Accuracy Improvement of Target Ship Position Estimation by a Shore Set-up Stereo Vision System

GAZİ KOÇAK

Kobe University
Maritime Faculty
Marine Engineering Department

修　士　論　文　要　旨

修士論文提出者	学籍番号	072W601W	氏名	KOCAK GAZI	講座	マリンエンジニアリング講座
修士論文題目	Accuracy improvement of target ship position estimation by a shore set-up stereo vision system					

Recently, safety ship navigation has become more important by the increment in the ship sizes and number of ships worldwide. Marine accidents and collisions are forming a great risk for environment and maritime sector, and this risk is gradually increasing by the increment in the number and size of the ships. The maritime authorities of countries want to have better control opportunity of ship traffic especially in strict areas as territorial waters and straights due to the risk of accidents and environmental pollution. For this reason the VTS's (Vessel Traffic Services) are developed and many people employed at VTS Centers to be able to control the traffic flow. Even more, worldwide maritime authority IMO (International Maritime Organization) constitutes new laws to achieve safer ship navigation and safer sea areas. However there is not enough preventive action taken and the frequency of maritime accidents is increasing.

For safer navigation and prevention of marine accidents, especially collision, ship detection has a great importance. There are already some detection devices as RADAR/ARPA (Automatic Radar Plotting Aid) and AIS (Automatic Identification System) on board ships. AIS is mandatory only for the vessels which are greater than 500GT. These equipments are capable of detecting ships which are greater than 500 GT in the range of 20 nautical miles. However, the ships smaller than 500 GT and which are not carrying AIS device induce a high risk of collision. Even more in case of small ships, the radar cannot distinguish the ship and sea clutter and may cause false detection. Therefore, these devices may not be enough for collision-free navigation especially in domestic waters where small vessels like fishing boats exist.

The accident inquiry reports show that the collisions are mostly caused by cargo ships and fishing vessels which are in the category we mentioned above. Therefore, this study has importance to detect ships in the range of about 3 nautical miles (about 5 km) -which is important to avoid from collision- which are small in size and not carrying AIS. The best way to do this is considered to be an artificial vision system that actually is carried out by human now. The inquiries show that most of the collisions take place in the range of 3 nautical miles from shore. One of the reasons for this is the fatigue of navigation officers. After a hectic work in port (loading and discharging), the navigation officer is already tired and has to carry out a navigation watch. Besides, near the port area the marine traffic is expected to be dense. In this case, the collision risk is increasing. By the application of such an assistant look-out system, the workload of seafarers will be reduced and safer navigation environment will be constituted.

（注）黒インクで1,000字以内に書くこと。

神戸大学大学院海事科学研究科

修士論文要旨	学籍番号	072W601W	氏　名	KOCAK GAZI

In maritime field, detection of ships and autonomous navigation is getting more important day by day due to reduction in the number of mariners and increasing number of ships. Application of computer vision and digital image processing techniques can be effective for this purpose. This subject has not been studied sufficiently especially in maritime sector, so far. Some limited number of studies already has been carried out, but an accurate and precise system has not been developed yet.

Stereo vision system is one of the most exploited systems to recover 3D information from 2D images. However, it has been applied almost for the land based robotic applications and not for maritime field. Therefore, stereo vision using digital images is one of the candidate methods for ship detection, 3-D location measurement and autonomous navigation of ships.

In the previous study, feasibility of a stereo vision system was studied to detect ships and acquire ship position data. In that study, camera calibration was emphasized. However, especially in the long range (by 5 km) data some errors were encountered. To improve the accuracy of 3D data, these errors should be removed.

One of the major causes of these errors is based on quantization errors. The quantization errors arise from use of digital images that has finite number of pixels. They affect the accuracy of the system especially in long range measurement due to increasing depth resolution with respect to range. That's why image plane quantization error elimination is one of the purposes of this study. Our solution to overcome the quantization error is calculation of corresponding points between stereo pair images with sub-pixel order accuracy. There have been several methods proposed for sub-pixel image matching. Intensity-based sub-pixel registration is preferred in this study.

Another important range error parameter is inaccuracy and vibration in disparity values of corresponding points. Refinement of the disparity values by low pass filtering (LPF) yields smoother values which results in reduction of 3-D measurement error to 1% of object's distance.

Applying the mentioned methods we observed a great improve in the accuracy of our stereo camera system. 77% of the range error is eliminated. In a determined range the Z dimension error of ±170 m. is reduced to ±35 m. The maximum measurable range with 150 m. range error is increased from 2600 m. to 5400 m. In the case of 1.5 km. of measuring distance the range error is reduced to ±12 m. which is acceptable for this range. Therefore; it is sufficiently available to detect and estimate ship position in the range of 20 meters to 5400 meters which is well matched with the purpose of the study of detecting and locating small ships in a close range with conventional video cameras.

神戸大学大学院海事科学研究科

発表番号	1	学籍番号	072W601W	氏名	KOCAK GAZI
題 目	Accuracy improvement of target ship position estimation by a shore set-up stereo vision system				
指導教員名	山本 茂広				

1. INTRODUCTION

Recently, safety ship navigation has become more important by the increment in the ship sizes and number of ships worldwide. Marine accidents and collisions are forming a great risk for environment and maritime sector. The maritime authorities of countries want to have better control opportunity of ship traffic. However there is not enough preventive action taken and the frequency of maritime accidents is increasing.

Detection of ships and autonomous navigation is getting more important day by day due to reduction in the number of mariners and for safer navigation and prevention of marine accidents, especially collision. There are already some detection devices as RADAR/ARPA (Automatic Radar Plotting Aid) and AIS (Automatic Identification System) on board ships. AIS is mandatory only for the vessels which are greater than 500GT. Therefore, the ships smaller than 500 GT and which are not carrying AIS device induce a high risk of collision. Even more in case of small ships, the radar cannot distinguish the ship and sea clutter and may cause false detection. Therefore, these devices may not be enough for collision-free navigation especially in domestic waters where small vessels like fishing boats exist. The accident inquiry reports confirm that the collisions are mostly caused by cargo ships and fishing vessels within the range of 3 nautical miles from shore.

Application of computer vision and digital image processing techniques can be effective for this purpose. Stereo vision system is one of the most exploited systems to recover 3D information from 2D images. Therefore, we propose a stereo vision system using digital images for detecting and 3-D location measurement of ships.

2. STEREO VISION SYSTEM AND STEREO ERROR

Especially in the long range (by 5 km) data 3-D error is increasing. To improve the accuracy of 3D data, these errors should be removed. One of the major causes of these errors is based on image plane quantization errors. The quantization errors arise from use of digital images that has finite number of pixels.

The model of our stereo system and error is illustrated in Figure.1. Assuming the p_l is true and p_r is detected with an error of ε, the 3-D location of P is estimated as P'. The estimated error of the location is obtained by the following equation,

$$\begin{bmatrix} X'-X \\ Y'-Y \\ Z'-Z \end{bmatrix} = \frac{\varepsilon Z}{L_B - \varepsilon Z} \begin{bmatrix} L_B/2+X \\ Y \\ Z \end{bmatrix}. \qquad (1)$$

The error ε in the camera coordinate system can be expressed in pixel coordinates by

$$\varepsilon = \frac{2\delta}{N}\tan\frac{\theta}{2}, \qquad (2)$$

where θ is horizontal angle of view, N is a horizontal image size and δ is the detected error of p_r expressed in pixels.

3. CAUSE OF ERROR AND PROPOSED SOLUTIONS

Our solution to overcome the quantization error is calculation of corresponding points between stereo pair images with sub-pixel order accuracy. There have been several methods proposed for sub-pixel image matching. Intensity-based sub-pixel registration is preferred in this study. In addition to that, refinement of the disparity values by low pass filtering (LPF) yields smoother values.

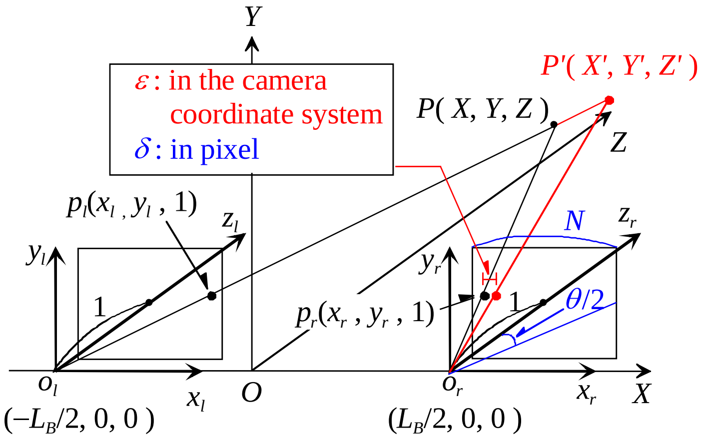

Figure.1 Normal stereo camera model

Applying the mentioned methods we observed a great improve in the accuracy of our stereo camera system. Figure 2 shows track of a ship measured by an experimental system, in which parameters are N=720, L_B=8.14m, θ=43°.In Figure.2 this improvement can be seen visually. 77% of the range error is eliminated. In a determined range the Z dimension error of ±200 m. is reduced to around ±35 m. This result indicates that we can reduce sub-pixel matching error δ to 0.11 pixels on image plane coordinate by means of the proposed method.

In addition, we can estimate the error in various conditions for the different values of N, L_B, θ *and* Z using equation (1) and (2). The maximum measurable range with 150m range error is 5400m in case of the experimental system. If we use a higher resolution camera, for example N=2048, and set the baseline length to L_B=25m, the range error can be reduced to 50m for an object at a distance of 5400m. Therefore, it is sufficiently available to detect and estimate ship position in the range of 20m to 5400m with 1% error of ship distance which is well matched with the purpose of the study and locating small ships in a close range.

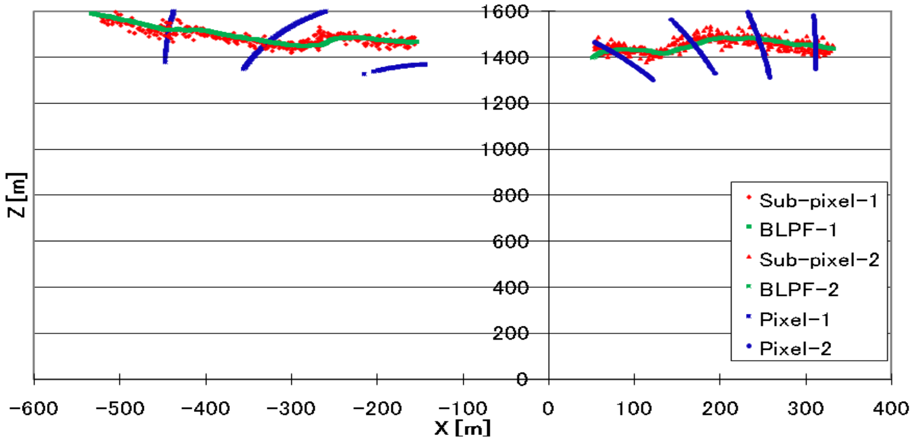

Figure.2 Step-by-step improvement of position estimation

4. CONCLUSION

In order to measure ship position by the stereo vision system, errors due to corresponding point matching was investigated and it was found that we can obtain corresponding points with an accuracy of about 0.1 pixels. This result indicates the possibility that we can measure ship distance with a sufficient accuracy, 1% error of ship distance.

修　士　論　文

Accuracy improvement of target ship position estimation by a shore set-up stereo vision system

（海岸設置ステレオビジョンシステムによる船舶位置計測の精度向上）

神戸大学大学院海事科学研究科

海事科学専攻

学籍番号　072W601W

氏　　名　KOCAK GAZI

（平成２１年７月２４日提出）

CONTENTS

CHAPTER 1

INTRODUCTION

Recently, safety ship navigation has become more important by the increment in the ship sizes and number of ships worldwide. International trade through shipping is significant, and expected to continue increasing over the next decade. Almost 90% of international trade is carried out via shipping. In 2006 global shipping was accounted as 30,686 billion tonne-miles. In 2007 it was 32,000 billion tonne-miles and was projected to grow in 2008 to more than 33,000 billion tonne-miles, more than a 3% per year increase. Marine accidents and collisions are forming a great risk for environment and maritime sector and this risk is gradually increasing by the increment in the number and the size of the ships. Maritime authorities of countries want to have better control opportunity of ship traffic especially in strict areas as territorial waters and straights due to the risk of accidents and environmental pollution. For this reason the VTS's (Vessel Traffic Services) are developed and many people employed at VTS Centers to be able to control the traffic flow. Even more, worldwide maritime authority IMO (International Maritime Organization) constitutes new laws to achieve safer ship navigation and safer sea areas. However there is not enough preventive action taken and the frequency of maritime accidents is increasing. Even more, the most encountered type of marine accidents is collision. Acccording to the report of Marine Accident Inquiry Agency more than 80% of accidents are due to human factors and mainly "improper lookout" and "non-compliance with marine traffic rules"[1]. Therefore it is clear that improving the lookout and piloting of the ship will have a key factor for decreasing the number of marine accidents.

For safer navigation and prevention of marine accidents, especially collision, ship detection has a great importance. There are already some detection devices as

RADAR/ARPA (Automatic Radar Plotting Aid) and AIS (Automatic Identification System) on board ships. AIS is mandotory only for the vessels which are greater than 500GT. These equipments are capable of detecting ships which are greater than 500 GT in the range of 20 nautical miles. However, the ships smaller than 500 GT and which are not carrying AIS device induce a high risk of collision. Even more in case of small ships, the radar cannot distinguish the ship and sea clutter and may cause false detection. Therefore, these devices may not be enough for collision-free navigation especially in domestic waters where small vessels like fishing boats exist.

The accident inquiry reports show that the collisions are mostly caused by cargo ships and fishing vessels which are in the category we mentioned above. Therefore, this study has importance to detect ships in the range of about 3 nautical miles (about 5 km) -which is important to avoid from collision- which are small in size and not carrying AIS. The best way to do this is considered to be an artificial vision system that actually is carried out by human now. The inquiries show that most of the collisions take place in the range of 3 nautical miles from shore. One of the reasons for this is the fatigue of navigation officers. After a hectic work in port (loading and discharging), the navigation officer is already tired and has to carry out a navigation watch. Besides, near the port area the marine traffic is expected to be dense. In this case, the collision risk is increasing. By the application of such an assistant look-out system, the workload of seafarers will be reduced and safer navigation environment will be constituted.

In maritime field, detection of ships and autonomous navigation is getting more important day by day due to reduction in the number of mariners and increasing number of ships. Application of computer vision and digital image processing techniques can be effective for this purpose. This subject has not been studied sufficiently especially in maritime sector, so far. Some limited number of studies already has been carried out, but an accurate and precise system has not been developed yet.

Stereo vision system is one of the most exploited systems to recover 3D information from 2D images. However, it has been applied almost for the land based robotic applications and not for maritime field. Therefore, stereo vision using digital images is one of the candidate methods for ship detection, 3-D location measurement and autonomous navigation of ships. Hence, this research is proposing a new method for automatic detection and accurate position estimation of ships through application of a stereo vision system.

1.1 Related Works

To achieve the autonomous vessel navigation (ground vessel, airborne vessel, planet discovery vessel, marine surface vessel and marine underwater vessel) the studies are focused on two major topics of obstacle detection and collision avoidance.

1.1.1 Vision-based Ship Detection

There are various studies carried out for ship detection. Except for the RADAR and SONAR detection techniques it's almost based on detecting from images. However, there are numerous imaging techniques as infrared (IR), SAR (Synthetic Aperture Radar), ISAR (Inverse Synthetic Aperture Radar), satellite, laser and optical (or digital) images. Most of the ship detection studies are exploiting satellite SAR imagery and optical satellite imagery. A comparatively new study is presenting the state-of-art in ship detection with SAR imagery [2]. In reference [3] a collision avoidance system is presented which integrates the radar and infrared (IR) image information and in reference [4] classification of ship types exploiting Forward Looking Infrared (FLIR) images is studied.

Besides, the number of digital image based ship detection studies is very limited. Reference [5] purposes a ship detection method through combining some image processing techniques to the sea images which are taken from a shore set-up single photo camera. Shimpo et al. studied real-time detection of ships by calculating optical flow and moving vector between the digital image sequences and integrated this with other navigational equipments and AIS [6][7][8]. In reference [9] a feasible navigational lookout support system applying image processing techniques to the image sequence from a single video camera is studied. In this study combination of region segmentation and optical flow techniques are emphasized. Yamamoto et al. has a unique study of real-time detection of ships from a shore set-up camera system using stereo image sequences [10] [11]. Santhalia et al. studied classification of ship types applying neural networks to digital images from a single stationary camera [12]. Luna et al. proposed another method for ship identification which needs side view of ship from some different spectral ranges (CCD, FLIR, image intensifier) and compares this view with the stored database using Concavity-Convexity Scale Space (CCSS) which is an improved model of Curvature Scale Space (CSS) [13]. Ju et al. studied detecting and extracting the moving ship from complex background by applying mixture Gaussian models [14]. Huang et al. proposed a new method for detecting and 3D modeling of ships based on real-time segmentation and a novel clustering algorithm [15]. Wang et al. used change detection, morphological operators and Connected Components Labeling for ship detection of a seaport surveillance system [16]. Phelp built a digital image monitoring system which detects the excessive motion of ships in port and integrating ship data with digital terrain map (DTM) to be able to safety management of the port operations [17].

1.1.2 Vision-based Target Ship Position Estimation

Smith et al.'s study is one of the oldest studies for the interpretation of ship images which gets the estimated location, orientation, dimensions, and heading of the ship using the spatial moments of the image [18]. References [10] and [11] are studying not only ship detection but also detected (target) ship's position estimation. Reference [19] and [20] shows a method to compute the course of target ship and distance between a target ship and own ship by means of a minute angle between the horizon and waterline using an image taken by a single video camera.

The topics of obstacle detection and localization are mostly studied in realizing unmanned (autonomous) vehicles. In maritime sector these are separated into two categories of unmanned surface vehicles (USV) and autonomous underwater vehicles (AUV). In reference [21] the obstacle detection and collision avoidance of a USV is carried out by categorizing the distance of target into two parts of near-field (180 -300 m) and far-field (more than 300m). Different techniques applied for these categories. For near-field obstacle avoidance raw radar, stereo vision, monocular vision, nautical charts and millimeter wave radar data are used. For the far-field AIS contacts, nautical charts and ARPA contacts are exploited. Dunbabin et al. proposed a self docking of USV using a vision system of a single camera, GPS and Inertial Measurement Unit (IMU). This vision system has three primary functions: (1) Target segmentation from the image, (2) correction for camera lens distortion, (3) transformation from image coordinates to global coordinates. However, target distance and length data is not so reliable according to the experimental results [22] [23]. Ebken et al. applied unmanned ground vehicles (UGVs) technology to USVs. The primary obstacle avoidance sensor is a digital marine radar system produce by Xenex Inc. The radar provides both the raw radar image and the contact track data. It was intentioned to primarily use the raw radar image for obstacle avoidance and add the ARPA contact data to the obstacle map. However, due to unreliable nature of ARPA and high minimum detectable range (around 100 m) of digital marine radar a stereo vision system is included to be able to detect the closer objects [24]. Huntsberger et al. proposed an autonomous surface vessel which consists of a dynamic planning engine, a behavior engine and a perception engine. The perception engine algorithms derived from those used onboard the Mars Exploration Rovers (MER) for passive stereo imaging, hazard detection/avoidance, and visual localization for navigation, combined with previous work by Spatial Integrated Systems in the areas of sensing and map-making. Camera models based on polynomial expansions used to correct camera/lens distortions are derived from a series of images obtained during a calibration procedure. A fast stereo algorithm developed at JPL (Jet Propulsion Laboratory) is used to generate a range map during the USV motion [25]. Reference [26] built an autonomous surface vehicle which has a navigation system that mainly consists of a monocular camera, a GPS and Inertial Navigation System (INS). Kalman filter is used for obstacle position estimation.

1.2 Objectives of the Research

In our study we propose a stereo vision system for accurate detection and localization of the ships. One of the advantages of stereo vision is you can both detect and calculate the location of the object. So it can be used not only for the marine surveillance system but also as a navigation tool for the ships.

In the application of stereo vision system for the detection of ships some major problems have to be sorted out. One of them is the localization of the ships. Due to some characteristic quantitative errors of image acquisition systems and imaging geometry some errors take place especially in the distant locations and exact location of the ship can not be measured. These errors shold be removed or decreased. Two kind of solutions are applied to reduce these errors.

1. Sub-pixel image registration
2. Disparity refinement

Sub-pixel registration is necessary to reduce the error which is based on insufficiency of image plane quantization. We will investigate the effect of quantization error and calculate the percentage of quantization error versus total error.

Disparity is the main parameter of the stereo vision to be able to calculate the 3-D data. However, disparity data may include some vibration due to matching cost and noisy image data. And this vibration will result in the 3-D measurement data to be erroneous. The value of 3-D error caused by disparity vibration will be investigated, too.

Proving the effects of quantization error and disparity vibration, these error will be eliminated or reduced to make the stereo 3-D calculation as accurate as possible.

1.3 Research Overview

In the previous study, feasibility of a stereo vision system was studied to detect ships and acquire ship position data. In that study, camera calibration was emphasized. However, especially in the long range (by 5 km) data some errors were encountered. To improve the accuracy of 3D data, these errors should be removed.

One of the major causes of these errors is based on quantization errors. The quantization errors arise from use of digital images that has finite number of pixels. They affect the accuracy of the system especially in long range measurement due to increasing depth resolution with respect to range. That's why image plane quantization error elimination is one of the purposes of this study. Our solution to overcome the quantization error is calculation of corresponding points between stereo pair images with sub-pixel order accuracy. There have been several methods proposed for sub-pixel image matching. Intensity-based sub-pixel registration is preferred in this study.

Another important range error parameter is inaccuracy and vibration in disparity values of corresponding points. Refinement of the disparity values by low pass filtering (LPF) yields smoother values which results in reduction of 3-D measurement error to 1% of object's distance.

- Chapter 2 explains the general features of stereo vision system and projective geometry.
- Chapter 3 shows how the quantization error occurs and to reduce this error by sub-pixel matching method.
- Chapter 4 points out the disparity value vibrations and application of Butterworth Low Pass Filter (BLPF) to refine the disparity data.
- Chapter 5 displays the experimental studies and results. Error analysis is carried out based upon the experimental results.
- Chapter 6 is conclusion and summary of the final results.

References

[1] Japan Maritime Accident Inquiry Agency : "Report on Maritime Accidents 2007".

[2] D. J. Crisp, "The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery", 2004, DSTO (Defence Science and Technology Organization) Information Sciences Laboratory.

[3] Y. Sato and H. Ishii, "Study of a collision-avoidance system for ships", Control Engineering Practice, Vol. 6, pp. 1141-1149, 1998.

[4] J. Alves, J. Herman and N.C. Rowe, "Robust Recognition of Ship Types from an Infrared Silhouette", CCRTS (Command and Control Research and Technology Symposium), June, 2004.

[5] K. Takahashi, Y. Kobayashi, M. Fujii, N. Shimbo, H. Ueda and K. Tsutsui, "Combined detection method in a sea surveillance system," IEICE Trans. on Inf. & Syst., Vol. E88-D, No.2, pp.230-237, February, 2005.

[6] M. Shimpo, R. Shu, M. Oshima, "Method for Real-Time Detection of Ships using Navigational Image Sequence", Proceedings of Asia Navigation Conference 2007, pp.105-113, Nov, 2007.

[7] M. Shimpo, M. Hirasawa, X. Yu, Y. Arvelyna, M. Oshima, "Method for Detecting and Tracking Ships Using Navigational Image Sequence as an Application of Integrated Navigation Systems", ION NTM 2006, pp.224-235, January, 2006.

[8] M. Shimpo, M. Hirasawa, A. Nakajima, R. Shoji, M. Oshima, "A Detection Method of Moving Ships by Image Processing as a Support System of AIS", Proc. ION NTM 2005, pp.438-446, January, 2005.

[9] M. Shimpo, R. Shu, S. Yamamoto, "Investigation of Image Processing Methods Capable of Supporting Navigational Lookout", Asia Navigation Conference 2008, November, 2008.

[10] S. Yamamoto et al., "Basic Study on Detection of Ships from Stereo Images", Journal of the Japan Institution of Marine Engineering, Vol.42, No.4, pp.713-718, July, 2007.

[11] S. Yamamoto and Myat Theingi Win, "A feasibility study on ship detection using stereo vision", Proc. of TECHNO-OCEAN 2006, 19th JASNAOE Ocean Engineering Symposium, 2006.

