



Autonomous Underwater Vehicle: Vision System



By Jack Reynolds

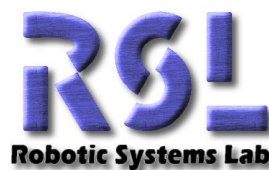
October 1998

Project Supervisors:

Mr Samer Abdallah	DE
Dr David Wettergreen	RSISE
Dr Alex Zelinsky	RSISE

Technical Supervisor:

Mr Martin Stonebridge	DE
-----------------------	----



Acknowledgments

There are many people that I would like to thank for helping me to perform this work.

I would like to thank in particular the entire AUV team (David Wettergreen, Chris Gaskett, Alexander Zelinsky, Dale Beswick, Samer Abdallah and Jon Kieffer) who were always willing to help me, no matter how busy they were with their own work. David especially was always available to help, even when on the other side of the world. They all made the project a great learning experience, not just in the area of work relating directly to the vision system, but also in many other diverse areas including learning systems, Kalman filters, amplifier design, and even component ordering and acquisition!

I am grateful also to Son Truong, Leon Mills, Martin Stonebridge and David Wettergreen for their assistance with mechanical design ideas. By the end of the project I was having nightmares about AutoCAD, but overall I think that during the year my mechanical design and technical drawing skills improved greatly.

I would also like to thank Gordon Cheng, Peter Kittler, Gerald Lee and everyone else who seemed to live in the computer labs, and provided motivation and someone to talk to when things were not running smoothly.

Finally, I would like to thank my family for providing constant encouragement and support throughout, and also for the many hours of spent proofreading reports and listening to seminar rehearsals.

Abstract

This thesis examines the design and implementation of a vision system for an autonomous underwater vehicle. The vision system consists of three cameras, two being a stereo pair, and the third being attached to a pan/tilt unit to allow it to track moving targets.

The goals for the development of the vision system were to:

- Design and build all of the hardware for the vision system. The main focus of work in this area was to design watertight camera housing to attach the two stereo cameras to the frame of the AUV.
- Interface all of the cameras with the on-board computer, and the optical fibre link to the surface.
- Develop software to allow the vision system to be used as a navigational aid for the submersible. To do this a range estimation system was implemented for use with the stereo cameras. A feature tracking system was also implemented to allow the AUV to track moving targets, either by moving the pan/tilt camera, or by moving the whole AUV.

Table of contents

Chapter 1 Introduction.....	1
1.1 Autonomous Underwater Vehicles	1
1.2 Underwater Vision.....	2
1.3 System Overview	3
1.4 Project Goals.....	4
1.5 Thesis Outline	4
Chapter 2 Vision System Hardware.....	6
2.1 Background.....	6
2.1.1 Stereo Camera Housing	6
2.1.2 Camera Housing Lens	7
2.1.3 Vision System Cabling	7
2.2 Stereo Camera Housing	7
2.2.1 Design Requirements.....	8
2.2.2 Solution.....	9
2.3 Lenses for Camera Housing and Cameras.....	10
2.3.1 Choice of Camera Housing Lenses	11
2.3.2 Choice of Stereo Camera Lenses	12
2.4 Data and Video Cables.....	13
2.5 Conclusion.....	14
Chapter 3 Image Processing.....	16
3.1 Image Enhancement.....	16
3.1.1 Edge Detection using Gradients	16
3.1.2 Edge Detection using Difference of Gaussians	18
3.2 Template Matching	19
3.3 Image Processing Testing.....	21
3.3.1 Edge Detection test.....	22
3.3.2 Template Matching Tests	22
3.4 Conclusion.....	23
Chapter 4 Software Design.....	24
4.1 Background.....	25
4.1.1 Range Estimation.....	26
4.1.2 Feature Tracking.....	29
4.2 Vision System Requirements	30
4.3 Range Estimation Implementation	30
4.3.1 Calibration	32

4.4 Feature Tracking Implementation.....	33
4.5 Integrating Feature Tracking and Range Estimation.....	33
4.5.1 Template Extraction	33
4.5.2 Template Updating	34
4.6 Software Testing.....	34
4.6.1 Template Update Frequency.....	34
4.6.2 Image Enhancement Benefits	38
4.6.3 Range Estimation Calibration.....	39
4.6.4 Range Estimation Accuracy at Varying Angles from Optical Axis.....	40
4.6.5 Effects on Range Estimation of Variations in the Baseline.....	42
4.6.6 Feature Tracking Tests	43
4.6.7 Combined Range Estimation and Feature Tracking Tests	44
4.7 Conclusion.....	45
 Chapter 5 Further Work.....	 47
5.1 Hardware.....	47
5.2 Software	47
5.2.1 Conversion of Code to VxWorks	47
5.2.2 Automatic Correlation Threshold Adjustment	48
 Chapter 6 Conclusion	 49
 References	 51
 Appendices	 A1
Appendix A Hardware Specifications.....	A2
Appendix B Derivation of Snell’s Law	A3
Appendix C Field of view underwater calculations.....	A5
Appendix D Calculation of Focal Length of Camera Lens.....	A6
Appendix E Calculation of feature tracking accuracy.....	A8
Appendix F Range estimation for converging cameras.....	A11
Appendix G Feature Tracking of an underwater target.....	A13
Appendix H Camera Housing Diagrams	A17
Appendix I AUV Endcap Diagrams	A25

List of Figures and Tables

Figures

Figure 1 – Autonomous Underwater Vehicle: Kambara	2
Figure 2 – Autonomous underwater vehicle vision system.....	4
Figure 3 - Degrees of freedom of stereo camera within its housing	8
Figure 4 - Degrees of freedom of stereo camera housings	9
Figure 5 - Section view of camera housing	10
Figure 6 - Distortion at water/glass/air interface due to light bending	11
Figure 7 - Images from lenses with 2.8mm (left) and 4.0mm (right) focal lengths	13
Figure 8 - Cabling diagram for complete vision system	14
Figure 9 - Example of the first derivative of an image.....	17
Figure 10 - Effects of noise on first derivative edge detection.....	17
Figure 11 - Calculation of the difference of Gaussian of an image.....	18
Figure 12 - Second derivative and difference of Gaussian of a noisy image	19
Figure 13 - Extraction of a template containing the target	20
Figure 14 - Correlation of the template on a new image	20
Figure 15 - Mask used for Gaussian blurring	22
Figure 16 - Difference of Gaussians with various threshold levels.....	22
Figure 17 - Correlation of a template with the image it originates from.....	23
Figure 18 - Kambara's software architecture [8]	24
Figure 19 - Image processing test program interface	25
Figure 20 - Stereo range estimation using parallel cameras	26
Figure 21 - Similar triangles used to estimate range for parallel cameras	27
Figure 22 - Range estimation with non-parallel cameras	28
Figure 23 - Feature tracking system	29
Figure 24 - Range estimation block diagram	31
Figure 25 - Calibration of stereo cameras	32
Figure 26 - Integrated feature tracking and range estimation systems	33
Figure 27 - Correlation values when target can and cannot be located.....	35
Figure 28 - Correlation values of left and right cameras.....	36
Figure 29 - Correlation levels for various updating techniques	37
Figure 30 - Correlation levels for template matching using enhanced images	38
Figure 31 - Graphs of range vs disparity (experimental and theoretical).....	39
Figure 32 - Test procedure for range estimation accuracy of off centre targets.....	41
Figure 33 - Error in range estimations as lens distortion increases	41
Figure 34 -Effect of varying baselines on range estimation accuracy.....	42
Figure 35 - Initial images from fish tracking test	43
Figure 36 - Correlation levels for underwater tracking of a target.....	44
Figure 37 - Light bending at water/air interface.....	A3

Figure 38 - Snell's law	A5
Figure 39 - Distortion at air/water interface	A5
Figure 40 - Calculation of focal distance	A6
Figure 41 - Error in estimated object position.....	A8
Figure 42 - Estimation of object position.....	A9
Figure 43 - Graph of object location vs pixel position.....	A10
Figure 44 - CCD error position vs error in position estimation	A10
Figure 45 - Schematic of converging cameras	A11
Figure 46 - Correlation values when tracking a target with no template updating	A13
Figure 47 - Correlation values when tracking a target with template updating	A15
Figure 48 - Templates added to the template buffer during the tracking	A15
Figure 49 - Camera Housing Parts	A17

Tables

Table 1 - Cameras in the vision system.....	7
Table 2 - Frame rates for various image enhancement routines.....	45
Table 3 - Camera Specifications.....	A2
Table 4 - Framegrabber Specifications.....	A2
Table 5 - Feature tracking of an underwater target without template updating	A14
Table 6 - Feature tracking of an underwater target with template updating	A16

Chapter 1 Introduction

1.1 Autonomous Underwater Vehicles

Only about 30 percent of the surface area of the Earth is dry land, the remaining 70 percent being covered by oceans. In comparison to dry land, very little is known about the underwater environment, with vast areas of the sea floor never having been explored and charted in a detailed way. As the world's population increases, the worldwide consumption of natural resources will also increase. Thus the exploitation of this unexplored environment will become inevitable. For people to better understand and manage this environment it is vitally important that accurate and detailed information can be obtained.

This is particularly important for Australia, which has a vast area of coastal waters containing a wide variety of unique plant and animal life. This environment must be preserved, not just because it is unique and irreplaceable, but also because it provides a multibillion dollar economic benefit to Australia.

Because of the large area of Australia's coastal seas, it is clearly impractical to manually explore it. However an automated method could be used, and this is one motivation for the development of an autonomous underwater vehicle (AUV). An AUV could be assigned the task of exploring a certain area, and, with limited human intervention, it could perform a detailed survey of that area, and then provide collected and processed data to the operator.

The most important aspect of the system just described is the fact that the AUV performs its task autonomously. Autonomy frees the human operators of the system to perform other important tasks while the submersible performs the routine and time consuming tasks by itself.

Another area where AUVs could be extremely beneficial is in the performance of monitoring and maintenance tasks in environments where it is dangerous or impractical to have humans working. For example, monitoring and cleaning the inside of a nuclear reactor vessel is a situation that is dangerous for human divers. However an AUV can be used in complete safety.



Figure 1 – Autonomous Underwater Vehicle: Kambara

1.2 Underwater Vision

For many years, there has been extensive use of vision systems as a navigational aid for surface vehicles. Nevertheless their use in navigation of underwater vehicles is relatively new. Traditionally, vision systems for surface vehicles have been used to locate features and estimate range to objects, and these functions should be equally useful for underwater vehicles. Underwater vision does, however, give rise to many problems that are not experienced elsewhere.

The first and probably the simplest problem to solve is providing additional protection from the environment that is required for equipment that is to be used underwater. The second and more significant problem concerns poor image quality in underwater environments.

Water tends to absorb light with longer wavelengths (yellow and red), while scattering light with short wavelengths (blue and green), which is why sea water appears blue or blue/green. The amount of absorption depends greatly on the purity of the water, but is also very dependent on the depth of the water. As the depth of water increases, the bandwidth of light that passes through the water decreases.[1] This causes a problem for underwater vision as the absorption of light makes it get dark very quickly as depth increases. Since different frequencies are absorbed at different rates, images tend to lose all colour information in certain parts of the spectrum, which can make some image processing algorithms ineffective.

Image quality can also be adversely affected by the scattering of light due to suspended particles in the water. Large particles such as silt cause light to be scattered in all directions, which causes the contrast in an image to decrease rapidly as depth increases, and it also causes blurring in images, making it hard to distinguish an object from its background.

1.3 System Overview

Currently the ANU is developing an AUV with the goal of getting it to perform simple tasks autonomously. Figure 1 shows the submersible, which has been named Kambara¹. The body of the submersible was designed and built by the University of Sydney, and consists of an aluminium frame 1.5 meters long, 1.3 meters wide, and 0.9 meters high. This frame supports 5 electrically powered propellers, which give the submersible a maximum velocity of 1 meter/second, and a maximum turn rate of 60°/second.

The main body of the submersible consists of two large aluminium tubes, one of which has a clear perspex dome on each end. This gives the submersible a total volume of approximately 117 litres.

The Research School of Information Sciences and Engineering (RSISE) is currently installing an on-board power supply, and a sensor suite, including a depth sensor, velocity sensors and gyros, and a set of three cameras. All of these sensors are attached to an on-board computer, which is a 740 Power PC running VxWorks. This full system is linked to the surface via several optical fibres, which transmit control data in both directions, and also transmit raw video from the on-board cameras. In the future, control and data communication may be accomplished using acoustic modems.

The on-board computer system gives the submersible three main methods of control. First, there is direct control, where the operator on the surface has full control over the vehicle. The second control mode is semiautonomous control, where the operator on the surface shares control with the on-board computer. The remaining control mode is autonomous control, where the on-board computer has full control over the operation of the submersible, without any external interference from the system operator.

For any of these control methods to be effective, accurate and timely information must be produced from the sensors and this thesis concentrates on how vision systems can assist in this task.

The vision system that is used to assist in the control of the vehicle is shown pictorially in Figure 2. The system consists of two Pulnix TMC-73 cameras, which make up a stereo pair, and a single Sony EVI-D30 pan/tilt camera. The two Pulnix cameras will be attached externally to the frame of the submersible, while the Sony camera is to be attached behind the front dome on the top of the submersible.

The Sony camera has a pan/tilt unit built into it, which is controlled using Visca, a language developed by Sony which allows commands to be transmitted over a serial link from a computer.

¹ Australian Aboriginal word for crocodile

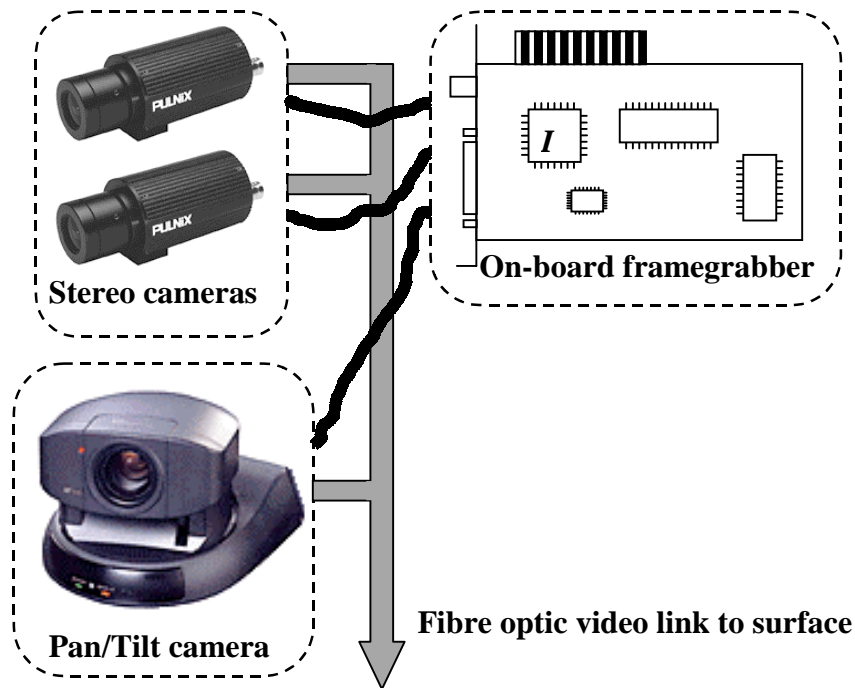


Figure 2 – Autonomous underwater vehicle vision system

1.4 Project Goals

The following broad goals were set for the development of the vision system.

- To plan and implement all of the hardware design required to install the vision system on the submersible. This includes the design of the camera housings to attach a pair of cameras to the frame of the submersible.
- To interface all the hardware with the on-board computer and power supply, as well as with the optical fibre link to the surface.
- To develop software that allows the on-board cameras to be used as a navigational aid for the submersible. This includes the implementation of stereo range estimation and feature tracking.

1.5 Thesis Outline

Chapter 2 outlines the work done in relation to the design and construction of the hardware for the submersible. It includes details of the design for the camera housing for the stereo cameras, including all of the associated cabling, together with details about the choice of camera lenses for the stereo cameras, and the reasons they were chosen. Details of tests that were performed to evaluate the effectiveness of the system are also included.

Chapter 3 gives some background information on some image processing algorithms that are required for the development of the vision system software.

Chapter 4 describes the software that was developed for the vision system. It includes details of the system requirements, and the algorithms that were implemented for this system. Also included are details of all of the tests that were performed to evaluate the effectiveness of both feature tracking and range estimation.

Chapter 5 contains details of suggested further work that can be done to improve the system, and some details on how to implement this work

Chapter 6 gives a summary of the results achieved.

Appendices A to J contain technical drawings for the hardware, as well as derivations for equations and graphs that are contained in the main body of the report. Technical specifications for all of the purchased hardware are also included.

Chapter 2 Vision System Hardware

There were three main areas of work in designing and building the hardware for the vision system. They were the design of the housing for the stereo cameras, determining the optimal lens for the front of the camera housing, and designing and building all of the cabling to link up the various components of the vision system. This work is described below. Work was also done designing and drawing an extension piece for the main body of the submersible. As this is not directly related to the vision system, no details on the design are included. However the technical drawings that were produced in relation to the extension can be found in Appendix I.

2.1 Background

As was described in Section 1.3, the vision system for the submersible consists of three cameras, attached to the on-board computer via a framegrabber card, and also linked to the surface via optical fibres. To install this system on-board the submersible, a number of components had to be built. They are described below.

2.1.1 Stereo Camera Housing

The first and most complex piece of hardware that had to be designed was camera housing for the stereo camera set. For reasons that will be covered in Chapter 3, the stereo cameras had to be separated by a certain fixed distance (called the baseline), and aligned precisely relative to each other. The baseline chosen for the stereo cameras was 300mm, which meant that the cameras could not fit into the main body of the submersible. It was therefore decided that each camera would sit in separate housing that would allow them to be attached to the external frame of the submersible. A decision was made to build the camera housings rather than purchasing off the shelf units. This was because commercially available units are expensive, and did not meet all of the requirements for stereo vision, particularly the requirements of allowing precise alignment of the cameras relative to each other. It was therefore decided to design camera housing that would meet these particular requirements and then have it manufactured at the Department of Engineering's workshop.

2.1.2 Camera Housing Lens

One of the most important aspects of the stereo camera housing was the design for the lens of the housing that the cameras look through into the water. One of the chief difficulties that occurs is due to the different indices of refraction of air and water, which causes distortion to occur in images that are produced through the air/glass/water interface present at the front of the camera housing. Therefore it is important to choose a lens that has good optical properties so as to minimise this distortion, as compensating for lens distortion in software, although possible, wastes valuable CPU resources, reduces the operating speed of the system, and hence diminishes the effectiveness of the vision system. For this reason considerable time was spent determining the optimal lens to use in the camera housing.

2.1.3 Vision System Cabling

For the vision system to work there must be a high quality transmission of images between the cameras and the on-board framegrabber. Table 1 shows the cabling that is required for each of the cameras in the system. The task of making all of these cables was complicated by the fact that the Pulnix cameras and the on-board Imagenation framegrabber each use custom plugs for all their interfacing. Also, since the stereo cameras are attached externally to the frame of the submersible, all of the connections to the Pulnix cameras must be made in a way that prevents water leaking into their housing, or into the main body of the submersible.

Item and description	Connections
2 × Pulnix Cameras For stereo vision. Each in camera housing attached on AUV frame	<ul style="list-style-type: none">• Power supply• Imagenation framegrabber (S video)• Coax to fibre converter (composite)
1 × Sony Pan/Tilt camera For feature tracking and direct operator control. Located in main body of the AUV	<ul style="list-style-type: none">• Power supply• Imagenation framegrabber (S video)• Coax to fibre converter (composite)• Serial port for Visca controls (RS232)

Table 1 - Cameras in the vision system

Sections 2.2 to 2.4 describe the solutions to each of the hardware problems that have been discussed.

2.2 Stereo Camera Housing

As described previously, one of the main design tasks for the submersible was the development of the camera housing of the stereo cameras that would allow the cameras to be attached to the frame of the submersible. The first step in the design process was to work out all of the requirements for the camera housing, and these are described below.

After the system requirements were defined, a design was developed to meet all of the requirements. It is described in Section 2.2.2.

2.2.1 Design Requirements

The following items were identified as requirements for the camera housing.

1. The camera within each housing must be kept dry, while the housing must allow access of up to 2 cables for power and video transmission. (Operating depths are up to approximately 30m).
2. The housing must be large enough to hold the two Pulnix cameras. (see Appendix A for camera specifications)
3. The housing must be able to be attached rigidly to the submersible frame, so that movement of the AUV does not cause the cameras to move independently thereby making feature tracking and range estimation difficult.
4. The cameras must be able to be adjusted within the housing, to align the cameras with respect to the lens, and the camera housing must be easily aligned on the submersible frame so that the stereo camera pair can be accurately calibrated.
5. The weight of the camera housings should be minimised, while the displaced volume should be kept as large as is feasible, to assist in the buoyancy of the AUV.
6. The camera housing must be resistant to damage from corrosion in salt-water etc. Also they should be able to withstand 'rough' handling without getting damaged.

The hardest specification to meet is requirement 4, to allow accurate alignment of the camera relative to the housing lens, and then to allow each camera housing to be accurately adjusted relative to one another.

Figure 3 shows the degrees of freedom that are required for the camera relative to the camera housing lens. The diagram shows that the camera needs to be able to translate in all directions (X , Y , Z) while it only needs to be able to rotate around the X and Y axes.

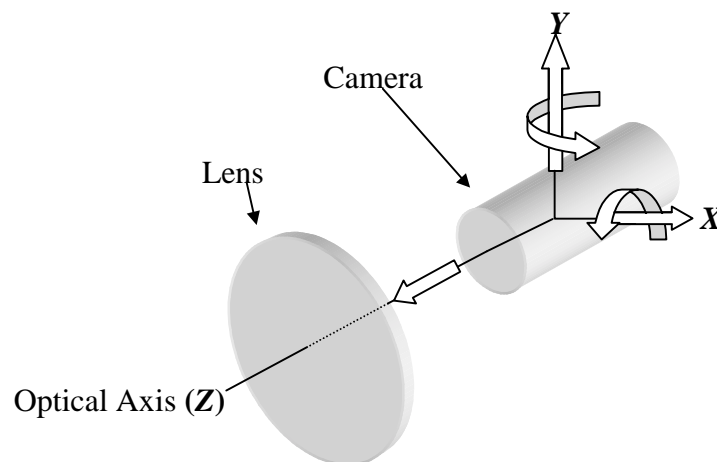


Figure 3 - Degrees of freedom of stereo camera within its housing

The reason for the translation motion is so that the centreline of the cameras can be aligned along the medial central axis of the camera housing lens, and also the optical axis is perpendicular to the face of the lens. The reason for this becomes apparent when the lens on the front of the camera case is changed from being say a planar to a spherical lens. In this case, if the camera is not aligned perfectly to the centre of the lens then a non symmetric distortion will appear in the output image.

A requirement for the stereo vision system is that the orientation of the cameras optical axes can be adjusted precisely. This allows for accurate range estimations to be made. Figure 4 shows the translations and rotations that are required for the camera housings relative to the submersible frame. In this diagram the cameras are pointing in the $+Z$ direction. The translation shown in the x direction is required so that the size of the baseline of the cameras can be controlled.

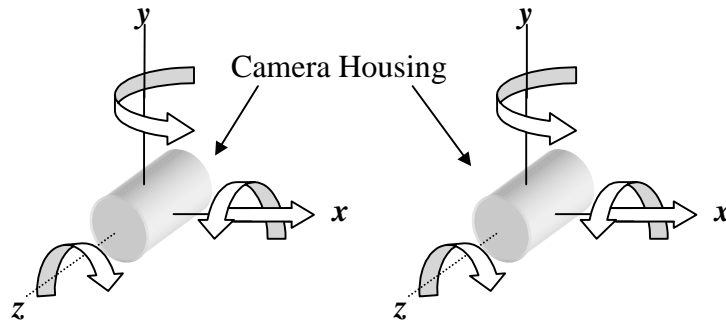


Figure 4 - Degrees of freedom of stereo camera housings

The degrees of freedom that are required by the camera housing are decided when the system is set up. Once this is done the two cameras are attached rigidly to the submersible. This means that all of the joints in the housing that allow for movement of the camera must be capable of being tightened to minimise the probability of accidental displacement.

2.2.2 Solution

Figure 5 shows the housing that was eventually designed and manufactured. This camera housing incorporates all of the required degrees of freedom, both for motion of the camera within the housing, and for motion of the housing on the submersible frame. A full set of technical diagrams can be found in Appendix G.

The camera inside the housing is attached to two plates that can rotate about the X and Z axes, and allow it to translate in the X and Y directions. The two plates are attached to a track that allows translation in the Z direction. To adjust the camera relative to the lens, the track to which these plates and the lens are attached can be removed from the rest of the housing.

The actual housing itself is attached to the submersible frame on a pivot that allows translation along the AUV frame to control the baseline length, and allows rotation around

the X and Y axes. Rotation about the Z axis is achieved by rotating the full camera housing within the loop that clamps it to the submersible's frame.

To minimise the weight of the camera housings, they were made out of aluminium 6061 which is light, has good machining, welding and anodising properties, and is corrosion resistant. After machining the housings were anodised to protect them from corrosion in a salt water environment.

Waterproofing of the housing was achieved using a set of rubber O-Rings. The points of entry for cables were kept watertight by using a set of metal fibre glands that are also used extensively elsewhere on the submersible.

The housings were made sufficiently large to accommodate the Pulnix cameras and also so that alternative larger Sony cameras could be housed instead.

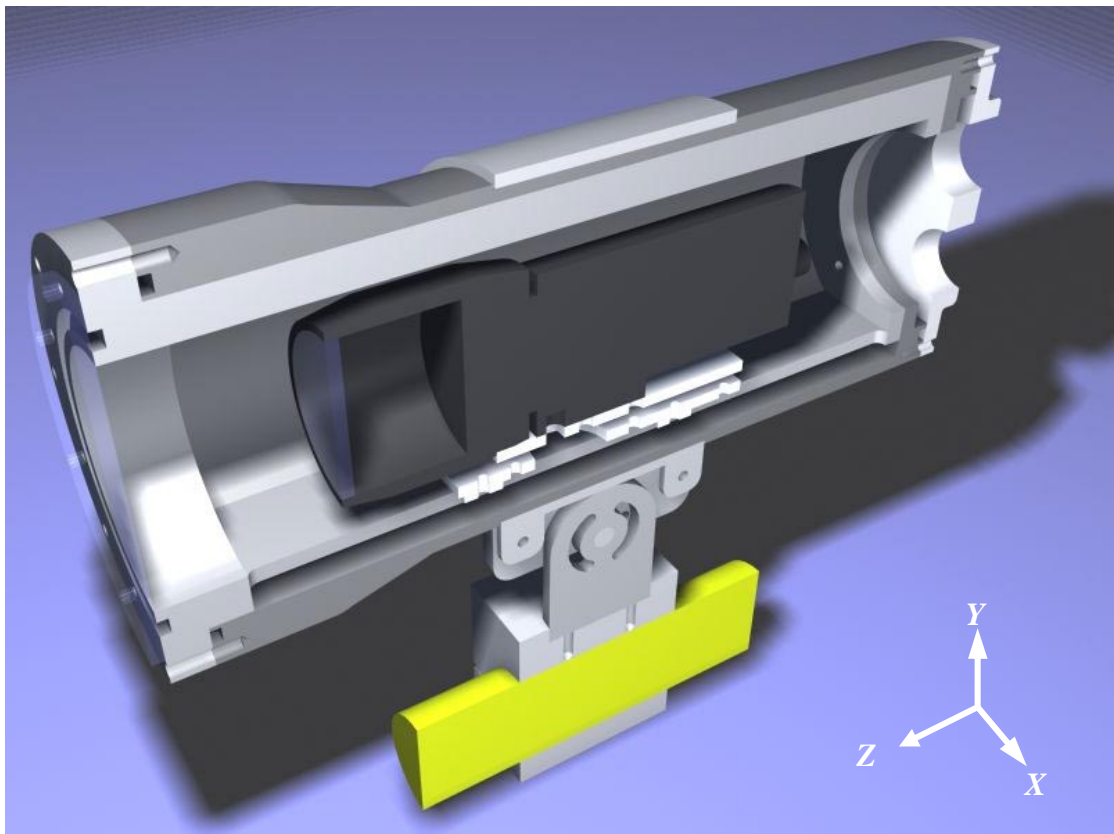


Figure 5 - Section view of camera housing

2.3 Lenses for Camera Housing and Cameras

When light passes through the interface between two materials that have different indices of refraction, the light is 'bent' due to the fact that it is travelling at different speeds in each material. This can be explained using Huygens' principle, which models light as waves, and states that:

*All points on a wave front serve as point sources of the spherical secondary wavelets.
After a time t , the new position of the wave front will be that of the surface tangent to these secondary wavelets.*

Using this principle Snell's Law can be derived, which explains how light behaves at the interface between two materials with different indices of refraction. Snell's Law is given in equation 2.1, and its derivation can be found in Appendix B.

$$N_1 \sin \theta_1 = N_2 \sin \theta_2 \quad (2.1)$$

Where N_1 and N_2 are the indices of the two materials, and θ_1 and θ_2 are the angles of incidence and refraction, measured from the normal to the material interface.

Snell's law illustrates that when light passes from one material to another, if the indices of refraction are different, then the light will be bent, which can cause distortion in images. At the air/glass/water interface there will be distortion due to the different indices of refraction of each of these materials.

It is therefore important to design the camera housing lenses to minimise this distortion.

