A New Approach for Safety Navigation of Ships Using a Machine Vision System

KOCAK, GAZI

(Degree)
博士（工学）

(Date of Degree)
2013-09-25

(Date of Publication)
2014-09-01

(Resource Type)
doctoral thesis

(Report Number)
甲第5968号

(URL)
https://hdl.handle.net/20.500.14094/D1005968

Doctoral Dissertation

A New Approach for Safety Navigation of Ships Using a Machine Vision System
(マシンビジョンシステムを用いた船舶の安全航行に向けた新しい取り組み)

July 2013

Graduate School of Maritime Sciences, Kobe University

Gazi KOCAK

A New Approach for Safety Navigation of Ships Using a Machine Vision System

Gazi KOCAK
B.Eng. Marine Engineering
Istanbul Technical University, Turkey, 2003

Submitted to
the Department *of* Marine Engineering
and the committee on
the Graduate School *of* Maritime Sciences
in partial fulfilment of the requirements for the degree of

DOCTOR OF ENGINEERING
in Maritime Sciences

Kobe University
July 2013

This page is intentionally left blank

Acknowledgments

First of all I would like to thank to my parents for their support to study in Japan and my wife to accept to come to Japan with me which was a brave decision. She suffered loneliness and a different culture for me and tolerated me in this stressful period.

I am thankful to my supervisor, Ass.Prof. Shigehiro Yamamoto for accepting me as a graduate student and his guidance through the study. As I observed Japanese people are the most kind and delicate society in the world. In this context I can figure out that Yamamoto Sensei has been tolerating me so much due to some culture difference. I am thankful to him for his tolerance through my stay in Japan.

I am grateful to Professor Katsutoshi Kozai and Professor Hiroaki Samata for taking their time to read my thesis and their interest to improve my study.

Dr. Cemil Yurtören encouraged me to study at Kobe University and guided for applying to Monbusho Scholarship. Dr. Serdar Kum helped me to start a life in Japan when I just arrived in Japan. I gratefully acknowledge them for their time and interests.

I would like to thank Professor Nobukazu Wakabayashi for his warm support to adapt Japanese culture and for his programming classes.

I am thankful to Dr. Hidenari Makino for his support and ideas about the study. He always encouraged me and tried to keep my motivation high. He shared precious information to get used to Japanese culture and become as a family to me.

I would like to acknowledge the help of Mr. Yasuhiro Nomura through my study. He spent his valuable time to help me about my study and he became a friend to me.

I am also thankful to Mr. Okan Duru, Mr. Emrah Bulut and all of my friends in Japan for their mental support and sharing good times.

This page is intentionally left blank

Contents

CHAPTER 1

INTRODUCTION

1.1 Background

International trade through shipping is significant. Almost 90% of international trade is carried out via shipping. The size of seaborne trade is increasing around 3% per year. In 2006 global shipping was accounted as 30.686 million 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. The growth of sea born trade from 1983 to 2010 according to Fearnleys Review (CO_2 Emissions, 2010) is shown in Fig.1-1.

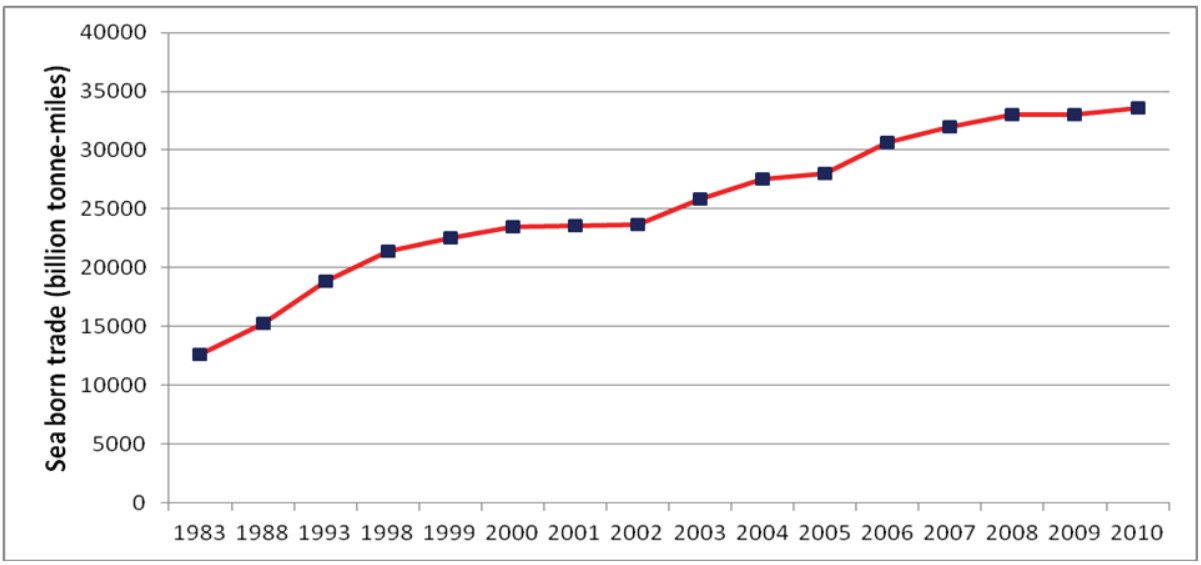


Fig.1-1 Growth in world seaborne trade (billion tonne-miles)

Increase in the seaborne trade affects the number and size of the ships globally. The total deadweight tonnage (dwt) of ships is increasing therewith the increase in the size of individual ships. In Fig.1-2 the increase of the number of ships by selected ship types from the Clarkson Research 2012 is shown (CWFS, 2012).

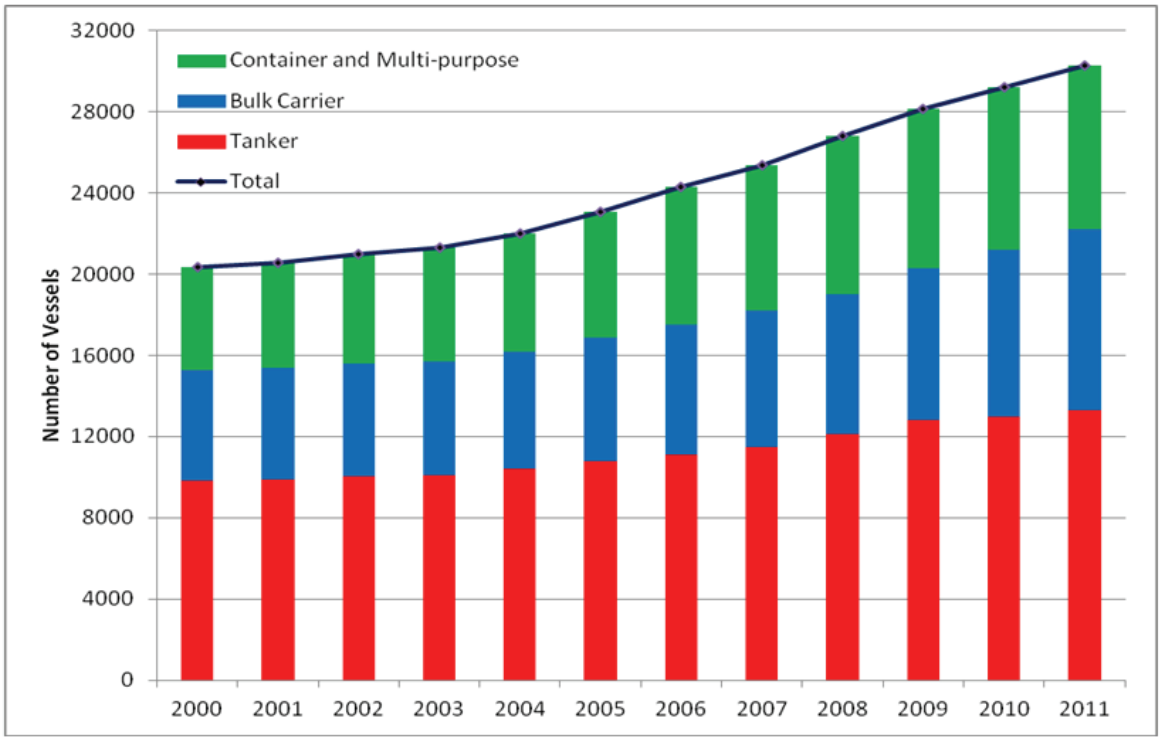
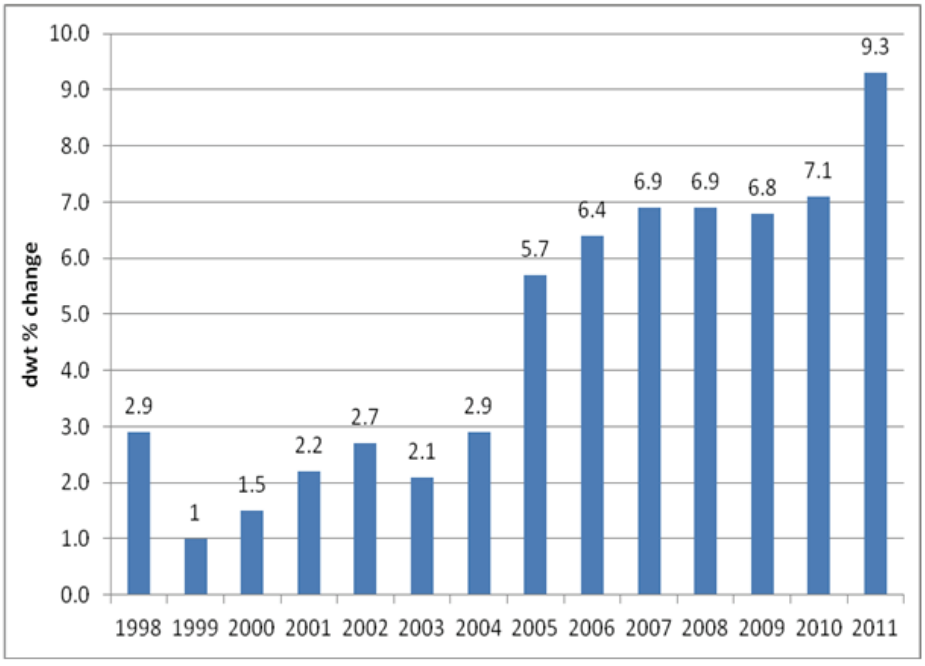
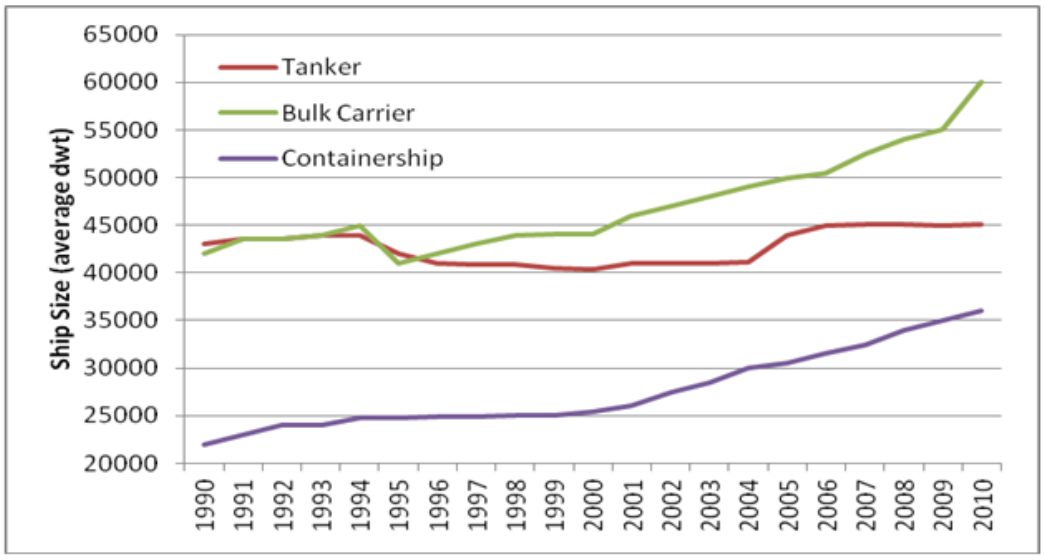


Fig.1-2 Increase in the number of ships by ship types

The increase of the ship dwt and the individual ship size from the statistics of ISL can be seen in Fig.1-3 (ISL-SSMR, 2011).



(a)



(b)

Fig.1-3 (a) World merchant fleet- Annual tonnage changes 1998-2011 (dwt-percent)

(b) World merchant fleet- Ship size development of selected ship types

Safety ship navigation has become more important by the increment in the ship sizes and number of ships worldwide. The researches show that there is not a significant change in the number of marine accidents. In Fig.1-4 there is a research outcome of Japan Coast Guard (JCG, 2007) which is showing the number of accidents from 1976 to 2007.

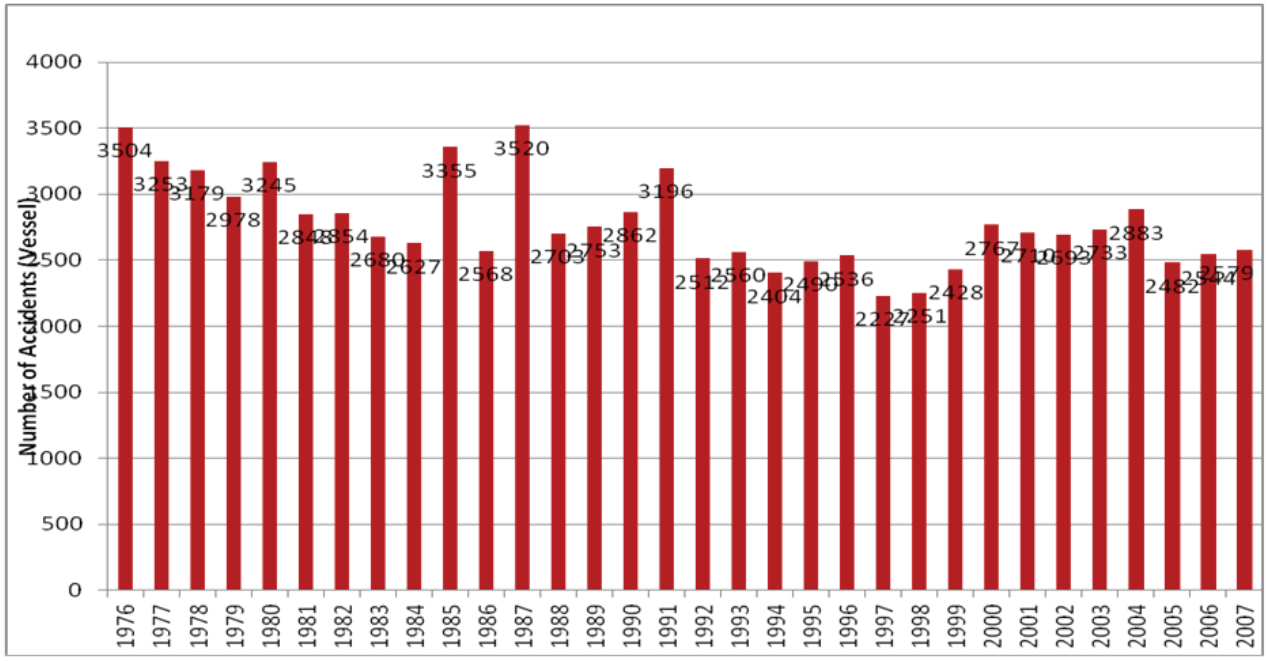


Fig.1-4 The number of maritime accidents

In this report, all of the maritime accident types are included. We would like to focus on the accidents which are related to safety navigation such as collision and contacts. The statistical data of European Maritime Safety Agency (EMSA) show that the number of collisions and contacts in European waters are forming the great part of all accidents which is displayed in Fig.1-5 (EMSA, 2010).

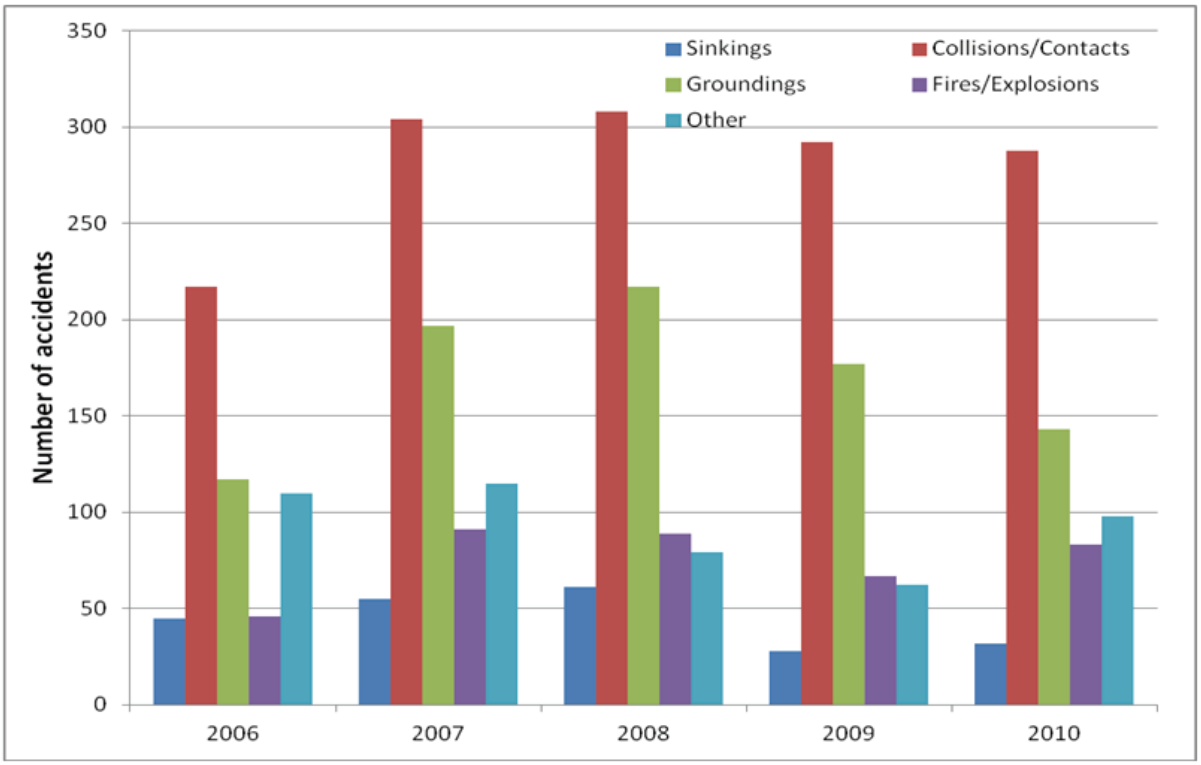


Fig.1-5 Number of accidents by different accident types

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 need 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.

Analysis of maritime accidents shows that the main reason of accidents is categorized as "human factor". According to the report of Marine Accident Inquiry Agency more than 80% of accidents are due to human factors and 54.3% of the collisions are due to "improper lookout (MAIA, 2007). Another research shows that the 54.4% of collisions between 2002 and 2006 are due to improper look-out (KMST). In the study (Furusho et al., 2011) the ratio of accidents due to improper look-out is almost same between 2000 and 2008. 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, especially collision.

For safer navigation through improved look-out system, 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. Radar is the essential tool for detecting other ships. AIS is a relatively new equipment and mandatory only for the vessels which are greater than 300GT. Therefore, AIS is capable of detecting ships which are greater than 300 GT in the range of 20 nautical miles. However, the ships smaller than 300 GT and which are not carrying AIS device induce a high risk of collision. Even more in case of small ships such as fishing boats, the radar cannot distinguish the ship and sea clutter and may cause false detection. The researches show that the most of the collisions are with such kind of small boats. For example, MAIA Digest which is a publication of Marine Accident Inquiry Agency Japan published a report with the title of "Collisions with Fishing Boats under Operation" in September 2008 (MAIA, 2008). This report says that 50% of the collisions reported between 2003 and 2007 are with the fishing vessels. This is shown in Fig.1-6. Another research carried out in Korea (KMST) is illustrated in Fig.1-7 which shows that 70.4% of collisions are with fishing boats.

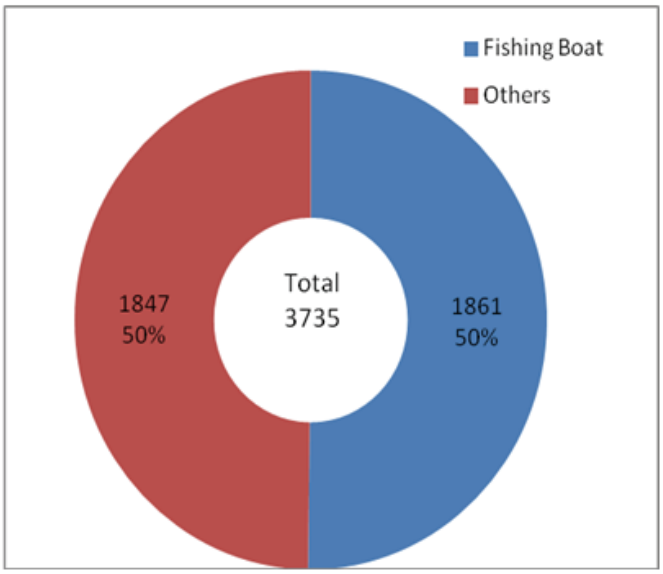


Fig.1-6 Ratio of collisions related to fishing boats (2003-2007)

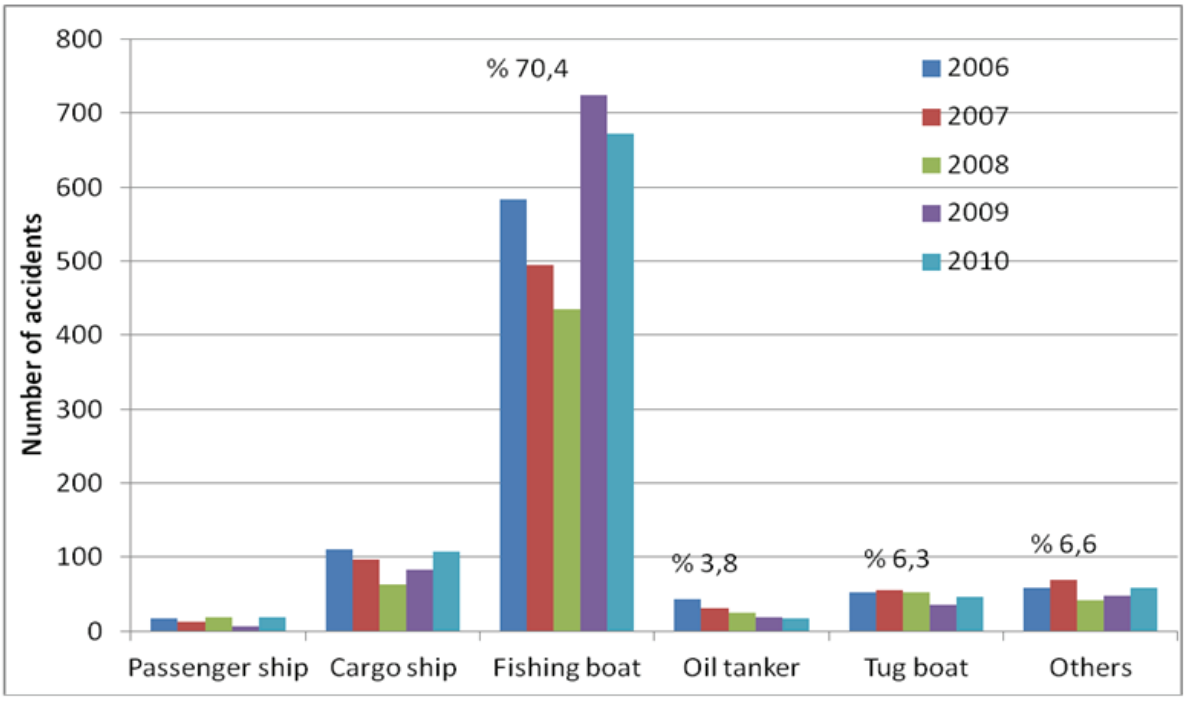


Fig.1-7 Number of collisions by ship type

Therefore, the existing equipment may not be enough for collision-free navigation especially in domestic waters where small vessels like fishing boats exist. And there is a necessity to detect the ships without AIS, which form the greatest ratio of collisions.

In this study I exploit a system to detect ships and other obstacles in the range of about 2 nautical miles (about 3 km) which is important to avoid from collision. The ships to be

detected may be large or small in size and may not carry AIS. The best way to do this is considered to be a 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 (MAIA, 2007). 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 field, so far. Some limited number of studies already has been carried out, but an accurate and precise system has not been developed yet. Furusho et al. showed that most of the collisions are in the good weather conditions and good visibility (Furusho et al., 2011). A camera system can be used under these conditions to automate and improve the look-out for safety navigation.

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, intelligent look-out system 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.

1.2 Literature Review

The topics of obstacle detection and localization are mostly studied in realizing unmanned (autonomous) vehicles. In the autonomous vessel the environment information is needed for navigation (localization and mapping). Obtaining the environment information through several sensors, the navigation planning is carried out in accordance with some localization and mapping algorithms. In this study I focus on obstacle detection, position estimation and tracking the estimated positions of targets using visual information. The obtained information is including stationary and moving obstacles which are useful for collision avoidance.

1.2.1 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), laser and optical (or digital) images. Most of the image based ship detection studies are using satellite SAR imagery and optical satellite imagery. A comparatively new study is presenting the state-of-art in ship detection with SAR imagery (Crisp, 2004). Sato and Ishii (Sato and Ishii, 1998) studied a collision avoidance system which integrates the radar and infrared (IR) image information and Alves et al. studied classification of ship types exploiting Forward Looking Infrared (FLIR) images (Alves et al., 2004). On the other hand, the number of digital image-based ship detection studies is very limited which is the topic of this study.

1.2.1.1 Maritime Surveillance with Single Camera

Takahashi et al. proposed a ship detection method based on combination of some image processing techniques to the sea images which are taken from a shore set-up, stationary single photo camera (Takahashi et al., 2005). Santhalia et al. studied classification of ship types

applying neural networks to digital images from a single stationary camera (Santhalia et al., 2008). Luna et al. proposed another method for ship identification which used silhouette of ship as an input and compared it with the stored database using Concavity-Convexity Scale Space (CCSS) which is an improved model of Curvature Scale Space (CSS) (Luna et al., 2005). However, in the real sea images it is very difficult to get such clear silhouette of ship images. Ju et al. studied detecting and extracting the moving ship from complex background by applying mixture of Gaussians models using stationary surveillance cameras (Ju et al. 2008). Wang et al. used change detection, morphological operators and Connected Components Labeling for ship detection from stationary cameras for a port area surveillance system (Wang et al., 2008). Qi et al. studied maritime surveillance system for detecting small surface objects calculating the dark channel image and considering atmospheric scattering model (Qi et al., 2010). Socek et al. applied an algorithm for object detection from a static camera for maritime surveillance (Socek et al., 2005). They obtained a probabilistic background model for foreground segmentation and then made color segmentation to detect the object. Voles et al. studied a target identification system for maritime surveillance using a static camera (Voles et al., 1999). In this study they applied anisotropic diffusion for filtering the image and used statistical differences for object detection. Then they improved the study by tiling the image applying variable size image windows and Mahalanobis distance measure for segmentation and obtaining the object region (Voles et al., 2000). Szpak and Tapamo (Szpak and Tapamo, 2011) studied detection and tracking of sea objects using level-set background subtraction and active contour tracking method. Their main purpose is maritime surveillance. Therefore they used stationary cameras with which detection and tracking of objects is much easier than moving cameras. Fefilatyev et al. applied a maritime surveillance system using a camera installed on a buoy (Fefilatyev and Goldgof, 2008; Fefilatyev et al., 2009; Fefilatyev et al., 2010). Their algorithm uses edges for detecting ships assuming all the

edges above detected horizon line belongs to an object which result in false detections when the horizon line is not straight due to waves. Bouma et al. used digital cameras and infrared cameras for surface vessel detection (Bouma et al., 2008). Their algorithm is firstly detecting the horizon line and then estimating the background using robust computer vision techniques. However, there is not enough explanation of the method in the paper. Finally, object is obtained by subtracting the image from background.

1.2.1.2 Single Camera Installed Onboard Ship

Smith and Wright'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 (Smith and Wright, 1971). They used only simulated images which are like overhead ship images but not a side view image of ships. Hayashi et al. studied a semi-automatic image ranging system for ship detection purpose (Hayashi et al., 1994-a, Hayashi et al., 1994-b). In this study, the minute angle and global geometric calculations using horizon line is exploited for range calculation. The system they developed is interactive and needs some data through an operator by mouse pointer. Sumimoto et al. studied image processing techniques for rescue operations of marine casualties (Sumimoto et al., 2000). Their idea is based on detecting the orange color in the image due to life saving equipment is usually in this color. 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 equipment as AIS (Shimpo et al., 2005; Shimpo et al., 2006; Shimpo et al., 2007). Another study of Shimpo et al. investigates the feasibility of navigational lookout support system applying image processing techniques to the image sequence from a single video camera (Shimpo et al., 2008). In this study combination of region segmentation and optical flow techniques are emphasized. In the study of Kang et al. ship target extraction is carried out after filtering the image and sea-sky line is obtained (Kang et al., 2006). They

applied a mean filter depending on the variance of local windows for smoothing and multiple histogram matching for object extraction. In this study a ship is assumed to lie on the sea-sky line and the algorithm works based on this assumption. However, this is not always true. Besides, their algorithm would fail in case of multiple objects or in case of rolling of the ship.

Martins et al. built an autonomous surface vehicle which had a navigation system that mainly consisted of a monocular camera, a GPS and Inertial Navigation System (INS) (Martins et al., 2007). Kalman filter is used for obstacle position estimation. 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 (Dunbabin et al., 2007, Dunbabin et al., 2008). Liu et al. designed a system composed of a GPS, an e-compass and an omnidirectional camera (Liu et al., 2008). They used visual attention algorithm for object detection which adaptively selected discriminative features and mean-shift algorithm for tracking objects. Gong et al. studied an autonomous surface vessel using an omnidirectional camera, Global Positioning System (GPS) and a gyro sensor (Gong et al., 2008). They used the image sequence data for stereo calculation assuming the vessel speed is fixed and 1.5 m/sec. They obtained the features by Shi-Tomasi algorithm and tracked these features by KLT (Kanade-Lucas-Tomasi) feature tracker. A matching residual-based outlier rejection is applied to eliminate surface reflections. Finally, N-ocular stereo algorithm is applied to get 3D point cloud. Wang et al. applied an adaptive segmentation algorithm for object segmentation which is based on information entropy theory (Wang et al., 2010). They calculate the local image complexity using information entropy theory and obtain the complexity image. Then applying median filter recursively to this image depending on the mean of the total image complexity to decrease

the complexity and when the complexity is under a certain threshold this image is binarized. Finally the object is segmented by thresholding the binary image.

1.2.1.3 Stereo Camera Installed Onboard Ship

Snyder et al. designed an autonomous river navigation system consisting of several sensors as Inertial Measurement Unit (IMU), Radar, Global Positioning System (GPS), Sonar, omnidirectional camera, 6 cameras to get 360°, pan-tilt camera, multi-camera stereo etc. They used an optical flow algorithm with Harris corner detector and normalized cross correlation for detecting and tracking features. They implemented an optic flow vector clustering based on RANSAC algorithm for motion detection. 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 (Ebken et al., 2005). In reference (Larson et al., 2006) 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. Another autonomous surface vessel is studied in (Zhang et al., 2009). This system has a Multi-Purpose Sensor System (MPSS) using a set of sensors that include stereo cameras, radars, compass, Global Positioning System (GPS), Inertial Measurement Units (IMUs) etc. The stereo vision component is used for obstacle detection and composed of two high resolution cameras with wide field of view and around 1 meter baseline. In the stereo

algorithm, the edges are obtained by Laplacian of Gaussians and non-maxima suppression is applied to get sub-pixel accuracy. Then a patch-based correlation algorithm is used for stereo matching. They obtain the 3D point clouds as an intermediate step and extract the objects in it as a post process. Huntsberger et al. proposed an autonomous surface vessel which consists of a dynamic planning engine, a behavior engine and a perception engine (Huntsberger et al., 2011). 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 dense stereo algorithm developed at JPL (Jet Propulsion Laboratory) is used to generate a range map during the USV motion which consists of four cameras. The four cameras actually set two stereo systems, two cameras one set looking forward right side and other two cameras the other one set looking forward left side both sets with 1 meter baseline.

1.3 Objectives of the Research

The main objective of the study is to get more environmental information for the safety navigation of the ship. This information may be integrated with other information from existing navigation equipment and make the environmental information of the ship more clear and understandable for the navigation officer. Considering the technologies for autonomous vessels, we thought that camera can be a promising device for this purpose. Learning from the marine casualty statistics that most of the collisions are in the good weather condition, the digital camera idea is strengthened. Of course, a system composed of infrared and/or thermal cameras with digital cameras may be stronger in the rainy or foggy

weather conditions and night time voyages. It may be a topic for future study. This study is focused on usage of digital cameras only.

In the literature there are various studies using cameras which are mentioned above. Most of these studies are focused on maritime surveillance and autonomous vehicles. And the algorithms are usually for very specific situations and assumptions. For example ships speed is assumed to be constant, the objects to be detected are assumed to be orange color (in case of emergency situations), the ships to be detected are assumed to be above sea-sky line etc. Actually it is assumed to be because of the difficulty of the problem. In case of the surveillance systems the cameras are stable and it is comparatively easier task to set-up the camera system and to detect the objects. When the cameras are onboard ship the complexity of the problem is increasing due to the dynamic structure of the environment. In the studies of autonomous vehicles the integration of many sensors is considered. The visual information is supported by sensors such as GPS, Inertial Measurement Unit (IMU) etc. The autonomous vehicles are usually small boats and their maneuverability is high. That's why they don't need a long range detection system. The visual systems of autonomous vehicles are usually designed for not more than 400 meters. In the literature there is a missing of vision-based system which is designed for navigational lookout support for marine vehicles. The study of Shimpo et al. (2008) is unique about the topic but it is using a single camera and only detection of objects is considered. I would like to contribute to the literature by a stereo vision-based navigational look-out support system which is capable of eliminating the problems of dynamic sea environment such as rolling and pitching of the ship, detecting the objects within a long range around 2 nautical miles, calculating the 3D locations with an acceptable accuracy, tracking them through the image sequences and obtaining the course of the tracked ships. Only digital cameras (not GPS or IMU integration) are considered to achieve this objective which is not realized in the literature yet.

In this study a stereo vision system for accurate detection and localization of the ships is proposed. The illustration of the proposed system is shown in Fig.1-8. The system consists of two cameras installed on the port and starboard side of the ship bridge area. The cameras are looking forward and configured nearly parallel (optical axis) to each other. One of the advantages of stereo vision is that you can both detect and calculate the 3D location of the detected objects. So it can be used to obtain environmental data such as stationary and moving obstacles around ship and their 3D locations as a navigational tool aid. Therefore one of the objectives of the study is to measure the feasibility of the stereo camera system for obtaining environmental information in marine environment and especially the target ships.