[12] G.K. Santhalia, S. Singh and S.K. Singh, "Safer Navigation of Ships by Image

Processing & Neural Network", 2nd Asia International Conference on Modeling & Simulation, 2008.

[13] Á. Luna, C. Miravet, D. Otaduy and C. Dorronsoro, "A decision support system for ship identification based on the curvature scale space representation", Proceedings of the SPIE, Electro-Optical Remote Sensing, Volume 5988, pp. 171-182 , 2005.

[14] S. Ju, X. Chen and G. Xu , "An Improved Mixture Gaussian Models To Detect Moving Object Under Real-time Complex Background", International Conference on Cyberworlds 2008, Sept, 2008.

[15] T. Huang, H. Ma and F. Li, "Segmentation And Recognition Of 3D Objects In Automatic Navigation", Fourth International Conference on Natural Computation (ICNC'08), March, 2008.

[16] Da-Jinn Wang, Wen-Sheng Chen, Thou-Ho (Chao-Ho) Chen, Tsong-Yi Chen, "The Study on Ship-Flow Analysis and Counting System in a Specific Sea-Area Based on Video Processing", IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing 2008, Page(s):655 - 658, Aug. 2008.

[17] D. Phelp, "Digital image monitoring to optimise safe port operation", 2nd CSIR Biennial Conference, CSIR International Convention Centre Pretoria, 17&18 November 2008.

[18] F.W. Smith and M.H. Wright, "Automatic Ship Photo Interpretation by the Method of Moments", IEEE Trans. Computers, vol. C-20, pp. 1089-1095, Sept. 1971.

[19] Y. Hayashi, N. Wakabayashi and T. Kitahashi, H. Wake, "Computation of the true course of a target ship from its image at sea," Proc. of 8th Congress of International Association of Institute of Navigation (IAIN), pp. 147-154, 1994.

[20] Y. Hayashi, N. Wakabayashi and T. Kitahashi, "Computation of the Distance between a target and own ship from its image at sea," Trans. IEE Japan, Vol.113-C, No.12, pp.1121-1129, December 1993 (in Japanese).

[21] J. Larson, M. Bruch, and J. Ebken, "Autonomous Navigation and Obstacle Avoidance for Unmanned Surface Vehicles", SPIE Unmanned Systems Technology VIII, 17-20 April, 2006.

[22] M. Dunbabin, B. Lang, and B. Wood, "Towards coordinated visionbased docking using an autonomous surface vehicle," in *Proc. Australian Conf. Robotics and Automation*, Brisbane, December 2007.

[23] M. Dunbabin, B. Lang, and B. Wood, “Vision-based Docking Using an Autonomous Surface Vehicle” 2008 IEEE International Conference on Robotics and Automation, CA, USA, May, 2008.

[24] J. Ebken, M. Bruch, and J. Lum, “Applying UGV technologies to unmanned surface vessel's”, In SPIE proc. 5804, unmanned ground vehicle technology VII, 2005.

[25] T. Huntsberger, H. Aghazarian, A. Castano,G. Woodward, C. Padgett, D. Gaines, “Intelligent Autonomy for Unmanned Sea Surface and Underwater Vehicles”Jet Propulsion Laboratory, California Institute of Technology, 2008.

[26] A. Martins, J. Almeida, H. Ferreira, H. Silva, N. Dias, A. Dias, C. Almeida, and E. Silva, “Autonomous surface vehicle docking manouvre with visual information,” in Proc. International Conference on Robotics and Automation, pp. 4994–4999, April 2007.

CHAPTER 2

IMAGE ACQUISITION AND STEREO VISION SYSTEM

Every image acquisition system performs some kind of transformation of 3D space (world) into 2D space (image). Finding the parameters of such a transformation is fundamental to describing the acquisition system. Perspective projective transformation is a good way of describing the behavior of real optics systems which can be described by a linear equation in a higher dimensional space of so-called homogeneous coordinates [1] [2] [3]. Figure 2.1 illustrates a simple perspective projection system.

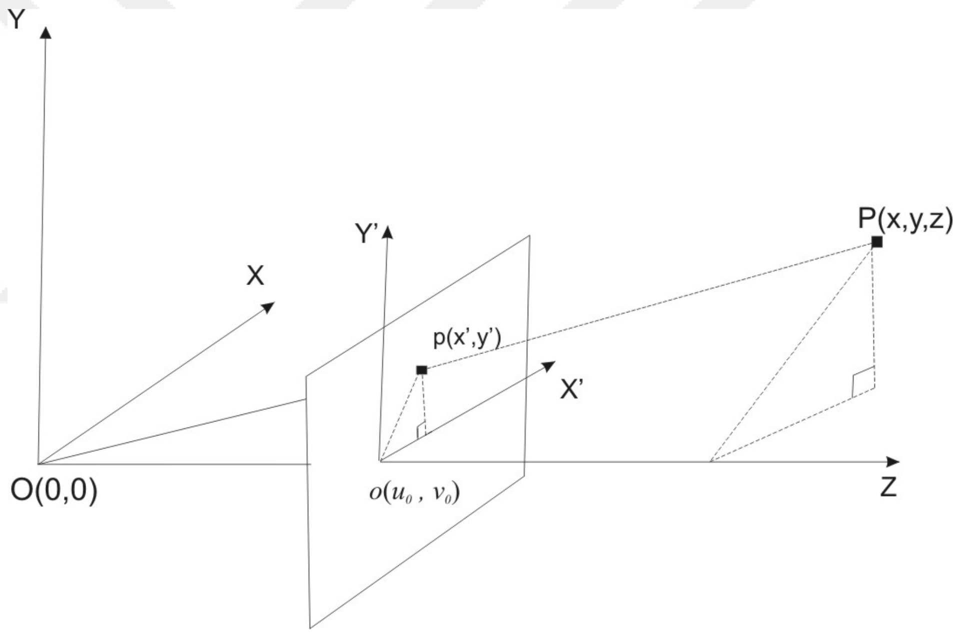

Figure.2.1 Perspective projection

The simplest form of real camera comprises a pinhole and an imaging screen. This mechanism makes an image possible to build up from world space to image space [1]. Two coordinate systems are explained to understand the mathematical relationship between the points in the world space and the imaging space. These are:

1. The external coordinate system which consists of two sub-coordinate systems of
   a. World coordinate system, and

b. Camera coordinate system.

2. The internal coordinate system or pixel coordinate system.

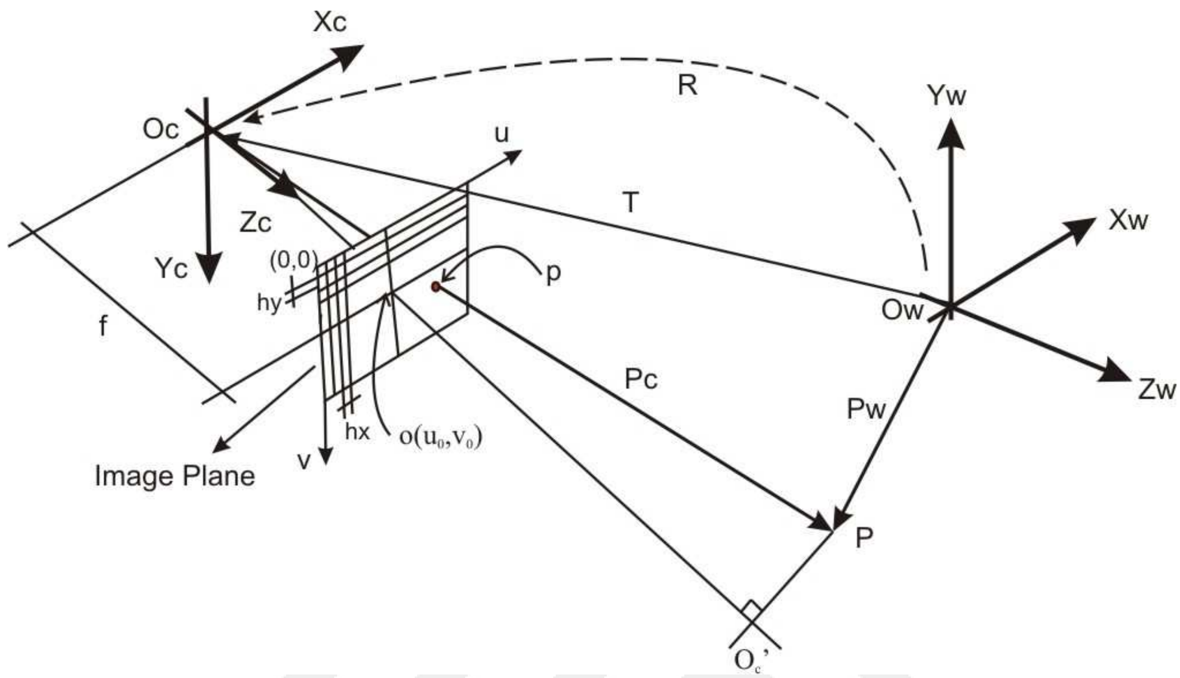

Figure.2.2 Pinhole camera model

In Figure.2.2 the point O_w is the central (or focal) point of world coordinate system which is denoted by the axis X_w, Y_w, Z_w. In our research, world coordinate system is used for estimating three-dimensional coordinates of marine vessels. Similarly, the camera coordinate system consists of X_c, Y_c, Z_c which has the origin of O_c. The relations between the world coordinates and camera coordinates are denoted by a rotation (R) and a translation (T). Another important part of the model is the image plane. The image plane is actually the CCD (Charge-Coupled Device) of the camera which is sensing the brightness values. So, the image is tessellated into rectangular elements which are called pixels. The image plane is parallel to the (X_c, Y_c) plane of the camera coordinates. The projection of the point O_c on the image plane in Z_c direction is called principal point $o(u_0, v_0)$ of image plane which is the center point of the image plane and expressed in pixels. The line through the O_c and the principal point is called principal axis or optical axis. The distance from the O_c to the principal point is the focal length. $P_w(X_w, Y_w, Z_w)$ is representing a point in world coordinates whereas $P_c(X_c, Y_c, Z_c)$ and $p(x,y,z)$ are representing the same point in camera coordinates and its projection onto image coordinates, respectively. Note that $p(x,y,z)$ is still using the units of camera coordinate system. $p(u,v)$ is used for denoting the same point in pixel coordinates after some transformations. The pixel coordinate system that we used can be seen in Figure.3.3.

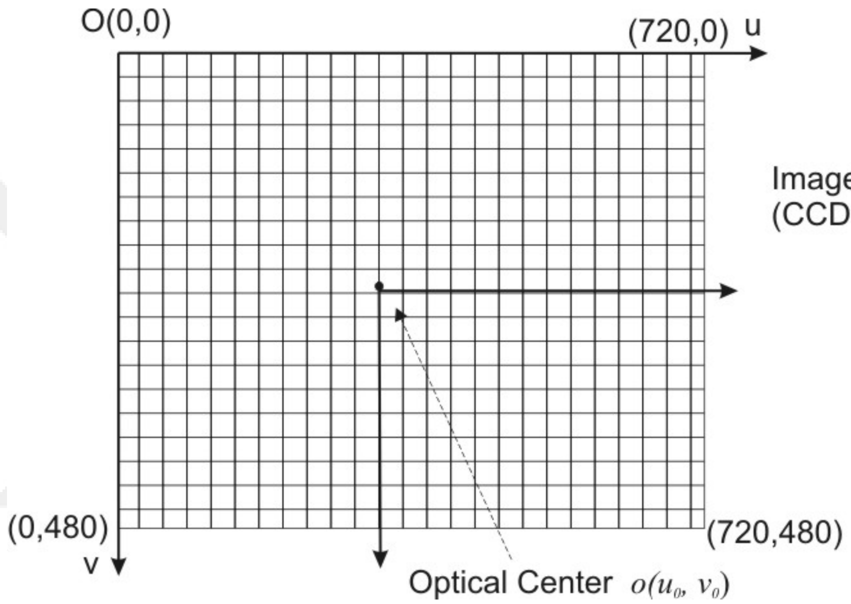

Fig.3.3 The pixel coordinate system.

The relations between the camera coordinates and the image plane (or pixel coordinates) are built by the similar triangles theorem of Thales as follows which are the foundations of the pinhole camera model:

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \qquad z = f \tag{2.1}$$

Here $P(X, Y, Z)$ are coordinates of point in camera coordiantes and $p(x,y,z)$ are the projection of the point onto the image plane (but still in units of camera coordiantes). We can assume that $f = 1$ as different values of f just corresponds to different scalings of the image [4]. In homogeneous coordinates these equations become:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2.2}$$

Incorporating the full camera calibration into the model it becomes:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \mathrm{K} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathrm{M} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2.3}$$

In case of standart stereo system the rotation matrix is identity matrix, theoritically. Therefore assuming the vertical and horizontal physical pixel sizes are equal to 1, scaling factor $\lambda \neq 0$, and the image center (principal point) overlaps with the camera optical axis the mapping matrix can be simplified to:

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{2.4}$$

In real images, the origin of the image coordinates differs from the principal point and the scaling along each image axis is different. Therefore a further transformation is carried out by matrix *K* which defines the intrinsic parameters of the camera. In other words, the product of *p*(*x,y,z*) by matrix *K* results in *p'*(*u,v*) of pixel coordinates. Evenmore, the world coordinates usually does not coincide with the camera coordinates and an euclidean motion (consists of rotation and translation) is described by matrix *M* which defines the extrinsic parameters of camera system. The product of world coordinate by matrix *M* gives the pose of the camera. *K* is independent of the camera position and represented by an uppertriangular matrix:

$$\mathrm{K} = \begin{pmatrix} f/h_x & s & u_0 \\ 0 & f/h_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{2.5}$$

where h_x and h_y are horizantal and vertical physical sizes of one pixel, f/h_x and f/h_y are scaling factors along the x and y axis of the image plane and s is skew between the axis (usually s = 0).

More practical representation of projective mapping from 3D to 2D is:

$$p' = MP_w \qquad \text{In homogeneous coordinates} \rightarrow \begin{pmatrix} \lambda u \\ \lambda v \\ \lambda \end{pmatrix} = M_{3x4} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{2.6}$$

where λ is the scale factor and M_{3x4} is called projection matrix which contains both interior and exterior camera parameters. Computation of the projection matrix is called the projective camera calibration.. In application firstly the projection matrix M is calculated and then intrinsic and extrinsic parameters are estimated by decomposition methods as QR decomposition or SVD decomposition. To eliminate the unknown scaling factor (or for normalization) we can take the ratios and get the pixel coordinates by:

$$u = \frac{M11X + M12Y + M13Z + M14}{M31X + M32Y + M33Z + M34} \qquad v = \frac{M21X + M22Y + M23Z + M24}{M31X + M32Y + M33Z + M34} \tag{2.7}$$

2.1 Extrinsic parameters

For a given point P, its coordinates related to camera and world coordinates are connected by the following formula [1]:

$$P_c = R\,(P_w - T) \tag{2.8}$$

where P_c is the expression of point P in camera coordinates, P_w expresses the same point in world coordinates, R is the rotation matrix and T is the translation matrix between the origins of these coordinate systems.

2.2 Intrinsic parameters

The intrinsic parameters can be summarized as follows:

1. The parameters related to projective transformation: Focal length (*f*)

2. The parameters that map the camera coordinate system into the image coordinate system:

   Origin of the image plane $o(u_0\ ,\ v_0)$, physical dimensions of pixels h_x *and* h_y, pixel coordinates $p'(u,v)$ and camera coordinates x, y of $p(x,y,z)$. The relation is formulated as:

$$x = (u - u_0)h_x \qquad\qquad y = (v - v_0)h_y \tag{2.9}$$

3. Geometric distortion parameters: Considering the radial distortion parameters k_1, k_2 and r new coordinates in pixel coordinate system becomes:

$$x_{corrected} = \frac{u}{1+k_1 r^2 + k_2 r^4} \qquad y_{corrected} = \frac{v}{1+k_1 r^2 + k_2 r^4} \qquad r^2 = u^2 + v^2 \tag{2.10}$$

Geometric distortions can be calculated through some different kind of polinomials, too. Contributing the extrinsic and intrinsic parameters we can express the projective transformation more deficitly:

$$p' = MP_w\ , \qquad p' = \begin{pmatrix} \lambda u \\ \lambda v \\ \lambda \end{pmatrix} = \begin{pmatrix} f/h_x & s & u_0 \\ 0 & f/h_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R_1 & -R_1 T \\ R_2 & -R_2 T \\ R_3 & -R_3 T \end{pmatrix} P_w \tag{2.11}$$

2.3 Stereo Image Acquisition

A common method for extracting depth information from intensity images is to acquire a pair of images using two cameras displaced from each other by a known distance called baseline *(L_B)* [5]. A standart stereo geometry is illustrated in Figure.2.4. It is the simplest stereo camera system with two identical cameras and the coplanar image planes in which optical axis of the cameras are parallel. Each image has its own image coordinate system (pixel coordinate system) with the origin of upper-left corner. Besides, the center point of images overlap with the camera optical axis. The distance between image center (principal point) and optical center is called focal length (*f*). Figure.2.5 illustrates the *X-Z* view of Figure.2.4. In this figure corresponding points and disparity can be seen more clearly.

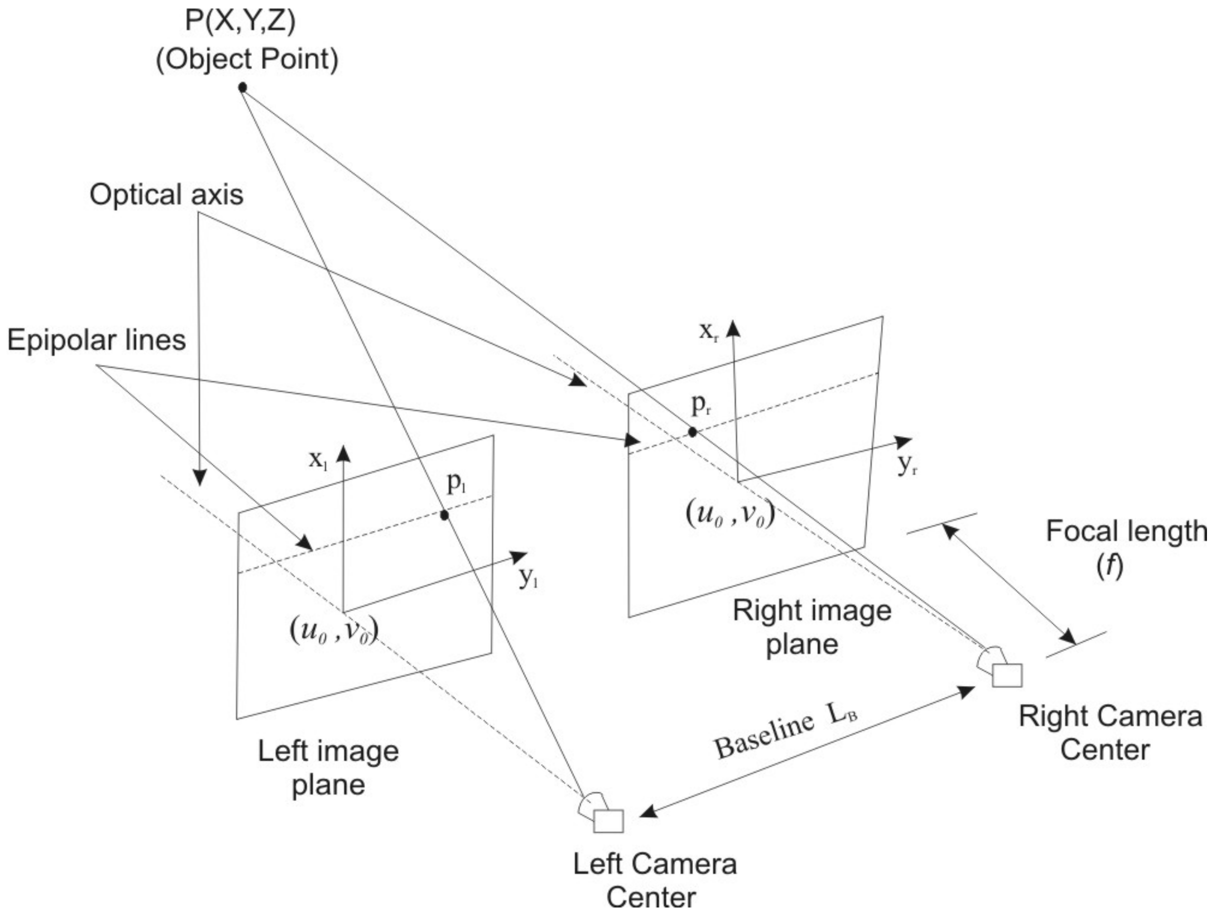

Figure.2.4 A standart stereo image acquisition system

Through the use of similar triangles , the range *(Z)* of the object can be calculated by equation:

$$Z = \frac{f.L_B}{x_l - x_r} \tag{2.12}$$

which is called triangulation. The projection of the scene object is different in left and right images. These are called corresponding points. The displacement between the locations of projections on the left and right image is called disparity which is the denominator of the equation (2.12) as well. Disparity and range are inversely proportional. When the object gets closer, disparity value increases. Notice that equation (2.12) is not valid for real environments and applications. To validate this equation three simplifying asumptions are made [6]:

1. Camera axis are assumed to be paralel,
2. The image planes are assumed to be paralel and coplanar, and
3. Lens is assumed to make projection without distortions and in accordance with the pinhole camera model.

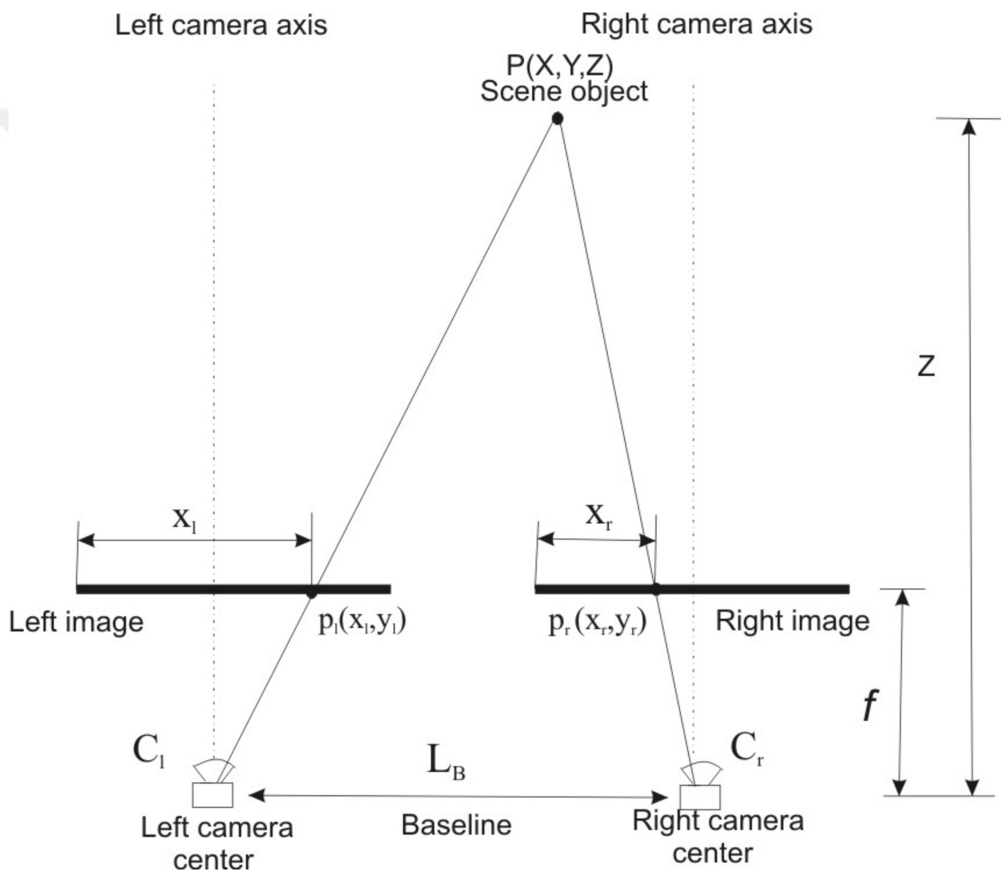

Figure.2.5 Standart stereo system viewed through X-Z axis

In standart stereo system theoritically the vertical disparity is zero. In other words, corresponding points lie on the equivalent vertical locations in image planes $(y_l = y_r)$. The plane passing through the object point in the scene and camera centers is called epipolar plane. The intersection of epipolar plane with left or right image plane defines epipolar lines. Epipolar plane and epipolar line are shown in Figure.2.6. In standart stereo system epipolar lines are paralel to the image scan lines or rows of pixel coordinates with constant y. Epipolar constraint is independent of scene structure and, only depens on the intrinsic parameters and camera relative positions. Epipolar line has a vital importance in case of defining epipolar constraint which decreases the corresponding point search to one dimension. That is because, the projection of any point in left image will lie on the epipolar line of the right image ($y_l = y_r$). In practice there may be a vertical disparity due to misregistration of epipolar lines.

