecting objects and their relative distance was investigated in this study. The aim was to implement obstacle detection within an area with moving objects and a moving vehicle. Objects were detected using disparity map generation. Different noise filtering techniques were applied to the disparity map to enhance the results. The noise filtering techniques tested for spatial aliasing, occlusion and low texture regions. To further improve the result, the disparity map of the current time frame was compared to that of its predecessor. The resulting disparity map showed large noise reductions but noise was still present.

With slow-moving vehicles, the improved algorithm gives promising results.

Stellenbosch University https://scholar.sun.ac.za

Uittreksel

Aanneemlikheid Studie van Stereo Visie as Aleen Lopende Sensor vir Hindernis Opsporing

("Feasibility Study of Stereo Vision as a Stand-Alone Sensor for Obstacle Detection")

A. Nolte

Departement Meganiese en Megatroniese Ingenieurswese,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: MIng (Meg)

Maart 2017

Die nut vir voorwerp opsporing en die bepaling van die relatiewe afstand van die voorwerpe word al hoe belangriker in die veld van robotika.

Die haalbaarheid om stereo visie te gebruik as metode om voorwerpe op te spoor en die relatiewie posisie te bepaal, is in hierdie studie ondersoek. Die doel was om voorwerp opsporing in 'n gebied met bewegende voorwerpe en 'n bewegende voertuig toe te pas. Voorwerpe is geïdentifiseer deur ongelykheidskaarte te gebruik. Verskillende geraasfilterings tegnieke is op die ongelykheidskaarte toegepas om die resultate te verbeter. Die tegnieke het vir ruimtelike aliasering, afsluiting en lae-tekstuurareas getoets. Om die resultate verder te verbeter, het ons die huidige ongelykheidskaart met die vorige ongelykheidskaart vergelyk. Die gevolglike ongelykheidskaart het baie minder geraas bevat, maar die geraas was nog nie heeltemal verwyder nie.

Met stadige voertuie gee die verbeterde algoritme belowende resultate.

Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations ...

- Dr. W.J Smit, my supervisor, for guidance, support and interest showed in my research.
- My family, for their ongoing support.

Stellenbosch University https://scholar.sun.ac.za

Contents

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

List of Figures

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

List of Tables

Stellenbosch University https://scholar.sun.ac.za

Nomenclature

Acronyms

CRS	Central receiver system
$LIDAR$	Light detection and ranging
SAD	Sum of absolute differences
$SONAR$	Sound navigation and ranging
$STERG$	Solar thermal energy research group
$SURF$	Speeded up robust features
$UAGV$	Unmanned autonomous ground vehicle
NCC	Normalize cross correlation
$ZNCC$	Variant normalize cross correlation

Variables

d	Disparity	[Pixles]
f	Focal length	[mm]
h	Half window size minus one	[mm]
s	Similarity score	[ ]
w	Window size	[Pixles]
X	Pixel location (column)	[Pixles]
Y	Pixel location (row)	[Pixels]
Z	Distance to a point	[mm]
θ	Rotation angle	[rad]
ϕ	Threshold	[ ]

Vectors and tensors

D	Disparity map
E	Essential matrix
F	Fundamental matrix
I	Image matrix
M	Matching points

Stellenbosch University https://scholar.sun.ac.za

O_s	Output similarity matrix
P	Location of a point
R	Rotation matrix
$\overrightarrow{\boldsymbol{T}}$	Translation vector
T	Template
W	Window template

Subscripts

c	Current
d	Direction
l	Left
L	Left
max	Maximum
min	Minimum
n	Number
p	Previous
r	Right
R	Right
$dmax$	Maximum disparity
$dmin$	Minimum disparity

Stellenbosch University https://scholar.sun.ac.za

Chapter 1

Introduction

Local navigation is a fundamental problem for mobile robots operating in real-world environments. Obstacle detection is a necessity for an unmanned autonomous ground vehicle (UAGV) to be able to roam freely. Generally robots use active sensors such as sonar and laser rangefinders for navigational purposes. Through our own personal experience, however, we know that it is possible to navigate locally using only our eyes for vision.

There are limitations to robots that mainly rely on combining wheel odometry and inertial sensing. Inertial sensors are prone to drift and wheel odometry is ineffective in rough terrain (Howard, 2008). Looking into the more advanced sensing capabilities, sonar is fast and cheap, but usually very crude. Laser range scanners are accurate, but slow and bulky (Agrawal and Konolige, 2006).

Vision systems are a relatively inexpensive approach to obstacle detection. They are light-weight and compact, and can provide mapping at a high frequency. The resolution of the images and the frequency can also be increased by using more expensive cameras.

This study investigated stereo vision technology to determine if it is viable for object detection as a stand-alone sensor. The system focused on detecting obstacles within the field of view to determine the relative position of the obstacles.

To achieve the goal, two webcams were used as stereo vision sensors. The cameras were mounted on a frame to align and keep them stable for use after calibration. The algorithm operated on a continuous feed from the cameras in the form of a recorded video and used disparity maps to generate the three dimensional (3D) coordinates of a detected object using triangulation. Triangulation determines the relative distance from the detected objects to the cameras.

Stellenbosch University https://scholar.sun.ac.za

1.1 Motivation

This study was part of the work done by the robotics research group in the Department of Mechanical and Mechatronic Engineering at Stellenbosch University, working with the Solar Thermal Energy Research Group (STERG).

Sensors served as the data acquisition tool for the object and distance estimation. Light detection and ranging sensors (LIDAR) can be used for object detection and distance estimation and are highly effective (Badino *et al.*, 2011), but are also very expensive. Cameras on the other hand, are far less expensive. Thus the question was raised whether it is possible to accurately detect objects when using only stereo vision.

1.2 Objectives

The main objective of the study was to determine if stereo vision is viable as a stand-alone sensor for object detection in a moving environment. In order to achieve the study's main objective, the following matters were addressed:

- Review literature on obstacle detection and distance estimation.
- Select a suitable method for detecting objects and estimating the relative distance towards the object.
- Determine if the results need to be improved by looking at noise within the data.
- Implement the resulting algorithm on a video feed to determine if objects are detected and distance is estimated.

1.3 Overview of the Algorithm

The purpose of the study was to detect an object in the field and to calculate the distance towards it with little noise in the results. A brief overview of the study's algorithm is provided in this section to help give a broad view of how the study was implemented.

The offline algorithm started by loading video recordings of a test field that had been created. The video recordings were video files from two cameras. The cameras had been positioned next to one another, and the study referred to them as the left-hand camera and the right-hand camera. After the video files had been loaded, frames were grabbed from the files and tested for a flash of light. The flash of light, created by a flashlight, indicated the start of the test. The start was identified by looping through the video files and searching for the flash. The first frame of the video file containing the flash was taken as the starting point. The flashes indicated that those frames had been captured

Stellenbosch University https://scholar.sun.ac.za

at the same time. By inducing a flash on the video, the cameras could be synchronised, after which the analysis of the gathered data could commence.

The first step in the process was to rectify the input images. With rectified images, the computational time was reduced. The next step was to compute the disparity maps. The disparity maps gave an indication of where objects were located. The disparity maps were used to determine the 3D locations of objects that had been detected. An motion compensation feature was used to reduce the noise in the disparity maps. The motion compensation was an input parameter for the temporal frame analysis. Temporal frame analysis compares disparity maps at two moments in time with one another. The disparity map that was improved was that of the current time frame. The improvement removed excess noise from the disparity map. Clearer disparity maps result in generating a clearer 3D reconstruction of the scene. The 3D reconstruction is a function of the disparity map and camera parameters. This process continued until the end of the video file. The flow diagram for the main algorithm is shown in Figure 1.1.

Stellenbosch University https://scholar.sun.ac.za

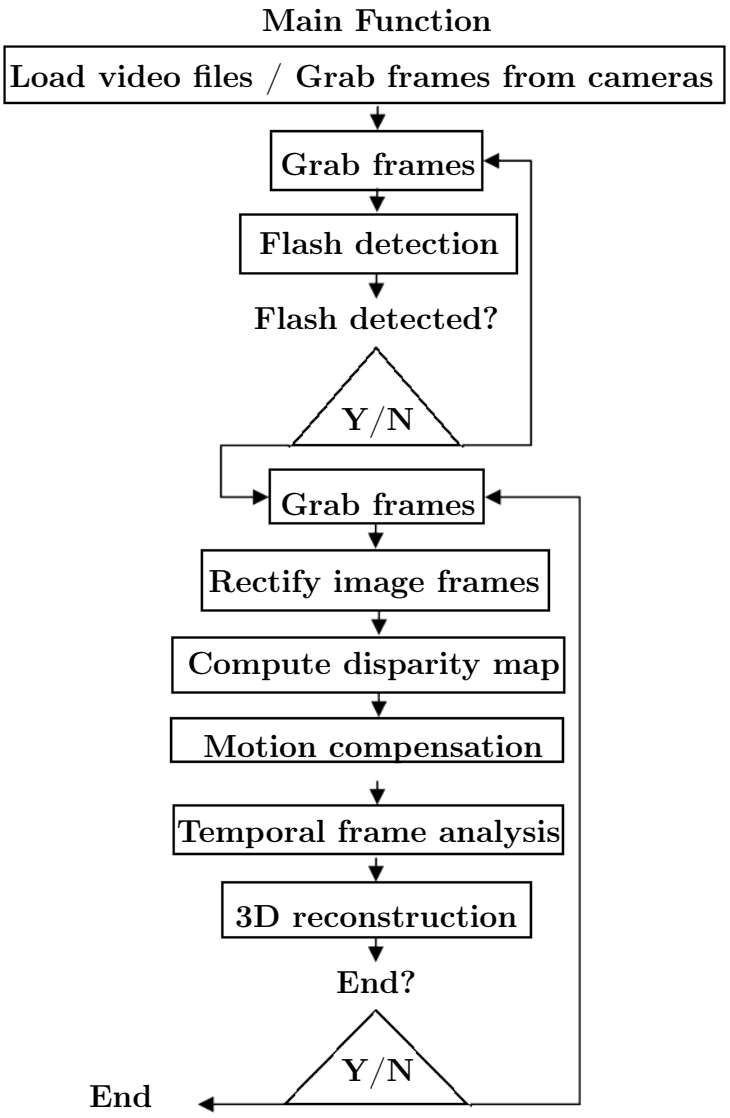


Figure 1.1: Flow diagram of the main algorithm. The process starts by grabbing frames from the video recordings, rectifying the images, computing the disparity map, determining the motion of objects, comparing the current disparity map with its predecessor to remove noise and finally creating a 3D reconstruction.

Stellenbosch University https://scholar.sun.ac.za

Chapter 2

Literature Study

Research was conducted to determine how stereo vision can be implemented for object detection and depth estimation. The literature study is discussed in this chapter and focused on the background, camera calibration, pinhole camera model, epipolar geometry, triangulation and disparity map generation and the problems that can arise with in the process.

2.1 Background

A large number of studies have been done that implement stereo vision and disparity maps for object detection and avoidance (Howard, 2008; Kostavelis *et al.*, 2009; Häne *et al.*, 2015). With stereo vision, object detection can be achieved by finding the same point in two images. With a matching pair of points the 3D world coordinates can be determined. Feature matching is a method that can be used to detect objects within a scene and is discussed by Laganière (2011). Feature matching finds the corresponding features in a second image, producing the pixel coordinates of matching features. The pixel coordinates of a feature can be used to determine the 3D world coordinates of the feature. The problem with feature matching is that it does not fill the entire body of an object with matching points, so that partial points of an object's body are detected as objects, and not the body as a whole. This is illustrated in Figure 2.1 where two images, one from the left-hand camera and one from the right-hand camera, are superimposed onto one another. A speeded-up robust features (SURF) feature matching technique was used to determine the matches between two images. The matching points are denoted with "+" for the right-hand image and "o" for the left-hand image with a line drawn between them. It can be seen that only partial points of the person in front were identified as matching points. Using only the matching features for a 3D reconstruction is not suitable for object detection.

The results of the matches were limited due to the objects not being the same in both images. The feature matching is very sensitive to the object's

Stellenbosch University https://scholar.sun.ac.za

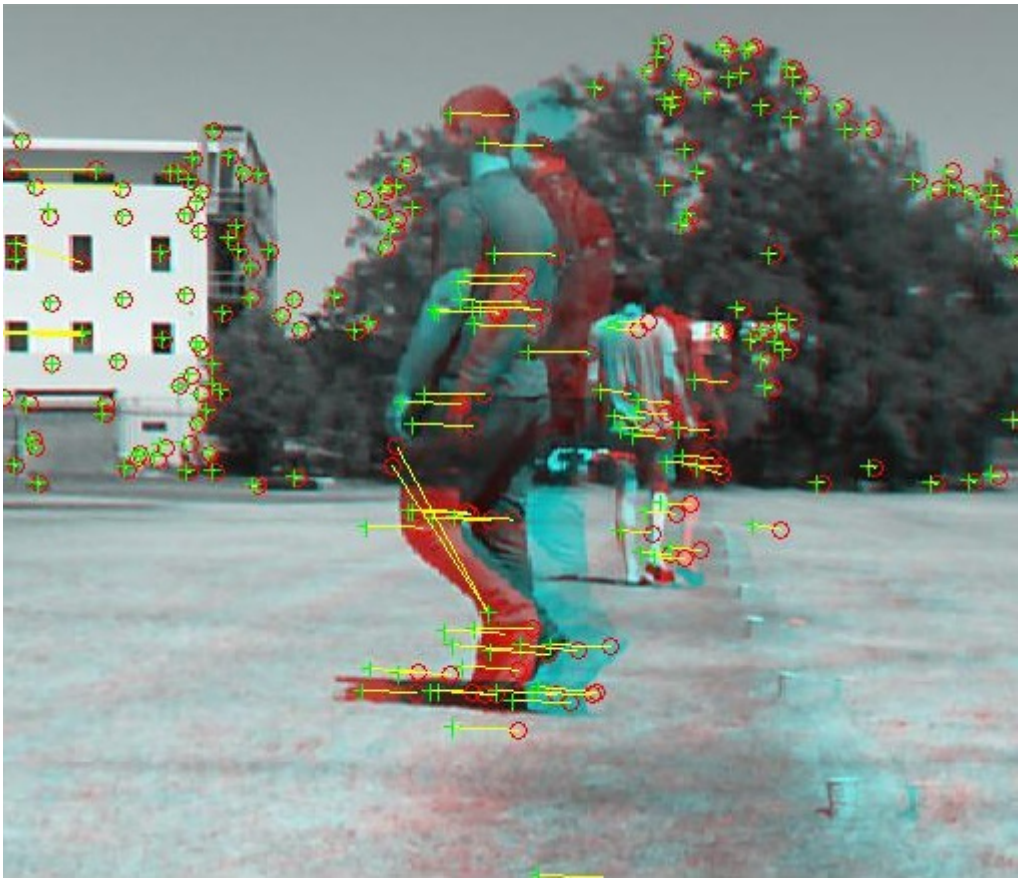


Figure 2.1: SURF matches found between two images, one from the left-hand camera with matching points "*o*" and one from the right-hand camera with matching points "+".

orientation, depending on the view of the object (Arnfred *et al.*, 2013). Take an object in the field and assume the object is close to the cameras. The camera on the left does not see the same features as the camera on the right, producing only a small number of matched features for that object. For this reason, disparity map formulation was implemented. (Pollefeys *et al.*, 2008) stated that it is possible to use real-time video and disparity maps to reconstruct a 3D environment and detect objects within a scene. Problems that occurred throughout the study included the large size of the video data that was processed, the large variability of illumination, the varying distance and orientation of the observed scene and the presence of objects that are hard to model, such as trees and windows. With disparity maps more matching points can be generated within the two images, resulting in a better 3D reconstruction.

Stellenbosch University https://scholar.sun.ac.za

2.2 Camera Calibration

As was discussed by Bradski *et al.* (2013), stereo calibration is the process of computing the geometrical relationship between two cameras. The calibration aims to find the rotation matrix (R) and the translation vector (T) that describes the location of the second camera relative to the first camera. These parameters are then used to formulate the essential matrix (E) and the fundamental matrix (F). The matrix (E) contains information about the rotation and translation that relates to the two cameras in their physical space. The matrix (F) contains the same information, with the addition of the intrinsic parameters of both cameras. The rotation and translation are illustrated in Figure 2.2, with the left-hand camera centre O_l, the right-hand camera centre O_r, the left-hand image plane I_l, the right-hand image plane I_r and the world point P. These parameters were used to determine the 3D coordinates of the objects that were detected.

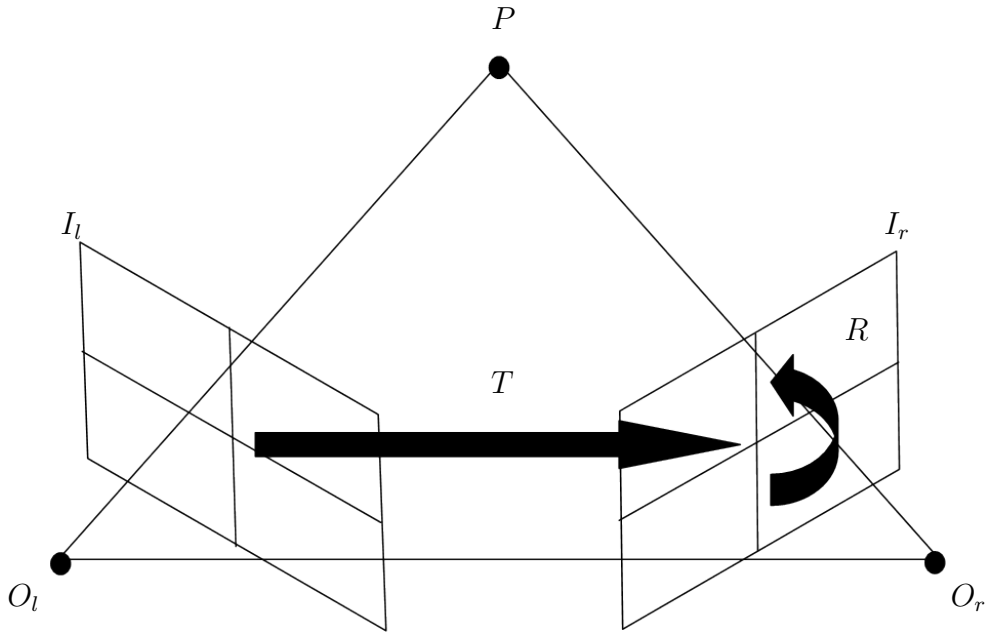


Figure 2.2: The geometry of stereo imaging indicating the translation and rotation (Bradski *et al.*, 2013).

2.3 Pinhole Camera Model

Bradski *et al.* (2013) also discuss the pinhole camera model. The pinhole camera model is a simple model of a camera that operates in such a manner that only a single ray of light reaches the camera. As a result, the image

Stellenbosch University https://scholar.sun.ac.za

on the image plane is always in focus. The size of the image relative to the distant object from which the ray is coming is given as a single parameter of the camera. The single parameter is the focal length and the distance from the pinhole aperture to the screen is the focal length. This is illustrated in Figure 2.3, where f is the focal length of the camera, Z is the distance from the camera to the object, X is the length of the object and x is the object's image on the image plane. From Figure 2.3 the relation for x can be formulated as seen in Equation (2.1).

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

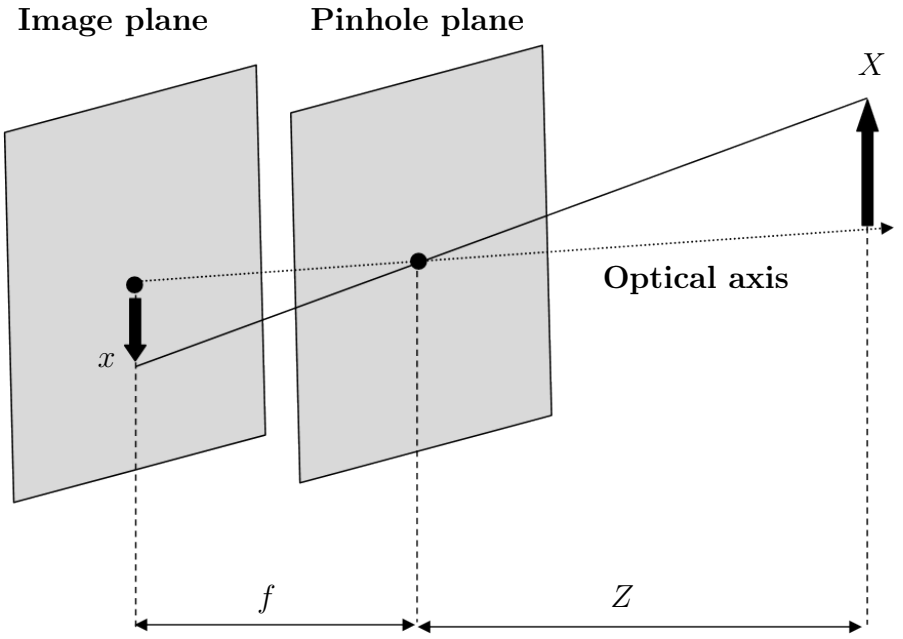


Figure 2.3: Pinhole camera model (Bradski *et al.*, 2013).

2.4 Epipolar Geometry

Rectification was used to create the disparity map. Using rectified images reduces the computational time of the disparity. As can be seen in Bradski *et al.* (2013), the basic geometry of stereo vision systems is known as epipolar geometry. The geometry combines two pinhole camera models and epipoles to determine the location of a visible point. Figure 2.4 illustrates the geometry concept. For every camera, there is a camera centre (O_l and O_r) and its corresponding projective plane, also known as the image plane. The physical real world point P is projected onto each of the image planes and is labelled

Stellenbosch University https://scholar.sun.ac.za

p_l and p_r. The points of interest are the epipoles, labelled as e_l and e_r. The epipoles are located on the image planes.

Epipoles are defined as the image of the centre of projection of the other camera. The concept is easy to understand when Figure 2.4 is inspected. The line drawn between the camera centres forms a point on the image plane. Those points are the epipoles. The plane in space formed by the point P and the two epipoles e_l and e_r is called the epipolar plane. The two lines that are formed by $p_l e_l$ and $p_r e_r$ are called the epipolar lines.

