FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Underwater Stereoscopic Vision with Convergent Cameras

João Paulo Meneses Aguiar

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Nuno A. Cruz

Second Supervisor: Andry M. Pinto

October 30, 2015

© João Aguiar, 2015

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MESTRADO INTEGRADO EM ENGENHARIA 2014/2015
ELETROTÉCNICA E DE COMPUTADORES

A Dissertação intitulada

"Underwater Stereoscopic Vision with Convergent Cameras"

foi aprovada em provas realizadas em 16-10-2015

o júri

Presidente Professor Doutor Paulo José Cerqueira Gomes da Costa
Professor Auxiliar do Departamento de Engenharia Eletrotécnica e de Computadores
da Faculdade de Engenharia da Universidade do Porto

Professor Doutor José Luís Magalhães Lima
Professor Adjunto do Departamento de Eletrotecnia da Escola Superior de
Tecnologia e Gestão do Instituto Politécnico de Bragança

Mestre Nuno Alexandre Lopes Moreira da Cruz
Assistente do Departamento de Engenharia Eletrotécnica e de Computadores da
Faculdade de Engenharia da Universidade do Porto

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.

João Paulo Meneses Aguiar

Autor - João Paulo Meneses Aguiar

Faculdade de Engenharia da Universidade do Porto

Resumo

A visão por computador tem vindo a ser cada vez mais utilizada no mundo da robótica subaquática para a reconstrução de ambientes marinhos, para o mapeamento do fundo oceânico ou até na procura de alguma informação relevante como *pipelines* ou destroços em águas profundas. De forma a serem adquiridos bons resultados, é realmente importante que a informação obtida tenha a melhor precisão possível. Esta dissertação focou-se na validação conceptual e na criação de um novo sensor de medição de distâncias subaquático baseado em câmaras rotativas, capaz de ser instalado num veículo subaquático. Foram realizadas experiências de forma a determinar, para um determinado objecto (a uma determinada distância), qual o ângulo que é mais adequado ao posicionamento relativo do objeto em interesse que se encontra no campo de visão, isto é, a convergência que resulta no menor erro de distância medida pelas câmaras. É estudado e analisado também o impacto de diferentes configurações das duas câmaras deste sistema estereoscópico no erro de uma nuvem de pontos tridimensionais. Para além disso, os resultados foram obtidos considerando dois cenários de teste: o sistema estereoscópico com *baseline* de 0.11m e de 0.29m. Os testes experimentais foram todos realizados num ambiente subaquático. Com este conceito de câmaras estereoscópicas convergentes, o erro na medição de distâncias obtido foi de cerca de 5% para distâncias compreendidas entre 1.125 e 3.125 metros. Um método de calibração para a estimação dos parâmetros extrínsecos foi também apresentado com o objectivo de ser implementado por forma a que o sensor possa então ser utilizado numa aplicação real, ajustando de forma automática os seus próprios parâmetros. O conceito de visão estereoscópica com câmaras convergentes que é apresentado nesta dissertação poderá vir a ser instalado num veículo subaquático no futuro.

Abstract

The computer vision is being increasingly used in underwater robotics in reconstruction of marine environments, seabed mapping or even in searching for some relevant information like pipelines or wreckage in deep waters. It is really important, in order to acquire good results, that the information obtained has the best accuracy as possible. This work focused in the conceptual validation and creation of a new underwater ranging sensor based in rotating cameras, capable of being installed in an underwater vehicle. It was made experiments to determine for a given object (at a certain distance), which angle is the most suitable for the relative position of an object located in the field of view, in other words, which converging angle results in the smallest distance error measured by the cameras. This paper studies and analyses the impact of the relative orientation of two cameras in regard to the error in the three-dimensional point cloud. Furthermore, the results were obtained by considering two testing scenarios: a pair of cameras with baselines of 0.11m and 0.29m. The experimental tests were all made in underwater environment. With this concept of stereoscopic convergent cameras, the range measurement errors obtained were as low as 5% for distances between 1.125 and 3.125 meters. A calibration method for extrinsic parameters estimation is also introduced with the objective to be implemented, so this sensor could be utilized in a real application, adjusting automatically its own parameters. The stereoscopic vision concept with convergent cameras that is presented in this document could be installed in an underwater vehicle in the future.

Agradecimentos

Este documento e todo o tempo dispensado na realização do mesmo é dedicado aos meus pais. Apesar de estarem longe, sempre me souberam moralizar, apoiar e sempre aceitaram todas as minhas decisões. Foi graças a eles que aqui cheguei e, por isso, dedico-lhes todo este docmento, visto que sem o apoio dos mesmos, muito dificilmente seria quem sou hoje.

Gostaria também de agradecer ao meu irmão, pelos conhecimentos que me passou ao longo deste percurso como estudante, a todos aos conselhos que me deu neste tempo e à sempre prontidão e disponibilidade que tem para comigo.

Deixo aqui também um agradecimento especial aos grandes amigos de preto que cá fiz para a vida. Não os vou nomear, pois eles sabem perfeitamente quem são. Foi graças a amigos como estes que vivi coisas que nunca antes tinha vivido e que provavelmente nunca mais irei viver com a mesma intensidade.

Por fim, mas não menos importante, gostaria também de agradecer ao prof. Nuno Cruz e ao Dr. Andry Pinto por acreditarem no meu trabalho, apesar da minha constante preguiça e demora na resposta aos emails.

João Aguiar

"The true sign of intelligence is not knowledge but imagination."

Albert Einstein

Contents

Abbreviations xvii

1 Introduction 1
- 1.1 Context and Motivation . . . 1
- 1.2 Goals . . . 2
- 1.3 Thesis Structure . . . 2

2 State of The Art 3
- 2.1 Underwater Range Measurement Sensors . . . 3
- 2.2 Stereo Correspondence Algorithms . . . 4
  - 2.2.1 Subtypes of Stereo Correspondence Algorithms . . . 5
  - 2.2.2 Some Stereo Correspondence Algorithms examples . . . 6
- 2.3 Camera Calibration . . . 8
- 2.4 Effect of Alignment Errors . . . 9

3 Acquisition of 3D Information 11
- 3.1 Camera Calibration . . . 11
- 3.2 Disparity to 3D coordinates . . . 14
  - 3.2.1 Brief Explanation of Stereoscopic Range Measurement . . . 14

4 Stereoscopic Range Measurement Error Analysis with Convergent Cameras 17
- 4.1 Stereoscopic System Based on Rotating Cameras . . . 18
- 4.2 Software Utilized for Stereoscopic Experiences . . . 18
- 4.3 Experience Scene Description . . . 19
- 4.4 Stereoscopic Correspondence Algorithms Performance Comparison . . . 20
- 4.5 Methods . . . 23
- 4.6 Results and Discussion . . . 25
- 4.7 Stereo System Sensitivity - Experimental Misalignment Effect . . . 30

5 Auto-Rotating Cameras for Precise Range Measurement 31
- 5.1 Explanation of Camera Extrinsic Parameters Matrix . . . 32
- 5.2 Our Approach . . . 33
  - 5.2.1 Extrinsic Calibration using IMUs Method . . . 33
  - 5.2.2 Extrinsic Calibration using Quadratic Function Approximation . . . 34

6 Conclusions and Future Work 41
- 6.1 Future Work . . . 42

References 43

List of Figures

2.1 Basics of local (window) correspondence algorithm where the squares represent pixels. The filled square is the pixel that is being compared and it is being used a 3x3 scan window. . . . 6
2.2 Representation of intrinsic parameters where f is focal length and p is the principal point [1]. . . . 8
2.3 Representation of extrinsic parameters where R is the rotation matrix and T represents the translation vector. . . . 8
2.4 Stereo camera rig angle error sources θ, γ and ϕ. [2] . . . 9

3.1 Comparison between original image and the respective undistorted. . . . 12
3.2 Comparison between original images of left and right views (on top) and the respective undistorted and rectified ones (on bottom), having left view as reference. 13
3.3 Comparison between original images of left and right views with cameras converged 11.02 degrees (top) and the respective undistorted and rectified ones (on bottom), having left view as reference. . . . 13
3.4 Rectified images black margin justification. . . . 14
3.5 Stereoscopy theory example. . . . 14
3.6 Image planes with parallel epipolar lines represented. . . . 15

4.1 Setup used in experimental tests. . . . 17
4.2 Software used in experiences with the two camera preview windows and the step controller slider. . . . 18
4.3 Image captured from the scene used in this experimental tests. . . . 20
4.4 Disparity maps of the tsukuba image processed by various algorithms (left column: groundtruth and SGBM; right column: BM and ELAS). . . . 21
4.5 Disparity maps of the underwater images processed by various algorithms (left column: original image and SGBM; right column: BM and ELAS). . . . 22
4.6 Diagram of the steps used to make objects distance measurement. . . . 23
4.7 Comparison between original images (on top) of both views and the respective undistorted and rectified ones (on bottom). . . . 24
4.8 Resulted disparity map of the pair of the previous images. . . . 25
4.9 Distance error from the several objects to the cameras with different angles (baseline = 0.29m). . . . 26
4.10 Distance error from the several objects to the cameras with different angles (baseline = 0.11m). . . . 28
4.11 Comparison of the best converging cameras case (smaller distance errors) with different baselines (black bars represents results with baseline = 0.29m and gray bars with baseline = 0.11m). . . . 29

5.1 Diagram of the steps used to make objects distance measurement. 31
5.2 Camera possible rotations. From left to right: rotation over X axis (camera side view), rotation over Y axis (camera top view) and rotation over Z axis (camera back view). 32
5.3 Camera with IMU on top. The red box represents the IMU sensor. 34
5.4 Angles and translation obtained in 0.29m baseline experimental tests. 35
5.5 Angles and translation obtained in 0.11m baseline experimental tests. 36
5.6 Results obtained in underwater range measurement using this quadratic approximation method (with baseline = 0.29m) where black bars represents the results of this experiment, the gray bars the results of the experiment of the last chapter and the smaller gray lines the standard deviation. 38
5.7 Results obtained in underwater range measurement using this quadratic approximation method. (with baseline = 0.11m) 39

List of Tables

2.1 Influence of camera angles misalignment in range measurement [2]. 10

3.1 Intrinsic calibration results. 11

4.1 Elapsed time in disparity map processing for each algorithm using an i5 2.4GHz processor with images having the resolution of 450x375 pixels. 21
4.2 Elapsed time in disparity map processing for each algorithm from underwater scene using an i5 2.4GHz processor with an image with the resolution of 1292x964 pixels. 23
4.3 Distance error with different cameras angles (baseline = 0.29m). 25
4.4 Distance error with different cameras angles (baseline = 0.11m). 27
4.5 The best angles for both baselines obtained in experimental results 29

5.1 Distance error using quadratic approximation with different cameras angles (baseline = 0.29m). 37
5.2 Distance error using quadratic approximation with different cameras angles (baseline = 0.11m) where black bars represents the results of this experiment, the gray bars the results of the experiment of the last chapter and the smaller gray lines the standard deviation. 37

Abbreviations and Symbols

1D	1 Dimension
2D	2 Dimensions
3D	3 Dimensions
AUV	Autonomous Underwater Vehicles
BM	Block Matching
CUDA	Compute Unified Device Architecture
DMP™	Digital Motion Processor™
ELAS	Efficient Large Scale Stereo Matching
FEUP	Faculty of Engineering of the University of Porto
GPU	Graphics Processing Unit
I2C	Inter-Integrated Circuit
IMU	Inertial Measurement Unit
LiDAR	Light Detection and Ranging
NCC	Normalized Cross-Correlation
OceanSys	Ocean Systems Group
PC	Personal Computer
PVC	Polyvinyl Chloride
SAD	Sum of Absolute Differences
SGBM	Semi-Global Block Matching
SNCC	Summed Normalized Cross-Correlation
SONAR	Sound Navigation and Ranging
SSD	Sum of Squared Differences
ToF	Time of Flight
WTA	Winner-Take All

Chapter 1

Introduction

1.1 Context and Motivation

In the last decades there have been several investigators trying to use the advantages that computer vision could bring to the aquatic environment, specially to improve perception. Nonetheless, the visual information that is acquired by underwater applications is often affected by several issues related to the severe physical conditions and the light propagation in deep waters.

The ability to understand its surrounding environment (with the highest detail as possible) is one important feature to turn the autonomous underwater vehicles (AUV) even more intelligent and independent. This is quite relevant for mapping underwater structures, obstacle avoidance and accurate location of itself using only its sensors.

Nowadays, there are technological solutions already available for robotic applications, such as Light Detection and Ranging (LiDAR), which calculates the distance using light propagation, or the Sound Navigation and Ranging (SONAR), that calculates the distance using sound waves. SONAR is the most used technology in underwater environments mainly due to the favourable properties of the sound propagation in water. Nevertheless, it presents advantages for greater distances, however for smaller distances the multiple reflections often result in noisy sensor data. The LiDAR sensor has a great precision but it produces poor results for greater distances in water due to the light propagation problems and absorption.

Another way to perform underwater range measurement is using stereoscopic cameras. It is considered a cheap solution and with this type of sensors, colour and visual texture may be preserved bringing more detailed data to the robot where sensors are installed.

This thesis follows the increasing use of cameras for underwater range measurements and, in this context, it presents and evaluates a novel concept for a reconfigurable stereo vision system. This document introduces a stereoscopic system that is being developed, which is based on convergent cameras. Therefore, this research studies the effect of changing the pose of the cameras in the acquisition and determines the accuracy of the point cloud that is expected to be obtained by the stereoscopic system. Moreover, the advantages of this non-conventional perceptual system are also presented. Experimental validations make it possible to analyse the performance of the

perceptual system in a realistic testing scenario. All the tests were performed in a water tank with some distinct objects at different ranges. Experimental results of the cameras with two different baselines (0.11m and 0.29m) are also presented. The camera rig is protected by an acrylic window from which the cameras can capture several frames and then analyse the distance to each object. Finally, a simple method for extrinsic camera calibration is proposed, in order to make this stereo rig system capable of being configurable and to work automatically in real-time in a real underwater vehicle.

1.2 Goals

The biggest goal of this thesis is to study, build and validate a non-conventional concept of a reconfigurable stereoscopic vision with 2 converged cameras for underwater applications. This system should compute a 3D point cloud in real-time from an underwater scene. In addition, a preliminary study comprises the comparison in terms of computational performance and quality of the disparity map that is obtained by some of the stereo correspondence algorithms that are currently available in the state-of-the-art. This study is relevant to evaluate which technique is more suitable for robotic underwater applications. A study of the advantages/disadvantages of converged cameras is also analyzed and discussed in this thesis. If this study brings some benefits to underwater range measurement, the system should also be capable to auto-calibrate every time a change of cameras pose happens so it could continuously work without interruption and more efficiently.