2.3.1 Choice of Camera Housing Lenses

If a planar interface were to be used, Snell's law tells us that along the optical axis of the camera there will be no distortion due to the housing lens, but as the angle of view increases away from the optical axis, the level of distortion increases. The reason for the distortion is the non-linear relationship between the angle of incidence of light, and its angle of refraction. This is defined by equation 2.3.5 and shown in Figure 6. (see Appendix C for details)

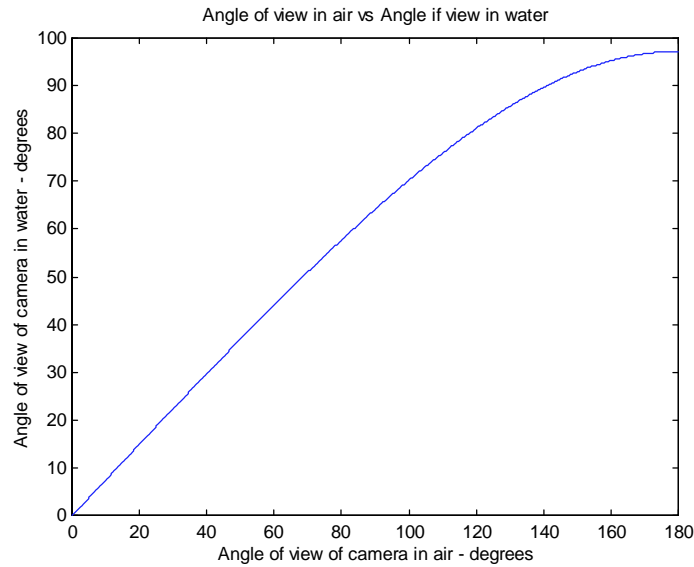


Figure 6 - Distortion at water/glass/air interface due to light bending

The graph in Figure 6 illustrates that for camera angles of view of up to about 100° the planar lens will produce relatively linear results, meaning that there will be minimal

distortion at the edges of the image. However, if the angle of view increases beyond this, the level of distortion becomes significantly worse. One possible solution to this problem, if a very large field of view is required, is to use a spherical lens for each housing. By placing the focal point of the camera at the centre of curvature of the lens, essentially the light must always pass through the lens perpendicular to its surface. This will mean that there is no distortion present in the image.

For the submersible, planar lenses were chosen because of the ease with which they can be machined, and because it was considered that 100° was a sufficiently large field of view for the stereo cameras. The principal reason why 100° was considered a sufficient field of view is because the output of the feature tracking and range estimation systems will be used to control the submersible's motion, and the motion that it produces will be directed at keeping the target in the centre of the images. A second reason for not using spherical lens is because of the difficulty and expense of machining spherical lenses with good optical properties.

Both of the lenses for the stereo camera housing are to be made out of polycarbonate, because of the ease of machining, it's material strength, and the good optical properties.

2.3.2 Choice of Stereo Camera Lenses

The Pulnix cameras that were used for the stereo rig were not supplied with lenses. Therefore it was important to choose the optimal lenses for the vision system. This was done with careful consideration of the size of the camera housing, the target field of view for the vision system, and the type of lenses that were attached to the camera housing. Since we are using planar lenses for the camera housing, it was possible to calculate the ideal focal length of the camera lenses to best suit the system. To perform this calculation the following assumptions were made.

- The lens on the front of each camera housing is planar.
- The vision system must track and estimate range to objects at a distance of 2m or less from the front of the submersible.
- An area 2m wide and 2m in front of the submersible must be visible in any single video frame.

Using this information it is possible to calculate that the optimum focal length of the stereo camera lens is 3.35mm. (see Appendix D for details). Lenses of this size are not standard, the closest available lenses having focal lengths of 2.8mm and 4.0mm.

The smaller the focal length, the larger the field of view of the camera. To ensure the requirement of being able to see an area $2 \times 2 \text{ m}^2$ in front of the camera meant that the best lens was the 2.8mm lens.

An area of concern when using such wide angle lenses is that the edges of the images will be distorted. This distortion is due to the spherical shape of the lens that is needed to

produce a short focal length. Figure 7 shows the images produced by the two different camera lenses, both looking at the same graph paper. The figure shows the borders of the image produced by the 2.8mm lens having significant distortion, which is not present in the image produced by the 4.0mm lens. Distortion was not considered to be a major problem since the outer edges of the images could be cropped, and all processing can occur on the relatively undistorted centre section of the image.

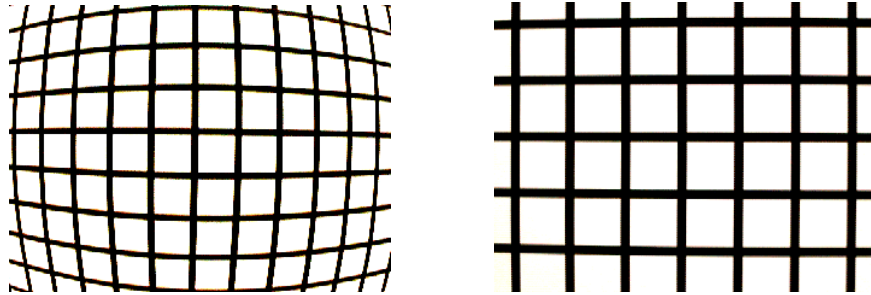


Figure 7 - Images from lenses with 2.8mm (left) and 4.0mm (right) focal lengths

Once it was determined that the 2.8 mm lens was suitable, it was possible to calculate how far from the centre of the camera housing lens the focal point of the 2.8mm lens must be placed to get the required 2m×2m field of view. We calculated this to be 35mm. (See Appendix E for details) Using the camera housing diagrams and the specifications of the Pulnix cameras it is possible to see that the camera will fit into the camera housing with the required amount of clearance.

2.4 Data and Video Cables

The final area of hardware design was to plan and build the cabling system for the vision system. Figure 8 shows the cabling that was required for the vision system, along with the hardware components that the cabling is linking. (Grey areas indicate existing hardware)

For the benefit of on-board image processing, considerable care was taken to minimise noise in all of the video signals that must be transmitted. This was done by using high quality components and by minimising the number of plugs and connectors used, as each plug/socket connection degrades the signal considerably.

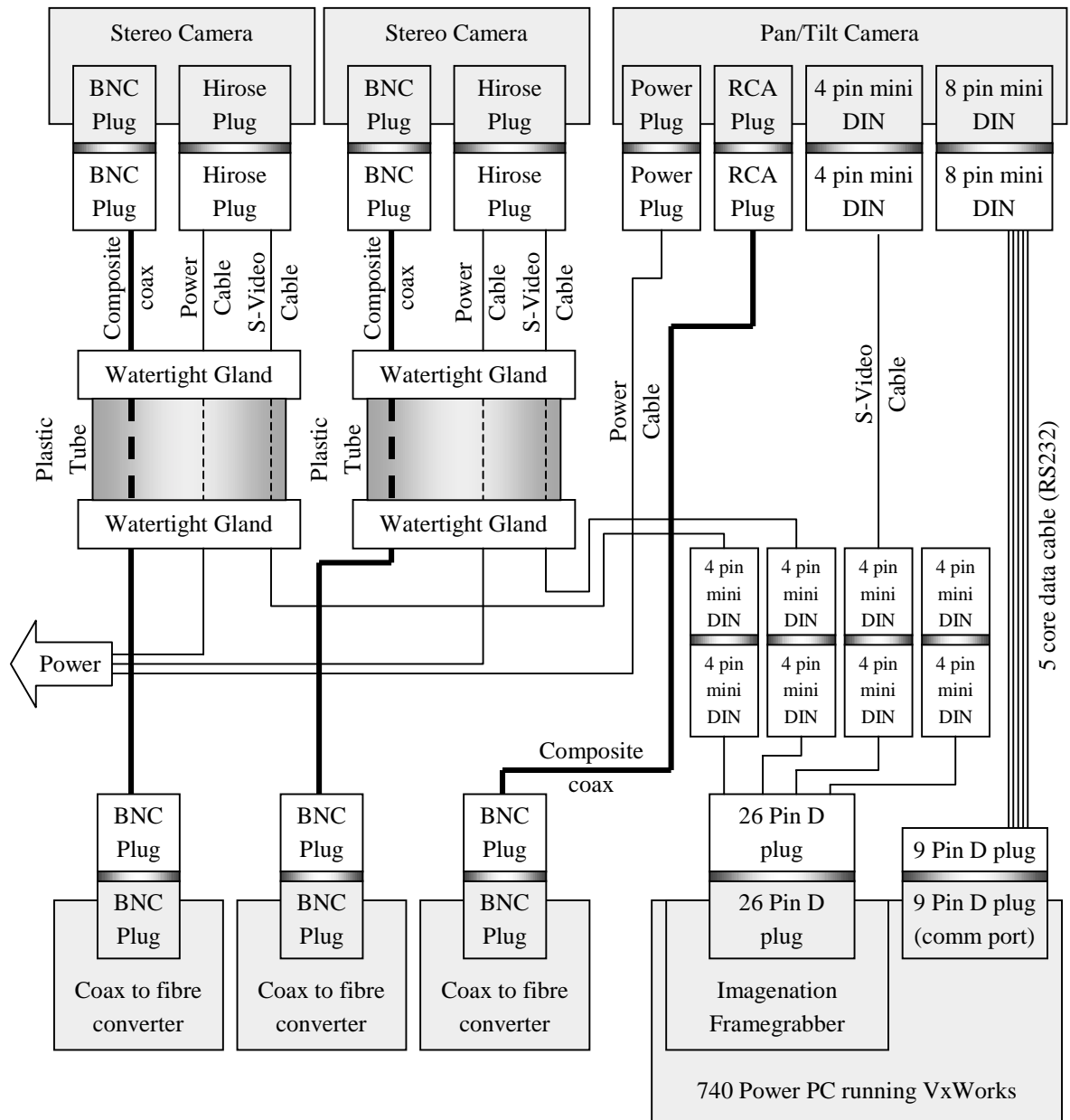


Figure 8 - Cabling diagram for complete vision system

2.5 Conclusion

All of the required hardware components for the vision system were designed and built. The camera housing was designed to allow the necessary degrees of freedom of movement, and to ensure that the cameras can be aligned easily and precisely relative to the housing lens. The camera housing can also be adjusted on the frame of the submersible so that the cameras can be calibrated as a stereo set. In addition all of the waterproofing and resistance to environmental damage requirements were met by making the casing out of aluminium that was anodised, and by using a series of rubber o-rings to ensure a watertight seal to protect the contents.

Our analysis has shown that the lens in the camera housing should ideally be spherical to minimise the distortion that is caused by light bending at the lens interface. However a planar lens was shown not to significantly increase distortion levels, and was much cheaper and easier to manufacture.

To get the required 2m×2m field of view for the stereo cameras, it was calculated that a 2.8mm camera lens would be the optimal choice for the stereo cameras. Lenses with this focal length give a wide field of view. However, at the edges of the images, the lenses produce a large amount of distortion. We believe that the cost of compensating for the distortion is less than the benefits gained from having a larger field of view.

To connect all of the cameras to the on-board computer components a significant amount of cabling had to be done. This was carefully designed in order to minimise the noise that was introduced into the signals. Upon completion this design was found to perform satisfactorily.

Chapter 3 Image Processing

This chapter outlines some image processing routines that are required for the stereo range estimation and feature tracking systems that are to be implemented.

An important aspect for both stereo range estimation and feature tracking is the algorithm that is used to track a target as it moves over a video stream. There are many ways to do this, and the one chosen to implement is called template matching, which is described later in the chapter.

However, because underwater vision suffers from many problems associated with poor image quality due to blurring and poor light levels, this chapter first examines possible means of improving the underwater image. There are several image enhancement methods available that are able to compensate for some of this lost image quality, and this chapter begins by examining some of those methods.

3.1 Image Enhancement

One method of image enhancement that can be useful underwater is edge detection. This involves locating all the places in an image where there is an abrupt change in colour. Edge detection is useful because it can compensate for blurring that results from light scattering due to particles in the water. Blurring is a problem for computer vision systems because it makes it difficult for automated systems to distinguish objects from their background.

3.1.1 Edge Detection using Gradients

A method for detecting edges in a picture is to examine that rate at which colours change in the image, which can be done by finding the derivative of the picture. An example of this is shown in Figure 9.

The image initially consists of a single white line on a black background. When the derivative is taken of that image, peaks occur at the edges of the white line. Taking the absolute value of this derivative gives the edges of the picture.

A simple method for approximating the first derivative is to apply equation 3.1 to all pixels in the image

$$deriv_{i,j} = \max(p_{i+m,j+n}) - \min(p_{i+m,j+n}) \quad (3.1)$$

where $m = [-1, 0, 1]$ and $n = [-1, 0, 1]$

This method examines a 3×3 mask, subtracts the minimum value in that mask from the maximum value, and records the difference in the centre of the mask.

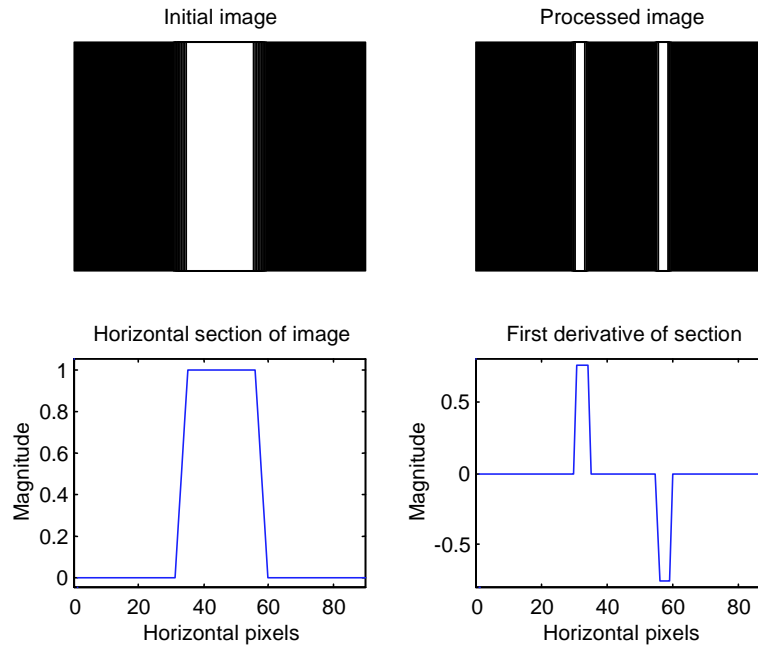


Figure 9 - Example of the first derivative of an image

This algorithm is effective at detecting edges but it is very sensitive to noise. As mentioned earlier, in underwater images a large amount of blurring occurs. This means that there are no sharp edges in the image and this limits the magnitude of the peaks on the first derivative picture. If there is noise in the unprocessed image, due to poor quality cabling or plugs between the camera and the framegrabber, then the resultant noise will not exhibit the same blurring as occurs uniformly over the rest of the image. This means that when the derivative is taken, the ‘edges’ around the noise are much more prominent than the actual edges in the image. Figure 10 illustrates this effect.

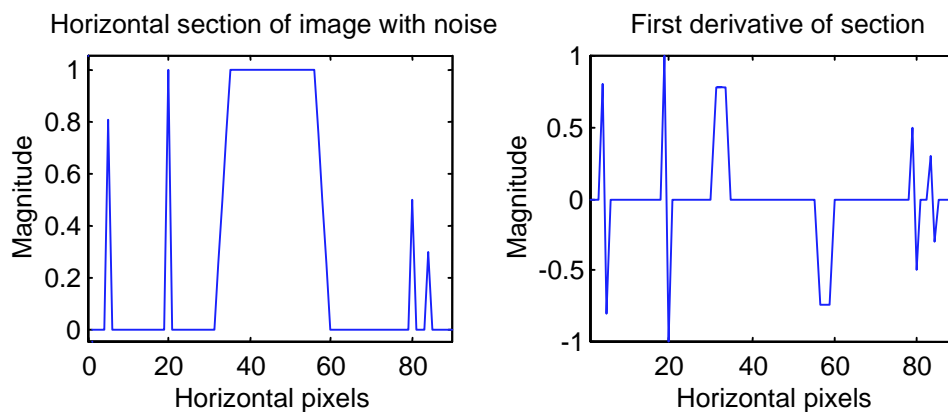


Figure 10 - Effects of noise on first derivative edge detection

Figure 10 uses the same section of the image used in Figure 9. However there are now 4 points of noise added to the image. The graph of the first derivative shows that the noise causes very large spikes to appear, making recognition of the real edges difficult.

This high sensitivity to noise reduces the effectiveness of this type of edge detection, so other image enhancing methods were examined. One possibility is the difference of Gaussians method, which is described next.

3.1.2 Edge Detection using Difference of Gaussians

Another technique for enhancing an image that can be used for edge detection purposes is to use the difference of Gaussians. This method involves performing two separate convolutions of the initial image with two different Gaussian surfaces, and then subtracting the result of one convolution from the other. This ends up producing a result similar to the second derivative of the image, which significantly enhances the suitability of images for template matching. The graphs in Figure 11 illustrate the calculation of the difference of Gaussians for the same image section used in Section 3.1.1.

We start with the original section of the image, and two Gaussian curves of different sizes (graphs (1), (2) and (3)). The initial image is convolved with each of the Gaussian distributions, shown in graphs (4) and (5). The resulting image that is produced using the small Gaussian curve is subtracted from the other image, to produce the difference of Gaussian image shown in graph (6).

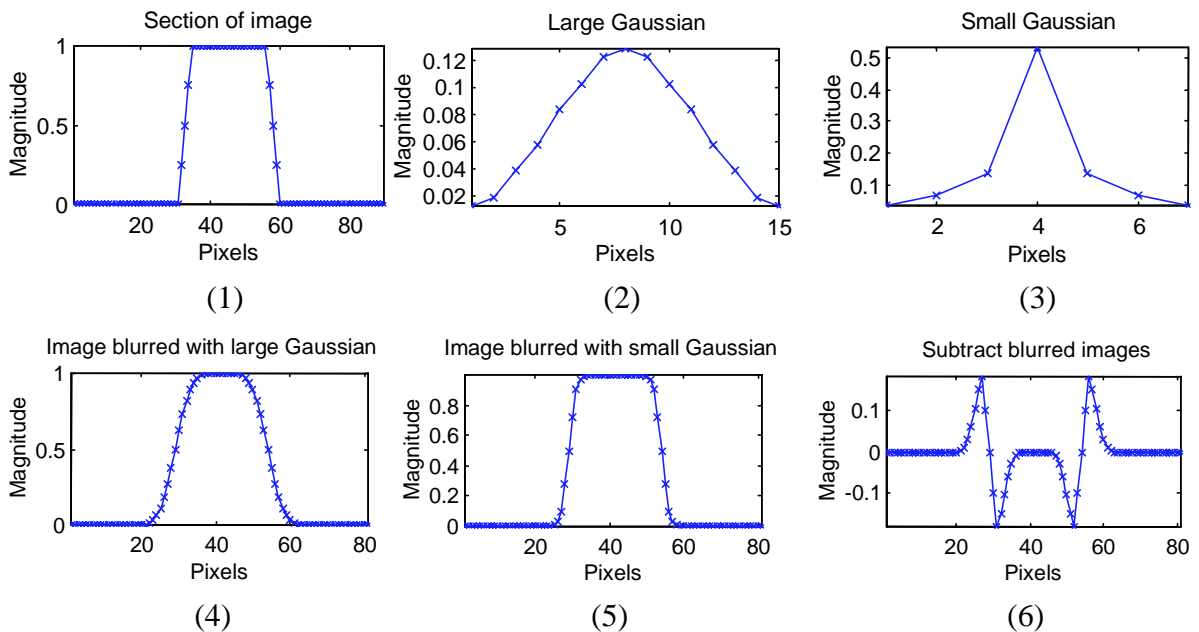


Figure 11 - Calculation of the difference of Gaussian of an image

This difference of Gaussian method is similar to the second derivative of the initial picture, however there is one important difference. The difference of Gaussian method is less sensitive to noise than a second derivative. The reason for this reduced sensitivity is because of the initial convolutions that were performed with the Gaussian distributions.

These convolutions have the effect of blurring the image, reducing the effect of any noise that was present in the image. The subtraction that is performed after these convolutions further decreases the magnitude of the noise, as the noise from one image is subtracted out of the other.

Figure 12 shows the effect of how the magnitude of the noise relative to the magnitude of the actual edges in the image in the difference of Gaussian method is much smaller than in the second derivative method.

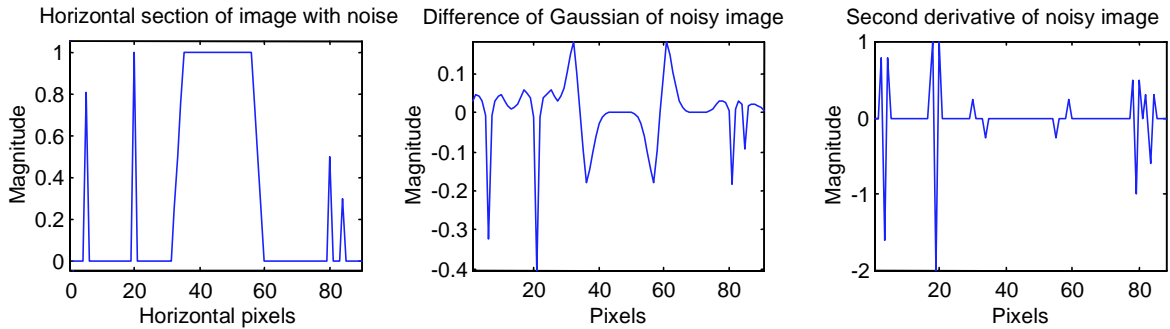


Figure 12 - Second derivative and difference of Gaussian of a noisy image

The noise minimisation feature of the difference of Gaussians made it the preferable image enhancement method to trail in conjunction with the chosen technique for tracking an underwater target, namely template matching, which is described in the following section.

3.2 Template Matching

There are a number of different techniques for tracking an object as it moves in a stream of images, and template matching is one possibility. Template matching involves extracting a small template of a region of interest in one image, and then trying to locate that template in all subsequent images. To locate the template in all future images the template is correlated with every position that it could have moved to in the image. This correlation examines how similar the template is to the original image that was recorded. After the correlation has been calculated for all possible regions where the template could be located, the position with the best correlation value is considered to be the new location of the template. This location becomes the new position of the region of interest. The following example illustrates how this is done.

Figure 13 shows the initial image, which contains a fish that we wish to track as it swims over a striped background.

The first step in template matching is to extract an initial template of the object that is to be tracked. Figure 13 shows a small section of the initial image that is cut away to become the template.

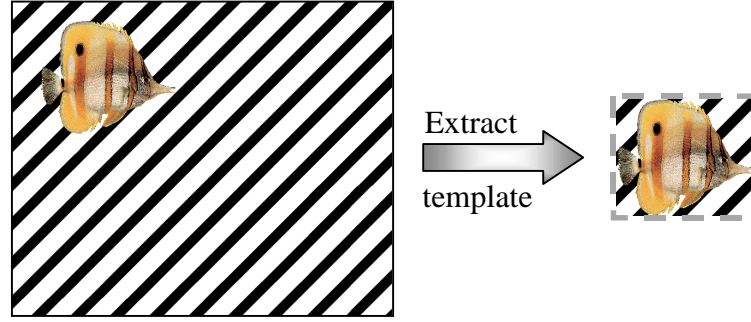


Figure 13 - Extraction of a template containing the target

At a later time the target (the fish) moves to a different location, as is shown in Figure 14. To locate the new position of the fish, the template that was extracted earlier is now correlated over the entire image, and the correlation values that the template has with the underlying image are recorded.

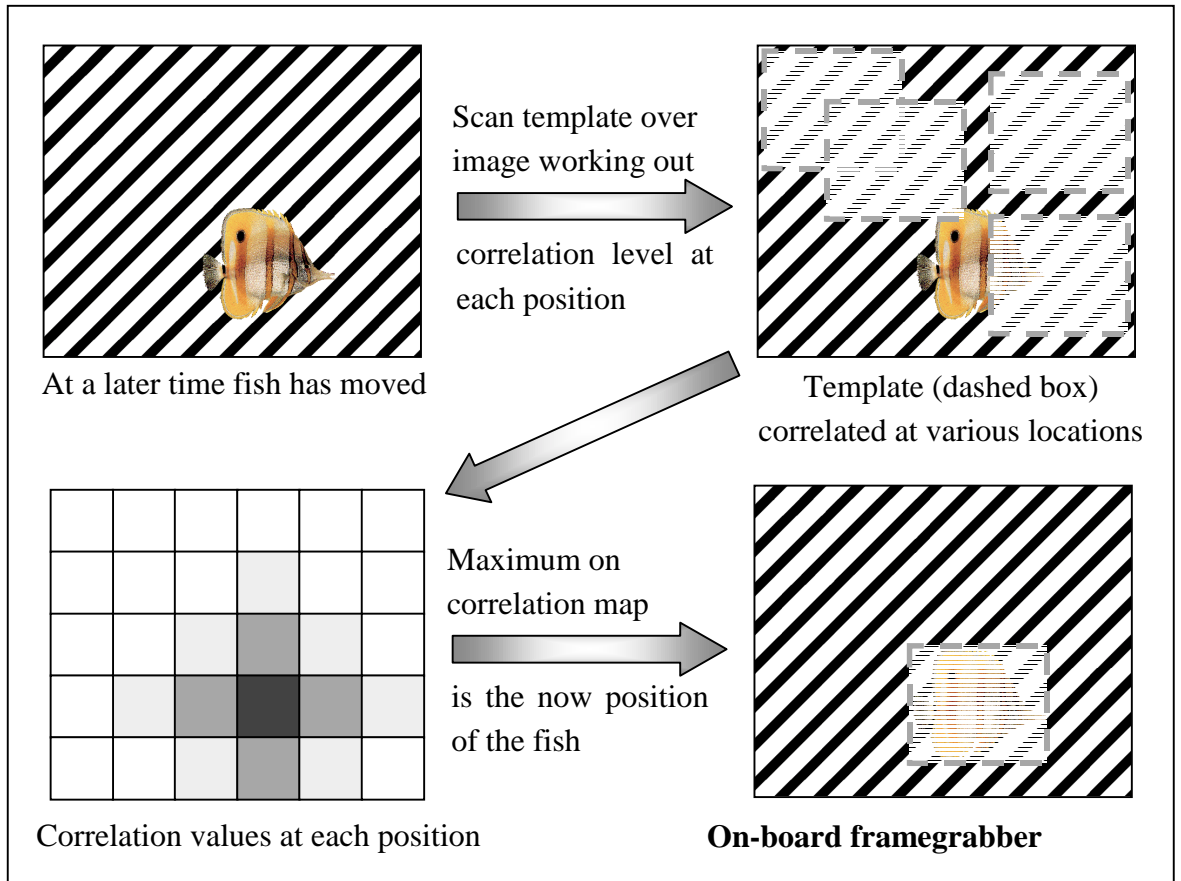


Figure 14 - Correlation of the template on a new image

The correlation value at any position (i,j) in the image can be calculated according to the equation 3.2, where m and n are the x and y dimensions of the template.

$$Correlation_{i,j} = \sum_m^{x=-m} \sum_n^{y=-n} |image_{i-x,j-y} - template_{m-x,n-y}| \quad (3.2)$$

The dimensions of the template are $2m \times 2n$, and the subscripts after *image* and *template* refer to specific pixels in the image and the template respectively.

This equation shows that the correlation value of the template at any position in the image corresponds to the sum of the absolute difference of each pixel in the template and the corresponding pixel of the image.

Using Equation 3.2 a ‘map’ of the correlation values of the template at every position in the later image can be generated. The magnitude of each point in this map corresponds to the degree of correlation of the template and the image at this position. As equation 3.2 illustrates, the better the template matches at any position, the lower the correlation value will be. A correlation of zero will correspond to a perfect match, and the worst possible correlation level for a $2m \times 2n$ template will be equal to:

$$\text{Worst correlation value} = 2m \times 2n \times 2^{(\text{bits per pixel})}$$

Once the correlation map has been produced, the location of the minimum value in this map corresponds to the new location of the target fish.

There are several problems which must be addressed for this system to work. The first has to do with the speed at which the object that is being tracked is changing. For example, if the fish that was being tracked in the last example, changed directions so that it was now facing the camera, the correlation value between the fish and the template would become poor, resulting in the target being lost. A similar problem develops if the camera and the target get closer together or further apart, since the size of the target in any image will get smaller or larger, while the template will remain unchanged. This results in gradually decaying correlation values, until the target is lost.

One method to overcome this sort of problem is to set a correlation threshold value, and if the template cannot be found anywhere on the image with a correlation value less than this threshold value, then the template is considered invalid, and should be replaced. The problem is that once the template cannot be located in the image, the target will then be lost, and new templates cannot be extracted. One solution to this is to maintain a buffer of templates, so that if the current template can no longer be found in an image, it is discarded, and a fresh template from the buffer is used. Hopefully, the new template will then be found with a correlation value below the threshold value, otherwise the process is repeated until the buffer is empty, at which point the target is lost. The concept of keeping a template buffer will be developed further in Section 4.5.2.

3.3 Image Processing Testing

The tests outlined in this section were undertaken to verify that the implementation of the image processing routines described in this chapter produced results that matched the

theoretical expectations. Further tests in Chapter 4 will evaluate the effectiveness of these routines in a real time situation.

3.3.1 Edge Detection test

Edge enhancement was implemented using a difference of Gaussians method. The image was initially convolved with the 3×3 mask shown in Figure 15, and then the initial image was subtracted from this blurred image.

1	2	1
2	4	2
1	2	1

Figure 15 - Mask used for Gaussian blurring

Thresholding was then performed on the image to highlight all of the edges. Varying the threshold value was found to be effective in changing how well the edge enhancement worked. For example, in the pictures shown in Figure 16, the same difference of Gaussian operator was applied, but different levels of thresholding were used which gave quite different results.