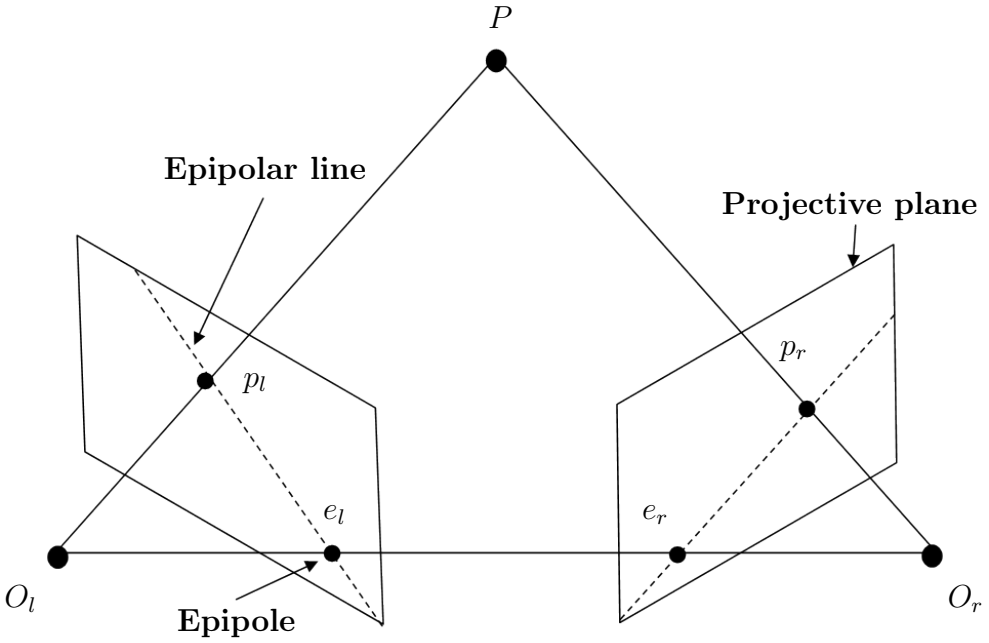


Figure 2.4: Epipolar geometry indicating the world point P, camera centres O_l and O_r, and the epipoles e_l and e_r (Bradski *et al.*, 2013).

When inspecting a world point P on an image plane, the point on the image plane can be located anywhere along the line of points formed by the point P and the camera centre O. This is because, with just a single camera, the distance from the camera to the point is unknown. Take the camera on the left, for example. The point P is seen by the camera on the left as p_l on the left-hand image plane. The point P can be located anywhere along the line formed by p_l and O_l. The point P is definitely on the line, but so are a lot of other possible points.

This is where stereo vision comes into play. The line formed by p_l and O_l is the line formed by p_r and e_r on the right-hand image plane. The concept of epipolar geometry can be summarised as per the list by Bradski *et al.* (2013):

- Every 3D point viewed by the cameras is contained in an epipolar plane that intersects each image plane in an epipolar line.

Stellenbosch University https://scholar.sun.ac.za

- Given a feature in one image, its matching view in the other image must lie along the corresponding epipolar line. This is known as the epipolar constraint.
- The epipolar constraint means that the possible two-dimensional search for matching features across an image pair becomes a one-dimensional search along the epipolar lines. Thus it is important to know the epipolar geometry of the stereo vision system.
- Another feature to take note of is order preservation. The order in which points occur in the first image is the same order in which they will occur in the second image.

2.5 Triangulation with disparity

Triangulation is also discussed by Bradski *et al.* (2013). Let us assume the stereo camera set-up is perfectly aligned. Both cameras are facing in the same direction and have parallel optical axes. Also assume that both cameras have the same focal length. If the baseline (B) is known, the distance between the cameras - the distance to where a point is located at - can be calculated by using the disparity of that point. This is illustrated in Figure 2.5, where Z is the distance of a point from the cameras, B the baseline and P is the real world point. The concept is formulated into Equation (2.2), where d is the disparity.

$$Z = f\frac{B}{d}. \tag{2.2}$$

2.6 Disparity

With a pair of stereo images, the disparity for any visible feature within the stereo pair can be calculated. If, for example, a feature is located at position $[X_l, Y_l]$ in the left-hand image and the same feature in the right-hand image is located at $[X_r, Y_r]$, then the disparity is the distance between these coordinates. This can easily be seen in Figure 2.6, where a stereo anaglyph of a pair of stereo images is shown. A stereo anaglyph occurs when two images are superimposed onto one another to indicate the difference between them. The location of the ear was taken as the point of interest. Equation (2.3) calculates the disparity (d) between the two points in Figure 2.6. Rectification allows us to only use the X coordinates to calculate the disparity.

$$d = [X_l] - [X_r]. \tag{2.3}$$

Stellenbosch University https://scholar.sun.ac.za

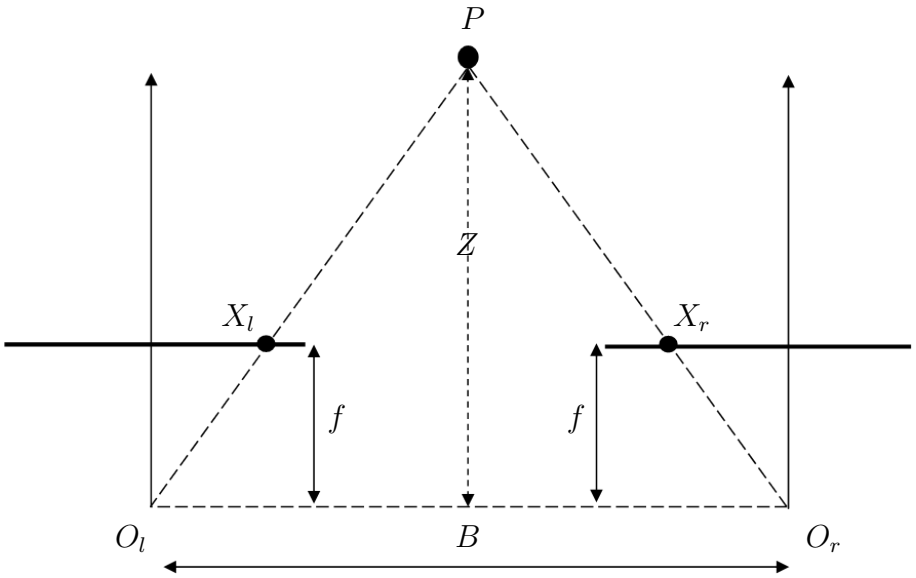


Figure 2.5: Aligned stereo rig and known matching points (Bradski *et al.*, 2013).

Howard (2008) is of the opinion that there are some operations that need to be performed before the disparities of a given stereo pair can be calculated. These operations entail rectification and the application of pre-filters. Rectification is done by aligning the epipolar lines of the left-hand image with those of the right-hand image. This transforms the multi-directional problem into a single-directional one, making it possible to search in one line of pixels for a match. The application of pre-filters improves the result of the disparity map as well. These pre-filters smooth the images while preserving the edges of the objects in the image. The pre-filters that were applied are the median and Laplacian-of-Gaussian filters.

Disparity maps are essentially matrices that are filled with each pixel's disparity value. The application of filters can improve the quality of a disparity map, but filters are not the only improvement that can be made. Several problems with the calculation of the disparity can result in lower-quality disparity maps. Corke (2011) states that there are many factors that can compromise the accuracy of disparity maps. The accuracy is affected by the amount of noise generated within the disparity map. The noise referred to is the mismatching of points within the image pair. These mismatches are features that do not exist or are wrong matches. These problems are the result of spatial aliasing, occlusion and low-texture regions, and are discussed in section 2.8.

2.7 Computing Disparity

Disparity map computation is discussed by Corke (2011). To determine the disparity of a point, a search range must be specified to indicate the expected

Stellenbosch University https://scholar.sun.ac.za

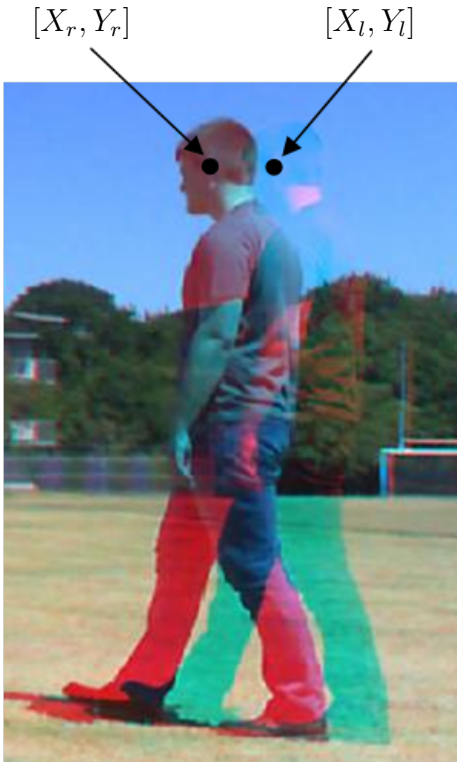


Figure 2.6: The images captured from the stereo pair are displayed over one another to indicate the matching features for disparity calculation. The point of interest is the ear of the person.

region in which the matching point can lie. The search range starts at the minimum disparity (d_{min}) and ends at the maximum disparity (d_{max}). These disparity ranges are selected by the user. With the use of rectified images, a multi-directional problem could be transformed into a single-directional problem, resulting in a one-directional search line. This implies that searches need only be done in a horizontal direction due to the epipolar constraint. A simple illustration of the disparity between two points can be seen in Figure 2.7, where a stereo anaglyph of the left-hand and right-hand images was used.

To determine the disparity for a pixel, a $w \times w$ window template was drawn around the pixel and the region was called T. Instead of a single pixel in the left-hand image being matched to a single pixel in the right-hand image, their templates were matched. The template around the pixel in the right-hand image was called W. The corresponding point was taken as the position where these templates were most similar. The output similarity ($O_s[X, Y]$), illustrated in Figure 2.8, was determined for each pixel by calculating the similarity scores (s) for the pixel pair as a function of their templates. This is shown in Equation (2.4).

Stellenbosch University https://scholar.sun.ac.za

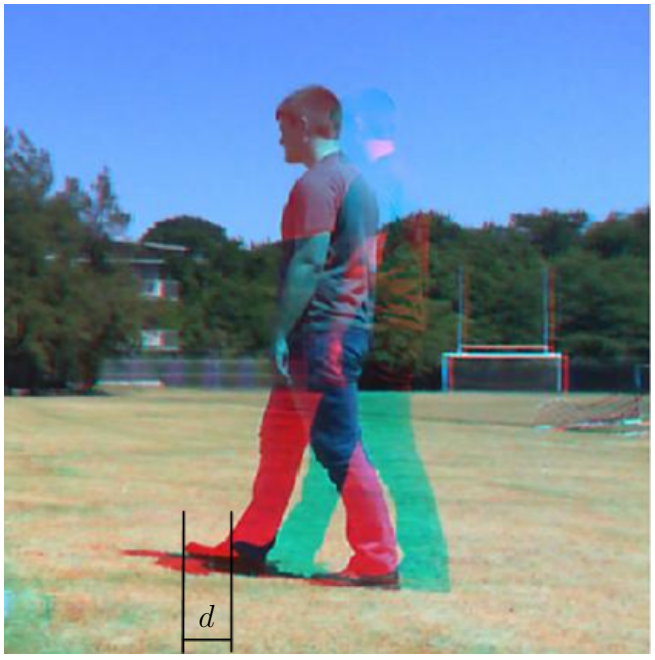


Figure 2.7: Stereo anaglyph of the left- and right-hand images indicating the disparity of a point.

$$O_s[X, Y] = s(T, W). \tag{2.4}$$

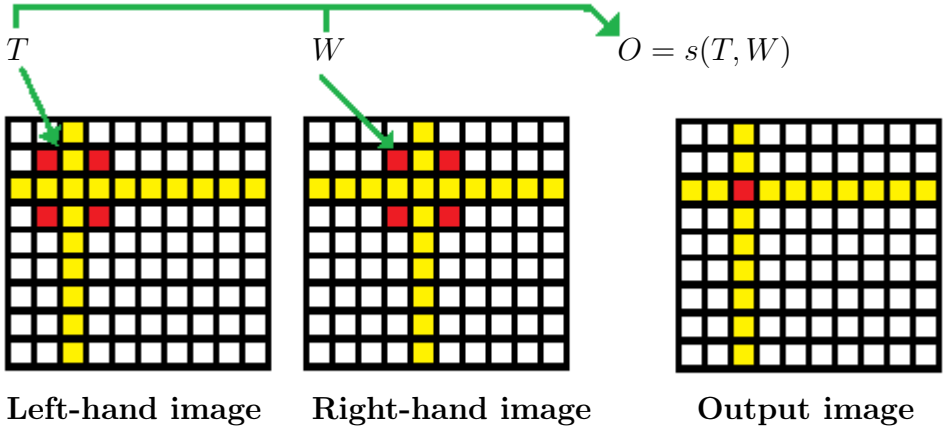


Figure 2.8: Template matching of two images to determine the similarity scores.

Fradi *et al.* (2011) tell us that it is important to select an appropriate window size. Large windows increase the probability of an error within the

Stellenbosch University https://scholar.sun.ac.za

matching pair by calculating the similarity over a larger area, using more values to determine the score. Large windows also blur the borders of objects. Small windows result in poor disparity estimates due to intensity variations between templates.

The disparity is the displacement along the epipolar line $d = L(X) - R(X)$, with L and R indicating the image and X being the horizontal pixel position. Taking a pixel in the left-hand image, (X_i, Y_i), the corresponding point will be located at (X_{i+d}, Y_i) with $d \in [d_{min}, d_{max}]$ in the right-hand image. Thus the template is shifted one pixel at a time from d_{min} to d_{max}. A template matching operation that calculates the similarity score can determine if the templates are similar. The template positioning is illustrated in Figure 2.9.

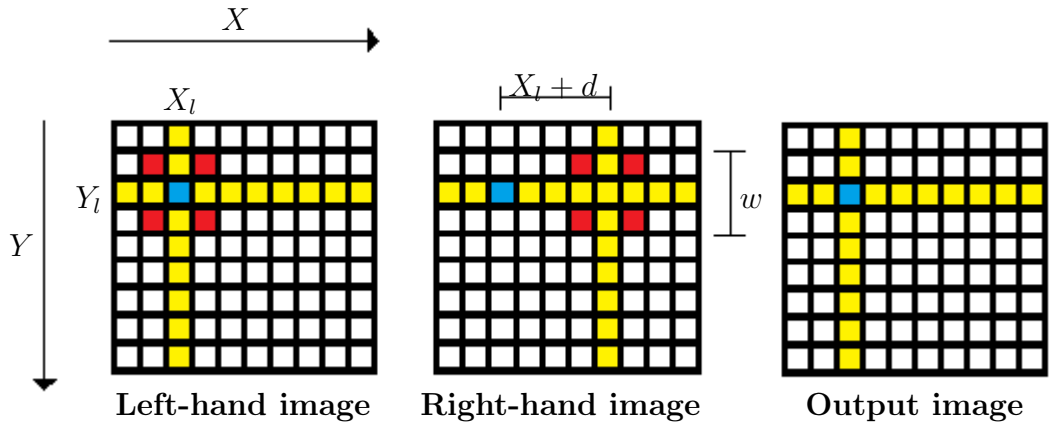


Figure 2.9: Template positioning with respect to the disparity.

The template is a square box with an odd side length, $w = 2h + 1$, centred around the point of interest, with h being half the length of the window minus 1. The easiest way to compare whether the templates are a match is to compute the sum of absolute differences (SAD) between the two templates, as seen in Equation (2.5). The SAD aims to achieve a similarity score of zero to indicate a perfect match. An alternative method for comparing the templates is to use the normalised cross-correlation (NCC). NCC is invariant to change in intensity, meaning that even if the one image is darker than the other, a relatively good match can still be found. Aditya *et al.* (2014) say that the lighting of a scene can have major effects on the calculation of the disparity map. Thus we aim to minimise the effect light can have by incorporating alternative methods to SAD. To account for offset in pixel values, each template's own mean value is subtracted from them, resulting in the NCC becoming the ZNCC. This can be seen in Equation (2.6). The Z- prefix denotes variants of the similarity measures. The ZNCC is invariant to intensity offset. The ZNCC and NCC

Stellenbosch University https://scholar.sun.ac.za

aim to achieve a similarity score of 1 to indicate a perfect match. In practice a value greater than 0.8 is deemed acceptable. These similarity calculations were discussed by Corke (2011).

$$s = \Sigma_{(X_i,Y)\in I}|I_1[X_i, Y] - I_2[X_i, Y]|. \tag{2.5}$$

$$s = \frac{\Sigma_{(X_i,Y)\in I}(I_1[X_i,Y]-\overline{I}_1)\cdot(I_2[X_i,Y]-\overline{I}_2)}{\sqrt{\Sigma_{(X_i,Y)\in I}(I_1[X_i,Y]-\overline{I}_1)^2\cdot\Sigma_{(X_i,Y)\in I}(I_2[X_i,Y]-\overline{I}_2)^2}}. \tag{2.6}$$

2.8 Problems with Template Matching

Problems found within the disparity maps and their solutions are discussed in detail by Corke (2011). The template-matching operation determines the best match for the template within the given disparity range. The disparity of the best match is then stored in the disparity map. If no suitable match was found, a zero value is stored in the matrix at the current location. The best match does not necessarily mean that the match is correct even if the similarity score is above 0.8. Incorrect matches can be identified by looking at the similarity score for the template across the entire disparity range. The three main problems that can occur are spatial aliasing, occlusion and broad peak, also known as low texture regions. These problems are illustrated in Figure 2.10. The perfect match is illustrated in Figure 2.10(a), where a single strong peak is visible. A single strong peak occurs when the similarity score for a given disparity range has one score that is higher than 0.8.

Multiple peaks can be seen in Figure 2.10(b). The amplitudes of the peaks are nearly similar and both qualify as a good match, indicating that the template was found twice in the reach region. The problem can be caused by regular vertical features within the scene, for example rows of windows or a picket fence. This is known as spatial aliasing and is illustrated in Figure 2.11. The template in the left-hand image can match any of the templates of the right-hand image. There is no real solution to the problem when only two cameras are used, but it can be detected. This problem can, however, be solved by using more than two cameras, resulting in more frames to which to compare the template. The ambiguity ratio was used to detect the problem. The ambiguity ratio, also called the peak ratio, is the ratio between the amplitude of the two highest peaks, calculated by dividing the second highest peak by the highest peak. High peak ratios indicate uncertainty and the disparity of that point should be discarded. The chance of detecting incorrect peaks can also be reduced by reducing the disparity range. Reducing the disparity range requires some knowledge of the expected range of objects. Small disparity ranges result in failure to detect objects that are close to the cameras. Some peak ratios are illustrated in Table 2.1. For good results, aim for a peak ratio of below 0.9.

Stellenbosch University https://scholar.sun.ac.za

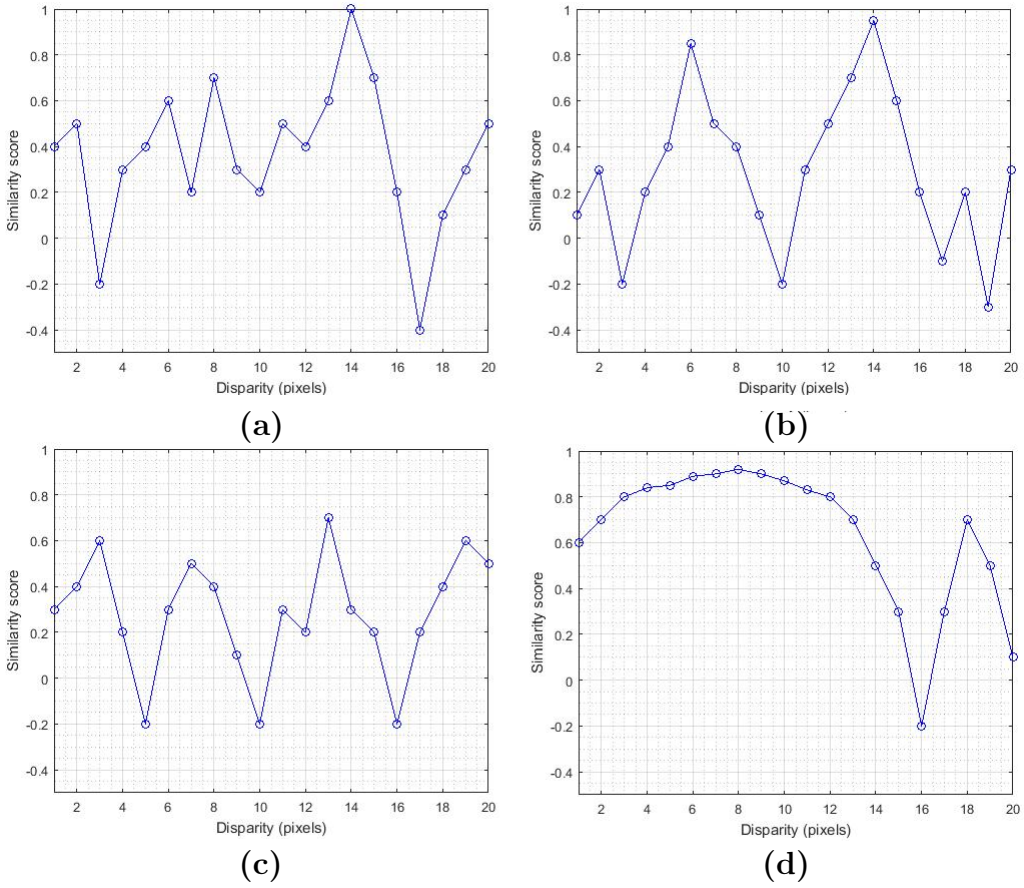


Figure 2.10: Typical examples of similarity scores versus disparity range. (a) Single strong peak; (b) Multiple peaks; (c) No strong peak and (d) Broad peak.

The similarity scores of occlusion can be visualised as a range of weak peaks, as is illustrated in Figure 2.10(c). Naturally this point will be discarded as a weak match. This problem occurs when the point from one image is not visible in the other image, and is illustrated in Figure 2.12. Points 3 and 5 are an example of occlusion occurring. Either of these points can be seen by only one camera, even though they are within the combined field of view. Points 1 and 7 can be viewed by only one of the cameras and will result in a weak match. The problem can also occur if the disparity range is too small. Occlusion can occur more frequently if the baseline increases. Testing for occlusion can be done by matching in two directions. This implies that instead of just matching pixels from the left-hand frame to the right, matching should be done from right to left as well. Start by finding the strongest match in the right-hand image of the pixel in the left-hand image. The point of the strongest match in the right-hand image is then searched for in the left-hand image. The point is considered valid if the point matches the original point.

Stellenbosch University https://scholar.sun.ac.za

Table 2.1: Disparity ranges for thresholds and the distance difference and average for the range

Peak ratios for different peak combinations		
Highest peak	Second highest peak	Peak ratio
1	0.8	0.8
0.8	0.6	0.75
0.8	0.7	0.875
0.9	0.85	0.94

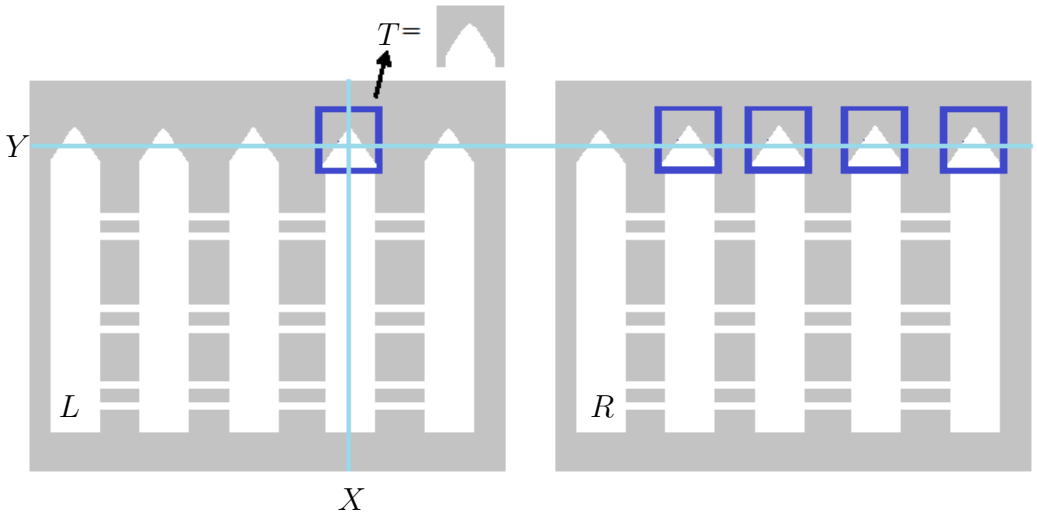


Figure 2.11: Spatial aliasing is illustrated in the form of a picket-fence problem. The template of the left-hand image can be any of the templates in the right-hand image.