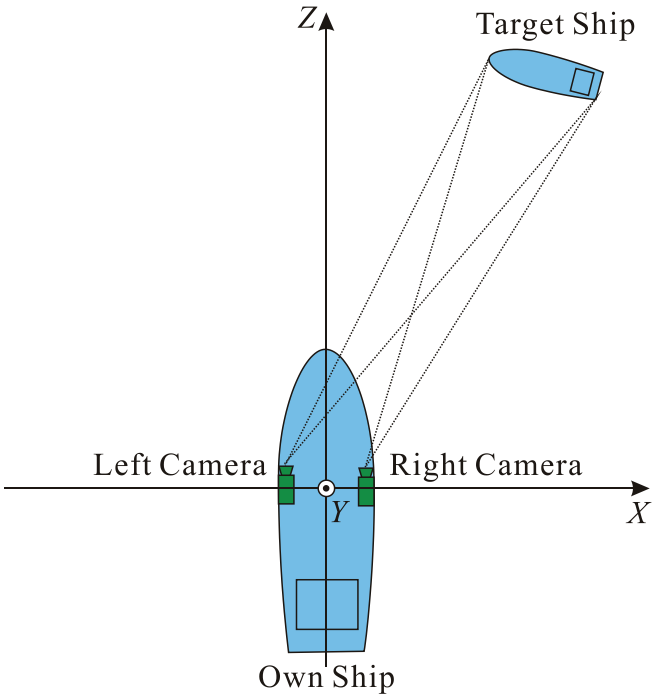


Fig.1-8. Concept of the proposed system

Detection of obstacles that lie on the ships course using visual information is one important point. The specular property of sea surface is a challenging problem for reliable detection. Another point is to calculate the distance of the detected objects accurately from the own vessel. In this case the effect of maritime environment such as ship movement (rolling and

pitching), and vibration may have important influence on camera set-up which should be considered for accurate calculation.

Only detecting the objects and calculating the 3D location is not sufficient for navigational safety. The course and speed of moving obstacles is also needed to take action according to COLREGs (International Regulations for Preventing Collisions at Sea). Therefore another objective is to get course and speed of target vessels. Tracking the 3D locations of detected objects through image sequences gives us the course and speed of them. As a result, this information can be used for increasing the situation awareness of the watch keeping officer or for an automatic action taking of ship control system in the future.

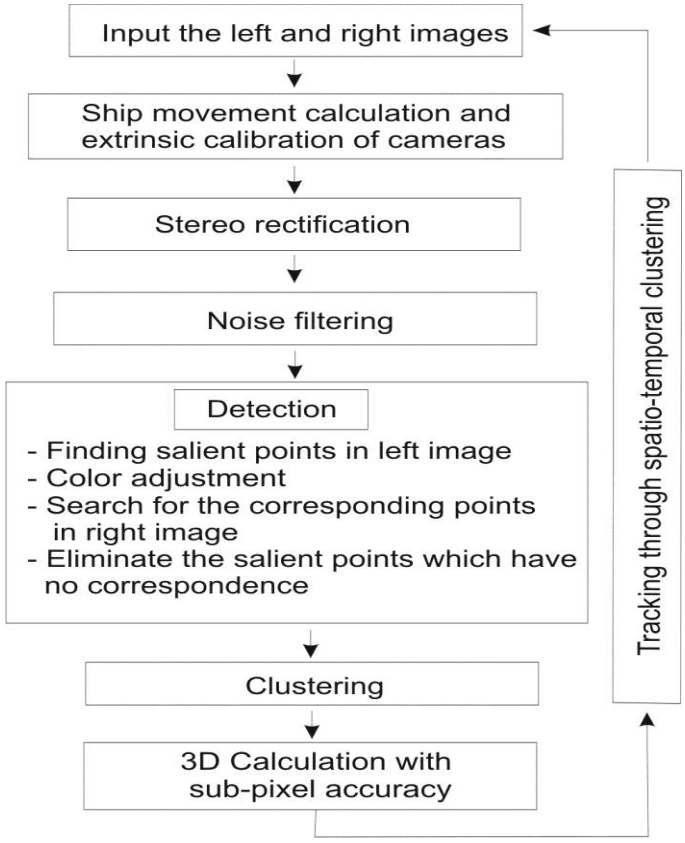


Fig.1-9 Algorithm flowchart of overall system

The algorithm for the overall system is illustrated in Fig.1-9. First, a pair of stereo images, which are referred to as left and right images for convenience, is captured. The movement of the ship is calculated using the image data which is explained in detail in Chapter 7. The

cameras need to be externally calibrated. The extrinsic calibration process is explained in Chapter 2 and its meaning in case of ship rolling is explained in Chapter 7. Then the images are rectified. Rectification of images is explained elaborately in Chapter 3. Noise filtering is one of the important steps to eliminate the noise in the images by some smoothing filters. The effects of different smoothing filters are analyzed in Chapter 5. In Chapter 3 the detection process is explained. The clustering step is studied in Chapter 6. 3D calculation with sub-pixel accuracy is explained in Chapter 4. The tracking step is actually extension of the clustering step to time domain which is called spatio-temporal clustering and the methodology is explained in Chapter 6. The result of the spatio-temporal clustering is the track of the ship location in 3D coordinates. Chapter 8 is describing some further process to improve the study such as some different stereo matching strategies and color space of images. In Chapter 9 the conclusion of the study is explained with the main contributions of the study and future plan is discussed.

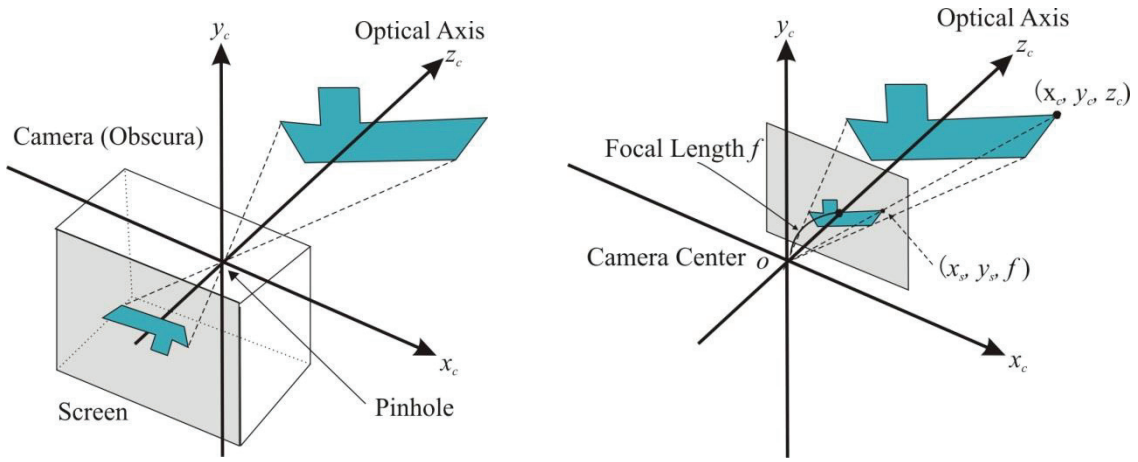
CHAPTER 2

CAMERA CALIBRATION AND STEREO VISION SYSTEM

2.1 Pinhole Camera Model and Perspective Projection

A camera is the device to transform visual information of 3D world to 2D images by performing some transformations. Various parameters are used in these transformations which are fundamental to describe the image acquisition system. The pinhole camera model is one of the most used and simple camera models with perspective projection to imaging screen. The transformation based on perspective projection is describing the behavior of real optic systems (natural systems such as human vision system) which can be described by a linear equation in a higher dimensional space of so-called homogeneous coordinates (Hartley and Zisserman, 2003; Faugeras, 1993).

Fig.2-1(a) shows a pinhole camera model. Light from an object is projected on the screen through a tiny pinhole. Let us define the *xyz* coordinates system whose origin is the pinhole and *z* axis is perpendicular to the screen. The *xyz* coordinate system is referred to as camera coordinate system. The distance between the origin and the screen is focal length. The *z* axis is also called as optical axis. Since the object is projected on the screen upside-down in the model, let us move the screen forward like Fig.2-1(b).



(a) (b)

Fig.2-1 Pinhole camera model and perspective projection

In the pinhole camera model, a point of (x_c, y_c, z_c) in the camera coordinate system is projected onto the point of (x_s, y_s, f), and we can obtain the following equation in matrix form for a real number λ

$$\lambda \begin{bmatrix} x_s \\ y_s \\ f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \tag{2-1}$$

This relation is referred to as perspective projection. This equation is obtained from similar triangles theorem of Thales. In this equation the values of both the camera coordinate point (x_c, y_c, z_c) and its projection (x_s, y_s, f) are expressed in the camera coordinate system. Writing the Eq.2-1 separately we obtain x_s and y_s at focal distance f as

$$x_s = f\frac{x_c}{z_c} \quad , \quad y_s = f\frac{y_c}{z_c} \tag{2-2}$$

2.2 Coordinate Systems

One of the difficult points of perspective projection is the coordinate systems. We have to constitute the mathematical relationship between some coordinate systems to represent the same point in each of them. Mainly we have to consider about four coordinate systems:

1) World coordinate system,
2) Camera coordinate system,
3) Image (pixel) coordinate system, and
4) Ship coordinate system.

World coordinate system is the main coordinate system that we would like to get the 3D reconstruction result of the points. In the experiments, the main purpose is obtaining the data in world coordinate system. The camera coordinate system is the coordinate system which accepts the optical center as the origin. And the image coordinate system is the 2D coordinate system of the image sensor. The relationship between these coordinate systems except for ship coordinate system is illustrated in Fig.2-2.

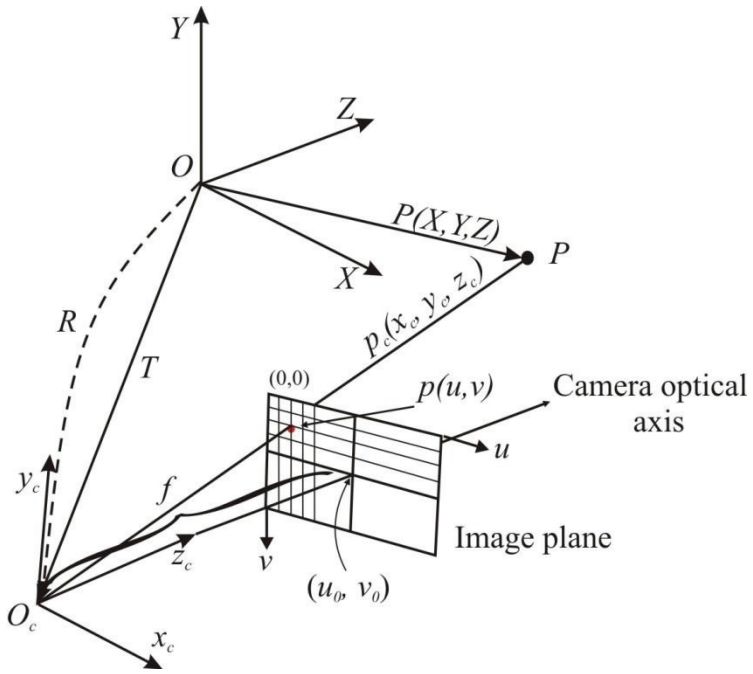


Fig.2-2 Coordinate systems

In Fig.2-2 the point $O(0, 0, 0)$ is the origin of the world coordinate system which is denoted by the axis *X, Y, Z*. In this research, world coordinate system is used for estimating three-dimensional coordinates of marine vessels. Similarly, the camera coordinate system is represented by (x_c, y_c, z_c) with the origin $O_c(x_0, y_0, z_0)$ which is also called optical center. The z_c direction is called optical axis. The relationships between the world coordinates and camera coordinates are denoted by a rotation (R) and a translation (T). Translation of camera center (T) is actually equal to (x_0, y_0, z_0). Another important part of the model is the image plane. The image plane is actually the CCD (Charge-Coupled Device) sensor 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 (optical axis) is called principal point (u_0, v_0) which is theoretically the center point of the image plane (but not in practice) and expressed in pixels. The distance from the O_c to the principal point is the focal length. $P(X, Y, Z)$ is representing a point in world coordinates whereas $p_c(x_c, y_c, z_c)$ and

$p(x_s,y_s,f)$ are representing the same point in camera coordinates and its projection onto image coordinates, respectively. Note that $p(x_s,y_s,f)$ is expressed in the units of camera coordinate system. In other words it is the location of image point in camera coordinate system. $p(u,v)$ is used for denoting the same point in image coordinate system which is shown in Fig.2-3 for a 720x480 resolution image.

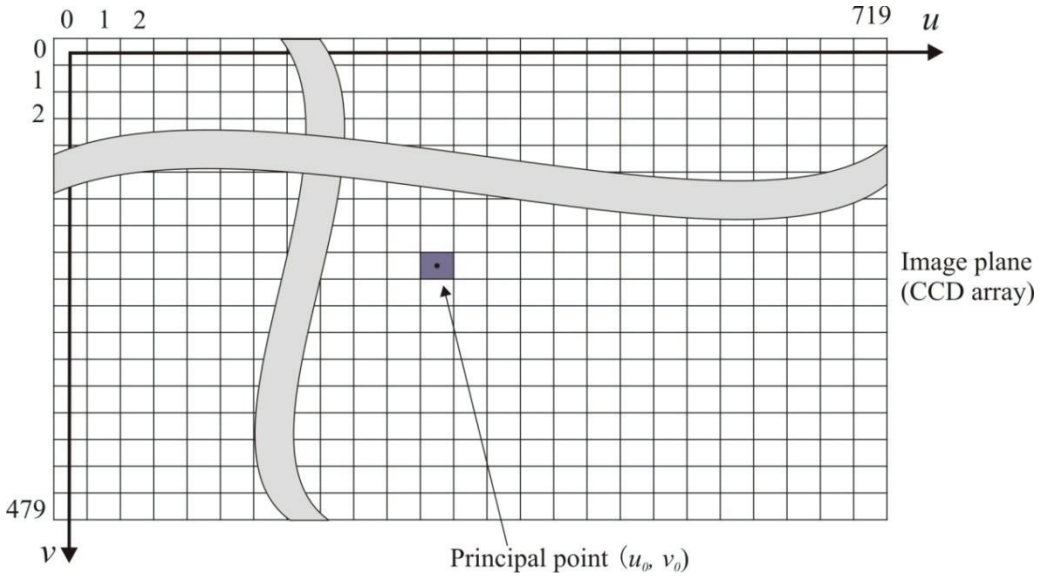


Fig.2-3 The image coordinate system considering sub-pixel locations

Ship coordinate system is denoted by $X_SY_SZ_S$ and it is very similar to world coordinate system. It is actually used for elimination of effect of dynamic environment such as ship motion (rolling and pitching) and vibration. In the case of no rolling, pitching and vibration the ship coordinate system overlaps with the world coordinate system. In case of such dynamics effects, these effects can be expressed in ship coordinate system. And the calculation in world coordinate system is possible using the constituted relationship between the ship coordinate system and the world coordinate system. It is the topic of Chapter 7 and the details are explained in the context of dynamic environment.

2.3 Camera Intrinsic Parameters

The unit of the camera coordinate system is “meter” due to convenience for 3D measurement, while the unit of the image coordinate system is “pixel”. The transformation from the camera coordinates to the image coordinate is carried out by

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \alpha_u & \beta \\ 0 & \alpha_v \end{bmatrix} \begin{bmatrix} x_s \\ y_s \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \tag{2-3}$$

where (u, v) is the projected point in the image coordinate system, and (u_0, v_0) is the principal point. We can express Eq.(2-3) in homogeneous coordinates as

$$f\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f\alpha_u & f\beta & u_0\\0 & f\alpha_v & v_0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_s\\y_s\\f\end{bmatrix} \tag{2-4}$$

Substituting this into Eq.(2-1) the following equation is obtained

$$f\lambda\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}f\alpha_u & f\beta & u_0\\0 & f\alpha_v & v_0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_c\\y_c\\z_c\end{bmatrix} \tag{2-5}$$

Generalization of equation by equating f to unity ($f=1$) is possible that the different values of f just correspond to different scaling of the image (Mohr and Triggs, 1996). Therefore,

$$\lambda\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}\alpha_u & \beta & u_0\\0 & \alpha_v & v_0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_c\\y_c\\z_c\end{bmatrix} \tag{2-6}$$

is obtained. α_u, α_v, β, u_0, v_0 are specific to a camera and are called intrinsic parameters. α_u and α_v are called scaling factor. β is the skew between the axes of image coordinates and (u_0, v_0) is the coordinates of image center (principal point). The inverse of the system which transforms camera coordinates to image coordinates can be expressed as

$$\begin{bmatrix}x_c\\y_c\\z_c\end{bmatrix}=\lambda\begin{bmatrix}\alpha_u & \beta & u_0\\0 & \alpha_v & v_0\\0 & 0 & 1\end{bmatrix}^{-1}\begin{bmatrix}u\\v\\1\end{bmatrix}=\lambda\begin{bmatrix}\frac{1}{\alpha_u} & -\frac{\beta}{\alpha_u\alpha_v} & -\frac{u_0}{\alpha_u}+\frac{\beta v_0}{\alpha_u\alpha_v}\\0 & \frac{1}{\alpha_v} & -\frac{v_0}{\alpha_v}\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}u\\v\\1\end{bmatrix} \tag{2-7}$$

2.4 Camera Extrinsic Parameters

In practice, the position of an object is usually represented in a fixed coordinate system irrespective of the camera (for example the world coordinate system). Therefore, we must constitute the transformation from the camera coordinate system to the world coordinate system. Denoting (X, Y, Z) as the position of the object in the world coordinate system, (X_0, Y_0, Z_0) as the position of camera center in the world coordinate system, and (θx, θy, θz) as the rotation angle around the X axis, Y axis and Z axis of the camera (in world coordinate system) respectively, we obtain the following equation:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R^{-1}\left\{\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}\right\} = \begin{bmatrix} R^{-1} & -R^{-1}T \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} R^{-1} & \boldsymbol{t} \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2-8}$$

where

$$R = R_X R_Y R_Z,$$

$$R_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_X & \sin\theta_X \\ 0 & -\sin\theta_X & \cos\theta_X \end{bmatrix}, \quad R_Y = \begin{bmatrix} \cos\theta_Y & 0 & \sin\theta_Y \\ 0 & 1 & 0 \\ -\sin\theta_Y & 0 & \cos\theta_Y \end{bmatrix}, \tag{2-9}$$

$$R_Z = \begin{bmatrix} \cos\theta_Z & \sin\theta_Z & 0 \\ -\sin\theta_Z & \cos\theta_Z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{t} = -R^{-1}\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \tag{2-10}$$

T and R are called extrinsic parameters.

2.5 Camera Calibration

2.5.1 Introduction

So far, the theoretical explanation of a camera system is explained. However, in practice there are some differences from the theory. For example, the principal point (u_0, v_0) is usually not at image center in practice despite it is in the center of the image in theory. Or the image coordinates are not perpendicular to each other which result in skew factor β. Therefore a procedure which is called camera calibration is carried out to calculate the camera intrinsic and extrinsic parameters. There are various studies carried out for camera calibration (Tsai, 1987; Zhang, 2000; Heikkila and Silven, 1997). In this study a calibration method using known 3D point coordinates is preferred.

Substituting Eq.(2-8), Eq.(2-9), and Eq.(2-10) into Eq.(2-6) yields the following equation:

$$\lambda\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KR\begin{bmatrix} I & -T \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = P\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \tag{2-11}$$

where K is intrinsic parameter matrix, T is translation, I is identity matrix and P is parameter matrix. Therefore, if the elements of the parameter matrix P are known, then the projection of the point (X, Y, Z) in the 3-D space to the point of (u, v) in the image space can be obtained. It is possible to assign $P_{34}=1$. Therefore, rearranging Eq.(2-11) gives the following equations:

$$\lambda u = p_{11}X + p_{12}Y + p_{13}Z + p_{14} \tag{2-12}$$

$$\lambda v = p_{21}X + p_{22}Y + p_{23}Z + p_{24} \tag{2-13}$$

$$\lambda = p_{31}X + p_{32}Y + p_{33}Z + 1 \tag{2-14}$$

Following equations are obtained by substituting Eq. (2-14) into Eq. (2-13) and Eq. (2-12)

$$Xp_{11} + Yp_{12} + Zp_{13} + p_{14} - uXp_{31} - uYp_{32} - uZp_{33} = u \tag{2-15}$$

$$Xp_{21} + Yp_{22} + Zp_{23} + p_{24} - vXp_{31} - vYp_{32} - vZp_{33} = v \tag{2-16}$$

Eq.(2-15) and Eq.(2.16) are only for one point. Then, n points in the 3D space and their projections in the image plane are observed. Let $P_i(X_i, Y_i, Z_i)$ and $p_i(u_i, v_i)$ (i=1,2,…,n) as these points, and substituting them into Eq.(2-15) and Eq.(2-16) gives:

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \\ v_1 \\ \vdots \\ v_n \end{bmatrix} \tag{2-17}$$

Because the number of unknown parameters is 11, we can obtain $p_{11}\sim p_{33}$ by least square method if n is greater than or equal to 6. Note that we cannot determine the parameters if all the n points are coplanar in 3D space. It is desired to use uniformly scattered points in 3D space to determine the parameters correctly.

In case of dynamic environment we express the ship coordinate system separately as explained in Section 2.2. In this case one more parameter R_s should be considered which

represents the rotation caused by ship motion. The relationship between the *XYZ* coordinate system and the *XsYsZs* coordinate system is expressed by

$$\begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix} = R_S \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2-18}$$

where

$$R_S = R_Z(\theta_Z) R_X(\theta_X) \tag{2-19}$$

Integrating the new rotation matix R_s we can express the overall relation as

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KR[R_s \quad -T] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2-20}$$

where p_{ij} (i=1, 2, 3, j=1, 2, 3, 4) are camera parameters obtained by following the above procedures. Equation (2-20) can be transformed into

$$\begin{bmatrix} p_{31}u - p_{11} & p_{32}u - p_{12} & p_{33}u - p_{13} \\ p_{31}v - p_{21} & p_{32}v - p_{22} & p_{33}v - p_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{14} - p_{34}u \\ p_{24} - p_{34}v \end{bmatrix} \tag{2-21}$$

When an object located at (*X*, *Y*, *Z*) is projected onto (u_r,v_r) and (u_l, v_l) in the right and left image, respectively, 3D location of the object (*X*, *Y*, *Z*) is calculated by solving the following the least square problem.

$$\begin{bmatrix} p_{31}u_r - p_{11} & p_{32}u_r - p_{12} & p_{33}u_r - p_{13} \\ p_{31}v_r - p_{21} & p_{32}v_r - p_{22} & p_{33}v_r - p_{23} \\ q_{31}u_l - q_{11} & q_{32}u_l - q_{12} & q_{33}u_l - q_{13} \\ q_{31}v_l - q_{21} & q_{32}v_l - q_{22} & q_{33}v_l - q_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_{14} - p_{34}u_r \\ p_{24} - p_{34}v_r \\ q_{14} - q_{34}u_l \\ q_{24} - q_{34}v_l \end{bmatrix} \tag{2-22}$$

where p_{ij} and q_{ij} (i=1, 2, 3, j=1, 2, 3, 4) are right and left camera parameters, respectively.

2.5.2 Separation of Intrinsic Parameters and Extrinsic Parameters

We need the intrinsic and extrinsic parameters of a camera separately for the further calculations such as distortion correction of lens as described later. Here, how to calculate the intrinsic parameters and extrinsic parameters separately from projection matrix *P* is shown.

First of all, each parameter is divided by $\sqrt{p_{31}^2 + p_{32}^2 + p_{33}^2}$, and the components become as follows:

$$P = \begin{bmatrix} \boldsymbol{p}_1^t & p_{14} \\ \boldsymbol{p}_2^t & p_{24} \\ \boldsymbol{p}_3^t & p_{34} \end{bmatrix}, \quad \boldsymbol{p}_1 = \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \end{bmatrix}, \quad \boldsymbol{p_2} = \begin{bmatrix} p_{21} \\ p_{22} \\ p_{23} \end{bmatrix}, \quad \boldsymbol{p_3} = \begin{bmatrix} p_{31} \\ p_{32} \\ p_{33} \end{bmatrix} \tag{2-23}$$

where, $\|\boldsymbol{p_3}\| = \sqrt{p_{31}^2 + p_{32}^2 + p_{33}^2} = 1$ as above formula manipulation. Under this assumption, we can calculate the intrinsic parameters as follows:

$$u_0 = \boldsymbol{p_1} \cdot \boldsymbol{p_3} \tag{2-24}$$

$$v_0 = \boldsymbol{p}_2 \cdot \boldsymbol{p_3} \tag{2-25}$$

$$\alpha_v = \|\boldsymbol{p}_2 \times \boldsymbol{p_3}\| = \sqrt{\left(\|\boldsymbol{p}_2\|\|\boldsymbol{p}_3\|\right)^2 - \left(\boldsymbol{p}_2 \cdot \boldsymbol{p_3}\right)^2} \tag{2-26}$$

$$\beta = \frac{\boldsymbol{p_1} \cdot \boldsymbol{p}_2 - u_0 v_0}{\alpha_v} \tag{2-27}$$

$$\alpha_u = \|\boldsymbol{p}_1 \times \boldsymbol{p_3}\| - \beta = \sqrt{\left(\|\boldsymbol{p}_1\|\|\boldsymbol{p}_3\|\right)^2 - \left(\boldsymbol{p}_1 \cdot \boldsymbol{p_3}\right)^2} - \beta \tag{2-28}$$

and,

$$\begin{bmatrix} \alpha_u & \beta & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R^{-1} \quad T] = P \tag{2-29}$$

so,

$$[R^{-1} \quad T] = \begin{bmatrix} \alpha_u & \beta & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} P \tag{2-30}$$

Using above equations the rotation array *R* and translation *T* are calculated.

2.5.3 Correction of Lens Distortion

The image is affected by the lens shape used in the camera and the shapes of objects may be distorted. So far, the calculations were carried out as the lens has no distortion. However, in practice the image of a straight line may be bent in image space due to radial distortion.

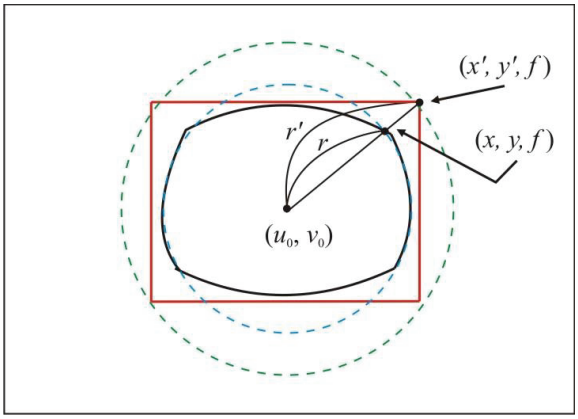


Fig.2-4 Lens radial distortion model

The effect of distortion can be better understood from Fig.2-4. Here, (x, y, f) is the actual projected point in the camera coordinate system and (x', y', f) is the distortion corrected point. r and r' are radial distance from principal point. r' can be calculated by

$$r' = r + k_1 r^3 + k_2 r^5, \quad x' : y' = x : y \tag{2-31}$$

$$r' = \sqrt{x'^2 + y'^2}, \quad r = \sqrt{x^2 + y^2} \tag{2-32}$$

If $k_1 = k_2 = 0$, there is no distortion. If k_1 and k_2 are positive, the distorted point is further away from the principal point. This is called barrel distortion model. In contrast, if both k_1 and k_2 are negative, it is spool distortion model. By the way, $x'/r' = x/r$ leads to

$$x' = x(1 + k_1 r^2 + k_2 r^4) \tag{2-33}$$

and $u' = \alpha_u x' + u_0$, $v' = \alpha_v y' + v_0$, $u = \alpha_u x + u_0$, $v = \alpha_v y + v_0$ (we can ignore β because it is very small) lead following:

$$(u' - u_0) = (u - u_0)(1 + k_1 r^2 + k_2 r^4), \quad r^2 = \left(\frac{u - u_0}{\alpha_u}\right)^2 + \left(\frac{v - v_0}{\alpha_v}\right)^2 \tag{2-34}$$

We can apply the same transformation for v and v' and we obtain the following equations:

$$(u - u_0) r^2 k_1 + (u - u_0) r^4 k_2 = u' - u \tag{2-35}$$

$$(v - v_0) r^2 k_1 + (v - v_0) r^4 k_2 = v' - v \tag{2-36}$$

For practical purpose, it is needed to measure a large number of $(u_i ; v_i)$, $(u'_i ; v'_i)$ $(i = 1, 2, .., n)$ and solve the following least-squares problem.

$$\begin{bmatrix} (u_1-u_0)r_1^2 & (u_1-u_0)r_1^4 \\ \vdots & \vdots \\ (u_n-u_0)r_n^2 & (u_n-u_0)r_n^4 \\ (v_1-v_0)r_1^2 & (v_1-v_0)r_1^4 \\ \vdots & \vdots \\ (v_n-v_0)r_n^2 & (v_n-v_0)r_n^4 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} u_1'-u_1 \\ \vdots \\ u_n'-u_n \\ v_1'-v_1 \\ \vdots \\ v_n'-v_n \end{bmatrix} \tag{2-37}$$

The camera intrinsic parameters were obtained using the distorted images. That's why it is difficult to obtain the distortion parameters. The distortion parameters are calculated using the intrinsic parameters obtained by camera calibration. Then distortion correction is carried out. Actually, this distortion correction is not sufficient. Therefore, the camera calibration is carried out again using these corrected images and new camera parameters are obtained. And then, the distortion parameters are calculated again using the new camera parameters which result in refined camera and distortion parameters. This procedure is recursively repeated until the error between calibration results converge to a threshold value which is small enough.

2.6 Stereo Vision System

2.6.1 Introduction to Stereo Vision

The basic studies of computational stereo are going back to 1960's and 1970's. Julesz (1960) studied depth perception of computer generated patterns. The stereo correspondence problem is studied through inspection of the human stereo vision system (Marr, 1974; Marr and Poggio, 1976; Marr and Poggio, 1977). Longuet-Higgins (1981) focused on computing relative orientation of cameras for 3D reconstruction from binocular stereo images. The advances of computational stereo especially in correspondence problem are summarized in Brown et al. (2003). A survey study about stereo vision methods is carried out by Koschan (1993). Recently, stereo vision is one of the most used methods for obtaining depth information from digital images which is using two cameras with a known distance between each other called baseline *(L_B)*. The fundamental theory of stereo vision is explained in detail

in many studies (Faugeras, 1993; Jain et al., 1995; Trucco and Verri, 1998; Hartley and Zisserman, 2003). The basic principle for stereo is the fact that projection of a 3D object point has a unique pair of image locations in two cameras viewing the same scene from different viewpoints. The projection of the object point is different in left and right images. These are called corresponding points. Therefore, given two camera images, it is possible to calculate three-dimensional location of an object point by triangulation if it is possible to find the image locations that correspond to the same physical point in real world. This is the most important problem of stereo vision which is called correspondence problem. A detailed study of correspondence problem for dense stereo matching is carried out by Schartein and Szeliski (2002). The displacement between the locations of projections on the left and right image is called disparity. Disparity and range are inversely proportional. When the object gets closer, disparity value increases. Some of the main reasons of correspondence problem are insufficiency of pixel level quantization, matching cost, occlusion, image noise, and lack of texture. In this study 3D error due to correspondence problem caused by pixel level quantization is eliminated by sub-pixel matching algorithm.

2.6.2 General Stereo Configuration and Epipolar Geometry

In stereo vision system, the cameras may be configured arbitrarily. The cameras may be convergent, parallel or in any other geometric layout. The important thing is obtaining the relative relation of the cameras such as translation of camera centers and relative orientations of cameras. One of the most important constraints of stereo vision is epipolar constraint which is defined using epipolar geometry. Epipolar geometry is defining the relation of corresponding scan lines of left and right images using relative relationship of cameras. 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 of a convergent stereo system are shown in Fig.2-5. In this

case, the corresponding points of left image lying through the epipolar line l are found through the epipolar line l' of right image.

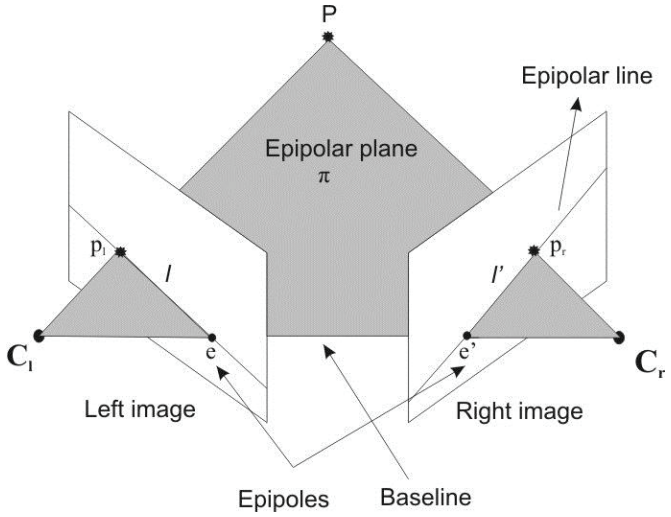


Fig.2-5 Epipolar geometry of convergent cameras

2.6.3 Standard Stereo Configuration

A standard stereo geometry is illustrated in Fig.2-6. 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 (pixel) coordinate system with the origin of upper-left corner. The distance between image center (principal point) and camera (optical) center is called focal length (f).

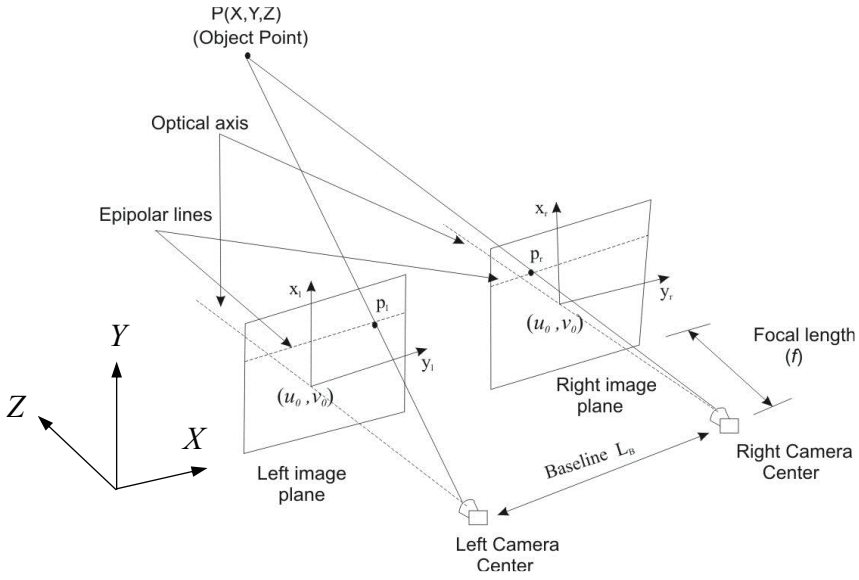


Fig.2-6 Standard stereo image acquisition system

Fig.2-7 illustrates the *X-Z* view of Fig.2-6. In this figure corresponding points and disparity can be seen more clearly.

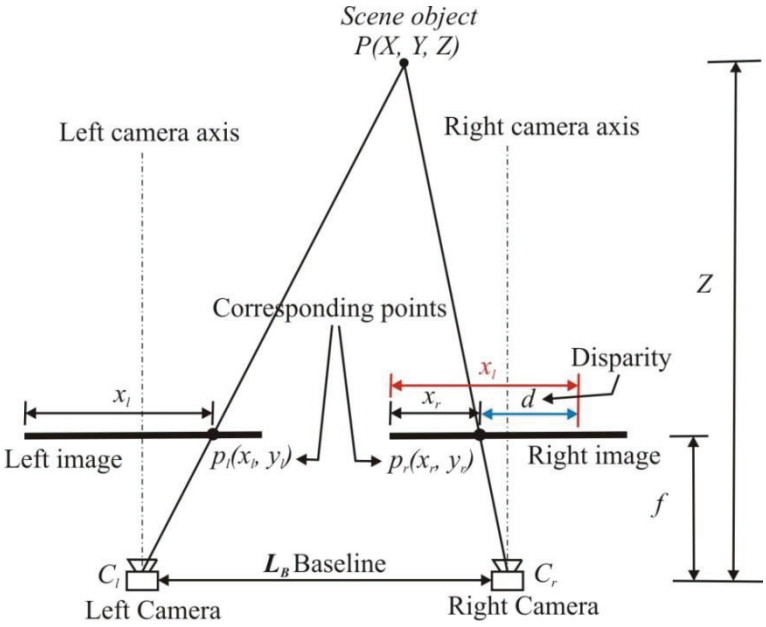


Fig.2-7 Standard stereo system viewed through X-Z axis

The range (*Z*) of the object can be calculated by equation which is using similar triangles:

$$Z = \frac{f.L_B}{x_l - x_r} \tag{2-38}$$

which is called triangulation.