1.3 Thesis Structure

This document is organized as follows. Chapter 2 presents the state-of-the-art related with underwater ranging sensors and with stereo vision principles, while chapter 3 introduces how the acquisition of 3D information is made and exposes the results of the cameras calibration process.

Chapter 4 presents the non-conventional stereoscopic system that is formed by the converged cameras. The main objective of this chapter is to analyse if the proposing system introduces benefits in underwater ranging measurements or not. A method that enables the automatic reconfiguration of the system is presented in chapter 5 and the conclusions of this document in chapter 6.

Chapter 2

State of The Art

This chapter introduces some stereo correspondence algorithms that were proposed recently by researchers. In addition, technology for measuring the range in underwater environments are also described.

2.1 Underwater Range Measurement Sensors

The most used sensors in underwater applications are the acoustic sensors due to the advantages that sound has in water. They are used in several applications like obstacle detection [3] and localization. The work [4] studies the limitations and capabilities of sonar technology in underwater localization. In the experiments, the authors use a robotic fish equipped with small, low-power sounder (buzzer) and microphones. They achieved an underwater localization resolution of 0.02m over a range of 10m.

Other sensors that are commonly used in range measurement are LiDAR sensors, despite of the light attenuation and absorption problems when submerged [5]. On the other hand, they are very useful for small and accurate distances and several investigators are already trying to attenuate the undesirable backscattering problem [6] [7].

LiDAR sensors can be separated in two different groups: the ones that work with triangulation and those that use ToF (time of flight). The former has higher resolution (less than 1mm) than ToF but only for short ranges (less than 1m). The latter is better for distances greater than 2.5m and has a 5mm precision range for 8m. In the work [8], and for a range of 10m, the LiDAR prototype based on ToF achieved a precision of 30mm (Jerlov Type III) during on-the-fly 3D reconstructions. There are experiments using both techniques discussed above with more accurate results [9].

Another possibility for distance measurement is the use of a pair of cameras (like stereoscopy). In this case, it is possible to use an active or passive technique. The former needs a light source like a laser [10], [11] or even structured light [12], [13] while the latter uses two similar cameras to determine the 3D coordinates.

The work presented in [12] concludes that structured light can be used in underwater environments since it presented good results for 3D reconstructions in low turbidity waters. However it

requires a projector and a housing capable of protecting both the camera and the projector itself when submerging the rig in water. The presence of a projector in this kind of system makes the range measurement easier due to the fact that a well-defined pattern is used, however, the power required by the projector device is substantial which reduces the autonomy of the robotic vehicle.

Passive techniques only require two cameras. For measuring the distance of a certain object within the scene first it is required the so called *stereo correspondence*, in which features of frames captured through the left camera are found in the right camera and then disparity (difference between right and left pixels) calculations are made.

The biggest disadvantage of using this technique is related to the underwater imaging problems. Some algorithms of image preprocessing were already created to help decreasing the backscattering problem [7]. In this case, the error increases as a function of the distance, having a measuring error of 0.28m for an object at 2.46m of distance due to the presence of scattering. Two different cameras were used in [14] for underwater ranging measurement. This solution presented good results since the average error was about 10% for distances under 5m.

Finally, the research presented in [15] demonstrates a similar comparison since the influence of the baseline distance is studied. As can be noticed, the range error increases with the distance between the cameras and the object, in this case using a 0.50m baseline it has an average range estimation error of 0.09m at a 5m distance increasing up to 3.2 at 50m.

2.2 Stereo Correspondence Algorithms

This section presents the theoretical principles behind the stereo correspondence algorithms. Stereo matching algorithms determine the similarities between the two image planes (in the stereoscopic case).

Several stereo matching algorithms were already created and a taxonomy and evaluation of this type of algorithms was made by D. Scharstein and R. Szeliski[1] to help other researchers to compare and evaluate their own algorithms. It is possible to consult in Middlebury website several detailed algorithms of this kind and its individual performance such as computational time and average error (the hardware used during the evaluation is also described). Furthermore, this website also offers some datasets of images for multi-view systems testing with their respective groundtruth images.

The following categorization [16] was based on four sub-steps where a large set of existing algorithms can easily be constructed:

1. Matching cost computation,
2. Cost (support) aggregation,
3. Disparity computation/optimization,
4. Disparity refinement.

[1] http://vision.middlebury.edu/stereo/

However, the steps taken depend on the algorithm. The local (usually called window-based) algorithms mostly use the first three steps and NCC (normalized cross-correlation), for example, they aggregate the first two steps into a single stage using a matching cost that is based on a support region. On the other hand, the global algorithms usually do not do the aggregation sub-steps. Instead, they combine the data obtained from step 1 with a smoothness term in order to compute the disparity. Local and global matching algorithms will be described better in section 2.2.1.

Relatively to the division steps used in this taxonomy, the matching cost computation is related to which matching metric is used to compute the similarities between two images, for example: "the matching cost is the squared difference of intensity values at a given disparity" [16]. The most common pixel-based matching costs are squared intensity differences (SD) and absolute intensity differences (AD). Other traditional matching costs like normalized cross-correlation (NCC), sum of squared differences (SSD) or even sum of absolute differences (SAD) are often used.

The second sub-step measures the correlation between intensity values inside a matching window of a reference pixel assuming that all that pixels of that window have similar disparities. This aggregation is only used in local methods. An example can be: "aggregation is done by summing matching cost over square windows with constant disparity" [16].

Relatively to the third sub-step (disparity computation and optimization), for local methods this is very simple because the most complex and challenging operations are located at sub-steps 1 and 2. The disparity is computed by minimizing the cost value associated with the minimum cost value. For example: "disparities are computed by selecting the minimal (winning) aggregated value at each pixel" [16]. These methods perform a denominated "winner-take all" (WTA) optimization for each pixel having the limitation of the uniqueness of matches only being enforced for the reference image. Contrarily to local ones, the global algorithms do most of its work in this step. They usually skip the step 2 and join the acquired data information from images with a smoothness term but this will be explained in next section.

Finally, the disparity refinement is the post processing part of the algorithm that has the objective of making the disparity result smoother. Commonly, these stereo matching algorithms conduct sub-pixel computations based on iterative gradient descent or fitting a curve to the matching costs at discrete disparity levels. A cross-checking (comparing left-to-right and right-to-left disparity maps) for occluded areas detection is another post-processing examples presented in some of these type of algorithms and lastly a simple median filter can also be applied in the resulting disparity map to eliminate some noise or to fill some holes on the objects.

As it was already said, this taxonomy was created by D. Scharstein and R. Szeliski, for a more detailed information about their stereo matching algorithms, refer to [16].

2.2.1 Subtypes of Stereo Correspondence Algorithms

The stereo correspondence algorithms can be divided into two types: the local and the global methods.

In the local methods, a matching between two dimensional windows on both views is made using a WTA approach. They are called "local" because, for each pixel on one view, a matching pixel is found on another view independently of the other pixels, in contrast with global methods. The local matching algorithms are usually faster than the global ones because of that particularity, nevertheless they often compute a disparity map suffering from lack of smoothness. Basically, the local methods try to find the most similar window (of pixels) of one view in the other, scanning a line in that last one. That phenomenon may be seen in figure 2.1.

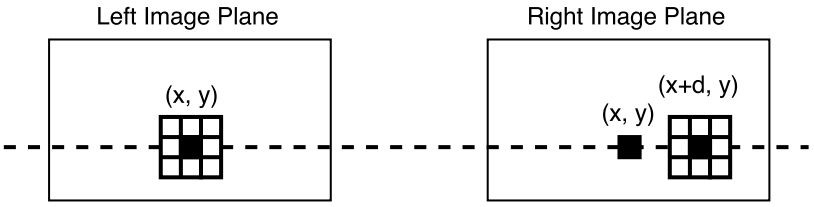

Figure 2.1: Basics of local (window) correspondence algorithm where the squares represent pixels. The filled square is the pixel that is being compared and it is being used a 3x3 scan window.

After the more similar window is found on the other view, by scanning between the minimum and the maximum disparity (these are usually parameters of the algorithms), represented by "d" in the image, the pixel (and consequently the window around it) with the lowest cost value is the chosen one (WTA).

The global matching algorithms rely in an energy-minimization framework. Therefore, the aim of this approach is to find the disparity "d" that minimizes the global energy function:

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \tag{2.1}$$

Where the E_{data} is the matching cost function like SAD, SSD or NCC, λ is a regularization parameter and the E_{smooth} represents the smoothness energy and it penalizes a disparity that is not smooth. One of the most used global energy method is the graph cuts. It presents good results, however it is computational inefficient since the process is slow for robotic applications.

2.2.2 Some Stereo Correspondence Algorithms examples

In this subsection some stereo correspondence algorithms and its functionality will be discussed. The website to evaluate the elapsed time needed to compute the similarities and to create a disparity map was used. In this way, it is possible to infer about the most suitable algorithm for real-time applications: the SGBM [17], ELAS [18] and SNCC [19] from this website having in consideration that they were the fastest techniques that do not require GPU implementations. These three algorithms are extremely useful for real-time applications, nevertheless more accurate

stereo matching algorithms are available in the literature, for instance, like the LCU[2] or TMAP[3]; however these methods are incredibly slow.

2.2.2.1 Semi-Global Block Matching (SGBM)

The SGBM is an algorithm that has the cost function similar to the global methods but in contrast with them is extremely fast and presents very good results.

This algorithm does not use a common matching cost function. Instead, it uses a function denominated as "mutual information" and more about this particular matching cost calculation can be viewed in Hirschmüller article [17]. The author chooses this specific matching cost calculation because of its capability for compensating radiometric differences of input images. The pixel-wise matching is also supported by a smoothness constraint that is usually expressed as a global cost function. The aggregation step is done by performing a fast approximation by path-wise optimizations in all directions. The disparity computation is made by searching for an equivalent pixel or a window of pixels in the same epipolar line of the other view just like a local correspondence algorithm. SGBM performs post-processing in the resulting disparity map in order to fill some gaps or to remove outliers from the resulted image.

Finally, an interesting capability from this algorithm is that it can compute the disparity map from two color input images, using the pixels intensity of the three channels.

2.2.2.2 Efficient Large Scale Stereo Matching (ELAS)

The ELAS algorithm has a different approach. What really distinguishes it from the others is the way how similarities are found. The authors want to develop an algorithm that is fast enough to work for real-time applications and at the same time that could compute good results. To do that, the algorithm starts to find the most reliable "support points" using a full disparity range. After that, these support points image coordinates are then used to create a 2D mesh using Delaunay triangulation. The support points are the matched pixels due to their texture and uniqueness (these parameters can be adjusted in the algorithm). The tricky part of this matching algorithm is the use of the initial correspondent points (the ones that are more reliable) to determine the disparity of unknown regions. Basically, this process is efficient by restricting the search to plausible regions.

2.2.2.3 Summed Normalized Cross-Correlation (SNCC)

The SNCC is a simple and fast algorithm that uses as a matching cost function known as Normalized Cross-Correlation. It is very used in image processing applications and consists essentially in subtracting the mean and dividing by the standard deviation of the template window that is being used as reference and of the subimage that is utilized for comparison. Just like the SGBM, a scan is done by the correspondent epipolar line of the other view, searching for the most correspondent window (subimage that is being compared) in relation to the template one. After doing all the

[2] Anonymous. CVPR 2015 submission 973
[3] Anonymous. ICCV 2015 submission 1667

similarity findings, a summation filter is applied directly on the result of the NCC filtering. With this the correlation values are averaged over the neighbourhood of each pixel at each disparity reducing the noise in the final map.

2.3 Camera Calibration

One extremely important step to acquire precise informations from images captured by cameras is the process of the camera calibration. This procedure is responsible for the calculation of the focal length, camera centres coordinates and even distortion coefficients (intrinsic parameters) and, in case of stereoscopic vision systems, for estimating the relative position of one camera in relation to other camera (extrinsic parameters). For a better comprehension of intrinsic and extrinsic parameters, figures 2.2 and 2.3 are respectively presented.

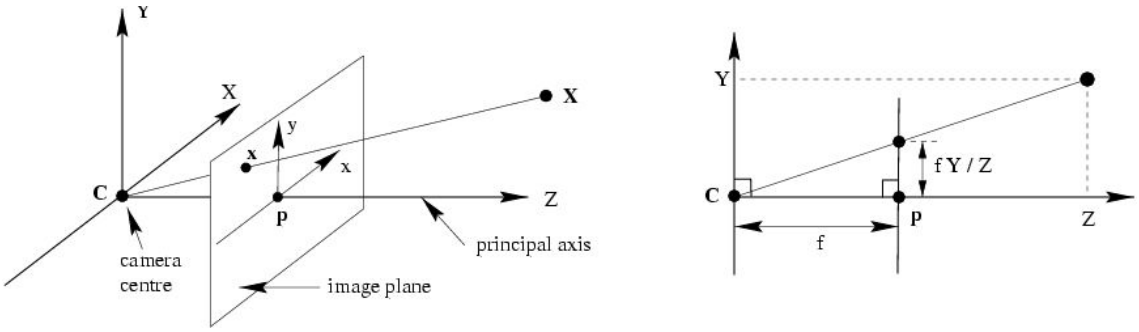

Figure 2.2: Representation of intrinsic parameters where f is focal length and p is the principal point [1].

When dealing with a camera calibration process, it is very important to understand and always have in mind the camera projection matrix (P) [1]:

$$P = K[R|t] \tag{2.2}$$

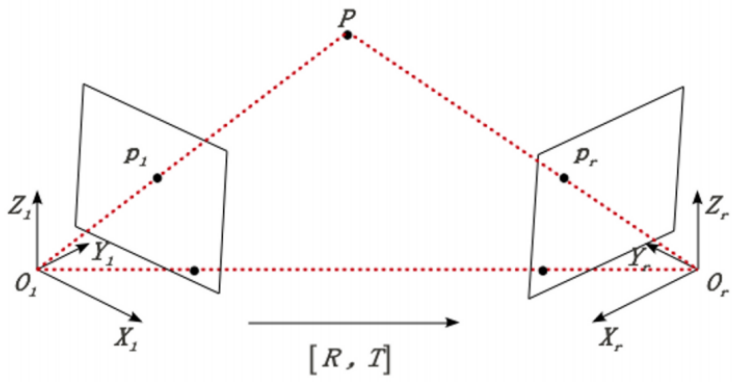

Figure 2.3: Representation of extrinsic parameters where R is the rotation matrix and T represents the translation vector.