The third problem that can occur is illustrated in Figure 2.10(d) and is the result of the similarity score in the form of a broad peak. The broad peak makes it difficult to see where the correct match is. This problem arises when the template has a very low texture, for example matching the sky, water or dark shadows. This can be compared to matching a template filled with black pixels to a range of templates that are all filled with black pixels. The measure of peak sharpness was implemented to handle these kinds of situations. The sharpness of the peak can be determined by looking at the angle θ it produces between the three points, as illustrated in Figure 2.13. The angle is a function of the horizontal displacement between the two adjacent points and the vertical displacement of the highest point to the adjacent points. The horizontal displacement of the points occurs in increments of one and will always be constant. If the horizontal displacement as a variable is removed

Stellenbosch University https://scholar.sun.ac.za

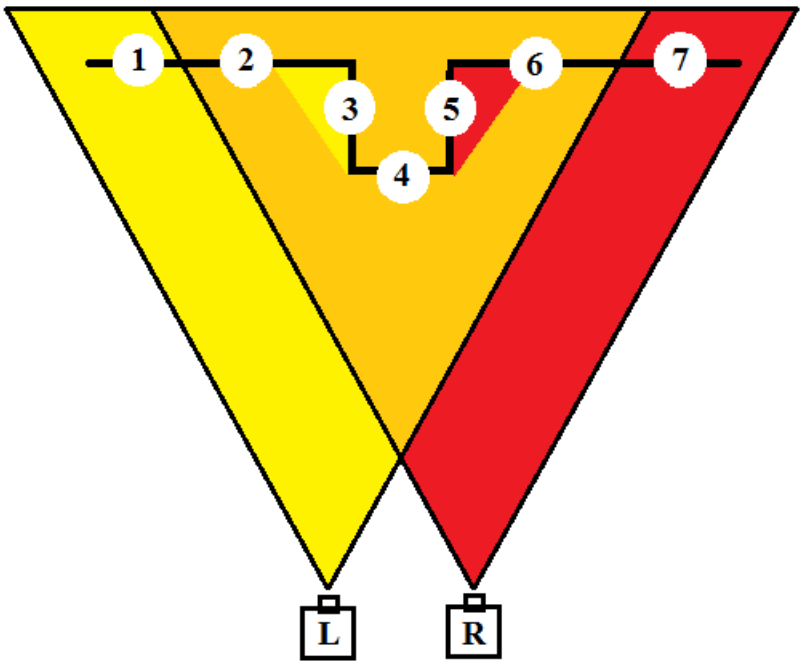


Figure 2.12: Occlusion in stereo vision. Points 1 and 7 fall outside the field of view of the combined vision. Point 3 is not visible to the camera on the right. Point 5 is not visible to the camera on the left (Corke, 2011).

from the angle calculation, only the vertical displacement is seen as a variable. The sharpness of the angle can therefore be determined by looking only at the vertical displacement between the points.

The sharpness of the peak was estimated by doubling the similarity score of the highest peak and then subtracting the similarity score of the two adjacent points, as shown in Equation (2.7). The parameters are the similarity scores of the peak (P), the adjacent left-hand point (P_L) and the adjacent right-hand point P_R. The equation gives the combined distance between the adjacent points and the peak point. The maximum achievable angle between the points is 45°, which corresponds to P being one and P_L and P_R being zero. The sharpness estimate was then compared to a threshold to determine if the peak was sharp enough. The sharpness needed to exceed a value of 0.3 to be deemed sharp enough as is indicated by the upper region in Figure 2.13.

$$PeakThreshold = 2P - (P_L + P_R). \tag{2.7}$$

A variety of elements influence the result of the disparity map. Increasing the baseline distance between the cameras increases the disparity, resulting in increased accuracy in depth estimation. The counterpart is that it gave rise to occlusion. The disparity search range also needs to be taken into consideration.

Stellenbosch University https://scholar.sun.ac.za

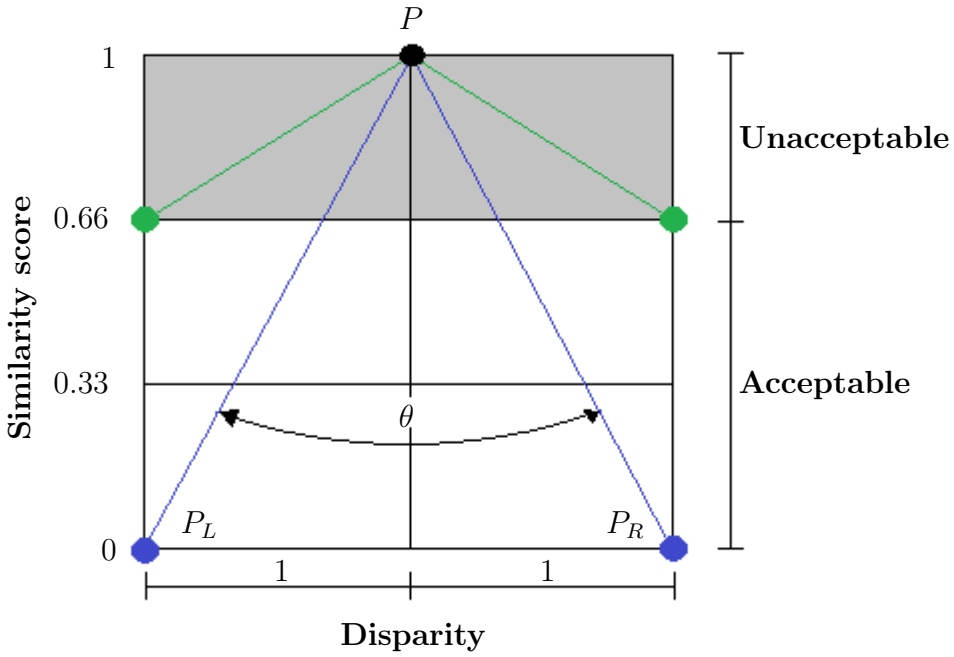


Figure 2.13: Comparing the similarity scores of the peak point to adjacent points. The threshold is in the top section of the figure.

Setting the maximum disparity range to a too large value increases the chances of spatial aliasing and increases the computational time. Setting the value too small will cause the object close to the cameras to generate incorrect and weak matches. The template size affects the computational time and the quality of the resulting disparity map. A large template size increases the computational time and gives a smoother disparity image. A small template size gives a noisier disparity map as it is susceptible to ambiguous matches.

Stellenbosch University https://scholar.sun.ac.za

Chapter 3

Calibration and Image Preparation

Stereo vision works on a principle similar to the human eyes. Being able to see the same object from both eyes enables us to estimate the distance we are from the object. Taking two images of a scene, from two camera viewpoints, the 3D coordinates of that object can be determined using triangulation. There are several problems that can prevent one from determining these coordinates. The first problem that can arise is if the location and orientation of the cameras relative to one another are unknown. Even if the object is visible in both images, calculating the coordinates will be very difficult. The second problem that can present itself is based on the camera properties. The focal length of the camera needs to be ascertained to determine the distance towards the object. The focal length of each camera can be determined using a calibration process.

Modern cameras have an autofocus option that can cause a problem. Autofocus adjusts the zoom of the camera as it focuses on an object in a scene. The zoom can either enlarge or shrink the size of an object, depending on whether it is zooming in or out. The difference in the size of an object between frames will change the pixel location of an object, thus changing the location of the detected object. To keep the object at the same location between frames, we disable autofocus. Setting the autofocus to manual will keep the focus fixed. Capturing consecutive images will show the same object at the same size in all the images.

The use of autofocus will lead to incorrect calibration parameters, even if the process is done correctly. For the purpose of this study, autofocus was disabled and the focal length was determined after the calibration had been done. In this chapter the calibration process, rectification and pre-filters that were applied before the images can be analysed are discussed.

Stellenbosch University https://scholar.sun.ac.za

3.1 Calibration

The camera system needs to be calibrated before the 3D world coordinates of an object can be determined. The output from the calibration process yields the positioning and orientations of the cameras relative to one another. The calibration was done using $MATLAB's$ stereo calibration toolbox and the necessary camera properties were obtained using this toolbox (The MathWorks Inc., 2013). The input for this application is a series of checkerboard images from both cameras and the size of the checkerboard cubes. The camera set-up needs to be placed in a fixed position and orientation relative to one another before the calibration images can be captured. Changing the location of these cameras during the image-capture process will result in a faulty calibration. The calibration object was a checkerboard and is illustrated in figure 3.1. The calibration needed a minimum of 10 image pairs, increasing the number of results in a more accurate calibration. The location and orientation of the checkerboard relative to the camera needed to change between each image pair. Adjusting the location of the checkerboard increased the accuracy of the calibration. The location had to change in distance away from the camera as well. Adjusting the orientation of the checkerboard allowed the calibration to take into account the lens distortions.

The calibration application looks for corners of the checkerboard squares by using one of $MATLAB's$ build in functions called $detectCheckerboardPoints$. The size of the checkerboard side lengths must not be the same. Having different side lengths allows the application to determine the orientation of the checkerboard. Knowing the orientation of the checkerboard, that all cubes are the same size and following the principle that the corners will occur in the same order in the left-hand image as they would in the right-hand image, the same points can be found in both image and the probability of incorrect matches is reduces.The checkerboard corner detection is illustrated in Figure 3.1.

The extrinsic parameters could be obtained once the calibration had been completed. The 3D extrinsic parameters could be plotted to give a visual confirmation that the calibration process had been successful. The plot, illustrated in Figure 3.2, provide the location of the camera centres and the locations of the detected checkerboards. Unsuccessful calibration could be detected by visual inspection. Checkerboards that appeared behind the cameras indicated a calibration error. The obtained rotation matrix and translation vector can be seen below.

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.9999 & -0.0118 \\ 0 & 0.0118 & 0.9999 \end{bmatrix}$$

$$T = \begin{bmatrix} -117.0797 & 1.0197 & 10.6807 \end{bmatrix}$$

Stellenbosch University https://scholar.sun.ac.za

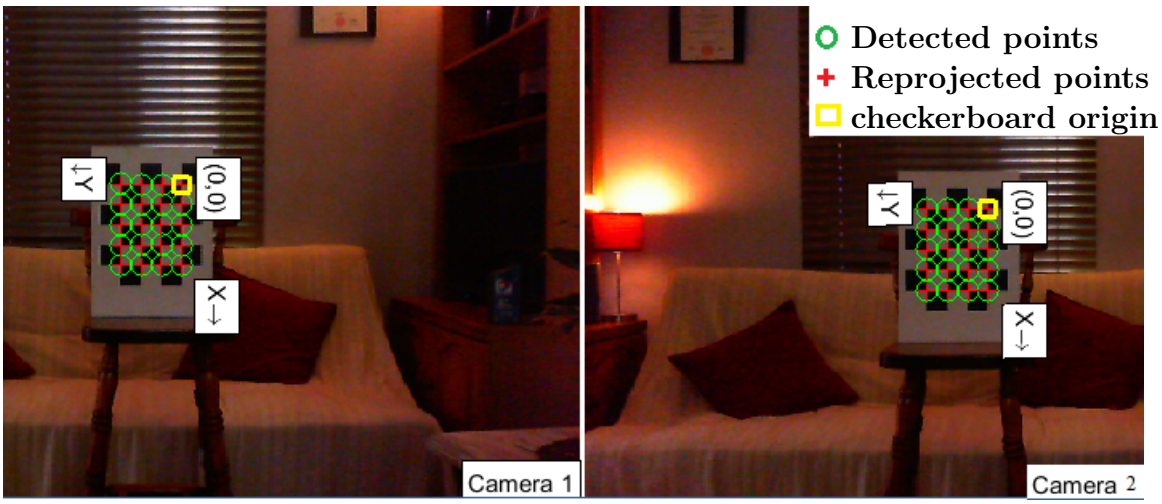


Figure 3.1: checkerboard cube detection.

3.2 Rectification and Pre-Filtering

After the extrinsic parameters were obtained from calibration, the images could be rectified. Rectification is the process where the epipolar lines of the left-hand image are aligned with the epipolar lines of the right-hand image, thus

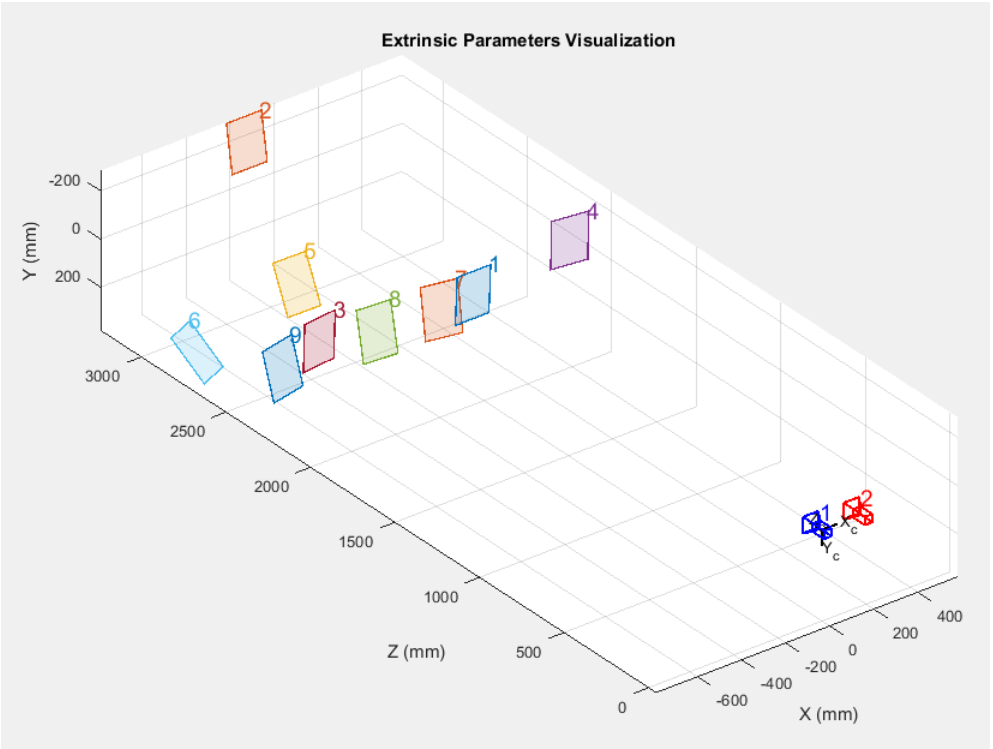


Figure 3.2: Extrinsic parameters visualisation.

Stellenbosch University https://scholar.sun.ac.za

creating a one-dimensional search line when searching for matching points. This rectification was achieved by rotating and transforming the images. An example of a rectified pair of stereo images is shown in Figure 3.3. The same image pair from Figure 3.1 was used to better illustrate the rectification.

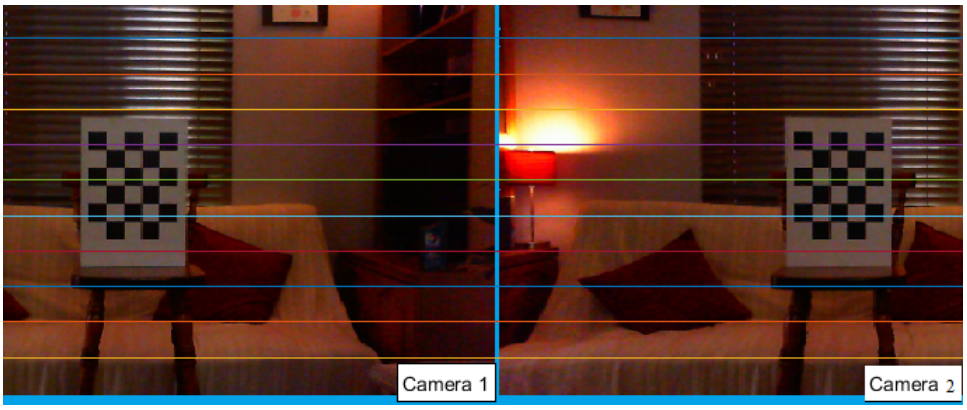


Figure 3.3: Rectified image pair.

Pre-processing of the images was done to improve the reliability of the disparity maps. The pre-processing techniques used were histogram equalisation combined with a low pass- and median filter. Histogram equalisation created an output image with high contrast. This was achieved by adjusting image intensities. Applying the histogram equalisation also introduced extra noise to the input image and thus needed to be filtered properly before the disparity map was calculated. The Gaussian filter was applied, followed by a median filter. The Gaussian filter reduced the noise content and smoothed the image while preserving the main features within the image. The median filter was used to remove the salt-and-pepper noise. Figure 3.4 illustrates a standard input image and one that has been filtered. It should be noted that there is a big difference in the left corner of the images. The edges of the building are amplified in the filtered image.

The setting values for the filters can be adjusted with some knowledge of the expected environment. With high texture regions, the standard deviation for the Gaussian smoothing can be increased to blur the image more, resulting in a reduction in texture. Histogram equalization improves the contrast of an image.Images that are too dark of too washed out will have a histogram with a pixel spread in a very narrow range. The histogram equalization aims to flatten the histogram, resulting in having more contrast across the image. By adjusting the number of bins the histograms shape can be adjusted.

With high texture regions the standard deviation of $MATLAB's$ imgaussfilt function was increased from 0.5 to 2. The histeq function's bins was changed from 64 to 32, to flatten the distribution of the intensities.

Stellenbosch University https://scholar.sun.ac.za

(a)

(b)

Figure 3.4: Pre-filtering of the images with histogram equalisation, Gaussian and median filters. (a) Unfiltered and (b) Filtered.

Stellenbosch University https://scholar.sun.ac.za

Chapter 4

Disparity Map

Disparity maps can be used to determine the 3D location of an object that is in the field of view. For a feature to be specified as an object, it must be visible in both the left- and right-hand images that were captured of the current scene. The disparity map is computed by looking at every pixel in the left-hand image and finding its corresponding point in the right-hand image. Images consist of a large number of pixels. It can be difficult to determine which pixel in the right-hand image is a good match for the pixel in the left-hand image. As was explained in Chapter 2, only partial points within the body of an object were obtained with feature matching. By generating a disparity map of the scene, the full body of all the features within a scene can be detected. The noise removal techniques that are implemented after the template matching followed methods discussed by Corke (2011) and were found to be used in general practice. In this chapter the aim is to determine the location of the best matching pixel in the second image and to determine the disparity. The application of noise-filtering techniques is also implemented and the resulting disparity maps are formulated.

4.1 Overview of the Disparity Map Algorithm

The structure of the disparity map function is discussed in this section. To compute the disparity map of a stereo image pair in this study, the images first needed to be rectified. After rectification, different pre-filters were applied to the rectified images to enhance the accuracy of the resulting disparity map. The algorithm started at the first pixel in the top left corner of the rectified left-hand image and drew a template around the pixel. Another template was drawn around the same pixel coordinate in the rectified right-hand image, but then shifted horizontally on the basis of the disparity range. These templates were tested against one another to determine the similarity score. After the similarity score for the current points had been calculated, it was stored in a vector that was be investigated later. The pixel coordinate in the right-hand

Stellenbosch University https://scholar.sun.ac.za

image was then shifted with one pixel to the right again and the process of similarity calculation and storing was repeated. This process started at the minimum disparity range and continued until the maximum disparity search range had been reached. The point with the best similarity score was taken as the disparity for the pixel coordinate in the left-hand image. To remove noise from the disparity map, the best similarity score needed to be determined and compared to a threshold to be deemed acceptable. The peak ratio was then calculated. The peak ratio was the ratio between the highest and second highest similarity scores. The peak ratio was compared to a threshold to be deemed acceptable. Finally, the sharpness of the peak was investigated. This test involved looking at the points next to the best score and determining the steepness of the angle between them. If the point passed all three tests, the disparity was retained. This process was repeated for all the pixels in the left-hand image. The flow diagram for computing the disparity map is illustrated in Figure 4.1.