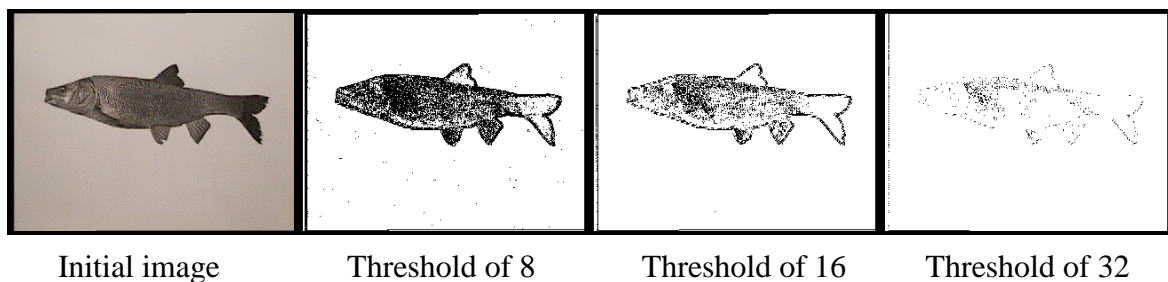


Figure 16 - Difference of Gaussians with various threshold levels

These results illustrate the importance of calibrating the edge enhancement system properly, as threshold values that are too high can be seen to remove an excessive amount of image data and values that are too low do not adequately highlight the edges.

3.3.2 Template Matching Tests

Correct operation of the template matching routine was tested by creating a simple template of a black spot on a white background, and then correlating it with the initial image it was extracted from. Figure 17 shows the initial image that the template was extracted from, as well as a copy of the template that was extracted. A graph of the correlation map for the region of the image where the template was extracted is also shown.

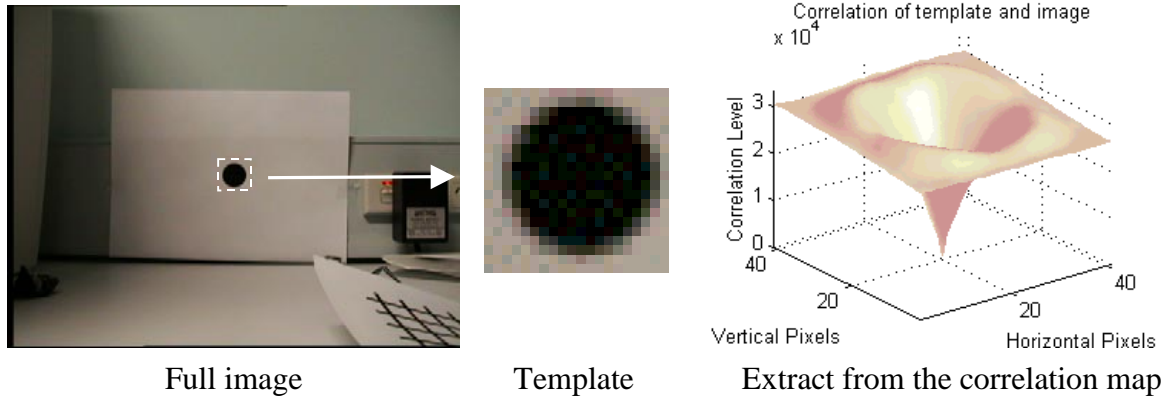


Figure 17 - Correlation of a template with the image it originates from

The graph shows that when the template is being correlated with the areas of the image around the spot, the correlation values are relatively constant at a level of about 30000. This 30000 corresponds to the correlation value of the black spot over the white background.

Once the template is correlated in a position where it partly overlaps the circle on the image, there is a region where the correlation value actually gets worse (i.e. a higher value). This corresponds to the additional mismatch that is produced from the white border of the template overlapping the black spot in the image. Once the black regions of the template and the image start to overlap, the correlation values start to decrease, until they reach zero. This corresponds to the position where the template is exactly over the location from which it was extracted.

The graph of the correlation matrix indicates that template matching is operating correctly, and that it matches with the theoretical predictions.

3.4 Conclusion

It was decided that edge detection using difference of Gaussians was the most suitable method for enhancing underwater images. The reason why edge detection was chosen was to compensate for blurring that occurs in all underwater images, due to light scattering off suspended particles.

Template matching is a technique for tracking moving targets in an image, and works by comparison of a small template of the target object with all parts of the image that the target is sought to be located.

Both difference of Gaussians and template matching were implemented, and found to provide results that agreed with theoretical expectations.

Chapter 4 Software Design

There are two main functions for the software that was designed. To estimate the range to objects in front of the AUV, and to track features as they move relative to the AUV. The output of the system can assist in the navigation of the vehicle. Figure 18 shows how the vision system fits in with the rest of the AUV's software.

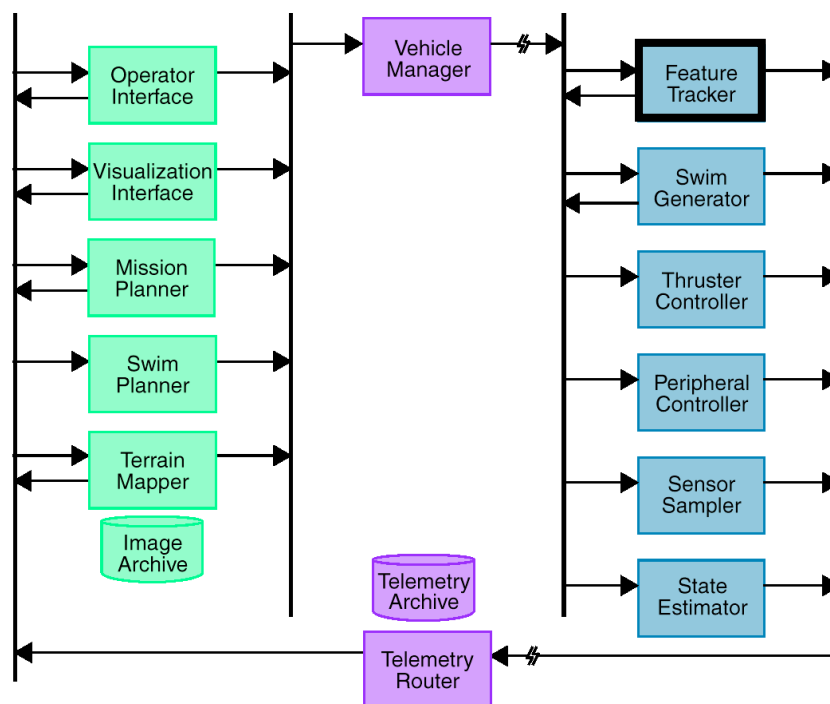


Figure 18 - Kambara's software architecture [8]

The left side of Figure 18 shows the parts of the AUV's software that are located on the surface, while the right side shows the on-board software. The diagram shows that the Feature Tracker provides feedback directly to the other on-board systems, as well as providing data to the surface. The information that is fed back from the Feature Tracker may be translated by the Swim Generator into thruster control signals, or even used to control other on-board devices via the peripheral controller.

As previously described, the submersible runs using the VxWorks operating system, therefore all of the software should be designed to run under VxWorks. However, because drivers for the framegrabber to run under VxWorks were not yet available, all development was done under a Windows NT operating system. For this reason all of the software was written into a test program that the framegrabber manufacturer provided for Windows NT.

This allowed testing of the effectiveness of the algorithms. All of the image processing code was written so that it be easily be converted to VxWorks at a later time, when the device drivers become available. Figure 19 below shows the output from the program that was developed for testing the vision algorithms. The initial program provided by Imagination only showed one image from a single framegrabber. It was modified so that it would show both images from a stereo pair, as well as display the results from the various image processing algorithms. The program allows the images from the cameras to be paused, and saved as bitmap images. It also allows for the calculation of the frame rate at which images are being processed, and for swapping between two video input sources.

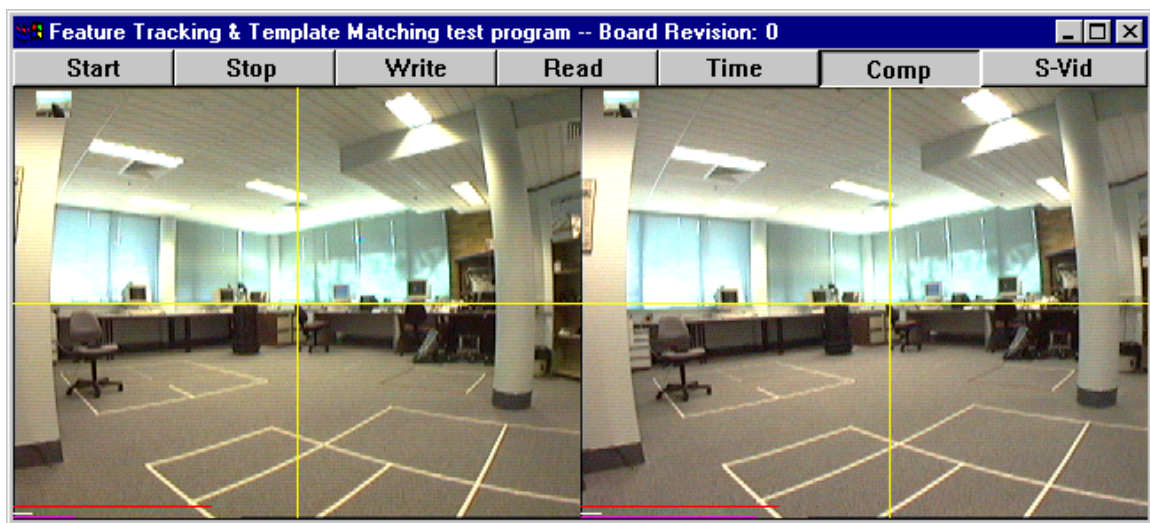


Figure 19 - Image processing test program interface

The Imagination framegrabber which will be installed on-board the submersible supports up to four video inputs, and allows for fast swapping between the inputs. This framegrabber is only available in a compact PCI format, which is incompatible with the system that was used for testing. To simulate the final target system, two PCI framegrabbers were installed in the testing computer, with each one allowing only a single input. Fast swapping between the framegrabbers was possible, which meant that the all the stereo algorithms could be tested.

It was initially proposed to devote some time to the development of drivers to control the motion of the Sony pan/tilt camera. However, it was found that drivers for the operation of the pan/tilt unit already existed under VxWorks and so only a small amount of time was needed to test their functionality.

4.1 Background

The use of the template matching routine described in Section 3.2 makes it possible to track a target as it moves over an image. This chapter describes how this template matching routine can be used to implement the range estimation and feature tracking

systems planned for the AUV, and begins with an explanation of how range estimation and feature tracking work.

4.1.1 Range Estimation

Range estimation involves the use of a pair of stereo cameras separated by a known distance to estimate distances from the front of the cameras to an object of interest. Figure 20 illustrates how this is done.

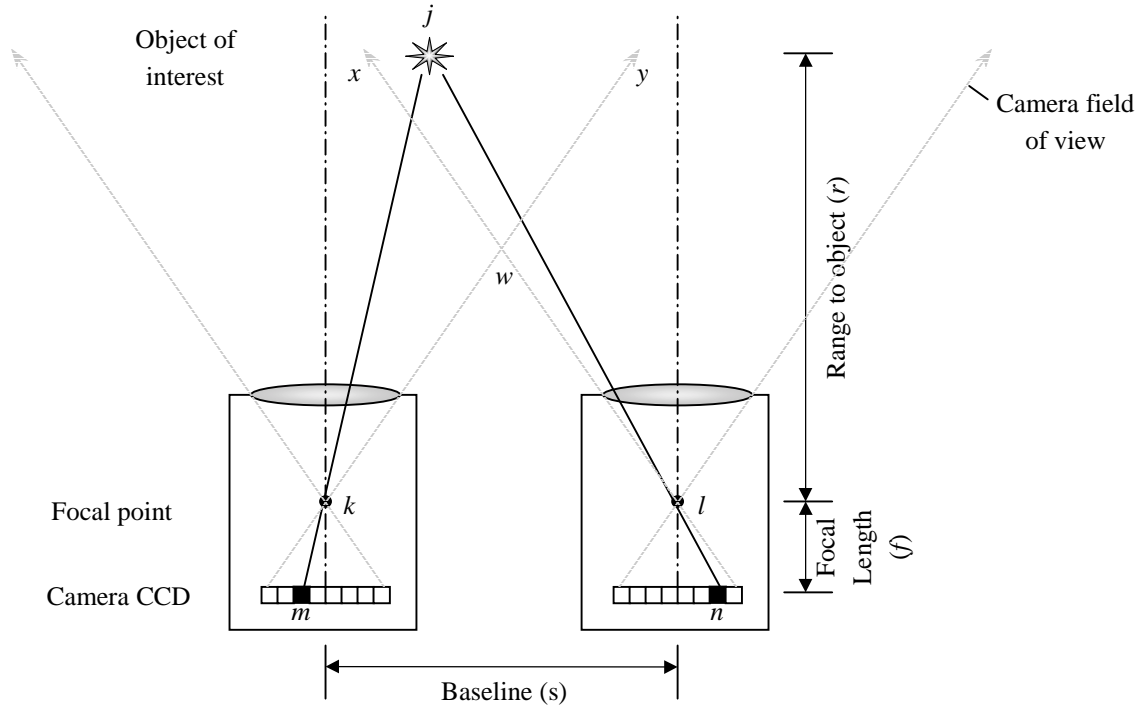


Figure 20 - Stereo range estimation using parallel cameras

In Figure 20, the two cameras are separated by distance s (called the baseline), the focal length f of each of the cameras is known, and the system must estimate the range to an object at an unknown distance r from the front of the submersible. By setting up the two cameras so that they are exactly parallel, similar triangles can be used to determine the range to the object (see Figure 21). The reason that this is possible is because the image of the object will appear on different pixels on each cameras CCD. By calculating the difference in the position in which the object appears in each CCD (the disparity), and using this with the known dimensions of each pixel, the following estimation can be made about the objects range.

$$\frac{\text{baseline}}{\text{Range to object}} = \frac{\text{Pixel disparity} \times \text{pixel dimension}}{\text{focal length}}$$

$$\therefore \text{Range to object}(r) = \frac{\text{baseline}(s) \times \text{focal length}(f)}{\text{Pixel disparity} \times \text{pixel dimension}} \quad (4.1)$$

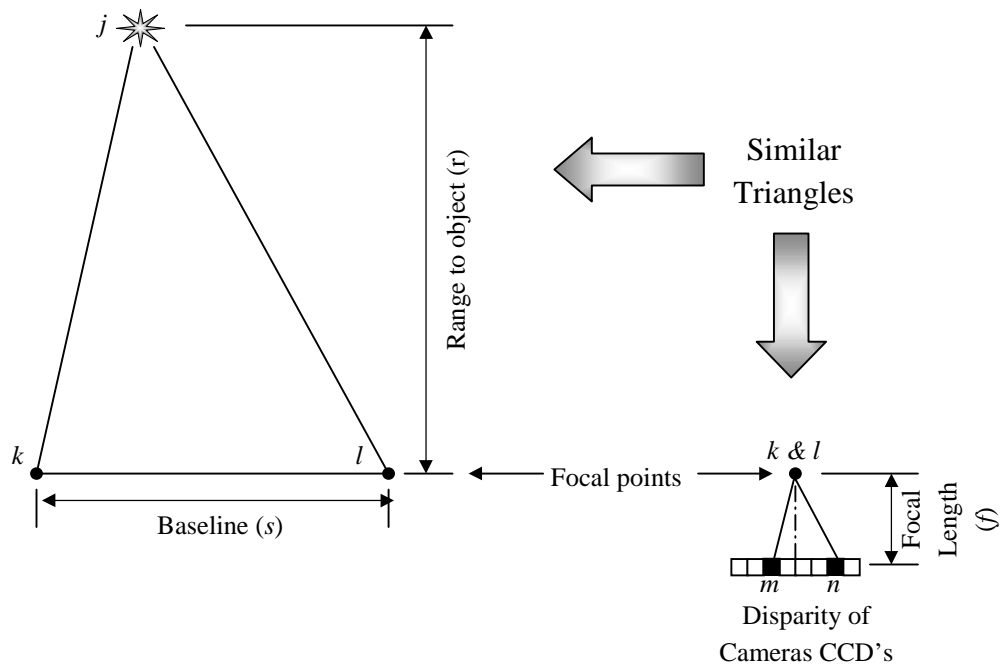


Figure 21 - Similar triangles used to estimate range for parallel cameras

From this analysis it can be seen that increasing the baseline of the cameras increases the pixel disparity. Since the pixel dimension is fixed, and the focal length is fixed, the effect of increasing the baseline will be to decrease the effects of small errors that may occur in the disparity calculation. (i.e. an error of 1 pixel will have a smaller effect on the range estimation for a system that has a larger separation distance.) This would suggest that the baseline should be as wide as possible. Theoretically, this would be true, but a problem develops due to the limited field of view of each camera. In Figure 20 only objects that are within triangle $w-x-y$ can be seen by both of the cameras, and so, as the baseline increases, the distance between this triangle ($w-x-y$) and the front of the cameras increases. To some extent this problem can be overcome by using lenses with very large fields of view, but lenses with very large fields of view generally have high distortion around their edges, thereby introducing other problems.

Another problem that develops when the length of the baseline increases is that template matching on both images gets increasingly difficult. This is because cameras that are separated by a large distance will be each viewing the target from a very different angle. Consequently, a template that correlates well on one camera's image may give a poor correlation on the other, as the target will present differently when viewed from different angles.

One difficulty with the system just described is the fact that each of the cameras needs must be adjusted so that they are exactly parallel for the equation to be correct. Setting and maintaining cameras to be exactly parallel is difficult, but it is also possible to perform range estimation using cameras that are not parallel. Figure 22 illustrates a system where

the cameras are not parallel, having their optical axes converging at a point D_c from the baseline.

Appendix F gives the derivation of equation 4.2. The equation can be used to estimate the range to an object in front of the convergent system if the distance to the convergence point of the optical axes is known.

$$r = s \times \left\{ \frac{1}{\tan^{-1} \left(\tan^{-1} \left(\frac{2D_c}{s} \right) - \tan^{-1} \left(\frac{d_1}{f} \right) \right)} + \frac{1}{\tan^{-1} \left(\tan^{-1} \left(\frac{2D_c}{s} \right) - \tan^{-1} \left(\frac{d_2}{f} \right) \right)} \right\}^{-1} \quad (4.2)$$

The location of this point can be found by measuring the range to an object that appears on the same pixels of the image from each camera. More details about how to do this calibration will appear in Section 4.3.1.

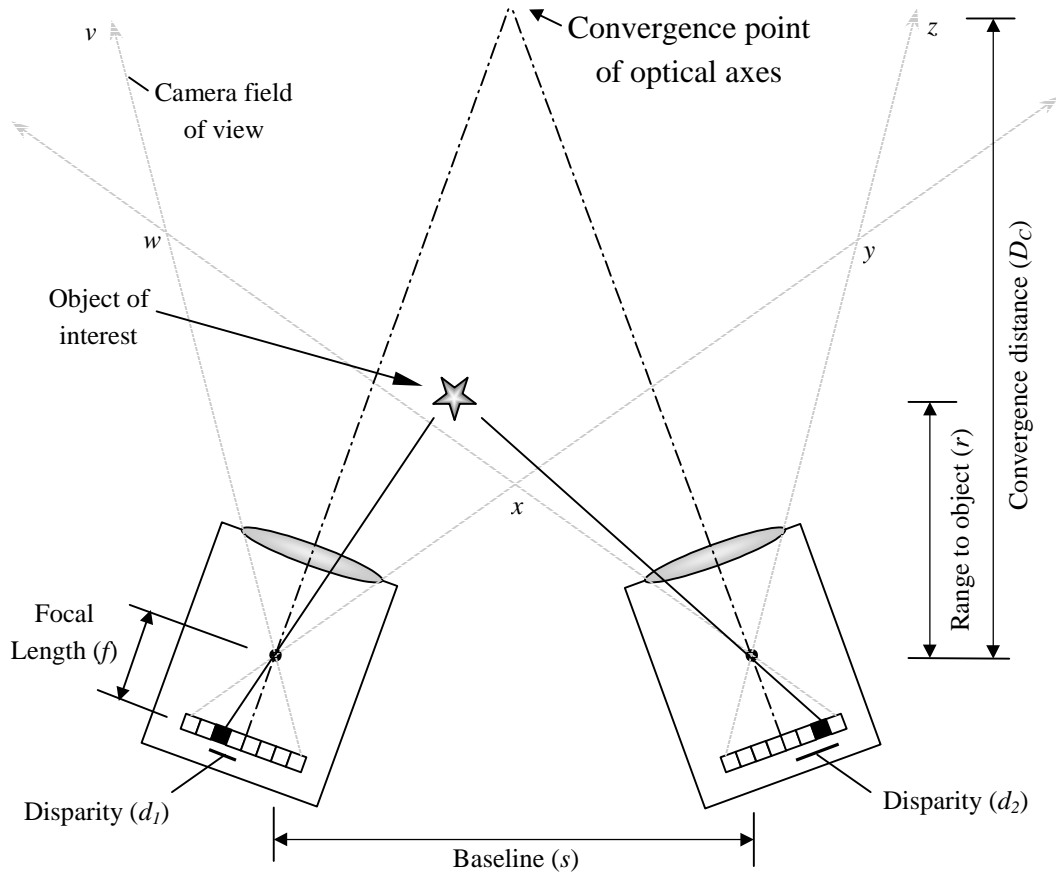


Figure 22 - Range estimation with non-parallel cameras

There are several advantages to designing a system that uses cameras that are not parallel. The first is that even for relatively large camera baselines, objects close to the cameras can remain in the field of view of each camera. This can be seen by observing that the area v - w - x - y - z (the field of view common to both cameras) is much larger, and much closer to the

focal point of the cameras in Figure 22 than in Figure 20, even though the cameras in Figure 22 have a larger separation distance. The capability of having a longer baseline leads to another advantages of the convergent system, namely that a larger baseline has more accuracy in range estimation.

The most important aspect in the implementation of any of the range estimation systems is the ability to accurately locate the object of interest in both the left and right images. This is difficult because the images from each of the stereo cameras can be very different due to the different position and orientation of the cameras, and slight variations in the performance of each of the cameras. There are many methods that have been developed to overcome these problems, and the most important will be explained in Section 3.2.

4.1.2 Feature Tracking

Unlike range estimation, feature tracking only requires that a feature of interest be tracked as it moves over a single video stream. If this can be done, then using this information together with the range estimation from the stereo system, the 3D position of the object that is being tracked can be determined. Figure 23 illustrates how changes in an object's position translate into varying positions on the cameras CCD, and varying ranges.

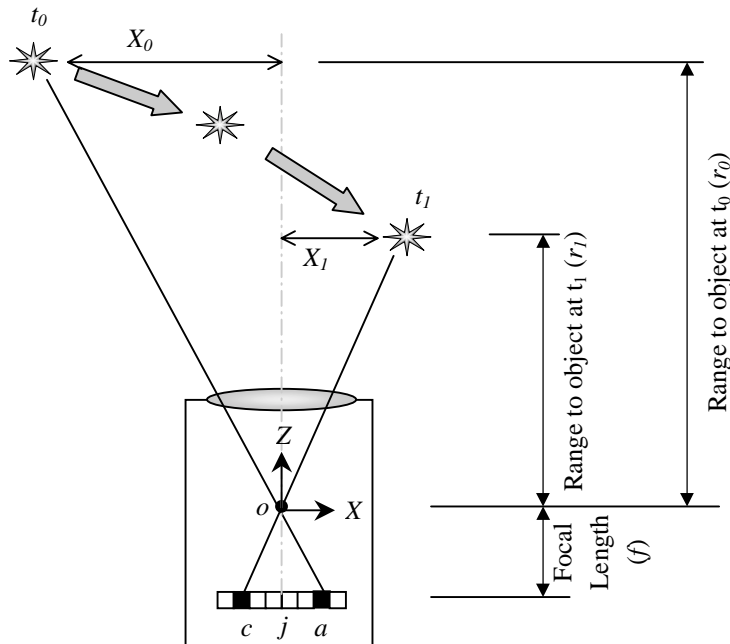


Figure 23 - Feature tracking system

If the range of the object in the Z direction is known, similar triangles can be used to calculate the X distance of the object (x_0), relative to the optical axis of the camera. For example, from Figure 23, the position of the object at time t_0 can be calculated as:

$$\frac{c - j}{o - j} = \frac{x_0}{r_0}$$

$$\text{rearranging gives } x_0 = \frac{(c - j) \times (r_0)}{o - j} \quad (4.3)$$

where the distance $(c - j)$ is equal to the pixel disparity from the centre of the CCD to the image of the object multiplied by a single pixel's X dimension.

This same calculation can be repeated in the Y - Z plane thereby allowing the Y dimension of the location of the object to be estimated.

Overall, it is possible to use feature tracking and range estimation to calculate the position of the object of interest relative to the camera system. Feature tracking has the same requirements as range estimation in that it needs the object of interest to be located in a continually changing stream of images. However, it is somewhat simpler because the image only has to be located in a single stream of mono images, rather than a stereo set. The method for tracking an object of interest that was chosen was template matching, as outlined in Chapter 3.

4.2 Vision System Requirements

As stated earlier, the two main functions of the vision system were to estimate the range to objects of interest in front of the cameras, and also to track the motion of that object as it moves relative to the AUV. The following requirements were set for the system.

1. To be able to estimate the range to objects at a distance of 2m or less from the front of the AUV with an error margin of less than 10%.
2. To be able to track the motion of objects at most 2 meters from the front of the AUV, as they move relative to the AUV.
3. To operate under VxWorks, using the Imagenation framegrabber, the Sony Pan/Tilt Camera, and the pair of Pulnix Cameras.
4. To perform feature tracking at a minimum of 5 Hz, and range estimation at a minimum of 2 Hz.
5. To perform all processing using colour images. Although grey scale processing is faster, the conversion from colour to grey can result in the loss of a large amount of information that might be useful.

4.3 Range Estimation Implementation

A block diagram of the system that was designed for range estimation is shown in the figure below.

At system startup the user is required to provide the coordinates of the object to be tracked. Using these coordinates, a template is extracted from one of the stereo images, both of

which have been enhanced using the difference of Gaussian method described in Chapter 3, to make the target easier to recognise from its background. The system tries to locate the newly extracted template in both the left and right stereo images. If it can be found, a range estimation is made, and the template matching is repeated on the next set of enhanced stereo images. If the template cannot be located in both of the stereo images, then it is no longer used for range estimation. Instead, it is discarded, and a new template is extracted from the template buffer if one is available. This new template is then used to try and estimate range.

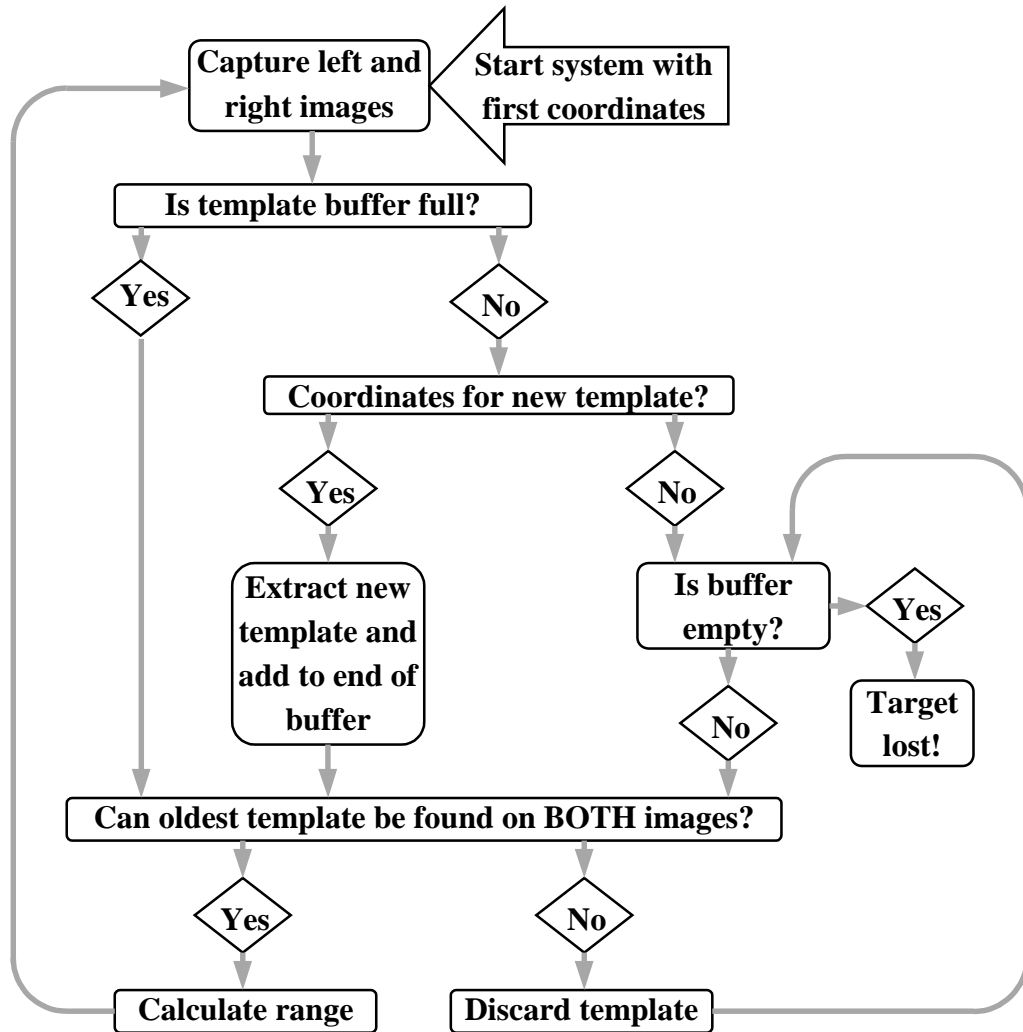


Figure 24 - Range estimation block diagram

An important aspect of this system is the method of adding templates to the template buffer. A method of doing this is to check at fixed time intervals to determine whether the current template correlates well with the current image. If it does, then a template is extracted from the position having the best correlation, and is added onto the end of the template buffer. Only extracting templates when the target is able to be located accurately in an image ensures that useful templates are kept in the buffer.