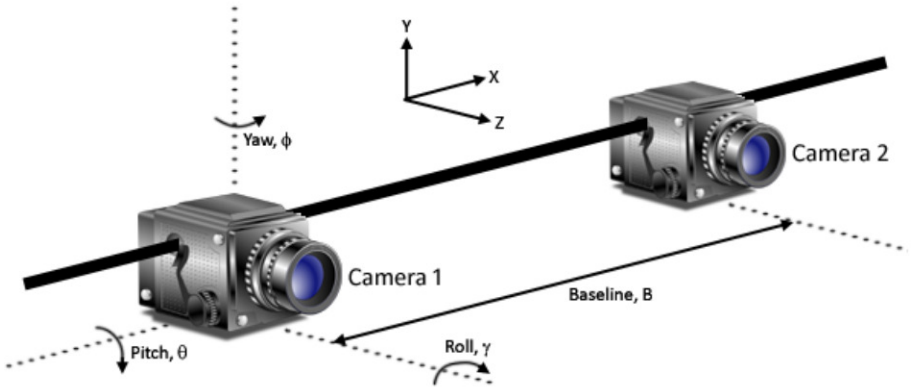

Figure 2.4: Stereo camera rig angle error sources θ, γ and ϕ. [2]

In this equation K is a 3x3 matrix with intrinsic parameters, R represents the rotation matrix and T the translation vector.

Intrinsic parameters and/or extrinsic parameters are obtained with a camera stereo rig and several algorithms were designed and could be divided in 4 different subsets: using a 3D reference object, a 2D reference plane, reference line or doing self-calibration.

For the 2D reference plane case, the approach of Z. Zhang [20] and J.Y.Bouguet [21] are the examples of two algorithms that retrieve intrinsic and extrinsic parameters with the presence of a calibration plane object with easy detectable feature points. In short, both algorithms consist in extracting the features in the calibration pattern and determine the intrinsic parameters. Posteriorly, the initial camera pose is estimated. Finally, Levenberg-Marquardt optimization algorithm is used to minimize the reprojection error.

Lastly, self-calibration is a method to retrieve the camera parameters and the main difference to the others is that self-calibration algorithms do not need a calibration object. It is only required to move the camera in a static scene and with the different views of the same scene it is possible to retrieve all parameters. It can be used with various static cameras with different views. Only three images are needed to obtain all the intrinsic and extrinsic values, assuming that the camera has fixed internal parameters. It is very advantageous in relation to the others if a calibration algorithm capable of readjusting camera parameters on the fly is needed, although it is not so precise like the 2D reference plane. An example of this method is present in this Santoro article [2], where a stereo rig system calibration is made based only in the correspondence principle. With this algorithm is possible to obtain yaw and roll angle errors below 0.01 degrees.

2.4 Effect of Alignment Errors

A very important fact that always has to be kept in mind in a converged stereoscopic system is the depth errors obtained from camera angles misalignment. Errors in the angles of the camera coordinate system will affect not only the triangulation process, but also the stereo correspondence algorithms, because most of them search for the similar point in the other view using the same

horizontal line. If a misalignment error exists, then the similar point might not be in that line and consequently the disparity map computation will not be so precise as desired. So basically, an error in yaw angle drastically affects the disparity and an error in range measurement appear. Furthermore, an error in pitch angle erratically misaligns the epipolar lines. Figure 2.4 depicts a stereo camera rig example with the three different rotation axis of the camera (yaw, pitch and roll). Table 2.1 exposes the error of each axis of the camera coordinate system and the influence of the baseline: where B is baseline, Z the true absolute depth, and X_2 and Y_2 represent the true 3D coordinate of the object in relation to camera 2. Furthermore, the authors of [22] concluded that the most critical errors were yaw, image sensor tilt, pitch, roll and baseline.

To achieve these equations it is assumed that camera 1 is perfectly calibrated as well as camera 2 except for the error source exposed in table 2.1. Knowing all of this information it is possible to conclude that in a stereo system with converged cameras the precision of the yaw angle is very important in order to obtain precise 3D coordinates, otherwise the acquisition of information from a scene is completely useless because of its error and uncertainty.

Table 2.1: Influence of camera angles misalignment in range measurement [2].

Error Source	**Range Measurement Error**
Yaw Error $\Delta\phi$	$\frac{\Delta Z}{\Delta\phi} \approx -\frac{Z^2}{B}(1+X_2)$
Sensor Tilt $\Delta\phi$	$\frac{\Delta Z}{\Delta\phi} \approx -\frac{X_2^2}{B}$
Pitch Error $\Delta\theta$	$\frac{\Delta Z}{\Delta\theta} \approx \frac{Z^2}{B}(X_2 Y_2)$
Roll $\Delta\gamma$	$\frac{\Delta Z}{\Delta\gamma} \approx \frac{Z^2}{B}(Y_2)$
Baseline Error ΔB	$\frac{\Delta Z}{\Delta B} \approx -\frac{Z}{B}$

Chapter 3

Acquisition of 3D Information

This chapter introduces how the 3D information is generally acquired using stereo vision. The calibration of cameras will be explored and the theory behind the transformation of the frames captured by the cameras to 3D coordinates will also be exposed.

3.1 Camera Calibration

An important step to obtain precise 3D coordinates from image planes is the calibration process. Camera calibration consists in obtaining intrinsic (focal length, camera centre and distortion coefficients) and extrinsic parameters (relative position and orientation of the two cameras). In this document, the calibration was conducted based on Z. Zhang [20] and J.Y.Bouguet [21] methods (using the functions available in OpenCV library). A 6x10 chessboard pattern was used, where each square had 0.072m. The pattern was pasted in a rigid acrylic structure so it can be submerged in water. To retrieve the intrinsic parameters, two essays were made for each camera: first a calibration with the chessboard panel outside and inside of the water was performed. The results of the calibration are presented in Table 3.1. Figure 3.1 depicts an example of a captured frame and the respective undistorted image. It is important to notice that these cameras have a 6mm lens so it is expected to have a similar focal length. These two experiments were done just to confirm that the focal length of the cameras are affected by the environment condition (above and below the water). This preliminary analysis is quite relevant for this thesis since the value of the focal length must be defined properly depending on the environment that the stereoscopic vision system is capturing.

Table 3.1: Intrinsic calibration results.

Focal Length (mm)			
Air		Water	
Left Camera	Right Camera	Left Camera	Right Camera
6.094	6.167	8.092	8.080

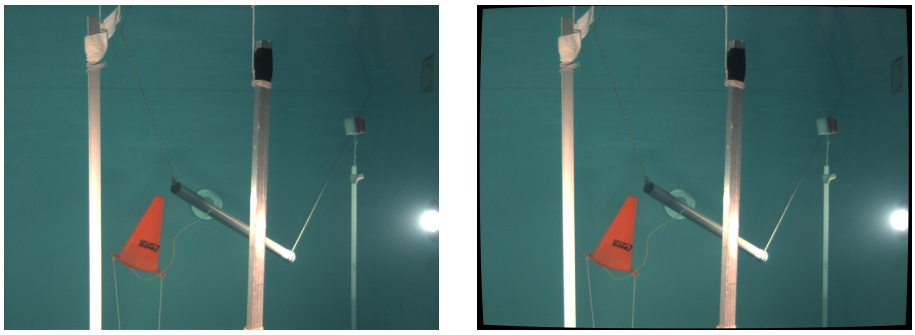

Figure 3.1: Comparison between original image and the respective undistorted.

As it can be observed, the focal length in underwater is approximately 1.33 times greater than in air which is precisely the value of the refractive index of water [12].

The extrinsic calibration was conducted in this thesis with the same acrylic structure with chessboard pattern. The acrylic pattern does not produce good results when estimating the extrinsic parameters since the pattern should be placed too far from the stereo system (due to its size 0.72 x 0.43 meters) in order to appear in the field of view of both cameras. Therefore, the cameras poses were determined with a smaller chessboard pattern. The smaller pattern is a 7x10 chessboard and each square has 0.034m. The extrinsic calibration in this thesis was always performed using that chessboard pattern and using the available OpenCV algorithms. A reliable extrinsic calibration is a crucial step for range measurement in a stereo vision sensor.

The rectification is one of the most important processes for extracting the 3D coordinates from the disparity obtained from image planes, because it aligns both images acquired vertically. Rectification is required to correct possible errors in alignment, which is the process of making epipolar lines parallel. This procedure is extremely important for converged cameras because it is a transformation operation that projects two images onto a common image plane. Figure 3.2 is an example of original frames (with parallel cameras) captured in both views (and represented on top of the figure where left image is the left view, and right image is the right view representation) with the result of the undistorted and rectified processes. Figure 3.3 illustrates the result from the undistortion and rectification procedures with cameras converged 11.02 degrees. The black margin of the undistorted and rectified images appears because of the transformation of the initial image planes (rectification) and due to the undistortion process. In figure 3.3 the black margin is larger than in figure 3.2 as a result of the rectification process (figure 3.4). Transforming both image planes into a common image plane, turning diagonal in horizontal lines, makes the resultant rectified pair of images with a black margin. In conclusion, for a stereoscopic system with convergent cameras, as long as the yaw angle becomes greater ("more convergent" cameras), larger margins will be originated.

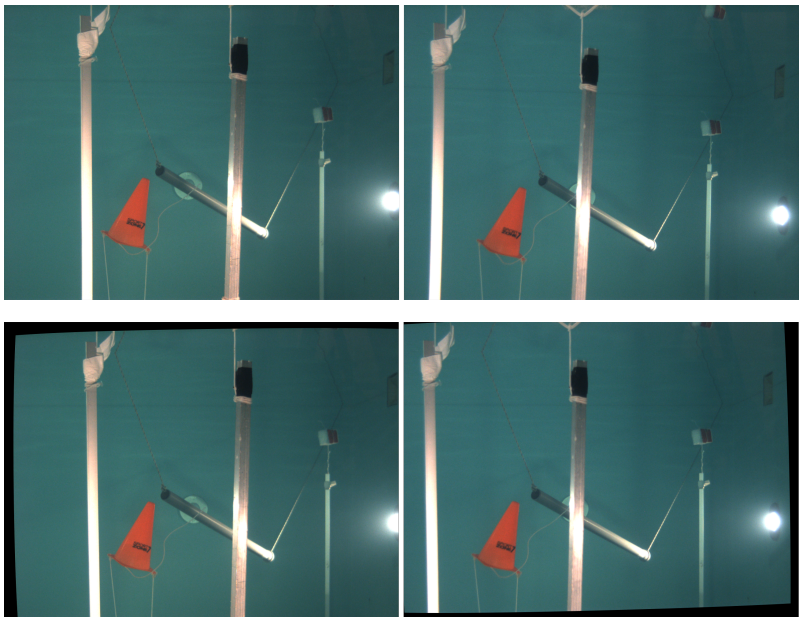

Figure 3.2: Comparison between original images of left and right views (on top) and the respective undistorted and rectified ones (on bottom), having left view as reference.

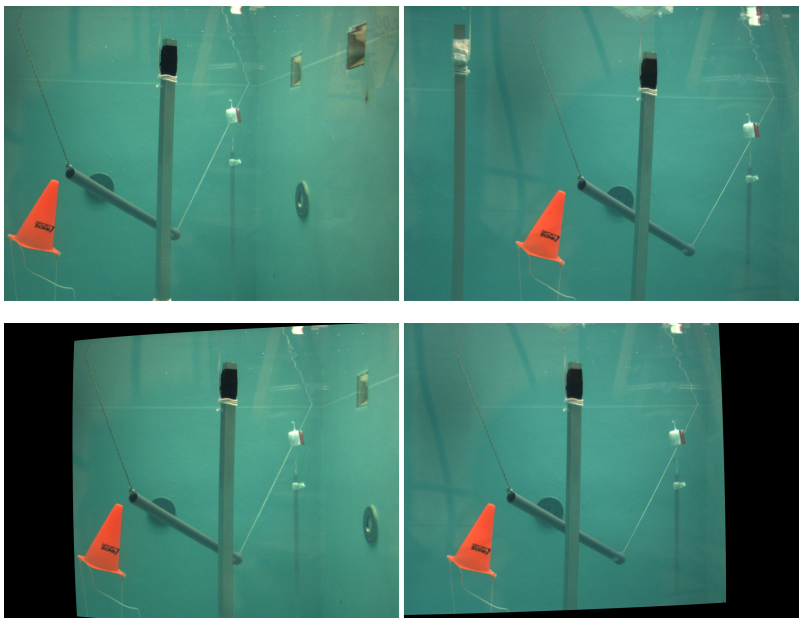

Figure 3.3: Comparison between original images of left and right views with cameras converged 11.02 degrees (top) and the respective undistorted and rectified ones (on bottom), having left view as reference.

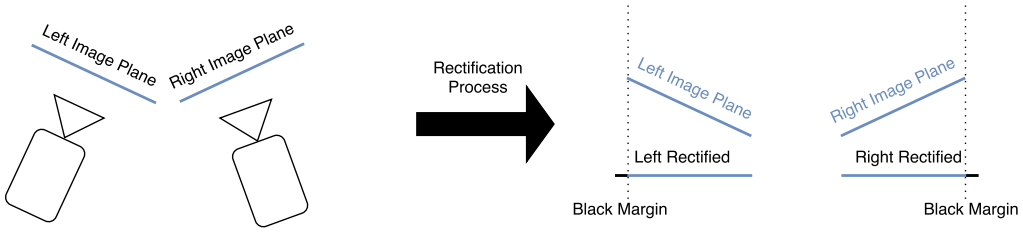

Figure 3.4: Rectified images black margin justification.

3.2 Disparity to 3D coordinates

It is possible to determine the distance to an object based on triangulation just like the human eyes. That theory will be briefly explored by contemplating the dynamic changes of the cameras' orientation.

3.2.1 Brief Explanation of Stereoscopic Range Measurement

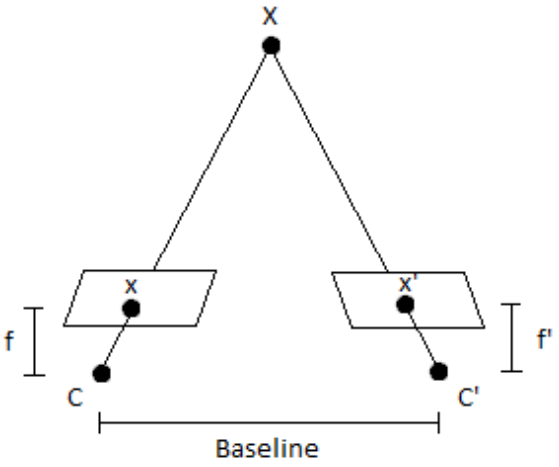

Figure 3.5: Stereoscopy theory example.