In standard stereo system the vertical disparity is zero, theoretically. In other words, corresponding points lie on the equivalent vertical locations in image planes $(y_l = y_r)$. In standard stereo system epipolar lines are parallel to the image scan lines with constant y (as shown in Fig.2-6) 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 real environments it is very difficult to realize an ideal standard stereo configuration. Even a small misalignment of the cameras will cause invalidity of equations. Therefore, some additional calculations are carried out for obtaining the standard stereo configuration using the camera extrinsic parameters explained in camera calibration. Knowing the relative orientation of the cameras that forms extrinsic parameters, it is possible to get the images of

rotated cases using image processing techniques such that the cameras are in standard stereo configuration. This process is called stereo rectification. One of the main advantages of rectification is reducing the search space of corresponding point to 1D. The rectification should result in same scan lines of left and right images to be on the same line. In other words, the epipolar lines of left and right images should lie on the same line. Therefore, one image point of left image should be on the same scan line of right image with a difference of disparity.

CHAPTER 3

BASIC STUDY

This chapter is explaining the study of (Yamamoto and Win, 2006, and Yamamoto and Win, 2007) which is the foundation of my study. (Yamamoto and Win, 2006) carried out a feasibility study for detecting ships using a stereo camera system. That study is the basic of my study. That's why I need to explain that study of (Yamamoto and Win, 2006, and Yamamoto and Win, 2007) in this whole chapter as basic study and the improvements are discussed in the other chapters.

3.1 Experimental Camera Configuration

The experiment was carried out at the Akashi Kaikyo Bridge to detect the ships from actual sea images. The experimental devices used in this study are two identical 3CCD Panasonic NV-GS300 video cameras with a zoom lens. The view angle of the cameras is maximum 43 degrees and the preferred resolution of the images is 720x480. The cameras are set up nearly parallel directions of camera axis with 8.14m base length between. A tripod is also used for the calibration purposes. Fig.3-1 shows the camera set-up and sample stereo images.

Fig.3-1 Experimental stereo configuration and obtained left and right images

The theoretical representation of experimental camera configuration is illustrated in Fig.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 point $P(X, Y, Z)$ in the world coordinate system is projected onto the $p_r(x_r , y_r)$ in the right image coordinate system and $p_l(x_l, y_l)$ in the left image coordinate system,

respectively. Extrinsic parameters of the cameras, which include locations and orientations of cameras, should be calculated to be able to measure the 3D location.

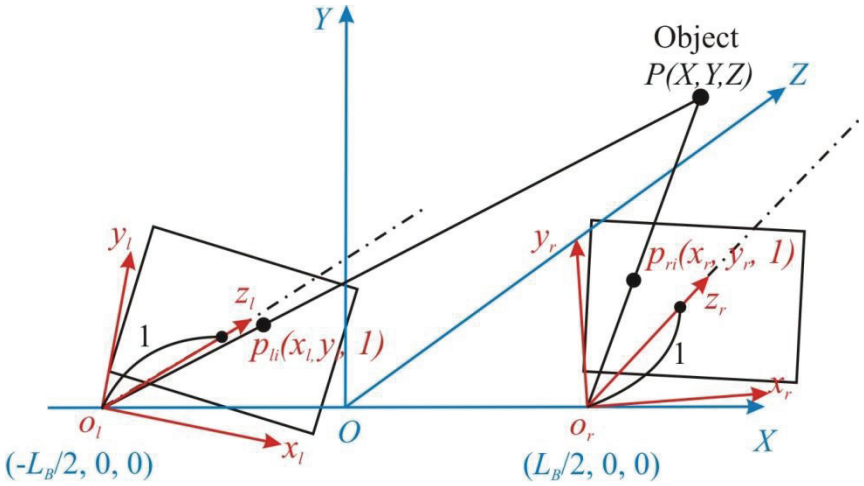


Fig.3-2 Experimental stereo configuration

3.2 Ship Detection Algorithm

The ship detection algorithm used in this study is explained elaborately in (Yamamoto and Win, 2007). The images used in this study are RGB color images. At first step, the points with the high intensity gradient are assumed to be salient points and these points are searched only in the red channel of left stereo image. In Fig.3-3 an intensity profile of red channel of an RGB image scan line is illustrated. In this image the solid line is the intensity values through horizontal pixel locations and the dashed line is the moving average of intensity values. Then candidate salient points are determined as the intersection of moving average values and the intensity values (points A, B and C in Fig.3-3). However, the intensity values are very unstable and this result in too many candidate points. To overcome this problem the area formed by the intersection of the intensity values and moving average line is considered which is shown as blue area in Fig.3-3. If the area before and after intersection points are large enough they are determined to be a candidate point (points A and B in Fig.3-3) which means the intensity change at this point is high enough. In our implementation the areas greater than 30 pixels are considered to be interest points.

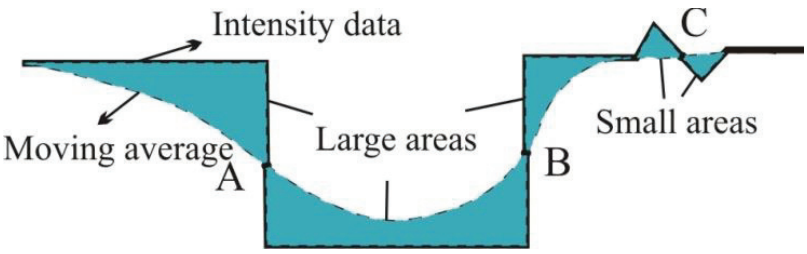


Fig.3-3 Detection algorithm using intensity profile

3.3 3D Measurement

After the camera calibration is carried out and the orientations of cameras are obtained, three-dimensional measurement of a detected ship point 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 coordinates is extracted or detected in one of the 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 (x_l, y_l). The next step is finding the corresponding point in the right image. The rectification of the stereo images is carried out to reduce the corresponding point search to 1D. An example of rectification result is shown in Fig.3.4.

(a)

(b)

Fig.3-4 (a) Original images (b) Rectified images

Before corresponding point search, the color adjustment between left and right images is carried out for reliable corresponding point search. Color adjustment is carried out so that the mean value of each color channel of the right image agrees with that of the left image. For example, the red values of the right image are adjusted by

$$R_{r,new} = R_r + \frac{(R_{l,ave} - R_{r,ave})(255 - R_r)}{255 - R_{r,ave}} \quad (R_r \geq R_{r,ave})$$

$$R_{r,new} = R_r + \frac{(R_{l,ave} - R_{r,ave})R_r}{R_{r,ave}} \quad (R_r < R_{r,ave}) \quad , \tag{3-1}$$

where *Rr* and *Rr,new* are, respectively, the original and adjusted red values of the right image; *Rl,ave* and *Rr,ave*, respectively, are the mean values of the red channel of the left and the right images. Then the corresponding point of this point (x_r, y_r) is searched in the other image through a similarity measure or matching cost called sum of squared differences (SSD) as shown in Fig.3-5. The equation of similarity measure with SSD is as follows:

$$\begin{aligned} R(d) = \sum_{i,j \in W} & ([R_l(i,j) - R_r(i-d,j)]^2 \\ & + [G_l(i,j) - G_r(i-d,j)]^2 \\ & + [B_l(i,j) - B_r(i-d,j)]^2) \end{aligned} \tag{3-2}$$

where *R*, *G* and *B* are RGB intensity values. The subscripts *l* and *r* indicate the reference (left) and observed (right) images, respectively. *W* is the area defined around the candidate point in the reference image. The pixel location *d* with the minimum value of *R* is determined

as the matching point. This is the vital point of the 3D measurement, because the accuracy of the 3D measurement is depending on the accuracy of the corresponding point.

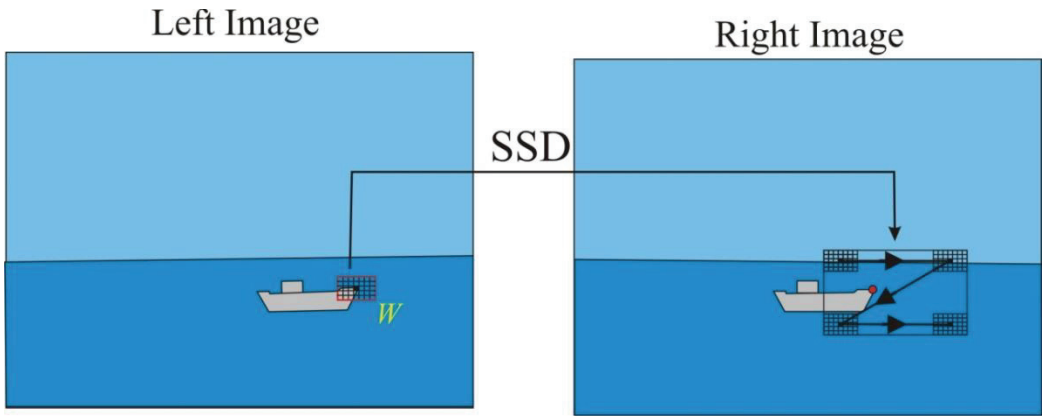


Fig.3-5 Corresponding point search

An experimental sample of stereo corresponding point pair is shown in Fig.3-6.

Fig.3-6 A sample corresponding point matching with SSD similarity measure

As mentioned before the sea images comprise high noise content which results in finding a false point as matching result. For example some sea waves can cause this. To confirm the matching point to be true a reverse matching process is carried out as shown in Fig.3-7. In this case the matched point in the right image is searched in the left image and if the both locations are same this point is determined as true matching point.

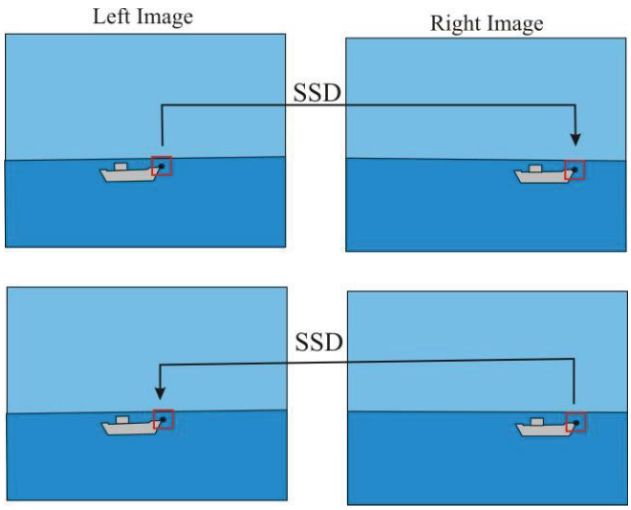


Fig.3-7 Verification of matching

To be able to calculate the 3D location of a point in world coordinates, we use the pixel coordinates of that point (x_l, y_l) and (x_r , y_r) which are the projections on the left image and right image, respectively. Having the pixel coordinates of interested point in both images the rotations of these points are compensated by the previously calculated rotation matrix to obtain the standard stereo configuration in which both cameras become parallel in Z-axis. This is called normalization and the normalized points are denoted as ($^n x_r$,$^n y_r$) for the right camera and ($^n x_l$,$^n y_l$) for the left camera. Besides, the rotation matrix for the right and left images are denoted as $R_r(\theta_{Xr}, \theta_{Yr}, \theta_{Zr})$ and $R_l(\theta_{Xl}, \theta_{Yl}, \theta_{Zl})$, respectively. Then the normalized points are calculated by

$$\begin{bmatrix} {}^n x_l \\ {}^n y_l \end{bmatrix} = \frac{1}{r_{l31}x_l + r_{l32}y_l + r_{l33}} \begin{bmatrix} r_{l11}x_l + r_{l12}y_l + r_{l13} \\ r_{l21}x_l + r_{l22}y_l + r_{l23} \end{bmatrix} \tag{3-3}$$

$$\begin{bmatrix} {}^n x_r \\ {}^n y_r \end{bmatrix} = \frac{1}{r_{r31}x_r + r_{r32}y_r + r_{r33}} \begin{bmatrix} r_{r11}x_r + r_{r12}y_r + r_{r13} \\ r_{r21}x_r + r_{r22}y_r + r_{r23} \end{bmatrix} \tag{3-4}$$

where r_{rij} and r_{lij} represent (i, j)th component of the rotation matrices R_r and R_l, respectively. Finally, the location of the object in the world coordinate system is obtained by the following equation.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{L_B}{{}^n x_l - {}^n x_r} \begin{bmatrix} ({}^n x_l + {}^n x_r)/2 \\ {}^n y_l \\ 1 \end{bmatrix} \tag{3-5}$$

3.4 Tracking Ships through Image Sequences

A point is manually selected to track the ship through the image sequences as shown in Fig.3-8(a). The reason for manually selecting the point is to see the feasibility of tracking after giving a point to track. Here, the ship is represented as one point which can be near bridge because it is the navigation center. This point is selected from the left image. The corresponding point of this point is found and 3D location of the ship is calculated as explained above. Then this point is found in the next left image (time *t*+1) by SSD and the new 3D location at time *t*+1 is calculated as illustrated in Fig.3-8(b). This procedure is carried out through image sequences which result in 3D tracking of ship location.

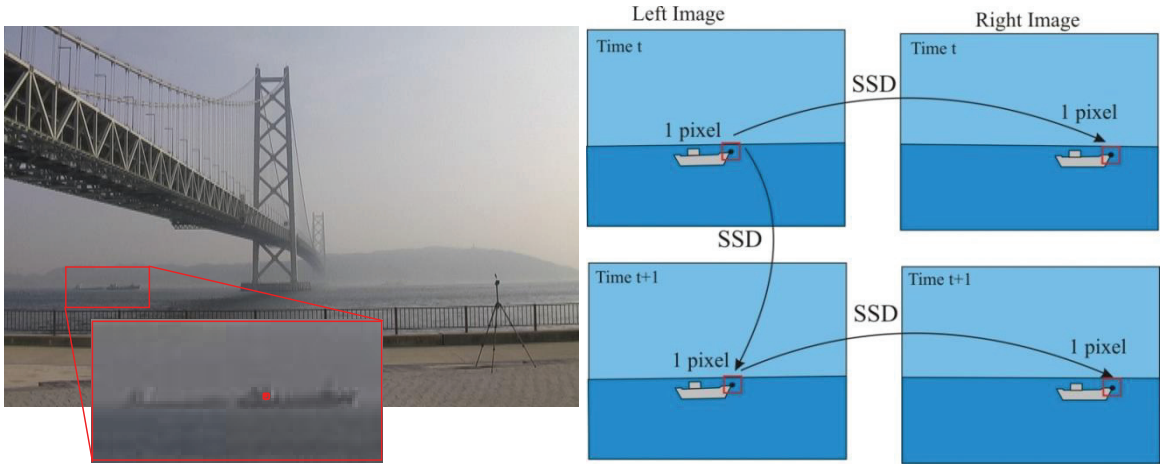


Fig.3-8 (a) Tracking point of ship (b) tracking through image sequence

So far, all the points are calculated in one pixel accuracy. Therefore, the disparity is also in one pixel accuracy. The tracking result of 3D locations of ships with one pixel accuracy calculation is shown in Fig.3-9. The blue dots are showing the 3D location of the ship for a time *t*. The ship is moving from left side of the image to the right side of the image (through +X direction). The ship is about 1.5 km far from the cameras.

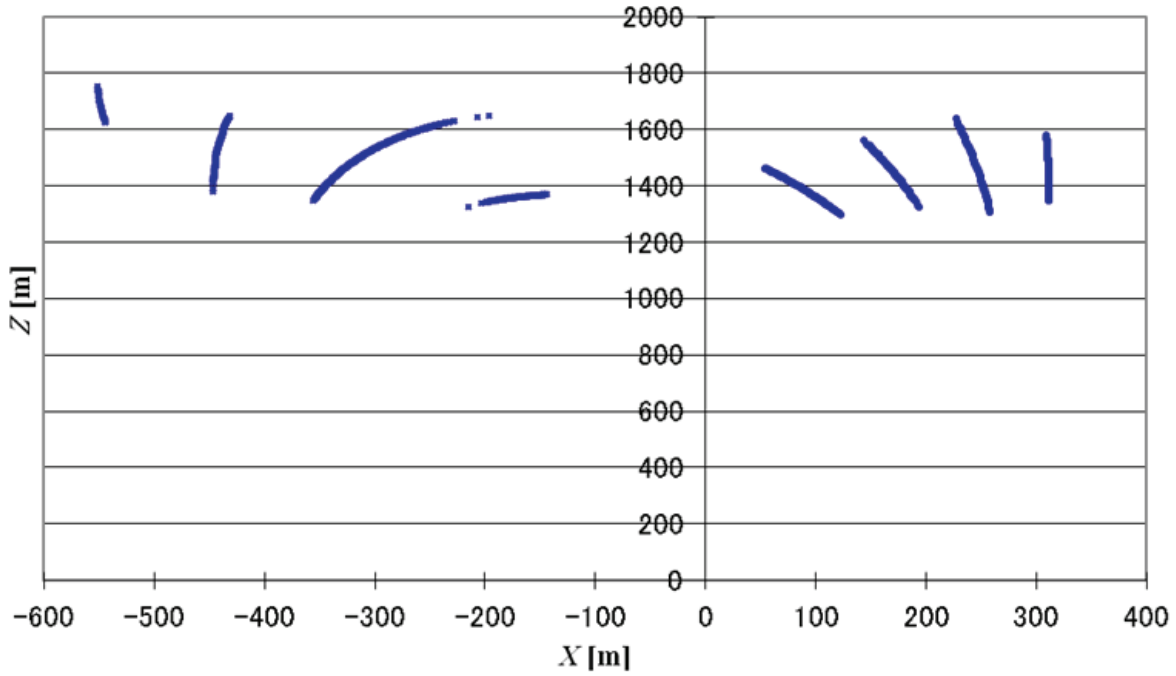


Fig.3-9 Tracking of ships with one pixel accuracy

It is seen in the figure that the ships trace is not continuous and there are certain gaps periodically. It means that the 3D location of the ship is not found truly for every point.

The basic study explained so far shows that it is possible to detect the objects, measure the distance of detected of objects with some accuracy and track a given point through image sequences. However, it needs to be improved which is done in my study. First of all, the 3D calculation accuracy is considered because the 3D location results were not satisfying. The main reason of this situation is insufficiency of one pixel level matching. The image represents the very dense real world data with a few pixels which result in too much loss of data. Especially images of far points are squeezed into one pixel of an image. One of the solutions to overcome this problem is finding the sub-pixel locations of points. Then 3D location should be calculated using these sub-pixel locations. The sub-pixel matching algorithm improved the results. Then we utilized the sequences of image data for further improvement of the 3D location calculation. In the sequences of images, the calculated 3D points of images at time t and time t+1 should be close to each other. But due to some vision problems they will not be very close and it will be a little bit noisy. We applied Butterfly low-

pass filter to improve the calculation which is a kind of averaging of data through the sequences of images.

Another problem is noisy sea images which are mainly due to specular sea surface. Some different smoothing filters are analyzed and an edge preserving smoothing filter is found to be more effective especially for detection purpose.

In the basic study the tracking is done manually by choosing a ship point in the beginning. It may be useful for a semi-automatic system in which the watch-keeping officer firstly selects a ship point manually. However, we would like to automate this process. It is done by clustering the detected points to represent fewer points to track and the center of these clusters are tracked automatically through the image sequences by a spatio-temporal clustering algorithm.

Another point is the environmental effects such as ship rolling and pitching to the camera system. This will affect the images obtained by the cameras and the geometric relation between camera and world coordinates will be changed. We proposed a system to calculate the ship motion utilizing the sea-sky line and eliminating the effect of ship motions such as rolling and pitching.

The algorithms and results are explained in the following chapters.

CHAPTER 4

ACCURACY IMPROVEMENT OF 3D LOCATION MEASUREMENT

4.1 Introduction to 3D Measurement Error

The images have finite number of pixels. The visual information of a wide area is compressed into this finite number of pixels. That's why some data is lost. This results in quantization error which is due to lack of image plane quantization (Blostein and Huang, 1987). This is one of the reasons of correspondence problem that the corresponding points can be represented in one pixel accuracy which is not sufficient. This affects 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, 3D 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 decreases (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 Fig.4-1.

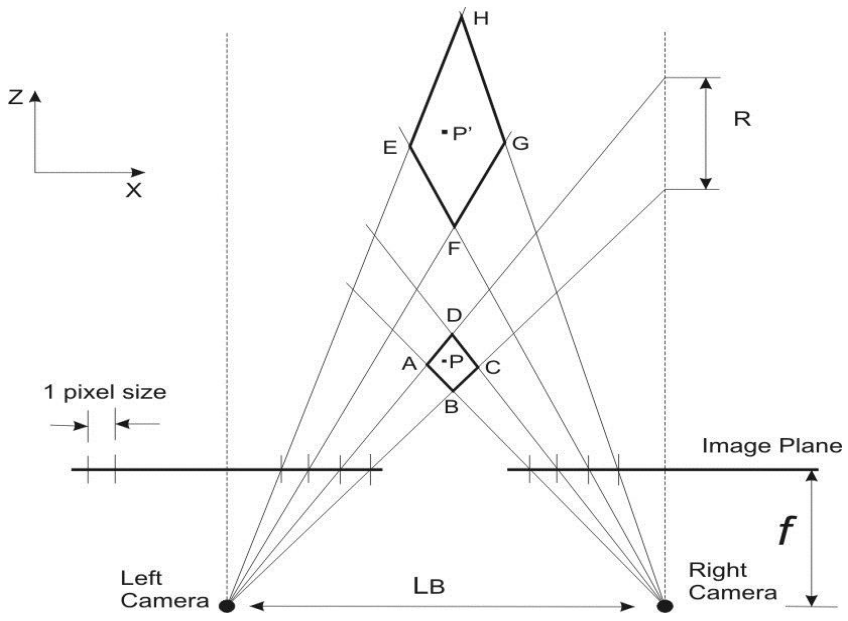


Fig.4-1 Image plane quantization, region of uncertainty and depth resolution (R)

Considering an ideal pinhole camera model and the quantization error, measured position of a scene point may be somewhere in a 3D volume due to finite resolution of the cameras which is called region of uncertainty (Blostein and Huang, 1987). In the Fig.4-1 ABCD quadrilateral is the X-Z direction view of region of uncertainty for point P. The exact location of the scene point can be anywhere in this region. As the scene point goes further in Z direction the region of uncertainty becomes greater. If the object is moved to point P' from point P, the region of uncertainty is increasing to quadrilateral of EFGH. In the case of long range object points, the spaciousness of region of uncertainty can be imagined. This causes greater errors in 3D recovery of long range points.

The accuracy of the 3D measurements with a stereo system is inversely proportional with the distance of the object. The stereo measurement error is discussed here. In the case of the ship is at the position of $P(0,0,D)$, calculation will be $P'(0,0,D+\Delta D)$ if each left and right cameras have an orientation error of $\Delta\alpha$. Fig.4-2 illustrates such kind of error situation.

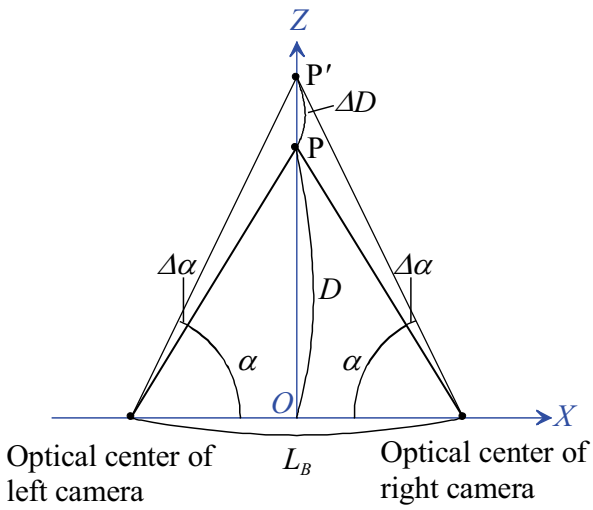


Fig.4-2 Error in case of $P(0, 0, D)$ [Courtesy of S. Yamamoto]

From the figure, D is calculated as

$$D = \frac{L_B}{2}\tan\alpha \tag{4-1}$$

Therefore contributing the error to the equation, it becomes

$$D + \Delta D = \frac{L_B}{2}\tan(\alpha + \Delta\alpha) \tag{4-2}$$

Equation (4-3) can be derived from (4-1) and (4-2).

$$\frac{\Delta D}{D} = \frac{(1+4k^2)\tan\Delta\alpha}{2k(1-2k\tan\Delta\alpha)} \tag{4-3}$$

where k is D/L_B.

Fig.4-3 illustrates the relation between k and the percentage error when $\Delta\alpha$ is equal to θ/w which means that $\Delta\alpha$ is the orientation error in case of 1 pixel detection error on image plane. Here, θ is the camera view angle and w is the horizontal resolution of the image. Therefore, to measure the location of an object with a camera of 90° view angle within 20% 3D error, the distance of the object should be not more than 100 times of the baseline length.

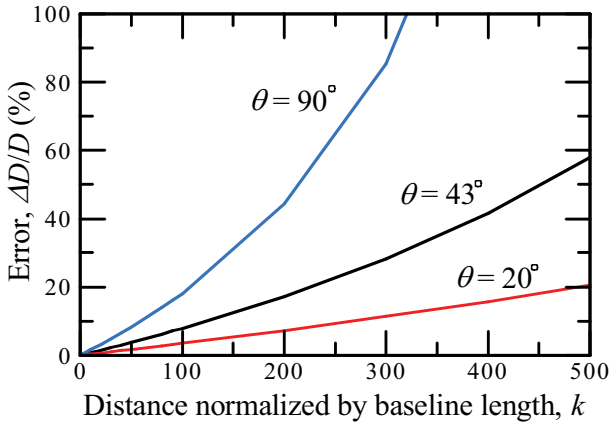


Fig.4-3 Relation between distance and error [Courtesy of S. Yamamoto]

It shows the difficulty of measuring 3D location of far objects accurately. Considering a case that a baseline (L_B) is 20m and an object distance (D) is 2000m, the calculated α is 89.7135° which is almost right angle. From this, it is easier to understand that the orientation error is expected to be quite small. The effect of very small angular error ($\Delta\alpha$) on 3D calculation using Equation (4-3) is shown in Fig.4-4. We can see from the figure that 0.2° of orientation error result in 4626m of depth error when the object is at 2000m distance and

baseline is 20m. And the 3D error is 6025 meters when the baseline is 8,14 meters and object distance is 1000 meters.

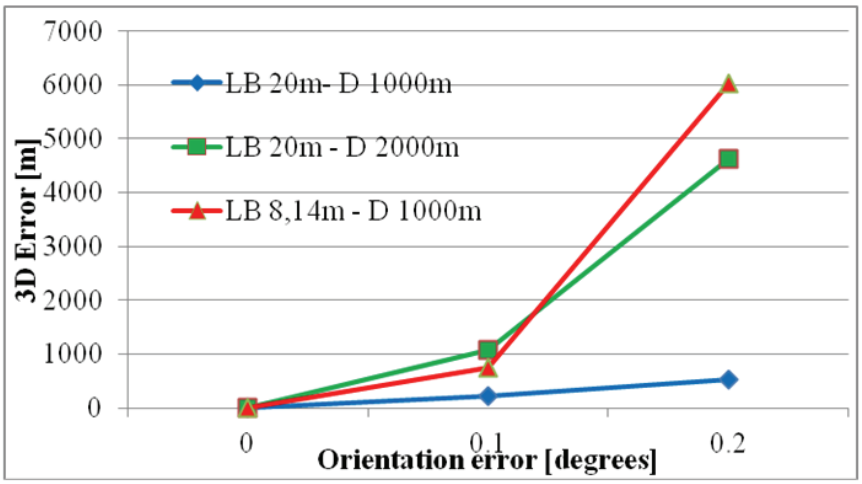


Fig.4-4 Effect of camera orientation error on 3D calculation

As a result, it is seen that the 3D calculation of far objects using stereo images is actually very difficult because even very small orientation error results in a large location error.

4.2 Concept and Estimation of Stereo System Error

The insufficiency of the one pixel level representation of points is shown in the previous Section. We need sub-pixel locations of image points to improve the correspondence quality. Following is theoretical explanation of error caused due to one pixel level matching.

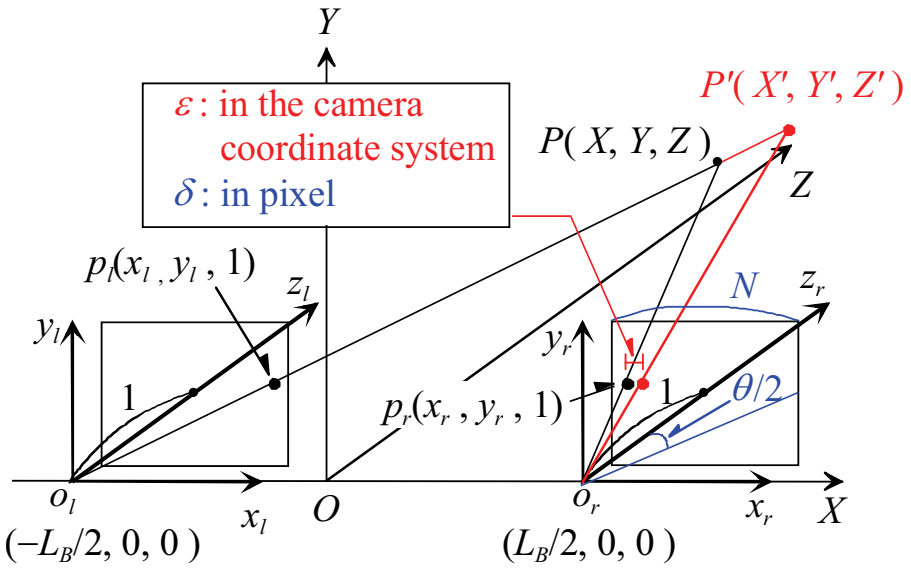


Fig.4-5 Concept of stereo 3D error due to correspondence problem

Fig.4-5 shows a standard stereo camera configuration 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*.

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 (X, Y, Z) coordinate of object is obtained with one pixel accuracy by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{L_B}{x_l - x_r} \begin{bmatrix} (x_l + x_r)/2 \\ y_r \\ 1 \end{bmatrix} \tag{4-4}$$

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 (4-4), the measured erroneous coordinate *(X', Y', Z')* is obtained by the following equation:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \frac{L_B}{x_l - (x_r + \varepsilon)} \begin{bmatrix} (x_l + x_r + \varepsilon)/2 \\ y_r \\ 1 \end{bmatrix} \tag{4-5}$$

From Equation (4-4) the following equations are extracted.

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

$$x_l - x_r = L_B / Z$$

$$y_r = Y / Z$$

Combining these equations with Equation (4-5) yields:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \frac{Z.L_B}{L_B - \varepsilon.Z} \begin{bmatrix} (\frac{X}{Z} + \frac{\varepsilon}{2}) \\ Y/Z \\ 1 \end{bmatrix} \tag{4-6}$$

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

$$\begin{bmatrix} X' - X \\ Y' - Y \\ Z' - Z \end{bmatrix} = \frac{\varepsilon Z}{L_B - \varepsilon Z} \begin{bmatrix} L_B/2 + X \\ Y \\ Z \end{bmatrix} \tag{4-7}$$

While 3D reconstruction of an image point the data is obtained in pixel coordinates first and then transformed to camera coordinates and world coordinates. That's why it is required to express the error in terms of the data of pixel coordinates, as well. From Fig.4-6 this process can be better understand.

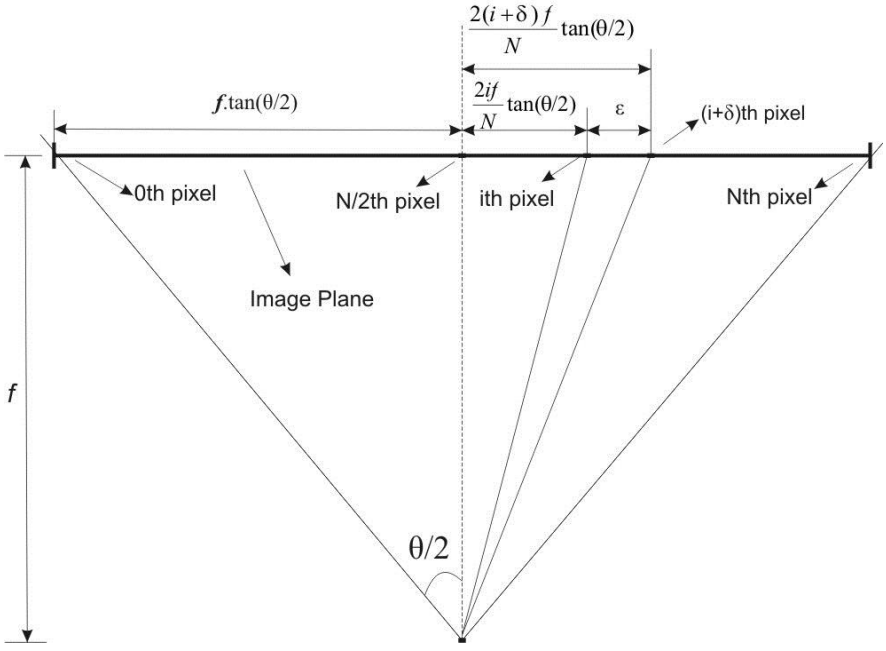


Fig.4-6 Correspondence error on image plane

Considering δ is the amount of shift in pixel coordinates due to correspondence error, the difference of i^{th} pixel and $(i+\delta)^{th}$ 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

$$\varepsilon = \frac{2\delta}{N}\tan\frac{\theta}{2} \tag{4-8}$$

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 Fig.4-7 the estimated error results for different parameters can be seen. We can figure out from the figure that increasing the horizontal resolution and decreasing the camera view angle result in decrease of the error. But it can be designed depending on the purpose. A wider view angle or higher resolution may be desired up to some situations but the computational complexity should be considered as well. That's why there is a trade-off between the parameters.