4.3.1 Calibration

Section 4.1.1 described two possible methods for estimating range using stereo cameras, either by having both cameras set so that they are parallel, or by having the optical axes of the cameras aimed at a target a known distance in front of the cameras. Both methods were tested, and the accuracy of each was then compared. The calibration method for both of these systems is essentially the same, and involves the following steps.

1. Separate the cameras by a known distance (the baseline of the system).
2. Place a target in front of the cameras (for the convergent system the target should be located at the desired point of their optical axes convergence, and for the parallel system the target should be as far away as is possible²).
3. Align the cameras so that the target appears on the same pixel on each image. This has to be done so that a horizontal line at the target position appears on the same row of pixels in each image. This is shown in Figure 25.



Figure 25 - Calibration of stereo cameras

The two cameras images on the left have been calibrated properly, because the cross appears on the same row and column of each image. The second pair of images have not been calibrated properly because the rows on which the cross appears on do not match up.

Having good horizontal calibration is important for getting good range estimation, because equations 4.1 and 4.2 can be applied directly to the system. However it is equally important to have vertical calibration correct. If two cameras are calibrated in the vertical direction, then once the template has been located on one of the stereo images, it will always lie on the same row on the other image, with only its horizontal position varying. However, if there is a vertical error in the initial calibration (as shown in Figure 25 (b)), then as the object moves further away from the cameras than the point where calibration was done, the vertical disparity of the object increases as well as the horizontal disparity. This means that in the well calibrated case, once the template is found in one image, only the same row on the other stereo image needs to be searched, while in the poorly aligned system, several rows above and below will also have to be searched. Since correlating templates with an image is a slow process, correlating a template on one row rather than several rows give the well calibrated system a major boost in performance.

² The minimum distance for this is a function of the length of the camera baseline, and the resolution of the cameras CCD's.

4.4 Feature Tracking Implementation

The feature tracking implemented works in much the same way as the range estimation, except that it is not necessary for a template to be located in both of the stereo images. For this reason the block diagram in Figure 24 is also applicable for feature tracking, except the images only need to be captured from one camera, and the template need only be located on one image and not both.

4.5 Integrating Feature Tracking and Range Estimation

Once both the feature tracking and the range estimation systems had been developed, they were integrated. Because both of the techniques work in similar ways, requiring an initially identified template to be located on a stream of moving images, they were able to be combined. One of the two cameras was used in both the range estimation and the feature tracking, while the other camera only assisted in the range estimation. Figure 26 illustrates the full system. The tracking occurs as fast as possible, with every frame that is grabbed from the left camera being used to perform feature tracking. Only after every n 'th left frame is the right frame also processed to obtain a range estimation. The magnitude of ' n ' will be dependant on a number of system variables especially the processing framerate and the frequency with which range estimations are required. Good calibration of the left and right cameras ensures that only one row of the image from the right cameras has to be processed to locate the template, as was described in Section 4.3.1. This system continues its operations until a template containing an image of the target can no longer be located on either of the images.

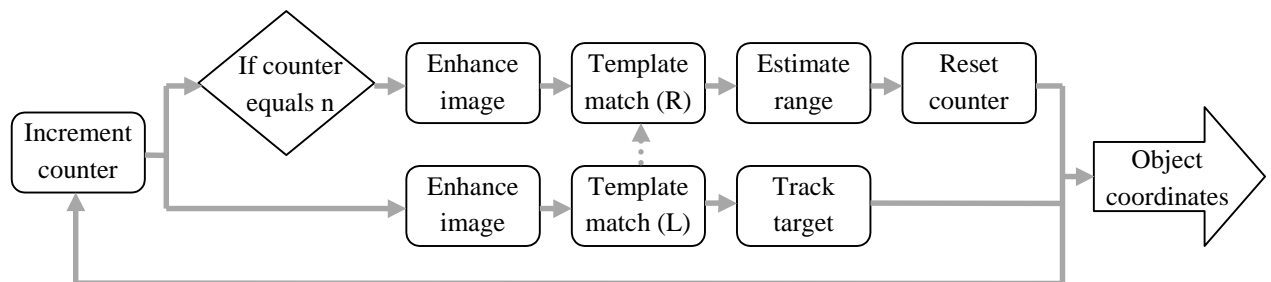


Figure 26 - Integrated feature tracking and range estimation systems

Several important design considerations for this integrated system are discussed in the following sections.

4.5.1 Template Extraction

It is important to note that in the integrated system templates are only ever extracted from one camera. The reason for this is that this camera is used for both the range estimation and feature tracking, therefore it is more critical that the target can always be identified in

this camera's images. It should be noted that in the system described in Figure 26 the choice of the left camera for use in both feature tracking and range estimation was purely arbitrary, and the roles of the cameras could easily have been reversed.

4.5.2 Template Updating

Another important design consideration was how frequently the template that is being used for range estimation and feature tracking is updated. There are several possible approaches that can be taken here.

The first and the simplest method is to keep updating the template whenever the current template that can be matched in the current image is within a certain threshold level. This method ensures that the template that is being used is likely to be very similar to the latest image of the target. It therefore makes the system able to track quickly changing targets, as the template is updated regularly to reflect the changes in the target. There is however one major problem with this system, which is due to the fact that it is very easy for the template to rapidly 'drift' off the target, due to several 'good' but erroneous correlations made in a row.

Another updating method, and one with the potential to solve the problem of templates drifting, is to use a 'first in first out' (FIFO) buffer to store templates which are recorded at regular intervals. The template that is being used for the matching is replaced with the oldest template from the buffer whenever the template's correlation performance exceeds a certain threshold value. This solves the problem of drift because it means that even if there is a small time period where erroneous estimates are made of the position of the target, and if a poor template is extracted during this period, the rest of the template buffer has to be cycled through before that poor template is used. By this time the target image will have probably changed significantly so the poor templates will no longer give good correlation values, and therefore will be quickly discarded. A comparison of these template updating techniques was done, and as reported in Section 4.6.1.

4.6 Software Testing

The following sections describe several tests that were carried out to evaluate the performance of the vision system software.

4.6.1 Template Update Frequency

Section 4.5.2 described two possible methods for updating templates, to get optimal performance of the correlation function when looking for a target. Both methods involve replacing the template if it cannot be located on an image to below a certain threshold value, and so tests were done to determine the threshold value. The threshold value is dependent on a number of environmental and system characteristics that will be described

later. However the method described below is provided to illustrate how the threshold value can be determined.

To calculate the threshold value a $25 \text{ pixel} \times 20 \text{ pixel}$ template was made of a target in an initial image, and the correlation level of that the template with the target on the following frames was also recorded. After that the target was removed from the following images, the correlation values that were now being produced were recorded. A comparison was made between the correlation values when a template could and could not be found in an image, and then a threshold value chosen. Figure 27 illustrates the results of this test. It should be noted that the images used in this test had no edge detection routines performed.

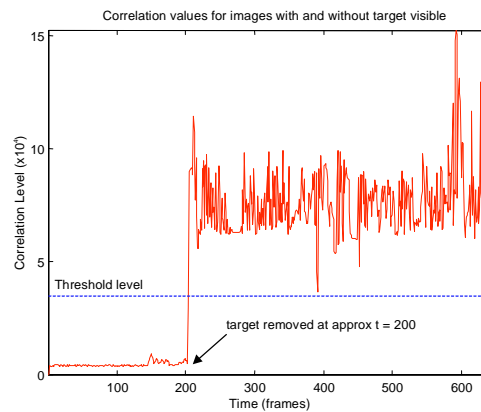


Figure 27 - Correlation values when target can and cannot be located

In this graph the target template was removed after the 200th frame. This caused the correlation values to increase dramatically. The average correlation value when the target was present in the image was about 4500, while the average value when the template was removed was 76000. Therefore a reasonable threshold value for deciding when a template could be considered to be no longer useful for correlation in an image was set near the midpoint of these two averages, at a level of 35000.

This threshold value should however be used with caution, as correlation values will vary greatly depending on the environment in which the cameras are operating. For example, if a target matches its background fairly closely then even when the target is no longer in the image, the correlation values for the empty images will be very similar to the levels when the target was present. This makes setting a good threshold value very important. This is important underwater where blurring and lack of light makes distinguishing a target from its background quite difficult.

Correlation values were also found to vary greatly even between similar cameras. This is possibly due to the physical settings on the camera. This factor is especially important to take into account for the stereo cameras, where a template extracted on one camera must be located on the other. The graph in Figure 28 shows the correlation values for an image that was extracted from the left camera, and then matched on both the left and right cameras. When the target was removed from both images similar correlation values were generated.

However when the target was present, the camera from which the template was extracted, (the left camera) consistently gave lower correlation value than the other camera.

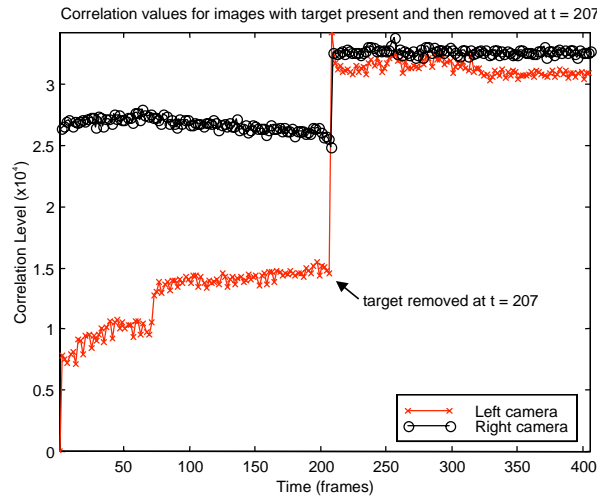


Figure 28 - Correlation values of left and right cameras

Once a threshold value had been selected, tests were done to evaluate the effectiveness of the two template updating methods discussed in Section 4.5.2. The methods were compared to a system that never updates the template that is initially extracted. The test was to view a target, and extract a template of it, and then slowly decrease the level of zoom of the camera, so that the size of the target appears to decrease in the image. This has the effect of making the target appear to be moving away, and this causes the size of the target to decrease. This was done to ensure that the target always appeared to be changing, as would be expected in most real systems. The time that the each method could keep tracking the target was measured. Figure 29 shows three graphs of the correlation levels.

In Figure 29(a) the initial template was used for all correlations, and was never updated. As would be expected, since the target is continually changing relative to the template, the correlation values gradually increase. The system was able to track the target for 51 frames (which corresponds to about 3.5 seconds), before the target was lost. The time to when the target was lost can be seen by the small decrease in the correlation value that occurred at that time. This corresponds to when part of the image other than the target has a better match with the template.

Figure 29(b) illustrates a scheme where the template is replaced every 40 frames, with no template buffering occurring. It can be seen that, the correlation values increase steadily as the target gradually changes to be different from the first template. However every 40 frames, a new template is selected, and immediately used. The use of new templates causes a sharp drop in the correlation value, corresponding to the near perfect match that the new template will have with the frames that follow. This method of tracking was able to track the target throughout the whole 12 second image sequence. The ‘template drift’ that was discussed in Section 4.5 was not observed here, possibly because the target, which

was changing shape due to changing zoom level, was always centered at the middle of the image, thus decreasing the chance of drift occurring.

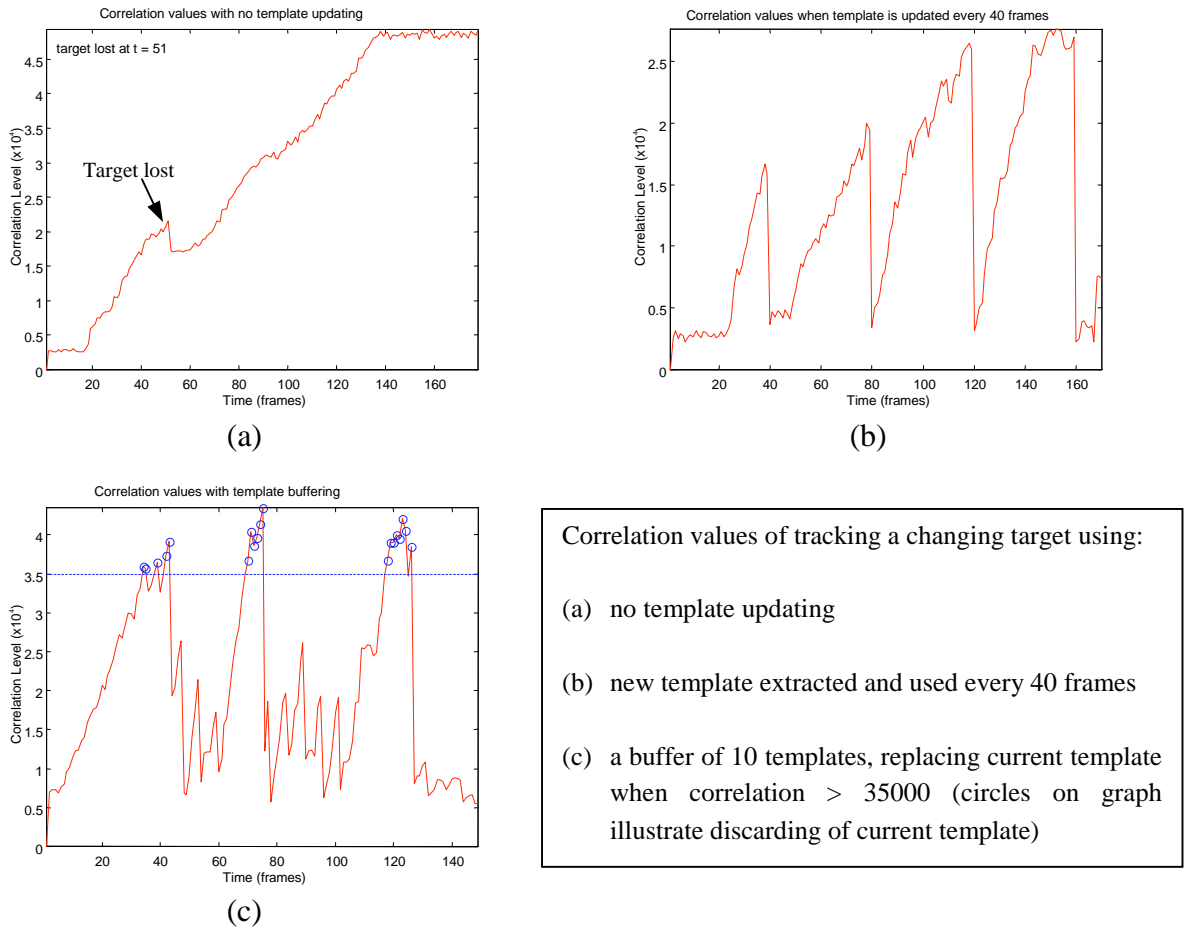


Figure 29 - Correlation levels for various updating techniques

Figure 29(c) shows the correlation values for a system that keeps a buffer of up to 10 templates, and replaces the current template when it is unable to be located to a threshold value of 35000. The template buffer was filled with new templates extracted from every 5th frame if the correlation level for the 5th frame is below a threshold value.

The graph shows three peaks where several templates from the buffer had to be tried before one could be found that would reduce the template correlation level to significantly below the threshold level. To get an adequate template from the buffer required trying a minimum of 5 templates from the buffer, suggesting that either the correlation threshold level could be decreased or the speed that templates are added to the buffer could be decreased, so templates in the buffer are used while they still have a chance of being valid. Despite this apparent inefficiency, it was found that the system was able to track the target for the full 12 second sequence.

Overall both methods of template updating were found to produce significantly better performance than a system which did not update the template. When comparing the two template updating techniques it was found that the additional complexity of keeping a FIFO buffer of templates did not produce any additional benefits over the system that

updates templates at a fixed time interval. It was observed that a buffer of templates, updated by reference to a correlation threshold, would reduce the chance of template drift, thus avoiding exposure of the major weakness of the fixed time updating method.

4.6.2 Image Enhancement Benefits

In the previous section various template updating techniques were evaluated for tracking a changing target. However in the tests, image enhancement was not performed, with templates being extracted and matched to raw images. Difference of Gaussians, described in Section 3.1.2, was chosen as a way to enhance images to try and improve the performance of the template matching. To evaluate the performance using enhanced images, a similar test to the one previously discussed was performed, using a FIFO buffer to store templates for the feature tracking. However, for this test the images had their edges enhanced using the difference of Gaussian technique. The following graph illustrates the correlation levels for this test.

The results of the experiment found that the target was lost quickly, with incorrect position estimates occurring after only 8 frames. This was a poor performance when compared with the similar test that had used the non enhanced images (see Figure 29(c)), which never lost the target, even after 160 frames.

There are a number of possible reasons for the poor performance of the template matching when the image had been enhanced. One possible reason is that the extra processing required to enhance the image caused the frame rate of the system to fall significantly, meaning that the shape of a target can vary more between frames. This makes it harder to track targets as there will be big differences in the correlation values between every set of consecutive frames that are processed.

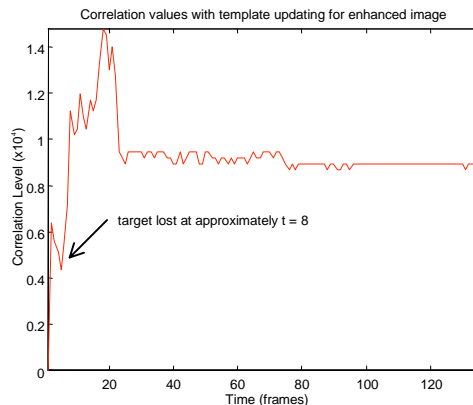


Figure 30 - Correlation levels for template matching using enhanced images

Another possible reason for the poor tracking performance is related to the information that is lost when the edges are detected on an image. All of the colour information is filtered away, and the surface texture information of objects is also lost. This makes tracking more difficult as there is now less to distinguish a target from its background.

The results of this test therefore suggest that enhancing the images to detect edges before template matching does not improve system performance.

4.6.3 Range Estimation Calibration

The stereo system was assembled, and then calibrated both with cameras parallel, and with the cameras optical axes converging at a point 2 meters in front of the cameras. A target was then moved along a track perpendicular to the cameras' baseline, passing through its midpoint. Using this setup disparity measurement calculations were taken at regular intervals, and a model was produced for the disparity to range relationship.

The graphs in Figure 31 shows the disparity data that was collected at varying distances from the cameras, along with a curve that was fitted to the data. For both the parallel axes and the converging axes tests, the camera baseline was 200mm. Using this distance, and the knowledge that each camera's focal length is 2.8mm and that the horizontal dimension of a pixel is 0.00635mm, equations 4.1 and 4.2 were used to calculate the theoretical disparities at these distances. The theoretical curves are also shown on the graphs.

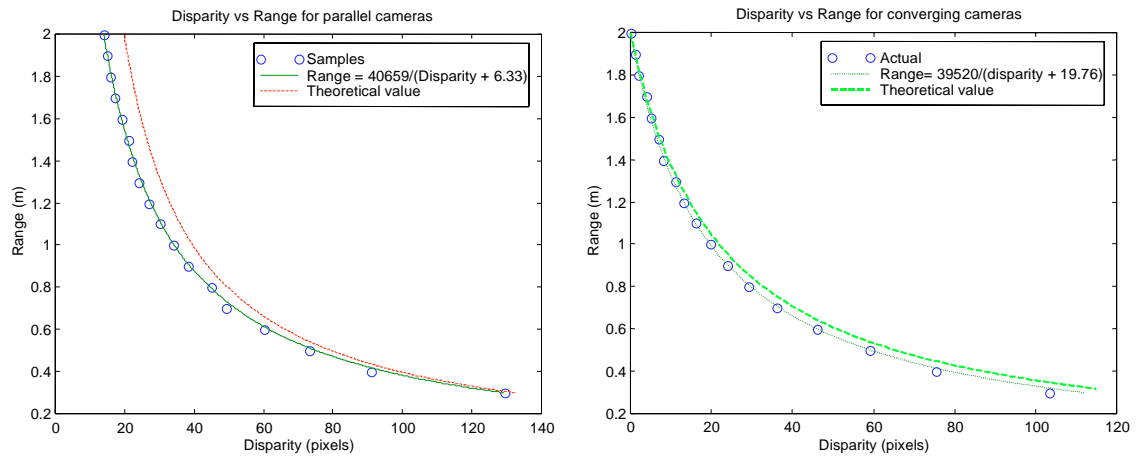


Figure 31 - Graphs of range vs disparity (experimental and theoretical)

As can be seen from the above graphs, a curve could be fitted very closely to the measured disparity/range values.

For the parallel cameras, the equation for range for the fitted curve was:

$$\text{range} = \frac{40659}{\text{disparity} + 6.33} \text{ mm} \quad (4.4)$$

This can be compared to the theoretical curve derived from equation 4.1 as:

$$\text{range} = \frac{200 \times 2.8}{\text{disparity} \times 2 \times 0.00635} = \frac{44094}{\text{disparity}} \text{ mm} \quad (4.5)$$

In this equation, disparity is measured in pixels, which is then multiplied by the pixel size for the cameras, and also multiplied by 2, to convert it to a dimension in mm. The reason

for the multiplication by 2 is because the image from the camera was subsampled to half its size to improve the system frame rate, meaning that some accuracy is lost, as the subsampling effectively doubled the size of each pixel.

A comparison between the two curves given by equations 4.4 and 4.5 show very consistent range results for large pixel disparities, but as the disparity decreased, the theoretical and experimental range estimations vary considerably. There are a number of reasons for this variation.

The first is that the finite resolution of the cameras' CCD's mean that although the cameras are calibrated so that their optical axes theoretically converge at 'infinity' they really are converging at a finite value. This means that as the disparity gets smaller, and the 'infinite' convergence point is approached, the errors get more and more prominent.

Other possible reasons for the variations in the equations are errors introduced by the lens distortion, and small errors in measurements of the disparity and range from the baseline.

For the converging camera system, it was found that the theoretical equation for the relationship between distance and disparity produced a good match with the measured values for the full range that the test was performed across. A curve with a similar form to the one for the parallel axis setup was fitted to the data, and it is given in equation 4.6.

$$\text{range} = \frac{39520}{\text{disparity} + 19.76} \text{ mm} \quad (4.6)$$

Computationally, equation 4.6 will run much faster than the theoretical equation given as equation 4.2, and since the results produced by both are very similar, equation 4.6 is the preferable equation to implement in software.

4.6.4 Range Estimation Accuracy at Varying Angles from Optical Axis

Once the stereo system was calibrated tests were performed to examine the accuracy of the range estimation when the target whose range is being estimated was near the edges of the image. This test was done to see the effects of the distortion that is present at the edge of an image has on the accuracy of the range estimation.

Figure 32 illustrates how the testing was done. Initially, a target was set a known distance from the cameras, on their centreline, and then the stereo cameras were slowly rotated by angle θ so that they were not pointing directly at the target. The pixel disparity for the templates was calculated at this new location, and then the range was estimated, using the model that was developed previously for the undistorted centre targets.

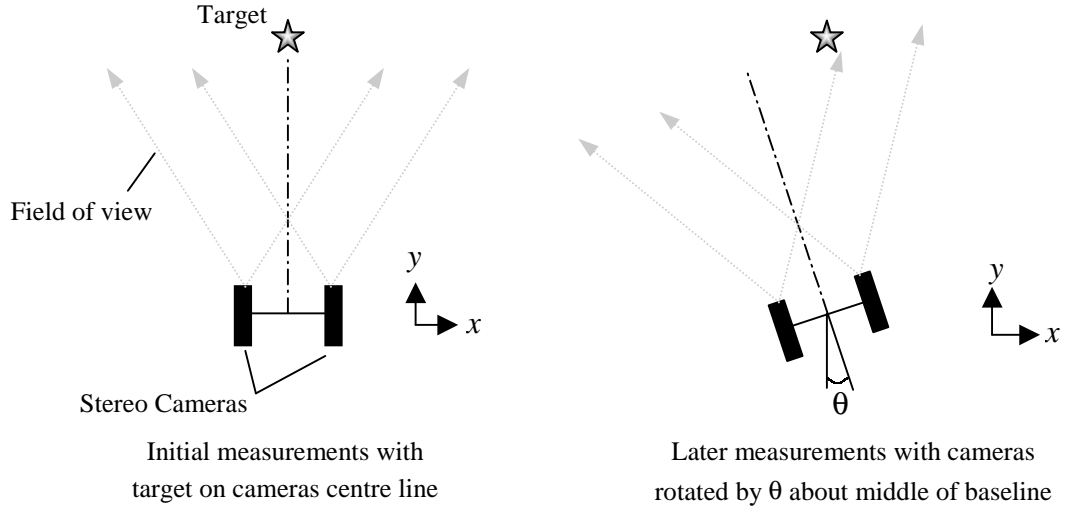


Figure 32 - Test procedure for range estimation accuracy of off centre targets

It was found that the cameras could be rotated by $\pm 30^\circ$ before the distortion from the camera lenses became too large for the object to be tracked in each image. Five sets of results were recorded at $\theta = 0^\circ, 7.5^\circ, 15^\circ, 22.5^\circ$ and 30° . At each of these points, 30 range estimations were made using the empirical model previously developed for the non parallel system, and the average was recorded. The range estimate that was calculated was then subtracted from the actual distance to the target to produce the graph shown in Figure 33.

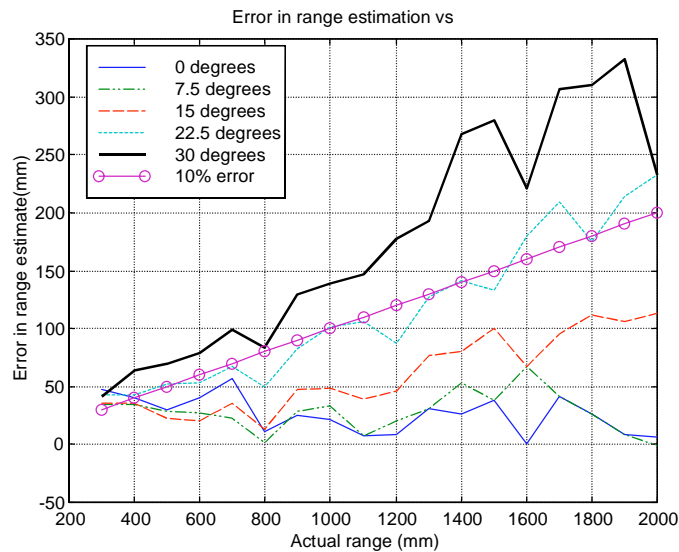


Figure 33 - Error in range estimations as lens distortion increases

As was expected, as the angle increased, the image distortion increased and the accuracy of the range estimation decreased. Also it was observed that as the range got closer to 2 meters (the distance that the system was calibrated to), the error in the range estimate became much larger. The reason for this large error can be explained by examining the equation that was developed to relate range to pixel disparity.

$$\text{range} = \frac{39520}{\text{disparity} - 19.7} \quad 4.7$$

From the equation we can see that when the disparity is small (theoretically it is zero at the calibration point), the range is extremely large. Hence a small error (i.e. 1 pixel) in the disparity, when the disparity value is small, results in a large error in the range calculation. Conversely, when the disparity is large, a small error in the disparity produces a small error in the range estimation. This is why the error the lens distortion adds to the image causes bigger errors as the range increases.

Theoretically, calibrating the stereo set to infinity (exactly parallel) should reduce this error, since the point where the disparity is zero is at infinity. This means that for the range over which this stereo system operates, the disparity values will be larger at all ranges, and therefore the error will be reduced.

The error should also be able to be scaled down by increasing the baseline of the cameras. As discussed in Section 4.1.1, a bigger baseline means that the disparity that is measured at any point is increased. However a variation of up to 15 degrees from the centreline results in errors in range estimation, which are below the target value of a maximum error of 10%.

4.6.5 Effects on Range Estimation of Variations in the Baseline

Section 4.1.1 described how the length of the stereo cameras' baseline affects the accuracy of the range estimation. A test was done to examine this effect. The accuracy of the range estimation was tested for camera baselines of 200mm and 460mm. To test the accuracy the cameras were calibrated to a convergence point 2 meters ahead, and then the accuracy of the range estimation using equation 4.2 was compared to the actual range to the target. This was performed for both the 460mm and the 200mm baselines, with targets located at varying distances along the centreline of the stereo cameras. Targets were also located at an angle of 15° from the centreline. (see Figure 32 for details)

Figure 34 shows the results from these tests.

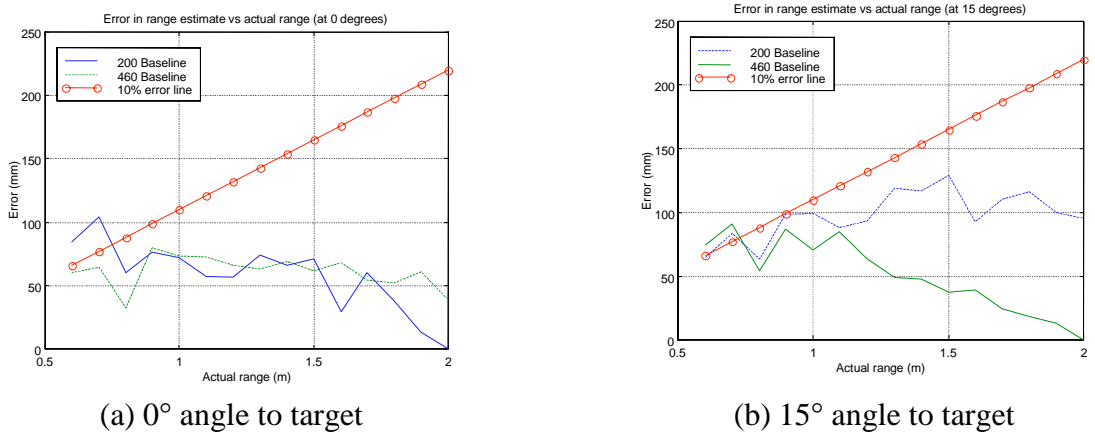


Figure 34 -Effect of varying baselines on range estimation accuracy

Figure 34(a) shows that when the target is on the centreline, the accuracy of the range estimation is approximately the same for both cameras. However when the target is at an angle of 15° from the centreline, the accuracy is significantly better for the large baseline cameras, especially for targets at distances near the camera calibration distance of 2 meters. The reason for this increased accuracy is because of the increased resolution that the wide baseline gives at all distances. This increased resolution means that a small error introduced by lens distortion is less likely to be rounded off to a large number in the wide baseline system. This illustrates the advantages of the wide baseline. What is not shown is the weakness of the wide baseline system, namely that the target cannot be tracked by both cameras once the target gets closer than 0.6 meters from the camera. In the small baseline system this distance was decreased to 0.3 meters.