Stellenbosch University https://scholar.sun.ac.za

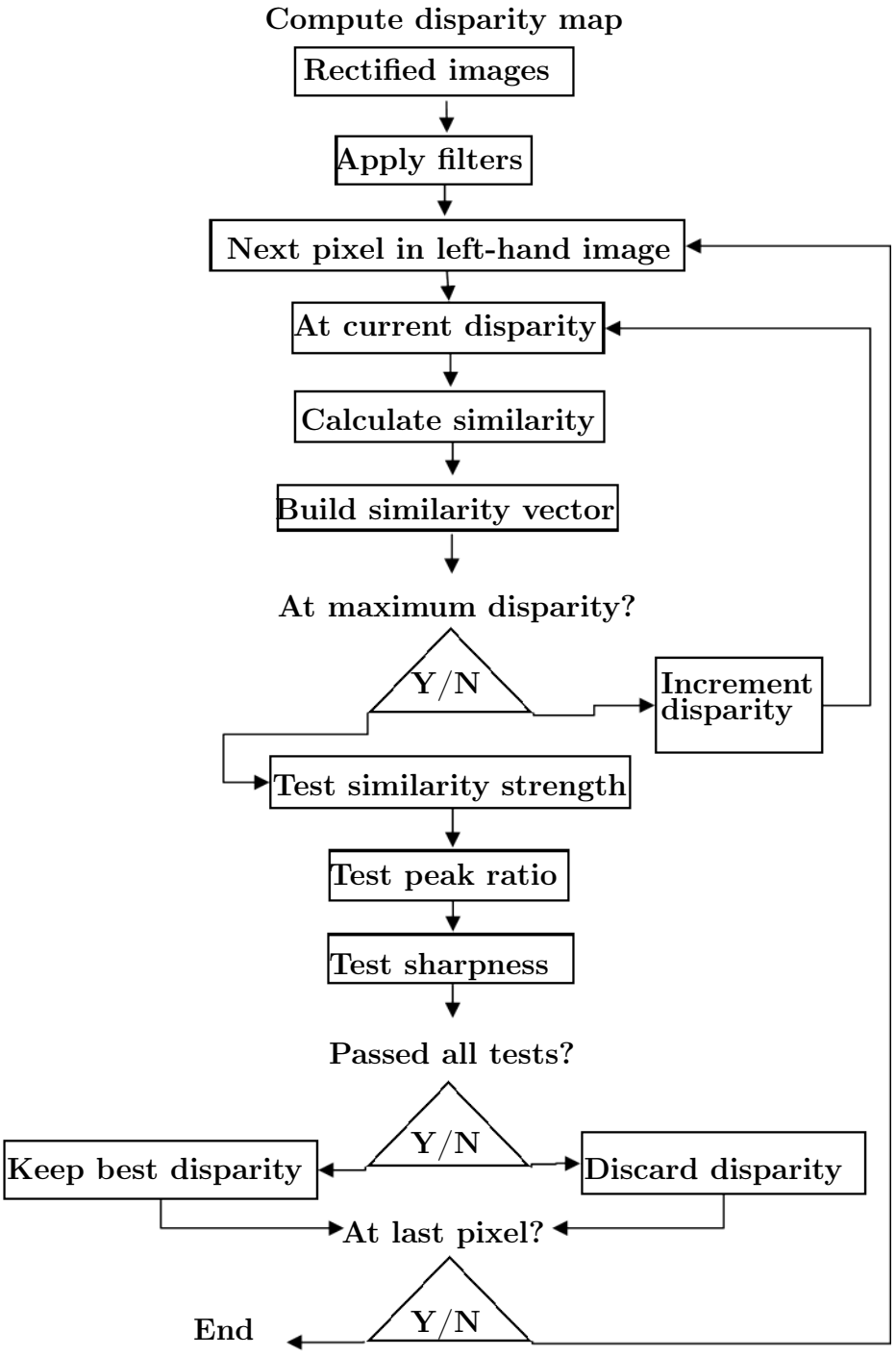


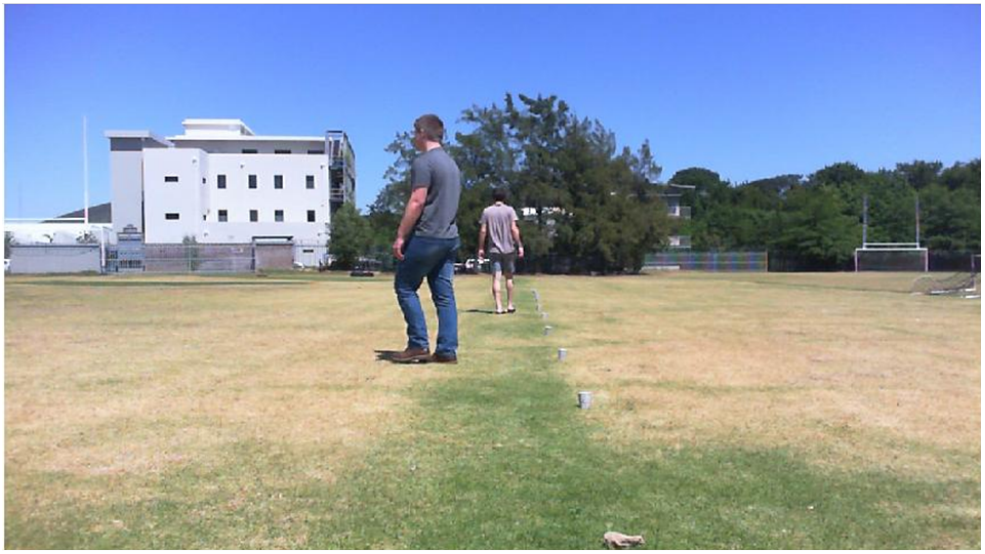
Figure 4.1: Algorithm flow chart for the disparity map computation.

Stellenbosch University https://scholar.sun.ac.za

4.2 Disparity Map from Template Matching

To establish a basis for comparison, the same input images were used to formulate a disparity map. These input images from the two cameras are illustrated in Figure 4.2, where an empty field was used to reduce clutter within the scene, making it easier to evaluate the resulting disparity maps. People were placed as objects that can be identified.

(a)



(b)

Figure 4.2: Original images taken by the cameras. (a) Left-hand image and (b) Right-hand image.

Stellenbosch University https://scholar.sun.ac.za

The disparity map was computed by comparing the block templates with one another, and the resulting disparity map can be seen in Figure 4.3. Comparing Figure 4.3 to Figure 4.2, the two people in the centre of the image can be identified as can the trees and building in the background. The grass can be identified as the floor in the disparity map. The colour of the floor ranges from a light yellow close to the cameras, and turns to a darker blue further away from the cameras. The colour of the pixel is an indication of the disparity value. Bright colours have large disparity values, indicating that the point is closer to the cameras. Darker colours have small disparity values, indicating that the point is far from the camera.

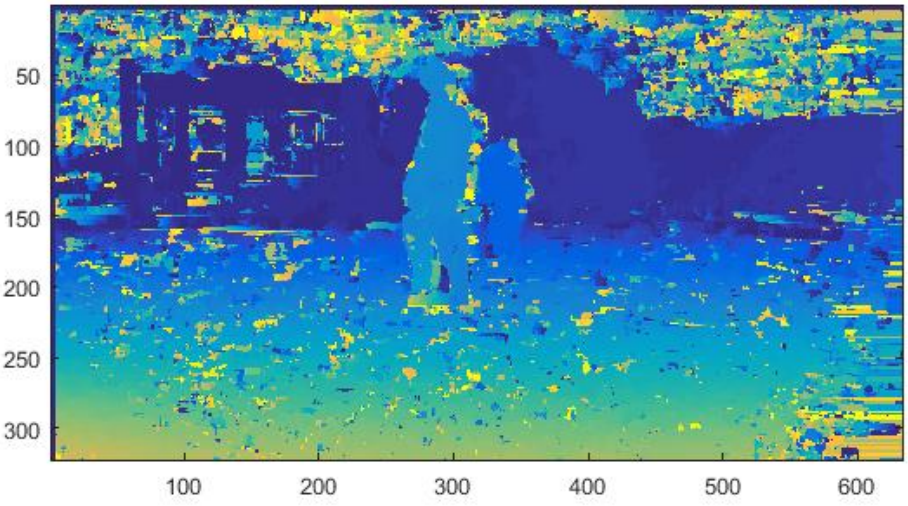


Figure 4.3: Disparity map after template matching.

Taking the floor of the image as a baseline for comparison, noise was identified within the disparity map. The noisy regions within the disparity map were identified as the bright yellow, orange or red pixels within the disparity map. These pixels generally have large disparity values. Noisy regions in the disparity map were easily identified through inspection. If the location where objects are positioned in the scene is known, noise can be seen by studying the disparity map and seeing which pixels do not fit. However, the computer can not accomplish this so easily. It can not identify what is an object and what is not.

To clear up some of the noisy regions, median and Gaussian filters were applied. These two filters remove noisy regions that were small in size. The effect is illustrated in Figure 4.4. It should be noted that most of the noise that was found in the top left quadrant of the image, the building, was removed. The noise still present in the image was grouped, resulting in a more even spread of the noise disparity value.

Stellenbosch University https://scholar.sun.ac.za

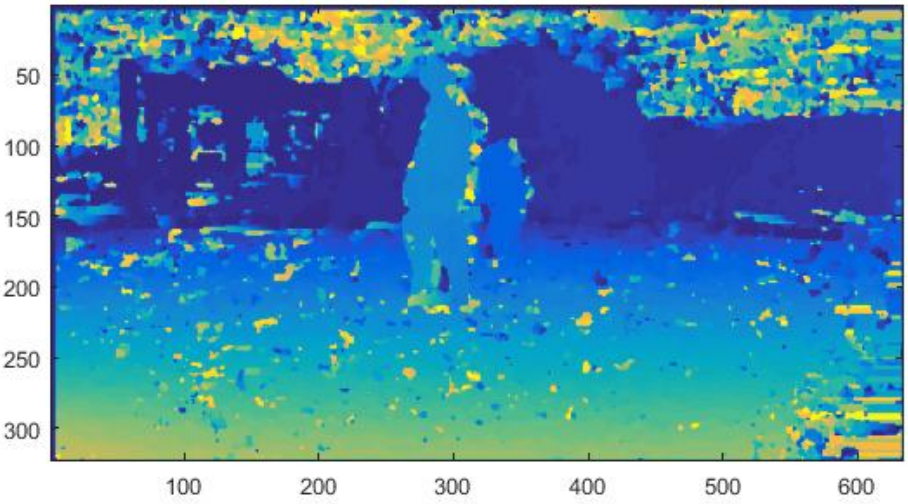


Figure 4.4: Disparity map after median and Gaussian filters were applied.

4.3 The Effect of Peak Threshold

After a clearer disparity map had been obtained, the next noise removal technique could be applied. The next noise removal technique that was implemented was the peak threshold. The peak threshold compared the best similarity score for the pixel to a predefined threshold value. The strength of the similarity score was dependent on the similarity score method that was used. The method used in this study was the ZNCC method. For the perfect match, the desired result was a similarity score of 1. The threshold was set at 0.8, and if the similarity score was more than the threshold, the point was deemed acceptable. Figure 4.5 illustrates the effect of the peak threshold. The threshold removed large portions of the noise. The most noticeable noise that was removed was within the floor region, indicating that the matches found within the floor region were weak matches. Large portions of the floor were also removed. If the grass regions in Figure 4.2 are compared with one another, it can be seen that there is a clear difference in colour between the images. It can also be noted that the colour of the grass field is not the same across the field. The grass patches dramatically change colour throughout the field. These changes in colour affected the matches that were found. Disparity values in the top half of the disparity map were also removed. Most of the removed values are based in the sky, where there is a clear transition from light blue to darker. The change in colour also affected the strength of the matches in a manner similar to that involving the grass.

The further away an object is from the cameras, the smaller it appears in an image. Keeping this concept in mind, the similarity score calculation is revisited. The similarity score compares a template in the one image with a

Stellenbosch University https://scholar.sun.ac.za

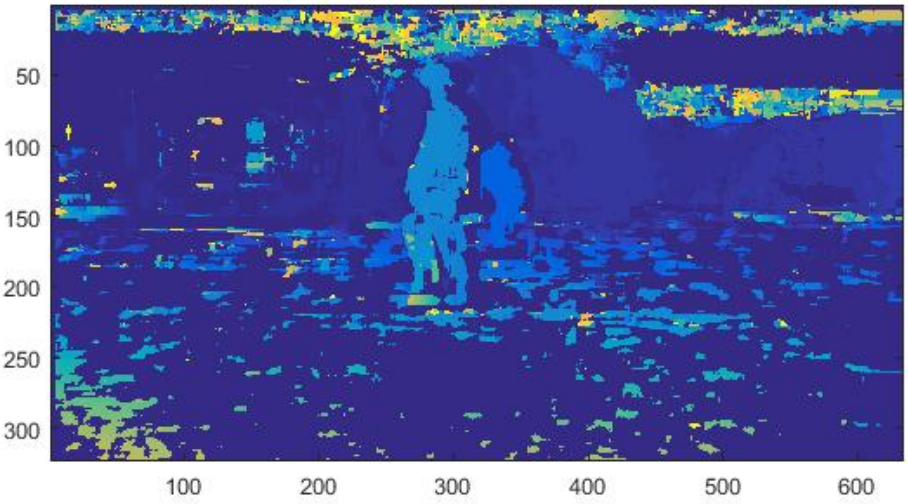


Figure 4.5: Disparity map after peak threshold.

template in the other. The effect that an object has on the template is reduced as the size of the object decreases. Taking a template with a small object in it, and comparing it to another template with the same small object in it, may not always show a good result due to the reduced effect of the object within the template. Thus comparing templates of faraway objects can result in a weak similarity score. The effect of the small object occurred in the top left quadrant of the disparity map. Most of the window features were removed due to this effect. The similarity score for small features was affected by the template size and the size of the object. The edges of the body that was removed are also due to a similar effect, where the pixels outside the body played a role in the similarity calculation.

4.4 The Effect of Peak Ratio

The peak threshold removed large portions of noise from the disparity map, referring to high disparity values located in the sky region, but there are still regions containing noise. The pixels within the disparity map were only tested to determine if their similarity score was strong enough. There was still uncertainty as to whether the match that was found was the correct one. The peak ratio was then implemented to determine if the match stood out from the other possible matches for that pixel. By computing the peak ratio between the two best matches a level of uncertainty was estimated for the pixel. Similarity scores that were found to be too close to one another in strength implied that both these points were possible matches for the relevant pixel, therefore uncertainty existed. The peak ratio of these two similarities was calculated

Stellenbosch University https://scholar.sun.ac.za

and compared to a peak ratio threshold. The peak ratio needed to be below the threshold to be deemed acceptable. Figure 4.6 illustrates the effect of the peak ratio. The threshold was set at 0.9. The peak ratio aims to remove points that were found with multiple strong peaks. The effect is clearly visible in the sky. Almost all of the pixels in the disparity map were removed. The similarity scores of the sky were close to one another due to its monotonous texture, implying multiple possibilities of good matches. The same effect can be seen in portions of the floor. It can also be noted that the window of the building was removed. The window could have been removed due to spatial aliasing.

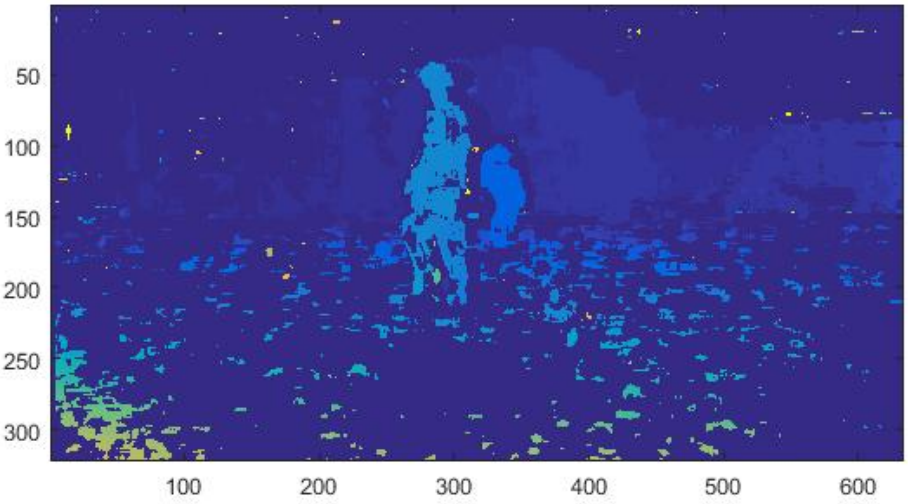


Figure 4.6: Disparity map with a peak ratio of 0.9.

Increasing the threshold value to 0.99 made the peak ratio more tolerant to noise. The disparity map with a peak ratio threshold can be seen in Figure 4.7. Less noise was removed compared to the threshold of 0.9. With a decrease in the threshold, the amount of noise left in the disparity map was reduced, but disparity pixels within the body of an object were lost. The effect can be seen by comparing Figure 4.7 to Figure 4.6. Precautions need to be taken when implementing the peak ratio. The user needs to determine what is more important. The current focus was noise removal, but the main goal was object detection. Reducing the threshold even more would therefore be undesirable. The residual noise that was left after the peak ratio comparison indicated that the similarity scores for the peak ratio of that pixel was acceptable.

Stellenbosch University https://scholar.sun.ac.za

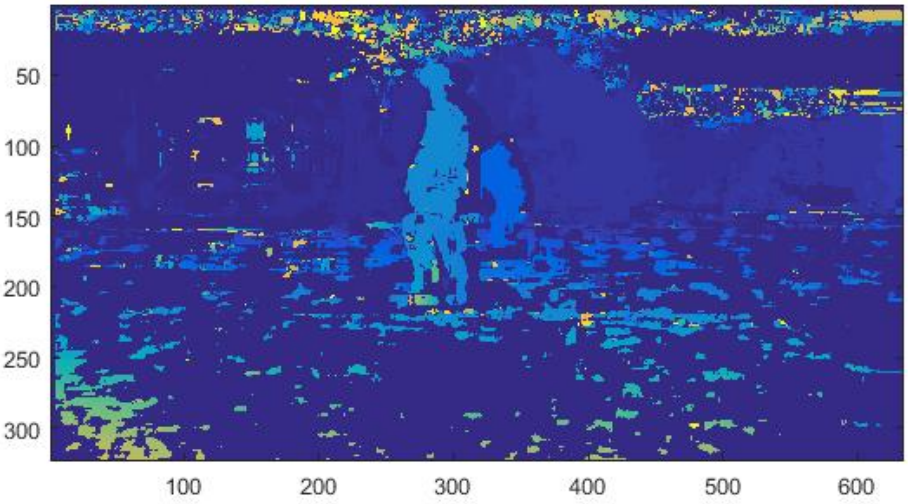


Figure 4.7: Disparity map with a peak ratio of 0.99.

4.5 The Effect of Peak Sharpness

The final noise removal technique that was applied to the disparity map of the image pair was the testing of the peak sharpness between similarity scores. The peak sharpness was determined by indirectly looking at the angle between the highest similarity score and its adjacent similarity scores. The calculated sharpness had to be larger than a threshold value. The resulting disparity map after the peak sharpness test can be seen in Figure 4.8. If Figure 4.8 is compared with its predecessor, Figure 4.6, it can be clearly seen that most of the residual noise was removed. By testing the adjacent points, low texture areas were detected. The pixels identified with low peak sharpness were removed. In Figure 4.8 the threshold was 0.3. Although most of the noise was removed, portions of the body were also removed. The inner body of the object lost most of its disparity pixels. Using the person closest to the camera as an example, most pixels were removed at the upper thigh and shoulder, where there were low textures. If the pixel loss of the person closest to the camera is compared to that of the person further away, it can be seen that there was much less pixel loss. The shirt of the person at the back has a lot more texture to it.

Increasing the threshold reduced the number of pixels deemed acceptable. This reduced the noise in the resulting disparity image even more, but the loss of object body was greater. The disparity map with a threshold of 0.6 can be seen in Figure 4.9. Although the resulting image had less noise, the body loss was too much. A more tolerable threshold for peak sharpness is preferred.

Implementing noise removal techniques cleared up unwanted pixels on the disparity map. Noise on the disparity map was removed as the noise removal functions were implemented, but at the expense of clarity within the body of

Stellenbosch University https://scholar.sun.ac.za

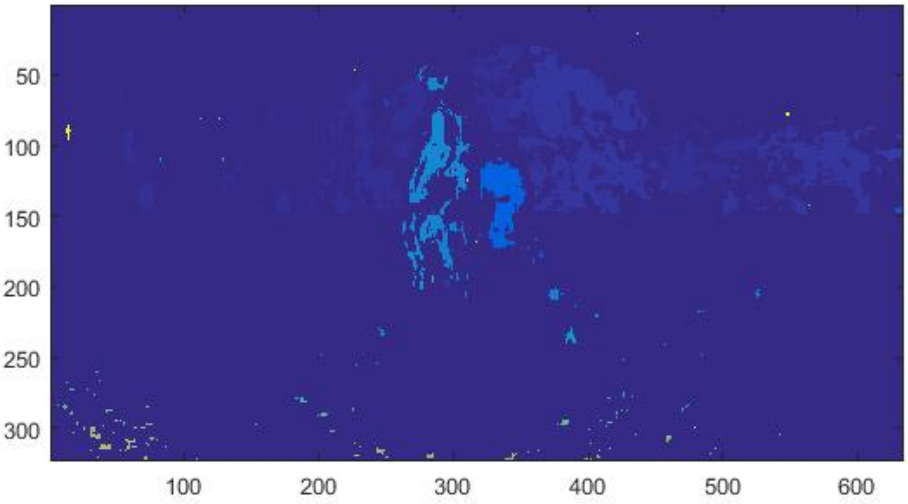


Figure 4.8: Disparity map with a peak sharpness of 0.3.

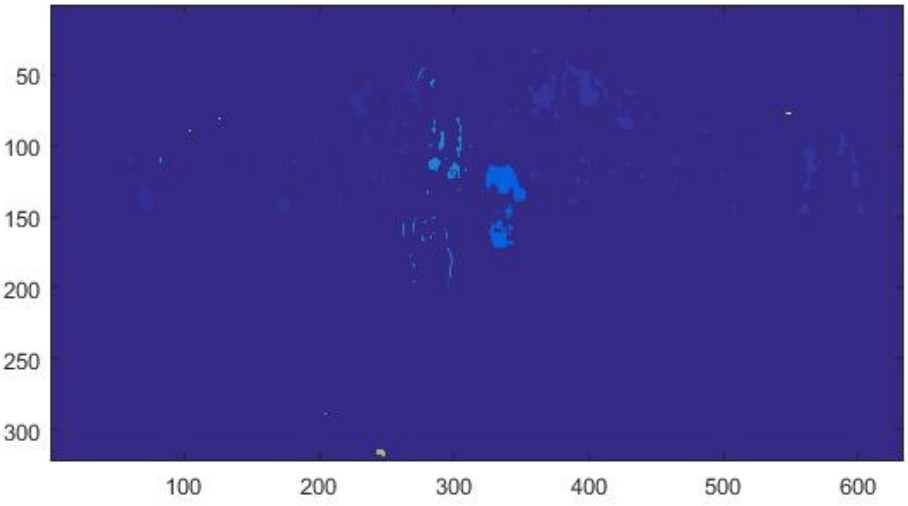


Figure 4.9: Disparity map with a peak sharpness of 0.6.

the detected object. The user must take this into consideration when implementing these techniques. The thresholds for these techniques depend on the main goal.

Stellenbosch University https://scholar.sun.ac.za

Chapter 5

Temporal Noise Removal

Noise will always be a problem with stereo vision applications. After the noise removal methods from Chapter 4 had been applied, there was still noise present in the resulting disparity maps. The noise can be in the form of false positives or incorrect disparity values commonly found at the edges of detected object. With the main focus being object detection and noise filtration, it was desirable to remove as much noise as possible without losing too much of the body of the object. The information obtained and used from the cameras thus far was the extrinsic parameters of the cameras and the images captured from the stereo vision set-up. The images supplied the data to detect possible objects in the path, while the extrinsic parameters helped to determine the 3D locations of these detected objects. Using the same information to cancel out noise would be difficult. To further evaluate the results, extra information was required. The extra information that could be implemented was the use of time. The possibility of using time as a variable to reduce the noise in the disparity map was investigated. This chapter discusses the theory of noise removal using time as a variable and the implementation of this method for fast and slow frame rates.

5.1 Temporal Frame Analysis Algorithm

The temporal frame analysis aims to remove noise from a disparity map by comparing the disparity values of the current time frame against those of the previous time frame. The inputs for this function were the disparity maps of the current and previous time frames, as well as the testing locations of where objects were located in the previous time frame. The algorithm started by looping through all the disparity values in the current time frame. The threshold for every pixel in the disparity map was calculated. This threshold was an indication of the disparity range in which a pixel in the previous frame was allowed to be. The initial comparison was done on the disparity pixels at the same coordinates. The disparity value was deemed to be correct if the dis-

Stellenbosch University https://scholar.sun.ac.za

parity of the previous frame was within the bounds of the threshold established by the disparity in the current frame. If the initial test fails, an alternative test is done. The second test involved testing the disparity of the current frame against the disparity of the previous frame, but at different coordinates. The coordinates were dependent on whether there had been movement in that section of the original rectified images. The same threshold test was done at the new location. The disparity is deemed to be acceptable if it passes the threshold test at the new location. If both of these test fail, the disparity at that pixel coordinate in the current frame is discarded. It should be noted that the previous disparity map used in the comparison was the disparity map before the previous frame went through this noise removal step. The reason for this was that, if noise is removed from the disparity map, some loss in the body of the detected objects occurs. Thus, by continuously removing small fragments of the object's body, the object may be removed entirely. Noise will be removed, but there will be no objects displayed in the disparity map. The program flow chart is illustrated in Figure 5.1.