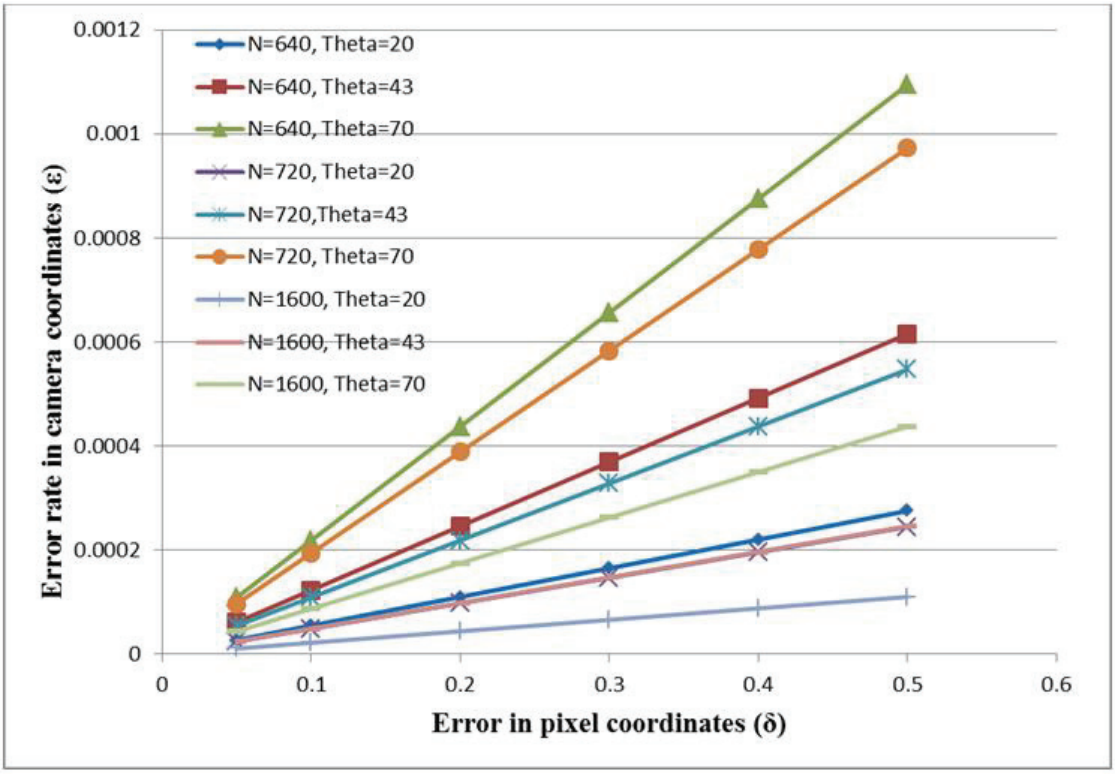


Fig.4-7 Stereo error in pixel and camera coordinates in different cases of view angle(theta) and resolution(N)

4.3 Matching with Sub-pixel Accuracy

Sub-pixel locations of the corresponding points can be calculated using similarity measure as a solution to decrease quantization error. In this study, similarity interpolation method of Shimizu and Okutomi is implemented which estimates the peak of a similarity function by parabola fitting (Shimizu and Okutomi, 2005). While estimating sub-pixel displacement through similarity interpolation, Sum of Squared Differences (SSD) between left and right images are calculated and the minimum value is selected as center similarity value $R(0)$. The preceding similarity becomes $R(-1)$ and the next is $R(1)$. The SSD values of $R(0)$, $R(-1)$ and $R(1)$ are calculated according to the following formula:

$$\begin{aligned} R(s) = \sum_{i,j \in W} & ([R_l(i,j) - R_r(i-d+s,j)]^2 \\ & + [G_l(i,j) - G_r(i-d+s,j)]^2 \\ & + [B_l(i,j) - B_r(i-d+s,j)]^2) \end{aligned} \tag{4-9}$$

where s is the shift value in pixel unit (-1, 0 or 1) from the corresponding point and d is the disparity. For instance, for s=1 $R(1)$ is obtained. Then, a parabola fitting is carried out whose minimum is the estimated sub-pixel position. Fig.4-8 is illustrating this situation.

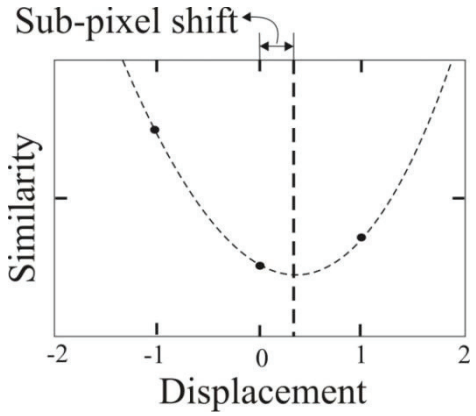


Fig.4-8 Parabola fitting of similarities

Following formula is the parabola fitting used for the calculation of sub-pixel shift ($\hat{d}$):

$$\hat{d} = \frac{R(-1) - R(1)}{2R(-1) - 4R(0) + 2R(1)} \tag{4-10}$$

where $R(-1)$, $R(0)$ and $R(1)$ are the similarity values obtained from Equation (4-9).

4.4 Tracking the Ships with Improved Disparity

After the corresponding point is found the sub-pixel location of the corresponding point is calculated as explained above and the 3D calculation is carried out using this sub-pixel information. The 3D location of ship is tracked through the image sequences as shown in Fig.4-9.

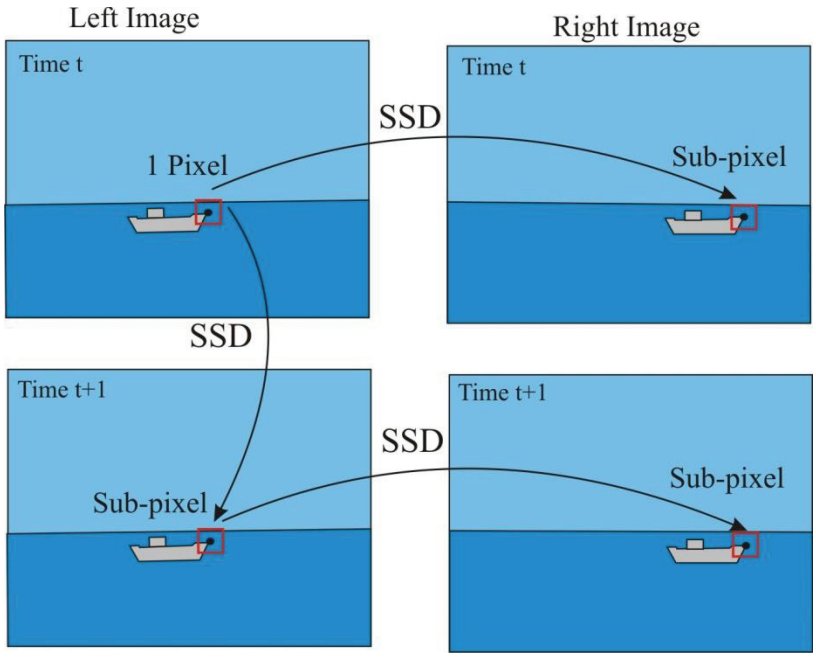


Fig.4-9 Tracking with sub-pixel accuracy

Sub-pixel matching is very effective method for decreasing the 3D error. In this study the detection range is quite far (around 2 nautical miles) and more accuracy is needed. In case of sub-pixel matching only a pair of stereo image set is sufficient. But we have a sequence of images in time series which can be utilized to decrease the error. The 3D calculation is based on the disparity of the same point in left and right images. In case of sequences of images, the disparity values of a tracked point through the sequence should be consistent. But due to the before mentioned problems of vision systems such as insufficiency of number of pixels (quantization problem) some noisy results will be obtained. This noise can be reduced by using the disparity data of many images in time series which is called disparity refinement.

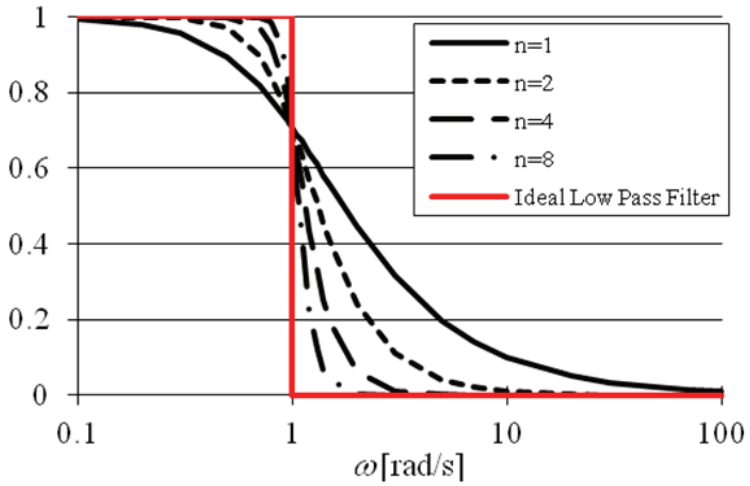


Fig.4-10 Butterworth low pass filter response (ω_c=1)

Disparity refinement is one of the method to achieve a more complete and accurate scene interpretation (Skulimowski and Strumillo, 2008; McKeown and Perlant, 1992). The correspondence data (or disparity) between stereo images is a kind of low frequency information (Frankot and Chellappa, 1988). Therefore, removing the high frequency information clears the disparity data from noises which result in a refined disparity data. A 2nd order Butterworth low pass filter (BLPF) is used for eliminating the noisy data from the measured disparity values which is actually taking the average value of disparities calculated in image sequence. The response of Butterworth LPF can be seen at Fig.4-10.

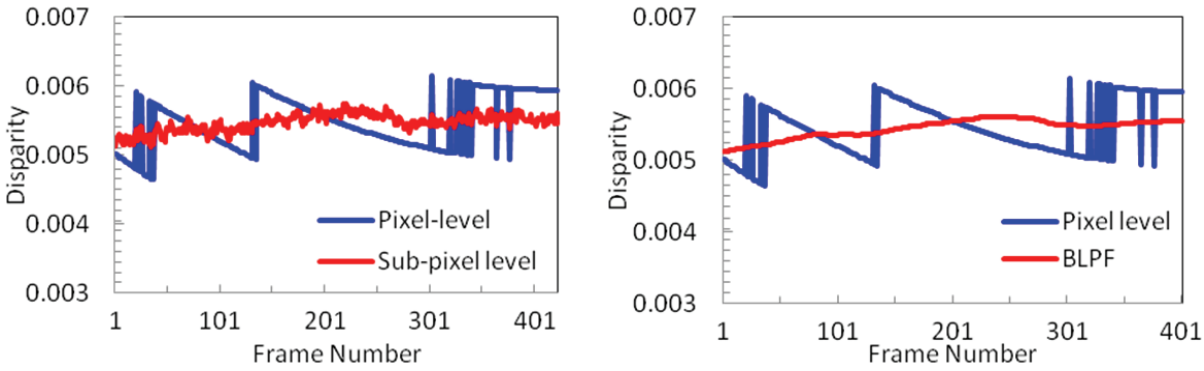
The transfer function of a BLPF is defined as:

$$|G(w)|^2 = \frac{1}{1+\left[\frac{\omega}{\omega_c}\right]^{2n}} \quad , \tag{4-11}$$

where ω_c is cut off angular velocity, and *n* is the filter order (Gonzales and Woods, 2008).

The effect of sub-pixel matching and low pass filtering on disparity data is shown in Fig.4-11. In the figure the blue line is the result of one-pixel level calculation of disparity through sequences of images. It is seen that the disparity values of one pixel level calculation is not consistent. In Fig.4-11(a) the result of the sub-pixel matching on disparity data is seen. The disparity data became much more consistent. However, it can be still improved. The result of

disparity refinement is shown in Fig.4-11(b) that it is smoother than sub-pixel disparity results.



(a) (b)

Fig.4-11 Effect of (a) sub-pixel matching and (b) low pass filtering on disparity

The 3D location tracking of one-pixel accuracy, sub-pixel accuracy and low pass filtered situations are shown in Fig.4-12. The Akashi Bridge is lying through the zero point of the *X* dimension. In the figure we can see that the one-pixel accuracy calculation is not continuous which result in great 3D calculation error while sub-pixel and low pass filtered data is continuous.

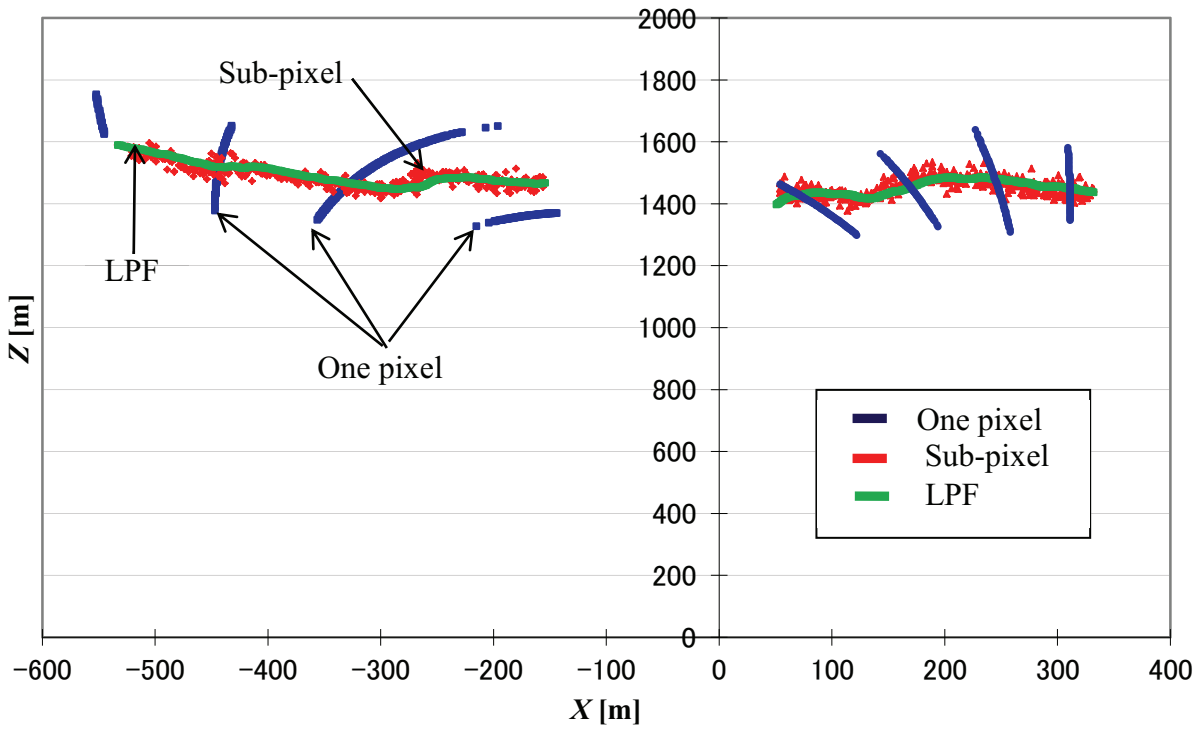


Fig.4-12 Improvement of 3D ship location tracking

In order to reduce errors, corresponding points between a pair of stereo images should be as accurate as possible. In other words, it is important to reduce the value of δ in Equation (4-8). Sub-pixel matching and low pass filtering are applied to decrease this error. The localization of a ship is shown in Fig.4-12. In the figure, the location of the ship's bridge is traced. The blue rectangles are results obtained by stereo pair matching in one-pixel unit, while the red triangles are those by sub-pixel matching method. The green line shows the smoothing result of sub-pixel matching by a low pass filter. *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.

There is one point we have to explain about the actual positions of ships. We attempted to measure the actual positions of the ships by using a surveying equipment (Topcon GPT-7500). However, this was not possible because the ships were in motion. Therefore, a stable point on the Akashi bridge is measured to show the accuracy of the 3D measurement of the stereo camera system. The projections of this point to the left and right images are shown in Fig.4-13. The 3D location of this point measured by the surveying equipment is (−108.694m, 193.502m, 917.499m) which are the *X*, *Y*, and *Z* distances, respectively. 3D measurement of this bridge point is performed by the stereo camera system for 50 images through image sequences. The average of these 50 calculations is (−110.733m, 195.06m, 930.171m). A comparison of the 3D measurement result of the surveying equipment and stereo camera system shows that the accuracy of the 3D measurement of the stereo camera system is acceptable.

A fitting line to the sub-pixel matching results is assumed to be the ships true course. 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.

Fig.4-13 The bridge point measured for 3D accuracy

The ships tracking data for 26 seconds, 260 frames are examined to evaluate the accuracy. As a result, the errors of one-pixel unit matching range from about −170 meters to about 210 meters. The values will theoretically become from -140 meters to 170 meters using Equations (4-7) and (4-8) 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 from −60 to 55 meters, and the low pass filter can reduce them to a range from −2 to 27 meters. This result indicates that sub-pixel matching with the low pass filter is at least 5 times more accurate than one-pixel unit. Therefore, it can be concluded that corresponding points between stereo pairs can be obtained with an error of ±0.1 pixels.

Based on the above arguments, the accuracy that can be realized is estimated as follows. Let N be 720 pixels, δ be 0.1 pixels from the above discussions, θ be 45 degrees. Then, since L_B is much larger than εZ when Z is less than 10,000 meters, and L_B is 8,14 meters, Equation (4-7) is approximated as

$$\frac{\Delta Z}{Z} = 1.1 \times 10^{-4} \times \frac{Z}{L_B} \tag{4-12}$$

Fig.4-14 illustrates the above results. For example, if the baseline length L_B can be set to be 20 meters, we can measure an object up to 5km with an error less than one per cent.

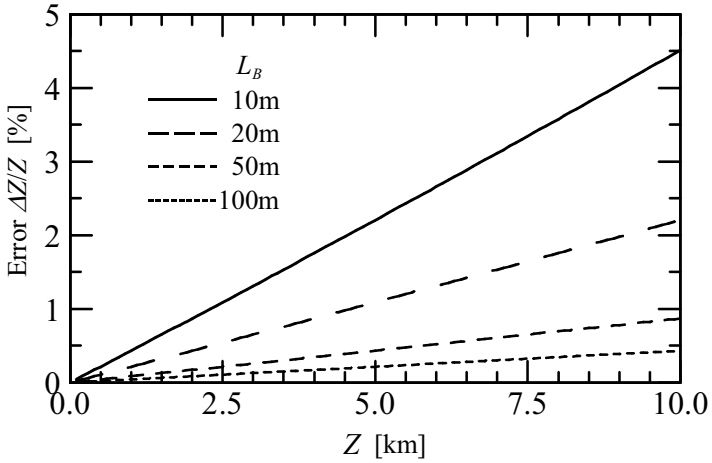


Fig.4-14 Estimated error

CHAPTER 5

COMPARISON OF SOME SMOOTHING FILTERS FOR REMOVING NOISE

5.1 Noise Filtering as Pre-process

One of the disadvantages of sea images is containing high noise due to the specular property of sea surface. We see this by detecting the edges. Edge detectors respond to local changes in images. The edges are the important information of an image especially for computer vision tasks such as object detection. Despite the sky and sea surface visually appear so similar, edge detection results are completely different. There are many edges in the image of sea surface which are mostly noise. Mostly used edge detectors are Sobel, Prewitt, Roberts and Canny edge detectors. Sobel edge detector is used in this study due to its simplicity and effectiveness. Horizontal and vertical edges are obtained by applying the convolution masks displayed in Fig.5-1.

-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1

Fig.5-1 Sobel convolution mask for horizontal (left) and vertical edges

Fig.5-2 Edge detection by Sobel filter

The result of Sobel edge detection on sea images is shown Fig.5-2. It can be clearly seen that there are abundant edges on sea surface which are noise while only a few edges detected on sky. This is the result of glares on sea surface.

During detection of the objects on the sea surface these edges of sea surface is also considered as a part of object which is an undesired situation. That's why elimination of this noise through a smoothing filter is required as a pre-process. There are various smoothing filters applied in image processing. The effects of following filters are considered:

1. Median Filter (7x7)
2. Gaussian Filter (3x3)
3. Multi-scale Gaussian Filter
4. Bilateral Filter

Median filter is a non-linear filter which is eliminating the noise by taking the median of the sorted intensity data. Gaussian filter is applying 2D Gaussian kernel function to pixel intensity values of the images which have a blurring effect. In image processing Gaussian filter masks are obtained from Gaussian kernel to simplify the process. The Gaussian mask used in this study is shown in Fig.5-3. Multi-scale Gaussian filter is recursive application of Gaussian filter. Applying the Gaussian filter 2 times results in 2^{nd} scale Gaussian filter and 3 times result in 3^{rd} scale Gaussian filter.

1	2	1
2	4	2
1	2	1

Fig.5-3 3x3 Gaussian filter mask

The above mentioned filters are smoothing the image regardless of edge data. Bilateral filter is smoothing the image noise while preserving the edges (Tomasi and Manduchi, 1998). The bilateral filter is a non-linear filter where output image J_s is a weighted average of the

input image *I* and the weighting for each pixel *p* is determined by the spatial distance and the intensity difference between center pixel *s* and pixel *p* within window Ω. Therefore, the value of pixel *s* is affected mainly by pixels that are spatially closer and have a similar intensity value. The output image is obtained by

$$J_s = \frac{\sum_{p\in\Omega} f(p-s)g(I_p - I_s)I_p}{\sum_{p\in\Omega} f(p-s)g(I_p - I_s)} \tag{5-1}$$

where the function *f* for spatial domain and *g* for intensity domain are usually Gaussian functions. The *f* and *g* functions used in bilateral filter are:

$$f(p-s) = e^{-\frac{1}{2}(\frac{p-s}{\sigma_d})^2} \tag{5-2}$$

$$g(I_p - I_s) = e^{-\frac{1}{2}(\frac{I_p - I_s}{\sigma_r})^2} \tag{5-3}$$

After removing the noise in the images by mentioned filters, the detection algorithm explained in Chapter 3 is carried out and the salient image points are detected. The 3D locations corresponding to these image points are calculated through the stereo theory explained before.

5.2 Effect of Smoothing Filters on Sea Images and Ship Detection

The results of Sobel edge detection after application of median filter, Gaussian filter, 2nd, 3rd, and 4th scale Gaussian filters and bilateral filter are shown in Fig.5-4. The size of the filter mask for median filter is 7x7 and the Gaussian filter is 3x3. The greater size of filter results in greater effect of smoothing. But the number of calculations will be greater which will slow down the algorithm. In the figure the most effective filter to remove noise is seen to be median filter. The effect of the Gaussian filter is increasing by increasing the scale of the filter because it is same as increasing the size of the filter. The bilateral filter is also effective even if it is not as good as median filter. In this study we also consider the detection purpose in which filtering has important effect.

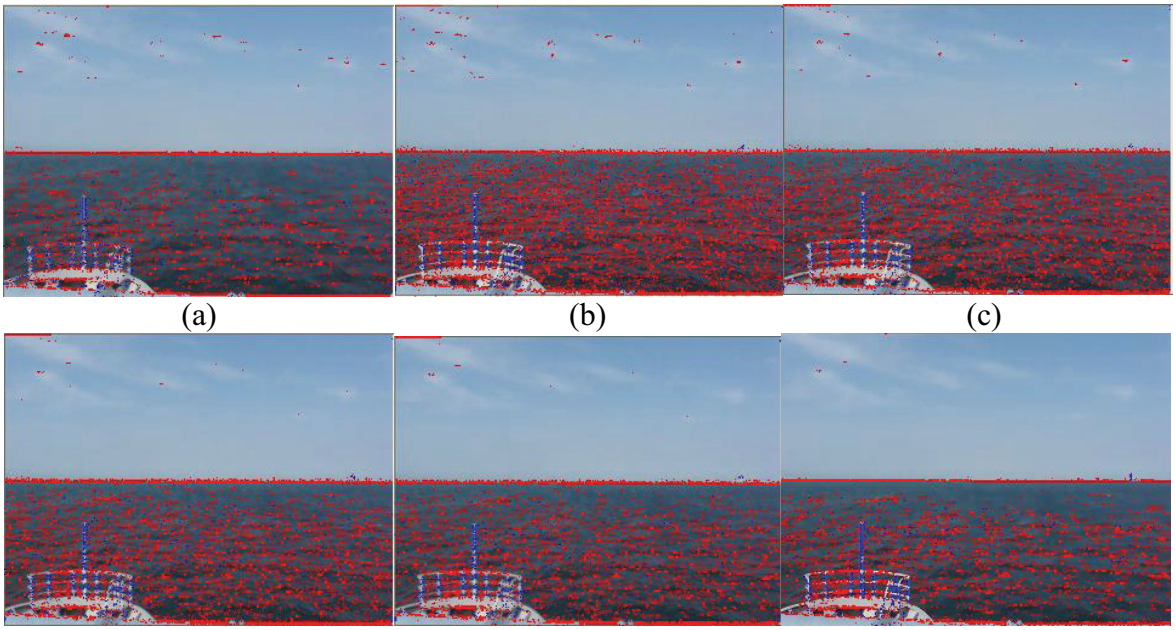


Fig.5-4 Effect of smoothing on edge detection (a) Median filter (b) Gaussian filter (c) 2nd scale Gaussian (d) 3rd Gaussian (e) 4th scale Gaussian (f) Bilateral filter

In Fig.5-5 the detection algorithm of a real sea image is illustrated. The vertical axis shows the intensity values. The intensity value of an image may span between 0 and 255. The horizontal axis is the horizontal pixel location (1 to 720). The blue line is the intensity profile of one scan line and the red line is moving average. The green areas are the formed areas to be considered for saliency and the black points are candidate salient points.

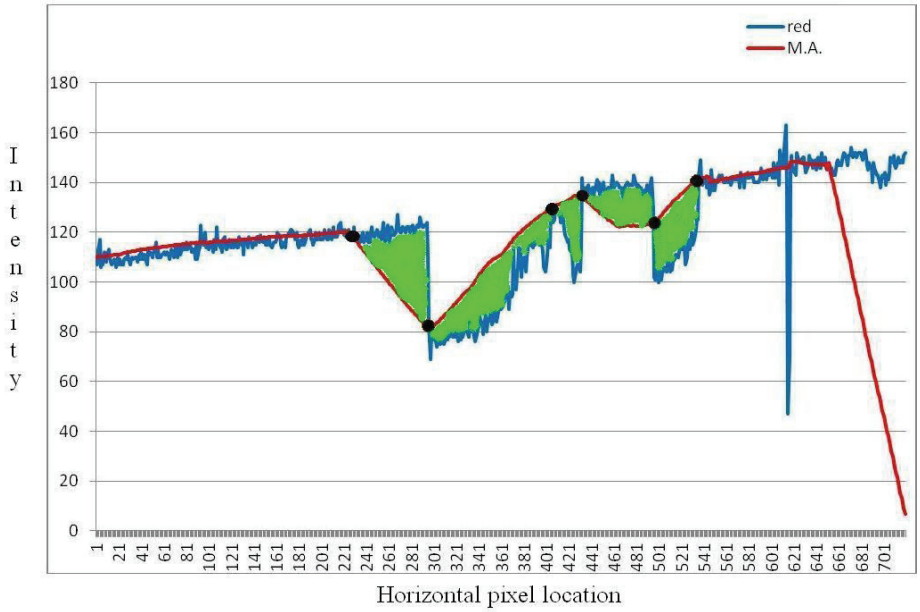


Fig.5-5 Detection algorithm on a real sea image

Smoothing the images is an important step. However, some edge data is lost while removing the noise. That's why a suitable smoothing filter should be preferred. The results of two types of smoothing filters to ship detection are compared from the previously mentioned filters due to their efficiency. One of them is 7x7 median filter and the second one is bilateral filter. The results of median filter and bilateral filter on a ship image are shown in Fig.5-6. While the original image is noisy, the filtered images are smoothed. However, in the median filtered images the edges of ship hull are deteriorated which results in difficulty of detection. Besides in bilateral filtered image the ship shape is preserved while the noise is removed.

Fig.5-6 Effect of filtering on ship image a) Original b) Median filtered c) Bilateral filtered

Application of different kind of smoothing filters affects the intensity profile and hence detection quality. The intensity profile of a real sea image and its filtered forms are shown in Fig.5-7. Actually, only one scan line is not enough to show the total effect but only some local effects are displayed to show the algorithm.

The intensity values of a scan line are changed after the application of smoothing filter which affects the detection quality. The scan line after application of median filter and bilateral filter are shown in Fig.5-7. In the figure there are a bridge pier and two ships. One of the ships is closer and has a clear image while the other is farther and not so clear. The important point for detection is shown with the black circle that the median filtered image cannot detect the salient point while the bilateral filtered image can. Besides, the median filter is truncating the sharp edges while the bilateral filter preserves them.

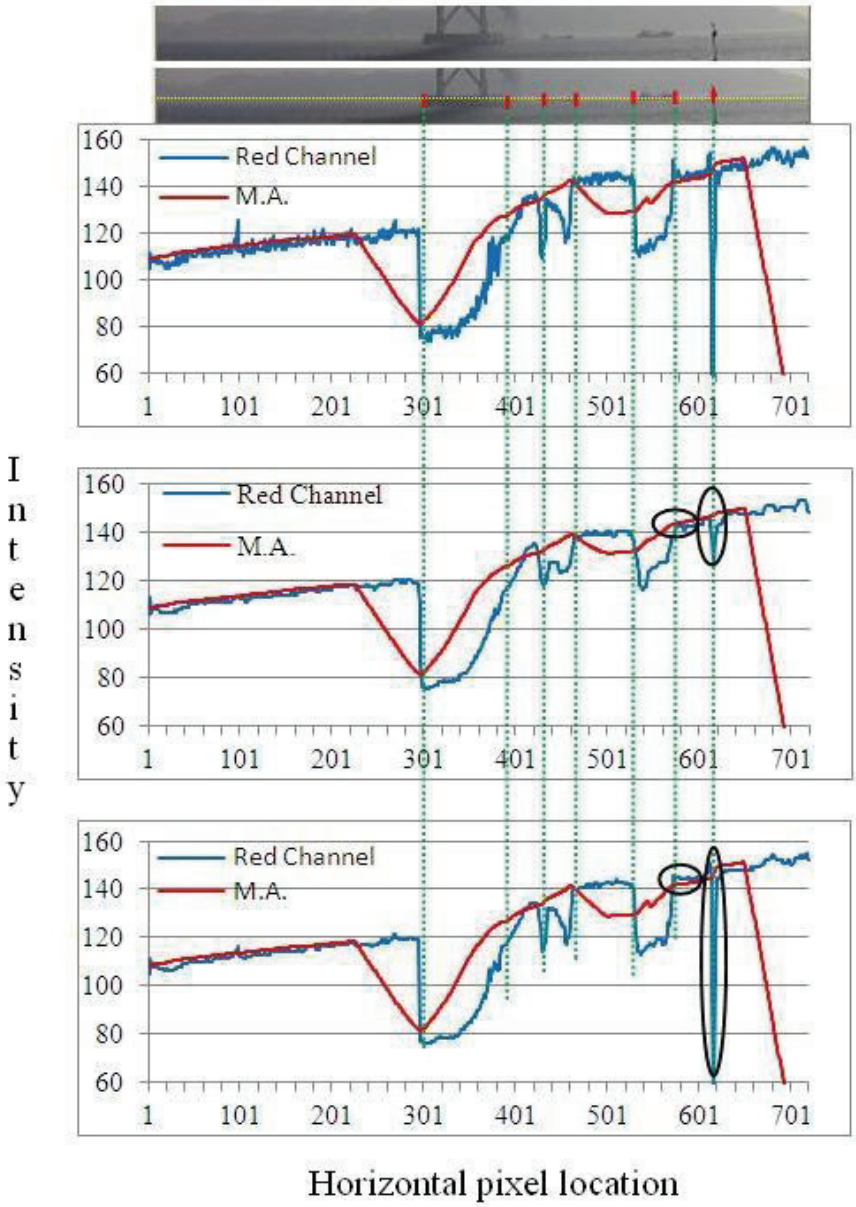
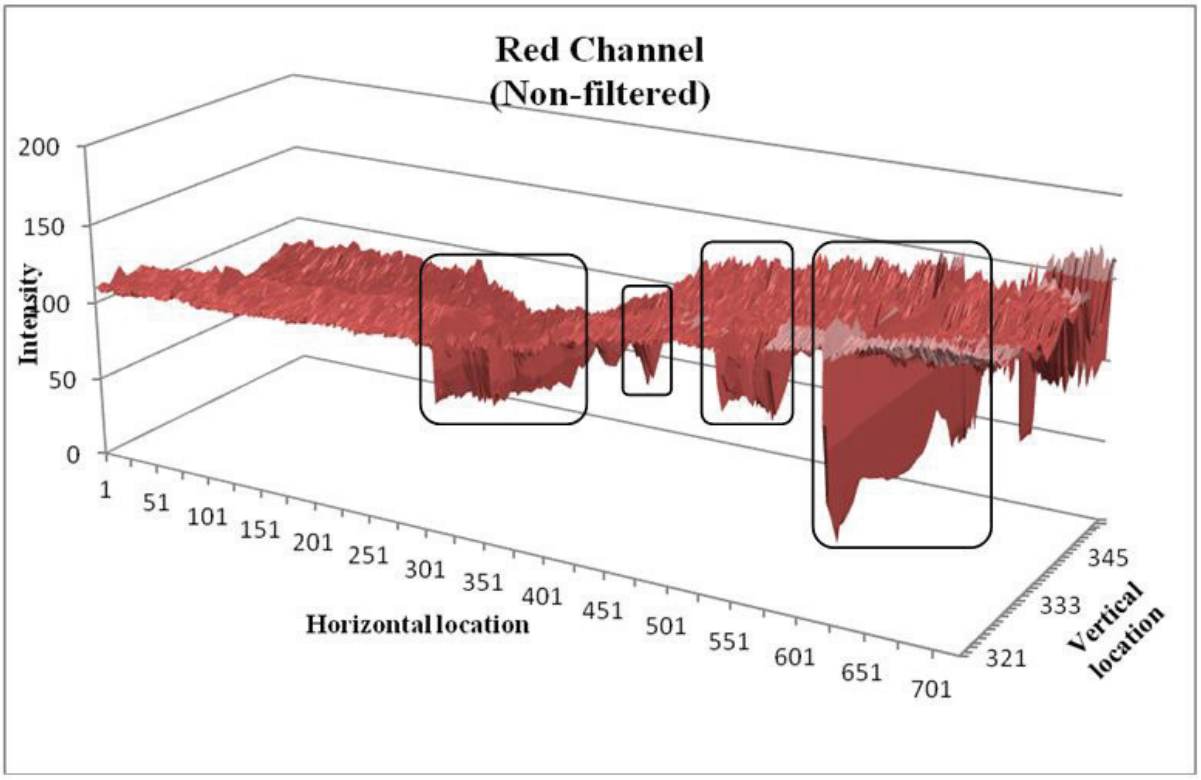


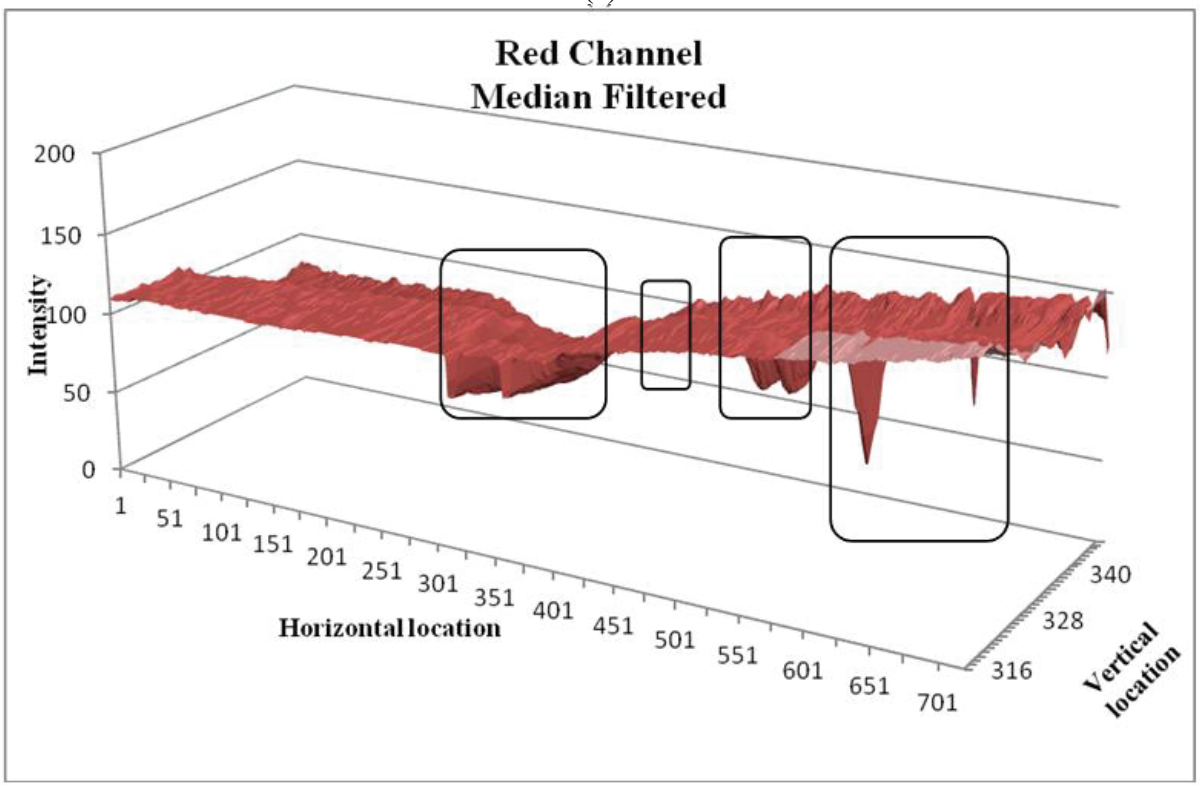
Fig.5-7 Intensity profile of a sea image. Top: Non-filtered, Middle: Median filtered, Bottom: Bilateral Filtered

The intensity profiles of several scan lines are better to understand the smoothing effect. The scan lines between 320 and 355 (vertical image location) of image shown in Fig.5-7 are displayed in Fig.5-8 with the filtered versions. In the non-filtered intensity data which is displayed in Fig.5-8(a) the surface is generally so noisy that there are many spikes. The edge points are shown by black rectangles on the image where there are sudden and big changes in intensity data. In Fig.5-8(b) the effect of median filter is seen. The surface is generally much smoother but some of the data in the edge locations are lost which are also shown with black

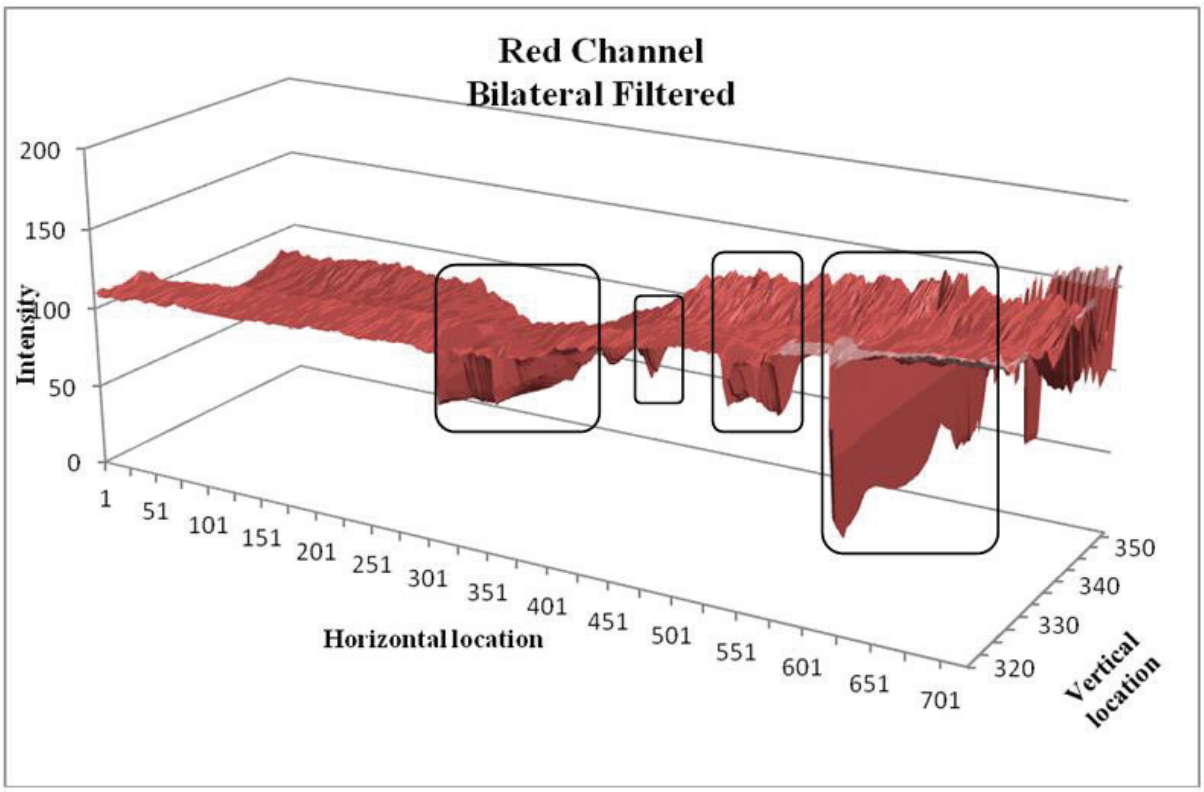
rectangles. In Fig.5-8(c) the effect of bilateral filter is shown. In the figure the surface is smooth and the edges are preserved. It can be seen by comparing the black rectangles.