4.6.6 Feature Tracking Tests

A test was done to evaluate the overall performance of the feature tracking routine, using video of a fish recorded at an aquarium. The tape could be played back into the framegrabber to compare the performance of different feature tracking methods. Although the water that fish was swimming in was clearer than would be expected in a natural environment, and the lighting was significantly better, the test was still interesting because it allowed evaluation of the feature tracking in a 'semi-natural' environment, unlike all previous tests of slow moving and easily controlled targets.

Feature tracking was tested with and without template updating, using unenhanced images, and one test was done using enhanced images.

The initial image of the target can be seen in its unenhanced and difference of gaussian forms in the figure below. The template that was used for tracking in all the tests was 25×20 pixels big, and was centred on the fish's eye in the first frame. Appendix G shows the tracking at several stages throughout the video sequence.

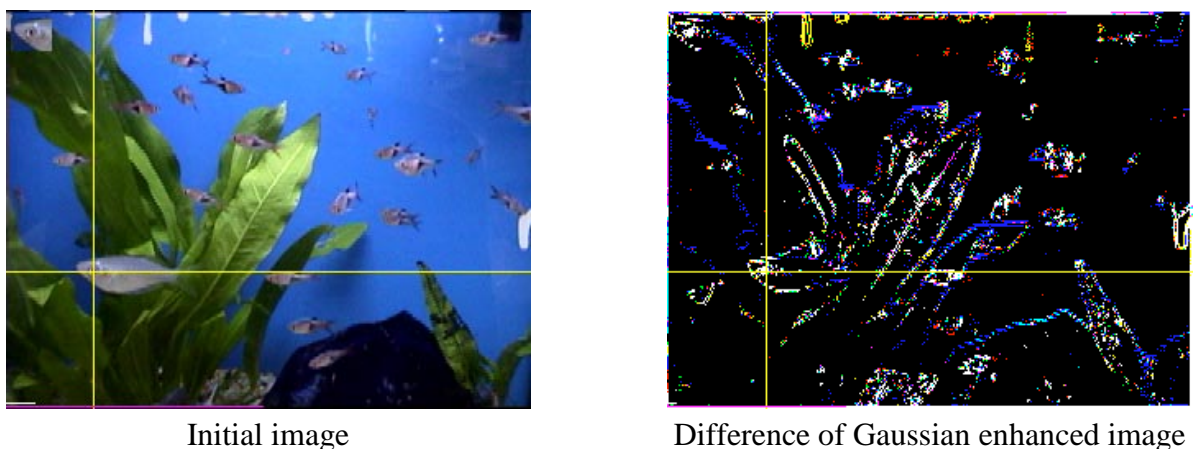


Figure 35 - Initial images from fish tracking test

It was found that both the system that updated templates from a FIFO buffer and the system that used a single template captured from the first frame, were able to track the

target for the first 22 seconds of the video. Tracking was lost in both cases when the fish changed directions quickly, causing all template matching to fail. The graphs below illustrate the correlation levels, and show the point where the target was lost.

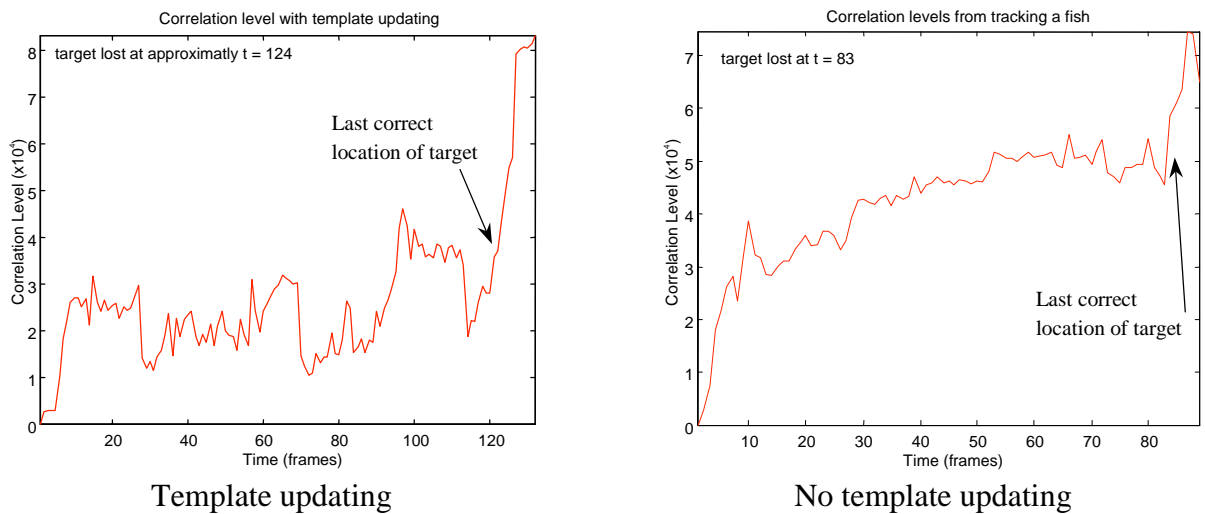


Figure 36 - Correlation levels for underwater tracking of a target

From these results it can be seen that the system that updates the template is better, as there is a much bigger jump in correlation levels when the target is lost. This is important since it means the computer is able to detect automatically when it has lost the target.

When the image was enhanced, and then the target was tracked, it was found that the tracking only lasted for about 3 frames, confirming the results from Section 4.6.2, that edge detection removes too much information from an image to provide any benefits to tracking. In this test a 3×3 and a 1×1 Gaussian surfaces were used when calculating the difference of Gaussians, and each of the Red, Green and Blue channels were enhanced independently. (A 1×1 Gaussian blurring means that the image does not get modified in any way). Various threshold levels were tested after the Gaussians had been subtracted, but they did not result in improved performance.

4.6.7 Combined Range Estimation and Feature Tracking Tests

Once the range estimation and feature tracking tests were done, the performance of the integrated system was evaluated by examining frame rates for different system configurations. The following table summarises the results obtained, when running the software on a Pentium II 300, with 64Mb ram. All image processing was done on 24 bit colour images.

A comparison of tests 1 and 2 illustrates the large performance cost of applying the difference of Gaussians edge detection over the full image, with the frame rate for full sized NTSC images (640×486) being reduced to less than one quarter of the unprocessed speed. This dramatic reduction in framerates was expected to be one of the reasons for the poor performance of feature tracking when the images were enhanced, as was described in section 4.6.2.

A possible solution to this reduction in framerate may be to use the difference of Gaussians only on regions of the image where templates are being extracted or correlated, rather than applying it to the whole image. Framerate could also be improved by decreasing the number of colours in each image. For example, processing a 16 bit image would be significantly faster than the 24 bit images used in these tests.

Test #	Description	Template size	Template search area	Frame rate (640×486)	Frame rate (320×243)
1	No processing	-	-	29	35
2	Edge detection (full image)	-	-	7	25
3	Template matching	20×25	10×10	18	28
4	Template matching	20×25	20×20	10	12
5	Template matching	20×25	30×30	5	6
6	Template matching	10×13	30×30	14	19
7	Template matching, edge detection	20×25	20×20	6	10
8	Template matching, FIFO buffer	20×25	20×20	10	12

Table 2 - Frame rates for various image enhancement routines

Tests 3 to 6 illustrate the tradeoff that is present between template size, search area and frame rates. Tests 3 to 5 show that increasing the search area causes the framerate to decrease significantly, and tests 5 and 6 show that increases in the template size also cause a decrease in framerate. Optimal system performance will require a tradeoff between template size and search area to keep the framerate high enough to be able to track fast moving targets.

4.7 Conclusion

Throughout this chapter the method for estimating the three dimensional coordinates of an object relative to the submersible have been developed. It was found that using a stereo set of cameras, the range to a target closer than 2 meters from the submersible could be estimated with an error of less than 10%.

A number of methods of calibrating the stereo system were investigated. It was found that calibrating the cameras so that their optical axes converged at a point of known distance from the cameras allowed an empirical equation to be derived to relate pixel disparity to range, and the results from this equation matched the theoretical derivation closely. When the cameras were set to be parallel, an empirical equation could also be developed that related pixel disparity to range, but this empirical equation did not produce as close a match to its theoretical values as occurred in the converging case.

It was also found that if the cameras were calibrated with a wide baseline they gave more accurate range estimates, especially in cases where a systematic error was introduced into the images due to lens distortion.

Both the feature tracking and the range estimation require a template matching routine to track a target over a stream of images. Tests were done to evaluate the optimal performance of this template matching routine, and it was found that performance was maximised when the templates were updated whenever they could not be correlated on an image to a certain threshold value. The update was done using a buffer of templates, and when a new template was needed, the oldest one from the buffer was always used. The use of a buffer decreases the chance of template drift occurring.

The difference of Gaussians was initially considered to be a potential method of enhancing images to increase the performance of a template matching algorithm. However it was found that, in all of the test cases performed, it caused a decrease in performance of the algorithm. The possible reasons for this was frame rate due to the time required to apply the difference of Gaussian, and also because edge detection caused a large amount of useful colour information to be lost.

Chapter 5 Further Work

There are several areas of the vision system that require further work to be performed on them. Most of this work relates to the software design and implementation. There is a small amount of work relating to the hardware design area that could also be done. The following sections outline the work that is possible.

5.1 Hardware

The main area where further work can be done relating to the hardware design of the vision system is to have a pair of spherical lenses manufactured for the stereo camera housing. As discussed in Section 2.3.1 this type of lens will give the vision system optimal performance, removing all the distortion that was present when a planar lens was used on the housing. The magnitude of the distortion that the existing planar lenses produce in comparison to the ‘ideal’ lenses could then be examined, as well as the effect that this distortion has on the system performance.

5.2 Software

There are several aspects of the vision system that require further work, and these are described in the following sections.

5.2.1 Conversion of Code to VxWorks

Although all of the code referred to has been written for a Windows NT environment, most of the image processing algorithms should be able to be ported to VxWorks without too much modification. This would then allow the whole vision system to be tested underwater on the submersible, which could lead to an evaluation of how suitable the vision system is as a navigational aid. It will also allow the integration of the feature tracking and the range estimation routines with the control of the Sony pan/tilt camera.

Image Enhancement for an Underwater Environment

When the vision system has been integrated into the submersible, the performance of the software algorithms will be able to be evaluated in an underwater environment. Vision algorithms that are effective for above water systems will probably prove to be ineffective

underwater, due to poor light and lack of image contrast, so there is considerable scope for investigating ways to improve these algorithms for their new environment.

The main method of enhancing images that was investigated in this report was to detect edges using a difference of Gaussian operation. This was found not to enhance the performance of the feature tracking and range estimation systems. However there are many other image enhancement methods that may provide better results, some of which are described below.

The first is to enhance images by locating the zero crossing points. This involves blurring the image, possibly by applying the difference of Gaussian method discussed in Section 3.1.2. Once this is done, the image depth is converted to binary, by setting all positive values to 1, and negative values to 0. The zero crossing points are then the pixels on the boundaries between the regions of zero and the regions of one.[2] This method is considered worth evaluating for underwater vision because it is particularly effective on blurred images, and on noisy images, both of which are very real problems for the AUV vision system.

Histogram equalisation is another possible way of enhancing the underwater images that may be beneficial. Histogram equalisation would essentially involve examining the colour histogram of an underwater image, and if it is found that the fill histogram lies in a very small region of the spectrum, it can be scaled out, so that it covers the full spectrum. This has the effect of making it much easier to recognise objects that are initially hard to distinguish from their background because of poorly lit environment, as the equalisation will have the effect of separating colours that initially were very similar. This should improve the system for underwater vision.

5.2.2 Automatic Correlation Threshold Adjustment

The final area of further work that will be discussed is to modify the template matching algorithm so that it periodically resets the correlation threshold. As was discussed in Section 4.6.1, the correlation levels that are produced when an template of a target can and cannot be located on an image varies greatly according to many factors, including lighting, template size, and camera and lens settings. This means that a correlation threshold level that may be suitable for determining if a target can be accurately located in one environment may fail completely in giving accurate results a short time later, due to very small different environment.

To overcome this problem an algorithm could be developed that, at regular time intervals, correlates the current template with sections of the image where the target is known not to be located. Then, by comparing these correlation values, with previous 'good' correlation levels, an updated correlation threshold value could be obtained to help make it easier to automatically tell if a target has been lost.

Chapter 6 Conclusion

This report examined the development of a vision system for an autonomous underwater vehicle. Work was performed in two areas, the design and building of some hardware for the vision system, and the designing and implementation of software for range estimation and feature tracking.

The hardware design work was broken down into three sections. The first was the designing of camera housings to attach a stereo camera set onto the frame of the AUV. Housings were designed that had all of the desired properties, being able to provide waterproof casing for a camera, and allow the camera to be accurately calibrated within this casing.

The next section of hardware design was to determine the optimal lens for the front of the housing, so as to minimise distortion that occurs at this interface. It was found that a spherical lens would be the ideal solution, but a planar lens would provide a satisfactory substitute, while being considerably cheaper and easier to machine and install.

The final section of hardware design related to the vision system hardware was to design and build cabling to interface the cameras with the on-board computer, and to provide power and control signals to the pan/tilt unit. All the cabling was designed and built, and found to perform satisfactorily.

The design and implementation of the vision system was the other area of work that was performed. Both range estimation and feature tracking systems were implemented and tested.

The range estimation implemented gave varying levels of accuracy, with performance decreasing as the distortion from the camera lenses increased. The system developed was found to be capable of range estimation with at least 90% accuracy for a 30° field of view, and with at least 80% accuracy for fields of view up to 60°.

Feature tracking employing two different template update systems was compared with feature tracking without a template update. Both systems were found to be significantly more effective than the system that did not update the template.

Of the two updating systems, one used buffer of templates which provided replacement templates whenever the current template could not be matched with an image to a certain threshold value, while the other system updated the template at a fixed time interval. Both

performed well but in view of the capacity of the buffer system to overcome the problem of template drift, it was concluded that this system was to be preferred.

Edge detection algorithms using the difference of Gaussian method were implemented to assess their capacity to improve the performance of the template matching for both feature tracking and range estimation. However the additional enhancement was found to reduce system performance. Reasons for this decrease in performance may have been due to reduced frame rates caused by the extra image processing, or because large amounts of useful colour information was discarded in the edge detection process.

References

- [1] Glover, T., Harwood, G., Lythgoe, J., *A Manual of Underwater Photography*, 1977, Academic Press, London.
- [2] Gonzalez, R., Woods, R., *Digital Image Processing*, 1993, Addison-Wesley, Sydney.
- [3] Halliday, D., Resnick, R., Walker, J., *Fundamentals of Physics Extended*, Fourth Edition, 1993, Wiley, New York.
- [4] Horn, B., *Robot Vision*, 1989, McGraw-Hill, New York.
- [5] Kingslake, R., *Applied Optics and Optical Engineering*, Volume IV, Part 1, 1967, Academic Press, New York.
- [6] Myler, H., Weeks, A., *The Pocket Handbook of Image Processing Algorithms in C*, 1993, Prentice Hall, New Jersey.
- [7] Nishihara, H., *Practical real-time image stereo matcher*, 1984, SPIE Proceedings Vol 449
- [8] Wettergreen, D., Gaskett, C., Zelinsky, A., *Development of a Visually-Guided Autonomous Underwater Vehicle*, Oceans 98.

Appendices

Appendix A	Hardware Specifications.....	A2
Appendix B	Derivation of Snell’s Law	A3
Appendix C	Field of view underwater calculations.....	A5
Appendix D	Calculation of Focal Length of Camera Lens.....	A6
Appendix E	Calculation of feature tracking accuracy.....	A8
Appendix F	Range estimation for converging cameras.....	A11
Appendix G	Feature Tracking of an underwater target.....	A13
Appendix H	Camera Housing Diagrams	A17
Appendix I	AUV Endcap Diagrams	A25

Appendix A Hardware Specifications

A-1 Cameras

Camera	Pulnix TMC-73	Sony EVI-D30
Video outputs (NTSC)	S-Video (12 pin Hirose plug), Composite (RCA)	S-Video (4 Pin mini DIN) Composite (RCA)
Lens	CS Mount attachment included.	F = 5.4 to 64.8mm, F1.8 to F2.7, horiz angle 4.4° to 44.8°
CCD	1/3" color CCD	1/3" color CCD
Power Requirements	12V DC, 190mA	12 – 14 V DC, 850mA
Dimensions	40 × 85 mm	142 × 109 × 164 mm
Mass	120g	1200g
Additional Features		Pan tilt unit – range 100° horizontal, 25° vertical Visca control language

Table 3 - Camera Specifications

A-2 Framegrabber

Framegrabber	Imagenation PXC200 (CompactPCI)
Inputs	Three standard multiplexed video inputs (NTSC/PAL/SECAM/S-video);
Maximum resolution	640 x 480 (NTSC) and 768 x 576 (PAL/SECAM)
Color output formats	YCrCb 4:2:2, RGB 32, 24, 16, 15 and 8-bit dithered
Camera power	+12 VDC output

Table 4 - Framegrabber Specifications

Appendix B Derivation of Snell's Law

Huygens' principle, which models light as waves, and states that:

*All points on a wave front serve as point sources of the spherical secondary wavelets.
After a time t , the new position of the wave front will be that of the surface tangent to these secondary wavelets.*

Using this principle Snell's law, which explains what happens when light passes from one material to another, can be derived. This is done using a model of light passing through a water/air interface can be made. Figure 37 (a) illustrates light waves in air that are approaching water with an angle of incidence θ_a and a wavelength λ_a . When the edge of the wavefront enters the water, the wavelength of the light is decreased to λ_w , due to the reduced velocity of light in water. This means that over the next time period t , the light in water travels a shorter distance than the light in air. This can be seen in Figure 37 (b).

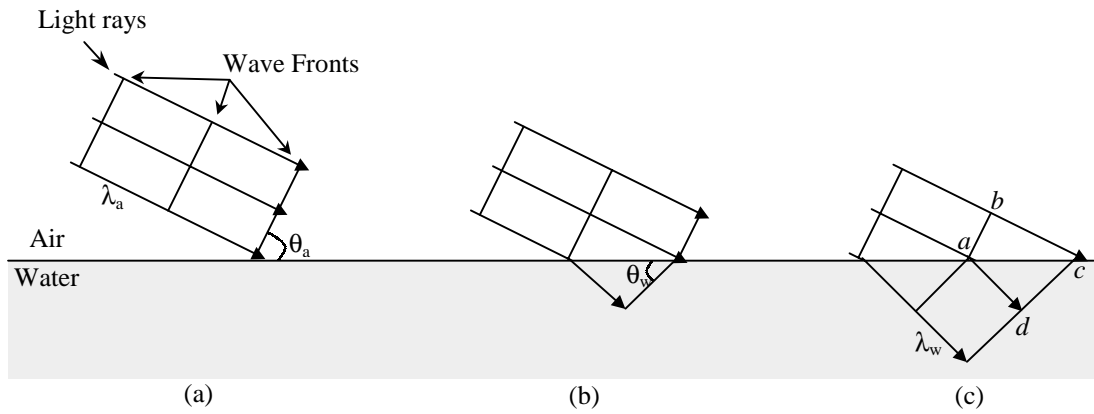


Figure 37 - Light bending at water/air interface

Trigonometry can then be used to calculate the relationship between the angle of incidence (θ_a) and the angle of refraction (θ_w). First, the change in wavelength is directly proportional to the velocity of light in the material. ie

$$\frac{v_a}{v_w} \propto \frac{\lambda_a}{\lambda_w} \quad (\text{B.1})$$

We know that the index of refraction for a material is equal to:

$$N_{\text{material}} = \frac{v_{\text{air}}}{v_{\text{material}}} \quad (\text{B.2})$$

Therefore we can say:

$$\frac{v_a}{v_w} \propto \frac{\lambda_a}{\lambda_w} \propto \frac{N_w}{N_a} \quad (\text{B.3})$$

From Figure 37 (c) we can see:

$$\sin(\theta_a) = \frac{bc}{ac} = \frac{\lambda_a}{ac} \quad (\text{B.4})$$

$$\sin(\theta_w) = \frac{ad}{ac} = \frac{\lambda_w}{ac} \quad (\text{B.5})$$

Solving (B.4) and (B.5) together, and substituting (B.3) gives:

$$\begin{aligned} ac &= \frac{\lambda_a}{\sin(\theta_a)} = \frac{\lambda_w}{\sin(\theta_w)} \\ \therefore \frac{v_a}{\sin(\theta_a)} &= \frac{v_w}{\sin(\theta_w)} \\ \therefore N_a \sin(\theta_a) &= N_w \sin(\theta_w) \end{aligned} \quad (\text{B.6})$$

Equation B.6 is called Snell's law. This equation describes what happens when light is passing from a material with a low index of refraction to a material with a higher index. The light will be bent towards the normal at the material interface. This has the effect of decreasing the field of view when looking from air into water.

Appendix C Field of view underwater calculations

Snell's Law gives the relationship between angle of incidence and refraction for the interface between two surfaces with different refractive indices (N_1 and N_2) as:

$$N_1 \sin \theta_1 = N_2 \sin \theta_2 \quad \text{C-1}$$

Using the above equation the reduction in the field of view of a camera through a glass/air/water interface can easily be calculated as follows.

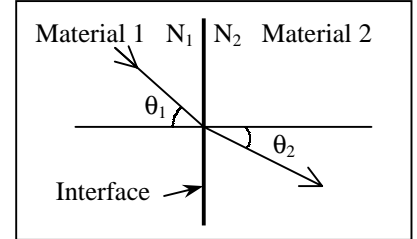


Figure 38 - Snell's law

$$N_{\text{air}} \sin \theta_{\text{air}} = N_{\text{glass}} \sin \theta_{\text{glass}}$$

$$N_{\text{glass}} \sin \theta_{\text{glass}} = N_{\text{water}} \sin \theta_{\text{water}}$$

Combining these gives:

$$N_{\text{air}} \sin \theta_{\text{air}} = N_{\text{water}} \sin \theta_{\text{water}} \quad \text{C-2}$$

Where $N_{\text{air}} = 1.0$, and $N_{\text{water}} = 1.333$. As can be seen from the equation for a planar interface the index of refraction of the glass interface is irrelevant to the performance of the system.

From equation C-2 it can be easily calculated that if θ_{air} is set to 90° (the maximum angle that it can be), the largest that θ_{water} can be is 48.75° . This means that the maximum possible angle of view through a planar lens is $2 \times 48.75^\circ = 97.5^\circ$. ie if a camera that is rated as having a 180° field of view (in air) is used, the widest field of view in water is 97.5° . Figure 39 shows this relationship between angle of view of a camera in air, and the angle of view in water, when using a planar interface between the camera and the air.

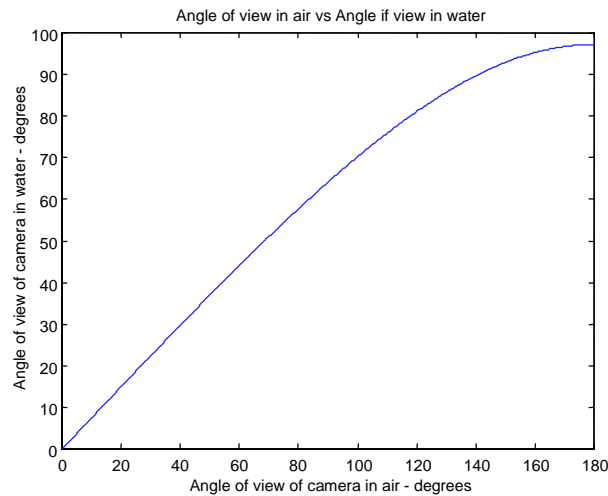


Figure 39 - Distortion at air/water interface

Appendix D Calculation of Focal Length of Camera Lens

The specifications for the lenses that are to be used for the stereo cameras can be calculated when some assumptions about the system performance are made.

- Assume that we want to be able to track objects 2 meters in front of the AUV
- Assume that 2 meters of movement of the object relative to the AUV will keep the object in view of the cameras.

The following information is also known about the system:

- The size of the camera CCD for the Pulnix cameras 3.6×4.8 mm
- The diameter of the lens at the front of the camera housing that the camera is looking through is 60mm
- A planar lens is initially going to be used on the camera housing

Using this information, the focal length (f) of the lens required can be calculated, and the offset of the camera lens from the lens at the front of the camera case (x) is also able to be found. Figure 40 illustrates this.

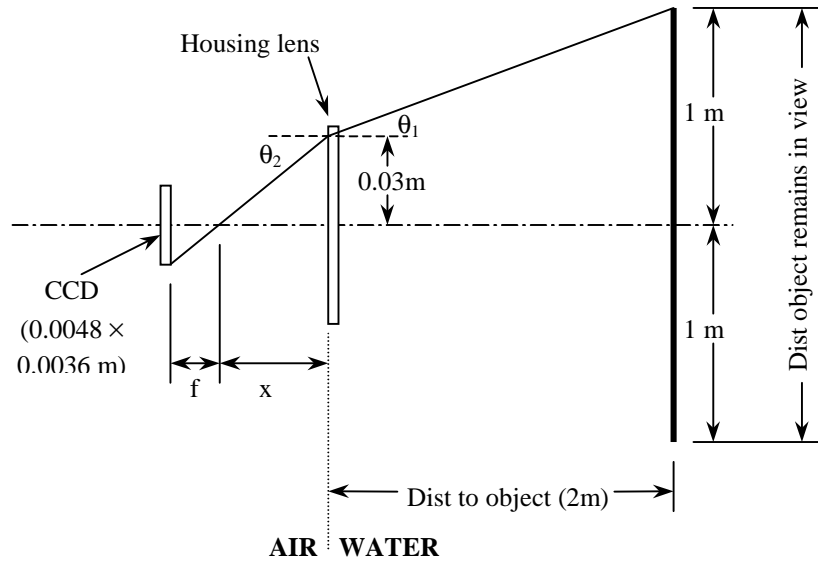


Figure 40 - Calculation of focal distance

Initially θ_1 can be calculated as:

$$\tan(\theta_1) = \frac{\text{dist which object remains in view} - \text{housing lens diameter}}{2 \times \text{distance to object}} \quad \text{D-1}$$

$$\therefore \theta_1 = \tan^{-1}\left(\frac{1-0.3}{2}\right) \approx 25.87^\circ$$

Now θ_2 , the angle of incidence at the air side of the air/water interface, can be found using Snell's Law (see Appendix C) as:

$$\theta_2 = \sin^{-1}\left(\frac{N_w}{N_a}\sin(\theta_1)\right) \approx \sin^{-1}\left(\frac{1.333}{1.0}\sin(25.87)\right) \approx 35.58^\circ \quad \text{D-2}$$

Now the focal length of the lens can be calculated as:

$$\tan(\theta_2) = \frac{\left(\text{horizontal dimension of CCD}/2\right)}{\text{focal length } (f)} \quad \text{D-3}$$

$$f \approx \frac{\left(4.8/2\right)}{\tan^{-1}(35.58)} \approx 3.35\text{mm}$$

Similarly the distance from the camera lens focal point to the lens can be calculated as:

$$\tan(\theta_2) = \frac{\text{radius of camera housing lens}}{\text{distance } (x)} \quad \text{D-4}$$

$$x \approx \frac{30}{\tan^{-1}(35.58)} \approx 41.9\text{mm}$$

The above calculation shows that the lens that is required must have a focal length of 3.35mm. The closest standard camera lenses that are available are 2.8mm and 4.0mm. As the focal length of a lens decreases the angle of view increases, and so for a 2.8mm focal length CS-mount lens, the angle of view can be calculated to be approximately 86° . Examining the graph in figure 3, (the angle of view in air vs angle of view in water), it can be seen that an angle of 86° or less in air, has an almost linear relationship with the angle of view in water, suggesting that there will be minimal picture distortion for any lenses with focal length of 2.8mm or greater when a planar lens is used on the camera housing.

Appendix E Calculation of feature tracking accuracy

To estimate the possible errors that would be observed in an object that is viewed underwater, the following assumptions were made.

- The possible error in the object position projected on the CCD is limited by the size of each pixel on the CCD, which are approximately $6.35\mu\text{m}$ (H) \times $7.4\mu\text{m}$ (V).
- The object that is being viewed is a distance of 2 m from the front of the camera housing.
- The object is in water that has a refractive index of 1.33, and the refractive index of the air where the camera is located is 1.0.
- The camera housing lens is planar, so the distortion will increase as the edge of the image is approached.

Using these assumptions, the following diagram can be drawn to illustrate the way an error in the position calculation can occur.