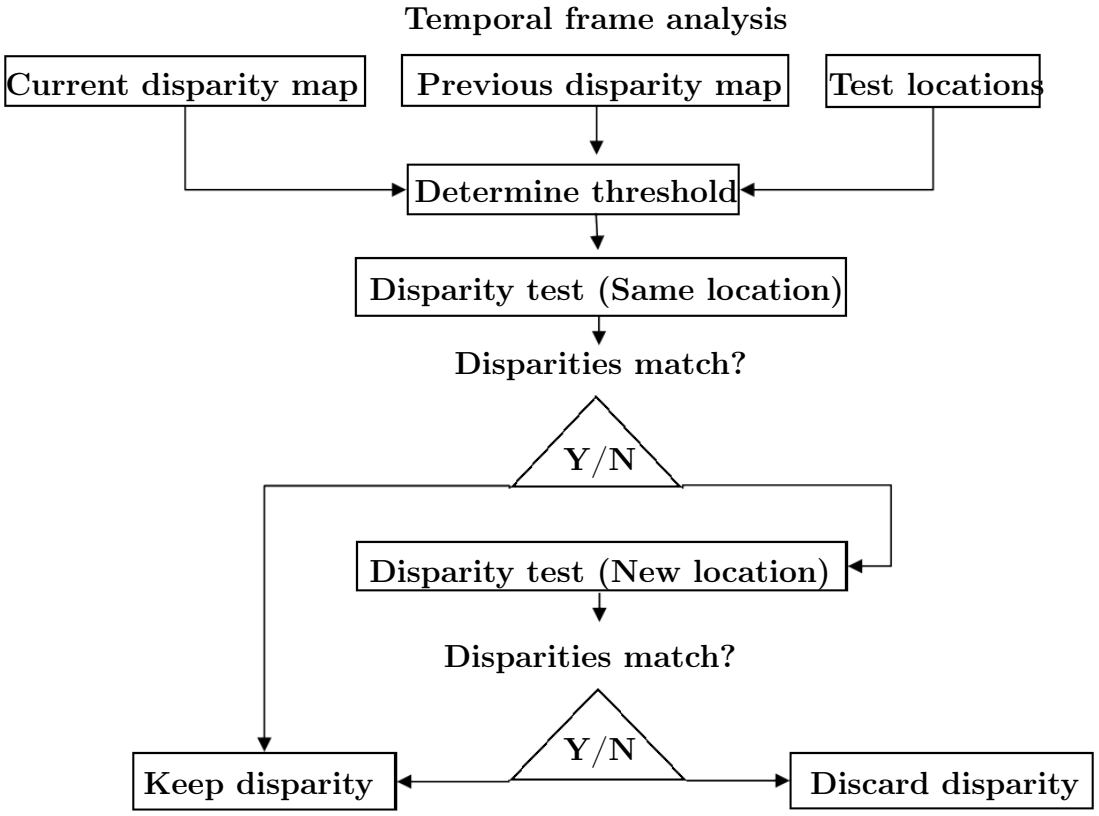


Figure 5.1: Program flow chart for the temporal frame analysis.

Stellenbosch University https://scholar.sun.ac.za

5.2 Theory of Temporal Noise Removal

Computing the disparity map of an image pair will deliver a set result. Repeatedly computing the disparity map of that same image pair will always result in the exact same disparity map. In theory, this approach is similar to having a stationary scene, but this is not exactly true. A video recording of a scene reveals slight changes in the input images as time progresses. The changes occur even if nothing was moved within the scene.

An inspection of the disparity maps generated from the video file showed that, although nothing had changed within the scene, the disparity maps were not the same. Small spots of noise appeared at different locations as time progressed. The different locations for the noise were investigated.

Taking a video recording inside a building removes the effect of nature, like wind forcing movement in the grass and trees. The recorded frames are now only affected by the lighting of the scene and the camera properties. The lighting of the scene affects the contrast and brightness values of the camera properties. The contrast and brightness affect the value that is stored in each pixel, resulting in different template values calculated with the similarity score calculation. In this study, when the camera properties were opened and the contrast and brightness parameters were set to manual instead of automatic, the parameters stayed fixed for the duration of the recording. After the parameters had been set to manual, there were still changes between the images as time progressed. This indicated that lighting played a large role in computing the disparity maps.

Depending on the scene, nature can also play a role. A video recording of a grass field was done to determine the effect of small moving particles. The biggest identifiable problems that caused noise were due to spatial aliasing and low texture regions. Trees were identified to show noise where the leaves were subjected to spatial aliasing. Even if the peak ratio is acceptable, it does not imply that the match is the correct one. Grass patches showed large disparity values that were incorrect. The disparity map of a plain grass field should have a low disparity value at the top of the matrix and steadily increase towards the bottom of the matrix, as the section of grass moves closer to the cameras. It was found that there were small regions in the matrix that had far larger values than their neighbours. These results can be due to low texture regions finding an unwanted match.

The theory is to calculate the disparity map for the current time frame D_n and compare it to its predecessor D_{n-1}. Where there are differences, it is likely that there was noise. If a difference was detected, it was compared to a threshold to be deemed to be noise.

Stellenbosch University https://scholar.sun.ac.za

5.3 Temporal Noise Removal Methodology

The focus of implementing the time steps is to reduce the noise within disparity maps. This process works on the principle of comparing one disparity map with another by looking at the difference in their pixel values. The initial comparison was done by subtracting the pixels at the same coordinates from one another.

The first problem that presented itself was the time between these disparity maps. Having a large time jump in a stationary scene may not cause any problems, because the matching will be done at the same coordinates, but what if there are moving objects in the scene? The moving objects will be eliminated from the disparity map if too much time has passed from one disparity map to another when the same coordinates in the disparity maps are compared. This indicates that, if the aim is real-time applications, the frame rate of the recordings as well as computational power of the processor plays a major role.

Depending on the application, the temporal noise removal can be implemented in disparity maps that have a high frame rate or a low frame rate. The general process for both of these situations is similar. The difference between the disparity maps is computed and then compared with a threshold.

The threshold ϕ can either be constant or a function of the disparity of the current pixel. Depending on the frame rate, two methods were proposed in the study for the implementation of the temporal noise removal. With a fast frame rate, a pixel of the first disparity map, $D_n[X, Y]$, was compared with a pixel in the previous disparity map, $D_{n-1}[X, Y]$, therefore at the same coordinates, as shown in Equation (5.1).

With slower frame rates, the location of the object in the previous frame must be determined. This is shown in Equation (5.2). The coordinates of the previouse location of the matching pixel $[X_m, Y_m]$ were used to determine the difference in disparity. Figure 5.2 illustrates the structure of the methodology.

$$Difference = D_n[X, Y] - D_{n-1}[X, Y]. \tag{5.1}$$

$$Difference = D_n[X, Y] - D_{n-1}[X_m, Y_m]. \tag{5.2}$$

5.3.1 Constant Threshold

The disparity of an object moving towards or away from the cameras will increase or decrease. This implies that it is not appropriate to just test for the exact same disparity value. Testing only for the exact same value means that the objects that move will be classified as noise and eliminated from the disparity map.

The first threshold that was investigated in this study was the constant threshold. The constant threshold is a vector ranging from $-\phi$ to ϕ. The difference computed must be within this region to be deemed acceptable.

Stellenbosch University https://scholar.sun.ac.za

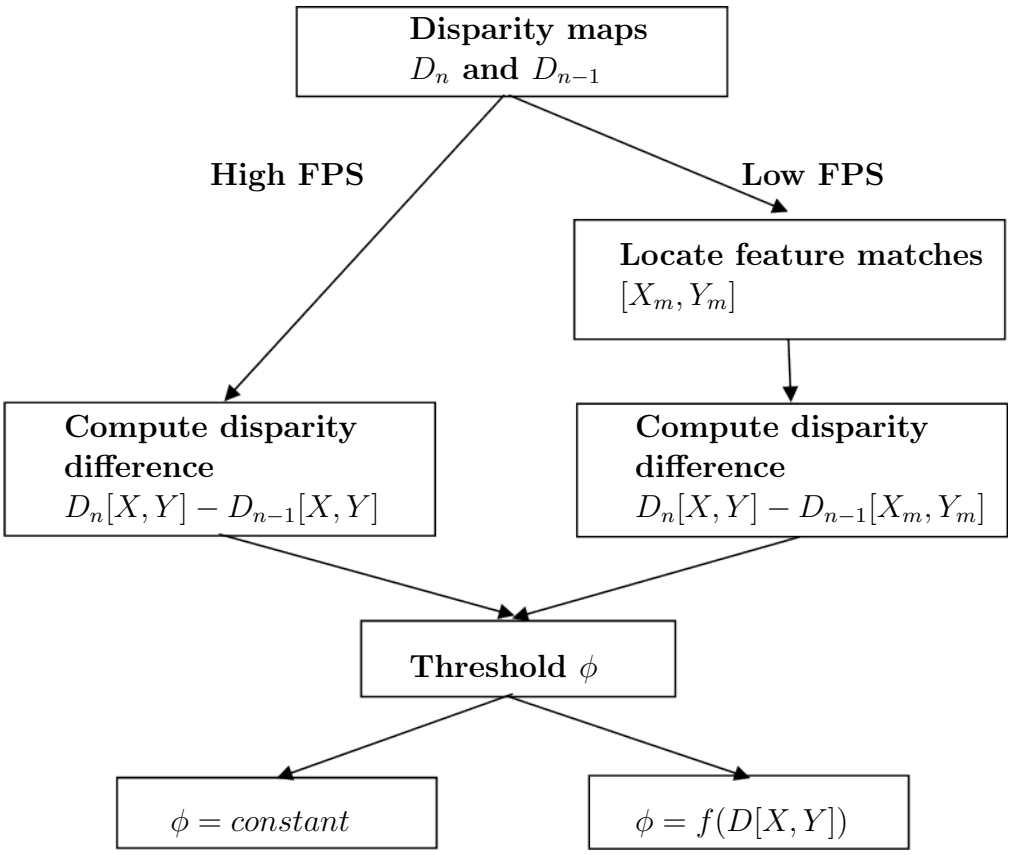


Figure 5.2: Structure of the temporal noise removal methodology.

The constant threshold is explained with the help of Figure 5.3, using a ϕ value of 3. Both points in the second column resulted in an acceptable match. The points in the left-hand column were not deemed acceptable and therefore the disparity map at those two coordinates was set to zero.

The current time frame disparity map that is used for illustration is the one from Figure 4.8. Figure 5.4 illustrates the effect of using a constant threshold value. Like the previous noise removal techniques, the temporal noise removal removed some of the noise within the image, but at the expense of loss in body. The left leg of the person closest to the camera was nearly removed.

5.3.2 Variable Threshold

The distance of an object from the camera affects the disparity value. Objects that are far away have small disparity values. Objects that are close to the cameras have large values. The calculated distance from the cameras to the object is a function of the disparity. With the current camera set-up, the maximum detectable distance was 212 m. Dividing the distance by a disparity gives us the distance for that disparity, as is shown in Equation (2.2). As the disparity value increases, the distance decreases, as is illustrated in Figure 5.5.

Stellenbosch University https://scholar.sun.ac.za

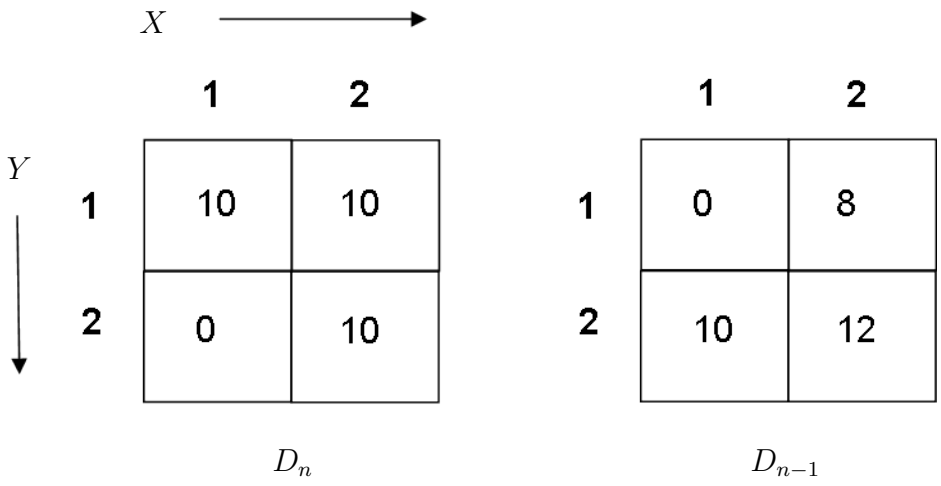


Figure 5.3: The current D_n and previous D_{n-1} disparity map with sample disparity values.

Figure 5.4: Temporal noise removal with constant threshold.

The solid line indicates the distance towards the camera for the given disparity range. Investigating small disparity values, shows that through incrementing the disparity value can show a large decrease in the distance. Incrementing an already large disparity value reveals a small decrease in the distance. Therefore, testing for a constant threshold value will not always be the best idea.

Assume a point with a disparity value of 4. Using the same constant threshold of 3 for this example, the allowable matching disparities can be in the range of 1 to 7. These disparity values correspond to a distance of 212 m and 30 m and are illustrated in Figure 5.6. That jump in distance is unacceptable.

Stellenbosch University https://scholar.sun.ac.za

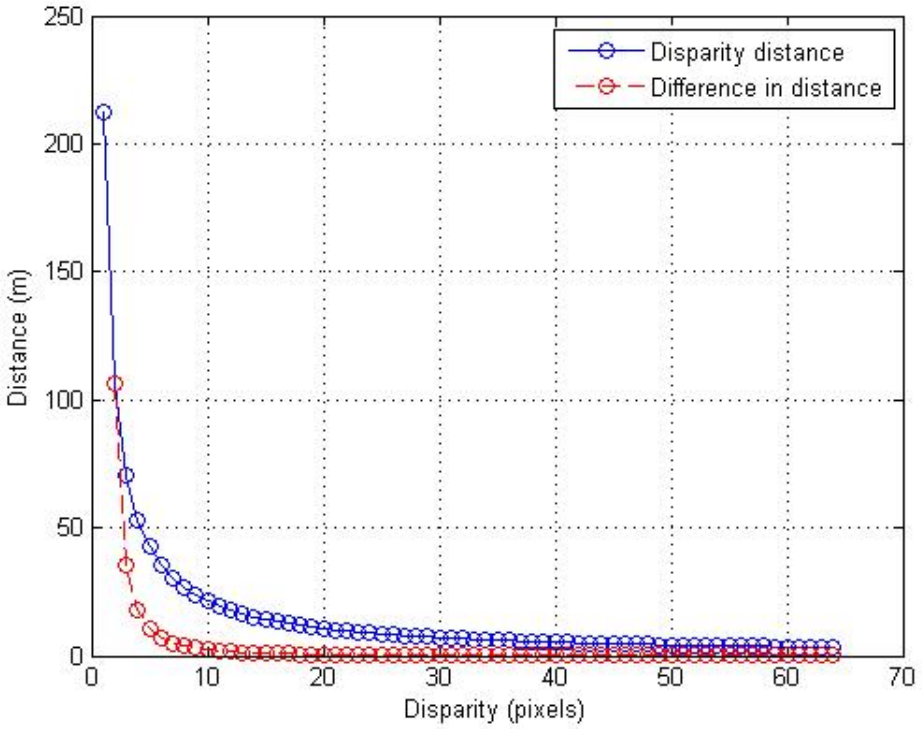


Figure 5.5: The distance away from the camera with respect to disparity is indicated with a solid line. The distance increment difference between the current disparity and the previous disparity is indicated with a dashed line.

The effect becomes much smaller with higher disparities. Taking a disparity value of 60, the distance difference between disparity 57 and 63 is only 0.4 m. This leads to the implementation of a threshold function that varies with the disparity of the current pixel under investigation.

The threshold as a function of the disparity of the current frame's pixel is shown in Equation (5.3). Depending on the disparity at a given point, a threshold can be determined. The disparity value of the matching coordinate in the previous frame must fall within the threshold to be deemed acceptable. These threshold ranges are proportional to how the disparity influences the distance jumped between disparities. The threshold calculation can be seen in Equation (5.4) and the list of disparity values affecting the ranges can be seen in Table 5.1. The threshold ranges from the minimum disparity ϕ_{dmin} to the maximum disparity ϕ_{dmax}. The disparity for the given point must fall between the maximum and minimum disparity value. The threshold is calculated by subtracting the current disparity from the maximum and minimum value of that range. Table 5.1 shows the total distance between the minimum and maximum disparity and the average distance per disparity for that range. The threshold for the disparities close to the cameras seems large, but is necessary.

Stellenbosch University https://scholar.sun.ac.za

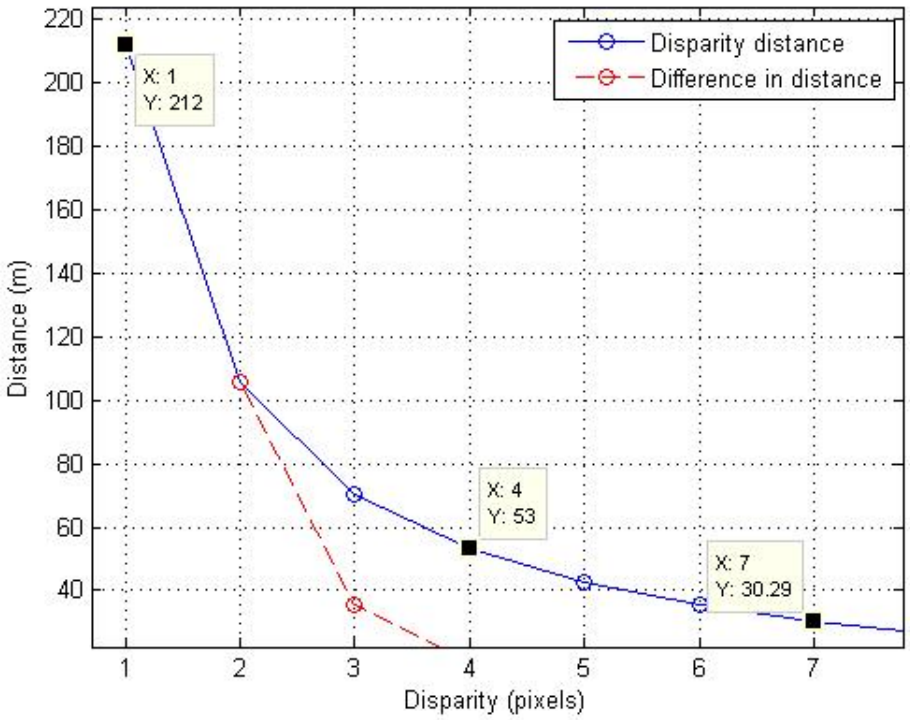


Figure 5.6: The distance difference between a disparity of 4 and the acceptable threshold boundaries with a fixed threshold of 3.

Objects that are close to the cameras change position much faster than objects that are far away. A fast-moving object can easily travel 1 m in a short time period. It should be noted that these thresholds do not apply only to a fast frame rate but to the slower frame rates as well. Having a smaller threshold range causes the slower frame rate analysis to remove objects that move at a fast pace towards the camera. The average distance between disparities increases as the threshold range decreases. The small increments in distance between disparities can be seen in Figure 5.7 which zooms in on the disparity distance difference in Figure 5.5. The small increments in the distance difference can be clearly seen.

$$\phi = f(D[X, Y]). \tag{5.3}$$

$$\phi = [\phi_{dmin} - D[X, Y], \phi_{dmax} - D[X, Y]]. \tag{5.4}$$

Figure 5.8 illustrates the effect of temporal noise removal using a scaling threshold. The scaling threshold removed some noise in the bottom left corner of the disparity map, but the trees in the right-hand background were removed. The threshold scaling had a larger effect on smaller disparity values, i.e. on disparity values farther away from the cameras.

Stellenbosch University https://scholar.sun.ac.za

Table 5.1: Disparity ranges for thresholds and the distance difference and average for the range

Disparity ranges and distance difference			
ϕ_{dmin}	ϕ_{dmax}	Distance difference (m)	Average distance per disparity
45	63	1.4	0.0778
37	45	1	0.125
32	37	0.9	0.18
29	32	0.68	0.227
26	29	0.85	0.28
20	26	2.4	0.4
14	20	4.5	0.75
10	14	6	1.5
5	10	21	4.2
2	5	63	21
1	2	106	106

5.4 Motion Compensation Algorithm

Noise can be identified by subtracting the disparity values at the same location in two disparity maps from one another, as the noise acts like a movement not located at the same pixel position in each frame. However, a problem occurs when an actual object is moving. The pixels related to the object is classified as noise as well. To resolve this, motion compensation was implemented.

To determine the location from which the object in the current frame came, an motion compensation algorithm was implemented in the study and formed part of the temporal noise removal algorithm. The argument was made that all that was needed to determine where the object came from was the location in the previous image. Feature matching was therefore introduced to determine the location of the points.

The motion compensation program started by taking the current and previously rectified images as inputs. SURF feature detection was used to determine all the feature points in both these images, followed by finding the matches between these features. The rectified images were broken up into small cubes, similar to a mesh grid. The cubes were then taken as individual elements of a matrix. The area inside the cube was investigated to determine if there were any matching features from the current frame within its boundaries. The average of these features was calculated, resulting in an average coordinate of matched feature points for the cube. The matching features of the points in the previous image were also calculated. The combination of these two averages provided an indication of the average direction in which the feature moved within in the cube. The average calculations were done for every cube within

Stellenbosch University https://scholar.sun.ac.za

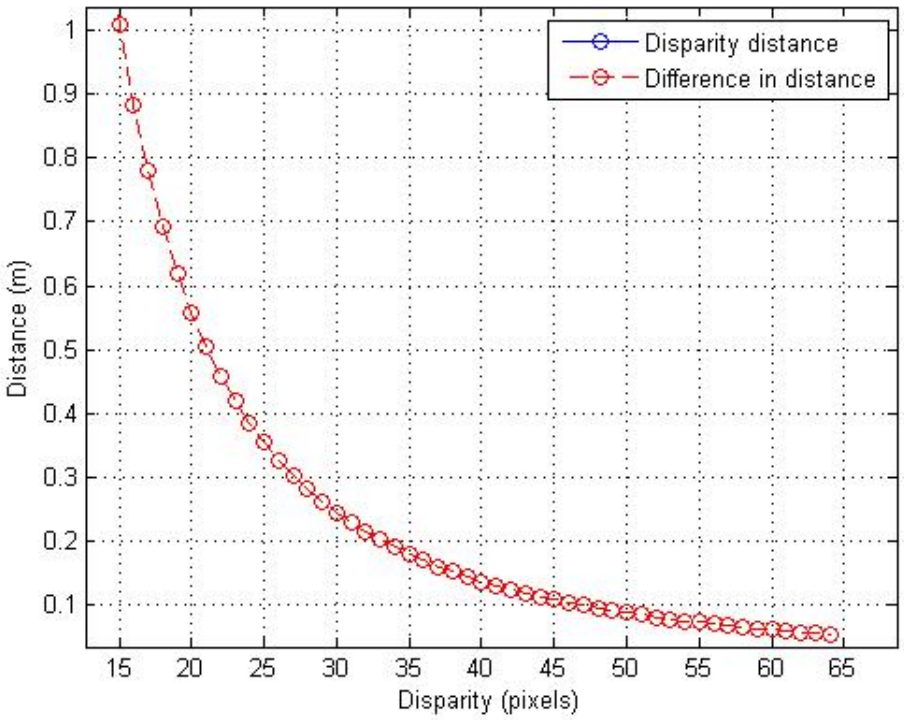


Figure 5.7: A zoomed-in section of the disparity distance difference in Figure 5.5.

the image. With the average for each cube, the existence of outliers could be determined. The outliers were removed by comparing the pixel distance between the feature pair with the largest distance between them to the average of the cube it belonged to. The coordinates were removed if they were classified as outliers and the average was recalculated. The algorithm flow chart for the motion compensation is illustrated in Figure 5.9.