In real environments it is very difficult to realize an ideal stereo configuration. Even a small misalignment of the cameras will cause invalidity of equations. Therefore, some additional calculations as distortion correction and error minimization by linear and unlinear estimation are carried out for reliable results which is called camera calibration. The calibration procedure consists of determining internal and external camera parameters which are mentioned before [7]. Therefore, a linear relation between the 3D real environment and the 2D image plane is constituted in homogeneous coordinates.

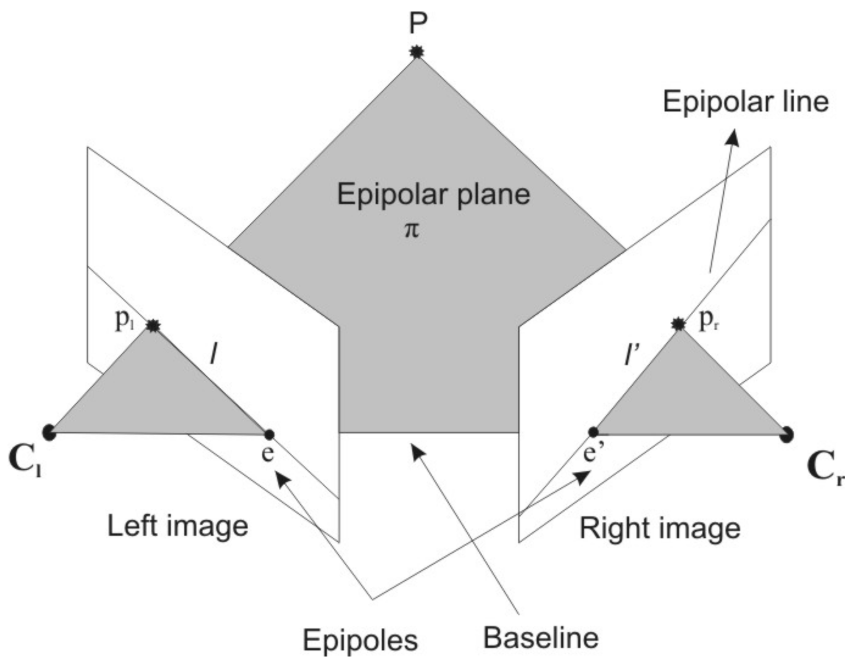

Figure2.6 Epipolar Geometry

2.4 Problems of stereo vision

Stereo camera system is used for calculating the three-dimensional structure of an object using two or more images taken from different viewpoints. The fundamental basis for stereo is the fact that projection of a single three-dimensional object point has a unique pair of image locations in two distinct cameras. Therefore, given two camera images, if it is possible to locate the image locations that correspond to the same physical point in real world, then it is possible to determine its three-dimensional location by triangulation [8].

The essential problems of stereo vision are:

1. Calibration problem,
2. Correspondence problem due to:
    a. Lack of image plane quantization.
    b. Matching cost and brightness quantization.
    c. Occlusions.
    d. Image noise.
    e. Lack of texture.
    f. Other vision problems.
3. Reconstruction problem.

The calibration problem of our stereo system is emphasized during the previous study and the results show that the calibration of the system is good enough [9].

The key problem in stereo is how to find corresponding points in the left and in the right image, referred as the correspondence problem [10]. In other words, matching the pixels in one image with their corresponding pixels in the other image. A detailed study of correspondence problem is carried out by Schartein et al. [11]. In this study we will focus on reducing the three-dimensional error due to correspondence problem.

References

[1] B. Cyganek and J.P. Siebert, “An Introduction to 3D Computer Vision Techniques and Algorithms”, Wiley Publication, 2009.

[2] R.I. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”, 2nd ed., Cambridge University Press, 2003.

[3] O. Faugeras, “Three-Dimensional Computer Vision: A Geometric Viewpoint”, MIT Press, 1993.

[4] R. Mohr and B. Triggs, “Projective Geometry for Image Analysis”, A Tutorial given at ISPRS, Vienna, July 1996.

[5] R. Jain, R. Katsuri and B.G. Shunck, “Machine Vision”, McGrawHill Inc., 1995.

[6] D.C. Huish, “Theoritical and Experimental Performance of a General Stereo Vision System”, Msc. Thesis in Utah State University, 2006.

[7] R.Y. Tsai, “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses”, IEEE Journal of Robotics and Automation, Vol. **RA-3,** No. 4, August 1987.

[8] M.Z. Brown, D. Burschka and G. Hager, “Advances in Computational Stereo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No: 8, August 2003.

[9] S. Yamamoto et al., “Basic Study on Detection of Ships from Stereo Images”, Journal of the Japan Institution of Marine Engineering, Vol.42, No.4, pp.713-718, July, 2007.

[10] A. Koschan, “What is New in Computational Stereo Since 1989: A Survey on Current Stereo Papers”, Technical University of Berlin, August 1993.

[11] D. Schartein and R. Szeliski, “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, International Journal of Computer Vision, Vol. 47, No. 1, pp. 7-42, 2002.

CHAPTER 3

SUMMARY OF THE PREVIOUS STUDY (OR BASIC STUDY)

In our experiment, two same digital video cameras, Panasonic, NV-GS300 are used to capture images with a resolution of 2048x1512 pixels and horizontal angle of view is about 43 degrees by "photo shot mode".

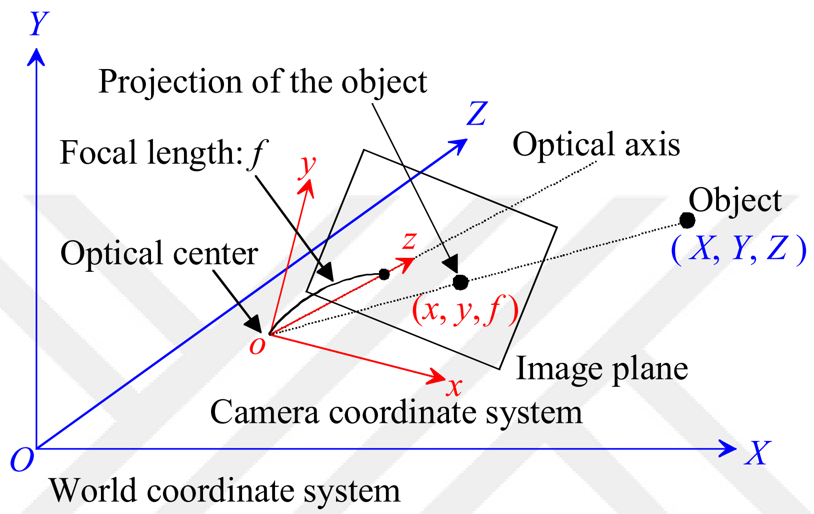

Figure.3.1 Definition of coordinate systems.

There are vaious camera calibration methods. In this study calibration from a known scene is preferred. According to the Figure.3.1 an object point $P_i(X, Y, Z)$ is projected onto the image plane of camera coordinates (x, y, f). And the mapping from camera coordinates to pixel coordinates of (u, v) is carried out by the following functions.

$$\begin{cases} x = f_x(u,v) \\ y = f_y(u,v) \end{cases} \tag{3.1}$$

Here, the intrinsic parameters of the camera are focal length f, and the functions f_x and f_y . Considering the extrinsic parameters of rotation (R) and translation (T) of camera coordinate system, which depends on the camera configuration, the following equation is obtained for calculation of distance in world coordinates:

$$P_{calculated}(X,Y,Z) = k_i R \begin{bmatrix} f_x(u,v) \\ f_y\,(u,v) \\ 1 \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}, \tag{3.2}$$

where (X_0, Y_0, Z_0) is the coordinates of camera center with regard to world coordinates and k_i is a positive real number. Considering the real environment and calculation assumptions there will be an error between the calculated world point and the real world point which can be formulized as

$$\boldsymbol{f}_i = k_i R \begin{bmatrix} f_x(u_i,v_i) \\ f_y(u_i,v_i) \\ 1 \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} - \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \boldsymbol{0} \tag{3.3}$$

If an adequate number of the points P_i $(i=1, 2, \ldots, N)$ and there projections onto the image p_i (i=1, 2, …, N) are observed, unknown parameters in the functions f_x and f_y can be determined by minimizing the following sum of squared errors, S.

$$S = \sum_{i=1}^{N} \boldsymbol{f}_i^T \boldsymbol{f}_i \tag{3.4}$$

Moreover, the geometric distortion is also need to be corrected. To do this, the following forms were adapted as the function f_x and f_y through trial and error.

$$\begin{aligned} f_x(u,v) &= a_0 u + a_1 v + a_2 uv + a_3 u^3 + a_4 uv^2 \\ f_y(u,v) &= a_5 v + a_6 u + a_7 vu + a_8 v^3 + a_9 vu^2 \end{aligned} \tag{3.5}$$

Extrinsic parameters of the cameras, which include locations and orientations of cameras, should be calculated according to our stereo camera configuration which is illustrated in Figure.3.2. The right and left cameras are set at (L_B/2, 0, 0) and (-L_B/2, 0, 0) in the world coordinates. An object $P_i(X_i, Y_i, Z_i)$ in the world coordinate system is projected onto the point $p_{ri}(x_{ri} , y_{ri} , 1)$ in the right camera coordinate system and $p_{li}(x_{li} , y_{li} , 1)$ in the left camera coordinate system, respectively. Then, the following equations are derived in regard to the right camera and left camera, respectively.

$$k_{ri} R(\theta_{Xr}, \theta_{Yr}, \theta_{Zr}) \begin{bmatrix} x_{ri} \\ y_{ri} \\ 1 \end{bmatrix} + \begin{bmatrix} L_B/2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} \qquad k_{li} R(\theta_{Xl}, \theta_{Yl}, \theta_{Zl}) \begin{bmatrix} x_{li} \\ y_{li} \\ 1 \end{bmatrix} + \begin{bmatrix} -L_B/2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} \tag{3.6}$$

where k_{ri} is a positive real number, θ_{Xr}, θ_{Yr} and θ_{Zr} are rotation angles of camera coordinates. Combining these equations the following equation is derived:

$$k_{li} R(\theta_{Xl}, \theta_{Yl}, \theta_{Zl}) \begin{bmatrix} x_{li} \\ y_{li} \\ 1 \end{bmatrix} - k_{ri} R(\theta_{Xr}, \theta_{Yr}, \theta_{Zr}) \begin{bmatrix} x_{ri} \\ y_{ri} \\ 1 \end{bmatrix} = \begin{bmatrix} L_B \\ 0 \\ 0 \end{bmatrix} \tag{3.7}$$

with which the rotation angles of both cameras can be estimated by least squares method, in the case of existence of adequate number of correponding point pairs.

3.1 Stereo Camera Cofiguration and 3D Measurement by Backprojection

The stereo system of this study is configured as it can be seen in Figure.3.2. After the camera calibration is carried out as mentioned above, three-dimensional measurement of a ship can be carried out. Actually this is similar to reverse process of camera calibration. In camera calibration the pixel coordinates are acquired due to a known 3D point in world coordinates. In contrast, while 3D measurement the world coordinate data of a point is calculated up to the pixel coordinate data which is linearly related by camera calibration. To do this, firstly the interested point –marine vessel in our case- in image (pixel) coordinates is extracted or detected in one of images which is determined as the reference image. Left image is the reference image in this study and the notion for the coordinates of the point is (u_l ,v_l). Then the corresponding point of this point in the other image (u_r ,v_r) is found by a similarity measure or matching cost. There are numereous matching techniques as sum of absolute differences (SAD), cross correlation (CC), sum of squared differences (SSD), ordinal measures etc. SSD was preferred for this study to find the corresponing point. The equation of similarity measure with SSD is as follows:

$$R_{SSD}(s) = \sum_{i \epsilon W}\big(f(i) - f(i-d)\big)^{2} \tag{3.8}$$

where *f(i)* and *f(i - d)* are the reference and observed images, respectively. This is the vital point of the 3D measurement, because the accuracy of the measurement is depending on the accuracy of the correspondence.

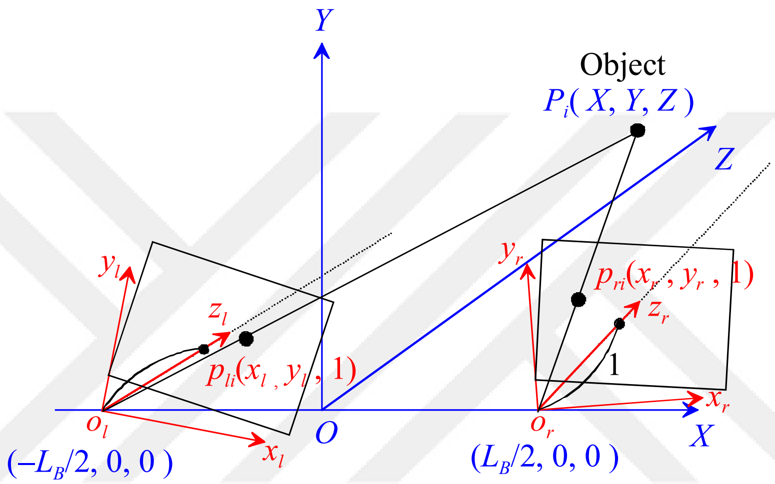

Figure.3.2 Camera configuration of stereo vision

To be able to calculate the 3D data in world coordinates, we have to follow the steps from pixel coordinates to camera coordinates, and finally from camera coordinates to world coordinates. Having the pixel coordinates of interested point in both images, these points are transformed to camera coordinates by the following equation:

$$\begin{bmatrix} {}^{n}x_r \\ {}^{n}y_r \end{bmatrix} = \frac{1}{r_{r31}f_x(u_r,v_r)+r_{r32}f_y(u_r,v_r)+r_{r33}} \begin{bmatrix} r_{r11}f_x(u_r,v_r) + r_{r12}f_y(u_r,v_r) + r_{r13} \\ r_{r21}f_x(u_r,v_r) + r_{r22}f_y(u_r,v_r) + r_{r23} \end{bmatrix} \tag{3.9}$$

$$\begin{bmatrix} {}^{n}x_l \\ {}^{n}y_l \end{bmatrix} = \frac{1}{r_{r31}f_x(u_l,v_l)+r_{r32}f_y(u_l,v_l)+r_{r33}} \begin{bmatrix} r_{r11}f_x(u_l,v_l) + r_{r12}f_y(u_l,v_l) + r_{r13} \\ r_{r21}f_x(u_l,v_l) + r_{r22}f_y(u_l,v_l) + r_{r23} \end{bmatrix} \tag{3.10}$$

where r_{rij} and r_{lij} represent (*i*,*j*)-component of the rotation matrices $R(\theta_{Xr}, \theta_{Yr}, \theta_{Zr})$ and $R(\theta_{Xl}, \theta_{Yl}, \theta_{Zl})$, respectively.

In these equations camera rotation is also compensated –in other words both cameras become parallel in Z-axis-, and focal lengths are unity. Finally, the location of the object in the world coordinate system is obtained by the following equation.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{L_B}{{}^{n}x_l - {}^{n}x_r} \begin{bmatrix} ({}^{n}x_l + {}^{n}x_r)/2 \\ {}^{n}y_l \\ 1 \end{bmatrix} \tag{3.11}$$

For the details of this study you can refer to the reference [2].

3.2 Relation Between the Ship Size and Measurable Distance

The distance calculation of the method depends on the resolution of a camera image, angle of view of the camera, size of a target ship and weather conditions and sea state. Relation among these factors except for weather conditions and sea states are derived as follows.

$$\frac{D}{L_s} = \frac{w}{2w_s \tan(\theta/2)} \tag{3.12}$$

where D is distance of a target ship, L_s is the apparent length of the ship along the X-axis, w is the horizontal resolution of image in pixels, w_s is the size of the ship's image in pixels and θ is the angle of view of the camera, as shown in Figure.3.3.

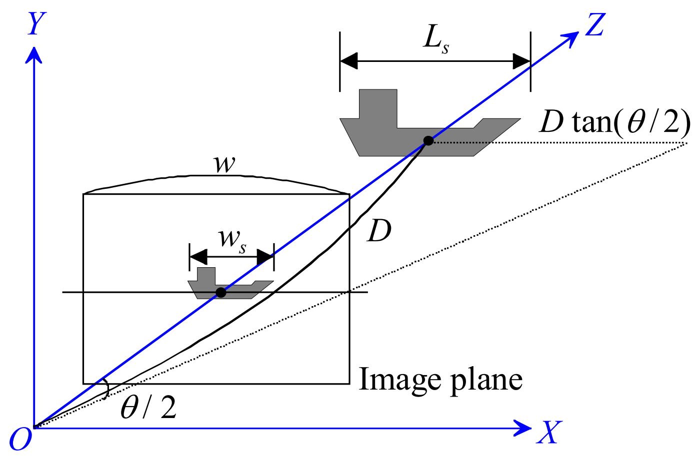

Figure.3.3 Relation among distance, horizontal resolution and angle of view

Assuming that at least 20 pixels of w_s required to detect the ship in the image, where w is 2048 pixels and θ is 43° -the camera used in this paper- distance about 130 times L_s is the limit to detect a ship. As an example, distance of a 50 meters of fishing vessel can be measured in 6500 meters at a clear weather condition.

References

[1] M.Z. Brown, D. Burschka and G. Hager, “Advances in Computational Stereo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No: 8, August 2003.

[2] S. Yamamoto et al., “Basic Study on Detection of Ships from Stereo Images”, Journal of the Japan Institution of Marine Engineering, Vol.42, No.4, pp.713-718, July, 2007.

CHAPTER 4

CONCEPT AND ESTIMATION OF THE STEREO SYSTEM ERROR

Recovering a three-dimensional scene's structure is one of the most important tasks in computer vision. Scene depth sensing technologies can be divided into active methods, such as radar, lidar, ultrasound scanners, or structured light projecting systems, and passive methods, which are predominantly based on stereoscopy [1]. The passive methods are particularly attractive as being the least interfering with the environment [2]. Many approaches have been proposed to model 3D objects and estimate depths using multiple input images. However, the issue of how to appropriately extract useful information in recovering depths from videos is still not addressed well [3]. The major issue is how to utilize disparate scene cues to achieve a more complete and accurate overall scene interpretation [4].

Figure.4.1 shows a normal stereo camera model and concept of the correspondence and 3D reconstruction error. Two cameras are placed at ($-L_B/2$, 0, 0) and ($L_B/2$, 0, 0), respectively, in the world coordinates system *O-XYZ* (Kocak et al. 2009).

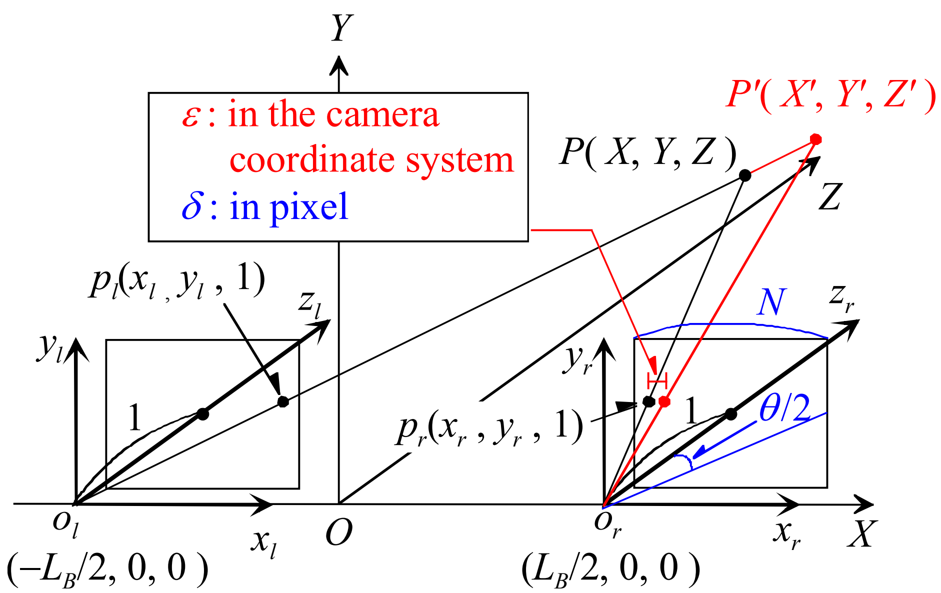

Figure.4.1 Concept of stereo 3D error due to correspondence problem

Focal lengths of both cameras are unity and their optical axes are parallel to Z-axis. In this model, an object located at $P(X, Y, Z)$ is projected onto $p_l(x_l, y_l, 1)$ and $p_r(x_r, y_r, 1)$ in each camera coordinate system, respectively. Then, the coordinate (X, Y, Z) is obtained by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{L_B}{x_l - x_r} \begin{bmatrix} (x_l + x_r)/2 \\ y_r \\ 1 \end{bmatrix}, \tag{4.1}$$

where L_B is a baseline length.

If the x-coordinate of p_r, which is the corresponding point of p_l, is detected with an error of ε, the 3D location of P is estimated as $P'(X', Y', Z')$. The amount of this error can be estimated. Adding this error parameter ε to the general reconstruction equation (Equation 1) the measured erraneous coordinate (X', Y', Z') is obtained by the following equation:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \frac{L_B}{x_l - (x_r + \varepsilon)} \begin{bmatrix} (x_l + x_r + \varepsilon)/2 \\ y_r \\ 1 \end{bmatrix}. \tag{4.2}$$

From Equation (4.1) the following equations are extracted.

$$x_l + x_r = 2X/Z$$

$$x_l - x_r = L_B/Z$$

$$y_r = Y/Z \tag{4.3}$$

Combining these equations with Equation (4.2) we obtain

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \frac{Z.L_B}{L_B - \varepsilon.Z)} \begin{bmatrix} (\frac{X}{Z} + \frac{\varepsilon}{2}) \\ Y/Z \\ 1 \end{bmatrix}. \tag{4.4}$$

The difference of the points $P(X, Y, Z)$ and $P'(X', Y', Z')$ gives us the estimated error in world coordinates by the following equation

$$\begin{bmatrix} X'-X \\ Y'-Y \\ Z'-Z \end{bmatrix} = \frac{\varepsilon Z}{L_B - \varepsilon Z} \begin{bmatrix} L_B/2+X \\ Y \\ Z \end{bmatrix} \tag{4.5}$$

While 3D reconstruction of a image point firstly we obtain the data in pixel coordinates and then transform it to camera coordinates and world coordinates. That's why we are required to express the error in terms of the data of pixel coordinates, as well. From Figure.4. 2 we can better understand this process.