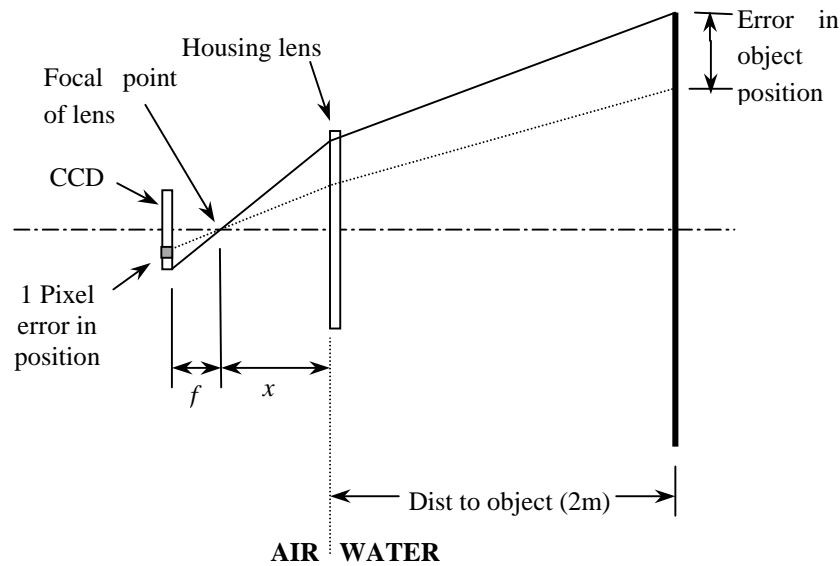


Figure 41 - Error in estimated object position

The camera lenses that were used had a focal length of 2.8mm. Using this information along with all of the known data about the camera housing, and the camera specifications, the size of any error that can occur can be estimated.

First, distance x (the distance from the camera focal point to the housing lens) needs to be calculated. This can be done using similar triangles, as is illustrated below.

$$\frac{\left(\frac{\text{horizontal dimension of CCD}}{2}\right)}{\text{focal length } (f)} = \frac{\text{radius of camera housing lens}}{\text{distance } (x)} \quad \text{E-1}$$

Substituting known values we get:

$$\frac{(4.8/2)}{2.8} = \frac{30}{x}$$

$$x = \frac{2.8 \times 30}{(4.8/2)} = 35mm$$

Using this information, the position where the object is estimated to be can be found. The following diagram illustrates how to do this.

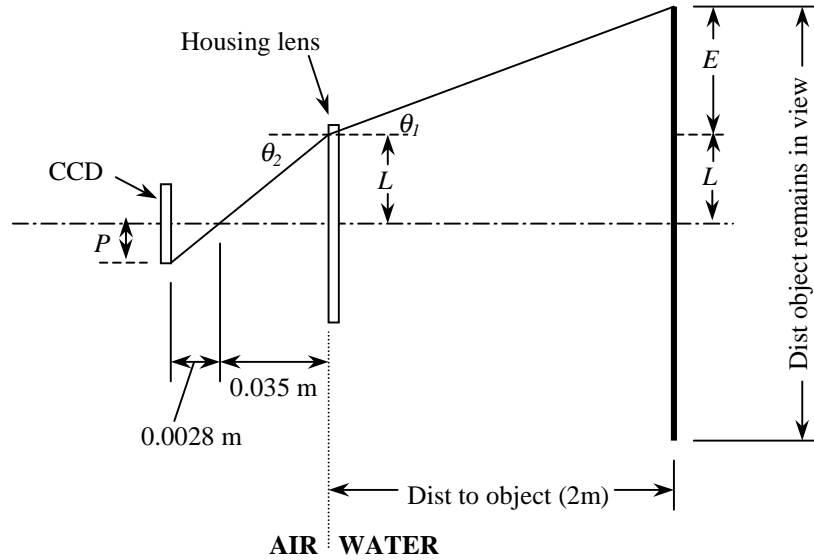


Figure 42 - Estimation of object position

To calculate the size of the error an equation must be found that relates the magnitude of E and L to the size of the CCD. The following equations do this.

$$\tan(\theta_2) = \frac{P}{0.0028}$$

$$\theta_2 = \tan^{-1}\left(\frac{P}{0.0028}\right) \quad \text{E-2}$$

Also, using similar triangles:

$$L = \frac{P \times 0.035}{0.0028} \quad \text{E-3}$$

Snell's law gives:

$$\theta_1 = \sin^{-1}\left(\frac{N_a}{N_w} \sin(\theta_2)\right) \quad \text{E-4}$$

Since the object of interest is 2 meters away:

$$E = 2 \times \tan(\theta_1) \quad \text{E-5}$$

Combining equations E-2, E-4 and E-5 we get:

$$E = 2 \times \tan \left\{ \sin^{-1} \left(\frac{N_a}{N_w} \sin \left(\tan^{-1} \left(\frac{P}{0.0028} \right) \right) \right) \right\} \quad \text{E-6}$$

Adding this to equation E-3 we get

$$L + E = \frac{P \times 0.035}{0.0028} + 2 \times \tan \left\{ \sin^{-1} \left(\frac{N_a}{N_w} \sin \left(\tan^{-1} \left(\frac{P}{0.0028} \right) \right) \right) \right\} \quad \text{E-6}$$

By setting P to an initial size, and then varying it by the size of one pixel on the CCD, a graph can be created of the position of an object on the CCD vs the actual position of an object 2m in front of the camera. This graph is shown below.

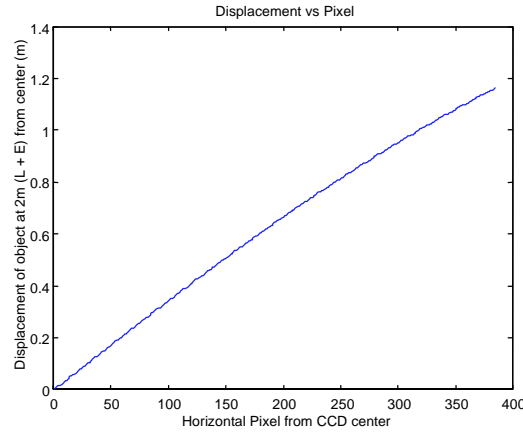


Figure 43 -Graph of object location vs pixel position

Now by examining the difference in the location of the object for every one pixel variation a value for the position error due to a one pixel error can be calculated. The following graph illustrates how this error decreases as the one pixel error gets further from the centre of the CCD.

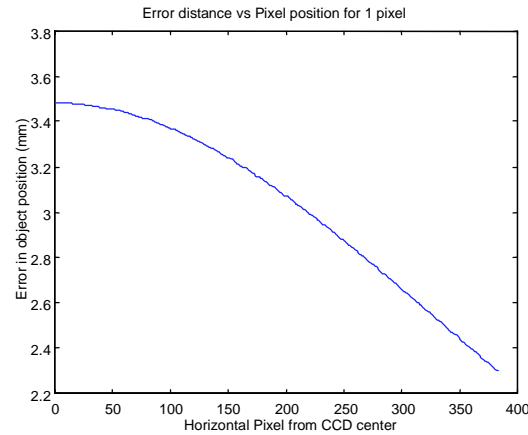


Figure 44 - CCD error position vs error in position estimation

Appendix F Range estimation for converging cameras

When using stereo cameras whose optical axes converge a known distance in front of the cameras, it is possible to derive an equation relating disparity (d_1 and d_2) to range (r). The following figure shows two cameras, which are separated by a baseline (s), and are converging at a point distance (D_c) from the baseline. It is assumed that the convergence point lies on the line that is perpendicular to the baseline, intersecting the base line at its mid point.

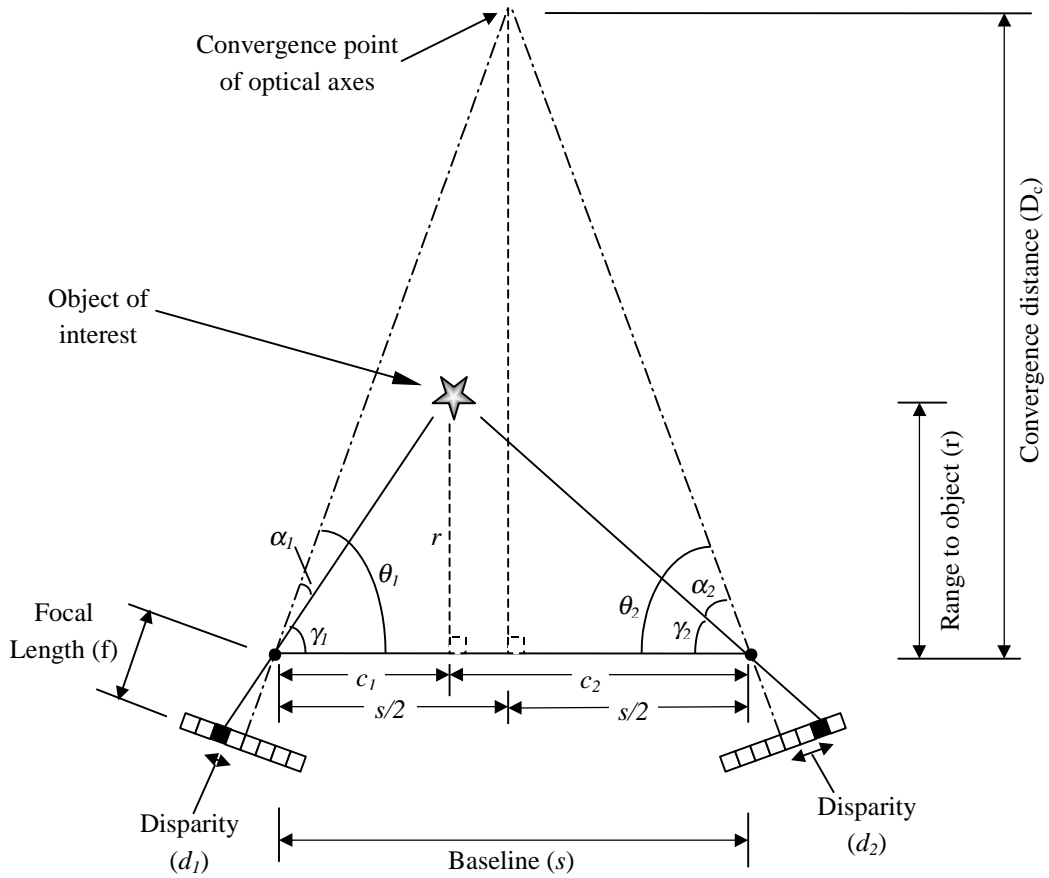


Figure 45 - Schematic of converging cameras

Since the focal length of the cameras is known, as is the disparity of the object on each CCD from the centre, it is possible to calculate the angles α_1 and α_2 to be:

$$\alpha_1 = \tan^{-1}\left(\frac{d_1}{f}\right)$$

$$\alpha_2 = \tan^{-1}\left(\frac{d_2}{f}\right)$$
(F-1)

We also know the separation distance of the cameras and the distance to the convergence point of the cameras' optical axes, so it is possible to calculate θ_1 and θ_2 as follows:

$$\theta_1 = \theta_2 = \tan^{-1}\left(\frac{2D_c}{s}\right) \quad (\text{F-2})$$

We can then show that $\gamma = \theta - \alpha$, which gives:

$$\begin{aligned} \gamma_1 &= \tan^{-1}\left(\frac{2D_c}{s}\right) - \tan^{-1}\left(\frac{d_1}{f}\right) \\ \gamma_2 &= \tan^{-1}\left(\frac{2D_c}{s}\right) - \tan^{-1}\left(\frac{d_2}{f}\right) \end{aligned} \quad (\text{F-3})$$

From the diagram we can also get the following relationships:

$$\begin{aligned} c_1 &= r \times \tan^{-1}(\gamma_1) \\ c_2 &= r \times \tan^{-1}(\gamma_2) \end{aligned} \quad (\text{F-4})$$

$$c_1 + c_2 = s \quad (\text{F-5})$$

Solving these for r , and substituting in F-3 gives:

$$\begin{aligned} s &= r \times \tan^{-1}(\gamma_1) + r \times \tan^{-1}(\gamma_2) \\ &= r \times \tan^{-1}\left(\tan^{-1}\left(\frac{2D_c}{s}\right) - \tan^{-1}\left(\frac{d_1}{f}\right)\right) + r \times \tan^{-1}\left(\tan^{-1}\left(\frac{2D_c}{s}\right) - \tan^{-1}\left(\frac{d_2}{f}\right)\right) \\ \therefore r &= s \times \left(\frac{1}{\tan^{-1}\left(\tan^{-1}\left(\frac{2D_c}{s}\right) - \tan^{-1}\left(\frac{d_1}{f}\right)\right)} + \frac{1}{\tan^{-1}\left(\tan^{-1}\left(\frac{2D_c}{s}\right) - \tan^{-1}\left(\frac{d_2}{f}\right)\right)} \right)^{-1} \end{aligned} \quad (\text{F-6})$$

This equation can then be used to relate pixel disparity to range for any calibrated system where the camera baseline, focal length and the distance to optical axes convergence point is known. It should be noted that the pixel disparity (d_1 and d_2) must be in the same units as the focal length (generally mm). This can generally be achieved by multiplying the pixel disparity by the pixel width.

Appendix G Feature Tracking of an underwater target

Although the camera housing was not built in time to perform underwater tests, it was possible to test the feature tracking using video recorded at an aquarium. The lighting and water clarity are much better than would be present in most natural lakes and seas. However the test was interesting because it allowed evaluation of the feature tracking in a 'semi-natural' environment.

Tests were carried out to examine the performance of feature tracking in an 'underwater' environment, and to compare examine the effectiveness of a system template buffering and a system in which the template never changes.

It was found that both methods were able to track the target fish for the first 22 seconds of the video sequence, but both immediately lost the target when the fish suddenly changed directions a short time later.

Figure 46 shows the correlation values that were recorded as the fish was tracked using template matching with no template updating, and Table 5 presents several frames recorded during the tracking.

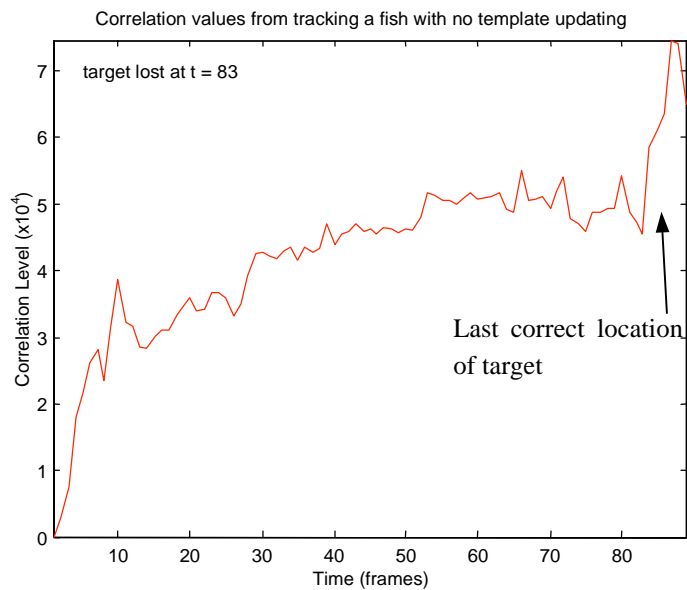


Figure 46 - Correlation values when tracking an underwater target with no template updating

From the table it can be seen that at about frame 84 the fish started to turn around, which caused the correlation levels between the template and the image to rise sharply, from approximately 46000 to over 70000 in less than five frames. It was at this time that the target was lost.

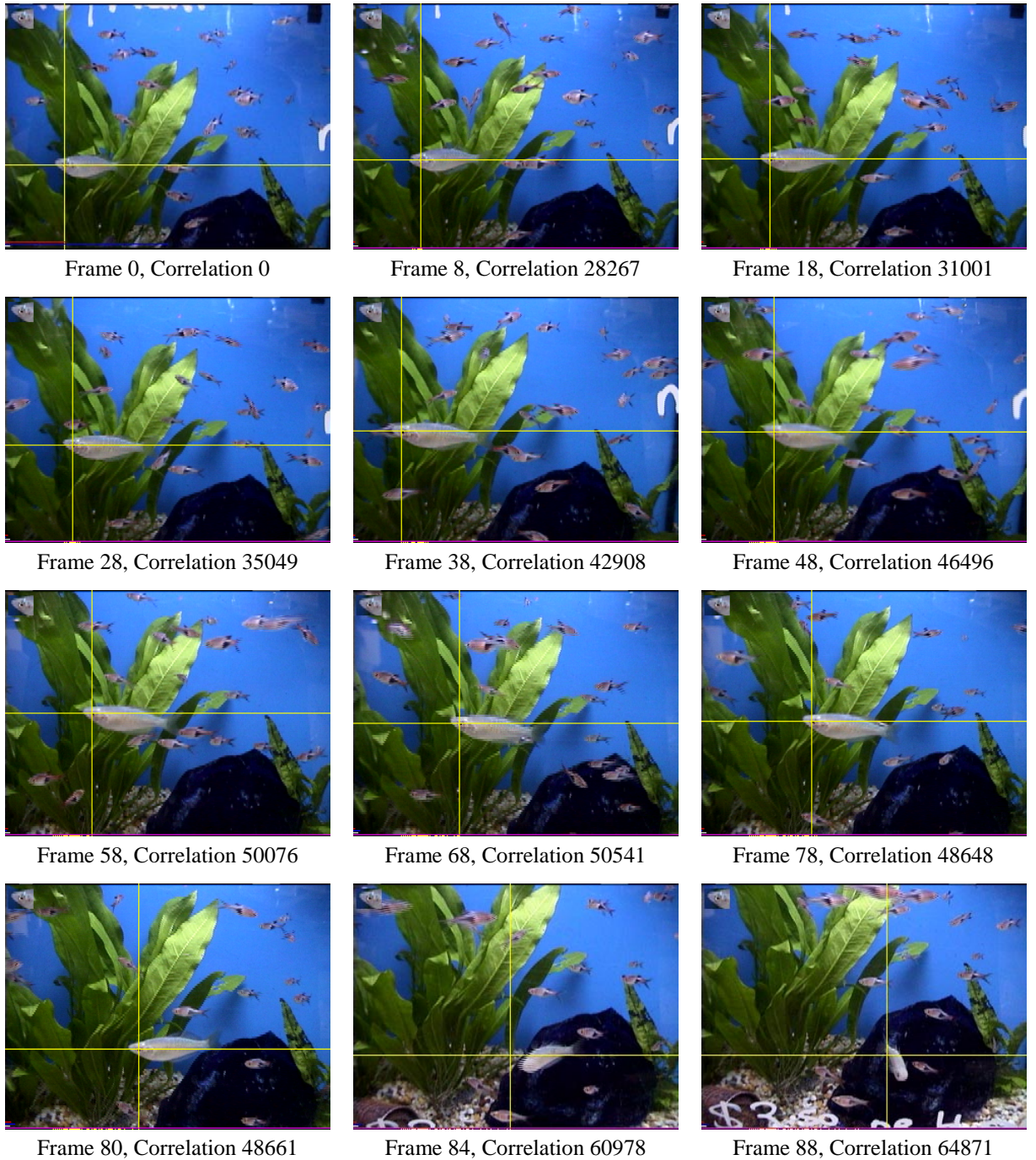


Table 5 - Feature tracking of an underwater target without template updating

Note that the template that is being matched is shown in the upper left corner of every image, and the cross represents the estimated location of the target produced by the template matching routine.

When tracking was performed on the same video using a system that utilised a LILO template buffer that could contain up to 10 frames, very similar results were obtained to the previous test, with the target being lost after approximately 22 seconds.

Figure 47 presents the correlation values that were recorded and Table 6 shows a sequence of frames recorded during the test.

The target was lost at the 124th frame, which corresponded to a jump in correlation values from approximately 40000 to 80000.

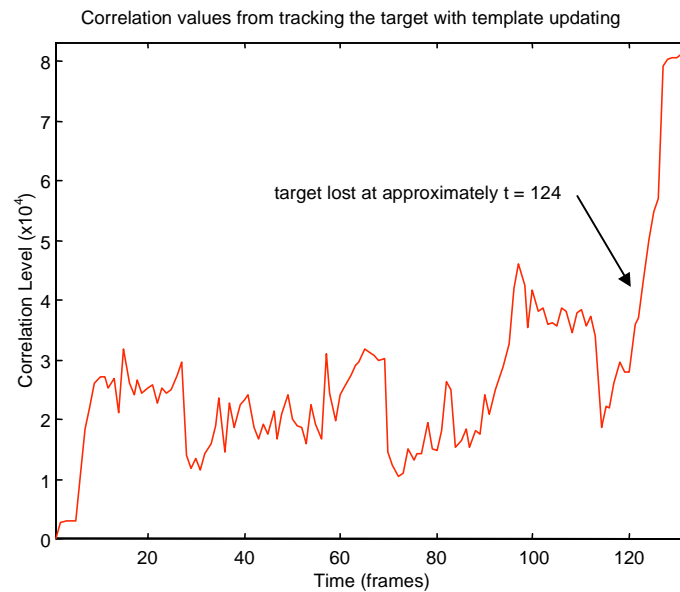


Figure 47 - Correlation values when tracking an underwater target with template updating

Although both tracking with and without template updating was able to track the target for the same length of time, the system that updated the templates produced results that are more suited for use in an automated system, as there is a much bigger change in threshold values from times when the target was located to times when the target was lost.

The template updating scheme also has the additional benefit that the changing template compensates for gradual changes in the image. Figure 48 illustrates how the templates that were stored in the buffer did vary. However they all kept the target in the center of the template. It is important that the templates that are collected give good representations of the target, and keep the target located in the same position in the template, as it reduces the chance of template drift occurring.



Figure 48 - Templates added to the template buffer during the tracking

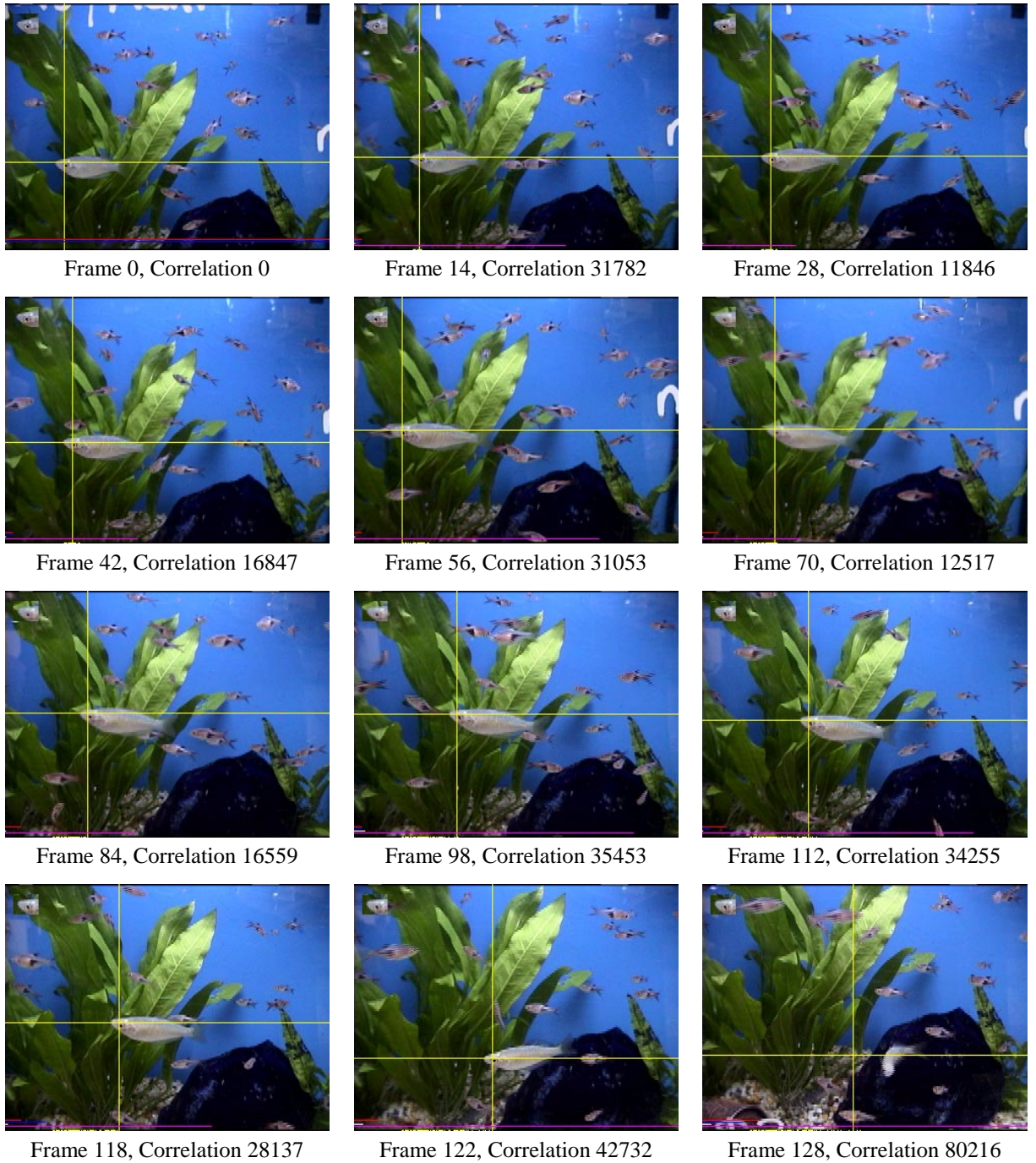


Table 6 - Feature tracking of an underwater target with template updating

Although both schemes tracked the target for the same time, the scheme that keeps a template buffer was considered to be the preferable scheme for a number of reasons.

The first is that produces a bigger variation in correlation levels when a target can and cannot be located in an image, which makes it easier to develop a system that automatically knows when the target is lost.

The second reason is that the template updating method is better able to track targets that change slowly over time, as it can gradually compensate for these changes.

Appendix H Camera Housing Diagrams

Drawing Number	Drawing Name
0001	Outside Case of Camera Housing
0002	Inner Frame of Camera Housing
0003	Frame Clamp – Upper and Lower
0004	Outer Case Clamp and Joint
0005	Camera Plates – Upper and Lower
0006	Lens and Back Plate
0007	Assembled View – Camera Housing

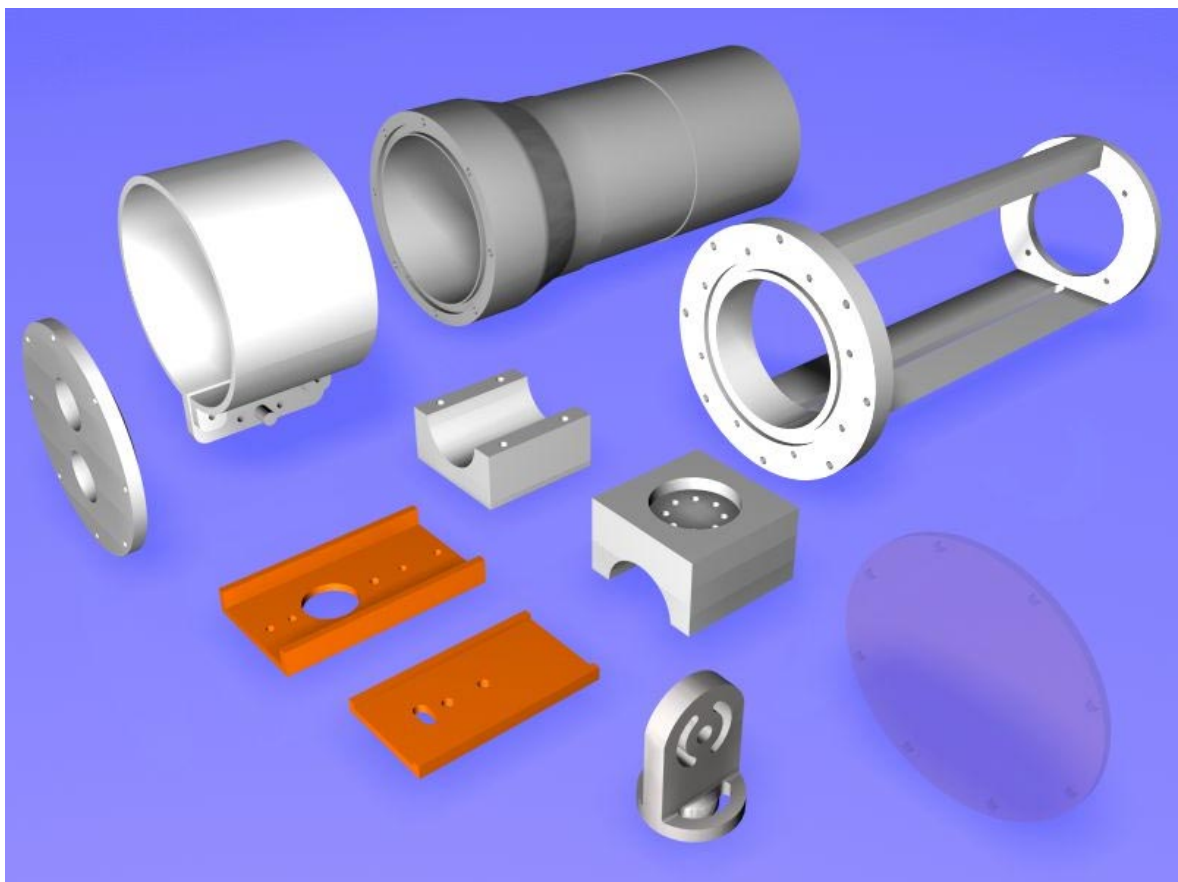


Figure 49 - Camera Housing Parts

8

7

6

5

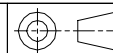
4

3

2

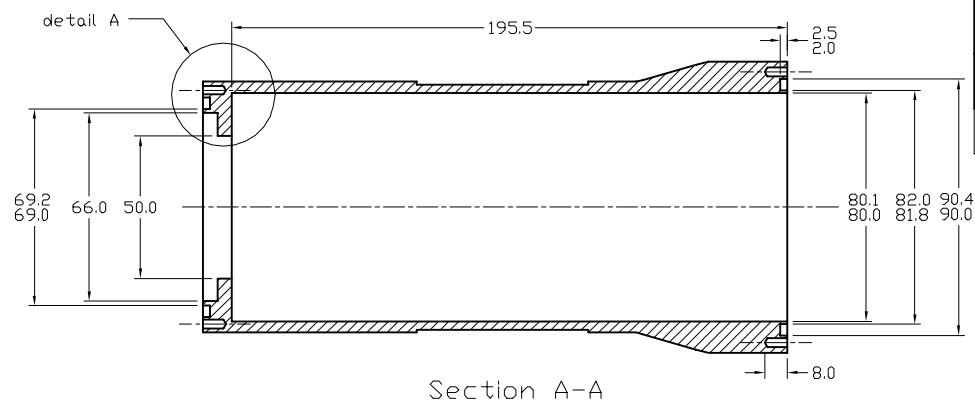
1

COMPUTER DRAWING - DRAWN TO AS1100

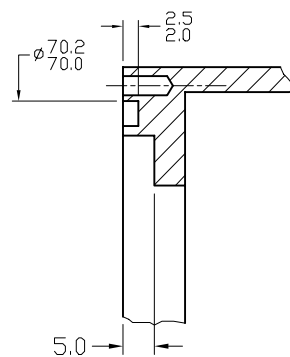
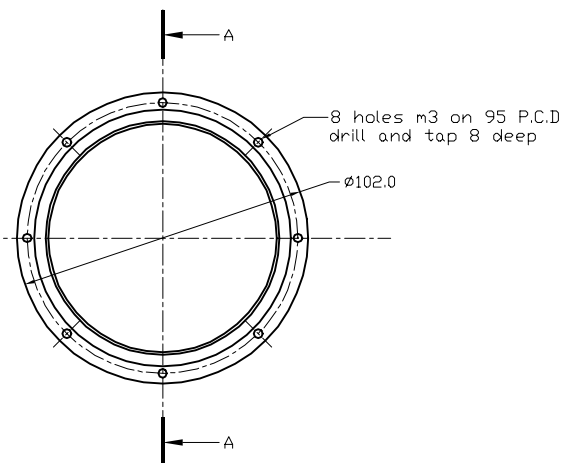
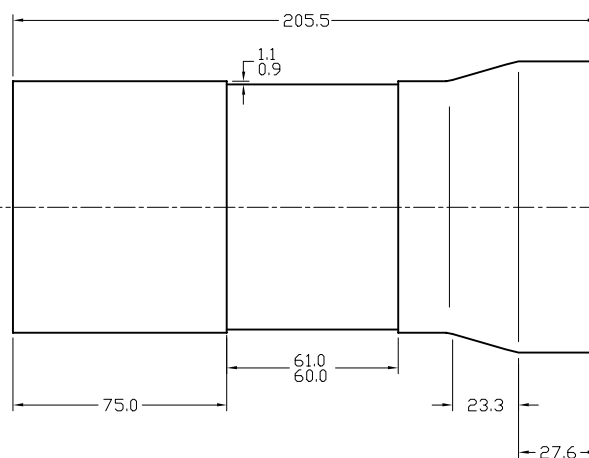
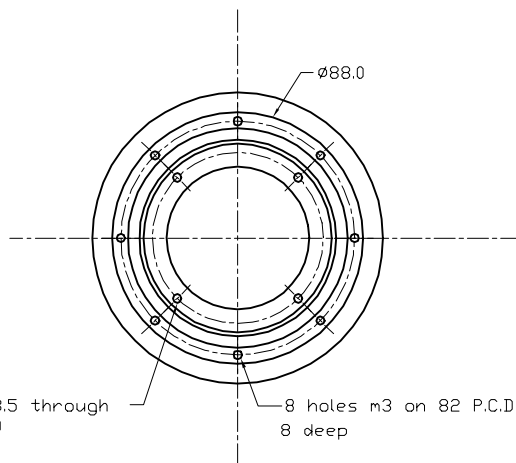
Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$ DIMENSIONS
MILLIMETRESSurface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