Figure 3.5 demonstrates an example of the stereoscopy principle, where X is the scene point, x is the left camera coordinate and x' is the corresponding coordinate in the right camera. In case of parallel cameras and after a full camera calibration (by knowing intrinsic and extrinsic parameters) it is possible to obtain coordinates of X using the Equation 3.1 where T_x is the horizontal translation

distance, C_x is the left camera center coordinate, C'_x is the right center coordinate and Z is the distance between cameras and the object.

$$Z = \frac{T_x f}{C_x - C'_x - d(x,y)} \tag{3.1}$$

Having both cameras calibrated, equation 3.1 can be used to calculate the distance to an object using the computed disparity (Figure 3.5).

Most of the stereo matching algorithms have the premise of getting the pair of images to be processed already with parallel epipolar lines. The rectification process could be better explained in figure 3.6.

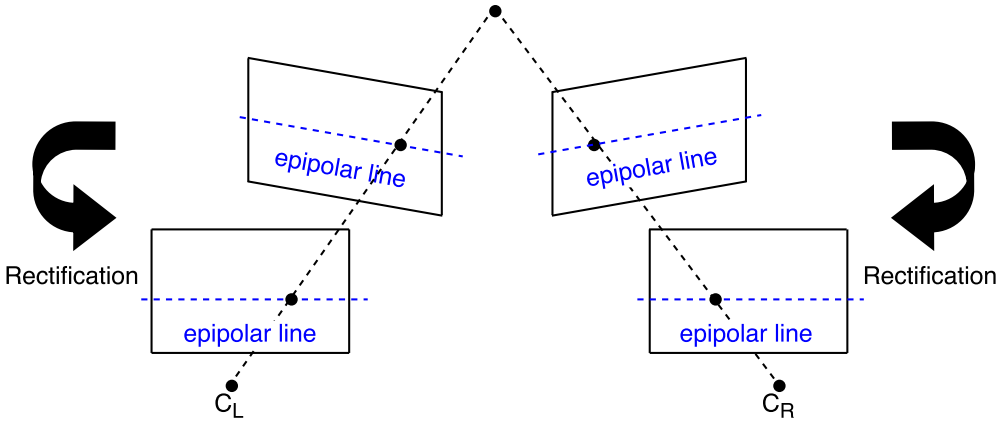

Figure 3.6: Image planes with parallel epipolar lines represented.

Most stereo matching algorithms works only with rectified images because this reduces the time needed to compute a similarity. Having the image planes with parallel epipolar lines, the search for the correspondent point in the other view is restricted to a 1D examination since it has to be in the same y-coordinate in both of views. Another important point is that after the rectification process and after transforming the image planes to parallel, the captured frames (originally with cameras rotated) can then be processed as if the cameras were initially parallel and then equation 3.1 can then be used to calculate the distance of each pixel or pixel window.

Chapter 4

Stereoscopic Range Measurement Error Analysis with Convergent Cameras

This chapter describes the main steps used in the experimental validations. The objective is to study the error of the range measurements that are obtained by converged cameras in different configurations. First the different components of the experimental set-up will be explained, as well as how the angle of the cameras is controlled. Posteriorly, the experiments' scene will be detailed and a comparison of stereo matching algorithms will be made. After that, all the steps utilized in this experience will be minutely explained and, at last, the system sensitivity will be introduced.

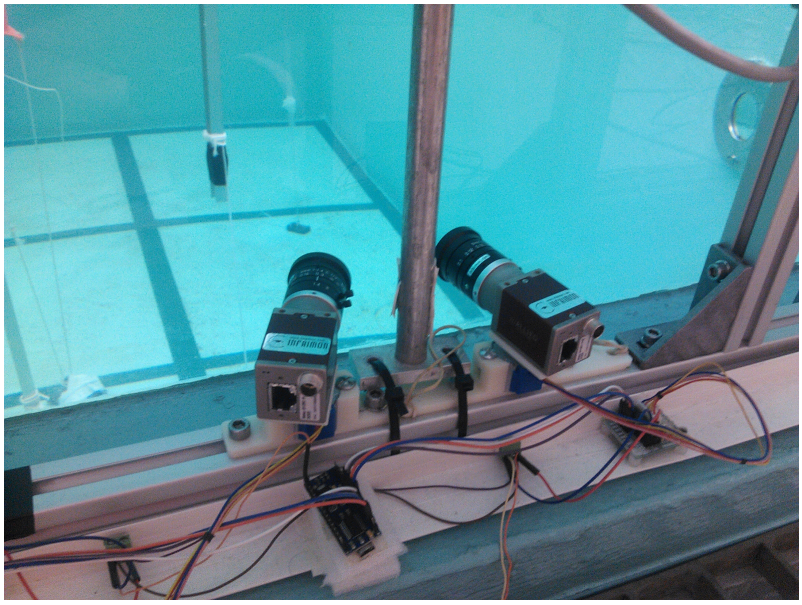

Figure 4.1: Setup used in experimental tests.

4.1 Stereoscopic System Based on Rotating Cameras

The stereoscopic system proposed in this thesis is presented in Figure 4.1 and it is possible to see that this rig is located on the outside of a water tank with an acrylic window. This system is composed by two Mako G125C cameras with 0.006m lens, two stepper motors, one microcontroller and the structure that supports the cameras mounted to the motors. With this supporting rig it is also possible to manually slide the cameras horizontally to easily change the baseline. Both cameras are supported by a plastic base that is attached to the stepper motors. The Arduino Nano is responsible for controlling the motors and consequently to converge or diverge the cameras according to the desired pose (for each camera). Therefore, this configuration can be modified according to the distance of the target. The idea of this experience is to study the impact of converged cameras, especially, if they cause additional errors in the overall accuracy of a common stereoscopic system.

4.2 Software Utilized for Stereoscopic Experiences

In all experiences present in this document a software was used. This one was created to preview both cameras in real time, take photos of the scene saving them to the PC, to which they were connected, and it was also possible controlling the steps of each stepper motor and consequently the rotation angle from both cameras.

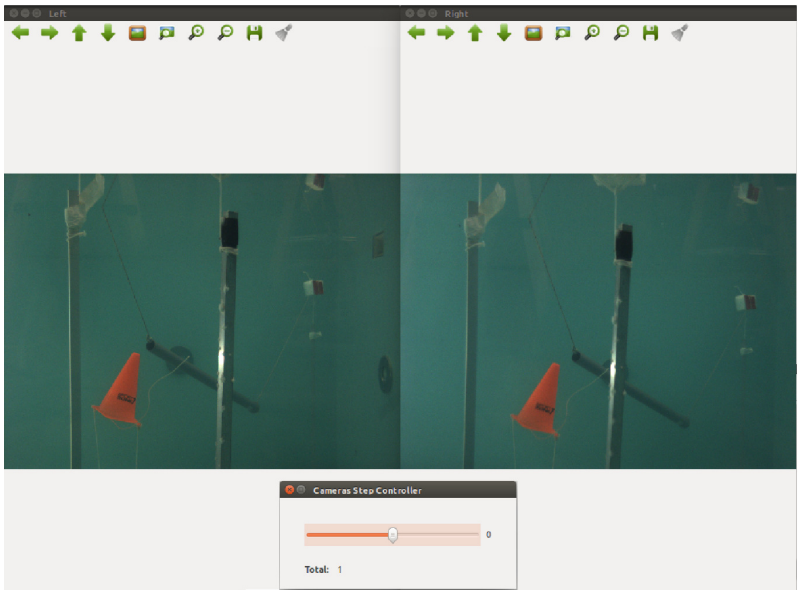

Figure 4.2: Software used in experiences with the two camera preview windows and the step controller slider.

Figure 4.2 depicts the software used in all experimental tests in this thesis and it is possible to see in this figure an example of the two views of the same scene. The slider was created to control the cameras rotations, where negative values represent steps to converge both cameras and

positive ones to diverge them. For example, if a "-1" was selected, one step should be applied to each camera, with left camera rotating to the right side (clockwise) and the other should turn to the left side (counter-clockwise), in order to converge as pretended.

The software was developed in Qt Creator[1] and in C++ language. For stereo algorithms the OpenCV library[2] has also been used. As it is known, in stereo vision the capture of images of both views at the same time is critical so this software is responsible to guarantee that premise. For steppers, the control was made modifying the slider position (that can be seen in figure 4.2) and with that value altered, a trama was then sent over serial port, in order to rotate the motors and consequently the cameras. In the future, it is intended to control the convergence/divergence of both cameras automatically based in features of the scene (distance to the target).

4.3 Experience Scene Description

The experimental tests were conducted in a water tank located at the OceanSys laboratory in FEUP. All the frames captured were taken using the same scene with objects in fixed positions. Various objects were submerged in the tank at different distances so conclusions about the accuracy of the distance measurement as a function of the orientation of the cameras and baselines can be made. Figure 4.3 depicts an example of the scene with the objects used for testing. The distances of two aluminum bars (at the same distance) and another one (the distant object), one cone and one PVC pipe were determined to be at 1.125m (the two bars that are at the same distance), 3.125m, 2.125m and 2.675m, respectively, from the cameras. The ground truth was established by measuring the distances with a tape-measure in a bridge structure over the water tank where objects were submerged. The distance between the cameras and the acrylic window located in the tank wall was also measured - 0.03m. All the experiments in this document were made with clear water and the laboratory lamps turned on, having the lights of the water tank deactivated.

[1] http://www.qt.io/ide/
[2] http://opencv.org/

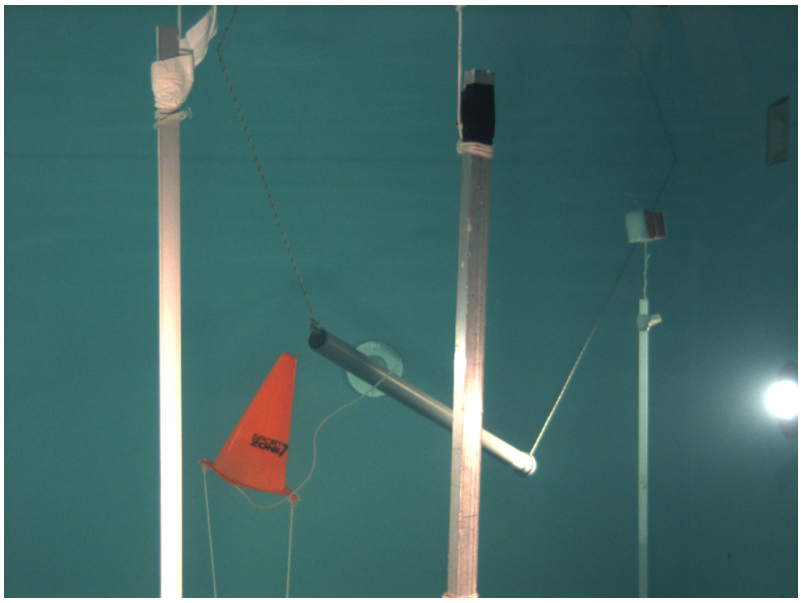

Figure 4.3: Image captured from the scene used in this experimental tests.

4.4 Stereoscopic Correspondence Algorithms Performance Comparison

The algorithms of stereoscopic correspondence are responsible for searching for images similarities. In this case it is very important to have an algorithm capable of finding as much similarities as possible having always in mind its processing time due to the real-time constraints of the robotic applications in real-time.

Three distinct algorithms were selected from the Middlesbury stereo website[3]: Block Matching, Semi-Global Block Matching (SGBM) [17] and Efficient Large-Scale Stereo Matching (ELAS) [18]. The criteria for picking these three was because they were the fastest ones excluding algorithms based in CUDA (Compute Unified Device Architecture) and they were also well classified in finding similarities. The BM was selected essentially to be used as a reference. The Middlebury stereo datasets were used for a first appreciation and the elapsed time in disparity map processing for each algorithm is depicted in table 4.1. All the elapsed time results in this section were obtained using an i5 2.4GHz processor. It is then possible to conclude based in table 4.1 that the BM is the fastest algorithm followed by ELAS and SGBM.

[3] http://vision.middlebury.edu/stereo/

Table 4.1: Elapsed time in disparity map processing for each algorithm using an i5 2.4GHz processor with images having the resolution of 450x375 pixels.

	Time (s)		
Image	BM	SGBM	ELAS
Tsukuba	0.007	0.096	0.052
Teddy	0.013	0.146	0.084
Cones	0.013	0.146	0.088

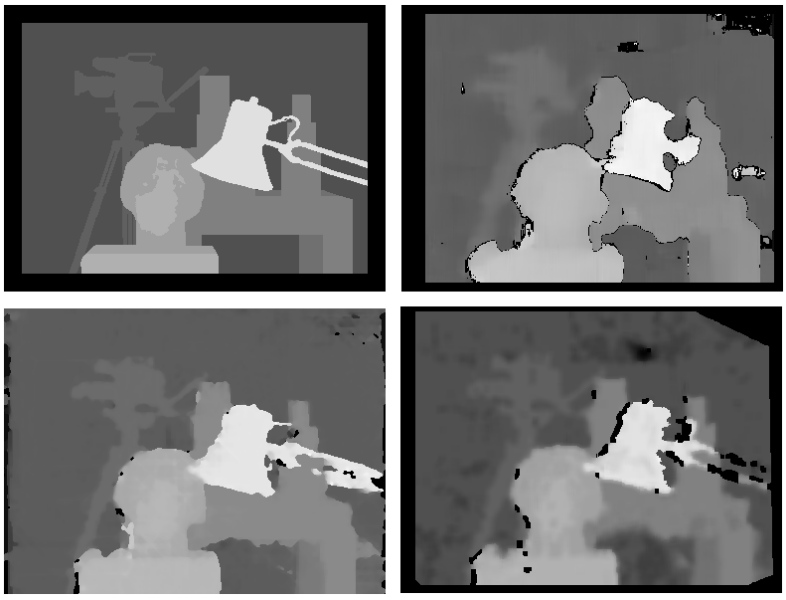

Figure 4.4: Disparity maps of the tsukuba image processed by various algorithms (left column: groundtruth and SGBM; right column: BM and ELAS).

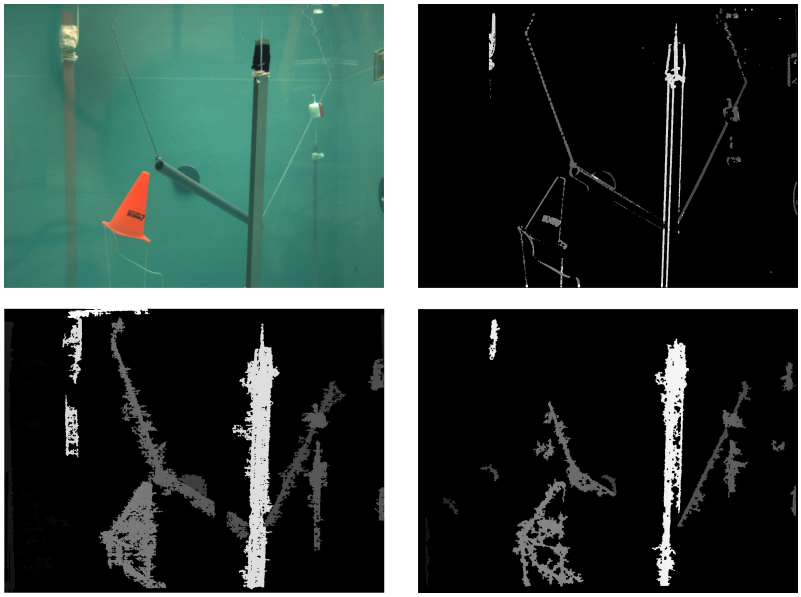

Figure 4.5: Disparity maps of the underwater images processed by various algorithms (left column: original image and SGBM; right column: BM and ELAS).