(a)



(b)



(c)

Fig.5-8 Intensity profiles of several scan lines for original and filtered images

The detected salient points are shown in Fig.5-9. It can be seen from the figure that bilateral filter is improving the detection quality. The detected ships are shown by yellow rectangles. Especially the ship which is on the left side is farther and the edges are weaker. This ship is better detected after bilateral filter. The positive effect of bilateral filtering can be seen more clearly on the tracking data which will be explained in Chapter 6.

(a) (b)

Fig.5-9 The detected points in left image after (a) median filter (b) bilateral filter

CHAPTER 6

CLUSTERING THE DETECTED POINTS AND TRACKING CLUSTERS

6.1 Collision Avoidance and Tracking of Ships

The statistical results of maritime accidents make the authorities and researchers to pay more attention to increase the navigational safety. It is desirable to have many choices for safety navigation which support the applicability of The International Regulations for Preventing Collisions at Sea (COLREGs) in which the target object location and course are essential information for obstacle avoidance. Fig.6-1 illustrates a crossing situation according to COLREGs rules. As it can be seen from the figure the action for the collision avoidance is depending on the location, course and the speed of the target vessel. Recently some studies are carried out for the application of COLREGs in autonomous surface vehicles. These studies assume that the obstacles are already detected and localized and, their algorithm makes localization and mapping in accordance with COLREGs rules (Perera et al., 2011; Benjamin et al., 2006; Statheros et al., 2008). However, the obstacle detection is a very complex step for collision avoidance. Especially, ships automatic detection becomes more important for the safety navigation of ships.

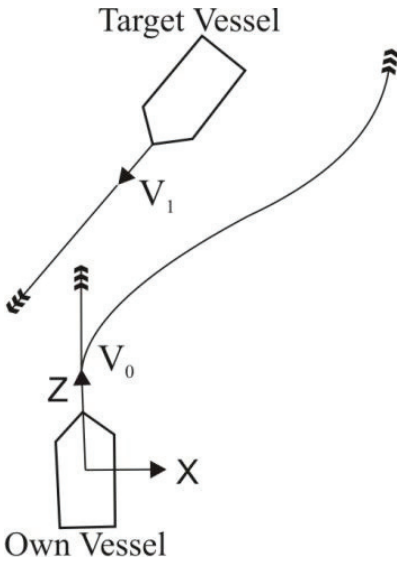


Fig.6-1 Crossing (Own vessel- give way)

Another important point is tracking the 3D location of automatically detected ships through image sequences to obtain the course of the ship for collision avoidance which is the topic of this Section. At the previous Sections the tracking of the ship was carried out by selecting a ship point manually. In that case only one point was used for representing a ship. However, it is very difficult for a computer to represent an object by one point. But it is possible to reduce the number of detected points to represent ships which makes tracking easier for computers. In this chapter, the detection algorithm explained in Section 3.2 is implemented for detecting obstacles automatically. The detection result will be many points which belong to the same object. Then, these detected points are clustered which represent the ship. By this way the number of points which represent a ship is decreased.

6.2 Clustering the Detected Points

The outline of the overall proposed method was explained in Chapter 1. It is summarized shortly here for capturing the place of clustering in overall context. First, a pair of stereo image, which are referred to as left and right images for convenience, are captured and the images are rectified. Then, vertical edges – edges are the important information of an image – are detected from the reference (left in this case) image after removing the noise from the image by a smoothing filter. The correspondences of the vertical edges are found in the right image and checked for trueness as explained in Section 3.2. The edges that cannot find the clear correspondence are removed. The remaining edges are assumed to be belonging to the detected objects.

As mentioned above, for a computer it is very difficult to represent an object with one point. Especially ship sizes have a great range from 5 meters to 350 meters. That's why the detected points are clustered. When the detection is completed, there will be many points which belong to several objects. For a better object representation and tracking performance the number of points representing the same object should be decreased as much as possible.

Therefore the detected points are clustered with a simple clustering algorithm which uses pixel location and disparity data which is calculated after detection. Every cluster is represented by values of the coordinates (*u, v*) and disparity. (*u, v*) is the coordinates of average of points in the cluster. So if a detected point has close values of the cluster, the average of them will become the renewal cluster. This is illustrated in Fig.6-2. If the distance between the red and green points is less than a certain threshold value, this point will be included in the cluster and the average value will be refreshed which represent the renewal cluster. When the whole image is scanned we obtain a number of clusters.

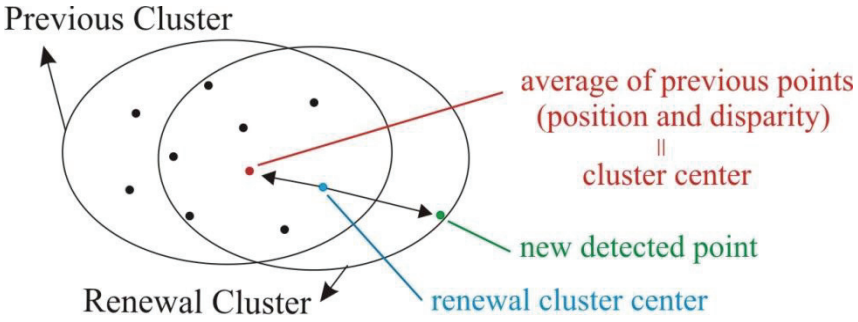


Fig.6-2 Clustering the detected points

The detected points and the cluster of these points can be seen in Fig. 6-3. The clusters are mainly located at bow and stern of the ship.

Fig.6-3 Sample of detected and clustered points

The clustering results of detected points in several illumination conditions are displayed in Fig.6-4.

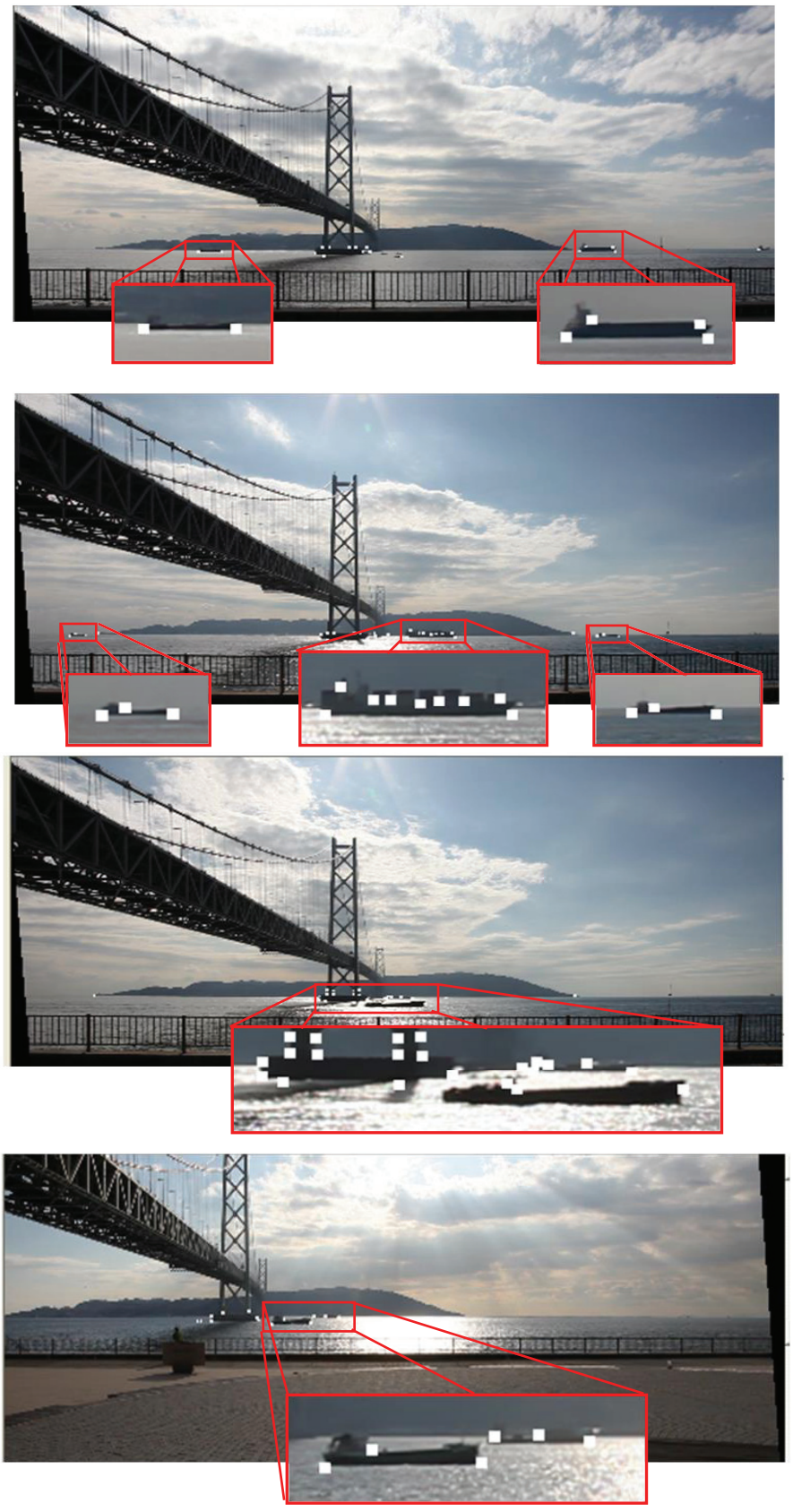
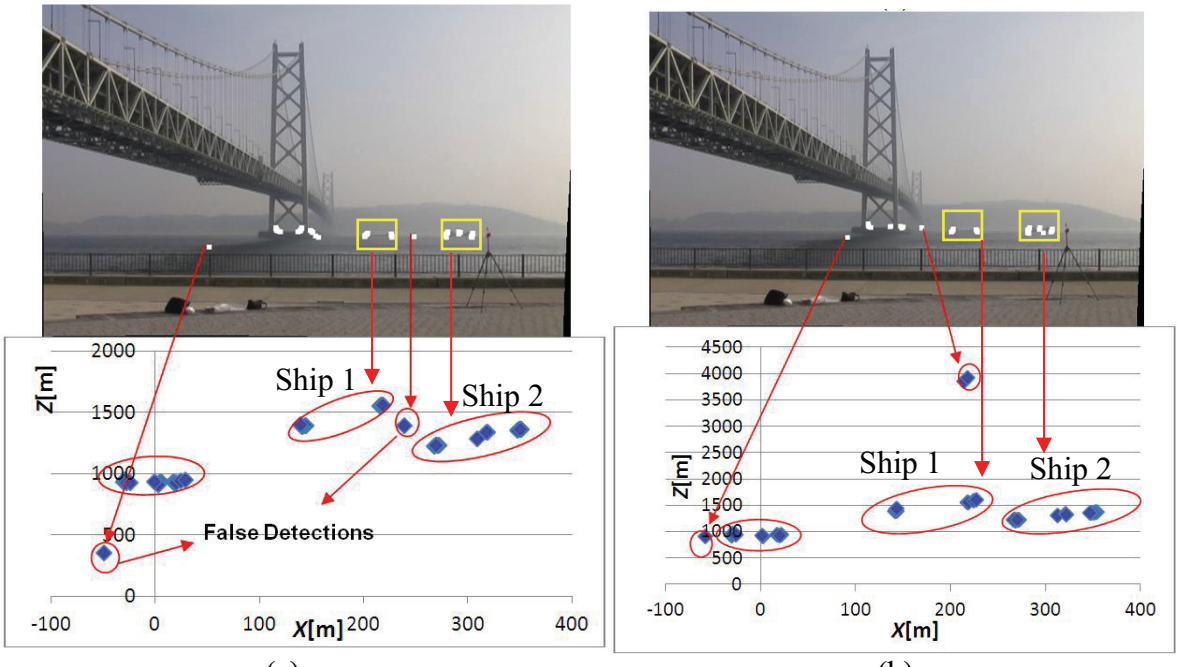


Fig.6-4 Clustering results of detected points under various illumination conditions

6.3 Tracking by Spatiotemporal Clustering and Feedback to Detection

After clustering the salient points, these points are tracked through image sequences. This is realized by the same procedure as clustering. The clusters of time *t* are saved. Then, detection is carried out in new image at time *t*+1. The detected points in the new image are integrated into the suitable existing cluster as it is a detected point of time *t*. When the cluster to which the salient points should belong doesn't exist, a new cluster is generated with them. On the other hand, when a cluster isn't updated several times in succession, it is removed. This results in tracking only stably detected clusters and elimination of unstable clusters.

The specular property of sea surface results in a noisy concept of sea images. This causes some reflectance points on sea surface to be perceived as an object point by computer. Despite smoothing by filtering it is not easy to eliminate all noise. It is possible to confirm the detection by tracking through image sequences which is improving the detection quality by elimination of such false detections. The tracking of 3D locations of certain object points will be stable and consistent while the arbitrary sea surface reflectance will change in successive images as shown in Fig.6-5.



(a) (b)

Fig.6-5 Changing positions of arbitrary false points (a) image at time *t* (b) image at time *t*+1

In the figure Ship1 is around 1500 meters in both images while Ship2 is around 1300 meters. However there are also some falsely detected points whose locations are extremely changed. While the false detected points are at about 380 meters and 1400 meters, in the following image their locations changed to 900 meters to 3900 meters. In this algorithm, feedback of the detection through tracking of the clusters yields elimination of false detected points because the false detected points will be out of the range of the clusters. In other words, the false point will be far from the cluster center and it will not be included in the cluster. To be included in the cluster it should be consistently detected. In Fig.6-6 the calculated 3D locations of detected points in Fig.6-3 are displayed in X-Z coordinates.

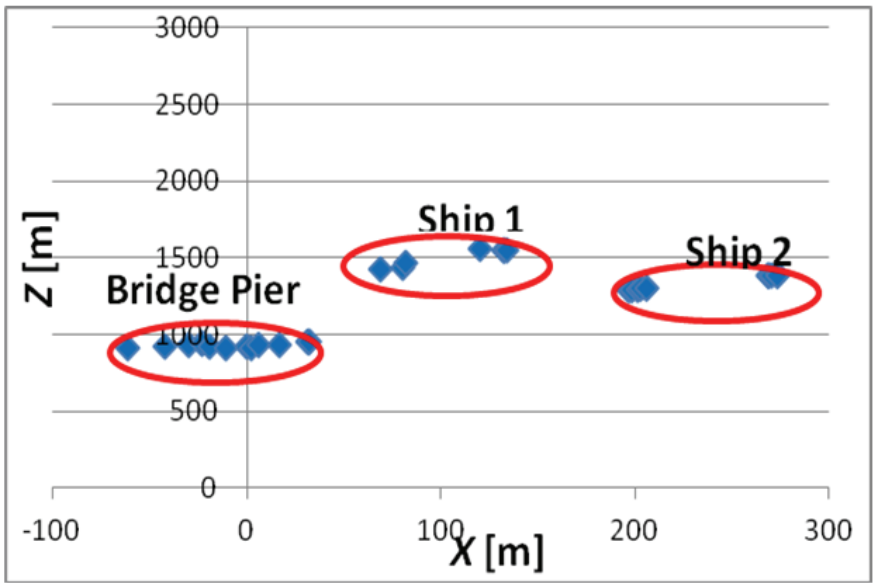


Fig.6-6 Calculated 3D locations of clusters

These points which represent one cluster are tracked through the sequences of images and the 3D locations are calculated for every image. In Fig.6-7 the calculated locations of all tracked clusters for median filtered images are shown for 300 images which mean tracking of 30 seconds.

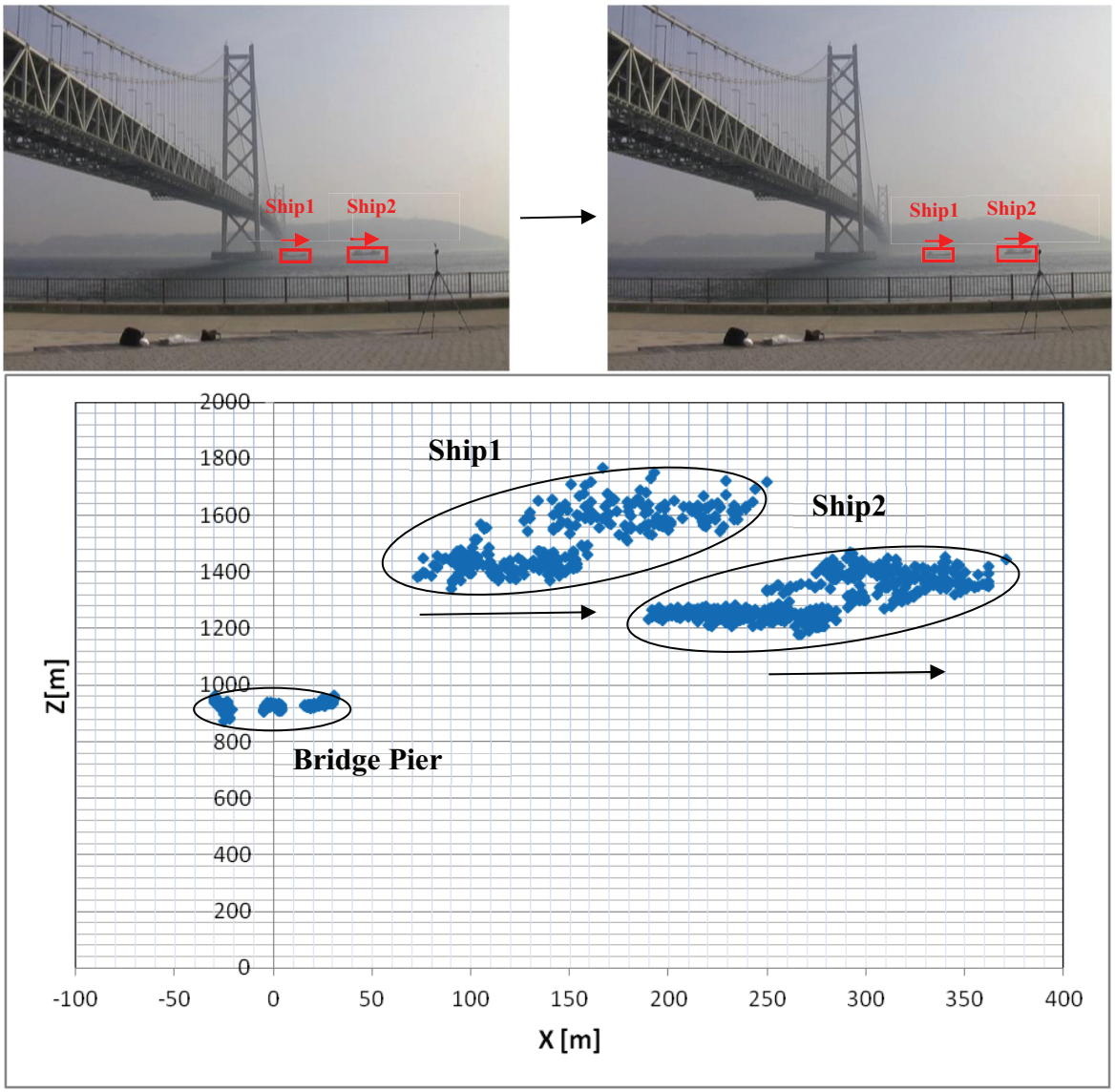


Fig.6-7 3D locations of tracked points for median filtered images

Implementing the clustering algorithm many clusters are obtained. Despite some few clusters are representing the same object they may seem as different clusters. However, tracking of these clusters for a dozen of seconds, the calculated 3D locations of these tracked points show a relationship between them. Visually it is easy to determine and distinguish the objects but it is very difficult for computers. So, we need to improve our clustering algorithm including the 3D locations of tracked points and constitute a relationship between the different clusters representing the same object as a future study.

Application of bilateral filter is improving the clustering and tracking results as a result of better detection. The tracking result of clustered points after bilateral filter is seen in Fig.6-8. Comparing the result of tracking with Fig.6-7 we can see the improvement of tracking through image sequences. The locations of cluster are not scattered as in Fig.6-8. They form a more line like figure which shows the course of the ship. This is very important for collision avoidance purpose.

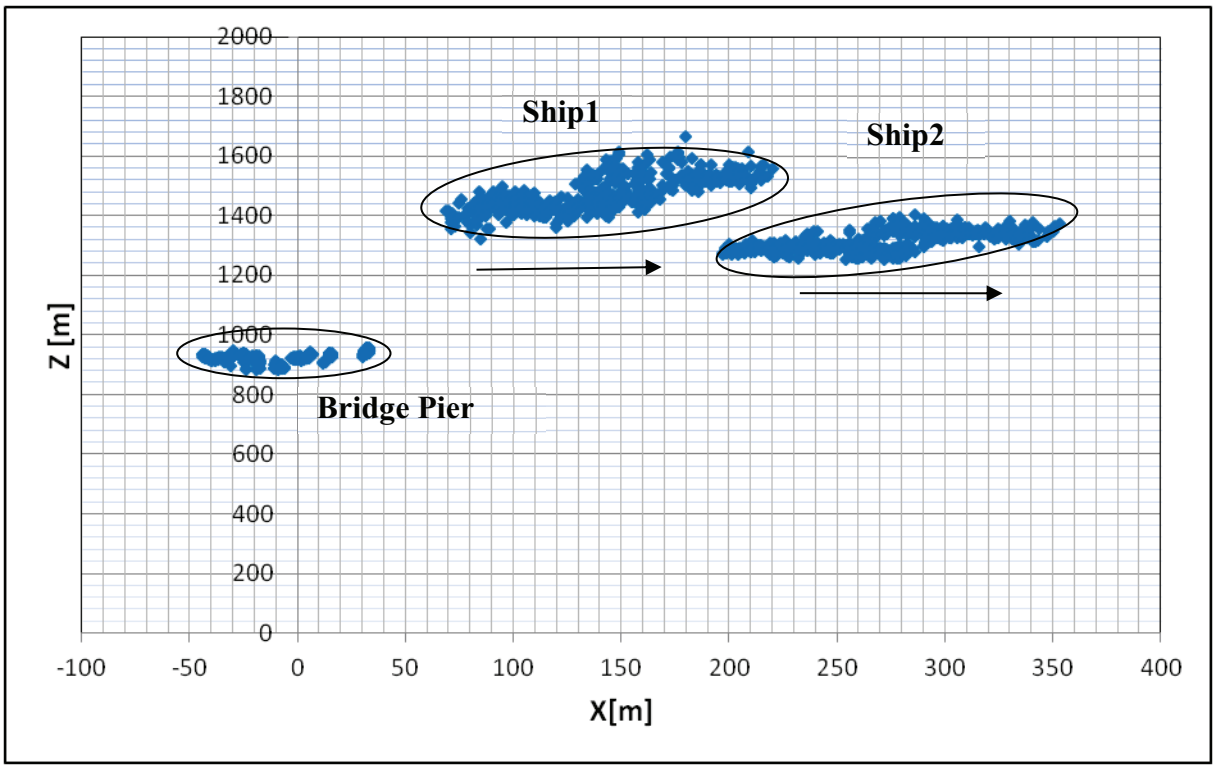


Fig.6-8 3D Locations of tracked points for bilateral filtered images

From the tracking results in Fig.6-8 we can visually constitute the relationship between tracked clusters and figure out that there are two ships moving through *X* axis which is shown by the arrow with a certain speed. The ship speeds are estimated as about 18 km/h for Ship1 and 20 km/h for Ship2.

In Fig.6-9 there are some more ship tracking results which are carried out in different illumination conditions and different ships.

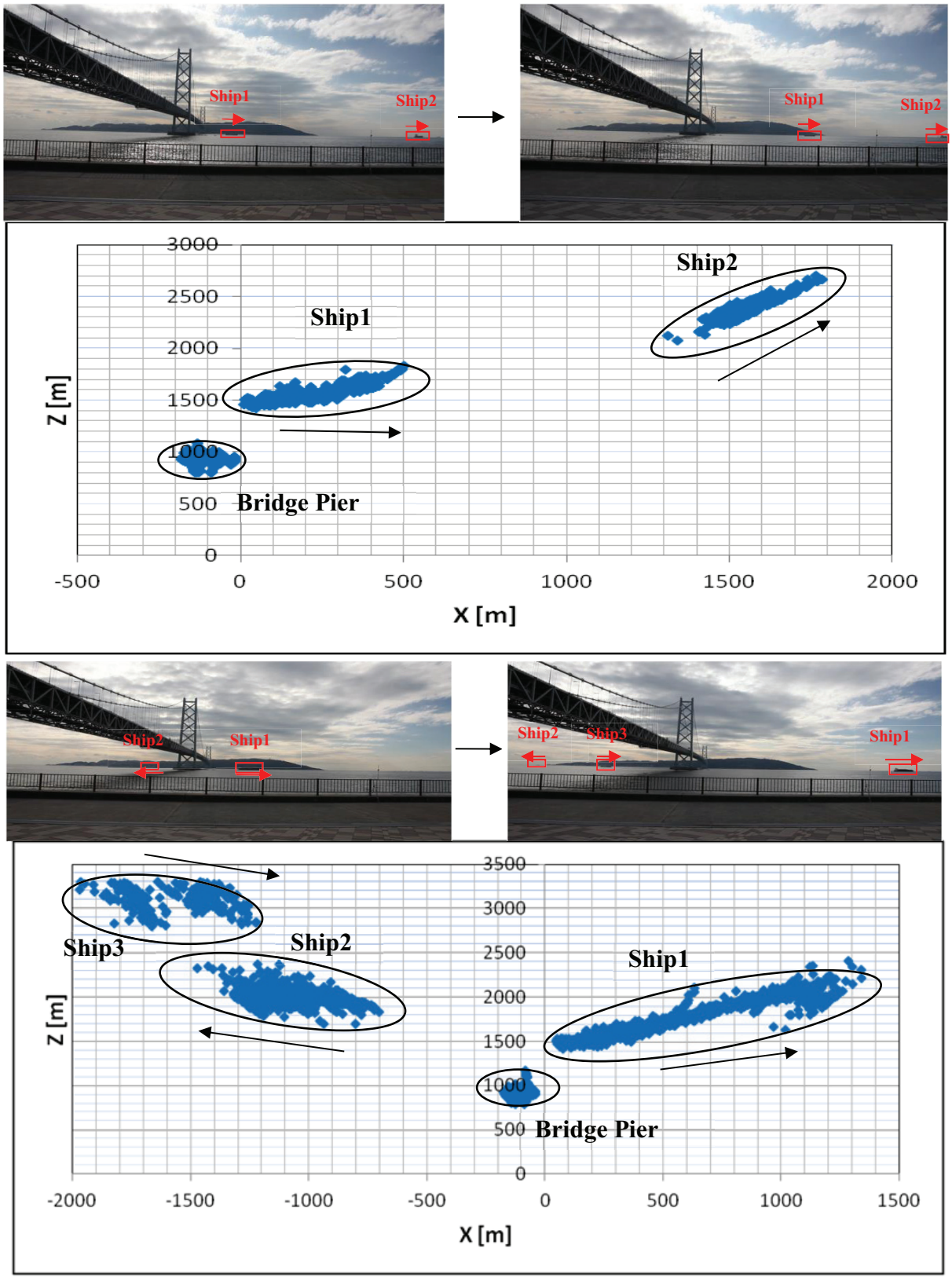


Fig.6-9 Tracking results with different image sequences

6.4 Tracking (+) and (-) Edges

One of the difficult points of ship detection is size of the ships. Nowadays the length of a ship can be greater than 300 meters. Therefore, it is a problem to represent one ship with one cluster in case of large ships. Therefore the ideal clustering should include stern and bow of the ships. Classifying the clusters as (+) and (-) edges can help us to achieve this. There will be some clusters -not one- representing the bow and some clusters representing the stern. The concept of (+) and (-) edges are illustrated in Fig.6-10. This is actually the concept of brightness gradient. If the brightness is changing from higher values to lower values at edge point this edge is a (+) edge and otherwise it is a (-) edge. Considering a ship shape in the sea, ship bow and stern may be represented as (+) and (-) edges which results in better representation of a ship and more accurate tracking results.

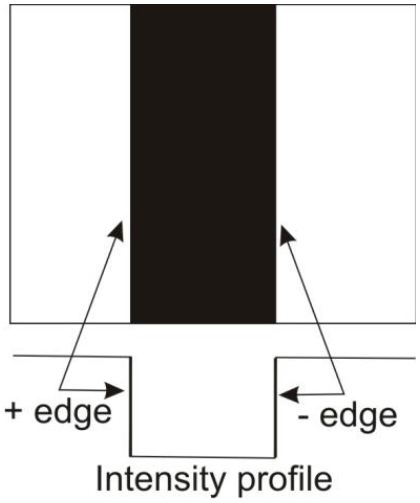


Fig.6-10 (+) and (-) edges

When the detected points are classified as (+) and (-) according to edge detection, the result will be as Fig.6-11. From this figure its seen that tracked clusters which seem as one part in Fig.6-8 can be divided into two parts. These represent the detected points in the stern side of ship and the bow side of the ship.

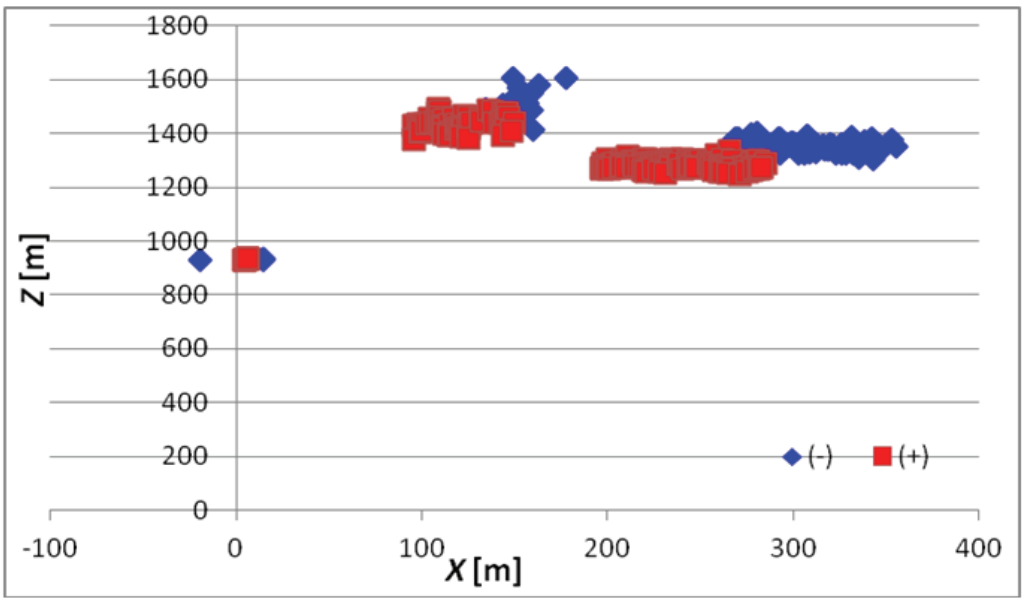


Fig.6-11 Tracking data of (+) and (-) edges

CHAPTER 7

ANALYZING THE INFLUENCE OF ENVIRONMENT

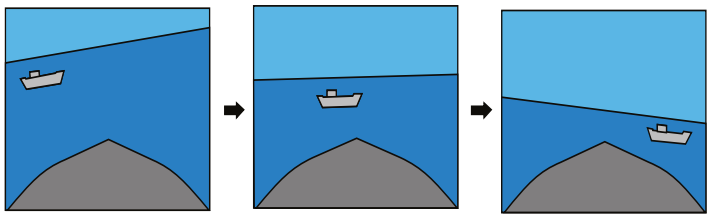
7.1 Background and Problem Statement

A ship is in an extremely dynamic environment. There are many parameters which may degrade the system quality such as ship motion (rolling, pitching and yawing), vibration due to strong wind, and structural vibration. In this section, application of the vision system to such a dynamic environment is inspected. At first, a principle to measure 3D location of an object under the unstable condition on board ship is described. It includes estimating the rolling and pitching angles of the ship to eliminate the negative effects of such movements and calculating the relative camera orientations for accurate 3D calculation of detected objects. Then, details of sea-sky line detection from sea images to obtain the rolling and pitching angle of the ship is described.