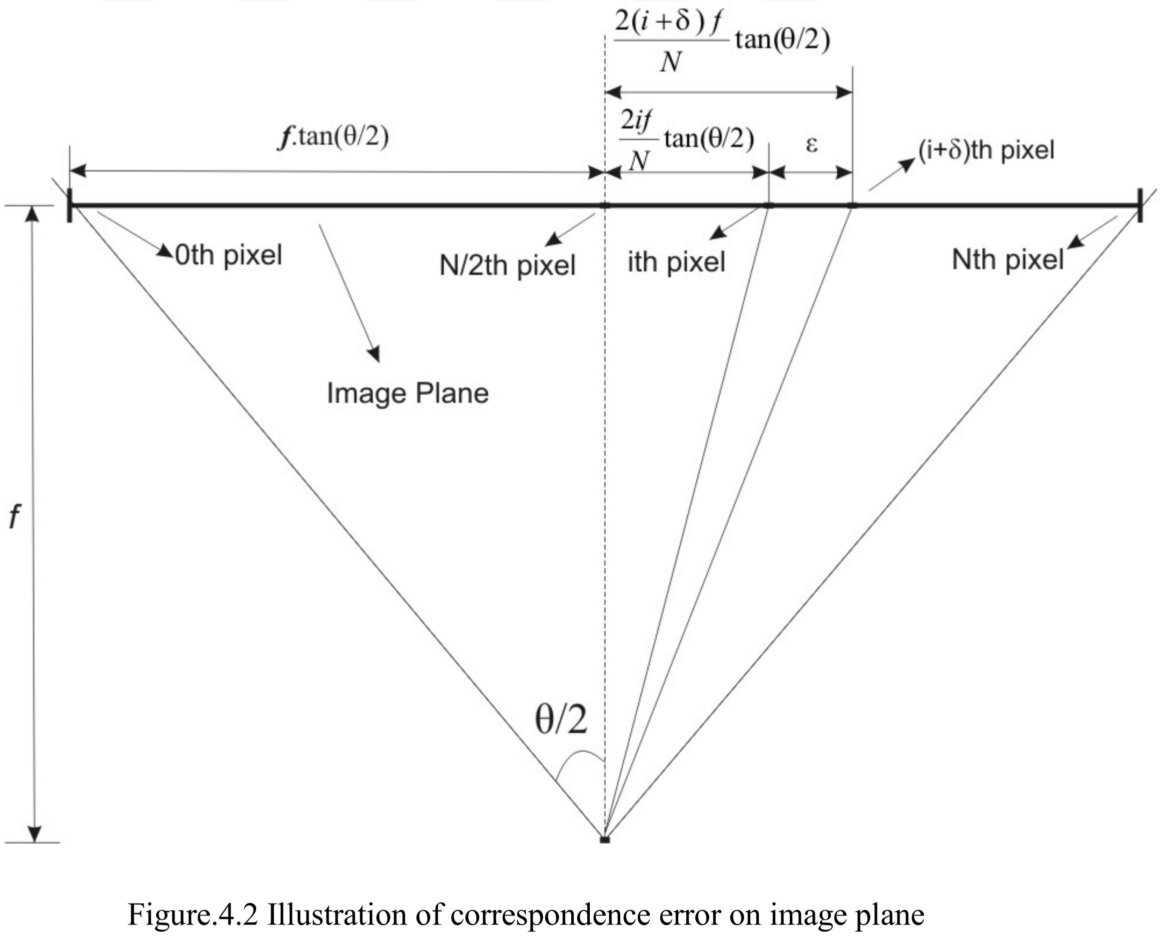

Figure.4.2 Illustration of correspondence error on image plane

Considering δ is the amount of shift in pixel coordinates due to correspondence error, the difference of ith pixel and (i+ δ)th pixel yields the error in pixel coordinates. When focal length f is normalized to 1 (or this image scale is chosen), the error ε in the camera coordinate system is expressed as shown by

$$\varepsilon = \frac{2\delta}{N}\tan\frac{\theta}{2}, \tag{4.6}$$

where θ is horizontal angle of view, N is a horizontal image size and δ is the detected error of p_r expressed in pixels.

In the case of the ship is at the $X = 0$ position of the world coordinates or at the position of $P(0,0,D)$, calculation will be $P'(0,0,D+\Delta D)$ due to error. Figure.4.3 illustrates such kind of error situation. From the figure D is calculated as

$$D = \frac{L_B}{2}\tan\alpha \quad , \tag{4.7}$$

Therefore contributing the error to the equation, it becomes

$$D + \Delta D = \frac{L_B}{2}\tan(\alpha + \Delta\alpha) \quad . \tag{4.8}$$

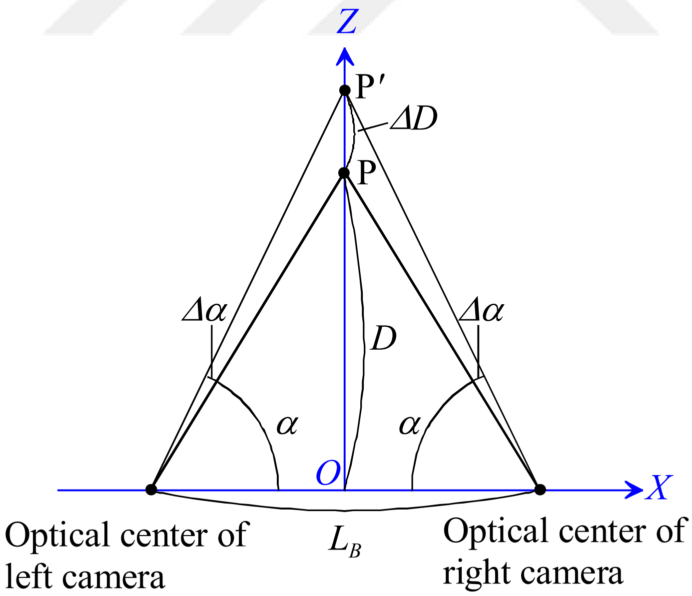

Figure.4.3 Error in case of $P(0, 0, D)$

Equation (8) can be derived from (6) and (7).

$$\frac{\Delta D}{D} = \frac{(1+4k^2)\tan\Delta\alpha}{2k(1-2k\tan\Delta\alpha)} \quad (4.9)$$

where k is D/L_B.

References

[1] M.Z. Brown, D. Burschka and G. Hager, “Advances in Computational Stereo”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No: 8, August 2003.

[2] P. Skulimowski, and P. Strumillo, 2008, “Refinement of depth from stereo camera ego-motion parameters”, *Electronics Letters,* Volume 44, Issue 12, Page(s):729 – 730.

[3] G. Zhang, J. Jia, T. Wong, and H. Bao, 2008, “Recovering consistent video depth maps via bundle adjustment”, *IEEE Conference on Computer Vision and Pattern Recognition*, Page(s):1 – 8.

[4] D.M. McKeown, and F.D. Perlant, 1992, “Refienement of disparity estimates through the fusion of monocular image segmentations”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings* Page(s):486 – 492.

CHAPTER 5

IMAGE PLANE QUANTIZATION AND SUB-PIXEL REGISTRATION

In the previous study the calibration of the cameras was mainly studied and the three-dimensional measurement was carried out based on one pixel accuracy of the image plane which resulted in some erroneous measurement. One of the major causes of the errors is based on correspondence error which is the key problem of stereo systems. In this part we will consider about the correspondence error due to lack of image plane quantization or in other words quantization error. The quantization errors arise from use of digital images that has finite number of pixels. They affect the accuracy of the system especially in long range measurement because of increasing depth resolution. Considering the depth resolution, image plane with accuracy of one pixel, possible three-dimensional reconstruction of adjacent points which are far from image plane will be more difficult. Increasing the distance from camera planes the accuracy of the depth measurement diminishes (or depth resolution increases) due to the geometrical limitations caused by the geometrical parameters of a stereo system. In other words, the corresponding measurement area of one pixel in world coordinates is increasing. It can be better understood from Figure.5.1.

That's why quantization error elimination is aimed to achieve more accurate correspondences of stereo image pairs. We applied sub-pixel matching as a solution to eliminate quantization errors. This is to calculate corresponding points between stereo pair images with sub-pixel order accuracy which yields decreased depth resolution and 3D reconstruction error. Some methods for sub-pixel image matching have been already proposed. Intensity-based sub-pixel registration is preferred for this study.

5.1 Spatial Image Plane Quantization

A crucial task that faces computer vision and other triangulation systems is the ability to obtain accurate three-dimensional position information in the presence of limited sensor resolution. Sensors for computer processing applications produce sampled, quantized data whose spatial resolution is determined by limits in device technology and bandwith. For the construction of accurate depth maps the resolution requirement is severe. Thus, to live within the constraints of limited spatial resolution, a greater understanding of positon error from image plane quantization is crucial [1].

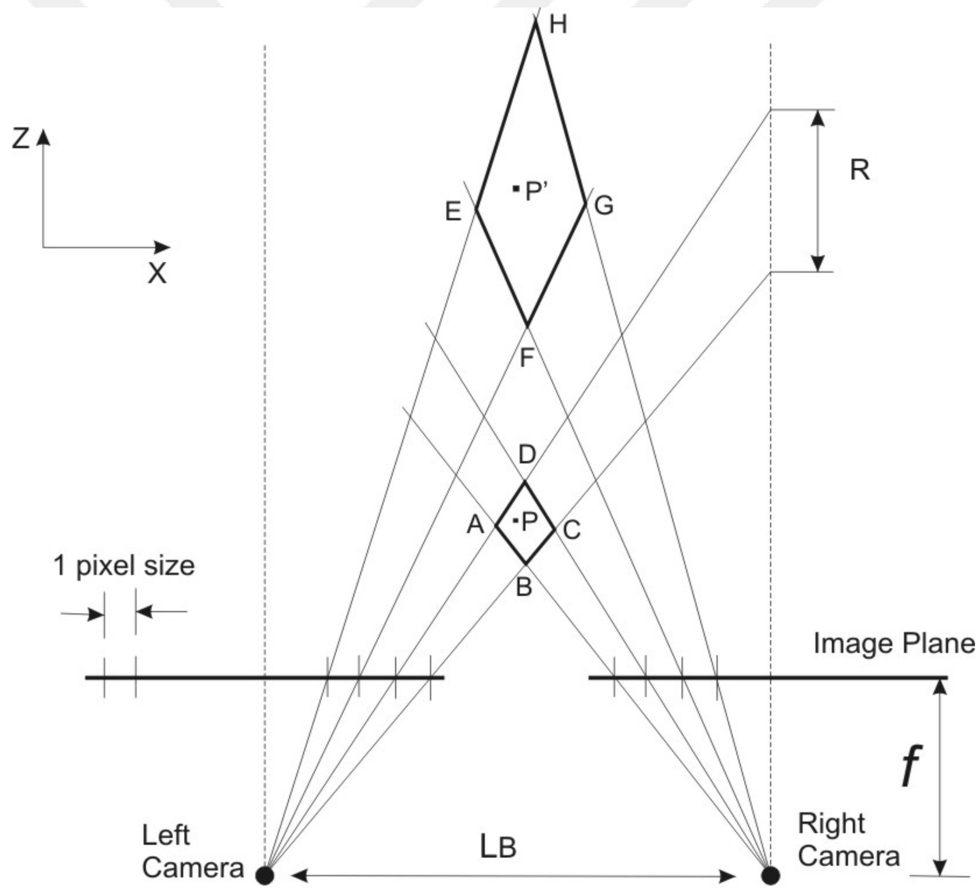

Figure5.1 Image plane quantization, region of uncertainty and depth resolution

In most of the stereo matching methods, it is assumed that one scene points' projection onto the image plane is excatly onto the integer valued image coordinates. However, in reality, the exact location of the projection onto the image plane is within sub-pixel level of ± 0.5 pixel. Considering an ideal pinhole camera model and the correspondence error due to spatial image plane quantization, a points'measured position maps to a small volume in 3D due to finite resolution of the cameras which is called region of uncertanity. Figure.5.1 illustrates this situation. In the figure ABCD rectangle is the X-Z direction view of region of uncertanity for point P. The exact location of the scene point can be anywhere in this region. Evenmore, it is seen that even a small change in the objects location in Z direction, has a great effect in that region. If the object is moved to point P' from point P, the region of uncertanity is increasing to rectangle of EFGH. In the case of long range object points, the spaciousness of region of uncertanity can be imagined. This causes greater errors in 3D recovery of long range points. This error has horizantal, vertical and range (Z direction) components but range error dominates over the horizantal and vertical error. That is why only range error is considered in this study.

5.2 Explanation of Sub-pixel Image Registration

Image matching is mainly concerned with finding a given target image in a reference image based on a pixel-by-pixel comparison of the target image with every possible sub-image of the reference image. Image registration is more concerned with how to make equivalent geographic points of a scene or objects in two images to be coincided, where the two images are taken from the same scene either by using different positions or different times [2]. In stereo vision, there are left and right images of the same scene from different positions (and different times in case of image sequences). In conventional type, two images are registered in pixel-by-pixel accuracy. However, because of the different aspects of the cameras, pixel-by-pixel comparison is not sufficient for accurate registration and causing quantization errors. In sub-pixel registration, it is achieved at an accuracy of a fraction of a pixel. Up to the type of sub-pixel estimation method 0.005 pixel registration accuracy can be achieved. Mainly, there are 5 types of sub-pixel displacement estimation [3]:

(1) Similarity Interpolation,

(2) Intensity Interpolation,

(3) Gradient-based Interpolation,

(4) Phase Correlation,

(5) Geometric Method.

In this study, we will exploit the similarity interpolation method of Shimizu and Okutomi which is based on intensity interpolations [4]. Similarity interpolation estimates the peak of a similarity function by fitting a parabola to the three indices near its extremum. Some different similarity measures as SAD (Sum of absolute differences), SSD (Sum of squared differences) or CC (Cross correlation) can be used. In this study we preferred SSD due to better results. While estimating sub-pixel displacement through similarity interpolation, SSD between left and right images are carried out and the highest value is selected as center pixel I (0). The preceding pixel becomes I (-1) and the next pixel is I (1). Then the SSD values of these pixels are also calculated according to the following formula:

$$R_{SSD}(s) = \sum_{i \epsilon W}\big(f(i) - f(i - d + s)\big)^{2}, \tag{5.1}$$

where $f(i)$ and $f(i - d + s)$ are the reference and observed images, respectively, and s is the shift value in pixel unit from the extremum point. It is clear that the R(-1) and R(1) is larger than R(0). The sub-pixel estimation is based on these SSD values. After calculating SSD values, a parabola fitting is carried out which’s centerline location is the estimated sub-pixel position. Figure.5. 2 is illustrating this situation.

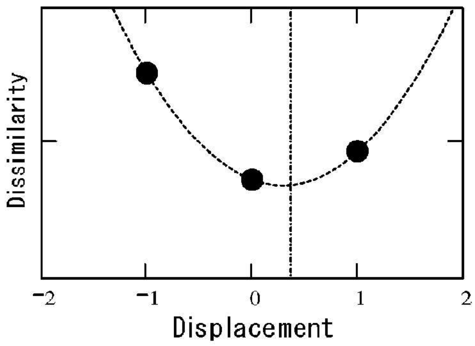

Figure.5.2 Parabola fitting of similarities

Following formula is used for the calculation of the parabola fitting:

$$\hat{d} = \frac{R(-1) - R(1)}{2R(-1) - 4R(0) + 2R(1)} \quad . \tag{5.2}$$

where R(-1), R(0) and R(1) are the similarity values obtained from Equation (5.1).

5.3 Experimental Results

In our experiment, two identical digital video cameras, Panasonic, NV-GS300, with a resolution of 720x480 pixels were used for image acquisition by "video mode" to get image sequence of 10 frames per second. The horizontal angle of view is about 43 degrees was constant in the following experiments. Figure.5.3 shows the experimental set up.

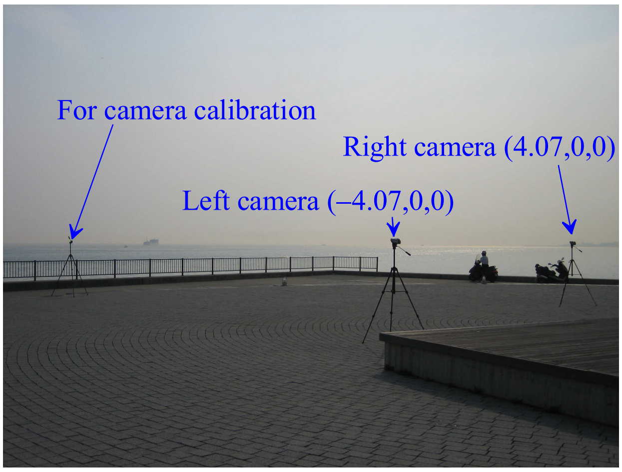

Figure.5.3 Experimental configuration of cameras

A degree of errors due to correspondence problem was verified using real images. The images were obtained at Akashi Strait. The baseline length is 8.14 meters. Calibration for intrinsic and extrinsic parameters of the cameras is briefly explained in Chapter 3.

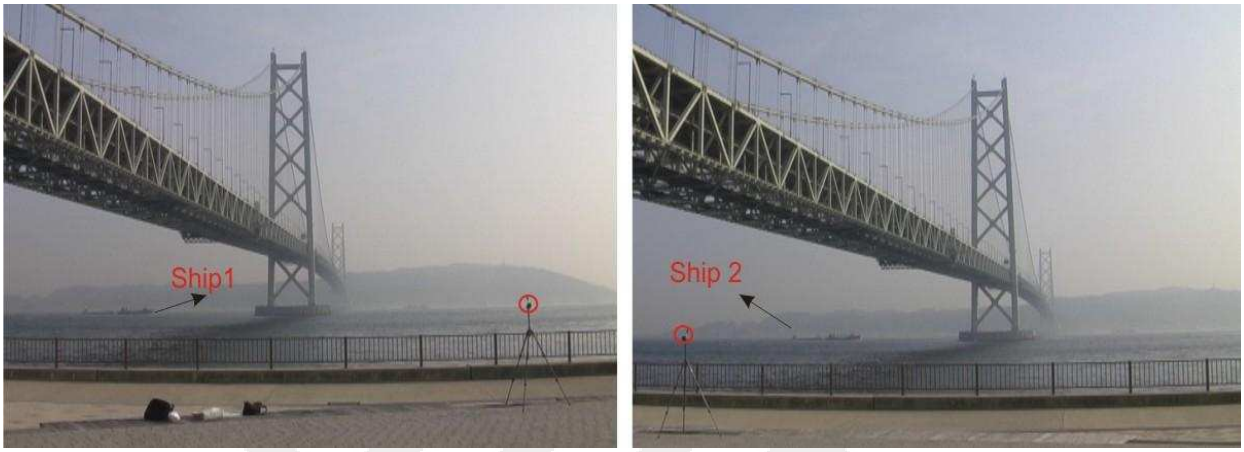

Figure.5.4 Sample stereo images of experimental configuration

Figure.5.5 Tracking point of ship in image sequences

Figure.5.4 indicates an example of stereo images of Akashi Strait and Figure.5.5 shows the tracking point of a ship in the image sequences. The corrsponding point of this tracking point in the right pair of the stereo images is detected by similarity matching using SSD. Figure.5.6 is one of the matching result of our experiments.

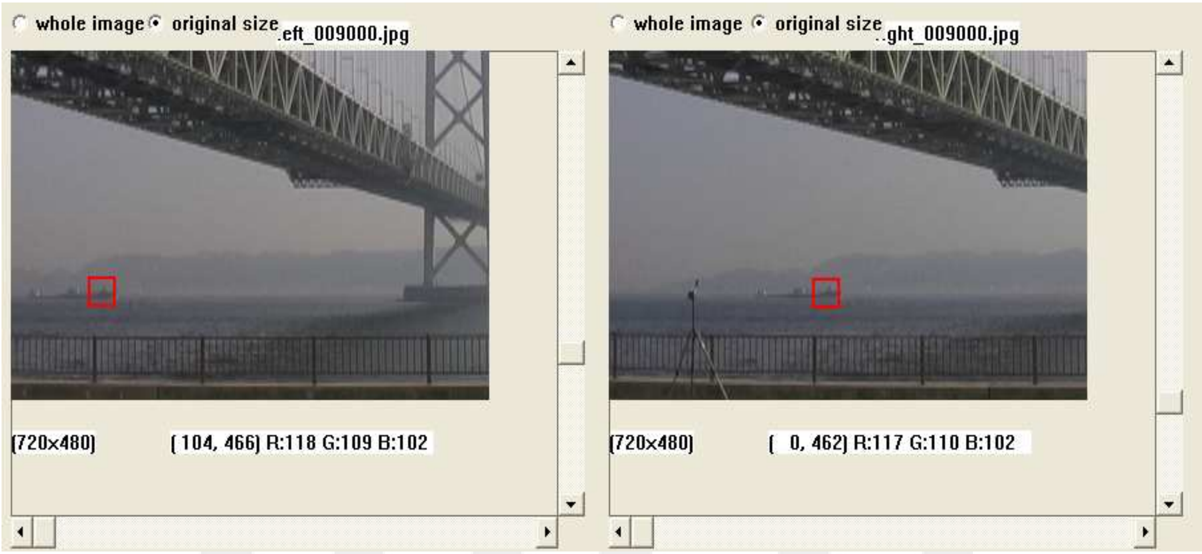

Figure.5.6 A sample image of image matching with similarity window

There are two ships in the right of the image. The localization of one of them which is calculated in one pixel accuracy of image plane is shown in Figure.5.7. In the figure, the location of the ship's bridge, the red point in Figure.5.5, is traced and plotted over about 38.5 seconds, 385 frames. The blue diamonds are 3D measurement results obtained by stereo pair matching. *X*-axis means the rightward direction of the image, and *Z*-axis does the depth direction. Since the actual locations of the ship are unknown and it navigates almost along the *X*-axis, errors of *X*-coordinate cannot be estimated. Therefore, those of *Z*-coordinate will be examined in the following.

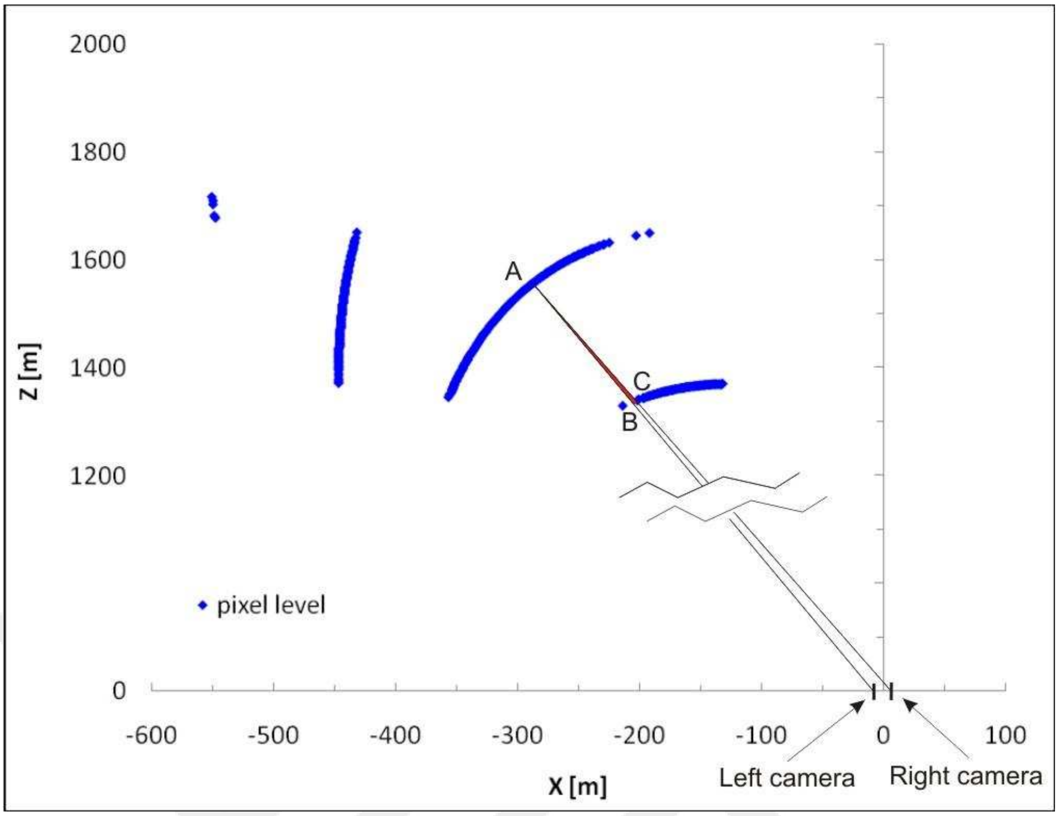

Figure.5.7 Illustration of quantization error in our experimental results

Despite it is not possible to illustrate the quantization error in the size of 3 km, it is tried to illustrate in the Figure.5.7. For understanding the error due to image plane quantization Figure.5.8 can be helpful comparing this figure with Figure.5.1.

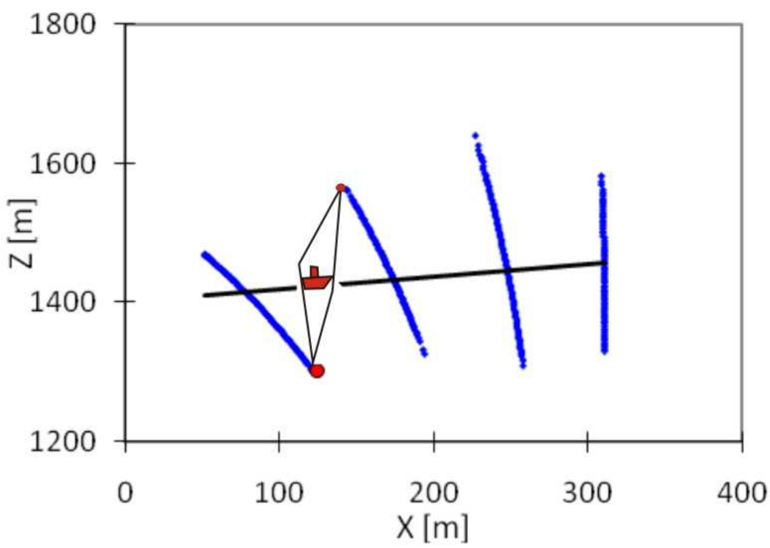

Figure.5.8 Experimental results of region of uncertainty

In Figure.5.8 we can see the experimental results of so-called region of uncertainty. In the figure the ships estimated true position is assumed to be on the fitting line of calculated data. As it can be seen from the Figure.5.7 and Figure.5.8 the measurement results are not continuous. This is because of before mentioned image plane quantization insufficiency. The exact location of point could not be calculated and one of the red points which are in region of uncertainty is calculated. Therefore the result of calculation is erroneous. Although we cannot find the systematic errors that are caused by, for instance, an improper camera calibration, the errors due to the stereo pair matching problem can be estimated as the difference between the measured values and the fitting line.