Figure 4.4 depicts the disparity map processed by the three previous algorithms configured with similar parameters and the groundtruth of the "tsukuba" image. After a visual inspection it is difficult to conclude which presents better results, so for supporting additional conclusions about these algorithms other image sequence was made using underwater frames (which is the environment where tests will be made) captured by the setup described in this thesis.

The results with underwater sequences are presented in figure 4.5 and the table 4.2 exposes the resultant elapsed time for the computation of the disparity maps from the underwater scene.

In this experiment, the SGBM clearly produces better results, in spite of being the slower one. Nevertheless, the computational time could be diminished by rising the minimum disparity value. With SGBM parameter bigger, the disparity map would not consider so longer distances, however it can be adjusted to a reasonable value using the follow equation:

$$Z = \frac{Bf}{Nx} \tag{4.1}$$

Z represents de distance between the cameras and the object, B is the baseline, f is the focal length, the maximum disparity level is represented by N and x is the pixel size of the image sensor.

In the figure 4.5 the cameras had a baseline of 0.11m and a focal length of approximately 0.008m in water (table 3.1). The pixel size of the image sensor is 3.75 μm[4] and for Z it can be attributed a reasonable value like 5 meters for example and the time is then reduced to 1.86s approximately. It takes more time than the other two algorithm, however it is more robust because it is not required too much alterations (in its parameters) in order to work in any environment. In contrast to the BM, the SGBM can easily distinguish easily the real objects of the scene with some

[4]Mako G125C datasheet - http://www.1stvision.com/cameras/AVT/dataman/Mako_DataSheet_G-125_prelim_en.pdf

Table 4.2: Elapsed time in disparity map processing for each algorithm from underwater scene using an i5 2.4GHz processor with an image with the resolution of 1292x964 pixels.

	Time (s)		
Image	BM	SGBM	ELAS
Underwater Scene	0.43	2.10	0.59

noise. The BM is a really fast algorithm for this purpose, but it can only find the edges of the objects (the most evident correspondence points) of the scene. BM do not presents then very good results for an object detection application, however it is more desirable for obstacle avoidance applications, because in table 4.2, it just needed 0.43 seconds to compute a disparity map from an image with the resolution of 1292x964 pixels.

With the ELAS, a good disparity map computation can really be efficiently made but the smoothness parameter must be adequately defined, which sometimes may be very difficult.

For a high precision application other algorithms can obviously be used, for instance, the global matching algorithms using cost function based in graph-cuts or belief propagation, for example, are usually great ones for high precision applications however, they take long time computing the disparity map.

4.5 Methods

The experimental results are explained and they are briefly represented in the diagram of the figure 4.6. Every time that the cameras were rotated an extrinsic calibration was required, and then the 3D coordinates were retrieved from the various objects within the scene. The complete process will be explained next.

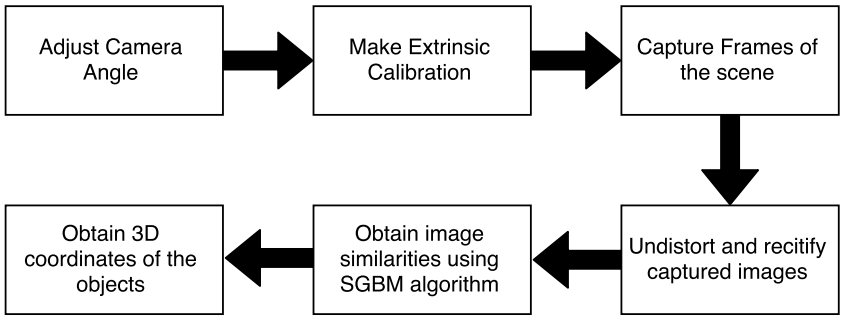

Figure 4.6: Diagram of the steps used to make objects distance measurement.

First, we converge the cameras to a predefined angle. Theoretically, a single step of the stepper

motors represent a rotation of 5.625 degrees. This thesis studied the error of the range measurement for 0, 1, 2 and 3 steps.

After rotating the cameras a new extrinsic calibration must be performed to obtain the respective rotation matrix and translation vector. The intrinsic parameters of all tests are presented obtained in section 3.1 and were fixed. The extrinsic ones were always discovered using the calibration methods of the section 3.1.

The next step consisted in taking multiple pictures of the scene. After that, these images were undistorted and rectified using the parameters obtained in the calibrations. The result of this process may be observed in figure 4.7.

Before getting 3D coordinates it is necessary to make the stereo correspondence. Having that in mind, the matching process was then processed by the SGBM algorithm and consequently disparity map can be seen in figure 4.8.

After applying this algorithm to the pair of images and calculating the distances between the cameras and the several objects in the scene (with triangulation principle) that distance is then converted to metric units and compared with the groundtruth values to evaluate the measurement error. The distance measurement using the cameras was done by selecting several surrounding pixels of the object to more stable solid results that could be comparable with the groundtruth values. This method process to retrieve the 3D coordinates was chosen to have more reliable results.

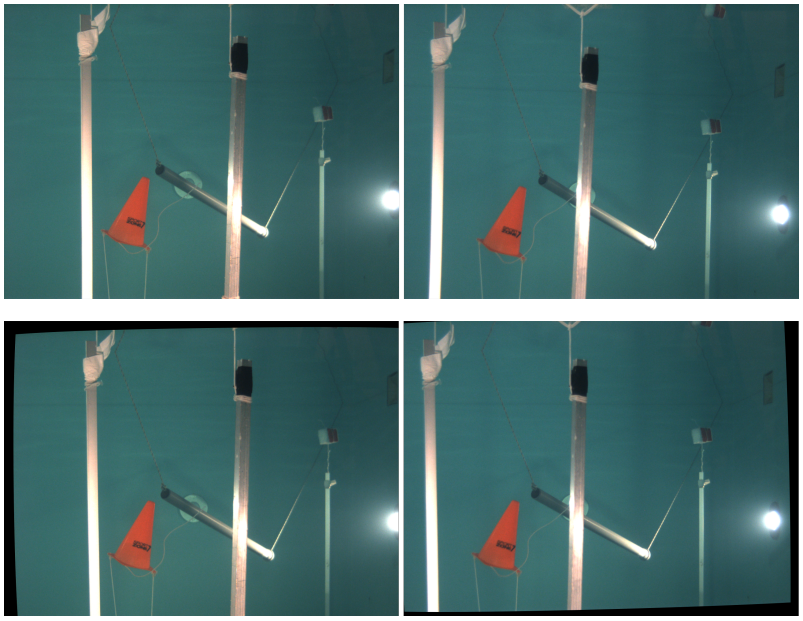

Figure 4.7: Comparison between original images (on top) of both views and the respective undistorted and rectified ones (on bottom).

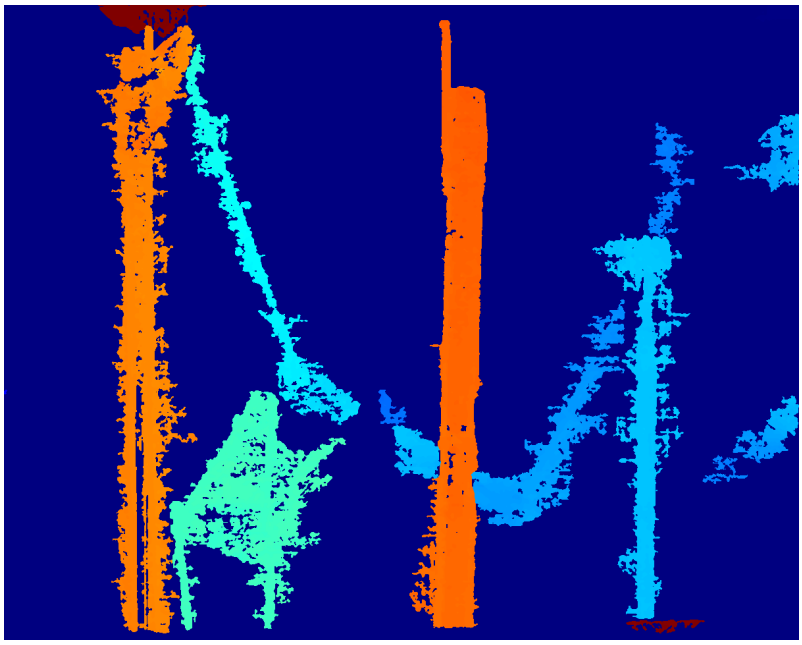

Figure 4.8: Resulted disparity map of the pair of the previous images.

4.6 Results and Discussion

In this section the results of the experiments are presented and discussed.

Table 4.3 and figure 4.9 demonstrate the distance error (in meters) for each object and considering different angles with a baseline of 0.29m. NV (stands for "not visible") is a parameter used for objects that are not within the visual range of both cameras at certain angles. A set of 20 different tests were conducted with 0, 1, 2 and 3 steps applied to each of the step motors, where 0 steps means no rotation and hence the cameras are parallel. The values in this table and figures are the average of angles obtained in the extrinsic calibration. The objective was to measure distances having cameras rotated with approximately 0, 10, 20 and 30 degrees but different angles were obtained because of the error introduced by the mechanical tolerances of structure (the weight of the cables, bases and cameras). Negative angles means divergent cameras. The distance error and standard deviation from various objects is presented in the following graphs. The bars represent the average errors and the smaller gray lines are the standard deviation obtained in the experimental results.

Table 4.3: Distance error with different cameras angles (baseline = 0.29m).

		Angle (degrees)			
Targets	Real Distance (m)	-1.1	6.3	12.3	16.8
Bar	1.125	0.089	0.084	0.083	0.091
Cone	2.125	0.079	0.113	0.149	0.190
Pipe	2.675	0.065	0.083	0.120	0.182
Bar	3.125	0.121	0.085	NV	NV

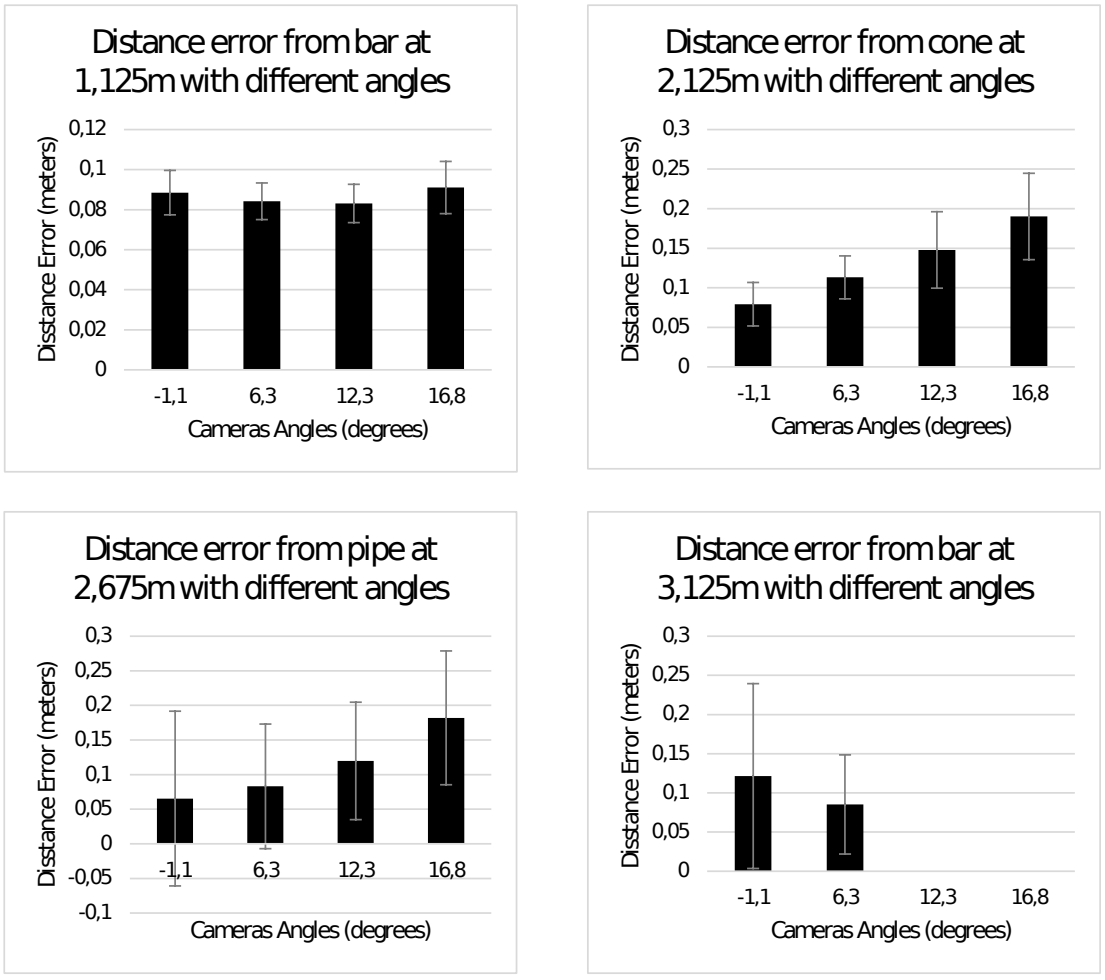

Figure 4.9: Distance error from the several objects to the cameras with different angles (baseline = 0.29m).

Table 4.4 and figure 4.10 present the results of the distance error with different camera angles for a baseline of 0.11m. For these tests a new bar was added at the same distance as the first one (1.125 meters) but centered for both cameras, because at this baseline the first bar was not visible when the cameras were parallel and it is desirable similar testing conditions (this bar is already present in figure 4.3).

Table 4.4: Distance error with different cameras angles (baseline = 0.11m).