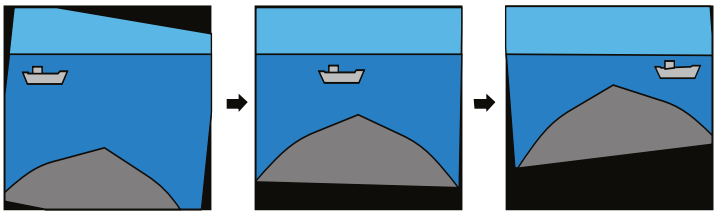
So far, the cameras were considered to be fully fixed and accuracy of the measurement was investigated. However, when the cameras are installed on board ship they will be exposed to vibration and movements caused by waves which result in camera shake and change in camera orientation which has a great effect on the accuracy of position measurement. Therefore it is really essential to obtain the camera headings precisely. To be able to obtain the camera headings we have to consider the movement of the ship and cameras separately. The cameras equipped on board ship are subjected to ship motion such as rolling and pitching which are same for both the ship and cameras. This can be eliminated by sea-sky line detection. However, they are also exposed to vibration different from ship's general motion and should be considered separately. The effect of vibration is small changes in camera orientation which result in huge 3D calculation error and need to be corrected in real-time. This can be realized by observing some rigid points on ship's hull. A method for

self-extrinsic calibration of cameras in a dynamic environment by integrating sea-sky line detection is explained in this section.

The main goal of the overall system is detecting ships and measuring their locations and finally their velocity and course by using a stereo vision system. However, the following two points should be considered in order to equip the stereo vision system on board ship. One point is rolling and pitching of the ship by waves. Since this leads to discontinuity and irregularity of projected images of target ships, it is difficult to track them. This is illustrated in Fig.7-1(a). In addition to that, even if we can track them, using the coordinate system fixed to the own ship also makes results of the 3D coordinates of targets discontinuous and irregular. Therefore eliminating the rolling and pitching effect of the ship is important. This is carried out by adjusting the images so that each sea-sky line in the images becomes horizontal and its vertical position in the images keeps constant as shown in Fig.7-1(b). After the adjustment of the images it will be possible to trace the ships much more easily. In order to realize this adjustment, precise detection of sea-sky lines becomes a key technique.



(a) Original images



(b) Adjusted images by sea-sky lines

Fig.7-1 Effect of adjustment on sea images

Of course, the cameras equipped on board ship are also subjected to ship motions such as yawing, surging and swaying. However, change of camera orientations with rolling and

pitching greatly affects locations of objects in the image as mentioned above, while a shift of camera positions with surging and swaying doesn't so much. Yawing motion of a ship is not so fast and relative positions of the target ship to the own ship are important. Therefore, we can ignore yawing motion by defining the world coordinate system based on heading of the own ship as indicated in Chapter 3.

Another point is small change of camera orientations. Equipment on board ship is in severe conditions. Therefore, there is an enough possibility that vibrations from an engine, shock caused by waves or warp of ship's hull change orientations of cameras a little even if the cameras are firmly fixed on the ship's hull. We have to measure positions of targets at a distance of a few kilometers from the point of view of prevention of sea accidents. On the other hand, a baseline length is 10 to 20 meters at most because the cameras should be set at the both side of the ship and the baseline length is limited with the width of the ship. In this study, we measure targets at a distance about a hundred times of the baseline length. Therefore, a very small change of camera orientation can cause a large error of 3D measurement as explained in Section 3.6. From these reasons, it is necessary to measure rolling and pitching angles of ships, and precise orientation of cameras to the ship's body.

7.2. 3D Measurement in Dynamic Environment

7.2.1 Coordinate Systems

In order to measure 3D locations of objects with a stereo vision system, intrinsic and extrinsic parameters of cameras are required which are explained in Chapter 2. The intrinsic parameters are related to camera internal geometry and they express mapping from 3D coordinates in the camera coordinate system which is fixed on the camera to image coordinates. On the other hand, extrinsic parameters are locations and orientations of the cameras in the world coordinate system. We have to define some coordinate systems and their relations as shown in Fig.7-2 which is needed to for the extrinsic camera parameters.

This is because the ship is in a dynamic environment and the coordinate systems have some relative changes due to rolling, pitching and yawing of ship and cameras which affect the camera extrinsic parameters and 3D calculation.

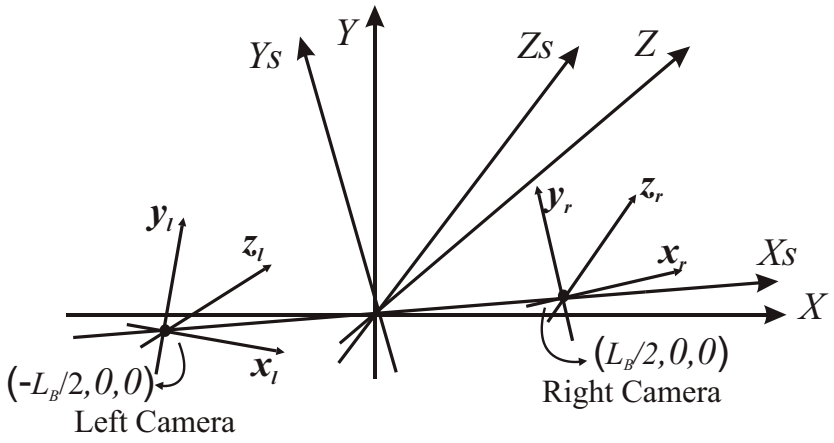


Fig.7-2 Coordinate systems in case of dynamic environment

First, we define the *XYZ*-coordinate system (world coordinate system) which is independent from cameras and ship body with *Z* axis fixed and points to the forward (through the bow of ship). In this coordinate system *XZ* plain is horizontal, *Y* axis is vertical to the sea surface and *X* axis is parallel to the sea-sky line. Then we define $X_SY_SZ_S$ coordinate system which is fixed to the ship body. When the ship is in stable condition (no rolling and pitching) the *XYZ* and $X_SY_SZ_S$ coordinates are overlapped. $X_SY_SZ_S$ coordinate system rotates around *Z* axis (rolling) at first, then, rotates around *X* axis (pitching) according to the motion of ship affected by waves. Since the objective of the proposed system is detection of other ships that have risk for a collision, knowing a relative direction and distance is sufficient and a yawing angle of the ship is not needed. This movement of the $X_SY_SZ_S$ coordinate system is same with the left and right cameras for which the coordinate systems $x_ly_lz_l$ and $x_ry_rz_r$ are defined and fixed to the cameras. The cameras are located parallel to each other at $(-L_B/2, 0, 0)$ and $(L_B/2, 0, 0)$ on the X_S coordinate of the $X_SY_SZ_S$ coordinate system as shown in Fig.7-2.

The main objective is to obtain 3D locations of the targets in the *XYZ* coordinate system. In order to achieve this, the locations and orientations of both cameras (i.e. extrinsic

parameters) in *XYZ* coordinate system should be known. At this time, two components of the camera orientations should be considered separately: posture of the ship, that is, relative rotation of $X_S Y_S Z_S$ coordinate system and *XYZ* coordinate system, and small movement of cameras, that is, cameras orientation in the $X_S Y_S Z_S$ coordinate system. Let's denote the ships rolling and pitching angles in *XYZ* coordinate system as θ_Z and θ_X, and cameras rolling, pitching and yawing angles in $X_S Y_S Z_S$ coordinate system as ϕ_z, ϕ_x and ϕ_y, respectively. Although there are two cameras, the same theory can be applied to both. Therefore, we denote them without distinction in the following sections.

The algorithm flowchart for 3D calculation in a dynamic environment is shown in Fig.7-3. Firstly, some stable points on ship hull are determined and their 3D coordinates are measured precisely. Then these points are tracked through image sequences for calculating camera orientations to ship body. Meanwhile, the sea-sky line is detected and used for calculating ships posture in world coordinates. Therefore, the detected ships 3D locations can be calculated by eliminating the variations occurred due to dynamic environment. These procedures will be described closely in the following sections.

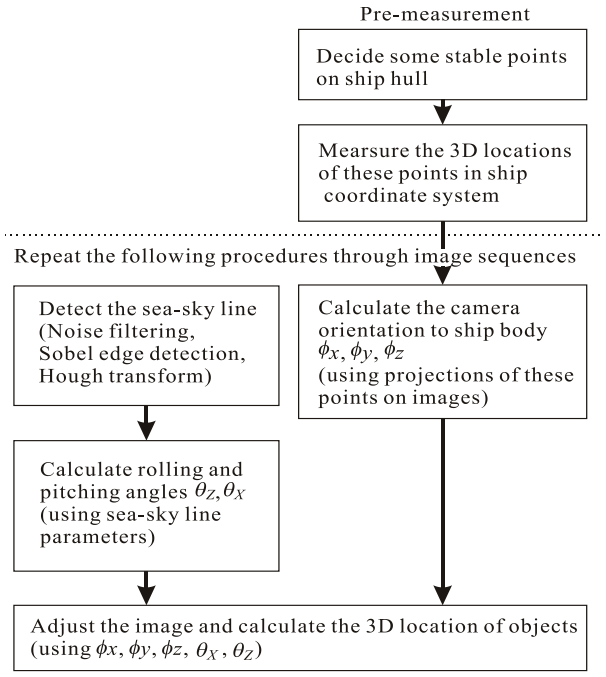


Fig.7-3 Algorithm flowchart for dynamic environment

7.2.2 Camera Orientation to the Ship

The relation between the ship coordinates and the image coordinates is expressed by

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KR[I \quad -T] \begin{bmatrix} X_s \\ Y_s \\ Z_s \\ 1 \end{bmatrix} \tag{7-1}$$

which is very similar to the formulations explained in Chapter 2.3 and 2.4 but in this case world coordinates are replaced with ship coordinates X_s, Y_s, Z_s and R is a rotation matrix between the camera and ship coordinates such as

$$R = R_x(\phi_x)R_y(\phi_y)R_z(\phi_z) \tag{7-2}$$

where $R_i(\phi_i)$ (i =x, y, z) are componential rotations of R around axis i by an angle of ϕ_i . The rotations ϕ_i are again between the camera coordinate system and ship coordinate system.

These intrinsic parameters K and extrinsic parameters R and T can be obtained by observing some known pairs of (X_{Si}, Y_{Si}, Z_{Si}) and (u_i, v_i) i=1,2,…,N. Since the number of parameters is eleven and Equation (7-1) gives two equations by eliminating λ, N must be equal to or larger than 6 when all the parameters are unknown. Once the parameters K and t are obtained, we can calculate the rotation matrix R by observing two or more known pairs of (X_{Si}, Y_{Si}, Z_{Si}) and (u_i, v_i) even if the camera rotates due to vibration etc. This can be realized by detecting at least two stable points on ship's hull.

7.2.3 Ship's Posture

In order to calculate θ_Z and θ_X, we detect the sea-sky line in the camera image and obtain its slope θ_H as shown in Fig.7-3. The sea-sky line is formulated as

$$v = \tan\theta_H(u - u_0) + v_H \tag{7-3}$$

where the parameters can be seen from Fig.7-4.

The following equations are expression of each row of Equation (2-6) separately.

$$\lambda u = \alpha_u x_c + \beta y_c + u_0 z_c \tag{7-4}$$

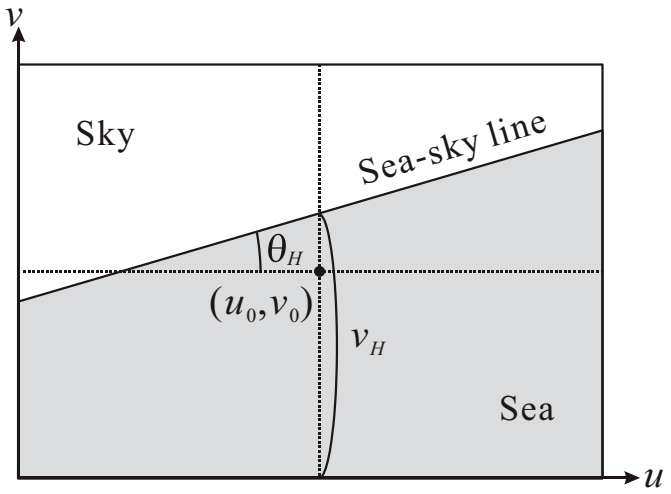


Fig.7-4 Sea-Sky line and its parameters

$$\lambda v = \alpha_v y_c + v_0 z_c \tag{7-5}$$

$$\lambda = z_c \tag{7-6}$$

Instituting Equation (7-6) in (7-4) and (7-5) we can express them as

$$z_c(u - u_0) = \alpha_u x_c + \beta y_c \tag{7-7}$$

$$z_c v = \alpha_v y_c + v_0 z_c \tag{7-8}$$

Then instituting Equation (7-3) into Equation (7-8) we obtain

$$z_c\{\tan\theta_H(u - u_0) + v_H\} = \alpha_v y_c + v_0 z_c \tag{7-9}$$

Reshaping Equation (7-9) we obtain

$$\alpha_u \tan\theta_H \, x_c + (\beta_u \tan\theta_H - \alpha_v) y_c + (v_H - v_0) \, z_c = 0 \tag{7-10}$$

The slope y_c/x_c means the reverse of camera rolling angle, expressed as

$$-\tan(\theta_Z + \phi_z) = \frac{\alpha_u \tan\theta_H}{\alpha_v - \beta_u \tan\theta_H} \tag{7-11}$$

and the slope y_c/z_c means the reverse of camera pitching angle, expressed as

$$-\tan(\theta_X + \phi_x) = \frac{v_H - v_0}{\alpha_v - \beta_u \tan\theta_H} \tag{7-12}$$

Finally the rolling angle θ_Z of the ship in the *XYZ* coordinate system is obtained by

$$\theta_Z = -\tan^{-1}\frac{\alpha_u \tan\theta_H}{\alpha_v - \beta\tan\theta_H} - \phi_z \tag{7-13}$$

and the pitching angle θ_X can be calculated by,

$$\theta_X = -\tan^{-1}\frac{v_H - v_0}{\alpha_v - \beta\tan\theta_H} - \phi_x \tag{7-14}$$

where v_H is v coordinate of the sea-sky line at $u = u_0$. ϕ_x and ϕ_z are assumed to be small enough.

7.3 Sea-sky Line Detection

7.3.1 Related Works

In literature, horizon detection studies are mainly classified into two approaches; region-based and edge-based. For the flight autonomy usually region-based methods are preferred due to the high intensity differences between the sky and ground image data which makes it possible to model the region. Some instances of this approach are based on covariance matrices and Eigen values (Ettinger et al., 2003), and linear discriminant analysis (Todorovic and Nechyba, 2004). In contrary the proximity of intensity values of sea and sky in images makes it difficult to model the region; therefore edge-based approach is preferred for this study.

Bouma et al. (2008) applied multi-scale edge detection on gray-scale and infrared images for the purpose of sea-sky line detection. However, the method they used for the horizon detection is not explained. Besides the horizon line is assumed to be a straight line. In case of a ship or an object lying through sea-sky line may affect the algorithm and there is not any example of such situation. All the applications were carried out in clearly strait sea-sky line images. Ji et al. (2009) used horizontal projection to reduce the search area and canny edge detector followed by a kernel Hough transform which operates on real-time speed. Jiang et al. (2010) applied histogram analysis and linear line fitting on gray-scale image for the same purpose. Fefilatyev et al. (2006) exploited machine learning techniques such as support vector machines for horizon detection. In their study they used very few images (only 20 images) to teach the sea-sky line to the system and the experimental images were so clear. In case of complex situations which are not taught to system will result in error.

The mentioned studies are usually using clear sea images where the sea and sky are easily separable. They also don't consider about the rolling and pitching of the ship and effect of different illumination conditions. In this study we show that an algorithm which can detect the sea-sky line in horizontal situation may fail when the ship is rolling. And the effect of change in contrast due to the chance of illumination conditions is considered.

7.3.2 Outline of Detection Method

The sea-sky line detection algorithm of this study is as follows: Firstly the image is smoothed with the bilateral filter. Then Sobel filtering is applied to obtain the edge data. The sea-sky line is assumed to be a linear line in the image. Therefore, the edge data obtained by the Sobel filtering is used for Hough transformation for linear lines which result in the detection of the sea-sky line.

7.3.3 Edge Detection

In order to detect the sea-sky line, we must firstly detect edges from the image. Edge detectors respond to local changes in images. Mostly used edge detectors are Sobel, Prewitt, Roberts and Canny edge detectors. In this study Sobel edge detector is used for obtaining horizontal, vertical and diagonal edges due to its simplicity and effectiveness. The Sobel operator performs a spatial gradient measurement of image brightness values. It computes the gradient at each pixel by 3x3 convolution kernels which are shown in Fig.7-5. The pixel is decided to be an edge when the gradient value is greater than a threshold value.

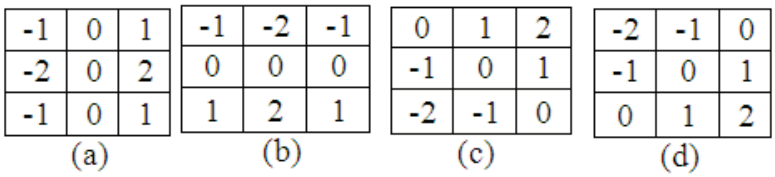


Fig.7-5 Sobel edge detector convolution masks; (a) vertical, (b) horizontal, (c) and (d) diagonal

The sea images are very noisy due to specular property of sea surface. And it causes abundant edges on sea surface. Despite the sky and sea surface visually appear so similar,

edge detection results are completely different as shown in Fig.7-6 that at sea surface there is abundant of edges while at sky there are only a few edges. Therefore, a good smoothing filter is crucial before applying Sobel filter.

Fig.7-6 Result of Sobel edge detection on non-filtered image

Another point is the edge detection effect in case of the rolling of the ship. Firstly, only first two Sobel masks shown in Fig.7-5 which are for horizontal and vertical edges are applied. The application of this Sobel filter is shown in Fig.7-6. We can see that the edges which are on the sea-sky line are nicely detected. However in case of the ship rolling, the sea-sky line lie in a diagonal form in the image. In this case the edge detection of only horizontal and vertical edges are not sufficient and the edges on the sea-sky line cannot be detected as shown in Fig.7-7 (a). This results in false detection of sea-sky line as shown in Fig.7-7(b) which will be explained in the following parts of this Section.

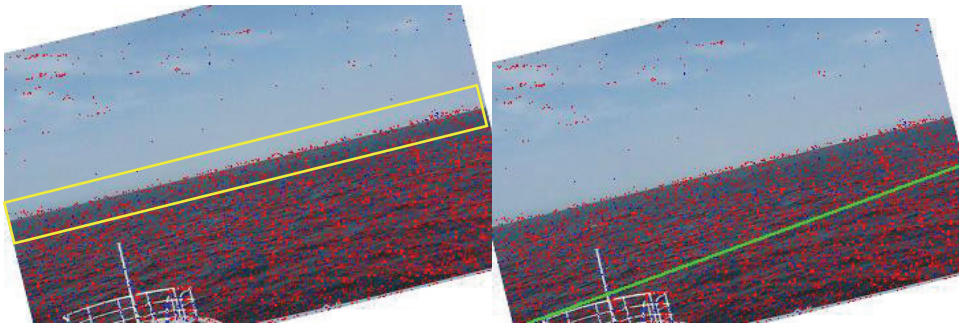


Fig.7-7 Edge (a) and sea-sky line (b) detection result in case of rolling of ship

That's why the diagonal components of the Sobel filter is included which are shown in Fig.7-5. The edge detection result including diagonal components is shown in Fig.7-8. In this case the edges on the sky are totally eliminated.

Fig.7-8 Sobel edge detection with horizontal, vertical and diagonal components

7.3.4 Effect of Smoothing Filters to Edge Detection on Sea Images

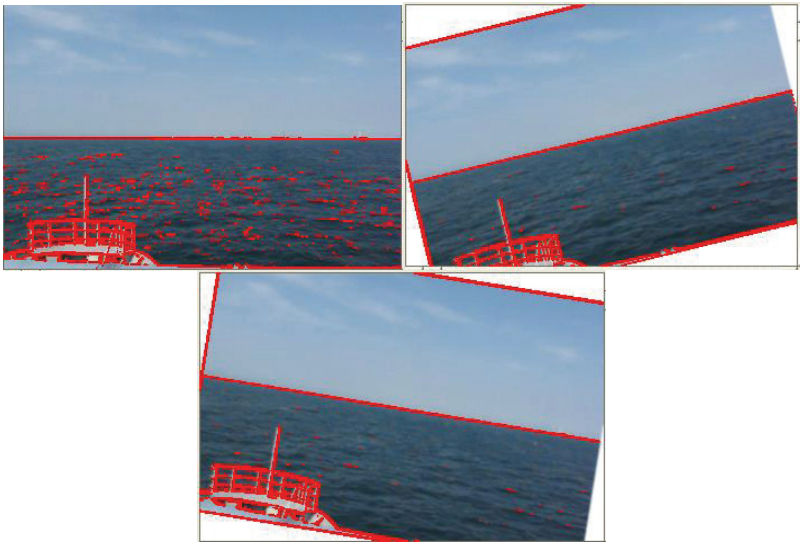
The smoothing effects of some filters to edge detection are compared. This is different from the edge detection results explained in Chapter 5 which uses only horizontal and vertical edges. Here the edge detection is using the diagonal components also as shown in Fig.7-5 and the situation of rolling of a ship is also considered. The compared filters are:

1. Gaussian filter,
2. Multi-scale Gaussian filter,
3. Median filter, and
4. Bilateral filter.

Application of a 3x3 Gaussian filter to a horizontal image, a clockwise rotated image and a counter-clockwise rotated image are shown in Fig.7-9.

Fig.7-9 Smoothing effect of Gaussian filter

Multi-scale Gaussian filtering is applied as iteration of Gaussian filters. So, we applied 3x3 Gaussian filter two more times and obtained the 2nd and 3rd scales. In Fig.7-10 multi-scale filtering and edge detection results for horizontal and rotated images are seen. Multi-scale Gaussian filtering is seen to be very effective to remove noise caused by sea surface reflectance. That is good for sea-sky line detection. However, its effect is not positive for the object detection purpose because the object edges are weakened by smoothing.



(a) 2nd scale

(b) 3rd scale

Fig.7-10 2nd and 3rd scale Gaussian filtered images and edges

Median filter is a non-linear filter which is also very effective for removing sea surface noise. The result of a 7x7 median filter and Sobel edge detector is seen in Fig.7-11.

Fig.7-11 Smoothing effect of median filter and obtained edges

Bilateral filter is a new type of filter comparing to the above mentioned filters. It is a kind of edge preserving filter as mentioned in Chapter 5. Recently real-time speed of bilateral filter is achieved (Yang et al., 2009). The result of smoothing effect of bilateral filtering is shown in Fig.7-12. The advantage of bilateral filter is its edge preserving property.

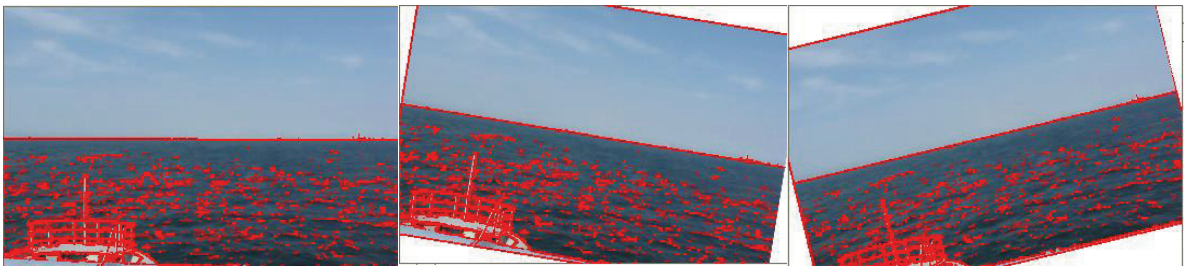


Fig.7-12 Smoothing effect of bilateral filter and obtained edges

Multi-scale Gaussian filter and median filter are very good for smoothing noise of sea surface. However, we have to consider the object detection purpose, as well. It is possible to use different filters for smoothing and for detection. However, filtering is an expensive operation and one filter is preferred if it is sufficient for both purposes. The bilateral filter is preferred in this study due to property of preserving edge data which is necessary for good object detection as explained in Chapter 4 and its sufficiency for noise removal.

7.3.5 Adaptive Threshold for Sobel Edge Detection

During the experiments we observed a problem of contrast change in the images. The amount of light is changing in different weather conditions and it directly affects the contrast of the image which results in the deterioration of Sobel edge detector. A threshold value is

needed for the edge detection by Sobel filter. For the high contrast images this threshold value should be high and for the low contrast images it should be low.

An adaptive threshold is applied to solve this problem. Firstly the image contrast is calculated. Image contrast is the difference between the lowest and highest brightness values. Therefore, firstly the lowest and highest brightness values are found and the difference is calculated. The value of this difference is used for determining the threshold value for Sobel filter. One of the three levels (25, 50 or 75) of threshold is chosen up to the calculated image contrast.

The effect of the adaptive threshold is shown in Fig.7-13. In Fig.7-13 (a) and (b) the threshold value is 75. In high contrast image the edges are detected nicely but in low contrast image no edges could be detected as in Fig.7-13 (b). In Fig.7-13 (c) and (d) the threshold value is 25. In high contrast image many edges are detected while in low contrast image edges are optimally detected.

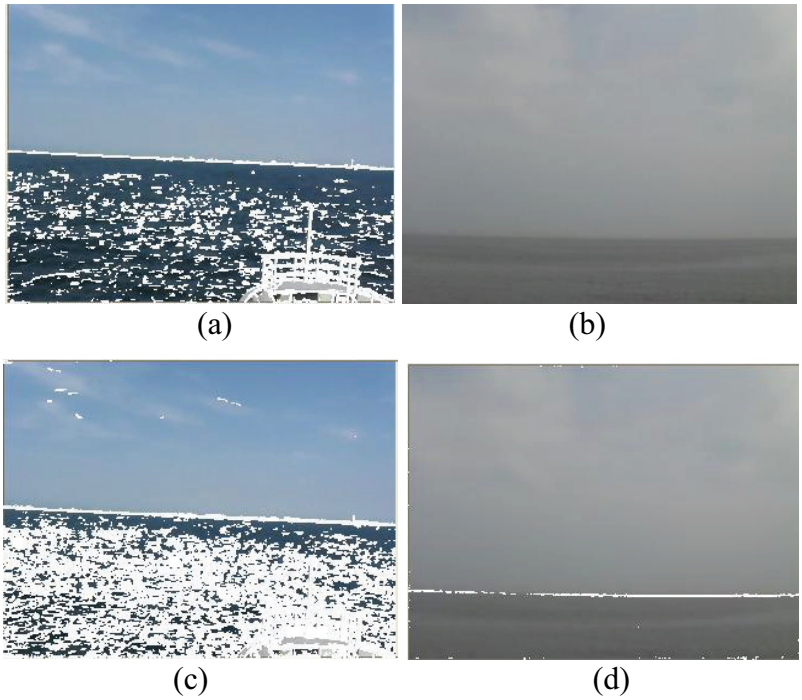


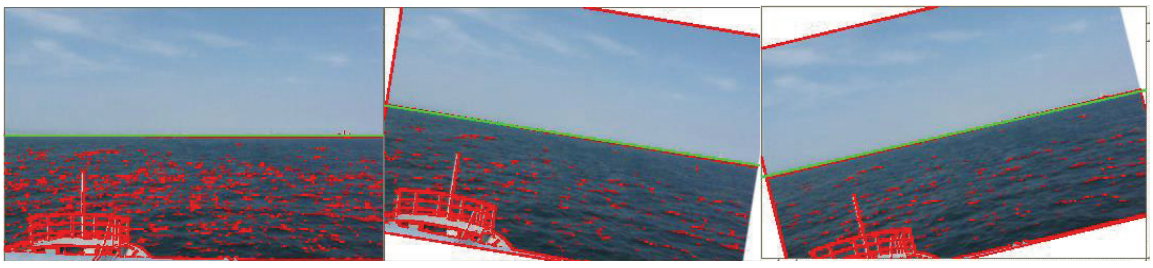
Fig.7-13 Relation of image contrast and edge detection; (a) High contrast, high threshold (b) Low contrast, high threshold (c) High contrast, low threshold (d) Low contrast, low threshold

7.3.6 Hough Transform and Sea-sky Line Detection

The Hough transform is one of the most popular algorithms for extracting straight lines from an image (Duda and Hart, 1972). The algorithm mainly processes a mapping between an image space and a parameter space. Each edge point in the image is mapped to the parameter space to vote for the parameters. The line is represented as

$$r = -(u - u_0)\sin\theta_H + (v - v_0)\cos\theta_H \tag{7-11}$$

where the parameter θ_H is the angle of the line and r is the distance of the line from the image center. The range of θ_H may span between [$-\pi/6$, $\pi/6$] due to the possible maximum rolling angle of a ship. It is limited to around 30 degrees depending on the ship type (France et al., 2001; Shen et al., 2005). r may span around the length of image diagonal. These parameters are discretized to accumulator cells and every (u, v) value of edges is mapped to this parameter space and collected in the accumulator cells. The corresponding parameter values for the peak values of the accumulator cell result in the detection of the straight line. Detecting the sea-sky line we obtain the two parameters θ_H and v_H which will be used for extrinsic calibration of the cameras as shown in Section 7.2.3. The results of sea-sky line detection for horizontal and rolling ships is illustrated in Fig.7-14. The green lines are the detected sea-sky line. In figure the calculated rolling angle of the sea-sky line is -9 degrees (minus means clockwise rotation) in the middle column images and 14 degrees in the right column images.

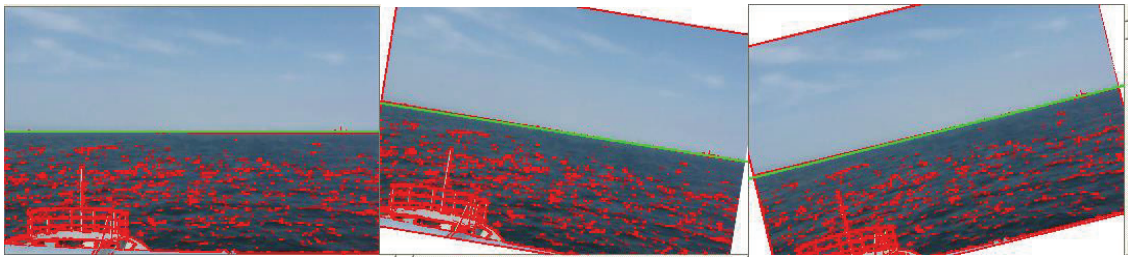


(a) 3x3Gaussian filtered

(b) 2nd scale 3x3 Gaussian filtered

(c) 3rd scale 3x3 Gaussian filtered

(d) Median filtered



(e) Bilateral filtered

Fig.7-14 Hough transformation results

The results show the efficiency of the sea-sky line detection algorithm in horizontal and rotated images.

7.3.7 Influence of Rolling and Pitching Angles and Their Errors

This section describes influence of errors in sea-sky line detection. At first, ship's rolling angle θ_Z is almost $-(\theta_H + \phi_z)$ from Equation (7-13) since α_u is nearly equal to α_v and β is small enough for a usual camera. Therefore, accuracy of θ_Z is mainly decided by that of θ_H

(slope of the sea-sky line). Next, since α_v of a high definition video camera with a lens of 70 degrees of view angle is about 1370 pixels, it is found that an error of 1 pixel in v_H causes an error of 0.04 degrees in θ_X from Equation (7-14) and it is very small.

Our objective is to measure 3D locations of a target ship. Fig.7.15 shows a calculation of influence of errors in rolling and pitching angles. The solid line indicates a track of the target ship. The own ship is rolling with an amplitude of 10 degrees. The rolling period is 8 seconds. It is also pitching with the same amplitude, but the period is 12 seconds. Circles in the figure show calculated locations in case that the rolling and pitching motions are neglected. It is difficult to estimate the course of the target ship from these erroneous locations. Crosses in the figure indicate locations calculated using rolling and pitching angles which have a random error of +1 or −1 degree. It is found that 1 degree of error does not affect the calculation of 3D locations so much and we can achieve this level of accuracy from the above argument.

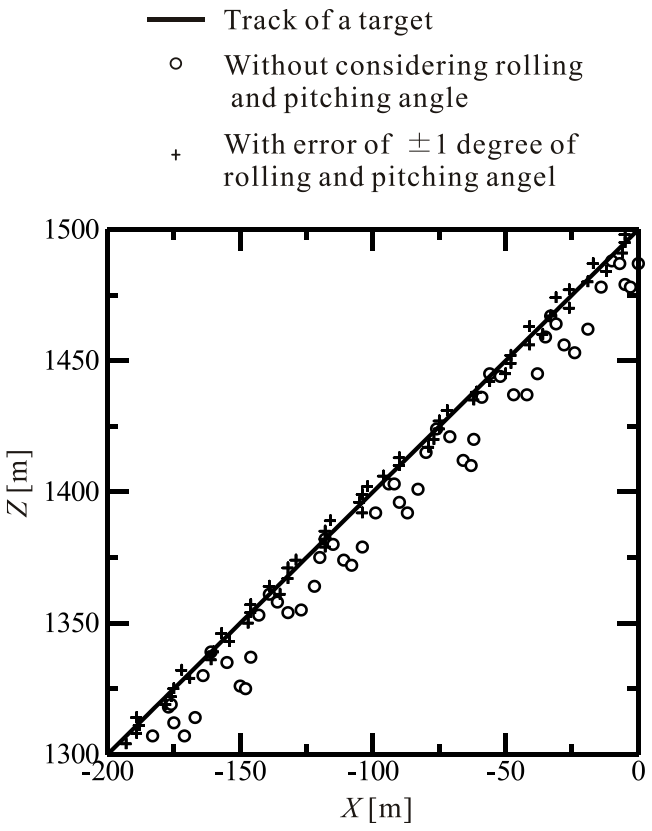


Fig.7-15 Influence of rolling and pitching angles

Obtaining the camera headings is a crucial step to be able to detect and trace the ships accurately using a stereo camera system equipped on board ship. This section described a method to realize it in a dynamic environment of the sea. The proposed method will realize precise measurement of 3D locations of target ships with a stereo vision system on board ship and we intend to confirm it by some total experiments on the sea as a future work.