Stellenbosch University https://scholar.sun.ac.za

Figure 5.8: Temporal noise removal with variable threshold.

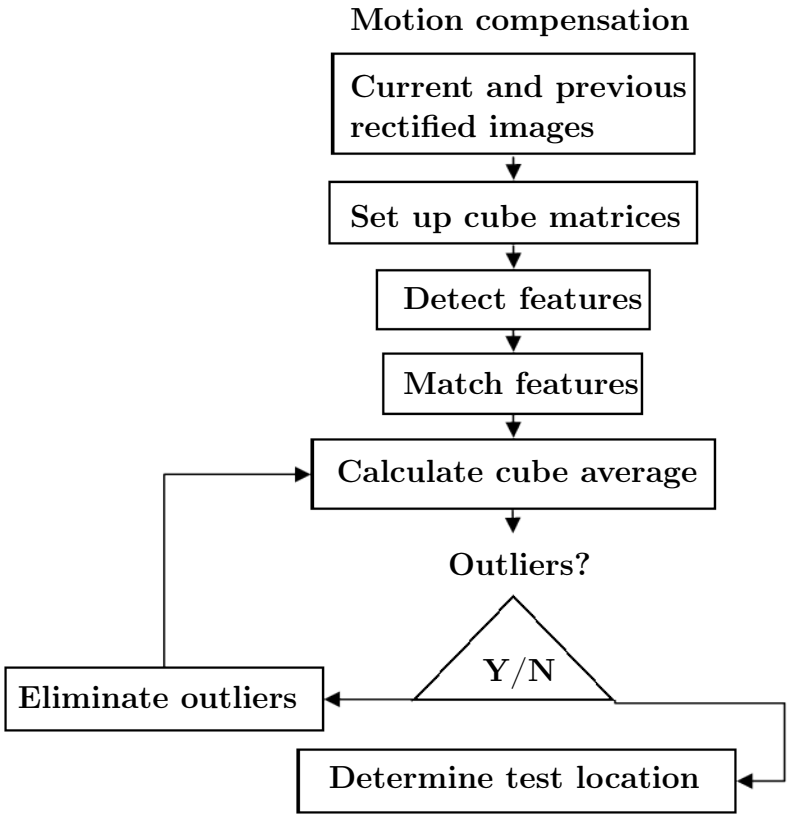


Figure 5.9: Algorithm flow chart for motion compensation.

Stellenbosch University https://scholar.sun.ac.za

5.5 Low Frame Rate Noise Removal

With a fast frame rate the temporal noise removal can be implemented by testing a pixel in the current disparity map with the same pixel in the previous disparity map. The drawback is that the edges of the moving object will be removed, because its not in the same location. With a low frame rate there will be large jumps in the location of the detected object. The jumps can lead to large portions of the object being removed from the disparity map, or - worst case - the entire object being removed. The object in the current frame needed to be identified as noise or not. To accomplish this, a feature detection step was implemented. The frame rate for the cameras was set at 30 frames per second. With the recording of the video frames, the time it takes to store the video and grab the next frame caused the video to not record at the specified frame rate. It was determined that the videos were recorded at 7.6 frames per second and is generally deemed as a slow frame rate.

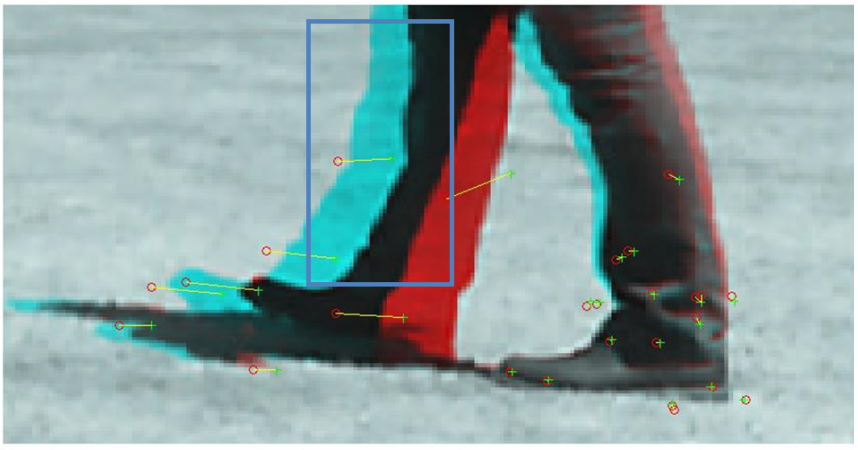
5.5.1 Matching Features

With feature detection it is possible to detect features that are in both the current and the previous frame. Feature detection was done on the rectified images of the left-hand camera. The computation of the disparity map used rectified images as input. By using the rectified images, the coordinate system of the feature detection was set on the same axis as the disparity map. Thus the feature locations were at the same position on the disparity map as in the rectified image. The rectified images also kept the size of the output image from feature detection the same as on the disparity map. For example, the dimensions of the image on which the features were located, 300×600, were the same as the disparity map's dimensions, 300×600. The disparity map was based on the left-hand image. When the disparity map was computed, it took a pixel in the left-hand image and searched for a match in the right-hand image. The disparity of the match was then stored at the location of the pixel in the left-hand image. Using the rectified left-hand image as an input ensures that the matching feature found will be at the same location as on the disparity map, stating that point $[X, Y]$ on the feature detection image is the same as point $[X, Y]$ on the disparity map.

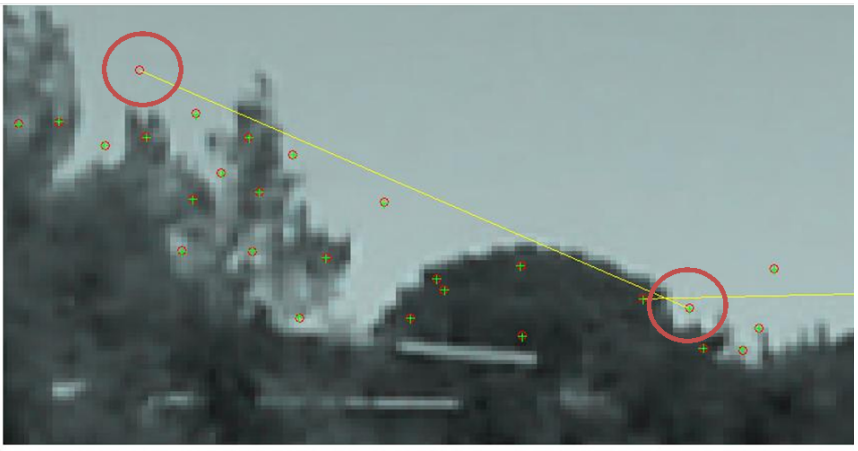
Matching the features of the current rectified left-hand image with the previous one, it is possible to see where the current features were in the previous frame. This provided us an indication of the direction the object was moving. Feature matching can not perfectly match all the features that can be identified with our own eyes. It only gives sparse points across the image, for example only some points on a leg will be deemed to be good matches, as illustrated in Figure 5.10(a). The rectangle drawn over the leg shows that only two points were found as matches in the entire region of the front leg. The matching point in the current figure was denoted with a circle and the previous match with

Stellenbosch University https://scholar.sun.ac.za

a plus sign. Not all the matches that were given were good matches. Some matching points can be across the image, shown in Figure 5.10(b). The circled points show a matching pair that was an incorrect match.



(a)



(b)

Figure 5.10: Matched features in the current and previous frame. (a) Only some parts of the leg show matching points, and (b) Incorrect matches that were found.

5.5.2 Sparse Matched Features

With the location of the point in the previous frame, the current disparity map was compared to the previous disparity map at these locations. It should be noted that this may be problematic. The first problem was that only some of the features were detected. Using only the detected features was not acceptable. Having some features of an object detected and testing only

Stellenbosch University https://scholar.sun.ac.za

the disparity values for those points, for example, resulted in the disparity map displaying only portions of the object. This is illustrated in Figure 5.11. Figures (a) to (c) show the matched features found between the previous and current frames. Figures (d) and (e) show the disparity map for the previous and current frame. Using only the feature matched to validate the disparity map of the current frame resulted in Figure (f). Only the matched feature had the correct disparity. The disparity maps of the rest of the objects were compared to points with no disparity value, resulting in the removal of that disparity.

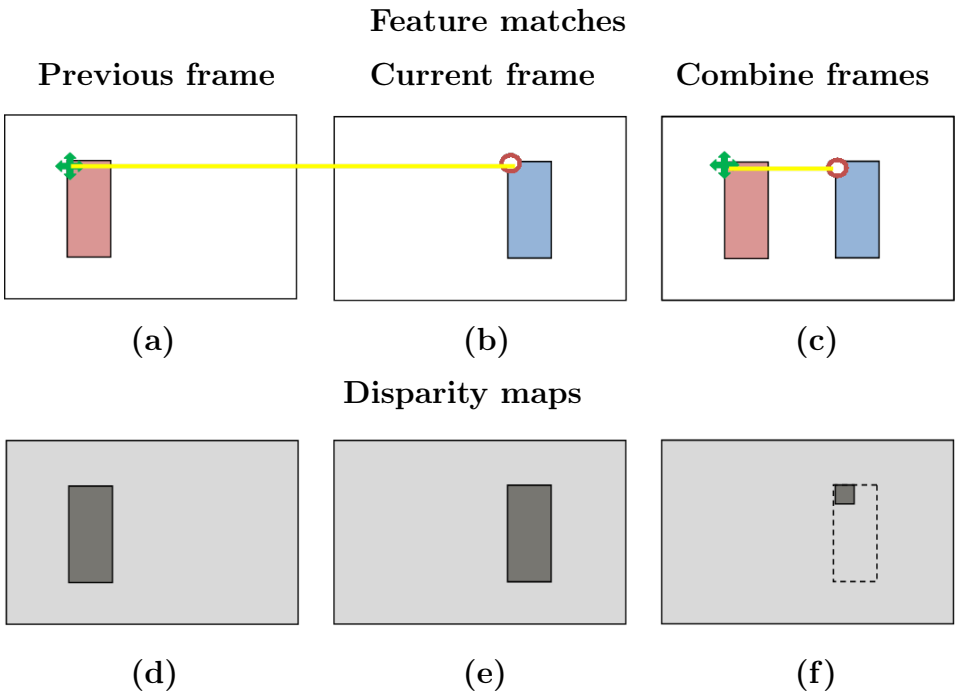


Figure 5.11: The detected matching features of an object and the resulting disparity map. (a) The previous frame with matching point; (b) The current frame with matching point; (c) The stereo anaglyph of the previous and current frame; (d) The disparity map of the previous frame;(e) The disparity map of the current frame and (f) The resulting disparity map.

The sparse matches were solved by splitting up the image into small cubes. The cubes served as a block in a matrix. By taking all the coordinates of the feature matches found in a cube and adding them together, and then dividing the sum by the number of entries in the cube, the average position for the entries were calculated. The calculation was done separately for the current frame and previous frame. The average of the feature matches in the current frame was calculated. The feature matches of the same points that were used in the average calculation for the current frame were used in the average calculation for the previous frame. If matches of the previous frame

Stellenbosch University https://scholar.sun.ac.za

fell within the cube, but the current frame's match did not, the point was not used for that cube. The average position was then stored as a coordinate in the matrix, at the location of the cube's location with respect to the image. For example, the average of the cube at the top left corner of an image was stored at location $[1, 1]$ in the matrix in the form of an X and Y coordinate. This was done for all the matched feature points found in the current frame, even if their matches for the previous frame were outside the cube. This resulted in two matrices, M_c and M_p. The assumption was made that all the pixels that fell within the cube of these matrix elements moved from M_p to M_c. Instead of only testing the matched feature points, all the pixels in the disparity map were then tested, based on the matrix elements. The pixels of the disparity map that fell within the area of the cube were not compared to the coordinates given by the element in the previous matrix M_p. This resulted in all the pixels being compared to the same pixel. The matrices M_c and M_p served as a direction vector for the pixels that fell within the cube. The direction is determined in Equation (5.5) where M_d is an indication of the direction where the match of the previous features can be found. The equation is shown in more detail in Equation (5.6), where its illustrated per coordinate. The direction was calculated by subtracting M_p from M_c. The magnitude or length of the direction vector was the distance from the coordinates M_p to M_c. The location of where the current disparity pixel searched for a match is shown in Equation (5.7), where the disparity location in the previous frame is the difference between the disparity location of the current frame and the direction from where the previous frame's match can be found.

$$M_d = M_c - M_p. \tag{5.5}$$

$$M_d[X, Y] = [(M_c[X] - M_p[X]), (M_c[Y] - M_p[Y])]. \tag{5.6}$$

$$D_{n-1}[X, Y] = [(D_n[X] - M_d[X]), (D_n[Y] - M_d[Y])]. \tag{5.7}$$

The regions of the rectified image that had no feature matches resulted in a matrix element pointing from the centre of the cube to the same centre. These points were tested as if there had been no movement, and resulted in the same disparity map that would have been obtained for the fast frame rates. The drawing of the cubes is illustrated in Figure 5.12 where the cubes are drawn across the stereo anaglyph of the current and previous frame.

The coordinates of the detected feature's previous location were determined with the aid of feature matching. The location was used to determine if the disparities matched. If the average of the cube that was calculated above is used as a direction vector, it can be seen that the coordinates of where the actual match was found may have shifted. The shift is dependent on how many entries were found in the cube. If these entries had different directions or different magnitudes, it would have caused a shift. To compensate for the shift, a template match-like feature like with the disparity map formulation

Stellenbosch University https://scholar.sun.ac.za

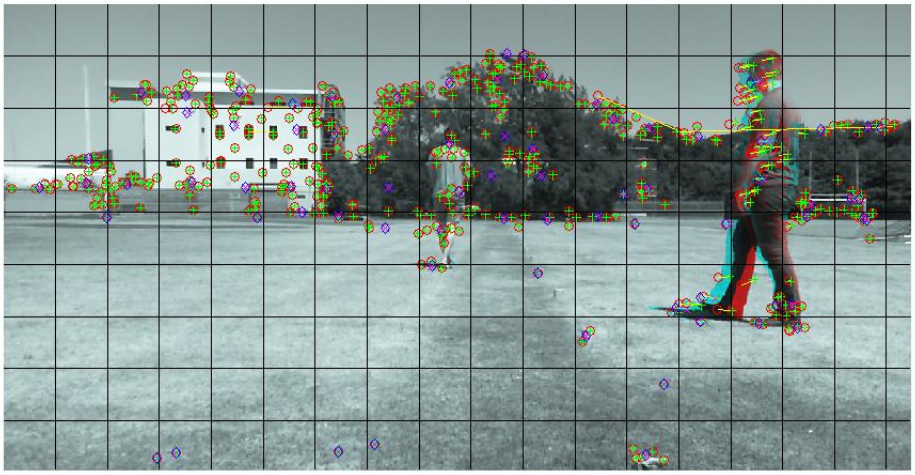


Figure 5.12: Rectangles are drawn across the stereo anaglyph of the current and previous frame to show the direction matrix. The matching points and the average of the cube are also shown.

was implemented. Instead of looking at a single pixel, the pixels around the point of interest were matched as well. If any of these points were a good match, the disparity was retained.

The direction vector was affected by the size of the cube. Having a large cube increases the number of entries to be taken into account. Large cubes can consist of multiple objects, each with their own direction of movement. An average direction has to be computed, which will find no matches if tested. Taking a small cube size will reduce the number of entries, making the search region more accurate, but at the expense of losing some parts of the features. The parts that will be lost are the parts where the cube has no matching features. Care must therefore be taken when selecting a cube size. The implemented cube sizes had dimensions of a ninth of the number of rows.

5.5.3 Outliers in Feature Matches

The second problem that was found was with outliers. Some of the matches that were found were not correct, having one point on the left side of the image and the other on the right. These outliers shifted the average dramatically. The outliers were eliminated by computing the average magnitude for the cube and comparing it to the largest magnitude of the points that were used to calculate the average. If the largest magnitude for the cube was more than three times bigger than the average, it was deemed unacceptable. If this was the case, the largest magnitude pair was removed and the calculation was repeated until it was acceptable. Testing if the largest magnitude was double

Stellenbosch University https://scholar.sun.ac.za

that of the average could, in some situations, remove some of the acceptable matches. Figure 5.12 illustrates the removal of the incorrect matches. An incorrect match can be identified in the top left corner of the image. The average coordinate for the current and previous image falling within the circle of the top left corner indicated that the outlier match was not taken into consideration.

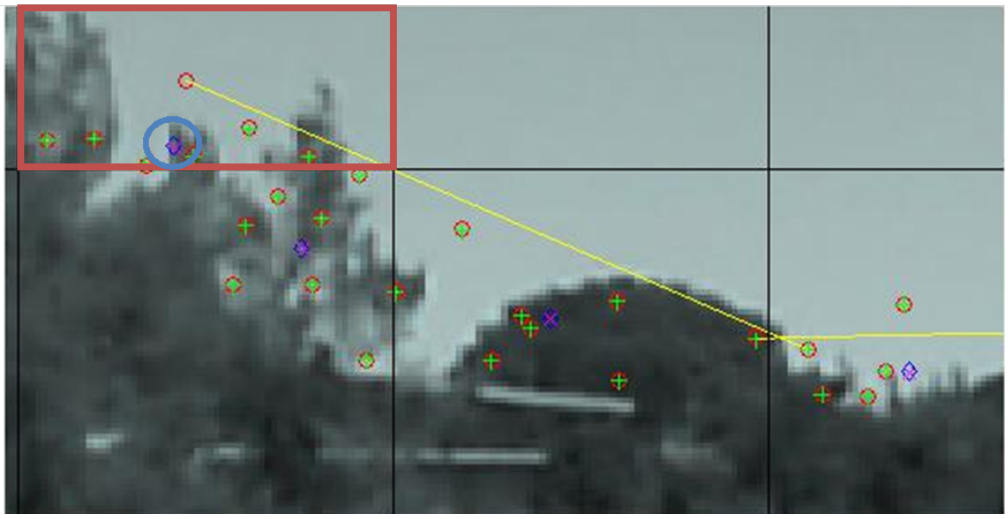
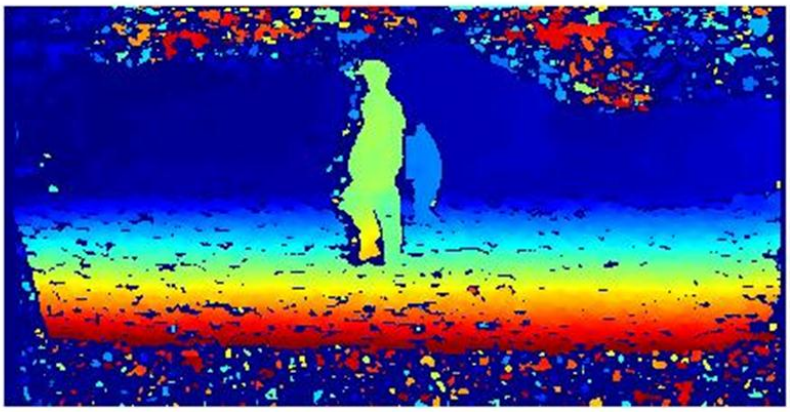


Figure 5.13: The effect of the outlier was removed from the average calculation. The average fell within the circle region as expected. The averages for the current and previous frame are indicated by the blue diamond and pink cross.

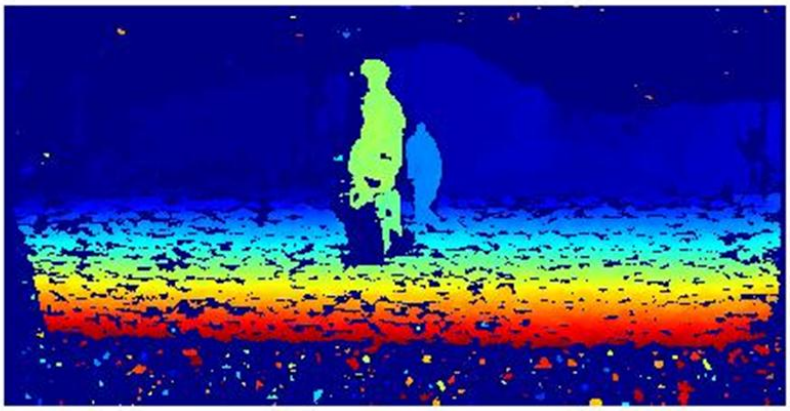
5.5.4 Resultant Disparity Maps

The temporal noise removal depends on the disparity maps formulated by the other noise removal functions. Having a disparity map that retains most of the body of the detected objects can improve the results of the temporal noise removal. To better illustrate the temporal noise removal, a disparity map with more noise and object body was computed by using MATLAB's disparity function. The resulting disparity map can be seen in Figure 5.14. The image at the top is the original disparity map and the image at the bottom is after temporal noise removal. The difference in the images can be clearly seen, especially in the sky region. Having large disparity values detected in the sky caused the 3D model to project them forward. These points were seen as definite objects in the path of the vehicle. The the number of pixels that found matches in the previous disparity map was calculated to give us an indication of the number of pixels identified as possible noise. It was determined that 49% of the pixels had found matches.

Stellenbosch University https://scholar.sun.ac.za



(a)



(b)

Figure 5.14: Disparity map with more clear features and the effect of temporal noise removal on it (a) Disparity map before temporal noise removal and (b) Disparity map after temporal noise removal.

Stellenbosch University https://scholar.sun.ac.za

Chapter 6

Results and Discussion

From the start of the algorithm to the end, major changes were detected in the resulting disparity maps. The process started off by accepting two image inputs from two cameras. These images were rectified and a disparity map was computed. The disparity map contained noise which is undesirable. The pixels of the disparity maps needed to be tested to filter out the noise. The initial filtering process was done by testing whether the disparity of the pixel was a good match or not. The new disparity map contained a lot less noise than its predecessor, but was still not deemed clear enough. The final step for noise removal was based on comparing the current disparity map with one that had been computed in the previous time step. From the resulting disparity maps, the 3D coordinates of the detected objects were determined. Additional experiments were also conducted to investigate the effect of the environment of the disparity maps. The additional experiments, 3D reconstruction, the accuracy of the results and problem areas are discussed in this chapter.

6.1 Additional Experiments

The initial experiment that was conducted was in an open field environment where objects were aloud to move but with a stationary camera set-up. The camera set-up was kept stationary to test the accuracy of the results. Six consecutive images of the initial open field test can be seen in Figure 6.1 and illustrates the difference between the resulting disparity maps as time progresses. The movement of the person in front can clearly be seen between each frame and the body retained most of it's volume. Noise appears and disappears between the frames and indicates that the temporal noise removal is not perfect.