		Angle (degrees)			
Targets	Real Distance (m)	0.0	4.0	11.0	19.7
Bar	1.125	0.051	0.038	0.099	0.090
Cone	2.125	0.066	0.283	0.287	0.026
Pipe	2.675	0.052	0.371	0.424	0.061
Bar	3.125	0.093	0.714	NV	NV

Finally it is possible to see in Figure 4.11 the comparison of the best convergent angle for each object (having smaller distance errors) with different baselines. For example, for the first bar with baseline equal to 0.29m the angle value used was when the cameras had a 12.5 degree angle and for the baseline equal to 0.11m, the parallel configuration was chosen because both were the best choices for that object with that specific baseline as it will be discussed next.

As we can see in figures 4.9-4.11 and tables 4.3 and 4.4 the distance error only grows in some cases with the convergence of the cameras. Sometimes when cameras are not parallel we get even more accurate results.

When baseline is equal to 0.29m (figure 4.9 and table 4.3) we got for a bar located at 1.125m with cameras converged with an angle of 12.3 degrees the more accurate distance measurement, having an error of about 7%. The 6.3 degrees rotation also presents a very similar error, so for an object at this distance a rotation between 6.3 and 12.3 degrees should be made to obtain the most precise distance measured with this sensor. For an object located at 2.125m, keeping the parallel setup is the best choice presenting a smaller error (4%). In pipe distance measurement the obtained results were not really good, even so an angle of 6.3 degrees is the one that presents less standard deviation and low error resulting in a 3% error. Finally with this baseline for a longer distance (3.125m) an error of 3% is obtained in this ranging measurement when cameras have a convergent angle of 6.3 degrees. So for small distances, an angle between 6.3 and 12.3 degrees should be used and for distant objects cameras should also converge as long as the object maintains visible.

When cameras have a baseline of 0.11m we expect obtaining a lesser angle than with the baseline equals to 0.29m to obtain best results with this baseline. With an analysis of figure 4.10 and table 4.4 it is possible to conclude that for a short distant object (1.125m) an angle of 4 degrees is the most suitable, despite of presenting a big standard deviation, resulting in a 3% error. For cone located at 2.125m the more precise results were obtained with cameras with an angle of 19.7 degrees. The measurements are presented with a really small error and standard deviation having only a 1% error. When objects become more distant using this baseline (pipe at 2.675m and bar at 3.125m) with the utilization of cameras with no convergence is what presents a smaller error,

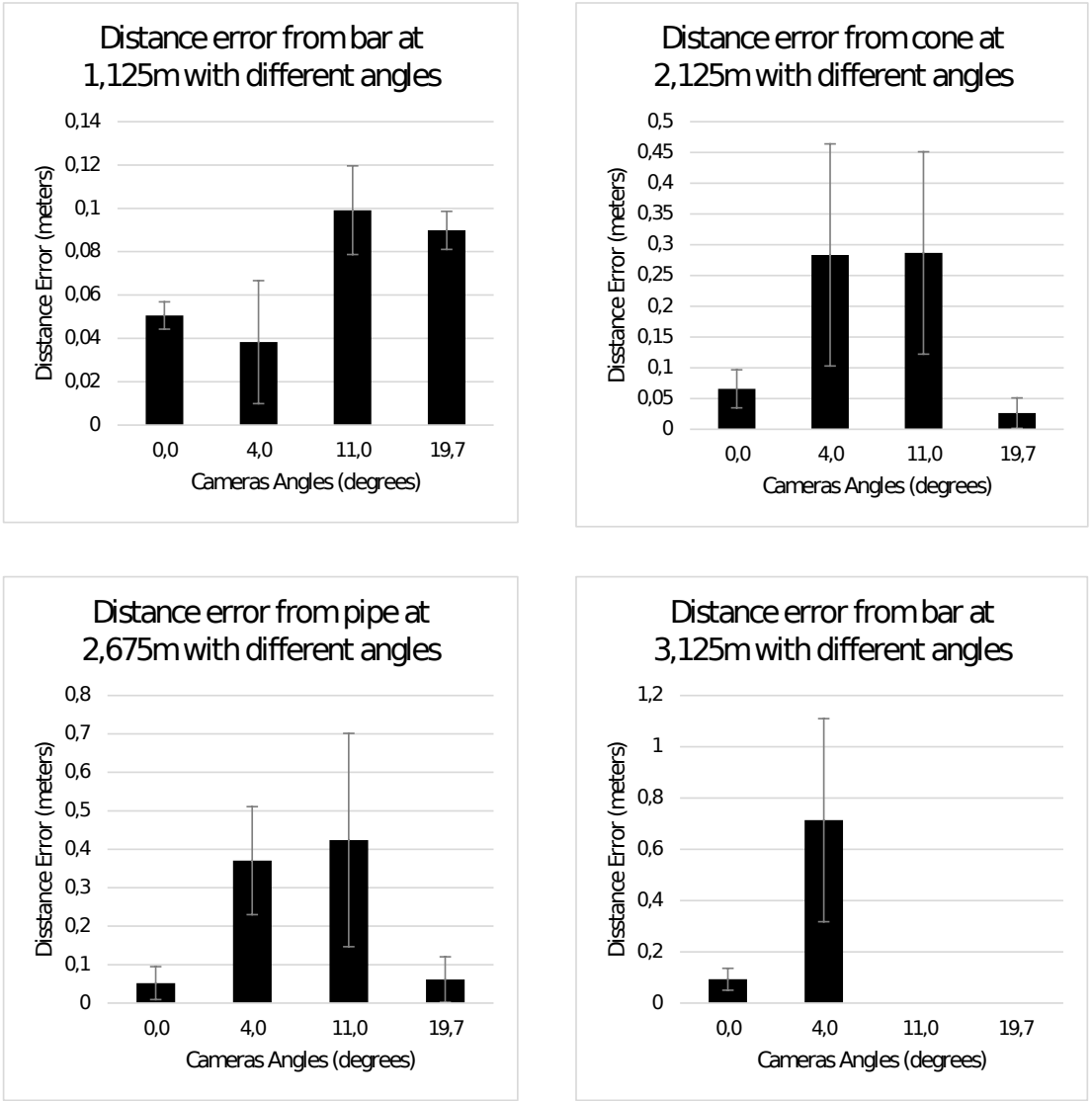

Figure 4.10: Distance error from the several objects to the cameras with different angles (baseline = 0.11m).

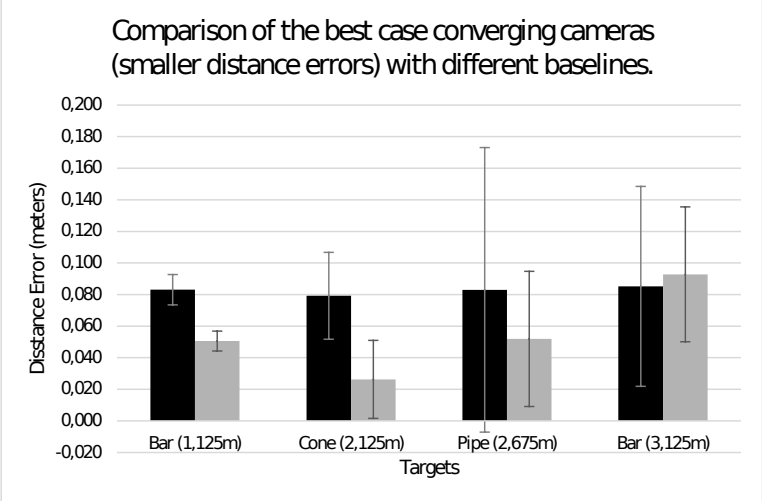

Figure 4.11: Comparison of the best converging cameras case (smaller distance errors) with different baselines (black bars represents results with baseline = 0.29m and gray bars with baseline = 0.11m).

2% and 3% respectively. In this specific case it is possible to assume that for distant objects, converging cameras do not bring any advantage. However, for objects located closer, an angle of 4 degrees should be used to compute more precise results. In this essays with baseline equals to 0.11m, more objects with small distances should have been considered and the cone result is inconclusive when compared with the bigger baseline results.

Finally observing figure 4.11 it is possible to see that the experimental results are not quite conclusive in some aspects, because of the big standard deviation. Still, based in our experiments in a water tank the usage of a smaller baseline is the more reasonable for getting more accurate results for distances between 1.125m and 3.125m. However, the average error of the 0.11m baseline is growing as long as the distance is getting bigger. At the same time the average error with baseline equals to 0.29m is keeping very stable and its standard deviation is becoming smaller which means that for short distances a small baseline is more precise but for long ranging measurements a bigger baseline will probably be a better choice. In table 4.5 it is presented a brief of the results obtained in experimental tests.

Table 4.5: The best angles for both baselines obtained in experimental results

	Best Angle	Distance Error (%)	Best Angle	Distance Error (%)
	Baseline = 0.11m		Baseline = 0.29m	
Bar (1.125m)	4.0°	3%	12.3°	7%
Cone(2.125m)	19.7°	1%	-1.1°	4%
Pipe (2.675m)	0.0°	2%	6.3°	3%
Bar (3.125m)	0.0°	3%	6.3°	3%

4.7 Stereo System Sensitivity - Experimental Misalignment Effect

In this section, the effect of misalignments in this stereo system or baseline in the depth calculation will be introduced. In state of the art this has already been explored, even so it will be presented in this section experimental results to evaluate the system sensitivity which is important to refer in any type of sensors.

It will be assumed that misalignments only occur in baseline or in yaw angle, because in this particular system these are the ones that are constantly changing. If the system is robust enough, the other possible errors should not exist and its values should even be always the same, so for more robust systems these misalignments errors (pitch and roll angle) could be clearly despised.

To evaluate the effect that error of the baseline, a program was created. That one estimates the depth of the several objects of the scene (with previously given coordinates) with cameras correctly calibrated (with the chessboard plane). After that calculation, the first value of the translation vector (T_x) is altered to have 0.01 meters more, and then that distance is calculated again. This can be done because this parameter does not change in the resultant disparity map, changing only the range measure value. Posteriorly, the error is calculated and a conclusion is made. In the case of this system, for 0.01m error of the baseline (with baseline originally equals to 0.11m) the depth error increases/decreases 0.089m. The general equation for this sensitivity analysis:

$$\frac{\Delta Z}{\Delta T_x} = \frac{0.00979Z}{B} \tag{4.2}$$

In yaw angle error, the procedure was similar to the previous one without having the automatic step, due the fact that this error makes an erratically rectification of the image planes and the same pixel may even not be located in the same object of the one calculated with cameras totally calibrated. So in this one for 1 degree error, the depth error for distances smaller than 4.20 meters (width of the water tank where tests were made) should be:

$$\frac{\Delta Z}{\Delta \phi} = \frac{0.016Z^2}{B} \tag{4.3}$$

To have a robust stereo vision system with convergence capability, it is really important to have a method to measure the angle and the baseline as much precise as possible, so good results can be computed, otherwise the parallel configuration should be possible be preferred, due the ease of measuring both parameters.

Chapter 5

Auto-Rotating Cameras for Precise Range Measurement

This chapter proposes a method to improve the range measurement precision in stereoscopic vision based in table 4.5, using the rotating capability of this stereoscopic system. Whenever the cameras alter their pose a new calibration must be made and this chapter will focus in two different approaches to estimate that poses differences. Briefly, the box "Make Extrinsic Calibration" of the figure 5.1 is going to be changed.

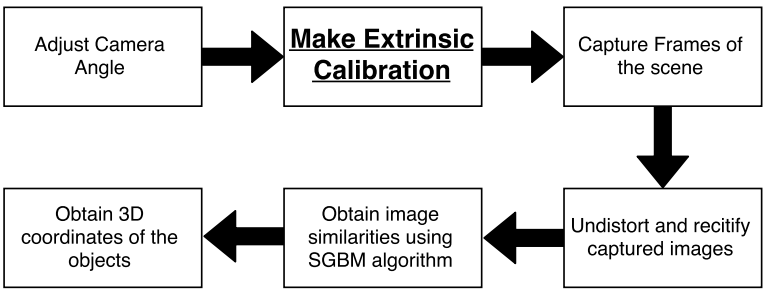

Figure 5.1: Diagram of the steps used to make objects distance measurement.

In the last experiment, the calibration was done using a chessboard plane. The objective of the proposed method is to extinguish the need of a well-defined object doing an automatic calibration. This calibration should be capable to acquire the extrinsic parameters with the cameras converged.

It should be noted that a new camera calibration is only required when cameras position change, because in that moment the new poses of the cameras must be estimated, the extrinsic parameters are obtained based in that estimation and then the map values are calculated. Having that with cameras maintaining its position, images can always be undistorted and rectified (remapped) with the same map values.

In the following section a briefly explanation of camera extrinsic parameters matrix is going to be presented because of its importance in a converged camera stereo rig calibration.

5.1 Explanation of Camera Extrinsic Parameters Matrix

As it has already been exposed in this document, camera projection matrix (equation 5.1) is composed by the intrinsic and extrinsic parameters matrices where the last is the one that is going to be explored in this section and is represented by the rotation matrix and translation vector $[R|t]$.

$$P = K[R|t] \tag{5.1}$$

The following equation (equation 5.2) is the extended version of the extrinsics matrix where the rotation matrix is calculated multiplying the 3 axes rotation matrices: $R = R_z R_y R_x$ and the translation vector is the distance between both cameras in the 3 different axes.

$$[R|t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \tag{5.2}$$

In order to make a calibration method capable of adapt automatically according to different camera poses, a manually change of this matrix is needed. To do that, is only necessary to have the distance of the cameras and the angles between them (in 3 axes) (figure 5.2).

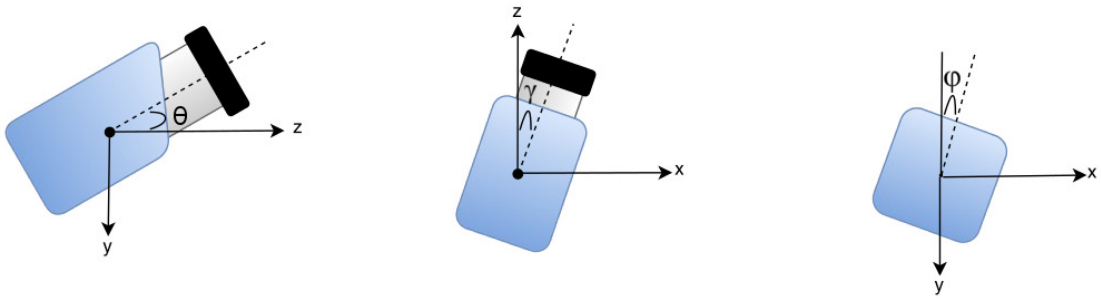

Figure 5.2: Camera possible rotations. From left to right: rotation over X axis (camera side view), rotation over Y axis (camera top view) and rotation over Z axis (camera back view).