CHAPTER 8

FURTHER CONSIDERATION ON DETECTION AND TRACKING

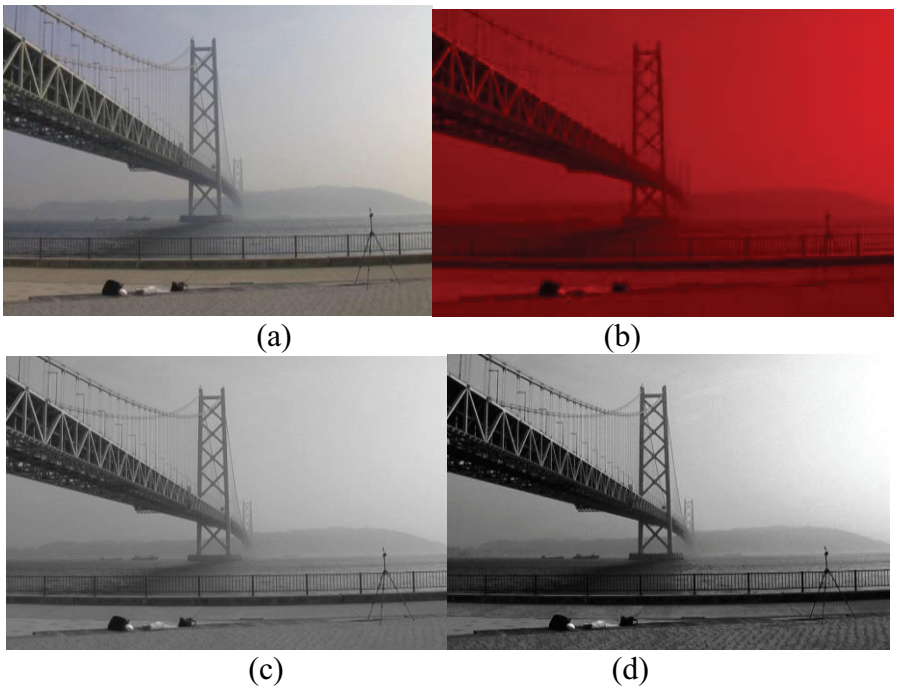
8.1 Detection

So far, the detection algorithm is applied on a red channel of a color image. In this section, the performances of detection algorithm on different channels of same image are considered to improve the detection quality. That's why the gray scale and histogram equalized gray scale of the color image are obtained. Transforming an image from color to gray scale is a trivial job using the following equation.

$$Gray = 0.3R + 0.59G + 0.11B \tag{8-1}$$

where R, G and B are red, green and blue components of color image (Pratt, 2007).

Histogram equalization is a method to enhance the visual details of an image. The image is enhanced by applying a transformation to the histogram values. The RGB color image, its red channel, gray level and histogram equalized gray level images are shown in Fig.8-1.



(a) (b)

(c) (d)

Fig.8-1 Images of different channel (a) RGB image (b) Red channel only (c) Gray level (d) Histogram equalized gray level

The magnified images of Fig.8-1 are shown in Fig.8-2 to be able to see the effect on ship.

(a) (b)

(c) (d)

Fig.8-2 Magnified ship images (a) Original (b) Red channel only (c) Gray level

(d) Histogram Equalized gray level

From Fig.8-2 we can figure out that gray level and histogram equalized gray level images are visually better than only red channel. The histograms of gray scale image intensity data and histogram equalized gray scale image intensity data are shown in Fig.8-3 to see the effect of histogram equalization.

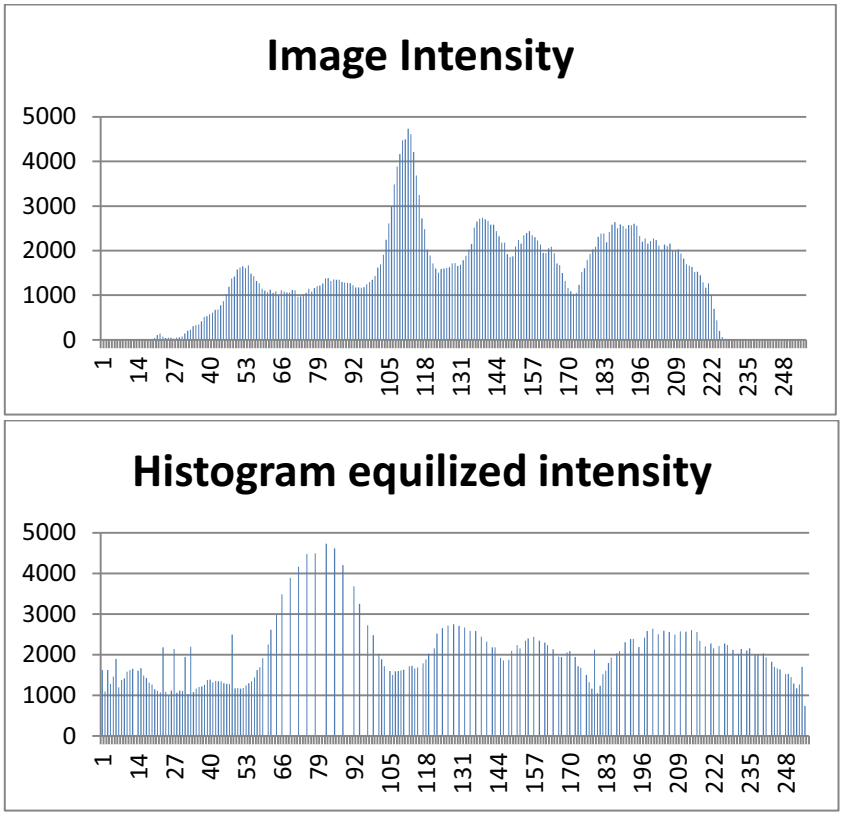
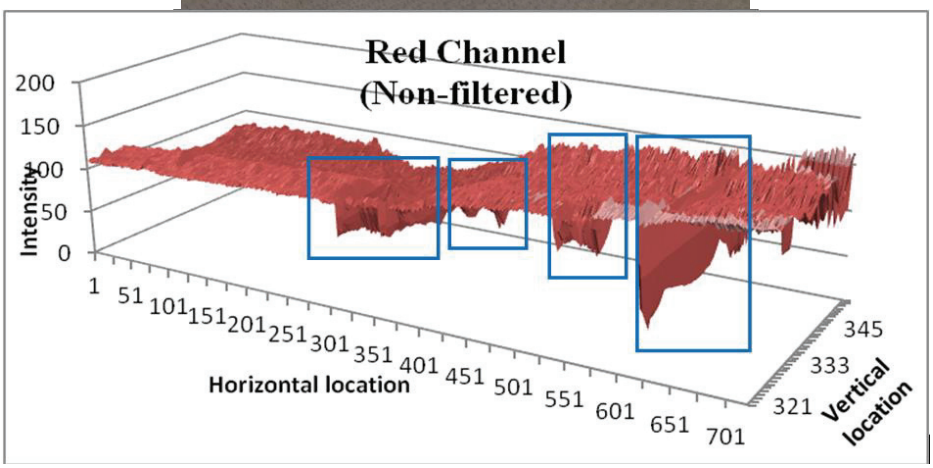


Fig.8-3 Histograms before and after equalization

Bilateral filter is applied to these images to eliminate the noise as explained in earlier chapters. The result of filtering to gray scale and histogram equalized gray scale images are displayed in Fig.8-4.

Fig.8-4 Application of bilateral filter to gray scale and histogram equalized image

The intensity data of a sea image (between orange dashed lines) is illustrated in Fig.8-5 for a better understanding of effect of processes.



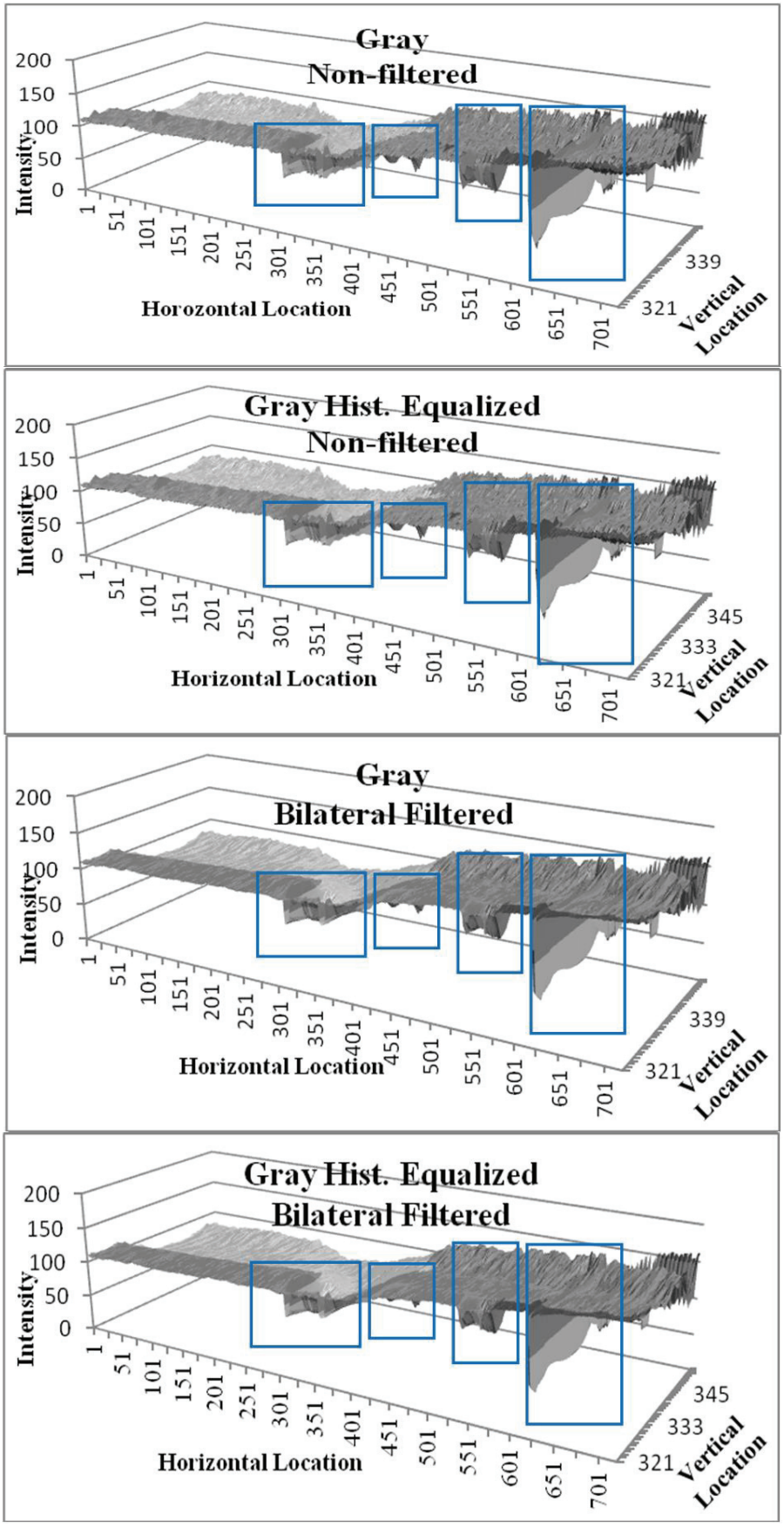
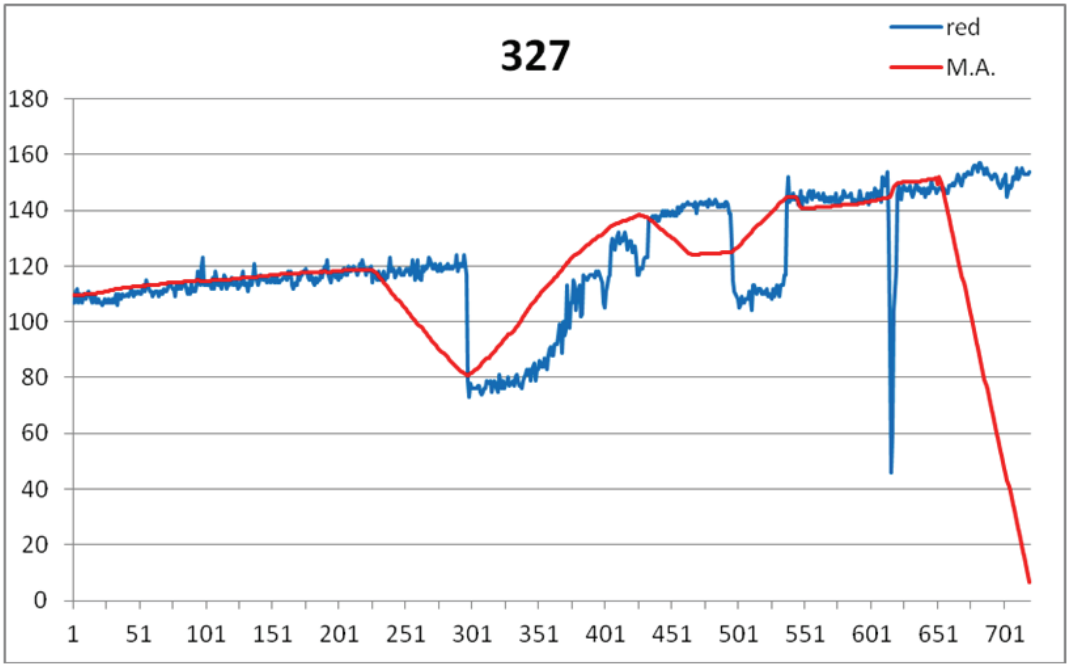
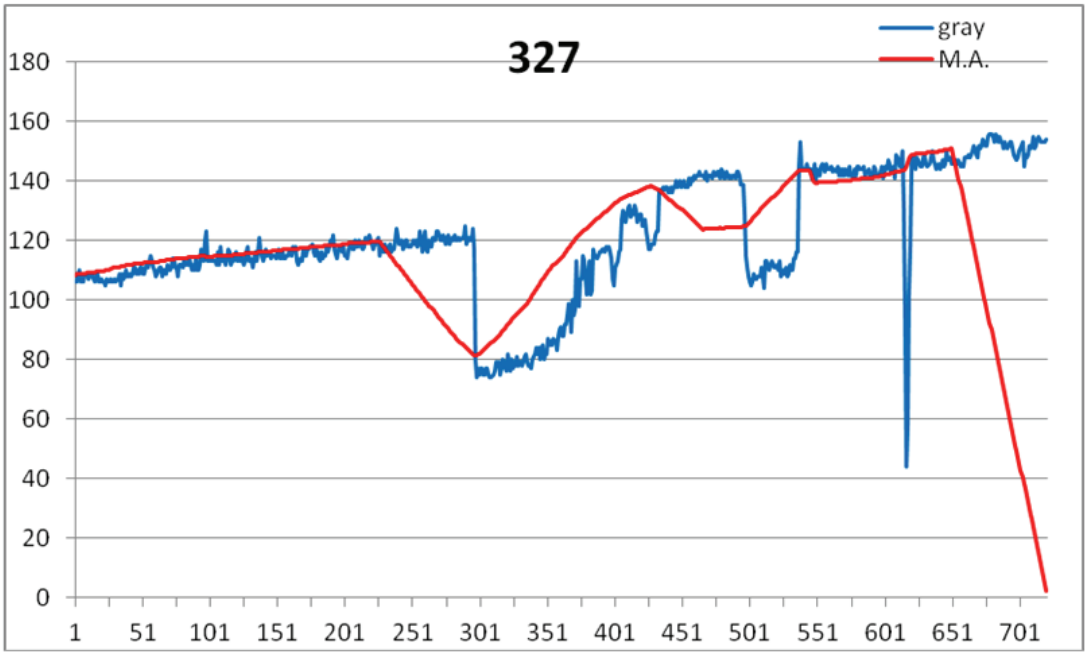


Fig.8-5 Intensity profiles of processed images

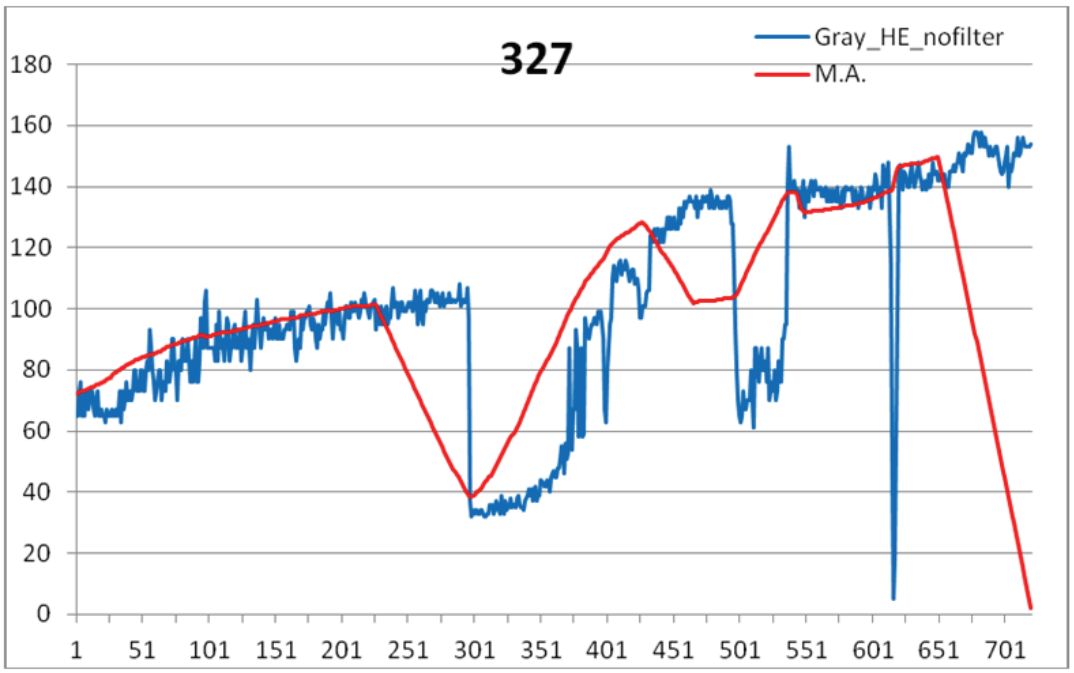
To show the effect of mentioned processes on detection algorithm only one scan line (327th row) of the images (yellow dashed line in Fig.8-5) are displayed in Fig.8-6. In the figure, vertical axis is the intensity data and the horizontal axis is pixel location through horizontal scan line.



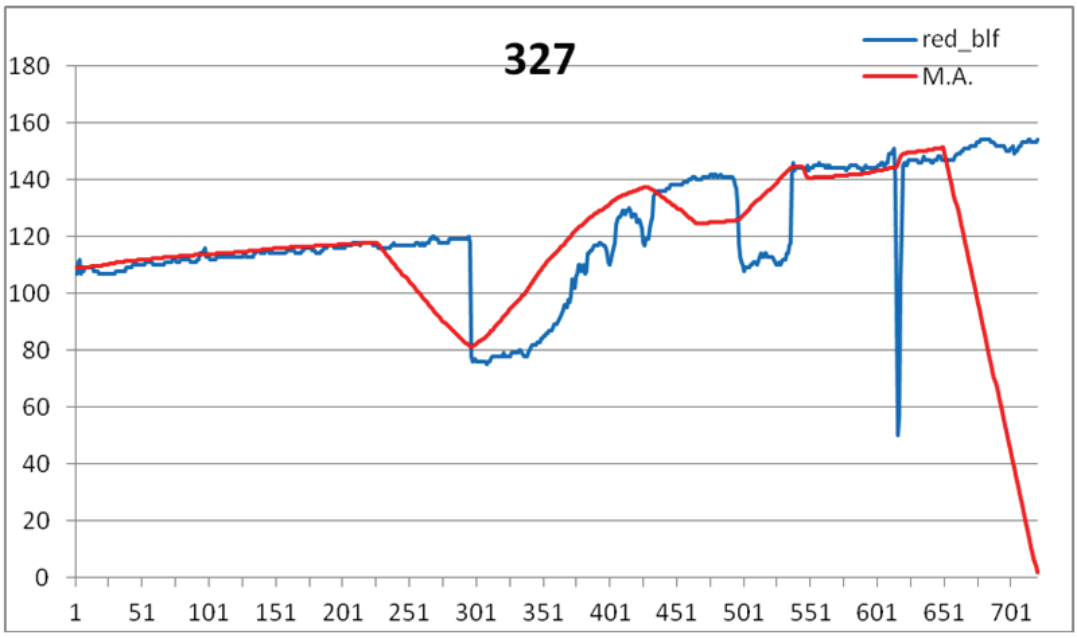
(a)



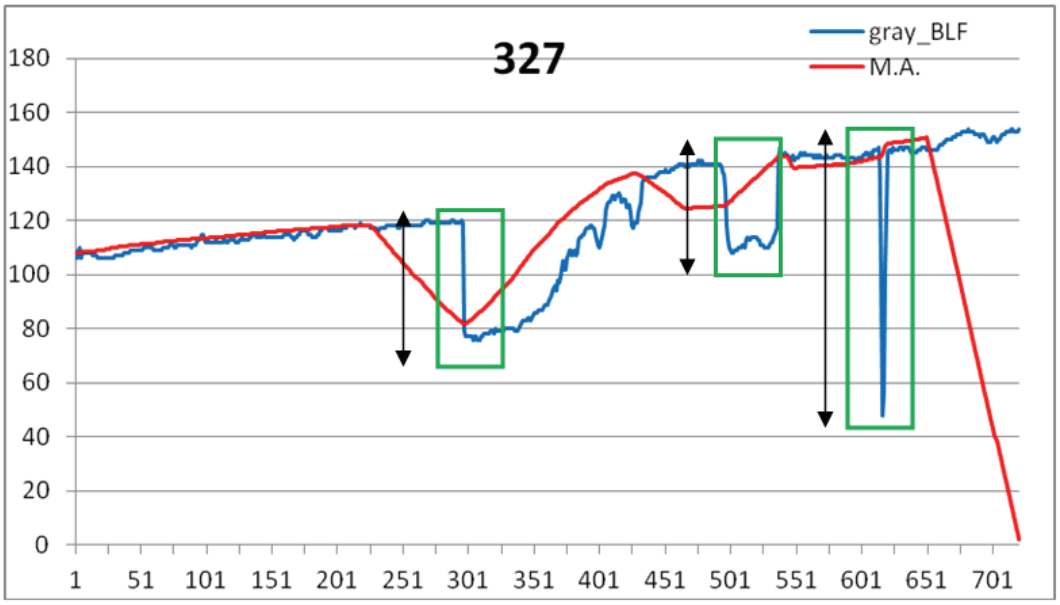
(b)


327
Gray_HE_nofilter
M.A.
180
160
140
120
100
80
60
40
20
0
1
51
101
151
201
251
301
351
401
451
501
551
601
651
701

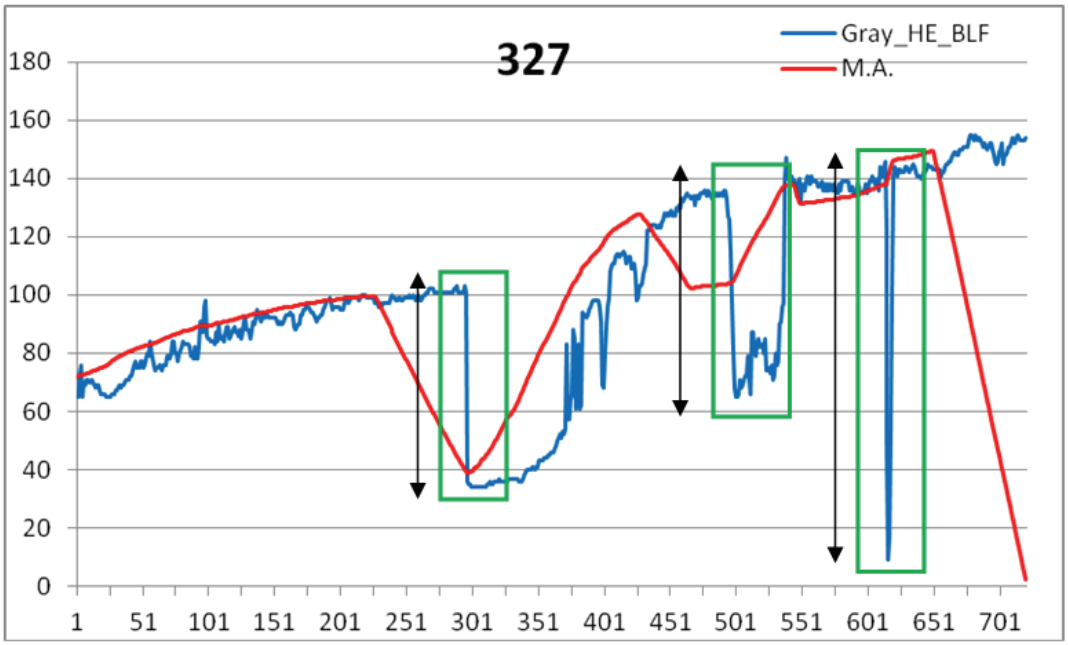
(c)


327
red_blf
M.A.
180
160
140
120
100
80
60
40
20
0
1
51
101
151
201
251
301
351
401
451
501
551
601
651
701

(d)



(e)



(f)

Fig.8-6 Effect of processes on detection algorithm (a) Red channel only (b) Gray scale (c) Histogram equalized gray scale (d) Red channel bilateral filtered (e) Gray scale bilateral filtered (f) Gray scale histogram equalized bilateral filtered

In Fig.8-6 we see that the intensity value of only red channel and gray scale images are almost same. Besides, the intensity gradients at edge locations are greater in the histogram equalized images which is expected to have positive effect on detection. The intensity

gradient change at edge locations are shown with green rectangles in Fig.2-7 (e) and (f). The detection results of some sea images are shown in Fig.8-7.

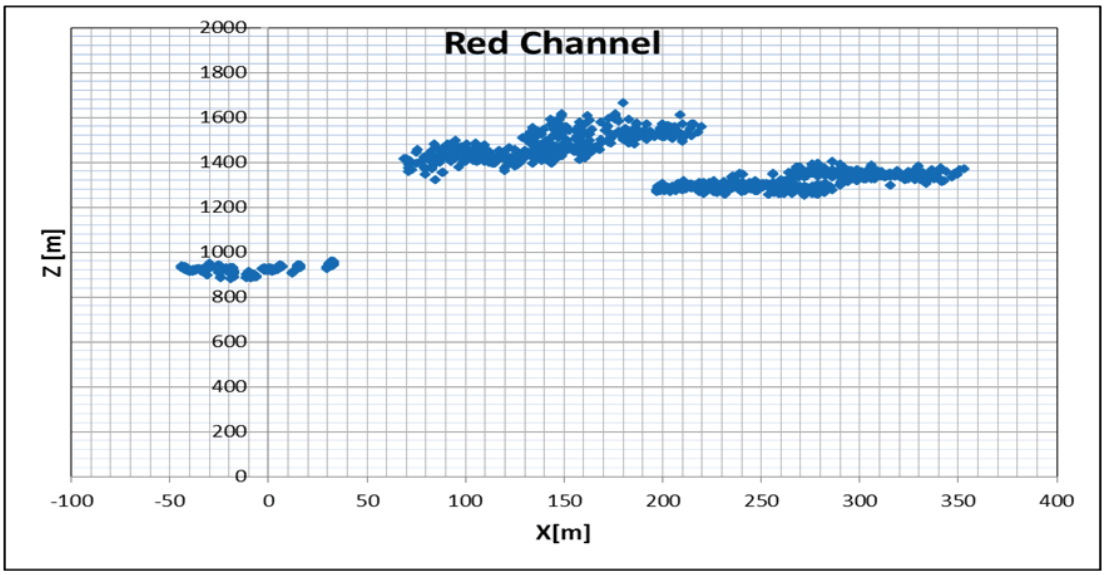
Fig.8-7 Detection results on (Top) Red channel only (Middle) Gray scale (Bottom) Histogram equalized gray scale

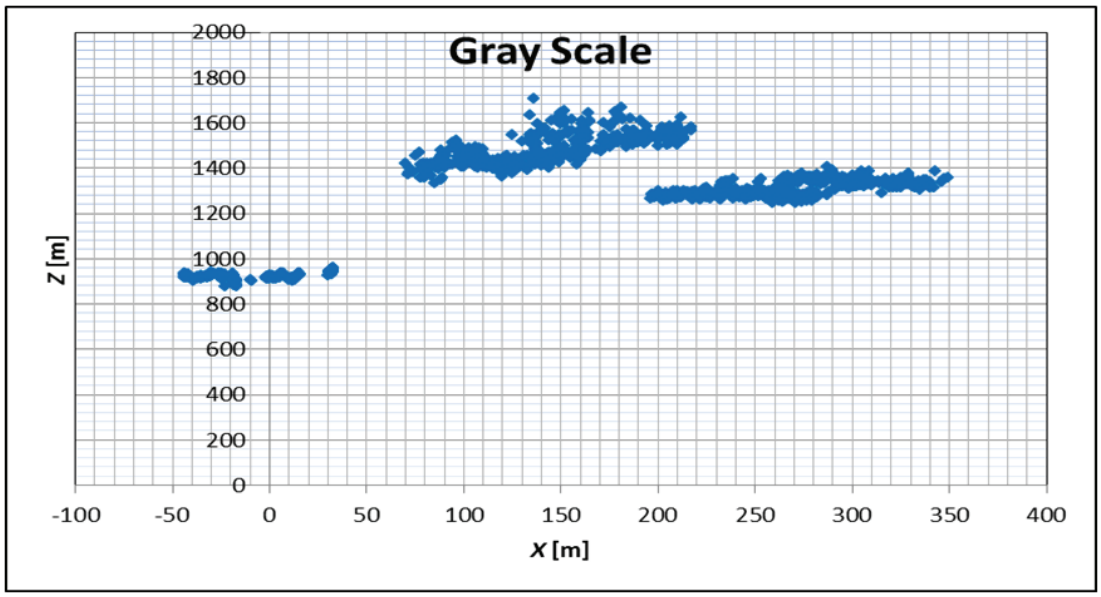
We can figure out that the detection result in this image is increased by histogram equalization but it is not as much as expected. The effect of histogram equalization is needed to be examined for different kind of sea images with different illumination conditions.

Actually, histogram equalization is using the obtained images under some illumination conditions and it cannot change the quality of the obtained image. But it can make it to view better for human visual perception. We may say that it is not same for computers. For the existing experimental image data the effect of histogram equalization is not as expected. Some more analyses with different illumination cases are considered as a future study.

8.2 Tracking

After the detection clustering is carried out and the 3D locations of these clusters are tracked through image sequences as explained in Chapter 6. The tracking results of red channel image, gray scale image and histogram equalized gray scale images are shown in Fig.8-8.





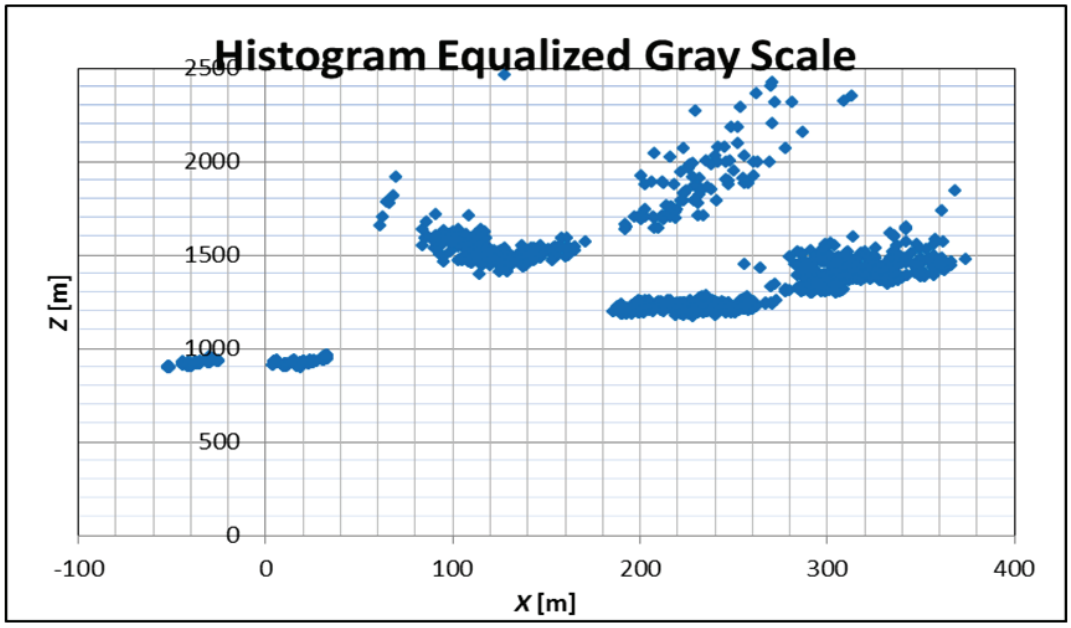


Fig.8-8 Tracking results of red channel (top), gray scale (middle) and histogram equalized gray scale (bottom) images

The tracking result of histogram equalized image is not as expected despite the detection result is better. The detected 3D locations are too much scattered. One reason of this result may be clustering algorithm. The number of detected points is increased in histogram equalized images. The effect of clustering algorithm on histogram equalized images may be studied later to increase the tracking accuracy. And the reason of bad tracking results should be investigated as a future study.

8.3 Effect of Some Matching Strategies on Tracking

The effects of some matching strategies are considered for increasing the matching quality and 3D location calculation. In Fig.8-9 we can see three images of a ship which are original (non-filtered) image, bilateral filtered image and median filtered image.

Fig.8-9 Non-filtered (top), bilateral filtered (middle) and median filtered (bottom) images

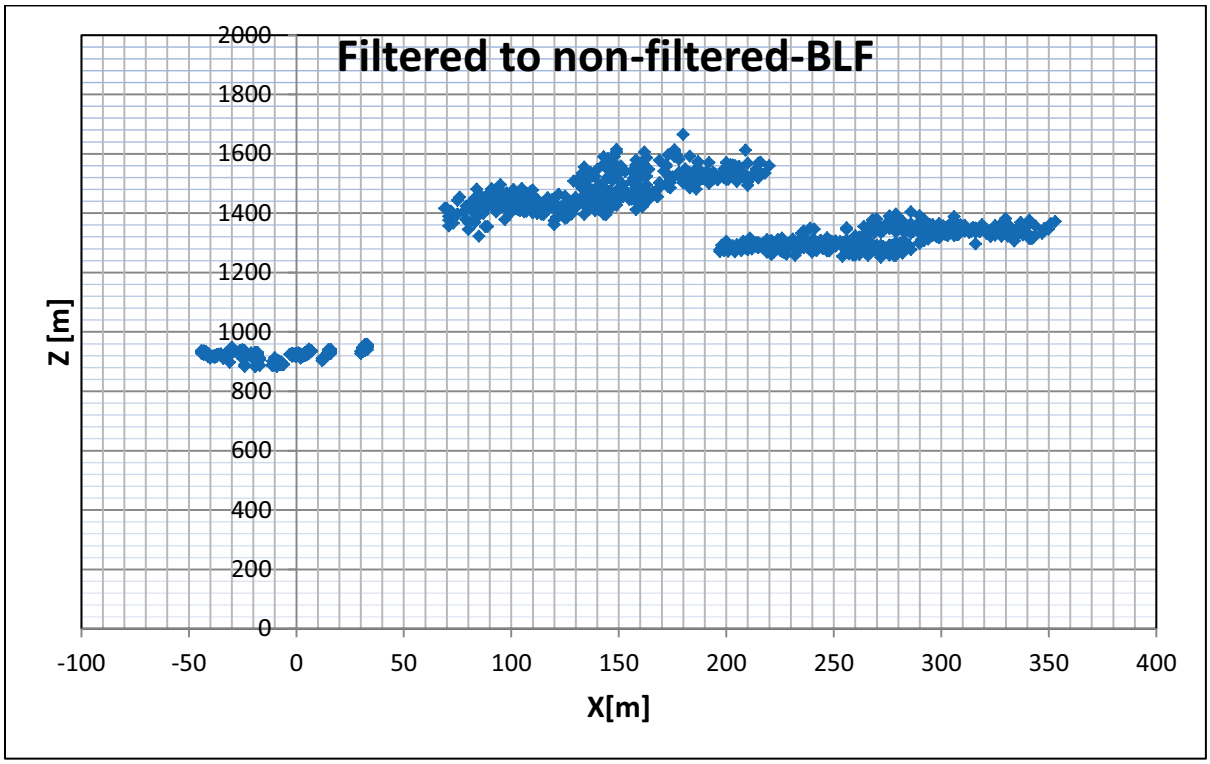
In the stereo matching corresponding point is searched to find a region which surrounds a detected point as the red square in the figure. We can see that filtering is affecting the intensity data which belongs to ship area and cause some deterioration of the ship shape. The effect of deterioration changes depending on the filter behavior. For example, median filter has more deterioration than bilateral filter which can be seen in Fig.8-10.