5.4 Results of Sub-pixel Matching

By application of sub-pixel matching we observed that region of uncertainty is reduced. The comparison of one-pixel level accuracy and sub-pixel level accuracy situations are illustrated to be able to clearly see the improvement in location measurement. In the figures the blue diamonds are results obtained by stereo pair matching in one-pixel accuracy, while the red squares are those by sub-pixel matching method [4]. The line fitting of the obtained data is also carried out to display the estimated true course of the ships.

One of the effective measure in the calculation of depth is disparity. The effect of sub-pixel registration to the disparity data is illustrated in addition to location figures. After sub-pixel matching the disparity data become more continiuous and reliable.

We will consider about four different situations to examine the effect of sub-pixel image matching. In Figure.5.4 we can see two ships in different size and different distance which are called Ship1 and Ship 2. The situations are as follows:

Situation 1: Ship 1 Before Akashi Bridge

Situation 2: Ship 1 After Akashi Bridge

Situation 3: Ship 2 Before Akashi Bridge

Situation 4: Ship 2 After Akashi Bridge

The foggy weather condition on the day of capturing images should be taken into consideration. If the weather was more clear, the accuracy of measurement would be much higher.

5.4.1 Situation 1: Ship 1 Before Akashi Bridge

Ship 1 is closer to the cameras (or has smaller range) and the image size (in pixels) is greater than Ship 2. That's why it is expected that the accuracy of the measured data of Ship 1 is better than Ship 2. In Figure 9 the discontinuity of one-pixel data is broken and continuous data is obtained by sub-pixel matching. The error is reduced due to closer data to the fitting line. We can say that great portion of the error was due to quantization problem.

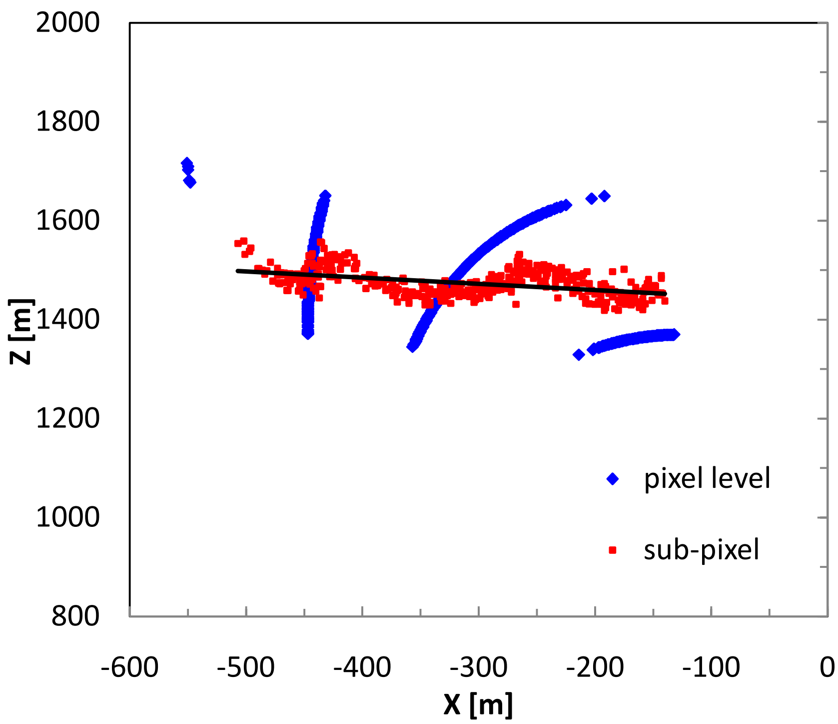

Figure.5.9 Effects of sub-pixel matching on range measurement: Situation 1

Despite the error is reduced, there is still some error, because it is impossible to completely remove the error caused by quantization problem. This error can be seen more clearly in Figure.5.10 which is the magnified image of Figure.5.9.

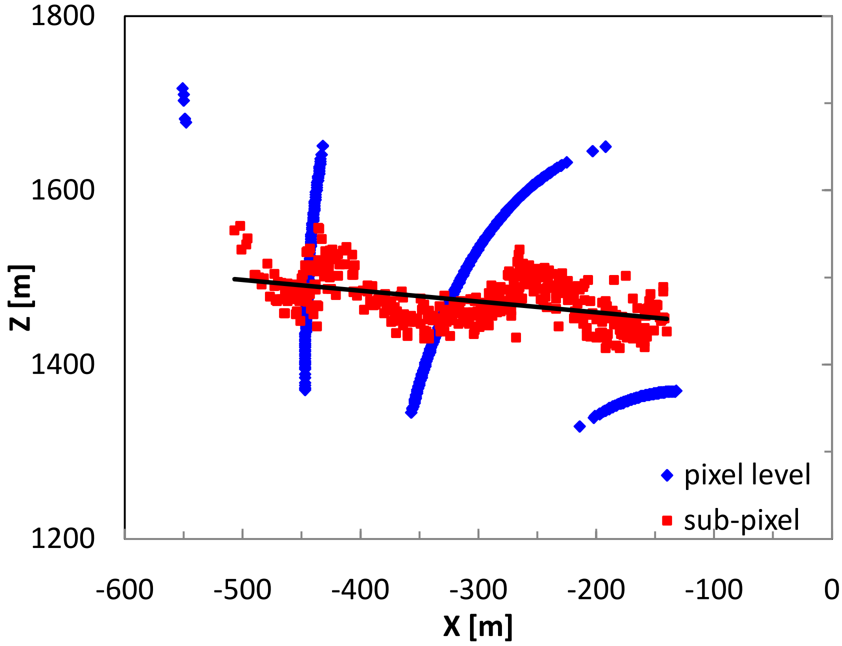

Figure.5.10 Magnified image of Figure.5.9

Figure.5.11 is displaying the effect of sub-pixel matching on disparity values. Disparity is inversely proportional to the range value. According to the figure we can say that the ship is moving almost parallel to the X axis of the cameras. After sub-pixel matching the sudden increase and decrase of the disparity values are removed.

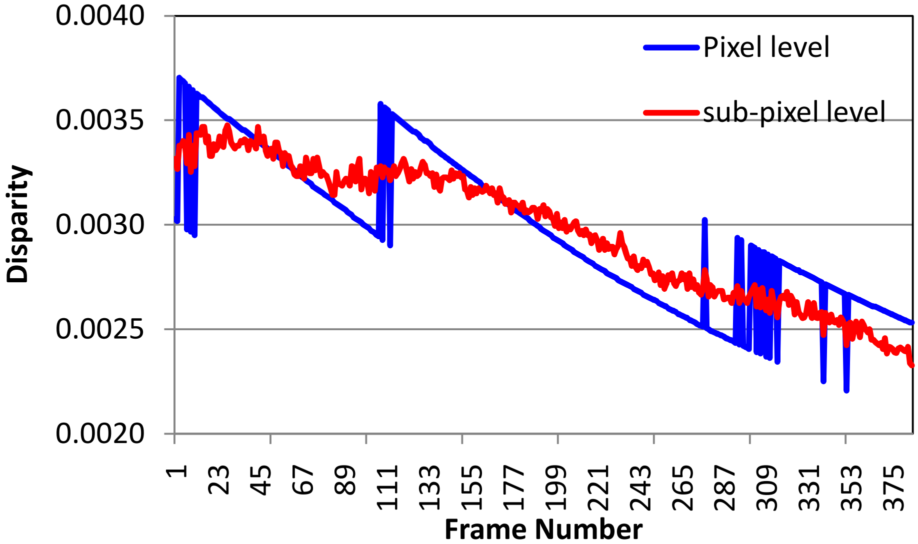

Figure.5.11 Effects of sub-pixel matching on disparity: Situation 1

5.4.2 Situation 2: Ship 1 After Akashi Bridge

The locational accuracy results of Situation 2 is very similar to Situation 1 as it can be seen in Figure.5.12.

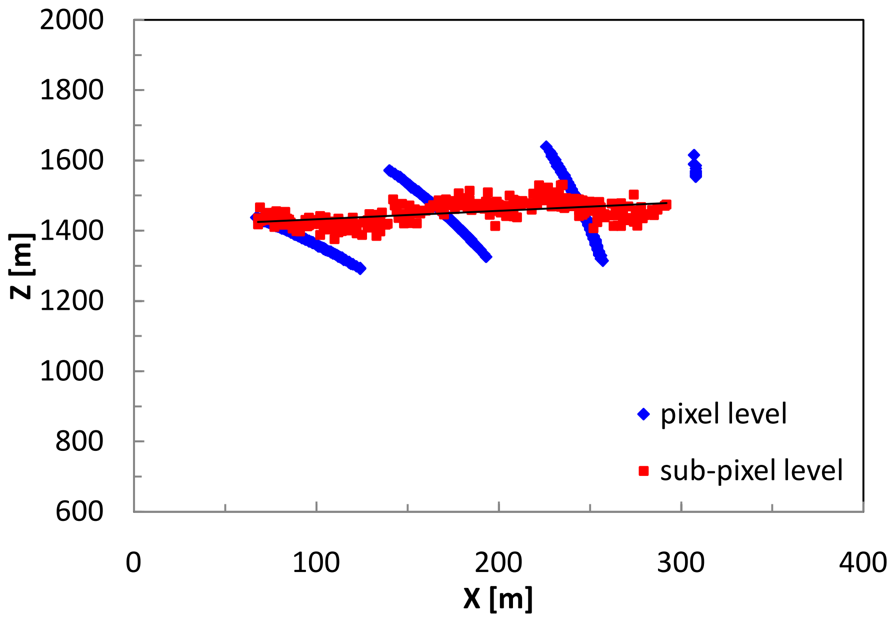

Figure.5.12 Effects of sub-pixel matching on range measurement: Situation 2

Checking the disparity values of Situation 2 in Figure.5.13 we see a larger disparity range comparing to Situation 1. It means that Ship 1 is slightly moving away from the X axis.

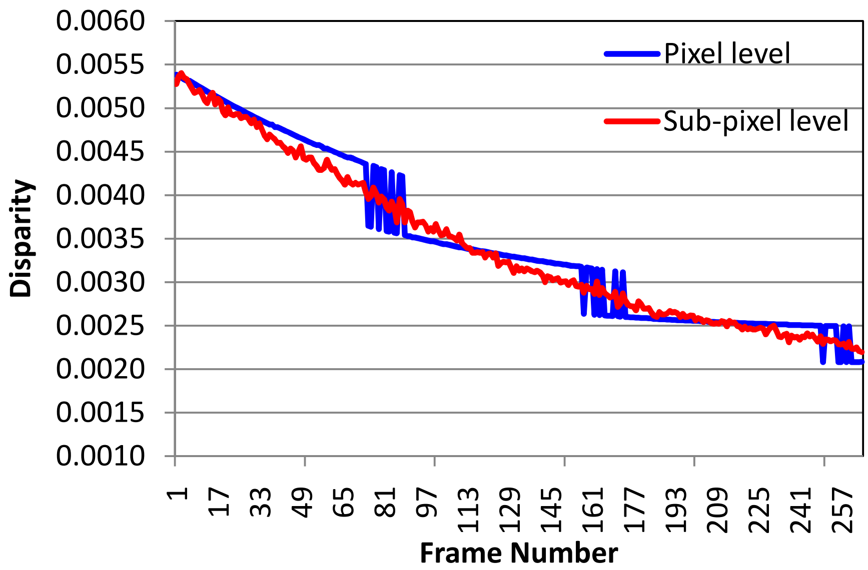

Figure.5.13 Effects of sub-pixel matching on disparity: Situation 2

5.4.3 Situation 3: Ship 2 Before Akashi Bridge

Ship 2 is smaller and more distant from the cameras. Adding the foggy weather condition the measurement is becoming more difficult. Figure.5.14 is describing the Situation 3. After -350 meters –getting closer the cameras- one-pixel matching and sub-pixel matching results are almost overlapping.

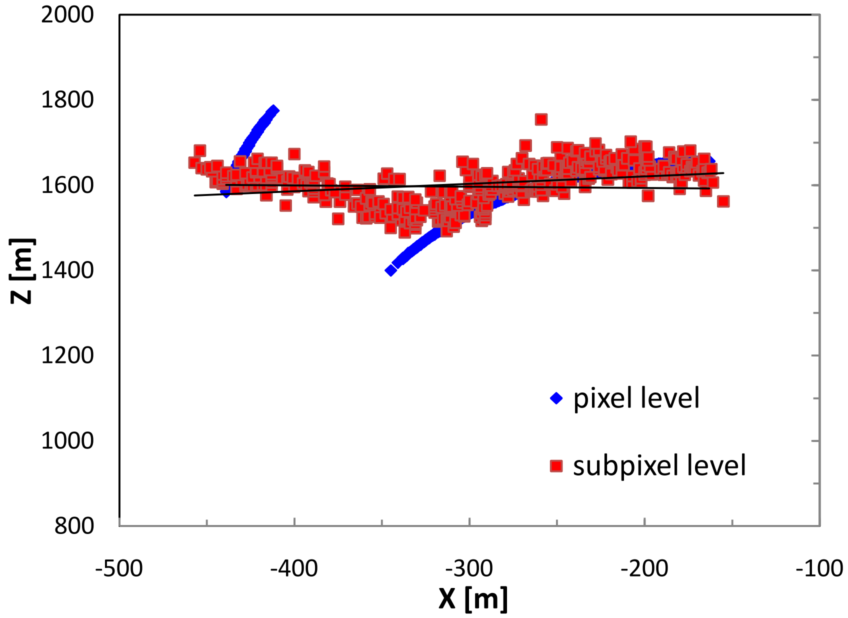

Figure.5.14 Effects of sub-pixel matching on range measurement: Situation 3

The disparity values have only one time jumping range from frame 61 to 101. After than is smoother than sub-pixel matching. This may be due to overlapping mentioned above.

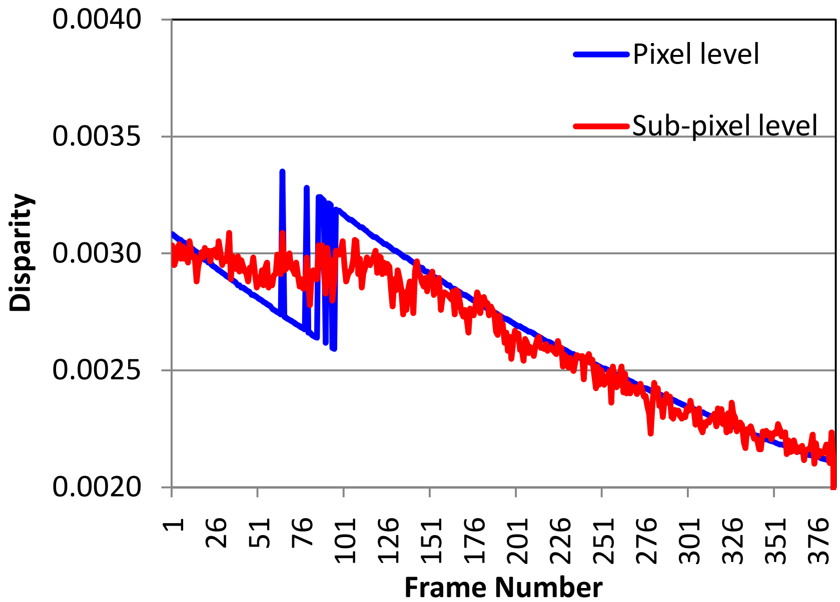

Figure.5.15 Effects of sub-pixel matching on disparity: Situation 3

5.4.4 Situation 4: Ship 2 After Akashi Bridge

The characteristics of Situation 4 is different than the privious three situations. Until the 100 meters there are many outliers observed. These outliers cause the line fitting be much different than the real case. Therefore, error measurement is also not reliable in this case. The main reason of the outliers is the error in the corresponding point finding. This is shown in Figure.5.18.

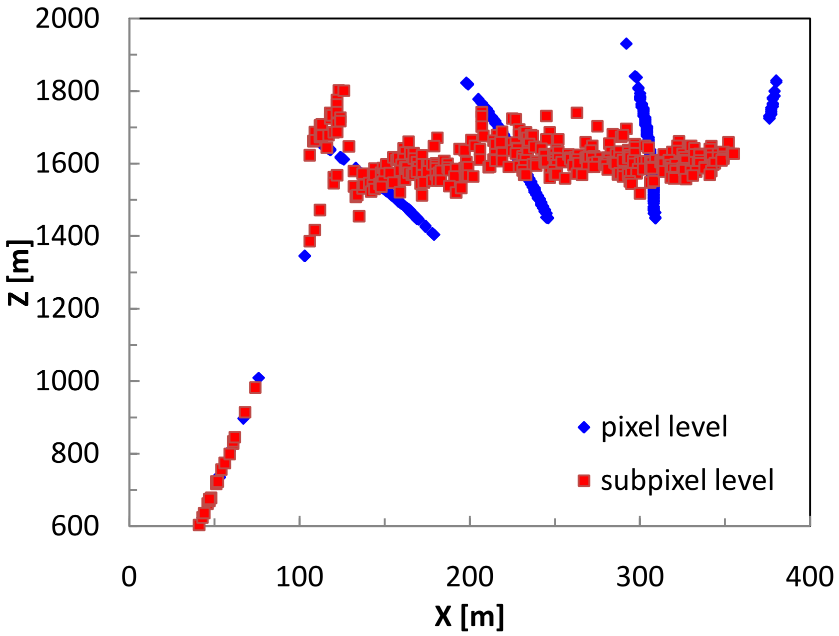

Figure.5.16 Effects of sub-pixel matching on range measurement: Situation 4

Due to corresponding point error, the disparity values are also abnormal as shown at Figure.5.17 . First 90 frames of disparity values have very strange characteristic.

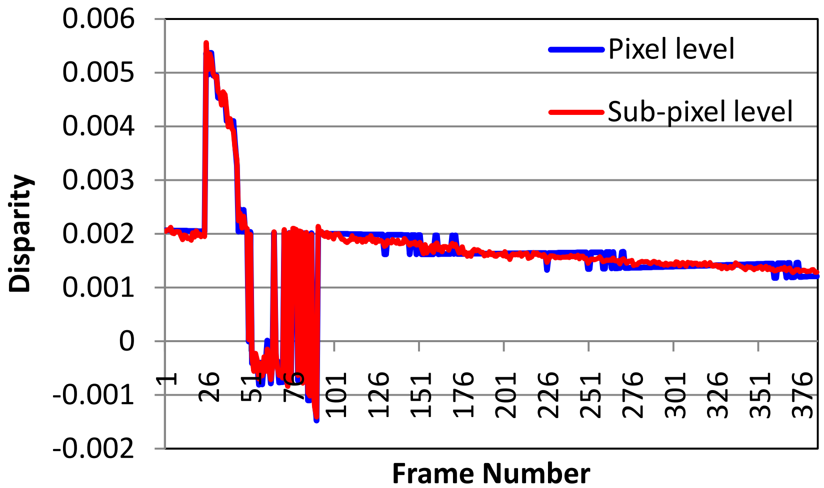

Figure.5.17 Effects of sub-pixel matching on disparity: Situation 4

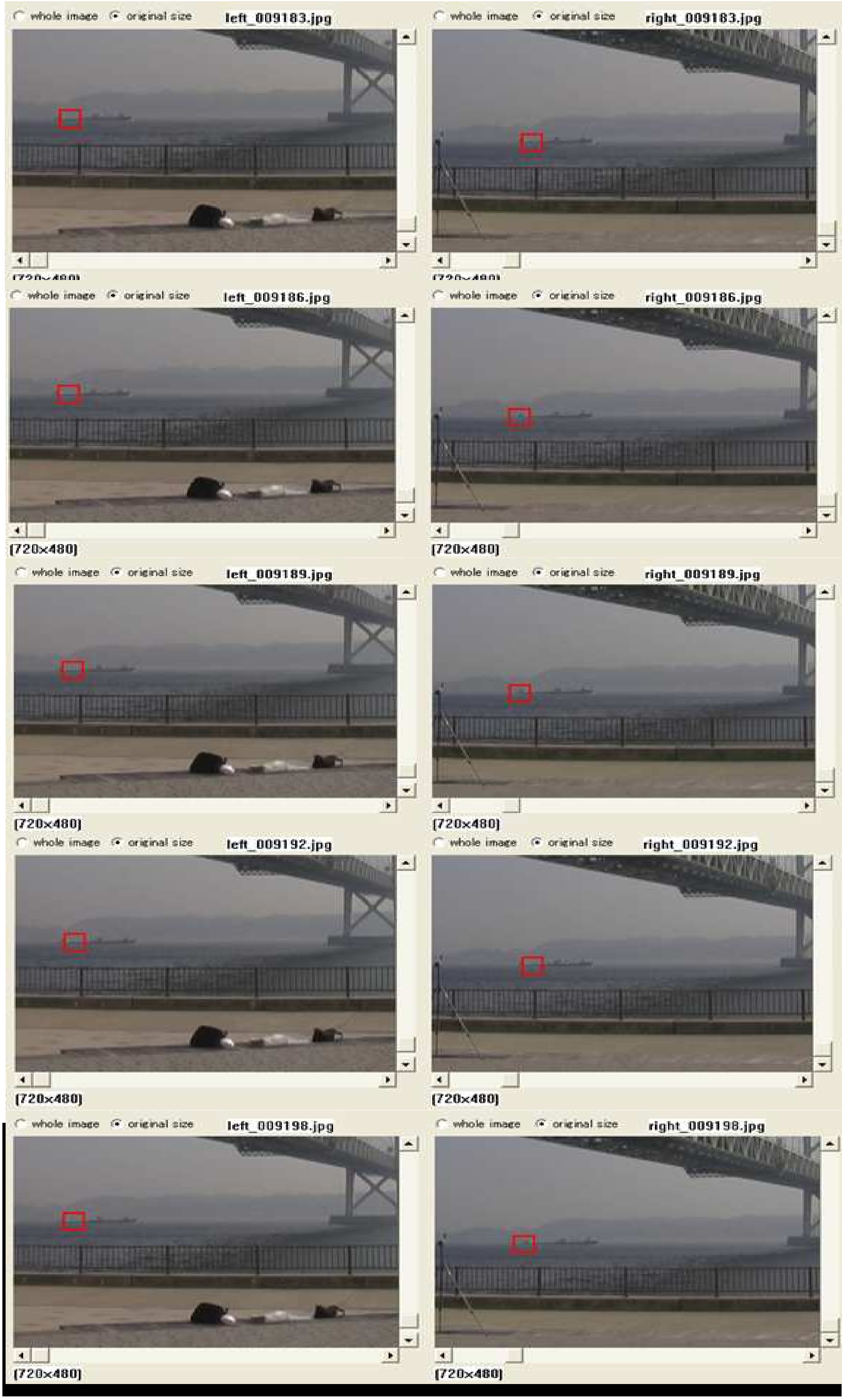

Figure.5.18 Correlation window jump

Finding the corresponding point in the right image away from the actual one causes significant calculation error. The images in Figure.5.18 are sequential frames. While the ship is moving the ship is tracked in these sequences and correlation is carried out every time. However, sometimes the calculation is not true due to image noise and specular characteristic of sea surface. And the calculated corresponding point is not true as shown in the figure. This error causes the disparity and location values to be very different from actual values and behave as outliers. This problem can be solved by outlier elimination algorithms or improving the matching cost.

As a result, the errors of one-pixel unit matching range from about −170 meters to about 200 meters. The values will theoretically become from -140 meters to 170 meters using equation (4.5) and (4.6) when δ ranges from −0.5 pixels to 0.5 pixels, and this almost agrees with the experimental result. On the other hand, the errors of sub-pixel unit matching are within ±50 meters, i.e. 0.014 as an error rate of objects distance. It is found that stereo matching in sub-pixel unit is at least 3 times more accurate than one-pixel unit. It means that about 70% of the error was caused by quantization error.