To obtain that angles and distances it is proposed in the following section a simple method that could be used having the objective to do an automatic extrinsic calibration without the need of the chessboard previously used. After obtaining the angles from the two cameras, the relation between them is calculated like this:

$$\theta = \theta_L - \theta_R; \tag{5.3}$$

$$\gamma = \gamma_L - \gamma_R; \tag{5.4}$$

$$\phi = \phi_L - \phi_R; \tag{5.5}$$

At last, the rotation matrix is finally defined:

$$R = \begin{bmatrix} cos(\gamma)cos(\phi) & -sin(\phi)cos(\theta) + sin(\theta)sin(\gamma)cos(\phi) & sin(\theta)sin(\phi) + cos(\theta)sin(\gamma)cos(\phi) \\ cos(\gamma)sin(\phi) & cos(\theta)cos(\phi) + sin(\theta)sin(\gamma)sin(\phi) & -sin(\theta)cos(\phi) + cos(\theta)sin(\gamma)sin(\phi) \\ -sin(\gamma) & sin(\theta)cos(\gamma) & cos(\theta)cos(\gamma) \end{bmatrix} \tag{5.6}$$

5.2 Our Approach

As discussed in section 4.6, converging cameras gives best results in some cases and that specific ones will be explored in our approach. It is pretended that both cameras rotate automatically to a specific angle in order to extract more precise 3D coordinates and consequently objects distances.

For this purpose, it was made an exhaustive analysis of the extrinsic parameters obtained in each essay made in chapter 4. The aim of this was to find a way to make an automatic calibration of the extrinsic parameters of the stereo system implemented in this document everytime a rotation of the cameras occurs. In this thesis, the intrinsic parameters of both cameras are maintained, thus the intrinsic parameters matrix is always the same. To make an automatic calibration for the extrinsic parameters, a quadratic approach was made, based in the results of the tables of section 4.6. A calibration method using two IMUs (Inertial Measurement Unit) was also tried, but quickly was discarded due to its low precision. These two experiments will be discussed next.

5.2.1 Extrinsic Calibration using IMUs Method

The idea for this calibration method was to add two IMUs to the setup (each of one placed in the top of each camera like the one in figure 5.3) to acquire the yaw, pitch and roll of each camera.

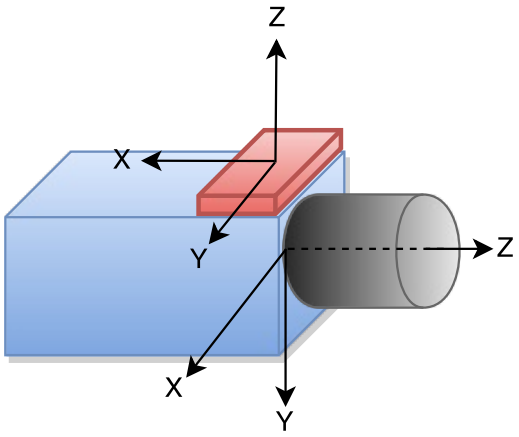

Figure 5.3: Camera with IMU on top. The red box represents the IMU sensor.

To test the precision of this type of sensors a calibration with the chessboard was done and then the angle values of each camera was saved. After that, a rotation was made and then the new angles were acquired by both sensors. The difference between the new angles and the ones obtained before the rotation summed with the initial extrinsic calibration using the chessboard should result in the camera pose. However, comparing the resultant angles from this approximation with angles obtained from another 2D calibration an error greater than 2 degrees was visible and then it was possible to conclude that this angle error was too big to make depth calculations. Even the stereo correspondence algorithms could not work and knowing all of that this method was discarded.

5.2.2 Extrinsic Calibration using Quadratic Function Approximation

Developing a camera calibration algorithm using a quadratic function approximation was the other approach used in this document. The experimental tests made in the last chapter were used as base in this approximation since it was done twenty essays and an estimation of the extrinsic parameters was done knowing all the rotation angles and translations that were made. The rotation matrix and translation vector were obtained using the openCV library functions and the chessboard previously described (for stereo system calibration) and afterwards the angles and distances between cameras were extracted from that extrinsic information acquired by the 2D camera calibration.

The figure 5.4 depicts the results obtained from the last experiment and the respective second order polynomial function representation with stereo system with a baseline of 0.29m. Figure 5.5 presents the results from the system having a baseline of 0.11m. Just like in the past chapter, black bars are the mean of the experimental results, the grey ones represent the standard deviation and the dotted line is the resultant quadratic approximation.

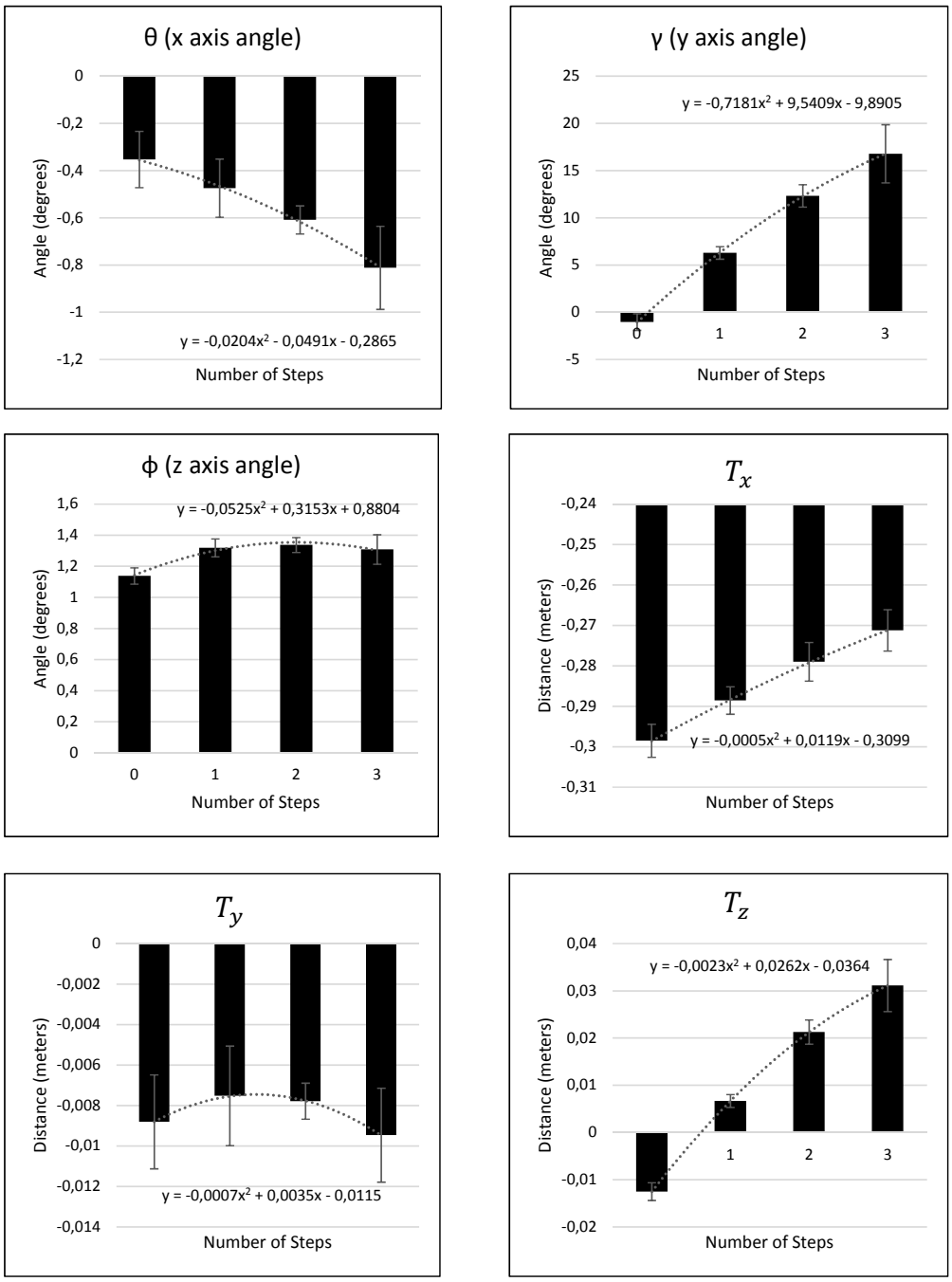

Figure 5.4: Angles and translation obtained in 0.29m baseline experimental tests.

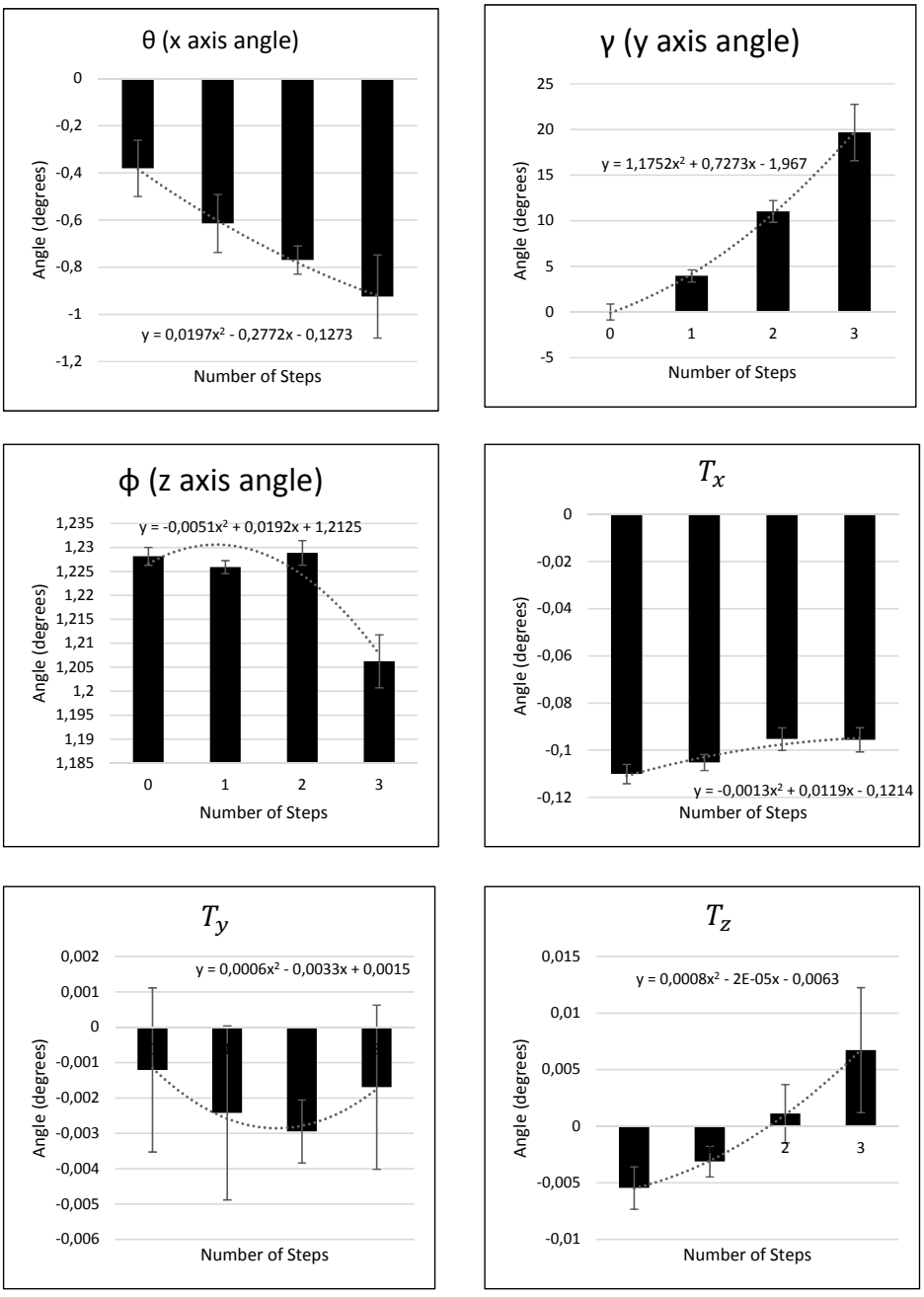

Figure 5.5: Angles and translation obtained in 0.11m baseline experimental tests.

Having the mean of all of these values is then possible to construct the extrinsic parameters

matrix as pretended. This method is as simple as fast which is really important in a real-time application and that occurs because it is only needed to solve some basic math operations. Nevertheless, this approach should present big errors in object distance estimation in non-perfect stereo systems. If it is used in a system that has high binary motors and low angle errors then this approach could be used and it should produce good results. In the case of the setup built in this thesis where the cameras sometimes do not verge as pretended, the range measurement results will probably be acquired with some errors, even so tests were made using the frames previously captured but this time using this quadratic approximation. The results will be next displayed and discussed.

5.2.2.1 Results and Discussion

In figure 5.6 and in figure 5.7 it can be seen the graphics that contains the range measurement errors obtained for each object using this method. In these tests, the frames obtained in the last chapter were used as test images to evaluate the error result in distance measurement and to be possible to compare the results with the obtained in the last chapter. The black bars are results of the error in object ranging measurement, the grey represent the results of the last chapter (with chessboard panel) and the grey lines are the standard deviation.

Table 5.1: Distance error using quadratic approximation with different cameras angles (baseline = 0.29m).

		Angle (degrees)			
Targets	Real Distance (m)	-1.1	6.3	12.3	16.8
Bar	1.125	NV	0.061	0.103	0.051
Cone	2.125	0.198	0.224	0.371	0.319
Pipe	2.675	0.301	0.302	0.562	0.493
Bar	3.125	0.460	0.465	NV	NV

Table 5.2: Distance error using quadratic approximation with different cameras angles (baseline = 0.11m) where black bars represents the results of this experiment, the gray bars the results of the experiment of the last chapter and the smaller gray lines the standard deviation.