3.6

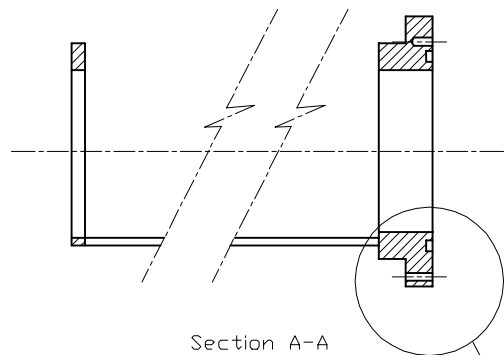


Section A-A



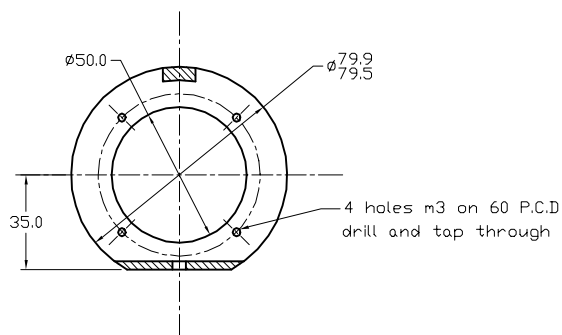
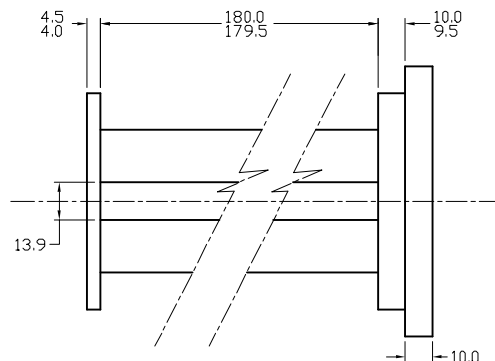
Detail A

A	--	Released for Fabrication	J.R.	03/08/98
Iss	Zone	Revision	Ref	DR Date
Drawn	Jack Reynolds	Scale	NTS	DO NOT SCALE
Date	July 7 1998	Material and Heat Treatment	Aluminium 6061 - T6	
Title	Outside case of camera housing			
Drawing No.	0001	Sheet	1	Of 7
THE AUSTRALIAN NATIONAL UNIVERSITY	DEPARTMENT OF ENGINEERING (FEIT)	Sheet Size	A3	
Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506				

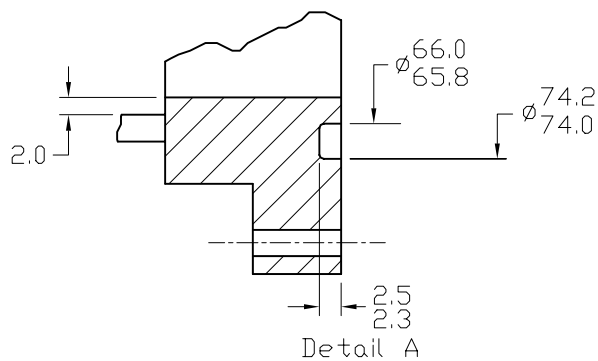
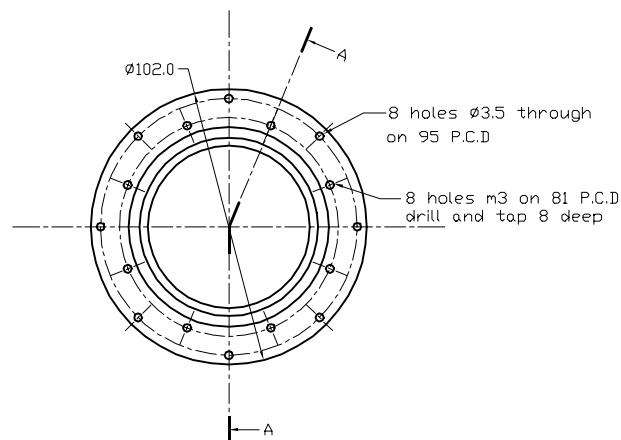
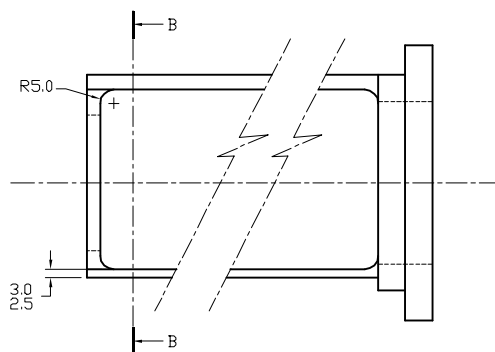


Section A-A

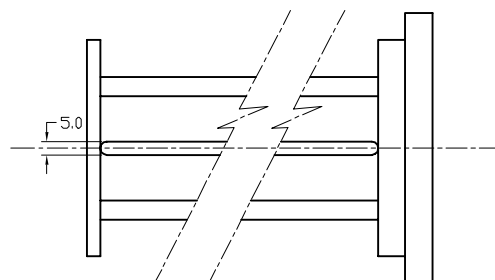
See detail A



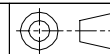
Section B-B



Detail A



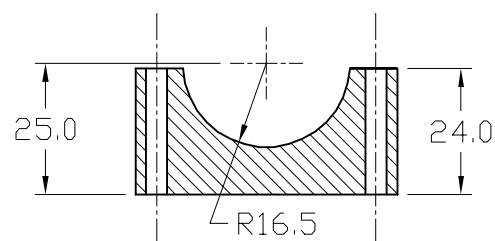
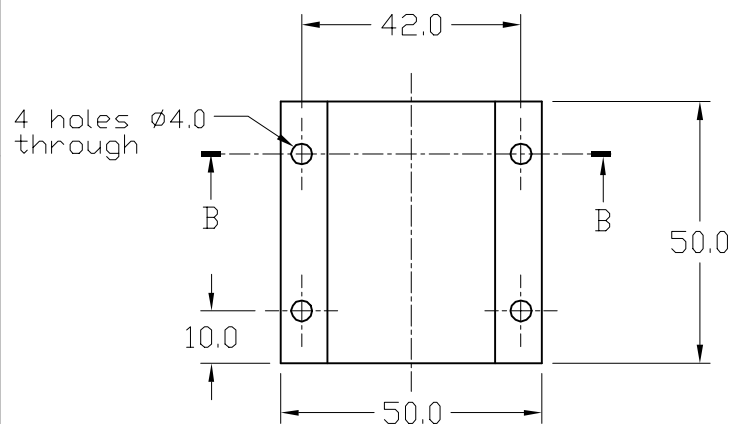
COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$ DIMENSIONS
MILLIMETRESSurface Finishes are in
micro metres Ra

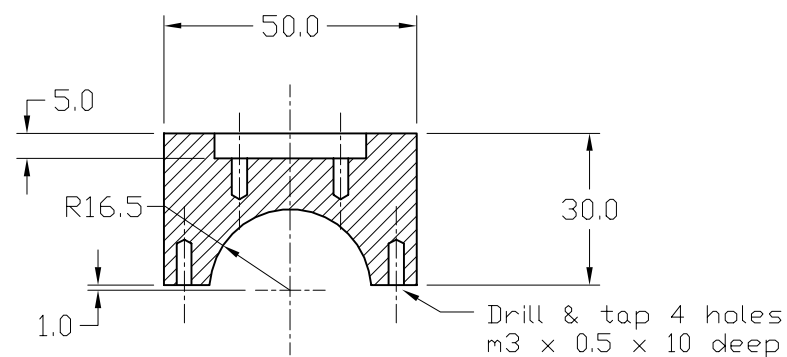
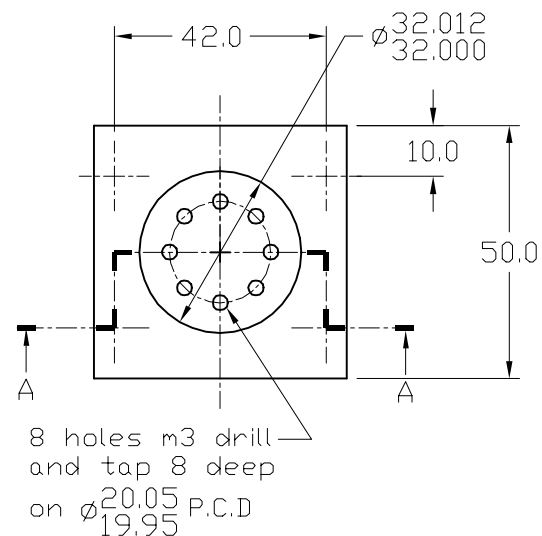
Remove all burrs & sharp edges

3.6

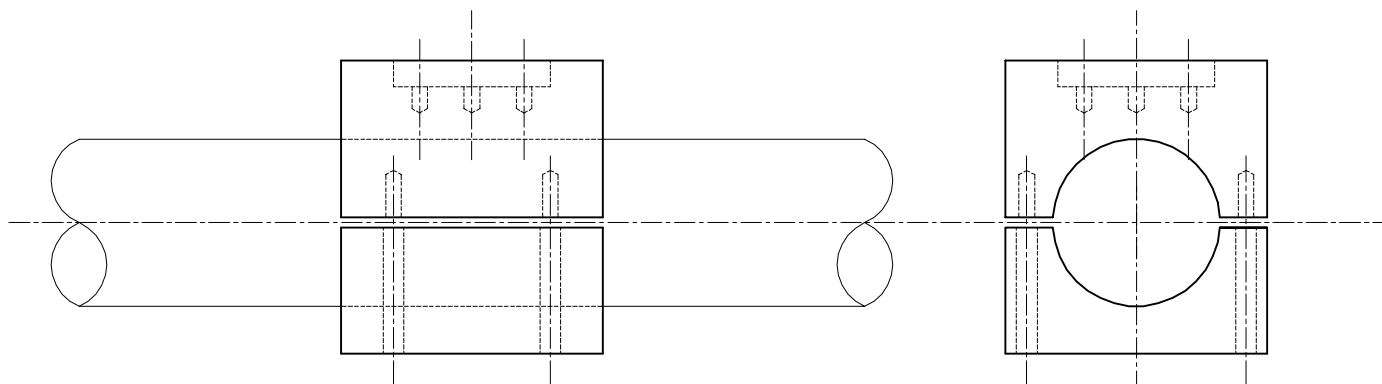
A	--	Released for Fabrication	J.R.	03/08/98
Iss	Zone	Revision	Ref	DR Date
Drawn	Scale	DO NOT SCALE	Date	
Jack Reynolds	NTS		July 7 1998	
Material and Heat Treatment				
Aluminium 6061 - T6				
Title				
Inner Frame of Camera Housing				
Drawing No.			Sheet 2	
0002			Of 7	
THE AUSTRALIAN NATIONAL UNIVERSITY			Sheet	
DEPARTMENT OF ENGINEERING (FEIT)			Size	
Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506			A3	



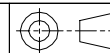
Section B-B



Section A-A



COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$ DIMENSIONS
MILLIMETRESSurface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

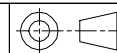
3.6

A	--	Released for Fabrication	J.R.	03/08/98
Iss	Zone	Revision	Ref	DR
Drawn	Scale	DO NOT SCALE	Date	July 7 1998
Jack Reynolds	NTS			
Material and Heat Treatment				
Aluminium 6061 - T6				
Title				
Frame Clamp Upper and Lower				
Drawing No.				Sheet 3
0003				Of 7
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506				Sheet Size A3

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
 Linear ± 0.5 Angular $\pm 1^\circ$

DIMENSIONS
 MILLIMETRES



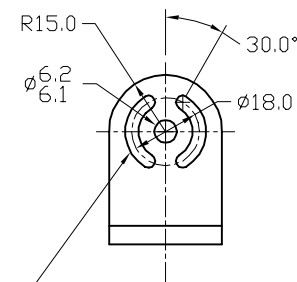
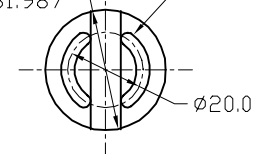
Surface Finishes are in
 micro metres Ra

Remove all burrs & sharp edges

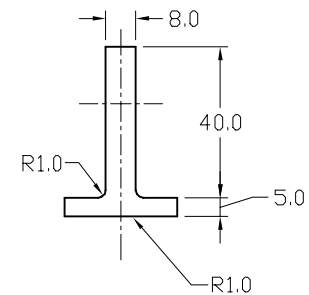
3.6

Radius to ctr lines
 on curved grooves
 1/8" dia groove

$\phi 32.000$
 31.987

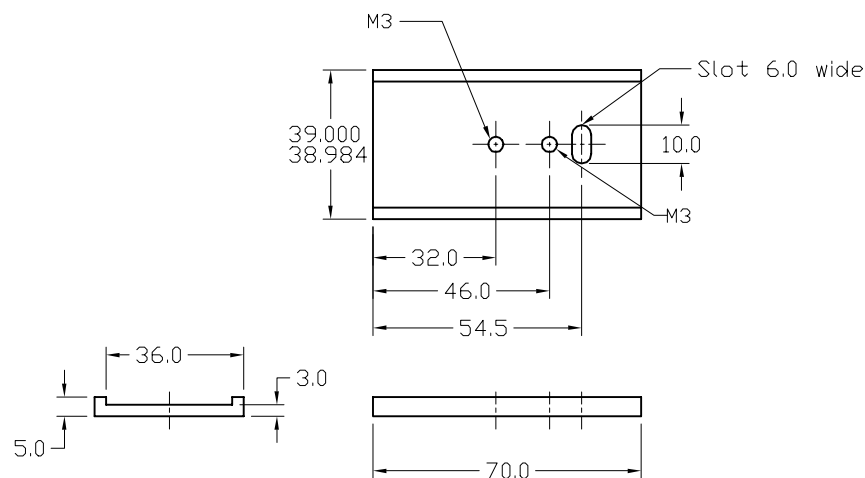
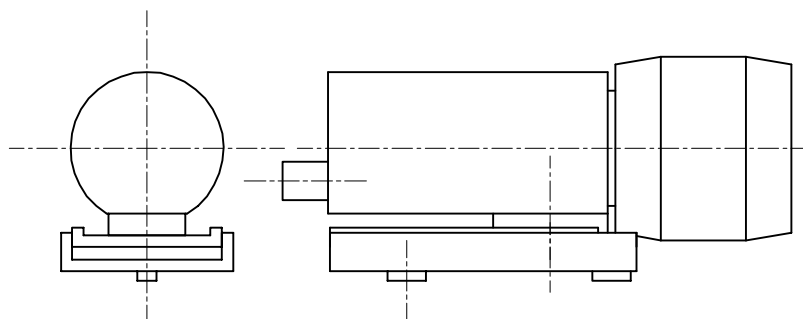


Radius to ctr lines
 on curved grooves
 1/8" dia groove

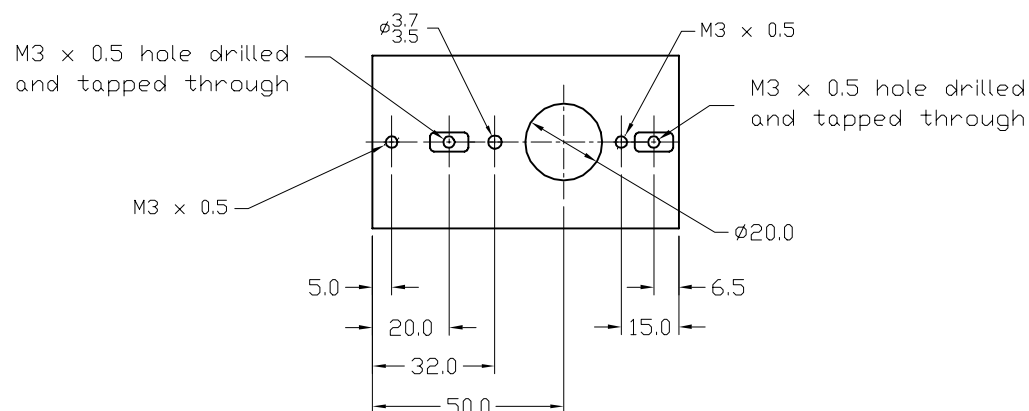


A	--	Released for Fabrication	J.R.	03/08/98
Iss	Zone	Revision	Ref	DR Date
Drawn	Jack Reynolds	Scale	NTS	DO NOT SCALE
Date	July 7 1998			
Material and Heat Treatment				
Aluminium 6061 - T6				
Title				
Outer Case Clamp and Joint				
Drawing No.	0004	Sheet	4	
		Of	7	
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506				Sheet Size A3

1



Upper Plate



Lower Plate

1

8

7

6

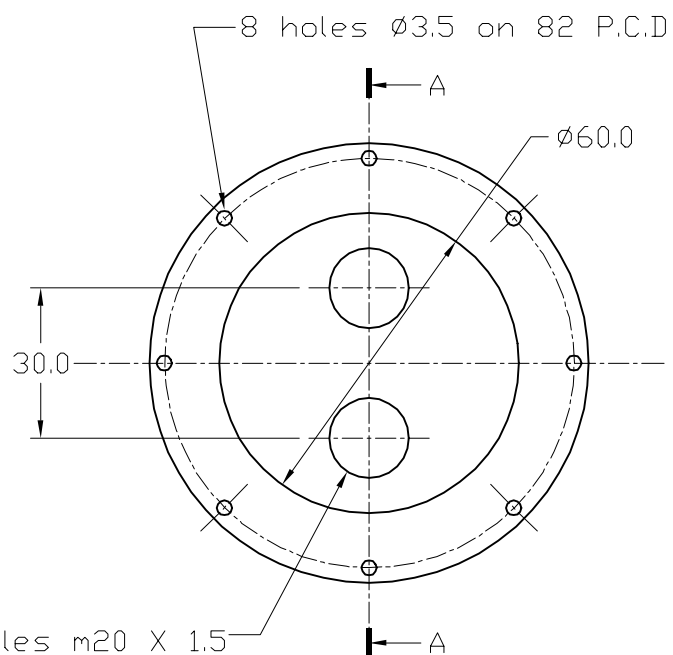
5

4

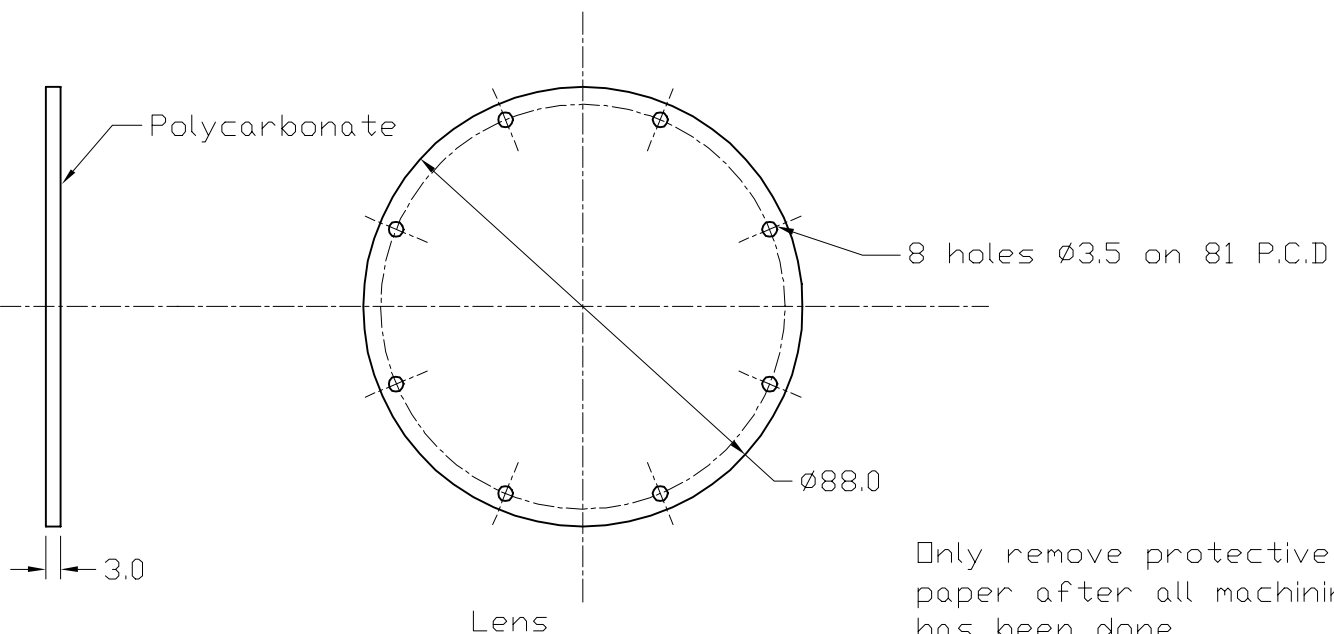
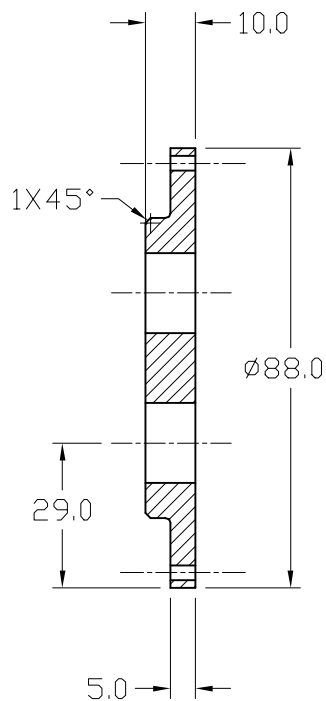
3

2

1



Back Plate

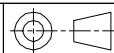


Only remove protective paper after all machining has been done

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.05 Angular $\pm 0.5^\circ$

DIMENSIONS
MILLIMETRES




Surface Finishes are in
micro metres Ra

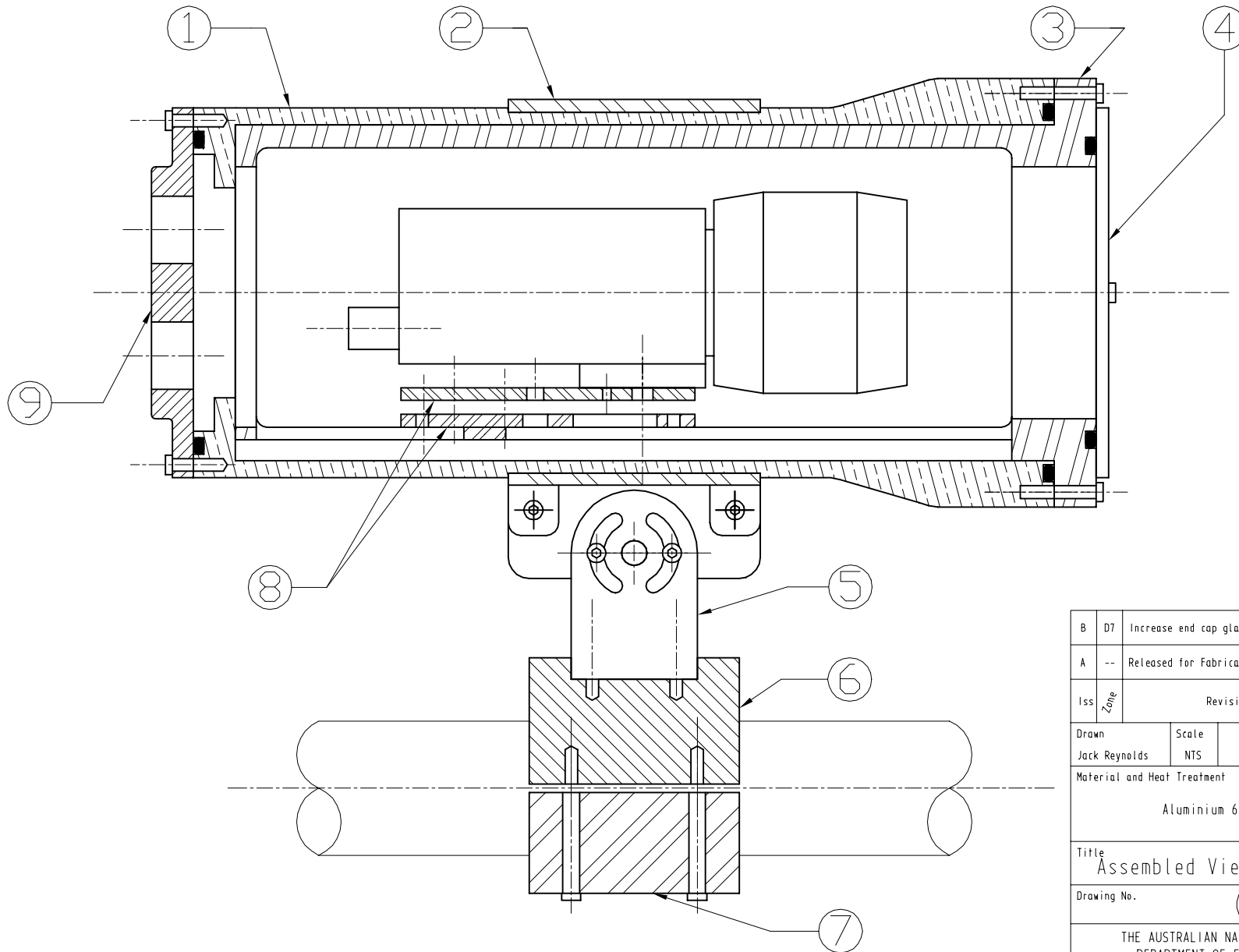
Remove all burrs & sharp edges



For all aluminium parts

B	D5	Increase size of holes on endcap		J.R.	15/09/98
A	--	Released for Fabrication		J.R.	03/08/98
Iss	Zone	Revision	Ref	DR	Date
Drawn	Scale	DO NOT SCALE		Date	
Jack Reynolds	NTS			July 7 1998	
Material and Heat Treatment					
Aluminium 6061 T6 and Polycarbonate					
Title					
Lens and Back Plate					
Drawing No.			0006	Sheet	6
				Of	7
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506					Sheet Size A3