References

[1] S.D. Blostein, and T.S. Huang, "Error analysis in stereo determination of 3-D point positions", *IEEE Trans. On Pattern Analysis and Mach. Intell.* Vol. PAMI-9, No.6, 1987.

[2] Q. Tian, "Sub-pixel Image Registration", Phd Study at University of South Carolina,1984.

[3] M. Shimizu and M. Okutomi, "Precise Sub-pixel Estimation on Area-Based Matching", Proc. 8th IEEE International Conference on Computer Vision (ICCV2001), pp.90-97, Vancouver, Canada, July 2001.

[4] M. Shimizu, and M. Okutomi, "Sub-pixel estimation error cancellation on area-based matching", *International Journal of Computer Vision* 63(3), 207-224, 2005.

CHAPTER 6

REFINEMENT OF DISPARITY ESTIMATES WITH LOW PASS FILTERING

Stereo reconstruction of dense depths from real images has long been a fundamental problem in computer vision. The reconstructed depths can be used by a wide spectrum of applications including 3D modeling, robot navigation, image-based rendering, and video editing. Although stereo problem has been extensively studied during the past decades, obtaining high-quality dense depth data is still a challenging problem due to many inherent difficulties, such as image noise, texture less pixels, and occlusions [1].

Two main constraints to be able to calculate disparity are similar intensity and smoothness. In other words if there is not occlusion, disparity map should vary as slowly as possible [2]. Correspondence between stereo image pairs provides low frequency information [3]. It means that if the measured disparity values have vibration, i.e. disparity values are not smooth, there exist a noise which means high frequency information. Removing these noise will result in a more reliable disparity estimation. By applying low pass filter (LPF), variable distance reconstruction of dynamic object is feasible despite restrictions in disparity variance.

We preferred to use a 2nd order Butterworth low pass filter (BLPF) for eliminating the noisy data from the measured disparity values. The response of Butterworth LPF can be seen at Figure.6.1.

The transfer function of a BLPF is defined as:

$$|G(w)|^2 = \frac{1}{1+\left[\frac{\omega}{\omega_c}\right]^{2n}} , \qquad (6.1)$$

where ω_c is cut off angular velocity, and n is the order of the filter [4].

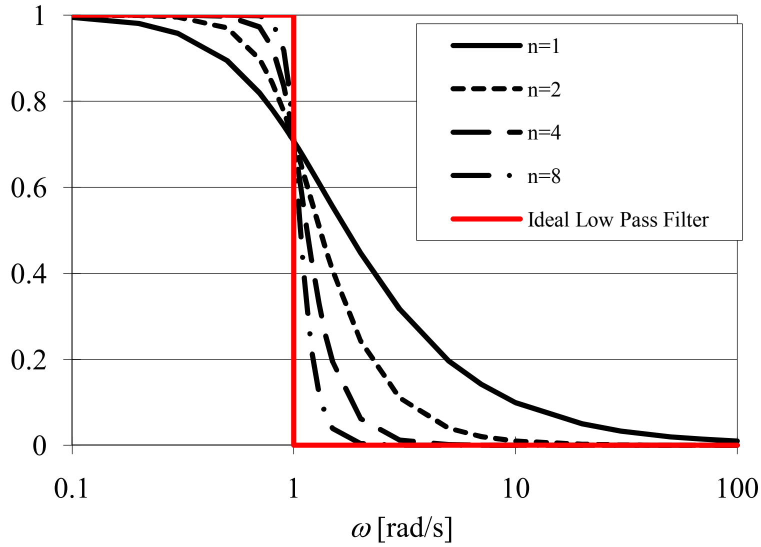

Figure.6.1 Butterworth low pass filter response (ω_c=1)

6.1 Results of Low Pass Filtering

We will examine the effect of BLPF on location measurement and disparity values in four situations as done for sub-pixel matching. The situations are as follows:

Situation 1: Ship 1 Before Akashi Bridge

Situation 2: Ship 1 After Akashi Bridge

Situation 3: Ship 2 Before Akashi Bridge

Situation 4: Ship 2 After Akashi Bridge

These situations are illustrated using the calculated data. In the figures pixel level, sub-pixel level and BLPF accuracy results are compared. 3D location and disparity values in different situations are illustrated to show the step by step improvement through pixel-level matching, sub-pixel level matching and finally BLPF application. After application of BLPF the accuracy of location and disparity values are highly improved.

6.1.1 Situation 1: Ship 1 Before Akashi Bridge

Figure.6.2 shows the improvement of calculation in 3 steps. After BLPF application the result is seen almost overlapping with the fitting line and the error could be decreased more.

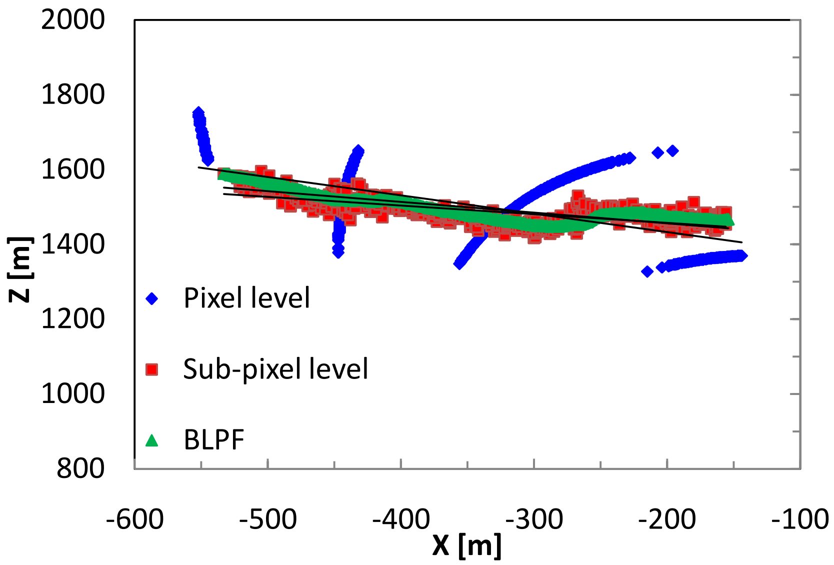

Figure.6.2 Effects of BLPF on range measurement: Situation 1

Looking at the Figure.6.3 which is zoomed form of Figure.6.2 we can see the situation more clearly and understand the characteristic of error.

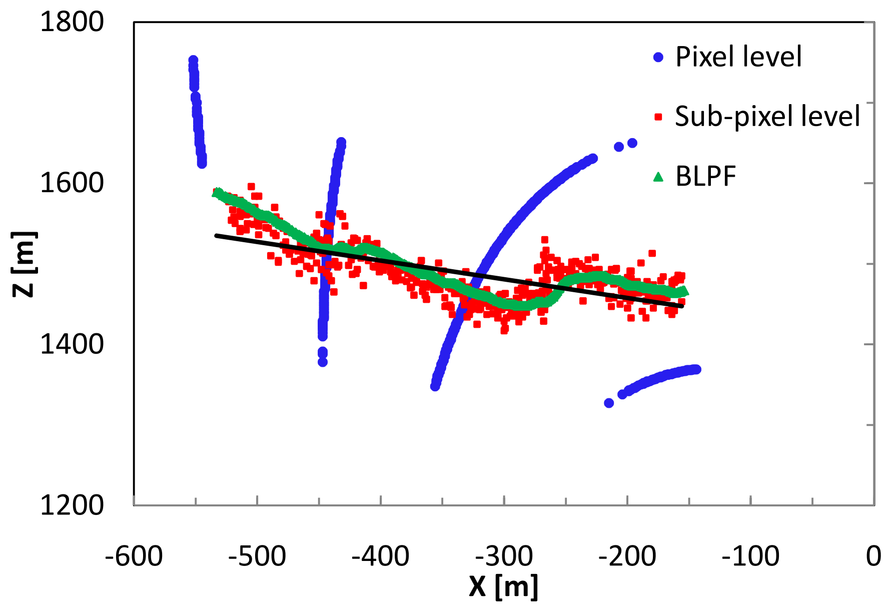

Figure.6.3 Magnified image of Figure.6.2

Figure.6.4 shows the same improvement for the disparity values. The vibration of the disparity values are eliminated and smoother disparity values obtained by BLPF refinement.

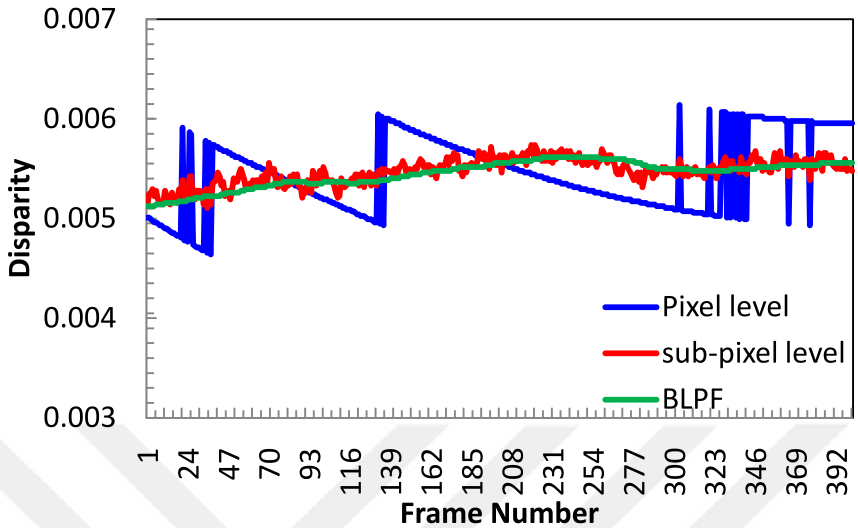

Figure.6.4 Effects of BLPF on disparity: Situation 1

6.1.2 Situation 2: Ship 1 After Akashi Bridge

Situation 2 is very similar to Situation 1. Only the location is different and there is not a problem of mismatching.

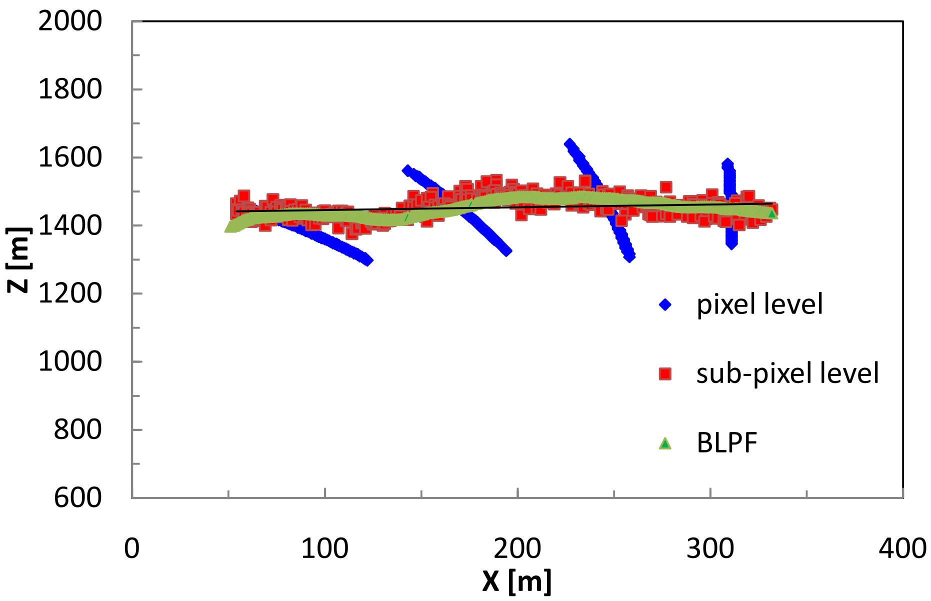

Figure.6.5 Effects of BLPF on range measurement: Situation 2

The disparity values of Situation 2 is almost same with the Situation 1 which means that the ship is moving parallel to X axis before bridge and after bridge.

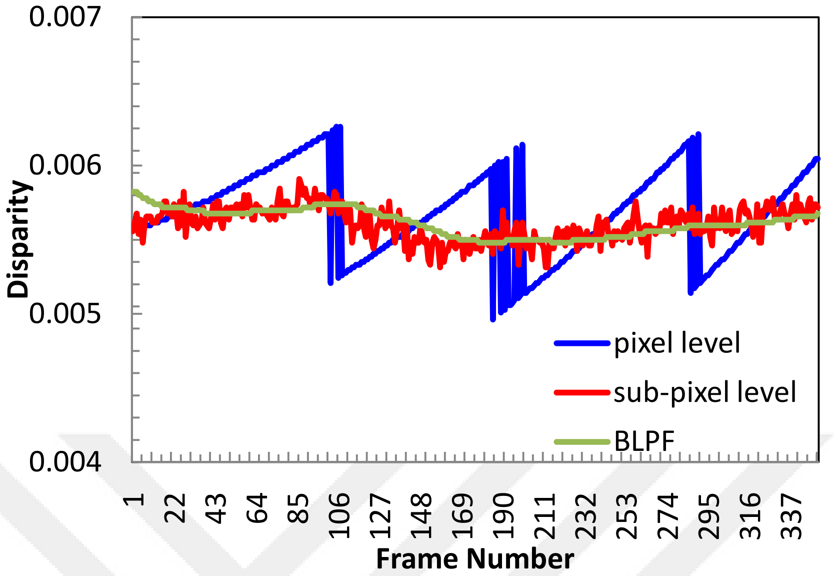

Figure.6.6 Effects of BLPF on disparity: Situation 2

6.1.3 Situation 3: Ship 2 Before Akashi Bridge

As the ship is distant from cameras and ship size is smaller, the possibility of wrong calculation is increasing. In Figure.6.7 we observe some error due to mismatcing. Even one outlier caused by mismatching have a great effect on the location and disparity data.

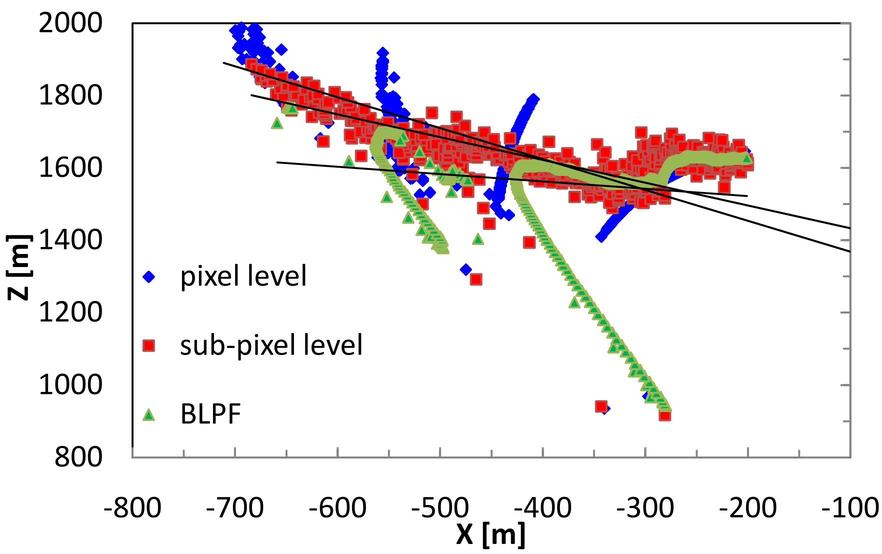

Figure.6.7 Effects of BLPF on range measurement: Situation 3

In Figure.6.8 the effect of BLPF on disparity is shown. We can see the effect of outliers to the disparity data.

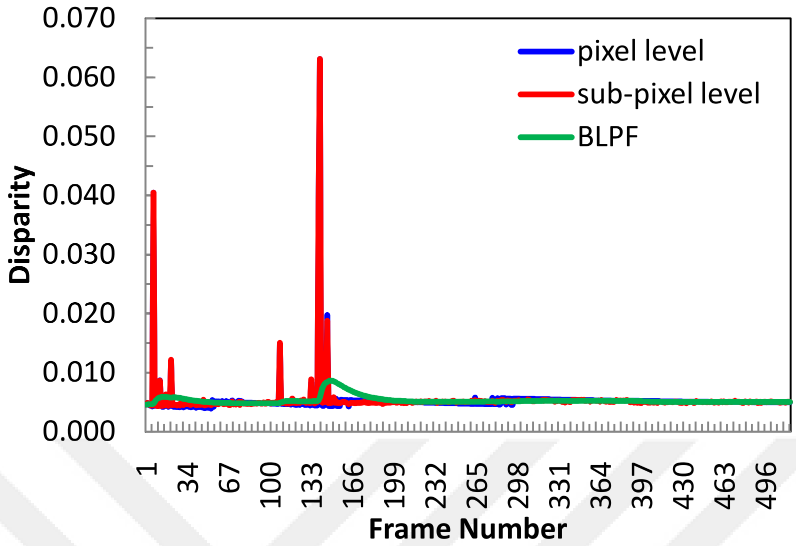

Figure.6.8 Effects of BLPF on disparity: Situation 3

6.1.4 Situation 4: Ship 2 After Akashi Bridge

The calculations of Situation 4 are quite good as Situation 1 and Situation 2. There is no outlier and refined values are almost overlapping the estimated true track of the ship as in Figure.6.9.

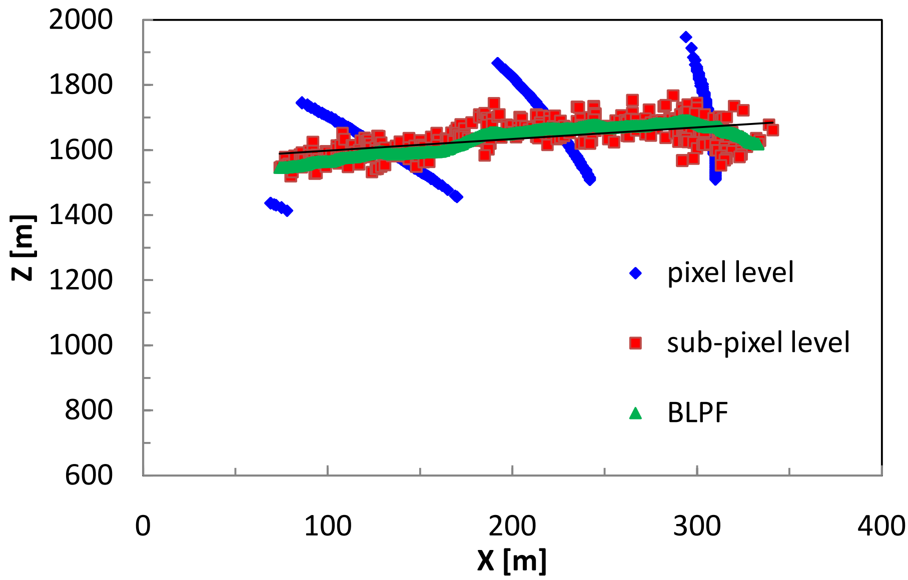

Figure.6.9 Effects of BLPF on range measurement: Situation 4

Refined disparity values are improved as expected in Figure.6.10.

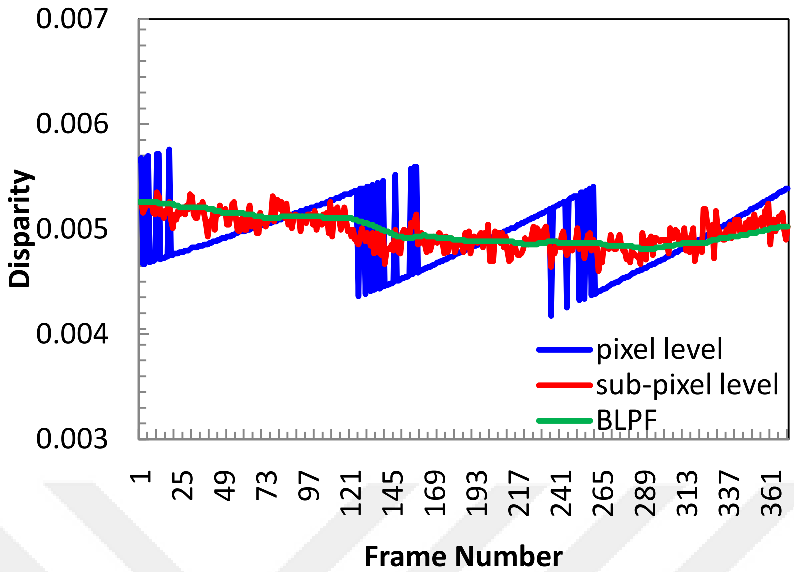

Figure.6.10 Effects of BLPF on disparity: Situation 4

As mentioned before, the correspondence data provides low frequency data. After removing the high frequency noise from this data by Butterworth LPF we obtain more smooth disparity values which results in better 3-D reconstruction values. In figures the effect of BLPF can be seen clearly. After application of BLPF the measured range error reduced to ±35 meters which is about 6 times smaller than one pixel level measurement. The rate of this error is reduced from 0.014 to 0.011 comparing to sub-pixel matching error of objects distance. When we crosscheck with the pixel level error, 77% of error is caused by quantization error due to mismatching of correesponding points. This means that 7% of the error is eliminated through disparity refinement by BLPF.

References

[1] G. Zhang, J. Jia, T. Wong, and H. Bao, “Recovering consistent video depth maps via bundle adjustment”, *IEEE Conference on Computer Vision and Pattern Recognition*, Page(s):1 – 8, 2008.

[2] S.T. Barnard, “Stochastic Stereo Matching over Scale”, *International Journal of Computer Vision* ,3, 17-32,1989.

[3] J.E. Cryer, P. Tsai, and M. Shah, “Integration of shape from shaping and stereo”, *Pattern Recognition* Vol.28, No.7, pp.1033-1043, 1995.

[4] Gonzalez, R.C. and Woods, R.E. 2008, Digital Image Processing, 3rd ed, Prentice Hall

CHAPTER 7

EVALUATION OF ERROR IN VARIOUS SITUATIONS

There are various stereo camera systems and configurations. One of the taxonomies of these configurations is wide-baseline and short-baseline stereo systems. Computer vision studies are mostly focused on short-baseline stereo studies and the number of wide-baseline stereo studies are limited. Even more the studies for evaluation of the stereo systems are usually theory based and experimental verification is done in a few studies. Some accuracy results of short-baseline stereo studies are as follows. The study expalined in reference [1] achieved the accuracy of ±5% at the range of 2.5 feet (0.76 meters). The stereo system constructed by Faugeras et al. [2] has an accracy of ±0.05 % at 2 meters range. Another study based on wide angle of view can achieve ±2.8 % of accuracy in 0.2 meters [3]. The range values of these studies are exteremely small for the case of ship detection. Our study is based on wide-baseline stereo system which is capable of measuring very long range (about 6000 meters) with small error (about 1%) comparing the above mentioned studies.

In our study the equation (4.6) for calculating the error is based on the error detected on the image plane. Therefore, fistly we detect the error of disparity on the image plane (in the unit of camera coordinates) and we increased the accuracy of image plane as much as possible. And then we will observe the effect of improvement in disparity data on 3D range measurement. Our accuracy improvement is carried out in two steps of sub-pixel matching and BLPF. Therefore we will consider about three situations:

1. One-pixel accuracy (basic study)
2. Sub-pixel accuracy
3. BLPF disparity refinement

The image plane error of the one-pixel accuracy is limited with ±0.5 pixels. After application of sub-pixel matching this error is reduced to ±0.15 pixels. Finally by application of BLPF disparity refinement the error is reduced to ±0.11 pixels on image plane.

Using the equations (4.5) and (4.6), error values of some different situations can be obtained. The values used in our study are N = 720 and θ = 43. δ is decreasing by the accuracy is increasing. For one-pixel accuracy δ = 0.5 pixel, for sub-pixel accuracy δ = 0.15 pixel and for BLPF refinement δ = 0.11 pixel values are feasible error values in image plane coordinates.

In Table.7.1 and Table.7.2 ε values of different view angle for changing error value (δ) are listed.