		Angle (degrees)			
Targets	Real Distance (m)	0.0	4.0	11.0	19.7
Bar	1.125	0.086	0.052	0.073	0.064
Cone	2.125	0.234	0.171	0.202	0.183
Pipe	2.675	0.277	0.289	0.416	0.221
Bar	3.125	0.418	0.431	NV	NV

A directly comparison between the chessboard calibration and this quadratic approximation can not be made, because the last one was created based in the results of the first one, so it is not so abnormal that in some cases the this algorithm calculate 3D coordinates with less error. Nevertheless, it is possible to make some conclusions about all the results obtained with this approximation. First of all, the standard deviation of this method is normally bigger than the one

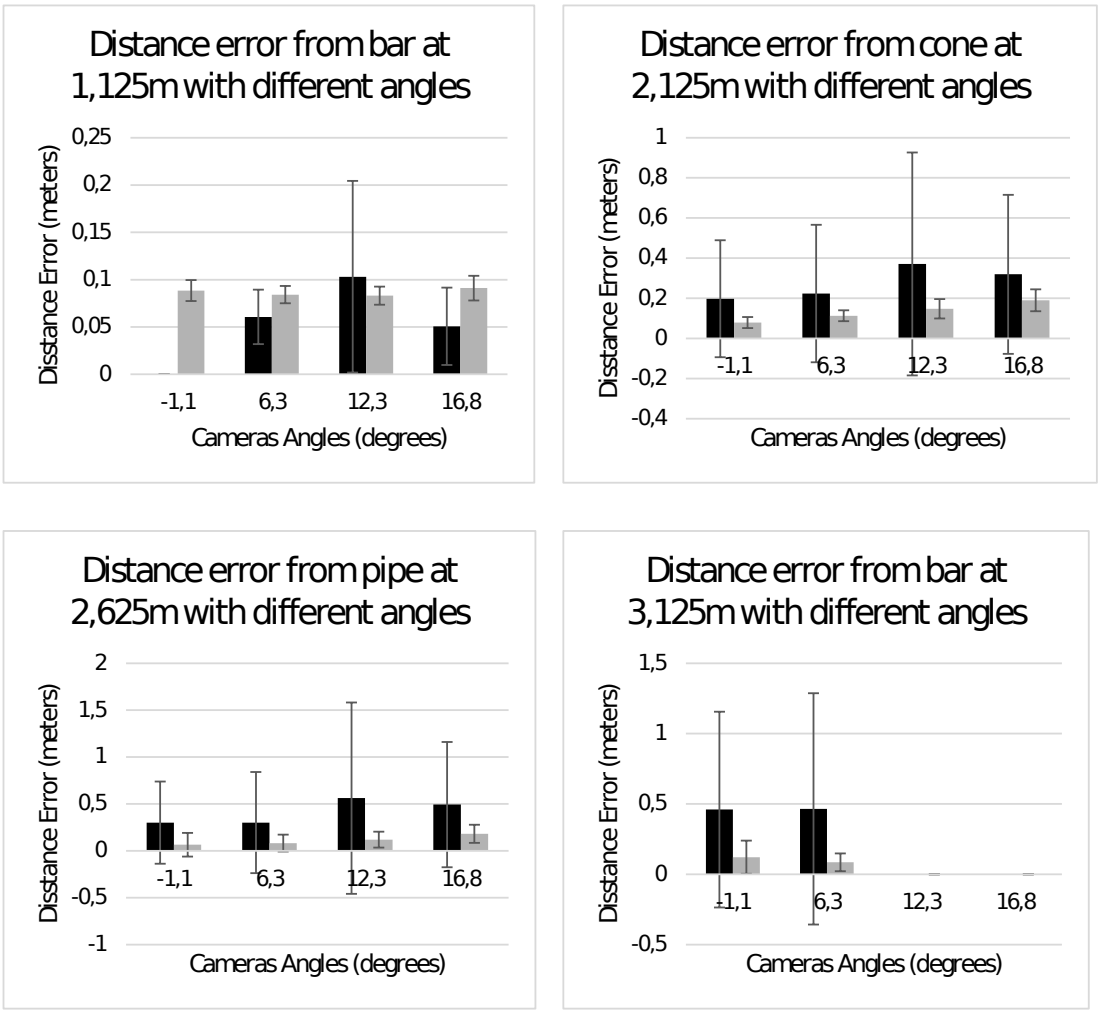

Figure 5.6: Results obtained in underwater range measurement using this quadratic approximation method (with baseline = 0.29m) where black bars represents the results of this experiment, the gray bars the results of the experiment of the last chapter and the smaller gray lines the standard deviation.

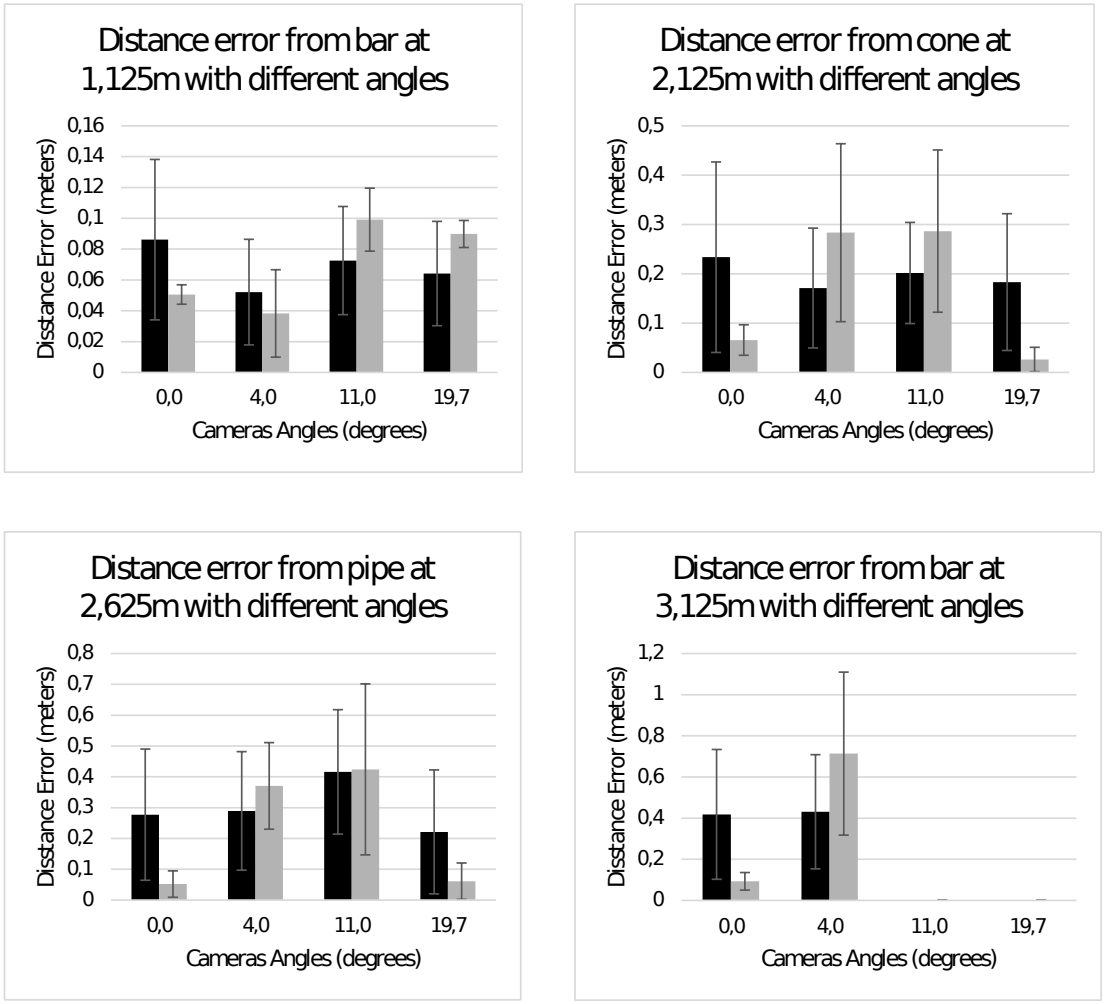

Figure 5.7: Results obtained in underwater range measurement using this quadratic approximation method. (with baseline = 0.11m)

of the chessboard plane calibration (except some particular cases like the distance from cone with baseline = 0.11m, with camera angle of 4.0 and 11.0 degrees) and that can be easily explained because of the fact that with the quadratic approximation method it is assumed that the steppers are perfect, that the cameras always rotate the same angle (or a very similar). When they rotate much more than the mean angle, an error from the yaw angle will be created and another one from the baseline (because they are completely dependent).

One more thing to have in consideration is that, based in these results, this method hardly can be applied in a real application. It presents errors bigger than 10% and for having a so great error, a static stereo vision system with an initial calibration with a chessboard, for example, is preferable. Nevertheless, with a much more robust setup, this approximation should possibly be applied for some applications, because if utilized in a stereo system that really have not some type of looseness, in other words, a system with a low angle error that rotates the most of times similar angles, the depth errors using this approximation should be much smaller bringing up results much more precise than the presented with our looseness setup.

Chapter 6

Conclusions and Future Work

In this thesis, it was explored the capabilities of one non-conventional stereoscopic system based in converged cameras. A setup with this principle was built and some experimentations were made. To evaluate the influence of the rotation of the cameras in range measurement, 20 essays were made with the system with two different baselines (0.11m and 0.29m), having always the same scene and same objects, in order to obtain solid results.

This experiment results show that for a baseline of 0.29m, the distance measurement error generally decreases when a 6.3 degrees rotation is made and for objects at longer distances (around 3 meters) the error can even be reduced in 1% decreasing from 0.121m to 0.085m of error. It is also expected that for bigger distances the use of a 6.3 degrees rotation could be even more beneficial to reduce the range measurement error.

For small baselines and closer objects, converging cameras to 4 degrees represents having a measured distance with a less error. So for objects located at really close positions in relation to the stereoscopic vision sensor a 4 degrees rotation should be the better choice. For objects with greater distances the parallel configuration is the more advantageous because presents results with a constant error of about 0.05m. After all of this, an automatic calibration method for extrinsic parameters estimation was also proposed, based in a quadratic approximation. The results obtained from this experiment were not so good as pretended. In this case, it can be concluded that for the setup used in this document, this approximation brings a great error in range measurement (bringing an average error of about 12%).

Finally, it can be said, based in these experiments that for application that require precision range measurement it is advisable to use for a baseline of 0.29m a fixed camera angle of 6.3 degrees and for a baseline of 0.11m the parallel configuration is the better choice, unless it is pretended to measure small distances (until 1.125m) and in that case a 4.0 degrees rotation is suggested. For the quadratic approximation for estimation of extrinsic parameters, it can be concluded, that the resultant distance measured with this type of approximation with a looseness setup is definitely a bad choice.

The proposed objectives of this thesis were all accomplished, but the obtained results were not so good as pretended. Some future work should be done, in order to corroborate some analysis

and some results, mainly in the extrinsic parameters estimation.

6.1 Future Work

Like it has already been said, some future work could be made, with the view of having more assertive and better results. For the camera convergence effect in the depth estimation, additional experimentations with vaster scenes should also be made to evaluate the effect of convergent cameras in bigger distances (for great baselines) and for small baselines a scene with a maximum distance of 1 meter (with objects at several distances) should also be considered.

A new and more robust setup should also be built with better stepper motors and even an encoder can be added to reduce the angles errors. The calibration method based in quadratic approximation could then be test with this setup and better results and conclusions should then be taken from this experimentation.

Finally, a self-calibration algorithm capable to estimate the camera parameters in underwater environments should be tested in a dynamic system like the one of this document.

References

[1] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[2] Michael Santoro, Ghassan AlRegib, and Yucel Altunbasak. Misalignment correction for depth estimation using stereoscopic 3-d cameras. In *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, pages 19–24. IEEE, 2012.

[3] Ji-Hong Li, Mun-Jik Lee, Won-Seok Lee, Jung-Tae Kim, Hyung-Joo Kang, and Jin-Ho Suh. Real time obstacle detection in a water tank environment and its experimental study. In *Autonomous Underwater Vehicles (AUV), 2014 IEEE/OES*, pages 1–5. IEEE, 2014.

[4] Stephan Shatara, Xiaobo Tan, Ernest Mbemmo, Nathan Gingery, and Stephan Henneberger. Experimental investigation on underwater acoustic ranging for small robotic fish. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pages 712–717. IEEE, 2008.

[5] Jules S Jaffe. Computer modeling and the design of optimal underwater imaging systems. *Oceanic Engineering, IEEE Journal of*, 15(2):101–111, 1990.

[6] Luke K Rumbaugh, Erik M Bollt, William D Jemison, and Yifei Li. A 532 nm chaotic lidar transmitter for high resolution underwater ranging and imaging. In *Oceans-San Diego, 2013*, pages 1–6. IEEE, 2013.

[7] Bing Zheng, Haiyong Zheng, Lifeng Zhao, Yongjian Gu, Lijun Sun, and Yuting Sun. Underwater 3d target positioning by inhomogeneous illumination based on binocular stereo vision. In *OCEANS, 2012-Yeosu*, pages 1–4. IEEE, 2012.

[8] Dan McLeod, John Jacobson, Mark Hardy, and Carl Embry. Autonomous inspection using an underwater 3d lidar. In *Oceans-San Diego, 2013*, pages 1–8. IEEE, 2013.

[9] Davide Moroni, Maria Antonietta Pascali, Marco Reggiannini, and Ovidio Salvetti. Underwater scene understanding by optical and acoustic data integration. In *Proceedings of Meetings on Acoustics*, volume 17, page 070085. Acoustical Society of America, 2014.

[10] Charles Cain and Alexander Leonessa. Laser based rangefinder for underwater applications. In *American Control Conference (ACC), 2012*, pages 6190–6195. IEEE, 2012.

[11] K Muljowidodo, Mochammand A Rasyid, N SaptoAdi, and Agus Budiyono. Vision based distance measurement system using single laser pointer design for underwater vehicle. *Indian journal of marine science*, 38(3):324–331, 2009.

[12] F Bruno, G Bianco, M Muzzupappa, S Barone, and AV Razionale. Experimentation of structured light and stereo vision for underwater 3d reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):508–518, 2011.

[13] Miquel Massot-Campos and Gabriel Oliver-Codina. Underwater laser-based structured light system for one-shot 3d reconstruction. In *SENSORS, 2014 IEEE*, pages 1138–1141. IEEE, 2014.

[14] Tom B Letessier, Jean-Baptiste Juhel, Laurent Vigliola, and Jessica J Meeuwig. Low-cost small action cameras in stereo generates accurate underwater measurements of fish. *Journal of Experimental Marine Biology and Ecology*, 466:120–126, 2015.

[15] Nicholas BW Macfarlane, Jonathan C Howland, Frants H Jensen, and Peter L Tyack. A 3d stereo camera system for precisely positioning animals in space and time. *Behavioral Ecology and Sociobiology*, 69(4):685–693, 2015.

[16] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42, 2002.

[17] Heiko Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, 2008. doi:10.1109/TPAMI.2007.1166.

[18] Andreas Geiger, Martin Roser, and Raquel Urtasun. Efficient large-scale stereo matching. In *Computer Vision–ACCV 2010*, pages 25–38. Springer, 2011.

[19] Nils Einecke and Julian Eggert. A two-stage correlation method for stereoscopic depth estimation. In *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pages 227–234. IEEE, 2010.

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

[21] J.Y.Bouguet. Matlab calibration tool. [Online; accessed 21-February-2015]. URL: http://www.vision.caltech.edu/bouguetj/calib_doc/.

[22] Wenyi Zhao and Nagaraj Nandhakumar. Effects of camera alignment errors on stereoscopic depth estimates. *Pattern Recognition*, 29(12):2115–2126, 1996.