To investigate the effect of the environment where the camera set-up is implemented, a dynamic experiment was done. The dynamic experiment involved moving the camera set-up as the video recordings were conducted. Three different scenarios were investigated, a hallway, a high texture environment and

Stellenbosch University https://scholar.sun.ac.za

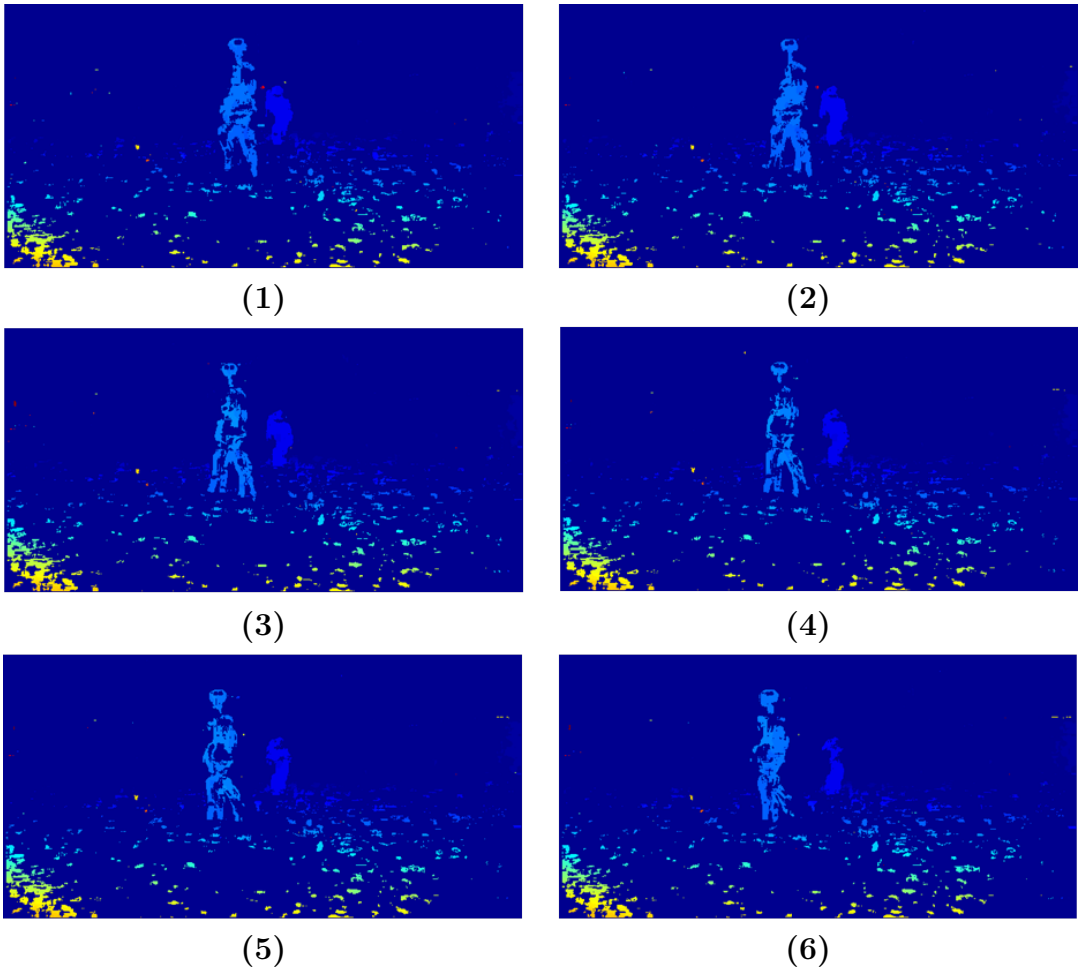


Figure 6.1: Disparity maps of an field as time progresses in sequence from 1 to 6.

a scenario where the camera set-up and the environment is moving.

6.1.1 Hallway

The first scenario was conducted in a hallway and aimed to investigate the effects of a cramp environment. A person was positioned close to the end of the hallway and was kept stationary. The camera set-up was moved towards the person. The disparity map formulation and noise filtering of two consecutive images are illustrated in Figure 6.2, where all the steps before the temporal noise removal can be seen.

Stellenbosch University https://scholar.sun.ac.za

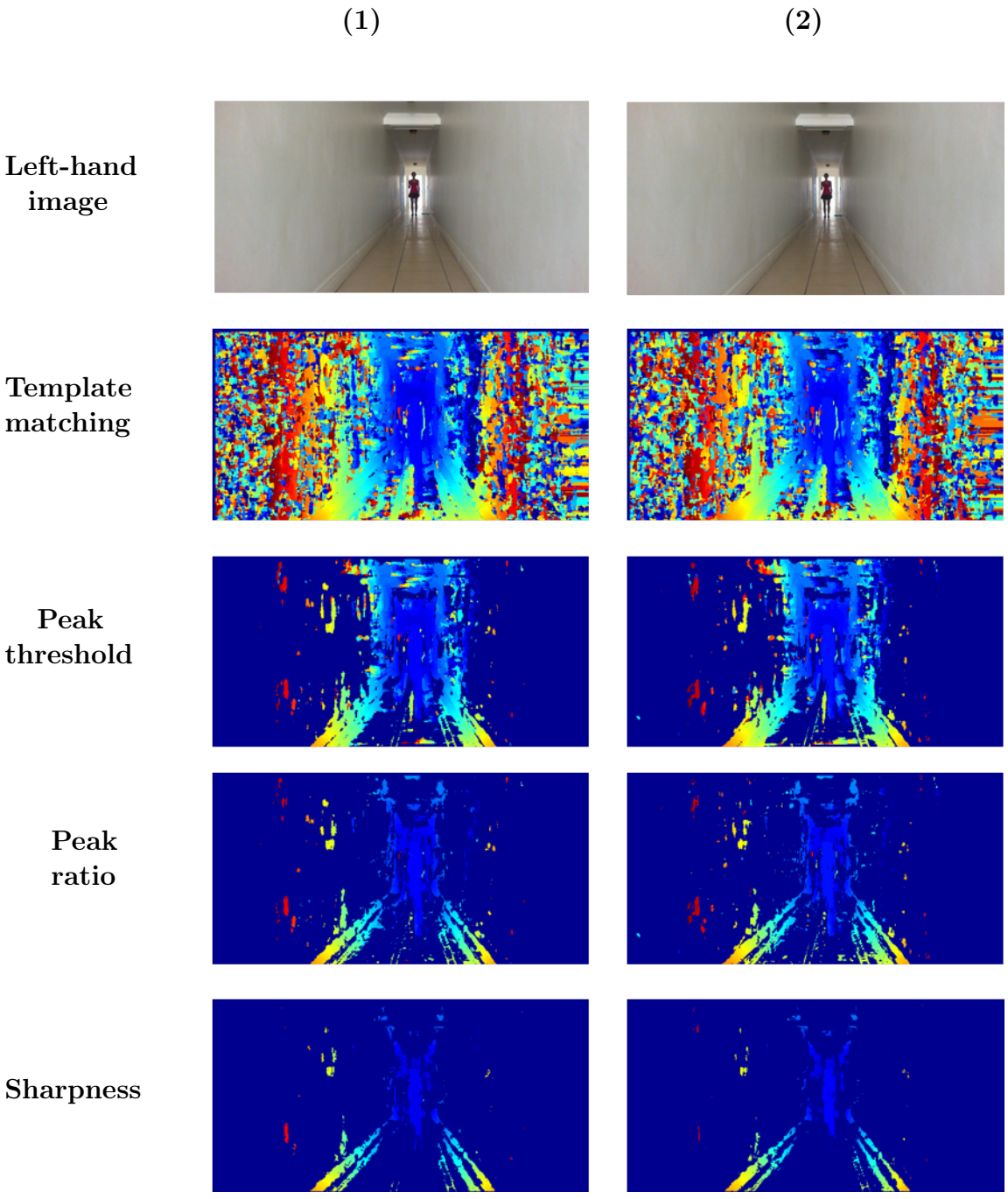


Figure 6.2: Disparity map formulation of a cramp environment indicating the results for template matching, peak threshold, peak ratio and sharpness for two consecutive images.

The template matching showed a variety of different disparity values on the walls close to the cameras, but with an underlining tone of disparity change in the direction away from the cameras. The peak threshold removed majority of

Stellenbosch University https://scholar.sun.ac.za

the disparity values of the walls, indicating that most of the matches did not have a high similarity score. The person in the middle of the hall is still visible at the end of the noise removal steps. The connection between the walls and floor can also be seen after the sharpness test. The temporal noise removal is illustrated in Figure 6.3. After the temporal noise removal, no clear indication of noise is still visible. The large disparity values on the sides of the walls are deemed correct, as they are scaling with the disparity values close to the cameras and can be compared to the connection between the walls and the floor.

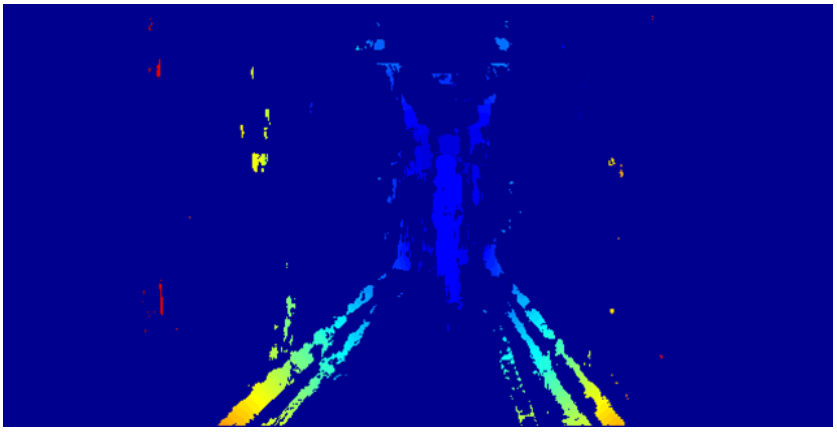


Figure 6.3: Temporal noise removal in a cramp environment.

6.1.2 High Texture Environment

The second scenario was conducted outside of a building's parking lot near bushes. The aim of the experiment was to determine the effect of high contrast regions close to the cameras. The camera set-up was moved in the direction of the bush. The disparity map formulation and noise filtering of two consecutive images are illustrated in Figure 6.4, where all the steps before the temporal noise removal can be seen.

Stellenbosch University https://scholar.sun.ac.za

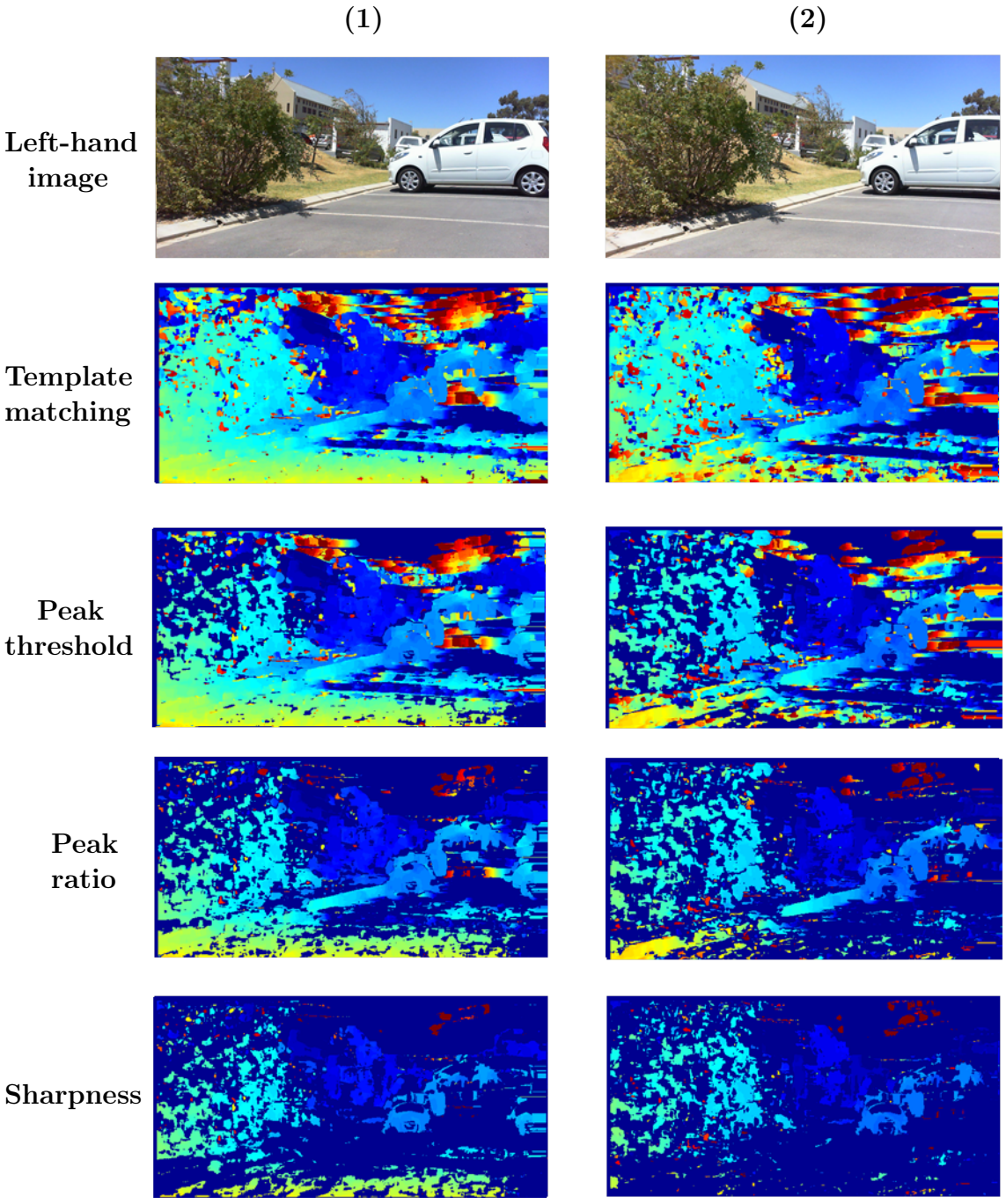


Figure 6.4: Disparity map formulation of a high texture environment indicating the results for template matching, peak threshold, peak ratio and sharpness for two consecutive images.

It was found that with high contrasts areas it was better to blur the image more with the pre-filters, resulting in high texture areas having less texture. Loss in the body of the bush can be noted after template matching. The

Stellenbosch University https://scholar.sun.ac.za

pavement was removed with the sharpness of the peak, because too many matches were possible good matches, resulting in a broad peak. The temporal noise removal for the high contrast images can be seen in Figure 6.5, where portions of the vehicle, the bush in the bottom left, the bush in the top left corner and the bush behind the vehicle can be identified. Some noisy pixels that had large disparity values were removed as well. The algorithm struggle with high texture regions, where a clearer disparity map of the bush is more desirable.

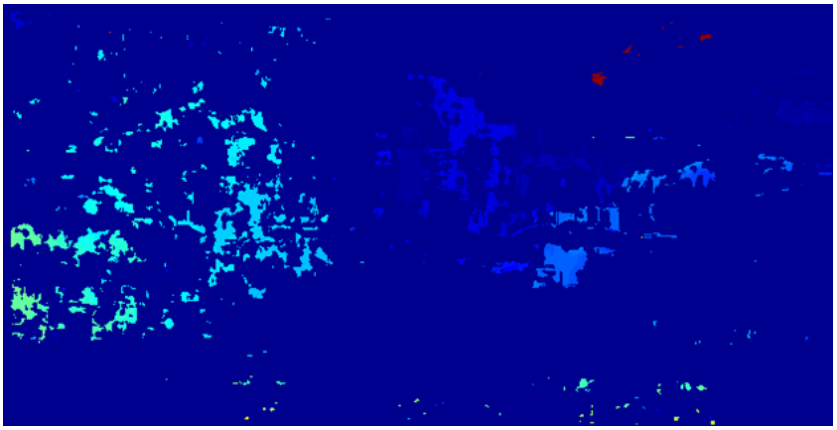


Figure 6.5: Temporal noise removal on a high texture environment.

6.1.3 Multi-Movement

The third scenario was to test the movement of the camera set-up with a moving environment. The disparity map formulation can be seen in Figure 6.6. In both time stamps the line in the middle of the road is visible throughout the noise filtration. The pixels that formed the line in the road had good similarity scores. The cars on the sides of the road is also visible after the sharpness test was applied. The car in the middle of the road was the only car that moved. The car in the middle is still visible in the final disparity map.

Stellenbosch University https://scholar.sun.ac.za

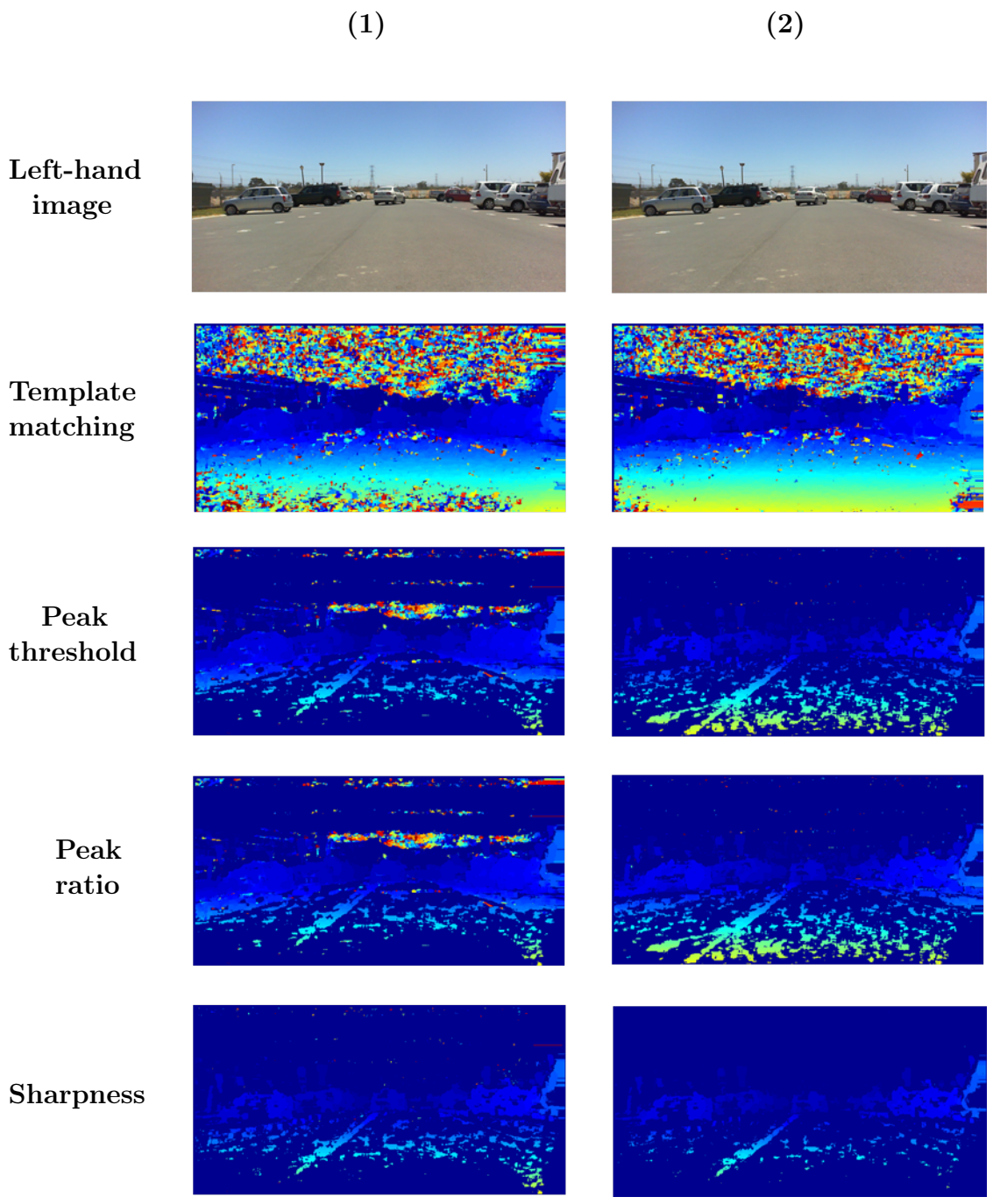


Figure 6.6: Disparity map formulation where both the camera set-up and environment moved and indicating the results for template matching, peak threshold, peak ratio and sharpness for two consecutive images. The vehicle in the back moved away from the camera set-up.

The temporal noise removal for the moving environment is illustrated in Figure 6.7. The line in the road is still visible and changing disparity values as

Stellenbosch University https://scholar.sun.ac.za

the distance from the cameras increase. The vehicles on the sides and in the middle of the road is still visible, indicating that the movement of the camera set-up still delivered acceptable results.

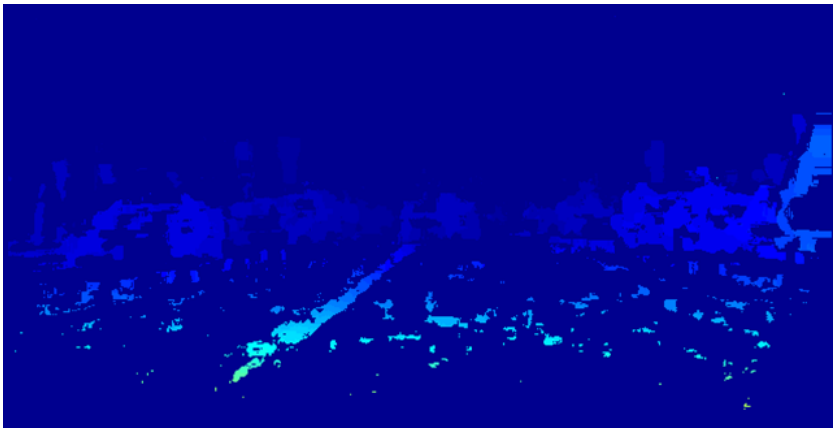


Figure 6.7: Temporal noise removal in a parking lot where the camera set-up and vehicle moved.

In the fourth test the camera set-up and vehicle moved towards one another, increasing the rate of change in the vehicle's location per frame. This test was done by placing the camera set-up in one lane and the vehicle in the other. The disparity map formulation and noise filtering of two consecutive images are illustrated in Figure 6.8. The vehicles on the side of the road can be identified in the disparity maps. The vehicle moving towards the camera set-up can be identified in the disparity maps as well.