Table.7.1 ε values for N=2048

θ (degree)	δ = 0.5	δ = 0.15	δ = 0.11
20	0.0000861	0.0000258	0.0000189
30	0.0001307	0.0000392	0.0000287
43	0.0001922	0.0000577	0.0000423
50	0.0002275	0.0000682	0.0000500
60	0.0002818	0.0000846	0.0000620
70	0.0003416	0.0001025	0.0000752
90	0.0004879	0.0001464	0.0001073

Table.7.2 ε values for N= 720 (used in this study)

θ (degree)	δ = 0.5	δ = 0.15	δ = 0.11
20	0.0002449	0.0000735	0.0000539
30	0.0003717	0.0001115	0.0000818
43	0.0005467	0.0001640	0.0001203
50	0.0006471	0.0001941	0.0001424
60	0.0008017	0.0002405	0.0001764
70	0.0009718	0.0002915	0.0002138
90	0.0013878	0.0004163	0.0003053

7.1 Evaluation of Maximum Measurable Range

Considering the calculated ε values we can get the maximum allowable range up to the tolerable error value in range data. In Figure.7.1 maximum measurable range for different base length and different angle of view, in the case of accepting the range error of 150 meters is illustrated. As the base length is increased the measurable range is increasing, too. In this figure when the θ = 20 and LB = 12 m the measurable range is 9759 meters. Besides when the θ = 43 and LB = 8.14 the measurable range is 5372 meters with tolerating the error of 150 meters.

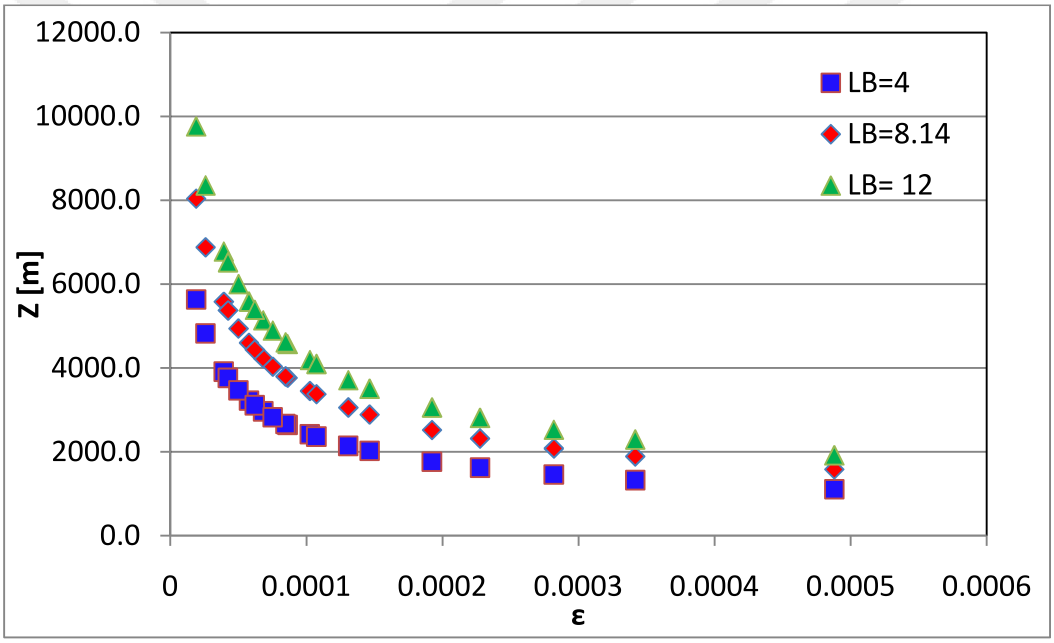

Figure.7.1 Measurable maximum range for N=2048, Ez=150m

Figure.7.2 is showing the maximum range values when N=2048, Ez=150m, LB = 8.14, and θ = 43 in the cases of one pixel level, sub-pixel level and BLPF refinement. By the accuracy is increased -or the δ value is decreased- maximum measurable range has a significant increase.

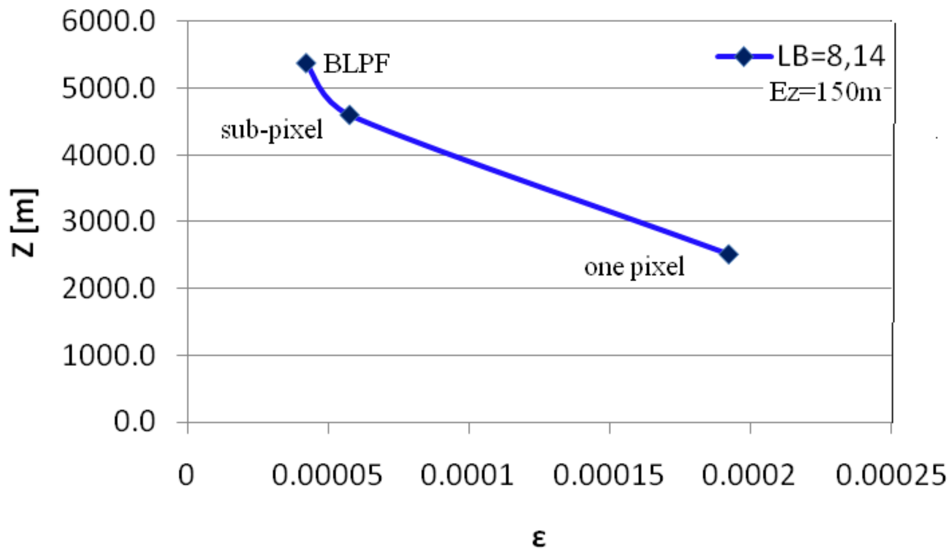

Figure.7.2 Maximum range, N=2048, Ez=150m, LB = 8.14, θ = 43

In Figure.7.3 the case that the range error is 50 meters is evaluated for different angle of view and base length. As the base length is increased and the angle of view is decreased the measurable range is increasing.

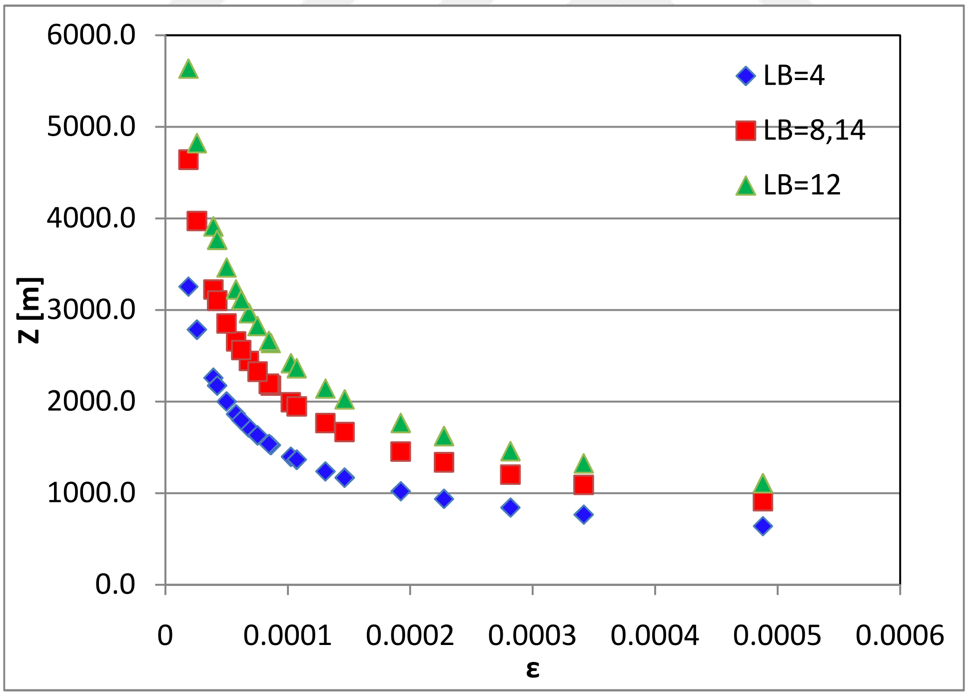

Figure.7.3 Measurable maximum range for N=2048, Ez=50m

In Figure.7.4 we observe that when the range error is decreased to 50 meters, measurable range is decreased from 5372 meters to 3101 meters in the case of BLPF.

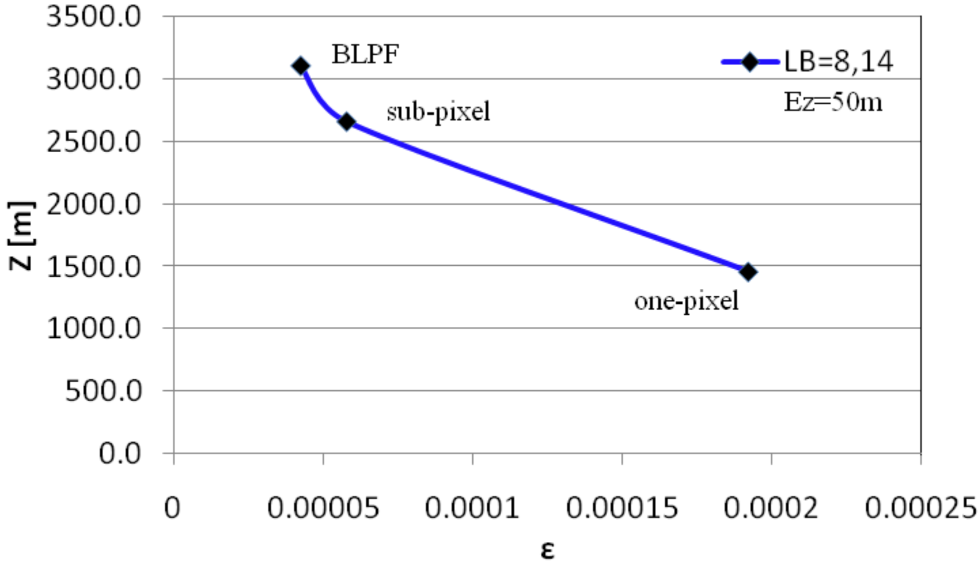

Figure.7.4 Maximum range, N=2048, Ez=50m, LB = 8.14, θ = 43

Figure.7.5 is illustrating the situation when the range error is desired to be maximum 35 m.

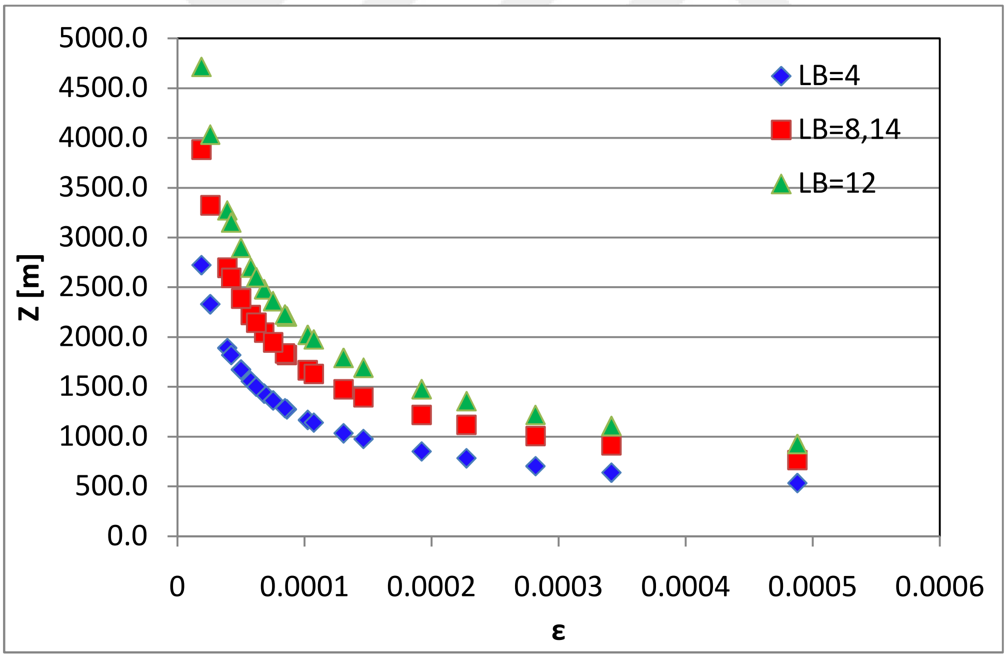

Figure.7.5 Measurable maximum range for N=2048, Ez=35m

In this case the measurable range decreases to 2595 meters which can be seen in Figure.7.6. Besides for one-pixel image plane accuracy this value is 1217 meters. Therefore by sub-pixel matching and BLPF range is increased from 1217 m. to 2595 m. for N=2048 pixels.

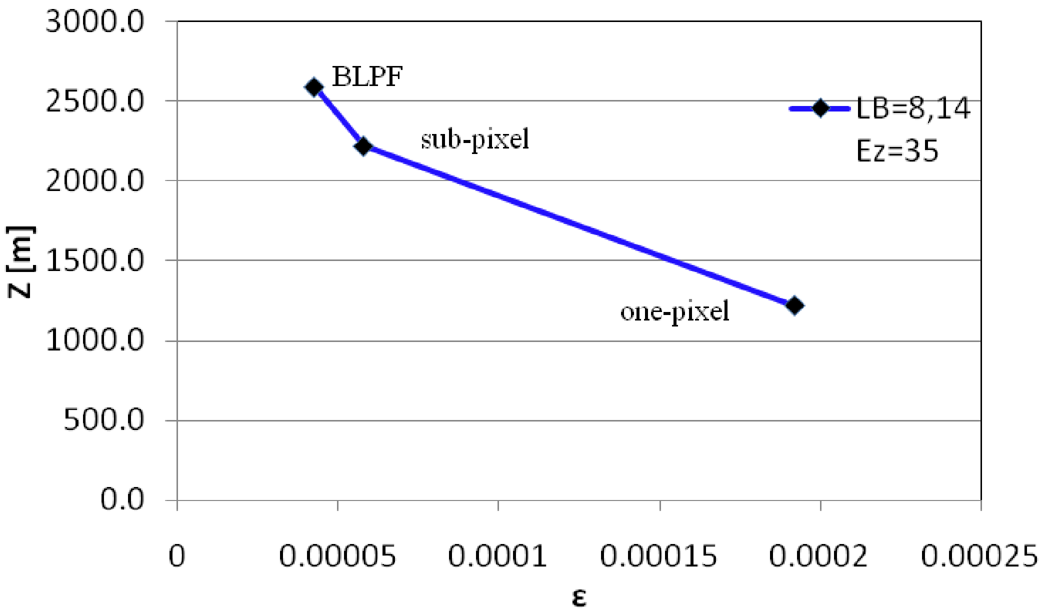

Figure.7.6 Maximum range, N=2048, Ez=35m, LB = 8.14, θ = 43

Despite the cameras used in this experiments have 2048 pixels of horizontal resolution in photo-shot mode we use video mode of 720 pixels. The following figures are for N=720 pixels. Comparing to N=2048, the calculated range values when N=720 are worse, but the calculation speed is faster which is very important in real-time applications.

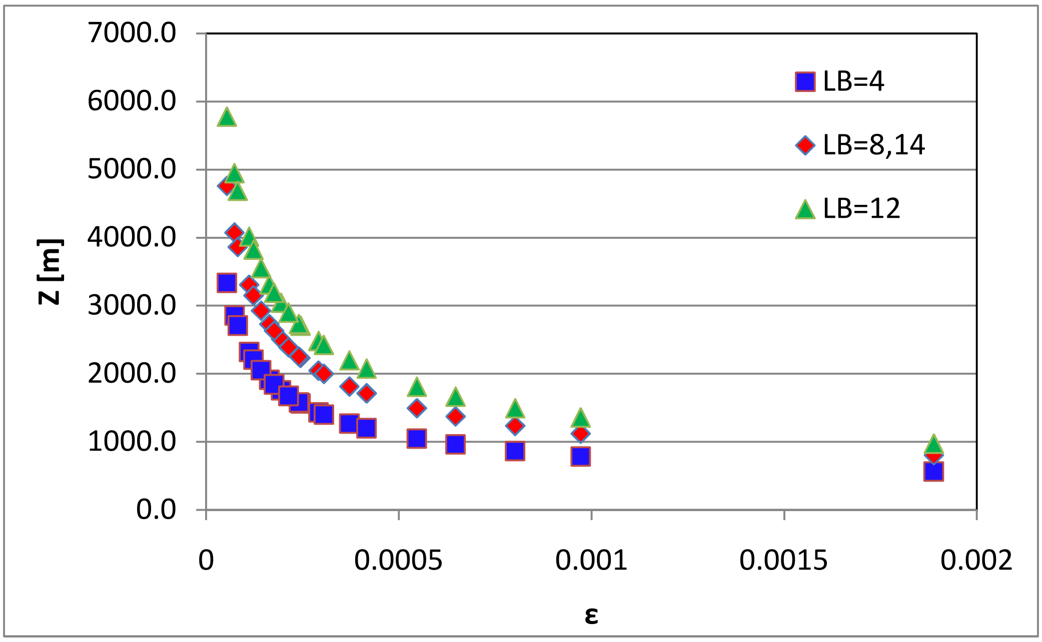

Figure.7.7 Measurable maximum range for N=720, Ez=150m

Figure.7.7 shows the maximum range for accepting the error is 150 meters in different base length. When the view angle is 20 degrees the maximum range is 5779 meters, however this view angle is too small. Our cameras have 43 degrees view angle which corresponds to 3150 meters of maximum range. Figure.7.8, Figure.7.10 and Figure.7.12 illustrates the situation of original values our experiments in the cases of different range accuracies. Figure.7.8 is for 150 meters range error case. Here, the measurable range is 3150 meters which was 5372 meters when N=2048.

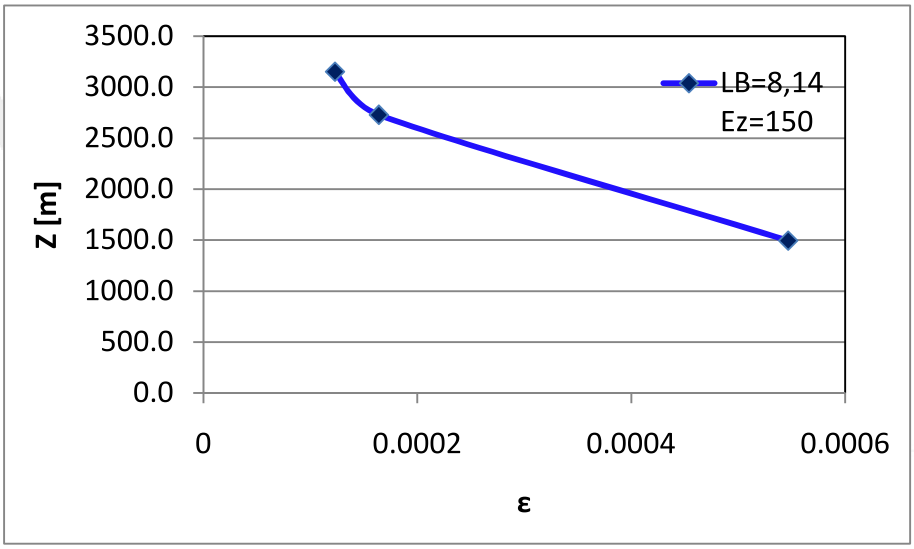

Figure.7.8 Maximum range, N=720, Ez=150m, LB = 8.14, θ = 43

In Figure.7.9 and Figure.7.11 the calculation when the range error of 50 meters and 35 meters are figured out respectively. Various situations of different view angle and base length data is considered. In Figure.7.9 when the view angle is 20 degrees 3336 meters can be measured while this value is 2791 meters in Figure.7.11. According to the original parameters of our experimental study, the maximum range is 1819 meters with error of 50 meters as shown in Figure.7.10. This value is decreased to 1521 meters when we desire lower error value of 35 meters. This can be seen in Figure.7.11.

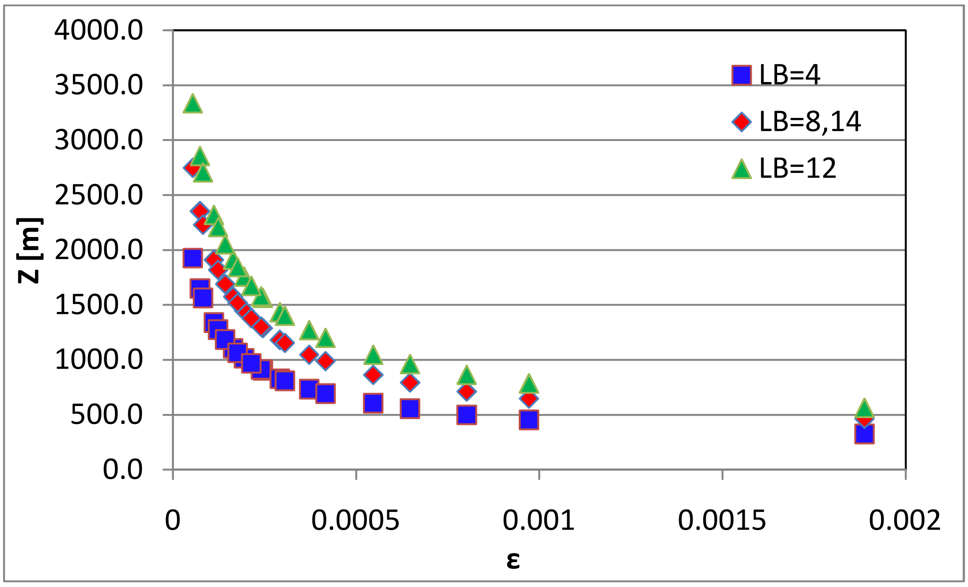

Figure.7.9 Measurable maximum range for N=720, Ez=50m

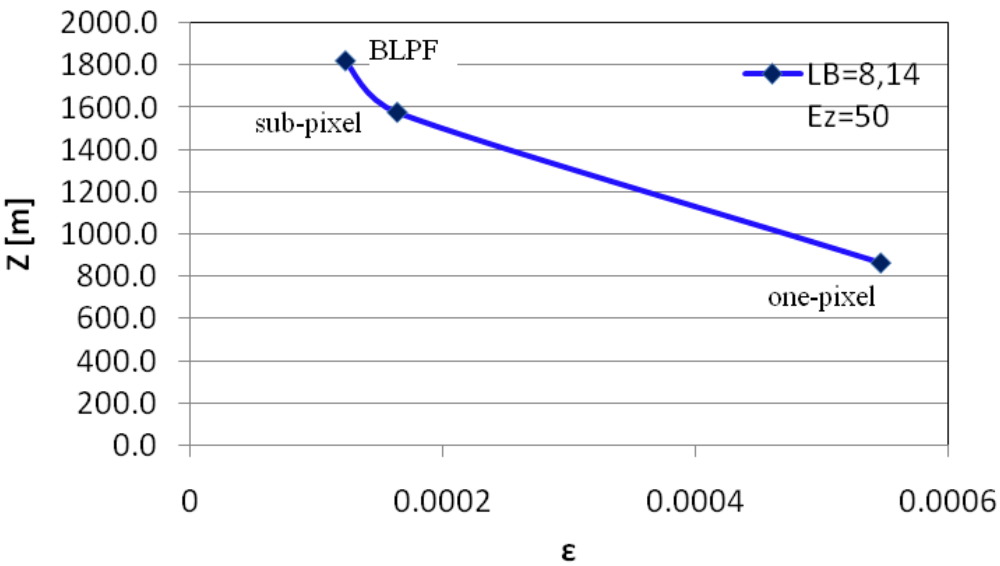

Figure.7.10 Maximum range, N=720, Ez=50m, LB = 8.14, θ = 43

In Figure.7.11 when the image plane accuracy is 1 pixel, the maximum range is 721 meters. By sub-pixel matching and BLPF refinement the maximum range is increased more than two times.