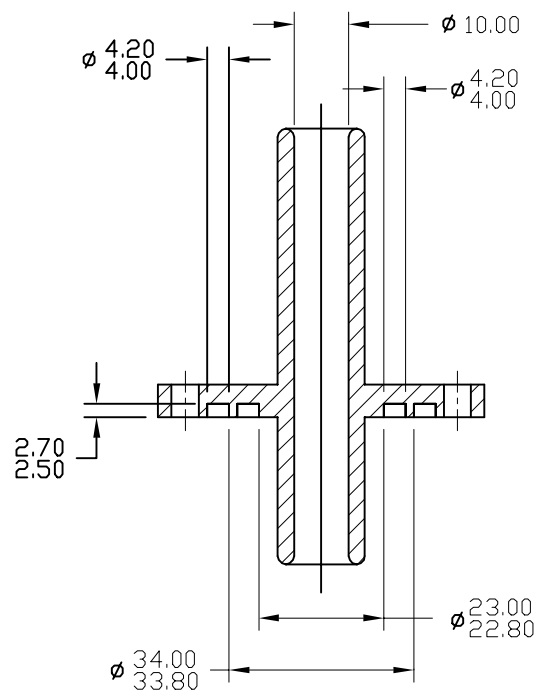
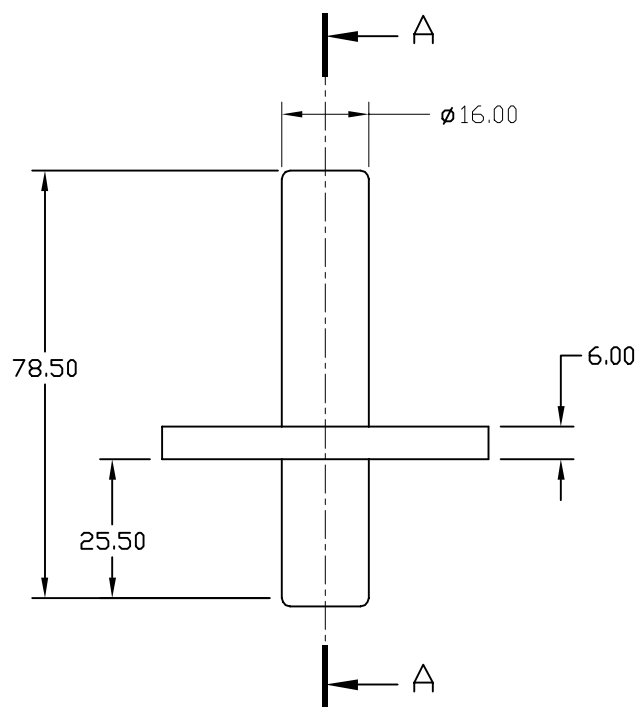
8		7		6		5		4		3		2		1	
Drawing No.	Number	Description	Drawing No.	Number	Description	Drawing No.	Number	Description	Drawing No.	COMPUTER DRAWING - DRAWN TO AS1100					
	1	Outer Case	0001	4	Lens	0006	7	Frame Clamp - Lower	0003	Tolerances unless otherwise specified Linear ± 0.05 Angular $\pm 0.5^\circ$		DIMENSIONS MILLIMETRES			
	2	Outer Case Clamp	0004	5	Outer Clamp Joint	0004	8	Camera Plates - Upper and Lower	0005	Surface Finishes are in micro metres Ra Remove all burrs & sharp edges					
	3	Inner Case	0002	6	Frame Clamp - Upper	0003	9	Back Plate	0006						



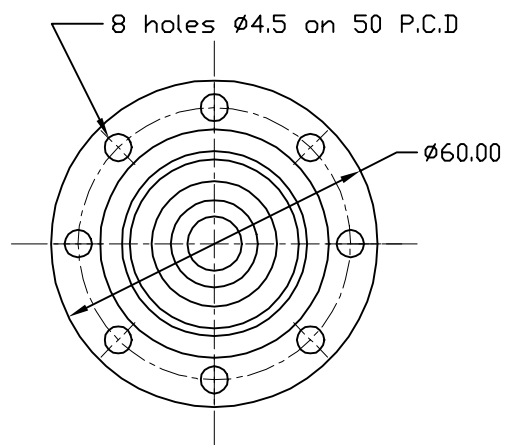
B	D7	Increase end cap gland holes		J.R.	15/09/98
A	--	Released for Fabrication		J.R.	03/08/98
Iss	Zone	Revision	Ref	DR	Date
Drawn	Jack Reynolds	Scale	NTS	DO NOT SCALE	Date
Material and Heat Treatment					
Aluminium 6061 T6 and Polycarbonate					
Title					
Assembled View - Camera Housing					
Drawing No.				Sheet	7
0007				Of	7
THE AUSTRALIAN NATIONAL UNIVERSITY					Sheet
DEPARTMENT OF ENGINEERING (FEIT)					Size
Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506					A3

Appendix I AUV Endcap Diagrams

Drawing Number	Drawing Name
0010	Fibre Cable Gland
0011	Gland Plate
0021	Cable Entry Plate
0022	Upper Arch
0023	Lower Arch
0024	End Plate
0025	Fastening Loop and Grips
0026	Endcap For Kambara



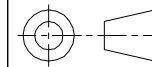
Section A-A



COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$

DIMENSIONS
MILLIMETRES



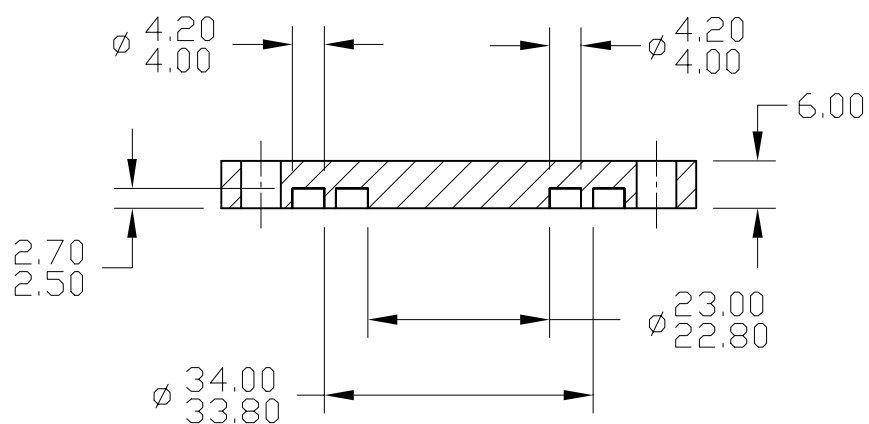
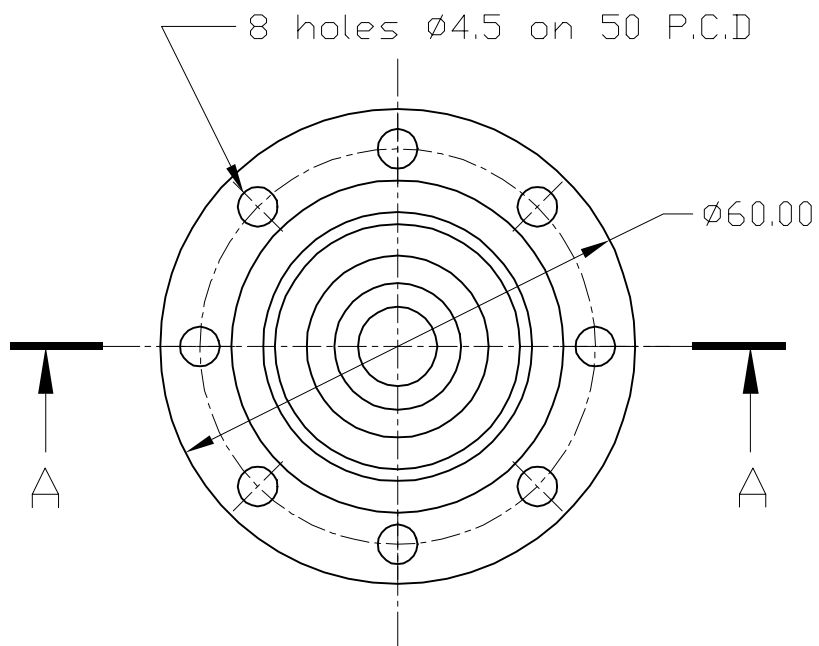
Surface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

1.6

0.5 x 45° chamfer on all edges

A	--	Released for Fabrication			J.R.	09/07/98	
Iss	Zone	Revision			Ref	DR	Date
Drawn		Scale	DO NOT SCALE			Date	
JR & DW		NTS				July 9 1998	
Material and Heat Treatment							
Aluminium 6061 - T6							
Title							
Fibre Cable Gland							
Drawing No.						Sheet	1
0010						Of	2
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506						Sheet Size	A4

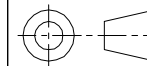


Section A-A

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$

DIMENSIONS
MILLIMETRES



Surface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

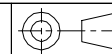
1.6

0.5 x 45° chamfer on all edges

A	--	Released for Fabrication	J.R.	09/07/98
Iss	Zone	Revision	Ref	DR
Drawn	Scale	DO NOT SCALE	Date	July 9 1998
JR & DW	NTS			
Material and Heat Treatment				
Aluminium 6061 - T6				
Title				
Gland Plate				
Drawing No.			Sheet	2
0011			Of	2
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506			Sheet	Size
			A4	

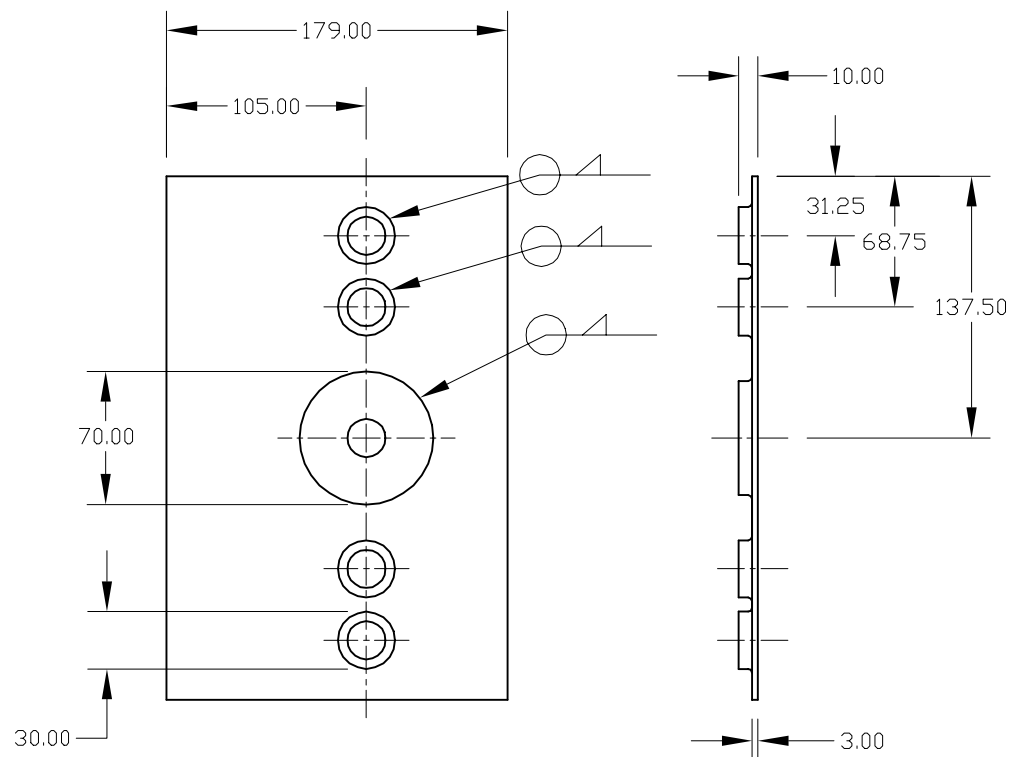
COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified

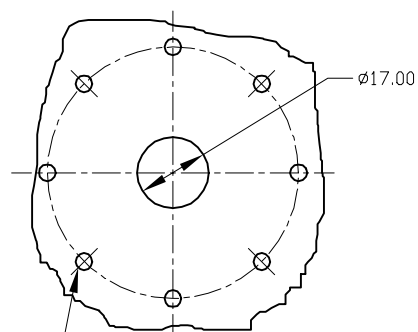
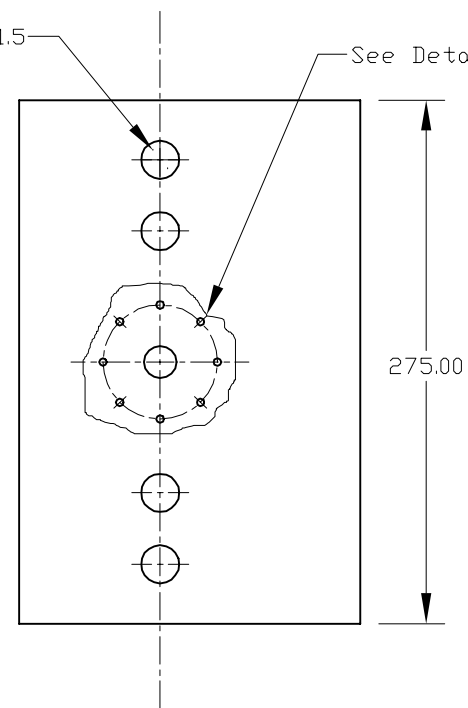
Linear ± 0.5 Angular ± 0.5 DIMENSIONS
MILLIMETRESSurface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

1.6

4 holes m20 x 1.5
fine pitch

See Detail 1

8 holes m4 on 50 P.C.D.
Drill and tap 8 deep

Detail 1

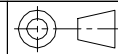
B	E2	Increased gland holes		J.R.	15/09/98
A	--	Released for Fabrication		J.R.	09/07/98
Iss	Zone	Revision		Ref	DR Date
Drawn JR & DW		Scale NTS	DO NOT SCALE		Date July 9 1998
Material and Heat Treatment					
Aluminium 6061					
Title					
Cable Entry Plate					
Drawing No. 0021				Sheet Of	1 6
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506				Sheet Size A3	

8 7 6 5 4 3 2 1

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$

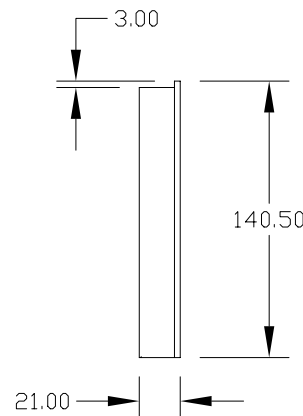
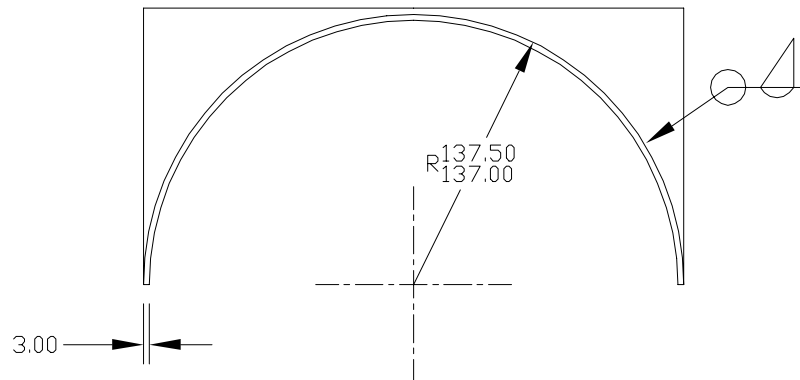
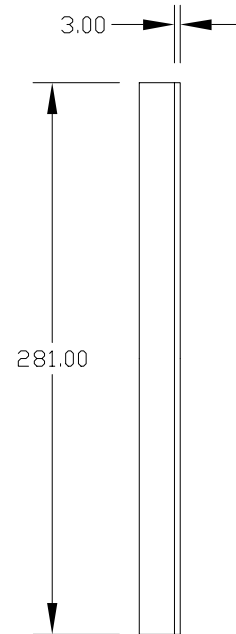
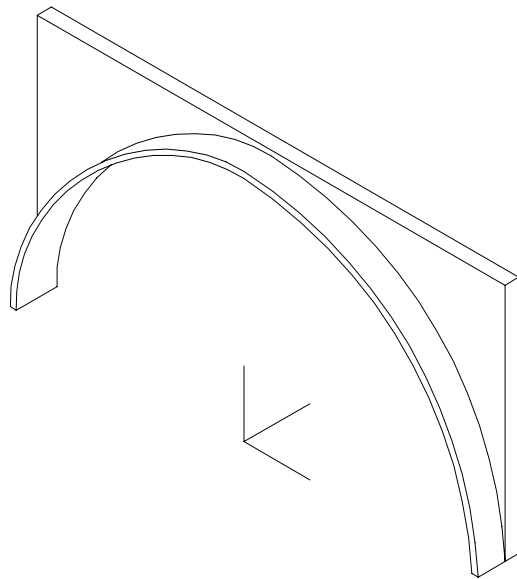
DIMENSIONS
MILLIMETRES



Surface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

1.6



A	--	Released for Fabrication		J.R.	09/07/98
Iss	Zone	Revision	Ref	DR	Date
Drawn	JR & DW	Scale	NTS	DO NOT SCALE	Date
Material and Heat Treatment					
Aluminium 6061					
Title					
Upper Arch					
Drawing No.				Sheet	2
0022				Of	6
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506				Sheet	Size
				A3	

Drawing No.

8

7

6

5

4

3

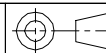
2

1

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$

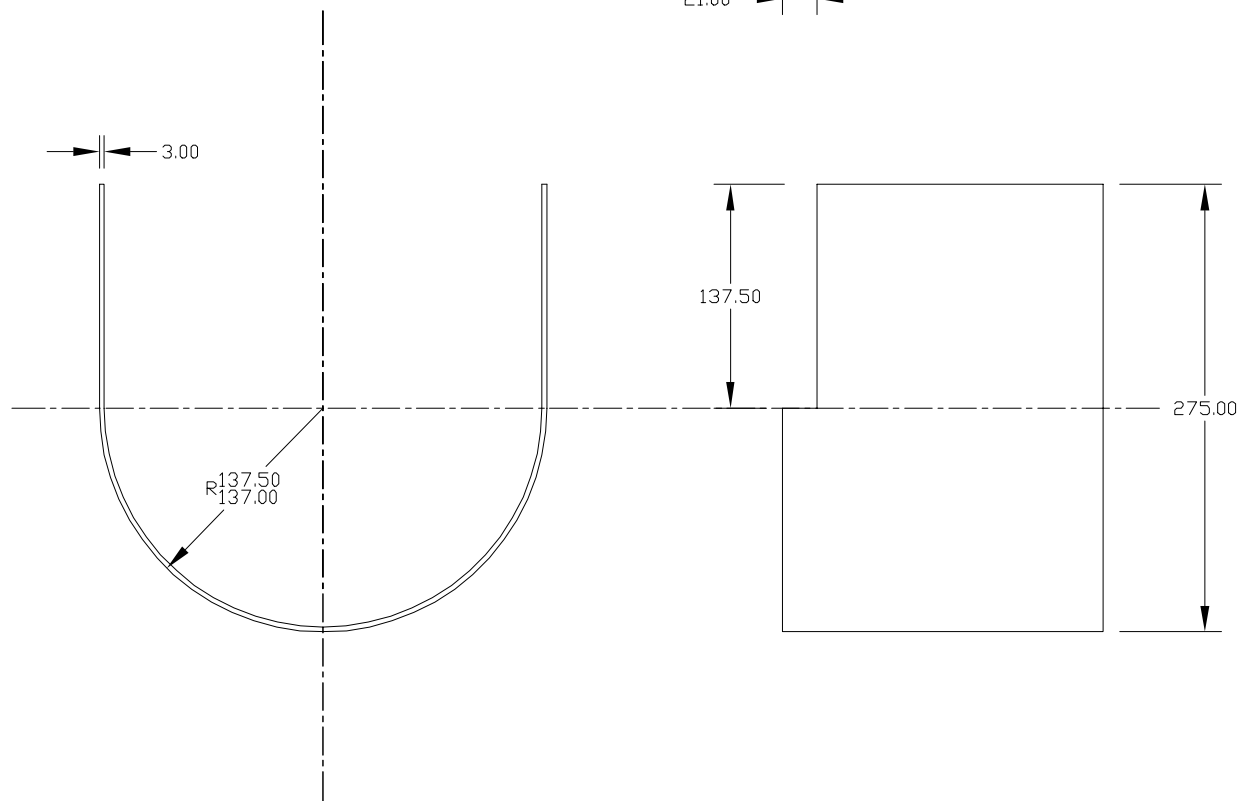
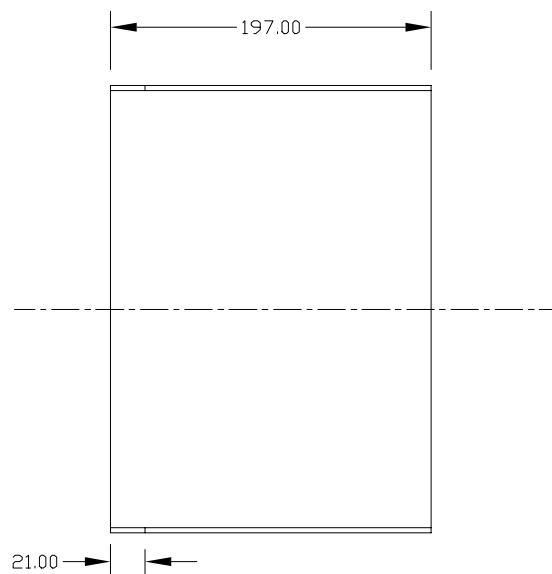
DIMENSIONS
MILLIMETRES



Surface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

1.6
▽



A	--	Released for Fabrication		J.R.	09/07/98	
Iss	Zone	Revision		Ref	DR	Date
Drawn JR & DW		Scale NTS	DO NOT SCALE		Date July 9 1998	
Material and Heat Treatment						
Aluminium 6061						
Title Lower Arch						
Drawing No. 0023				Sheet 4 Of 6		
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506					Sheet Size A3	

Drawing No.

8

7

6

5

4

3

2

1

3.00

275.00

275.00

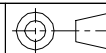
137.50

R137.50

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$

DIMENSIONS
MILLIMETRES



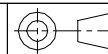
Surface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

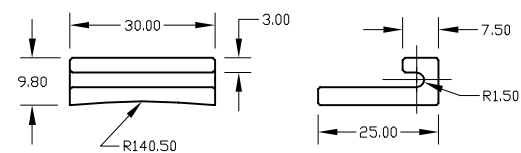
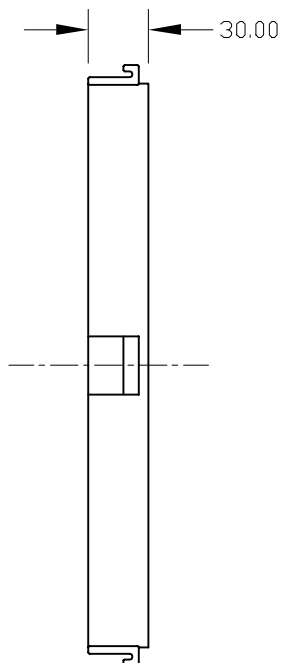
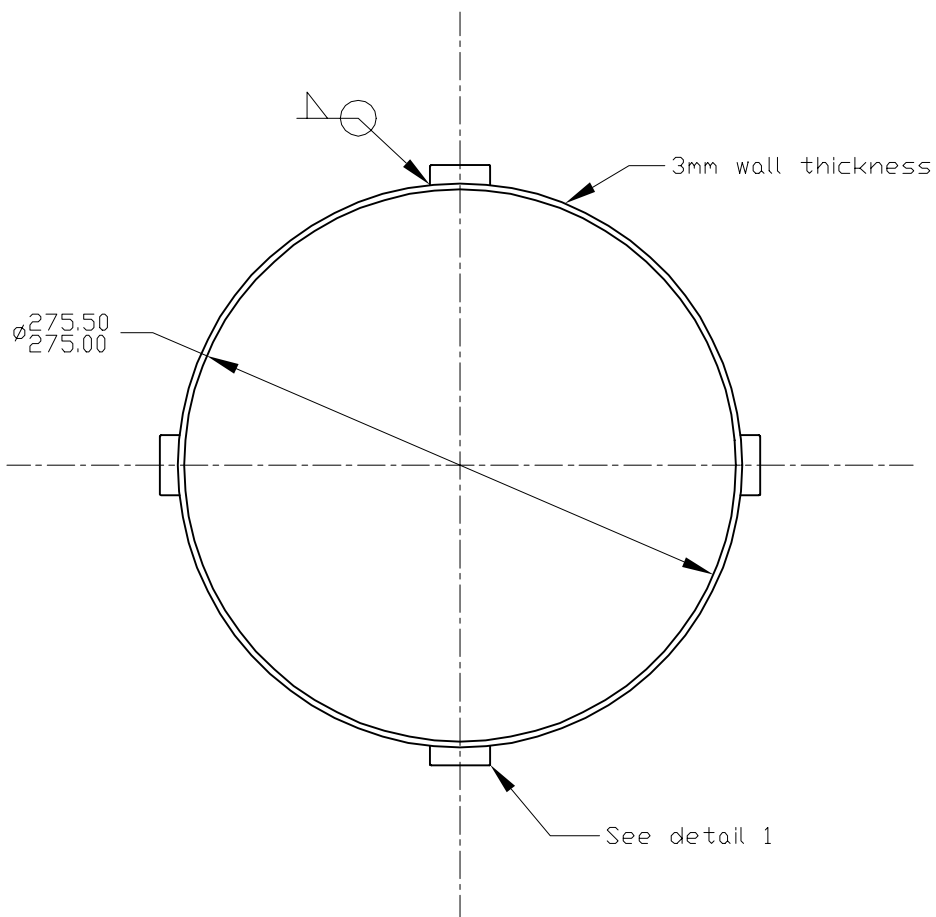
1.6
▽

A	--	Released for Fabrication			J.R.	09/07/98
Iss	Zone	Revision			Ref	DR Date
Drawn JR & DW		Scale NTS	DO NOT SCALE			Date July 9 1998
Material and Heat Treatment						
Aluminium 6061						
Title						
End Plate						
Drawing No.					Sheet Of	4 6
					Sheet Size	A3
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506						

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$ DIMENSIONS
MILLIMETRESSurface Finishes are in
micro metres Ra

Remove all burrs & sharp edges

1.6


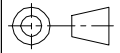
Detail 1

A	--	Released for Fabrication		J.R.	09/07/98
Iss	Zone	Revision	Ref	DR	Date
Drawn	JR & DW	Scale	NTS	DO NOT SCALE	Date
Material and Heat Treatment					
Aluminium 6061					
Title					
Fastening loop and grips					
Drawing No.				Sheet	5
0025				Of	6
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506				Sheet	Size
				A3	

COMPUTER DRAWING - DRAWN TO AS1100

Tolerances unless otherwise specified
Linear ± 0.5 Angular $\pm 1^\circ$

DIMENSIONS
MILLIMETRES

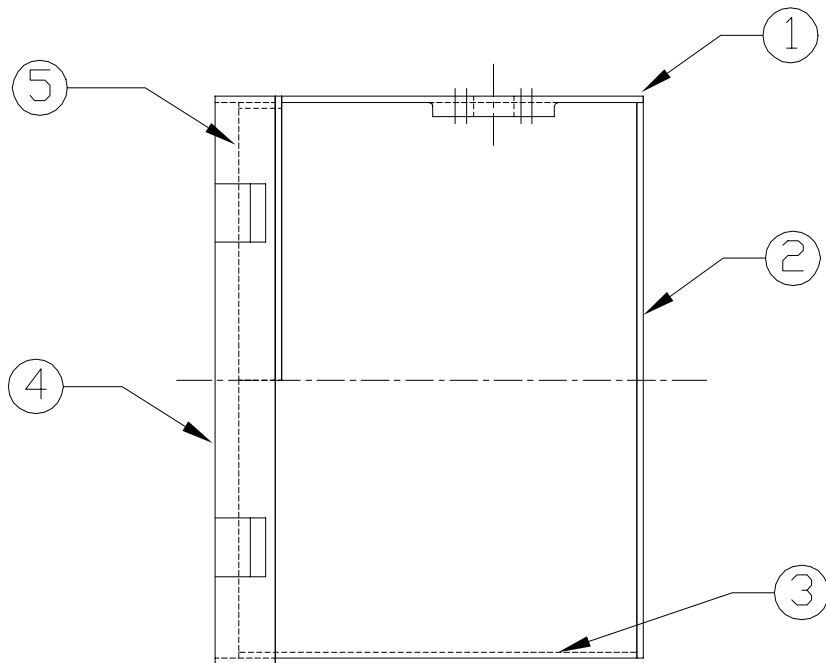


Surface Finishes are in
micro metres Ra

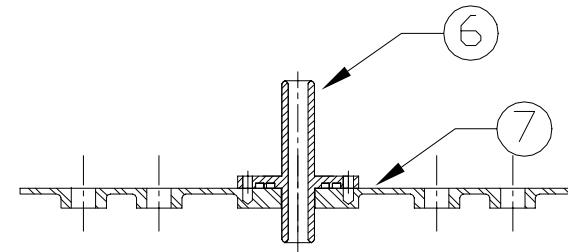
Remove all burrs & sharp edges

1.6
▽

No.	Description	Drawing No.
1	Cable Entry Plate	0021
2	End Plate	0024
3	Lower Arch	0023
4	Fastening Loop and grips	0025
5	Upper Arch	0022
6	Fibre Cable Gland	0010
7	Cable Entry Plate	0021



Assembled View



Sectional assembled view
of gland interface

A	--	Released for Fabrication		J.R.	09/07/98	
Iss	Zone	Revision		Ref	DR	Date
Drawn		Scale	DO NOT SCALE		Date	
JR & DW		NTS			July 9 1998	
Material and Heat Treatment						
Aluminium 6061						
Title						
Endcap for Kambara						
Drawing No.				0026		Sheet 6 Of 6
THE AUSTRALIAN NATIONAL UNIVERSITY DEPARTMENT OF ENGINEERING (FEIT) Canberra, ACT, 0200, ph 06-2493378, fax 06-2490506						Sheet Size A3