Stellenbosch University https://scholar.sun.ac.za

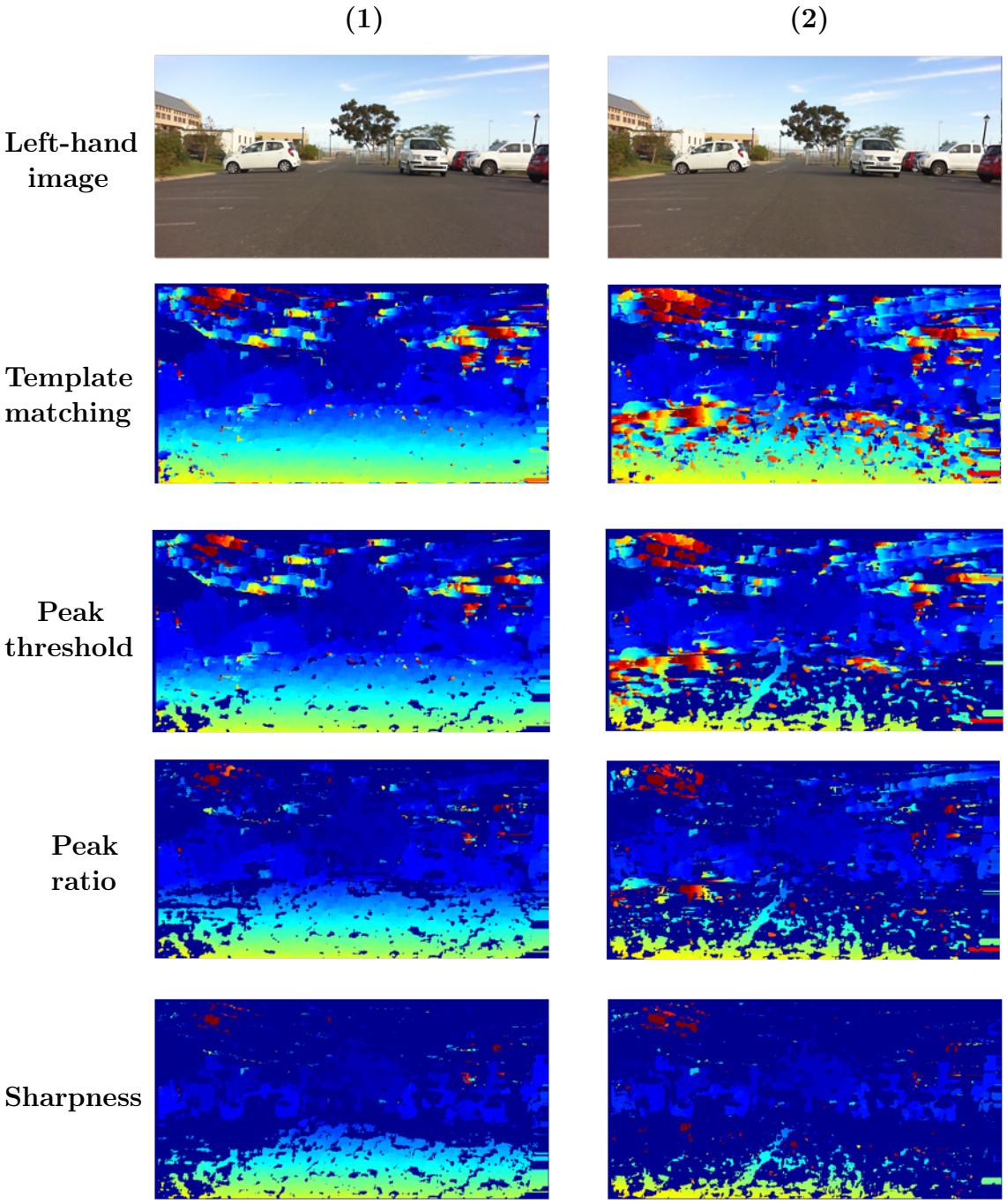


Figure 6.8: Disparity map formulation where both the camera set-up and environment moved and indicating the results for template matching, peak threshold, peak ratio and sharpness for two consecutive images. The vehicle and the camera set-up moved towards one another.

The temporal noise removal for the moving environment where the vehicle is moving towards the camera set-up is illustrated in Figure 6.9. The clouds

Stellenbosch University https://scholar.sun.ac.za

have noisy regions where large disparity values were present in both the current and previous disparity maps. This indicates that if the input disparity maps for temporal noise removal have noise in the same region, the noise will not be removed. To improve the results, a better formulation of the initial disparity maps are required. The moving vehicle is identifiable on the right hand side. The visible side of the vehicle can be seen and portions of the front end. It should be noted that the vehicle is closer to the cameras in the current frame then it was in the previous frame, but the pixel coordinates of the vehicle did not change dramatically.

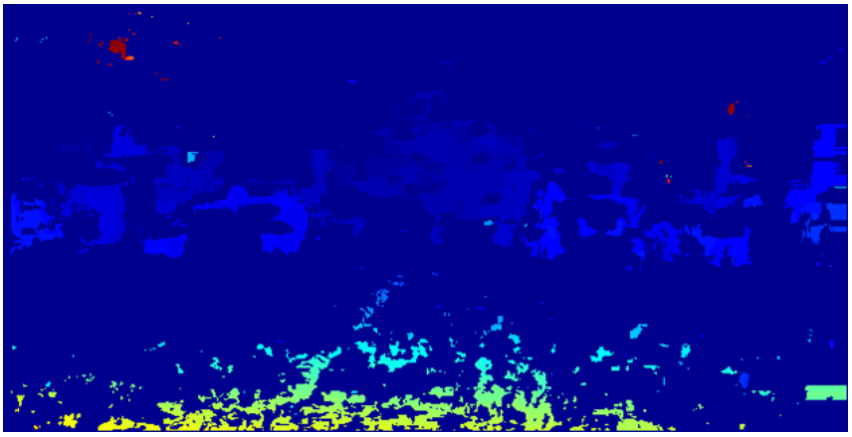


Figure 6.9: Temporal noise removal in a parking lot where the camera set-up and vehicle moved towards one another.

6.2 3D Map and Visualization

From the disparity maps it was possible to generate a 3D model of the field of view. The depth at which a pixel can be found in the 3D model was calculated from the pixel disparity and camera properties. The camera properties are the focal length and baseline. The 3D model of Figure 5.14 is illustrated in Figure 6.10 where the disparity map from Figure 5.14(b) was used. The axis of the model is in metres. The noise that was left in the disparity map is clearly visible in the model. The small blue particles scattered across the model can be linked to the sky in the original image.

Going back to the disparity map in Figure 5.4, a top-down view was generated of where the points were located in the environment. The frontal and top-down views are illustrated in Figure 6.11. The circles indicate where the people are located. The points at the bottom of the image are the noise that was found.

Large amounts of noise have removed from the disparity maps. To illustrate the amount of noise that was removed, Figure 6.12 was used to make a 3D

Stellenbosch University https://scholar.sun.ac.za

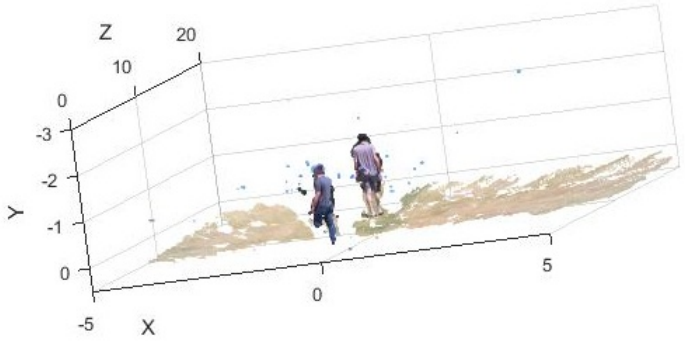


Figure 6.10: 3D model of Figure 5.14(b).

reconstruction. The 3D reconstruction can be seen in Figure 6.13, where the noise that is projected forward can be clearly seen.

6.3 Disparity Accuracy

The accuracy of the detected disparity was not perfect. The detected matches were filtered through a variety of techniques to decrease the uncertainty in the calculated disparity of that point. After the point had passed all the tests, there was a high probability that the point existed. The disparity value itself could also be inaccurate. Figure 6.14 illustrates the distance each disparity represents. Objects that were far away from the cameras will have a small disparity. A small disparity value points to a large distance. By incrementing a small disparity with a value of 1, a large jump in distance is obtained. Take, for example, a pixel with a disparity value of 2. Using the current camera set-up, the distance will be 106 m. Incrementing the disparity with 1, changes the distance to 70 m. That is a very large jump to make.

Using the method that was used in this study to calculate disparity, the disparity is determined by searching through the image, one pixel at a time. Thus the best match will have a disparity value of an integer. If the actual best match was between two pixels, or just a portion of the pixel, it will have the same integer value. For example, if an object is located at 80 m, the disparity value will most likely be 3, indicating that its at 70 m. This problem is more likely to occur with objects that are far from the cameras. With objects that are close to the camera, this effect if not so problematic.

Referring back to Figure 4.2, the cups on the ground served as markers. The first cup was at 4 m, with intervals of about 2.5 m per cup. The intervals were determined by using a rope that was about 2.5 m long. This implies that the person closest to the camera was about 6.4 m away and the second person was 12.5 m away.

Stellenbosch University https://scholar.sun.ac.za

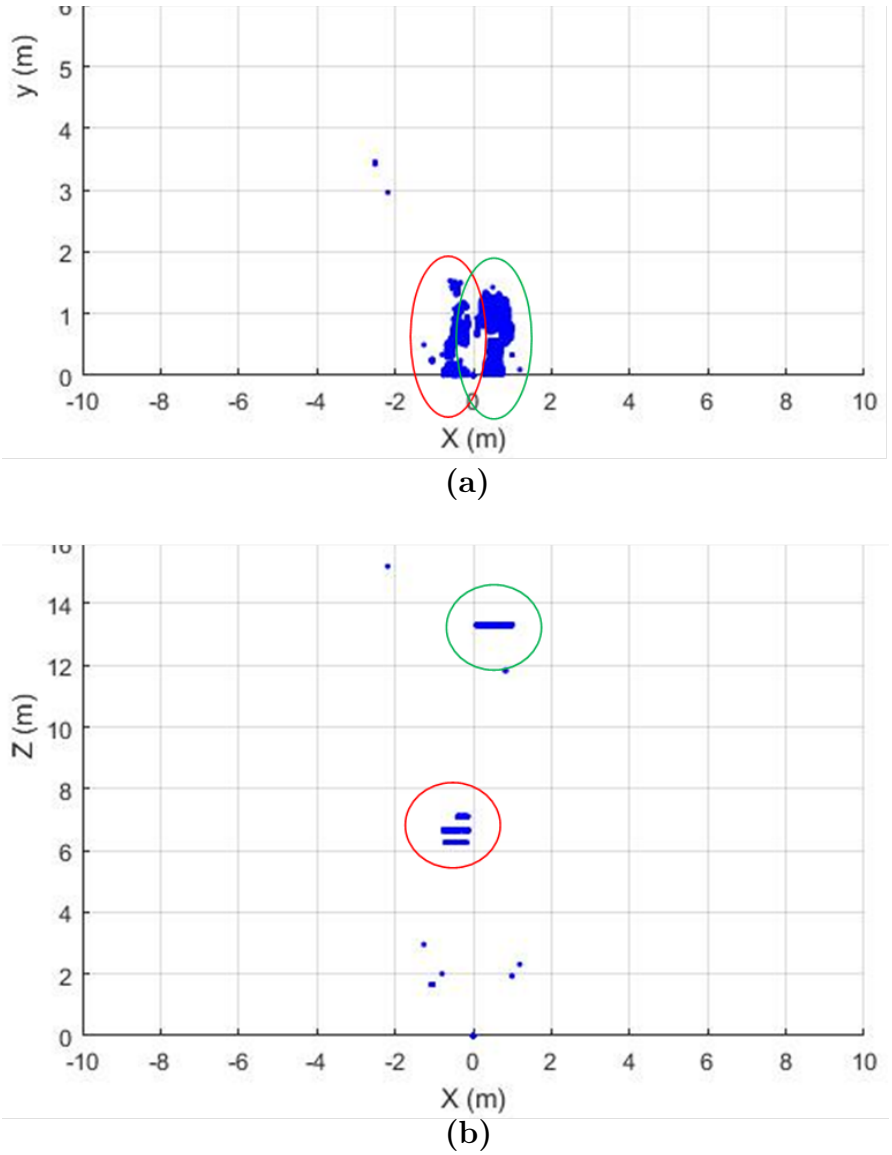


Figure 6.11: Resulting frontal and top-down views of Figure 5.4 (a) Frontal view and (b) Top-down view.

In Figure 6.11, inspecting the position of the man closest to the camera, the distance at which he was detected ranged from 6.6 m to 7 m. This indicated that he could be approximately 0.4 m wide. The calculated distances for the objects that were close to the cameras showed acceptable results. The disparity values for the distances of 6.6 m and 7 m were 32 and 30, indicating there was a disparity shift of 2 when locating the matching features with template matching in the disparity calculation.

The accuracy for the stereo vision set-up aims to have an accuracy of 0.5 m for objects that are within 15 m of the cameras. For larger distances away from the camera the accuracy will be reduced due to the difference in distance

Stellenbosch University https://scholar.sun.ac.za

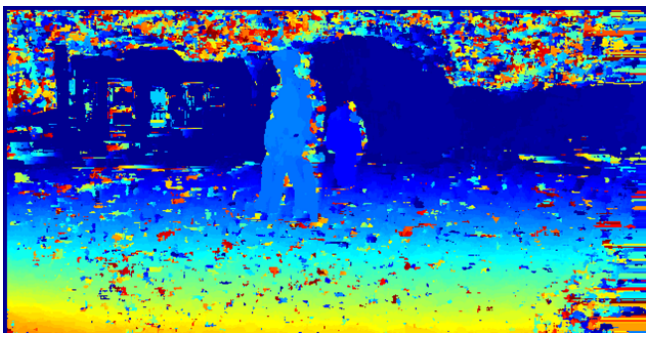


Figure 6.12: Disparity map without noise removal.

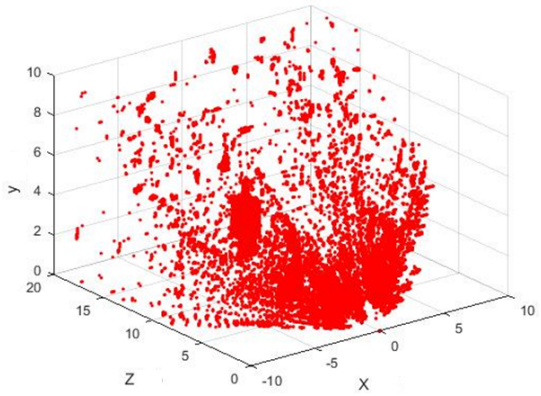


Figure 6.13: 3D model of Figure 6.12 indicating the amout of noise in the disparity map.

represented by each disparity value. The accuracy was deemed acceptable because we only need an indication of where objects are in the field of view.

The noise at the bottom of the top-down view was linked to the noise at the bottom of the disparity map. These noise particles were strong matches that were found for the grass features in the original images. Objects like the grass that are close to the cameras will generate large disparity values. The disparity values for those grass pixels that were found were the same as the maximum disparity. Disparity maps that have disparity values at the maximum disparity range located in the bottom regions of the map can be filtered out by setting the region of interest a set distance from the bottom on the disparity map. Noise with maximum disparity is most likely to occur in that region. The small portions of noise near the bottom of the disparity map can be removed by setting a region of interest within the obtained input images.

Comparing the disparity map of Figure 5.4, formulated by the disparity function of the study, to MATLAB's disparity map in Figure 5.14, a clear difference can be seen between the two. There is a lot more body in the de-

Stellenbosch University https://scholar.sun.ac.za

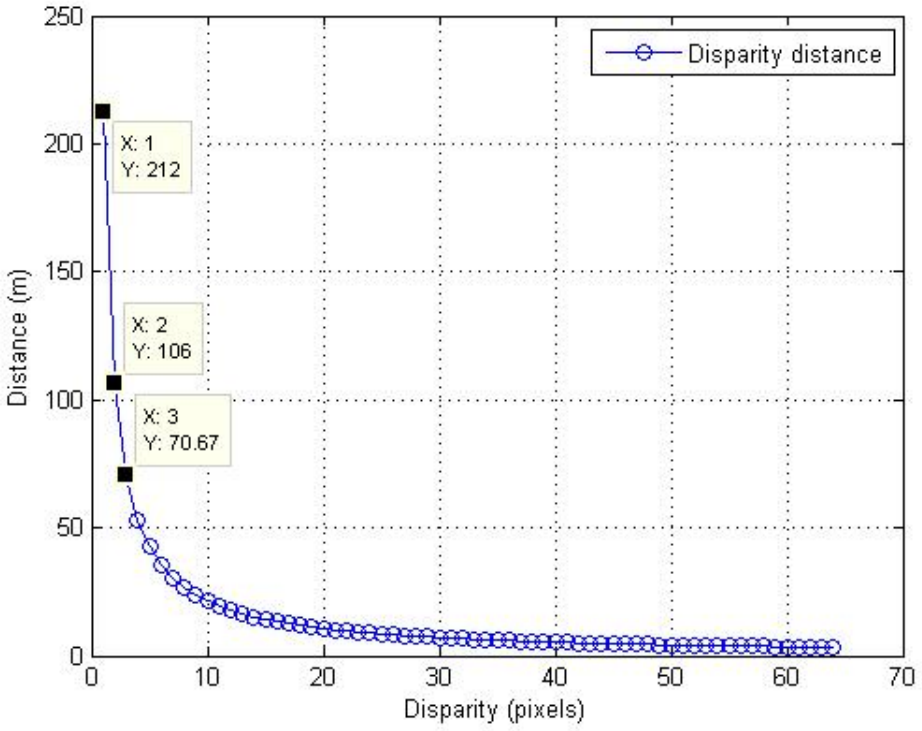


Figure 6.14: Distance represented by disparities.

tected objects in MATLAB's disparity map. Having more body for a detected feature will always be desirable, but less noise is for the objective. Comparing the locations of the detected object in Figure 6.10 and Figure 6.11(b) identified small differences in the detected distance. MATLAB's disparity function showed that the person in front was detected at between 6.4 m and 7 m. The person in the back was detected at 13.4 m. These results were very similar to those obtained with the study method.

To increase the accuracy of the stereo vision application, the baseline must increase. Increasing the baseline increases the maximum range of the objects that can be detected, but at the expense of minimum range. Thus, for the implementation of stereo vision it is recommended that you know the intended use so that the baseline can be determined.

6.4 Problem Areas

Disparity maps are influenced by a variety of factors that can decrease the quality of the resulting disparity map. Having a baseline that is to large can change the view angle of an object, resulting in lower similarity scores. To improve the similarity score for objects that are closer to the cameras, the baseline needs to decrease.

Stellenbosch University https://scholar.sun.ac.za

Temporal noise removal is affected by the recorded frame rate. If the frame rate is too low, the pixel coordinates of an object can change dramatically. With the object being in different locations, SURF features struggle to find positive matches for portions of the object's body. To improve the results of temporal noise removal, faster frame rates is recommended to reduce the pixel coordinate difference.

Stellenbosch University https://scholar.sun.ac.za

Chapter 7

Conclusion

The disparity map formulation underwent significant changes from the initial stage, with template matching, to the final stage after temporal noise removal. Large portions of noise were removed from the disparity map with the initial three threshold techniques that tested for spatial aliasing, occlusion and low texture regions. The implementation of temporal noise removal reduced the level of noise even further. The results from the temporal noise removal were significant when a disparity map with low threshold values was used for the initial tests.

Noise reduction in disparity maps can be done, but at the expense of loss in the body of the detected object. The loss in body can be reduced by adjusting the thresholds for each noise removal technique, but at the expense of less noise removal. The implementation of these techniques depends on what the user desires.

The temporal noise removal can be implemented at the same expense as that mentioned above. The temporal noise removal is affected by the quality of the input disparity maps. Having more object body in the disparity map will result in clearer objects after the temporal noise removal.

The detected distances of objects that were close to the cameras - less than 15 m - were deemed acceptable. Objects at longer distances could be detected but with uncertainty in the distance from the camera. The uncertainty was caused by disparity values that were generated as integers.

The implementation of stereo vision as a stand-alone sensor depends on the size of the area and the velocity at which the vehicle is moving. The implementation of stereo vision as a stand-alone sensor is not advised when working with a fast-moving vehicle. The detected distance of far away objects can possibly change dramatically. This will give the vehicle a false sense of where the object is located. The amount of stopping time required increases as the velocity of the vehicle increases. With slower-moving vehicles, the vehicle is given more time to react and the application can be deemed plausible.

Stellenbosch University https://scholar.sun.ac.za

List of References

Aditya, K.P., Reddy, V.K. and Ramasangu, H. (2014). Enhancement Technique for Improving the Reliability of Disparity Map under Low Light Condition. *Procedia Technology*, vol. 14, pp. 236–243. ISSN 2212-0173.
Available at: http://www.sciencedirect.com/science/article/pii/S221201731400067X

Agrawal, M. and Konolige, K. (2006). Real-time localization in outdoor environments using stereo vision and inexpensive GPS. In: *Proceedings - International Conference on Pattern Recognition*, vol. 3, pp. 1063–1068. ISBN 0769525210. ISSN 10514651.

Arnfred, J.T., Winkler, S. and Susstrunk, S. (2013). Mirror Match: Reliable Feature Point Matching without Geometric Constraints. *Pattern Recognition (ACPR), 2013 2nd IAPR Asian Conference on*, pp. 256–260.

Badino, H., Huber, D. and Kanade, T. (2011). Integrating LIDAR into stereo for fast and improved disparity computation. In: *Proceedings - 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2011*, pp. 405–412. ISBN 9780769543697.

Bradski, G., Kaehler, A. and Bradski, G. (2013). *Learning OpenCV*, vol. 53. ISBN 978-1-107-05799-1. arXiv:1011.1669v3.
Available at: http://eprints.utas.edu.au/4774/

Corke, P. (2011). *Robotics, Vision and Control - Fundamental Algorithms in MATLAB*. ISBN 3540221085. 978 3 642 20143 1.

Fradi, H., Dugelay, J.-l. and Antipolis, S. (2011). Improved depth map estimation in Stereo Vision. *Most*, vol. 33, no. 0.

Häne, C., Sattler, T. and Pollefeys, M. (2015). Obstacle detection for self-driving cars using only monocular cameras and wheel odometry. In: *IEEE International Conference on Intelligent Robots and Systems*, vol. 2015-Decem, pp. 5101–5108. ISBN 9781479999941. ISSN 21530866.

Howard, A. (2008). Real-time stereo visual odometry for autonomous ground vehicles. In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pp. 3946–3952. ISBN 9781424420582. ISSN 17594782.

Stellenbosch University https://scholar.sun.ac.za

Kostavelis, I., Nalpantidis, L. and Gasteratos, A. (2009). Real-Time Algorithm for Obstacle Avoidance Using a Stereoscopic Camera. *Student Eureka*.
Available at: `http://83.212.134.96/robotics/wp-content/uploads/2011/12/Real-Time-Algorith`

Laganière, R. (2011). *OpenCV 2 Computer Vision Application Programming Cookbook*. ISBN 9781849513241. `arXiv:1011.1669v3`.
Available at: `http://zenithlib.googlecode.com/svn/trunk/books/OpenCV_2_Computer_Vision_Ap`

Pollefeys, M., Nistér, D. and Frahm, J.-M. (2008). Detailed Real-Time Urban 3D Reconstruction From Video. *International Journal of Computer Vision*, pp. 1–43. ISSN 0920-5691.

The MathWorks Inc. (2013). Computer Vision System Toolbox. *(Mathworks) http://www.mathworks.com/products/computer-vision/*, pp. http://www.mathworks.com/products/computer–vision/.
Available at: `http://www.mathworks.com/products/computer-vision/`