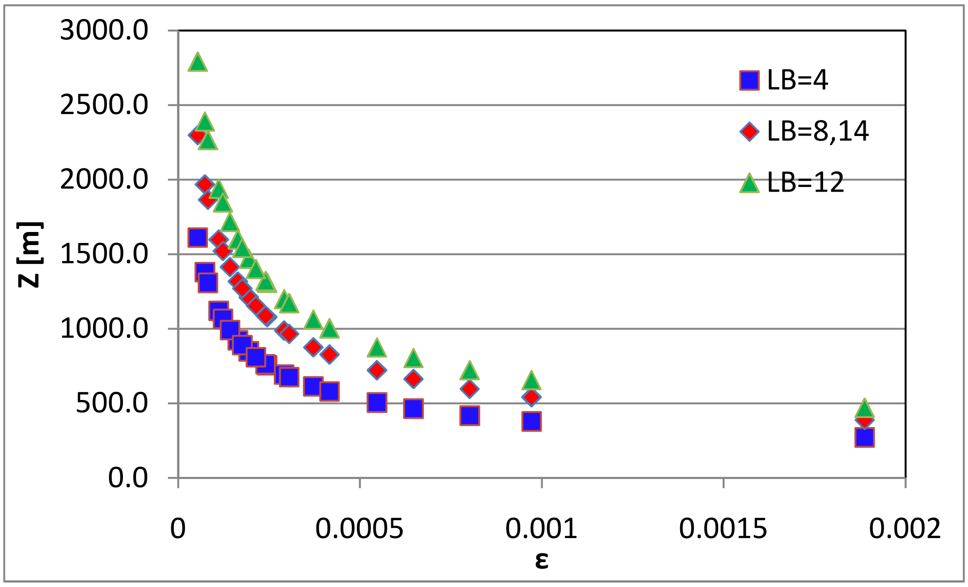

Figure.7.11 Measurable maximum range for N=720, Ez=35m

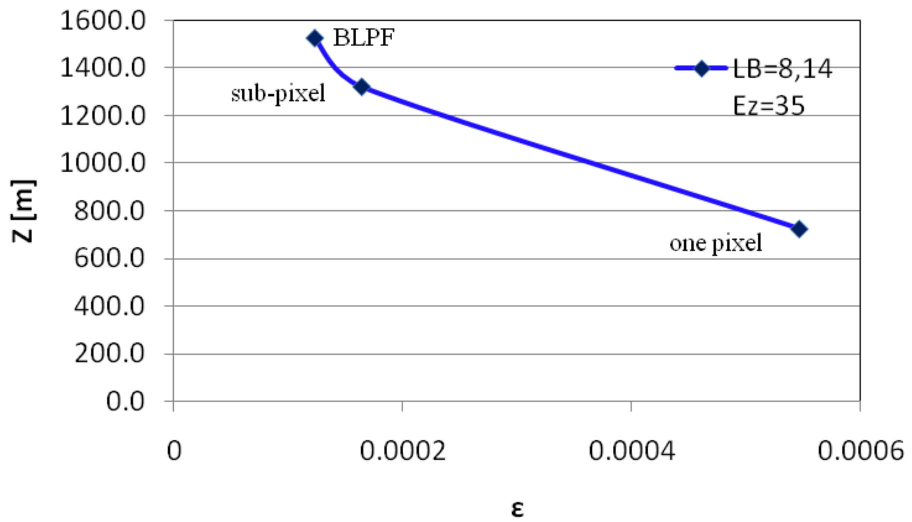

Figure.7.12 Maximum range, N=720, Ez=35m, LB = 8.14, θ = 43

The maximum measurable range has a very large range of values up to the different values of experimental parameters. While the maximum range is 9759 m. with N=2048, LB=12 m, θ= 20° and Ez is 150 m., it decreases to 1522 m. with N=720, LB=8.14, θ=43° and Ez=35 m.

7.2 Evaluation of Range Error

The figures above were showing the relation of ε and Z to see the maximum measurable range in different situations with error range of 150 m., 50 m., and 35 m.. Now we will show the occurred range error (Ez) up to the different range measurements. The relation of range error (Ez) and ε for different view angle, base length and range are showed through Figure.7.13 to Figure.7.18.

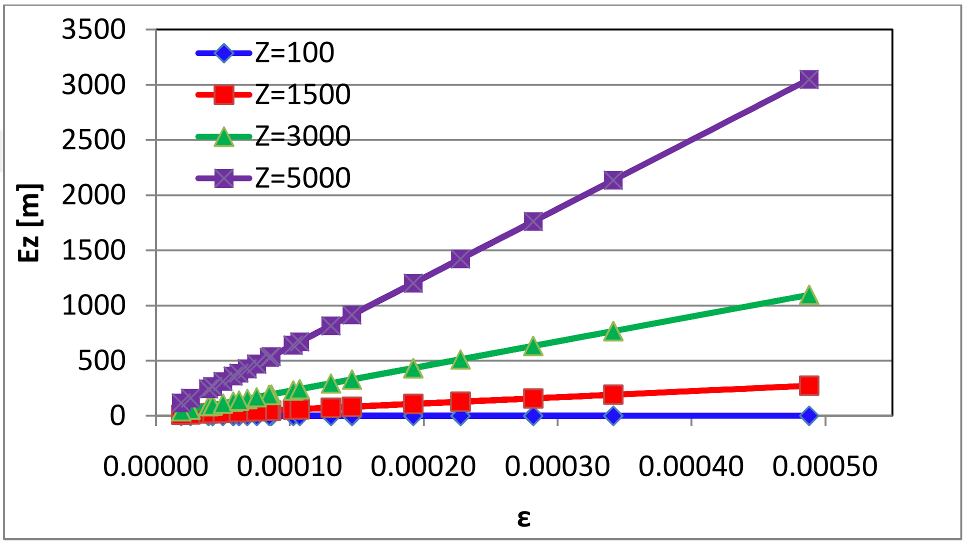

Figure.7.13 Relation of range error (Ez) and ε, N=2048, LB=4

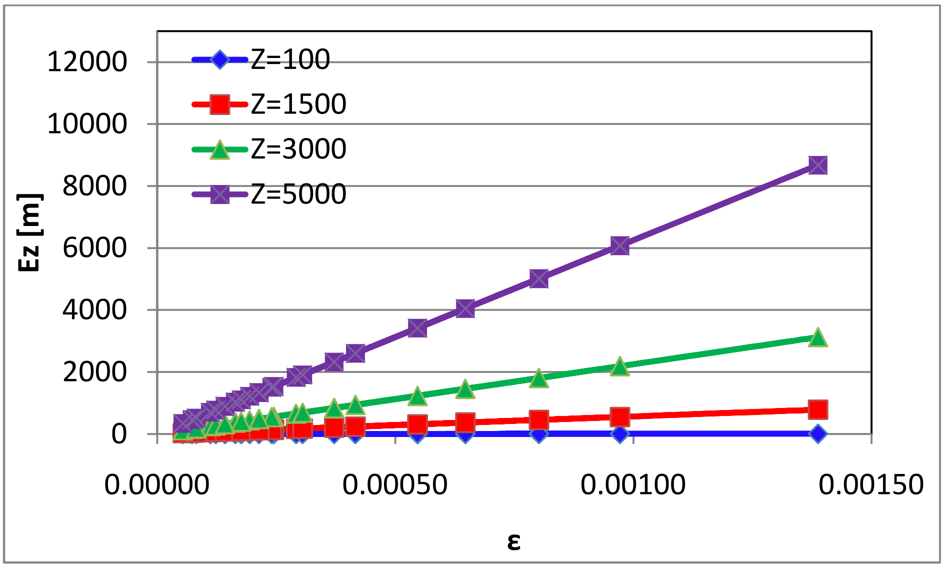

Figure.7.14 Relation of range error (Ez) and ε, N=720, LB=4

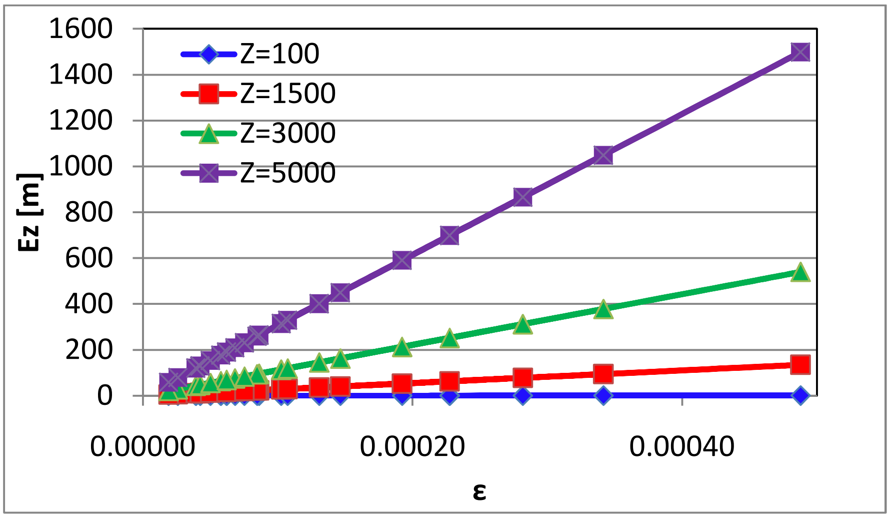

Figure.7.15 Relation of range error (Ez) and ε, N=2048, LB=8.14

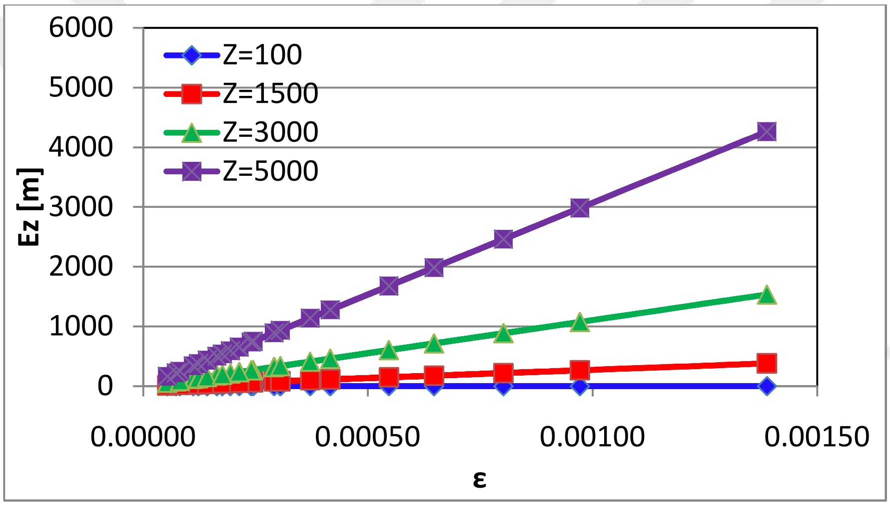

Figure.7.16 Relation of range error (Ez) and ε, N=720, LB=8.14

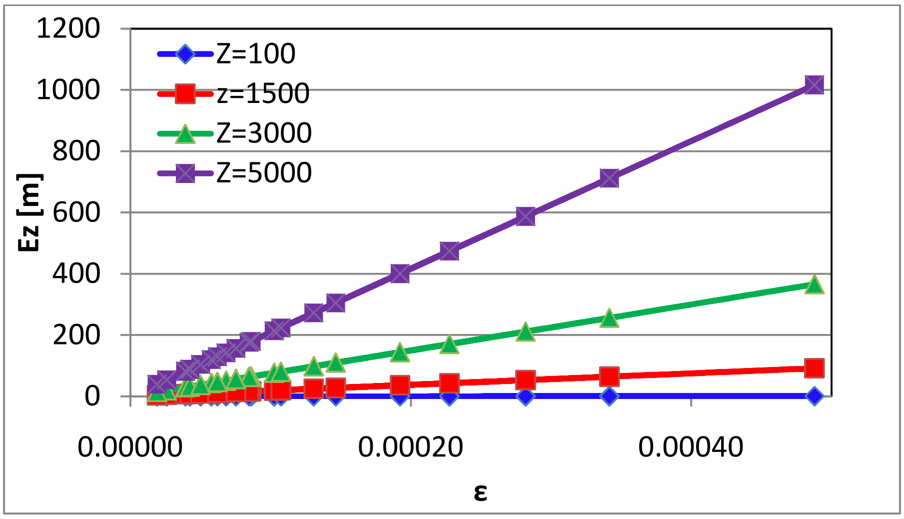

Figure.7.17 Relation of range error (Ez) and ε, N=2048, LB=12

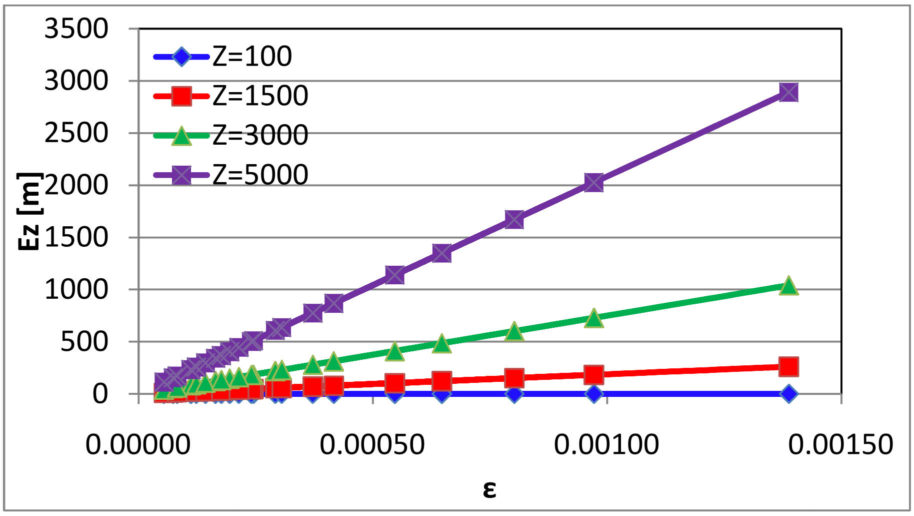

Figure.7.18 Relation of range error (Ez) and ε, N=720, LB=12

7.3 Evaluation of Range Error Percentage

Range error percentage calculation is carried out in two situations of N=720 and N=2048 for different view angles in different image plane accuracies. These are shown through Figure.7.19 to Figure.7.21. In these figures;

$$k = \frac{Z}{LB} \tag{7.1}$$

which is range value normalized by base length.

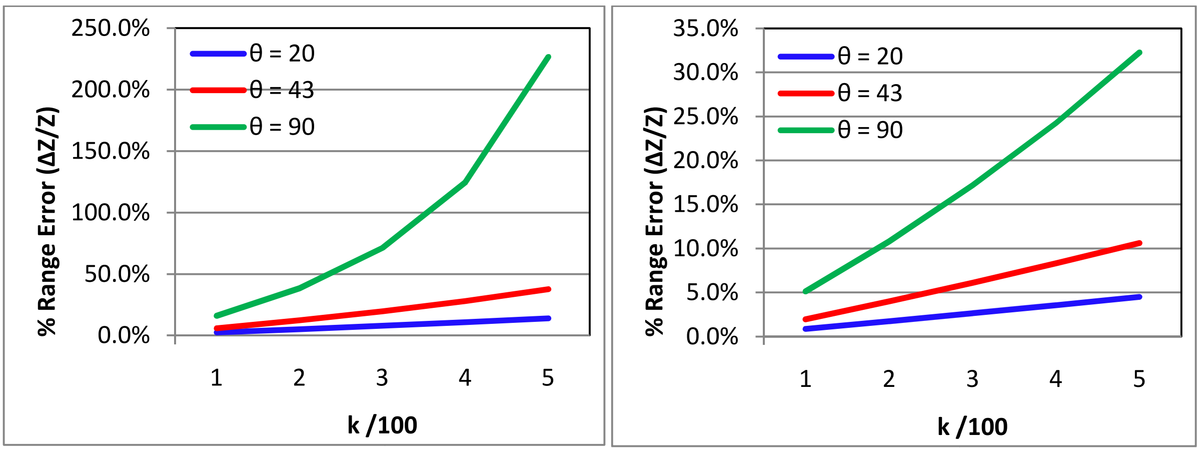

Figure.7.19 % error for one pixel accuracy, N=720 (left) and N=2048 (right)

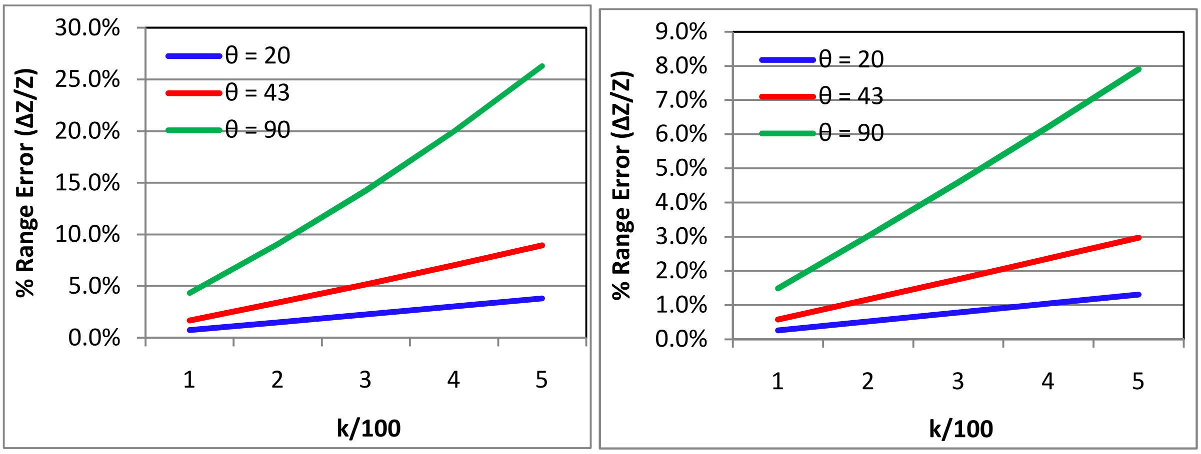

Figure.7.20 % error for sub-pixel accuracy, N=720 (left) and N=2048 (right)

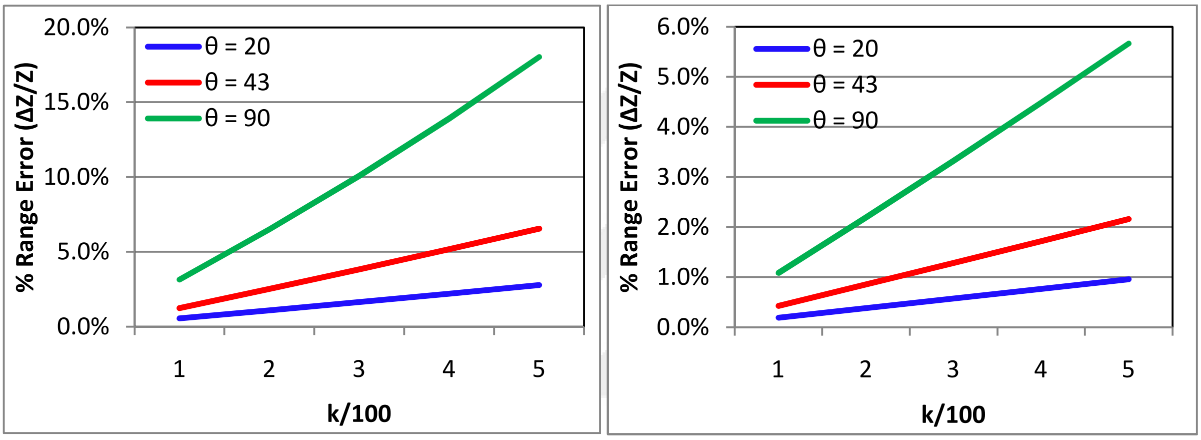

Figure.7.21 % error for BLPF refinement, N=720 (left) and N=2048 (right)

7.4 Minimum range

For detection and collision avoidance systems the minimum measurable distance (Znear) has also significance. In the case of stereo system minimum range measurement means that the image of same scene point can be taken by both of the cameras. The minimum range measurements in different situations are listed in Table.7.3. In our experimental set up the minimum measurable range is 20.68 meters. In other words, both cameras are able to capture the same scene point after 20.68 meters from camera centers.

Table.7.3 Minimum range measurements

θ \ LB	4	6	8.14	10	12
20	22.69	34.03	46.17	56.72	68.07
30	14.95	22.42	30.42	37.37	44.84
43	10.16	15.24	20.68	25.40	30.49
50	8.59	12.88	17.47	21.46	25.76
60	6.93	10.39	14.10	17.32	20.79
70	5.72	8.58	11.63	14.29	17.15
90	4.00	6.00	8.15	10.01	12.01

Figure.7.22 is illustrating the minimum range data in different situations.

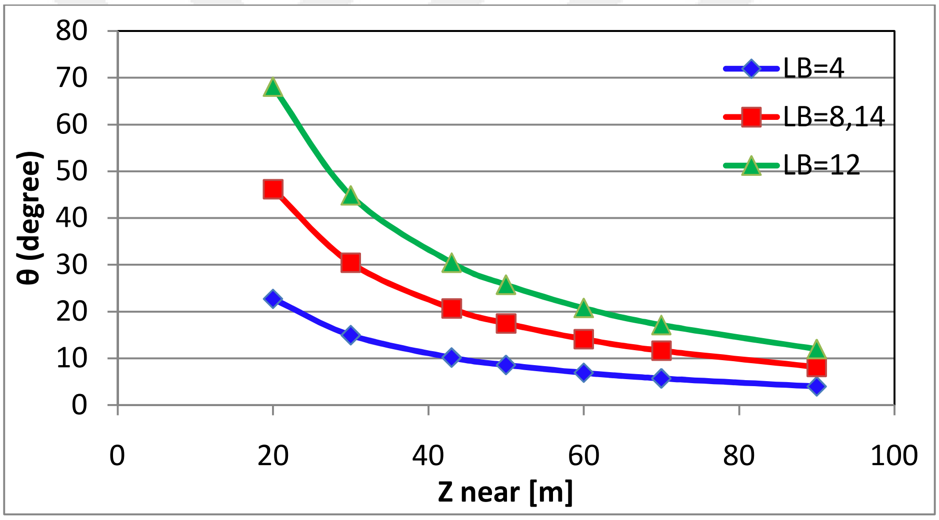

Figure.7.22 Minimum range (Z near)

References

[1] S.B. Marapane and M.M. Trivedi, “ Region and edge segment based stereo analysis for robotic tasks”, International Symposium on Intelligent Robots, Vol. 1571, pp.71-89, SPIE, January 1991.

[2] O. Faugeras and G. Toscani, “ The calibration problem for stereo”, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.15-20, June 1986.

[3] K.B. Johnson, “Development of a versatile wide-angle lens characterization strategy for use in the omnister stereo vision system”, Master Thesis, University of Tennessee, 1997.

CHAPTER 8

CONCLUSION

In this study, the accuracy of the stereo vision system to detect ships and the methods how to improve the accuracy is mainly investigated by theoretical and experimental considerations. Accuracy improvement is considered from point of view of stereo pair correspondence error. Especially error due to lack of image plane quantization and disparity vibration is considered. The effect of accuracy on maximum feasible distance to which stereo vision could be applied is analyzed. The sub-pixel accuracy of 0.11 pixel level is achieved in image plane coordinates which corresponds to 77% of the 3-D measurement error in metric coordinates. The experiments show that almost 70% of the error is eliminated by sub-pixel matching and the 7% of the error is eliminated by disparity refinement. In total 77% of the error is eliminated. In other words, in a determined range the Z dimension error of ±170 meters is reduced to ±35 meters.

Besides, the measurable maximum and minimum ranges are also shown. With our stereo camera configuration and camera intrinsic parameters it is sufficiently available to detect and estimate the position of ship in the range of 20 meters to 5400 meters which is well matched with the purpose of the study of detecting and locating of small ships in a close range with conventional video cameras. Comparing to previous study, the maximum feasible distance with 150 meters range error is increased from 2600 meters to 5400 meters. If the measuring distance is reduced to 1.5 km the range error will be reduced to ±12 meters, which is acceptable for this distance. The results indicate that we can measure distances of ships with about 1% error of ship distance. However, more accurate and larger range measurements can be carried out by changing the parameters which are described in the study. Nevertheless, the ability of a conventional video camera to get effective detections is limited and weather conditions, sea state and night time detection is not considered in this study. Besides, this system is considered as a support system to existing navigational equipment but not replacing them.

In our study and most of the studies, left image is considered as reference image and it is assumed true. And the relative error in the right image is calculated by sub-pixel matching. However, the projection location on the left image plane is also not exactly true (within the ±1/2 pixel level). Therefore, especially in the stereo image sequences it causes the error to continue in the following sequences of the image. A sub-pixel method which carries out the sub-pixel registration of both images can be considered as a future work to increase the accuracy.

Some of the noticed shortcomings of the study can be summarized as:

1. Disparity Gradient Restriction: A restriction in the disparity variance is mentioned. The measurement is good when the ship is going parallel to cameras (or constant disparity). However when the ship is getting closer or far from cameras (changing disparity values of ship) some delayed results occurred in the disparity refinement process. This should be improved and the measurement should be accurate in changing disparities, too. Removing this restriction and to be able to make accurate measurements in any range of disparity gradient is also considered.

2. Matching Outliers: During the matching process of small ship some outlier errors are occurred. This may be due to specular property of sea surface and foggy weather condition. These outliers can be eliminated applying RANSAC algorithm.

On the other hand, there are other causes of errors such as improper camera calibration. Therefore, the accuracy of the proposed system must be totally considered as a future work.

In conclusion, the proposed method on applicability of stereo vision for ship detection and position estimation is innovative, rather robust and satisfactory.