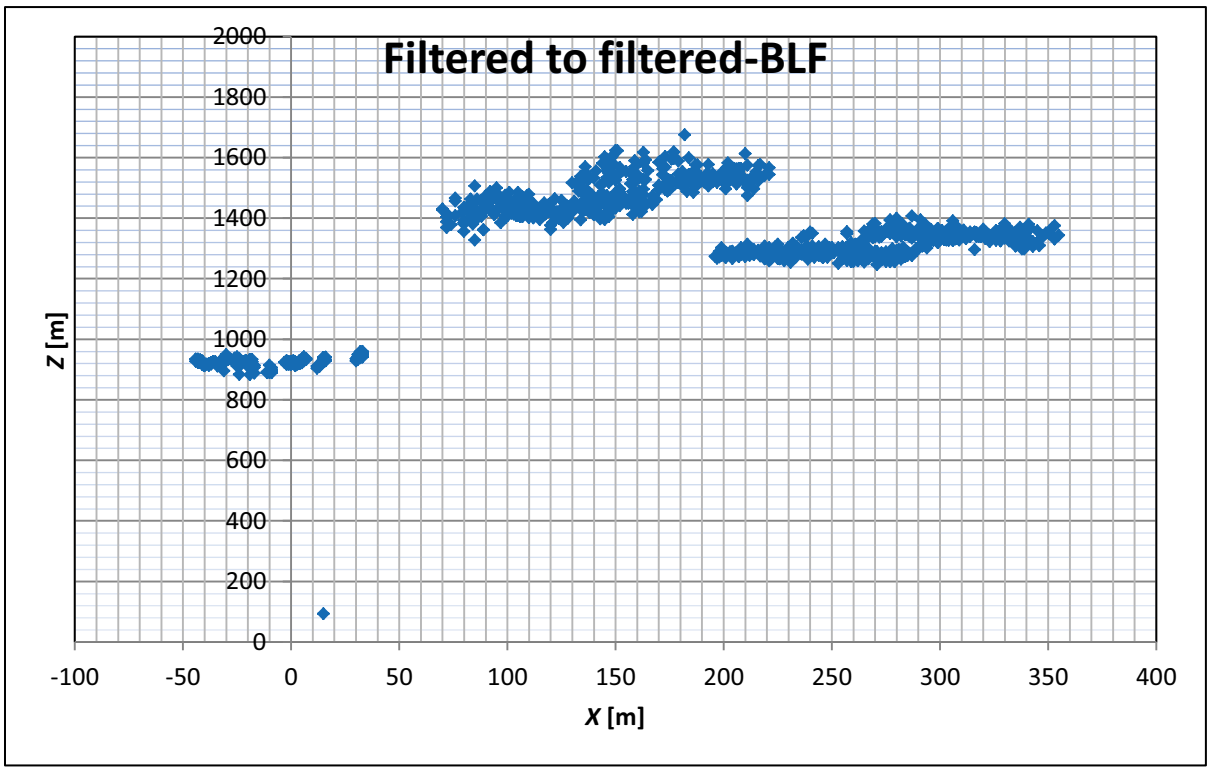
In this stereo algorithm the left image is the reference image. Filtering is important for the detection purpose. Therefore, the left image is filtered and detection is carried out to obtain the object points. Then these points are found in the right image by a similarity search as explained in Chapter 3. In this case the situation of right image should be considered. It can be filtered or non-filtered. Therefore, the matching between right and left stereo pair images is considered with respect to filtering. Three matching situations may be considered:

1. Filtered left image to non-filtered right image
2. Filtered left image to filtered right image
3. Non-filtered left image to non-filtered right image.

In case of non-filtered to non-filtered image matching, the detection is carried out after filtering the image. The location of detected point is used in non-filtered form of the left

image. Filter is bilateral filter through this section. Therefore the filtered image means the bilateral filtered image. The tracking results of these three situations are shown in Fig.8-10.





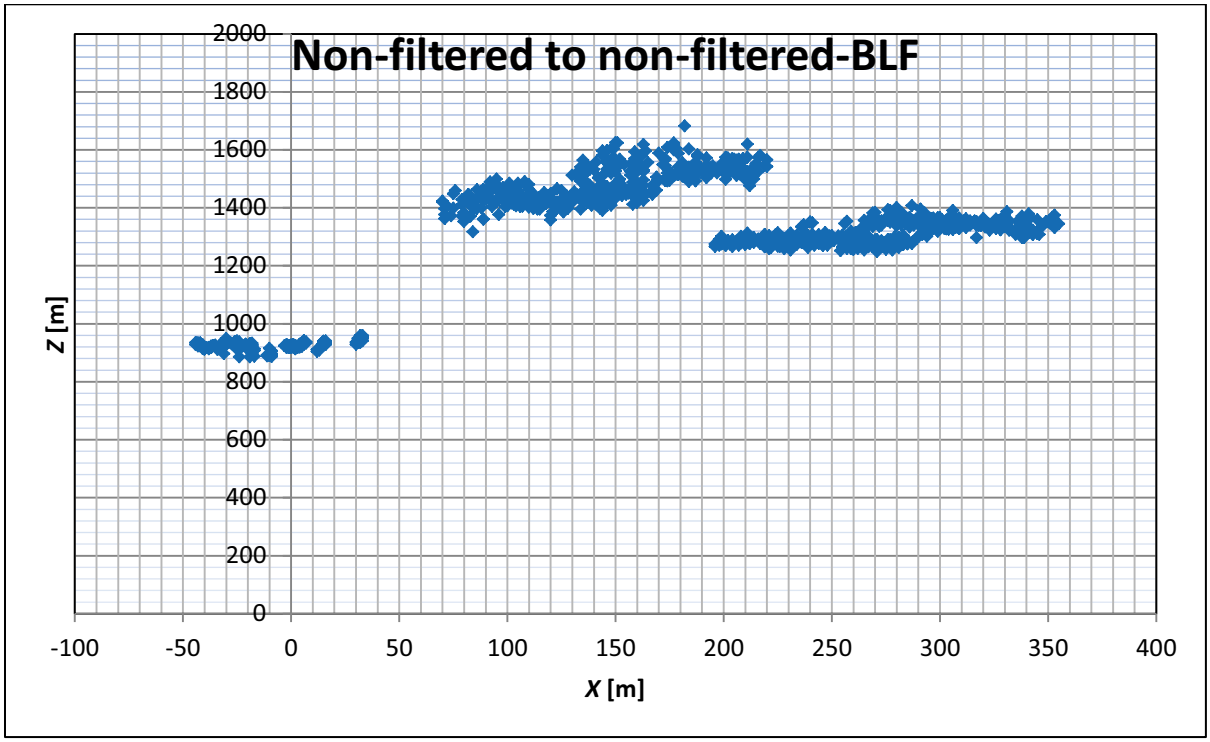
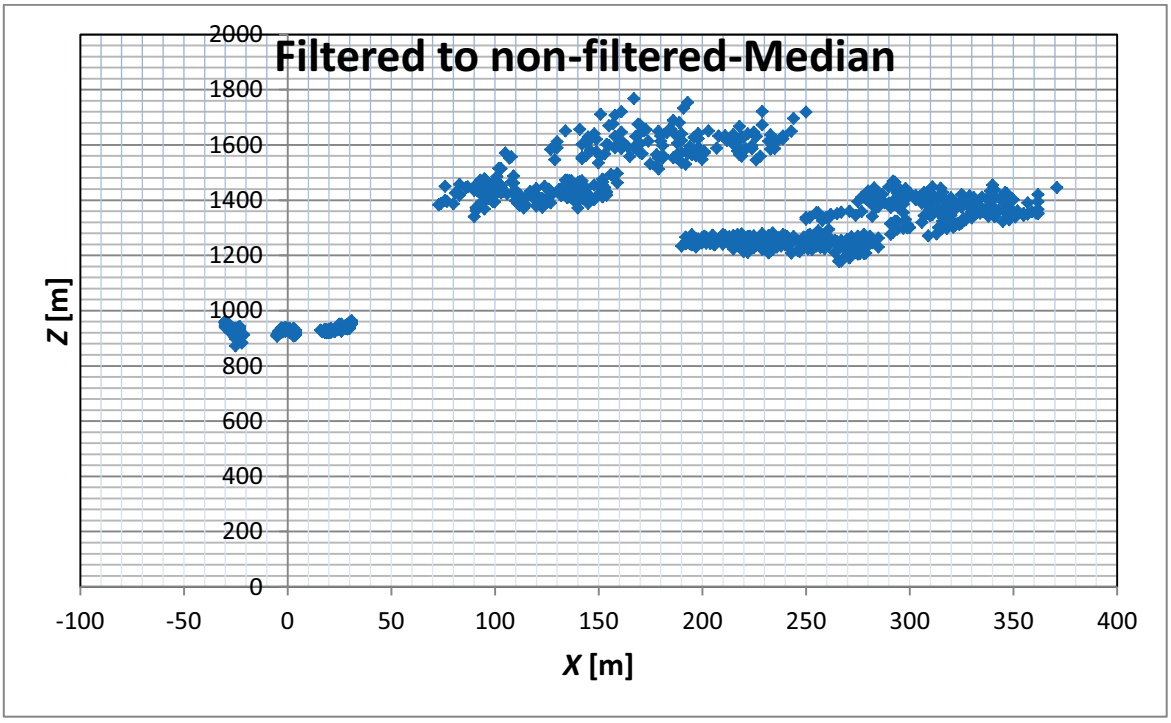


Fig.8-10 Tracking results of different matching situations

Inspecting the tracking results of different matching situations, it is seen that there is not a certain difference between them. One reason is that deterioration effect of bilateral filter is not much. However, in case of median filtered images a significant difference is observed which is displayed in Fig. 8-11.



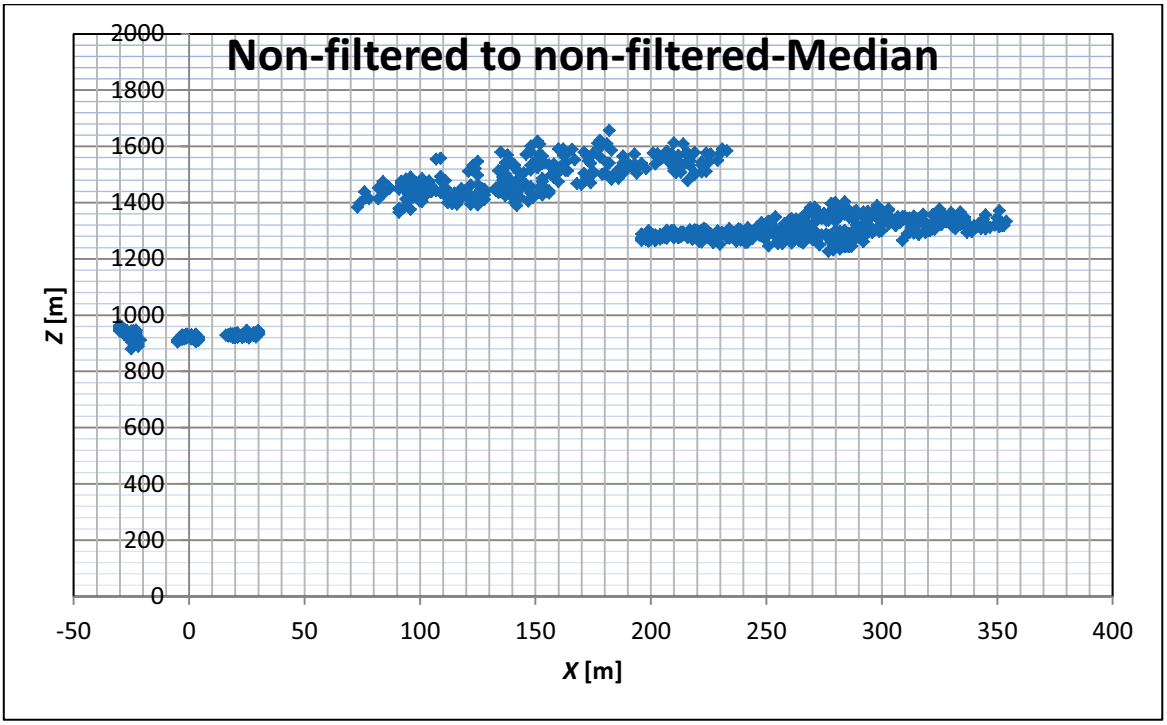


Fig.8-11 Tracking results for median filtered images

In the case of median filtered images the result of matching with non-filtered to non-filtered images is better. The main reason is that the median filtered images are deteriorated as explained in 8.3. Therefore, an image patch which is changed by median filter is searched through the right image. This may result in finding the corresponding point with a little bit error. When matching is carried out on both non-filtered left and right images, the image patches will be more similar (except for the difference caused by the different viewpoints of cameras). This situation is illustrated in Fig.8-12. The image patch around ship's bridge is compared in median filtered and non-filtered images. In the upper row of the figure median filtered left image and non-filtered right image are shown. In the lower row of the figure the left and right images are non-filtered. It is seen that the image patches of lower row is more similar comparing to image patches of upper row. Therefore, we can find the corresponding point with smaller error and the 3D calculation result will also be better which can be seen from Fig.8-11.

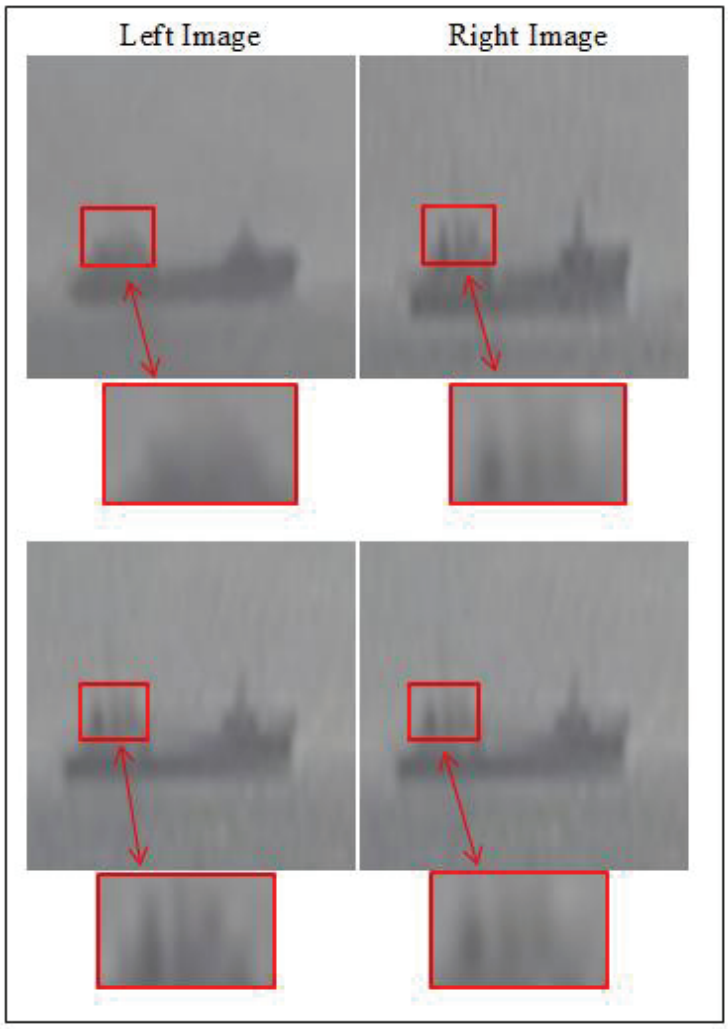


Fig.8-12 Effect of median filter to stereo matching

Considering the different types of filters and speed performance, non-filtered image to non-filtered image matching can be preferred. In this case the right image is not filtered because the computational cost is decreased and the computational speed is increased.

CHAPTER 9

CONCLUSION AND FUTURE RESEARCH

Obtaining the environmental data is very important for safety navigation. Statistical marine casualty analyses show that the existing navigation tools are not sufficient. And some tools are needed to improve the navigational information. Digital cameras are one of the most used tools for autonomous navigation of robots especially in the land based applications. Of course there are other sensors and the data fusion of these sensors is carried out and more reliable environmental data is obtained for localization and mapping of the robots. This can be used for ships, as well. In case of marine environment there are some disadvantages and advantages comparing to land based applications. For example sea surface glare is one of the problems which have great effect on object detection. Besides, sea images in open sea area are lack of texture. This is not good for camera self-calibration or stereo camera rectification. However, it is an advantage for ship detection because the ship can be easily separated from background.

In this study we described a method to detect, measure, and track ships from sea images using stereo vision for the purpose of safe navigation. The algorithm used in the method is mainly as follows.

1. Detect the point at which the intensity of the image significantly changes in the horizontal direction as salient points.
2. Find the corresponding points of the salient points between a pair of stereo images.
3. Measure the 3D location of the detected points.
4. Cluster the points.
5. Track the clusters through image sequences.

The merit of using stereo vision is not only that it can measure the 3D locations of objects, but also it is useful for the detection of the objects. That is, a corresponding problem between

a pair of stereo images can distinguish clear objects, such as ships, from unclear objects, such as waves. In addition, tracking the objects makes the distinction clearer because ships can continuously be detected through image sequences.

Main contributions of the study can be summarized as:

1. A wide baseline stereo configuration is studied for detection and tracking purpose. It is for capability of long range measurement. There are a few stereo vision studies for maritime field which are applied for autonomous surface vessels. These vessels are usually small boats which have high maneuvering capability. The maximum baseline was 1 meter for these studies which is planned to work in 400 meter range. This is because that range is sufficient for that kind of vessels to maneuver. However, the merchant vessel size has a very wide range between around 50 meters to 350 meters. And the maneuvering is more time taking. That's why longer range detection is necessary. Besides, the ship size is large and 8 meters baseline camera set-up is feasible.
2. The 3D location measurement is improved by applying sub-pixel matching and low-pass filtering. It is possible to calculate the range of ship with 1% error with a 720x480 resolution image at around 2500 meters with 10 meters baseline stereo configuration and 5000 meters with 20 meters baseline stereo configuration which is sufficient for locating ships. When the distance is decreased the accuracy is increasing due to quantization of images.
3. A theoretical explanation for the estimation of the accuracy of the 3D calculation in various conditions (view angle of the camera, baseline length, resolution of the image, distance of the target ship, etc.) is carried out. This result enables us to choose proper parameters to configure the stereo system for practical use.

4. The sea surface has a very dynamic texture which makes it very difficult to apply computer vision algorithms. Even more the glares on sea surface are big disadvantage comparing to land based computer vision applications. Effects of some different smoothing filters are examined for sea images to eliminate noise due to mentioned disadvantages.
5. Effect of the mentioned smoothing filters on object detection and horizon detection is shown. A tradeoff between smoothing the sea surface and detecting the objects is displayed. A smoothing filter which is preferable for both purposes is recommended.
6. A new clustering algorithm is proposed for clustering detected points. The clustering algorithm is capable of tracking the clusters through image sequences. The tracking results of clusters are analyzed.
7. The feedback of the detected points of new image to the previous clusters is constituted. Therefore, the false detections due to sea surface glare are eliminated which means the improvement of detection algorithm, as well.
8. The detected edges are classified as (+) or (-) edges for increasing the tracking accuracy.
9. A horizon detection algorithm which is invariant to rotation (by ship rolling) and illumination changes is shown.
10. An adaptive threshold is proposed for improving edge detection in different illumination situations.
11. Effect of ship movement and vibration to stereo calculation is studied. A methodology is proposed for elimination of such environmental effects on stereo system which is based on self-extrinsic calibration of cameras.
12. The detection algorithm is examined in color image, gray scale image and histogram equalized gray images. Effect of tracking is also shown.

13. The matching between stereo pairs or image sequences is examined for

- Filtered image to non-filtered image
- Filtered image to filtered image
- Non-filtered image to non-filtered image

Non-filtered image to non-filtered image matching (after detection) is found to be more effective for tracking the 3D locations of detected points. Even more it eliminates the necessity for filtering the second image.

Our experimental results indicate the potential of the proposed method, although there are some points to improve. First, the accuracy of the 3D measurement is sometimes insufficient, as it heavily depends on the precise calibration of a stereo camera system. That is, we must obtain intrinsic parameters such as the focal length and lens distortions of the cameras and extrinsic parameters, which include the locations and orientations of the cameras. Second, detected points are solely clustered and tracked by closeness of coordinates and disparity in the image. Using a Kalman filter or particle filter is considered to estimate the movement of the ship which will improve the tracking and will enable estimation of a ship's future course.

The rainy, foggy weather conditions and day time, night time situations are in consideration. A similar stereo configuration with infrared cameras is planned to investigate such kind of situations. Besides, we would like to carry out the same stereo configuration with wide angle lenses (~180°) which may obtain more visual information and monochrome camera to compare the efficiency with the current study.

Ship size is one of the problems for detection. It may span between 15 meters to 350 meters. In case of a large ship the distance between detected points may be very large. It is important for forming the instantaneous map. Making a map of environment which includes the tracking data of the ship and the whole ships 3D shape of that moment as shown in Fig.9-1 is planned as a future work.

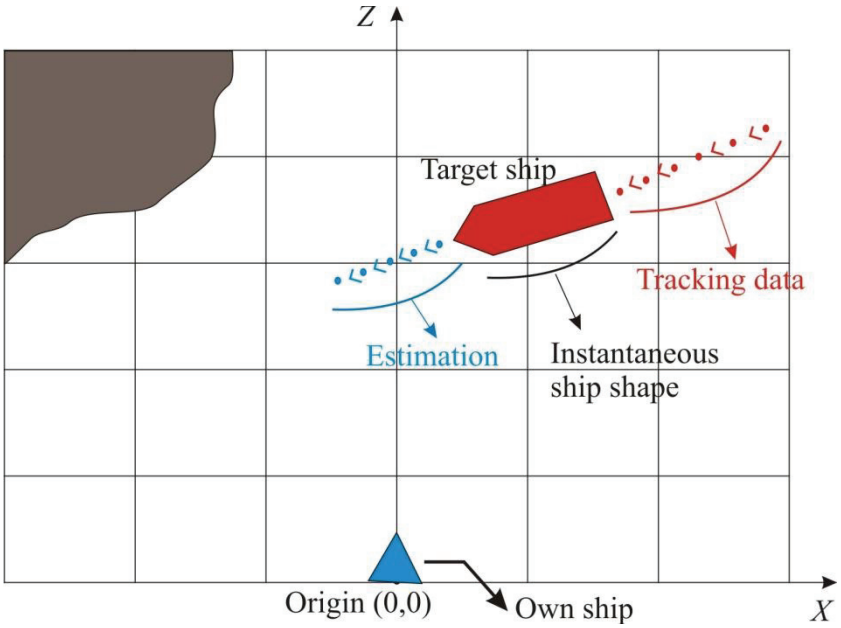


Fig.9-1 Map of ship shape and tracking data

Publication List

[Journal Papers]

1. Kocak G, Yamamoto S, Hashimoto T (2012) Analyzing Influence of Ship Movement to Stereo Camera System Set-up on Board Ship. Journal of Japanese Institute of Marine Engineers (JIME). Vol 47, No.6, pp. 111-118.
2. Kocak G, Yamamoto S, Hashimoto T (2013) Detection and Tracking of Ships using a Stereo vision System. Scientific Research and Essays, Vol. 8(7), pp.288-303. DOI: 10.5897/SRE12.318.

[Proceedings of International Conference]

1. Yamamoto S, Kocak G, Shimizu K (2009). Accuracy evaluation of localization of ships using stereo vision system. The 79th Conference of Japanese Institute of Marine Engineering, pp.41-42.
2. Kocak G, Yamamoto S (2009-a). A feasibility study on precise localization of ships using stereo vision. Conference of the Institute of Systems, Control and Information Engineers (ISCIE'09).
3. Kocak G, Yamamoto S (2009-b). Accuracy improvement of 3D location measurement of ships using stereo vision. 13th Congress of Intl. Maritime Assoc. of Mediterranean (IMAM 2009).
4. Kocak G, Shimizu K, Yamamoto S, Hashimoto T (2010) Detection and Tracking of Ships Using a Stereo Vision System. In: Proceedings of Techno Ocean Conference 2010, Kobe, Japan.
5. Kocak G, Yamamoto S, Hashimoto T (2011) Analyzing the influence of vibration to a stereo camera system set-up onboard ship. In: Proceedings of International Symposium on Marine Engineering (ISME), Kobe, Japan.
6. Kocak G, Yamamoto S (2012) Effect of image processing techniques on ship detection and tracking from digital cameras. In: Proceedings of International Symposium Information on Ships (ISIS 2012), Hamburg, Germany. Paper no: 0022.

REFERENCES

Alves J, Herman J, Rowe NC (2004) Robust Recognition of Ship Types from an Infrared Silhouette. Command and Control Research and Technology Symposium

Benjamin RM, Curcio J, Leonard JJ, Newman PM (2006) A Method for Protocol-Based Collision Avoidance Between Autonomous Marine Surface Craft. Journal of Field Robotics, 23(5), pages 333-346.

Blostein SD, Huang TS (1987) Error analysis in stereo determination of 3-D point positions. IEEE Trans. On Pattern Analysis and Machine Intelligence Vol. 9, No.6

Bouma H, Lange DJ, Broek SP, Kemp RAW, Schwering PBW (2008) Automatic Detection of Small Surface Targets with Electro-Optical Sensors in a Harbor Environment. In: Proceedings of SPIE, Electro-Optical Remote Sensing, Photonic Technologies, and Applications II, vol 7114

Brown MZ, Burschka D, Hager G (2003) Advances in Computational Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No: 8

CO2 Emissions 2010 Report. United Nations Framework Convention on Climate Change (COP16)--International Maritime Organization--Marine Environment Protection Committee

Crisp DJ (2004) The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery. DSTO (Defence Science and Technology Organization Salisbury, Australia) Information Sciences Laboratory.

CWFS, 2012. Casualty and World Fleet Statistics (January 2012)- Clarkson Research, January 2012

Duda RO, Hart PE (1972) Use of Hough Transformation to Detect Lines and Curves in Pictures. Comm. ACM, vol.15, pp.11-15

Dunbabin M, Lang B, Wood B (2007) Towards coordinated visionbased docking using an autonomous surface vehicle. In: *Proc. Australian Conf. Robotics and Automation*, Brisbane, Australia

Dunbabin M, Lang B, Wood B (2008) Vision-based Docking Using an Autonomous Surface Vehicle. IEEE International Conference on Robotics and Automation, CA, USA

Ebken J, Bruch M, Lum J (2005) Applying UGV technologies to unmanned surface vessels. In: SPIE proc. 5804, Unmanned ground vehicle technology VII

EMSA, 2010. Maritime Accident Review 2010, European Maritime Safety Agency

Ettinger SM, Nechyba CM, Ifju PG, Waszak M (2003) Vision-Guided Flight Stability and Control for Micro Air Vehicles. Advanced Robotics, vol.3, no.7, pp.617-40

Faugeras O (1993) Three-dimensional Computer Vision: A Geometric Viewpoint, MIT Press

Fefilatyev S, Goldgof D (2008) Detection and Tracking of Marine Vehicles in Video. In: Proccedings of 19th International Conference on Pattern Recognition (ICPR)

Fefilatyev S, Goldgof D, Lembke C (2009) Autonomous Buoy Platform for Low-Cost Visual Maritime Surveillance: Design and Initial Deployment. Ocean Sensing and Monitoring, Proc. of SPIE Vol. 7317, 73170A

Fefilatyev S, Goldgof D, Lembke C (2010) Tracking Ships from Fast Moving Camera through Image Registration. In: Proceedings of International Conference on Pattern Recognition

Fefilatyev S, Smarodzivana V, Hall LO, Goldgof DB (2006) Horizon Detection Using Machine Learning Techniques. ICMLA'06 Proc. Of 5th International Conference on Machine Learning and Applications, pp. 17-21

France WN, Levadou M, Treakle TW, Paulling JR, Michel RK, Moore C (2001) An Investigation of Head-Sea Parametric Rolling and its Influence on Container Lashing Systems. SNAME Annual Meeting

Frankot R, Chellappa R (1988) A method for enforcing integrability in shape from shading algorithms. IEEE Trans. On Pattern Analysis and Machine Intelligence, VOL 10**,** NO 4, pp. 439-451

Furusho M, Kawamoto K, Yano Y, Sakamoto K (2011)Visaul Condition at Seafor the Safety Navigation. International Journal of Marine Navigation and Safety of Sea Transportation, Volume 5, No 1, pp 59-64.

Gong X, Xu B, Reed, C, Wyatt C, Stilwell D (2008) Real-time Robust Mapping for an Autonomous Surface Vehicle using an Omni-directional Camera. IEEE Workshop on Applications of Computer Vision (WACV 2008), On Page(s): 1 – 6

Gonzalez RC, Woods RE (2008) Digital Image Processing. 3rd ed, Prentice Hall

Hartley RI, Zisserman A (2003) Multiple View Geometry in Computer Vision, 2nd ed. Cambridge Uviversity Press.

Hayashi Y, Wakabayashi N, Kitahashi T, Wake H (1994-a) An Image Ranging System at Sea. IEEE Position Location and Navigation Symposium, pp. 113-120

Hayashi Y, Wakabayashi N, Kitahashi T, Wake H (1994-b) 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

Heikkilä J and Silvén O (1997) A Four-step Camera Calibration Procedure with Implicit Image Correction. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1106 – 1112

Huntsberger T, Aghazarian H, Castano A,Woodward G, Padgett C, Gaines D (2011) Intelligent Autonomy for Unmanned Sea Surface and Underwater Vehicles. In: Proceedings of OCEANS 2011 On Page(s): 1 – 10

ISL-SSMR, 2011. ISL Institute of Shipping Economics and Logistics –Shipping Statistics and Market Review—Volume 55 No1/2 – 2011

Jain R, Katsuri R, Shunck BG (1995) Machine Vision, McGrawHill Inc.

JCG, 2007. Maritime accidents and rescues in 2007, Japan Coast Guard

Ji Zhong J, Su Y, Wang J, Hua R (2009) Robust Sea-sky-line Detection Based on Horizontal Projection and Hough Transform. CISP'09. 2nd International Conference on Image and Signal Processing, pp. 1-4

Jiang C, Jiang H, Zhang C, Wang J (2010) A New Method of Sea-sky-line Detection. IEEE 3rd Int. Symp. on Intelligent Info. Tech and Security Informatics, pp. 740-743

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

Julesz, B (1960) Binocular depth perception of computer-generated patterns. Bell system Technical Journal, Vol. 39, No. 5, pp. 1125-1162

Kang WJ, ding XM, Cui JW, Ao L (2006) research on Extraction of Ship Target in Complex Sea-sky Background. In: Proceedings of International Symposium on Instrumentation Science and Technology, Journal of Physics:Conference Series 48, pages 354-358.

KMST. Korean Maritime Safety Tribunal, Maritime Accident Statistics http://www.kmst.go.kr/eng/cms/cms.asp?code=DA

Koschan A (1993) What is new in computational stereo since 1989: A survey on current stereo papers. Technical report 93-22, Technical University of Berlin

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

Liu H, Javed O, Taylor G, Cao X, Haering N (2008) Omni-Directional Surveillance for Unmanned Water Vehicles. International Workshop on Visual Surveillance (VS2008)

Longuet-Higgins HC (1981) A computer algorithm for reconstructing a scene from two projections. Nature, Vol. 293, pp. 133-135

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

MAIA, 2007. Japan Maritime Accident Inquiry Agency : “Report on Maritime Accidents 2007”.

MAIA, 2008. Japan Maritime Accident Inquiry Agency, MAIA Digest: Collisions with Fishing Boats under Operation,September 2008

Martins A, Almeida J, Ferreira H, Silva H, Dias N, Dias A, Almeida C, Silva E (2007) Autonomous surface vehicle docking manouvre with visual information. In: Proc. of International Conference on Robotics and Automation, pp. 4994–4999

Marr D (1974) A note on the computation of binocular disparity in a symbolic, low-level visual processor. MIT Artificial Intelligence Lab. Memo 327

Marr D, Poggio T (1976) Cooperative computation of stereo disparity. Science 194, pp. 283-287

Marr D, Poggio T (1977) A theory of human stereo vision. MIT Artificial Intelligence Lab. Memo 451

McKeown DM, Perlant FD (1992) Refinement 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

Mohr R, Triggs B (1996) Projective Geometry fo Image Analysis, A Tutorial given at ISPRS, Vienna, July 1996

Perera LP, Carvalho JP, Guedes Soares C (2011). Fuzzy logic based decision making system for collision avoidance of ocean navigation under critical collision conditions. J Mar Sci Technol 16:84-89.

Pratt WK (2007) Digital Image Processing, Fourth Edition. Wiley Publication

Qi B, Wu T, He H, Hu T (2010) Real-time detection of small surface objects using weather effects. In: Proceedings of the 10th Asian conference on Computer vision (ACCV 2010)

Santhalia GK, Singh S, Singh SK (2008) Safer Navigation of Ships by Image Processing & Neural Network. 2nd Asia International Conference on Modeling & Simulation

Sato Y, Ishii H (1998) Study of a collision-avoidance system for ships. Control Engineering Practice, Vol. 6, pp. 1141-1149

Schartein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, Vol. 47, No. 1, pp. 7-42.

Shen Q, Zhang Y, Chen X (2005) Investigation of the Nonlinear Rolling of a Ro-Ro Ship with Slipping of Heavy Loads. Multibody System Dynamics, pp.189-203

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

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

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

Shimpo M, Shu R, Oshima M (2007) Method for Real-Time Detection of Ships using Navigational Image Sequence. Proceedings of Asia Navigation Conference 2007, pp.105-113

Shimpo M, Shu R, Yamamoto S (2008) Investigation of Image Processing Methods Capable of Supporting Navigational Lookout. Asia Navigation Conference November, 2008.

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

Smith FW, M.H. Wright MH (1971) Automatic Ship Photo Interpretation by the Method of Moments. IEEE Trans. Computers, vol. C-20, pp. 1089-1095

Snyder FD, Morris DD, Haley PH, Collins R, Okerholm AM (2004) Autonomous River Navigation. In: Proceedings of SPIE, Mobile Robots XVII, December, 2004, pp. 221-232

Socek D, Culibrk D, Marques O, Kalva H, Furht B (2005) A Hybrid Color-Based Foreground Object Detection Method for Automated Marine Surveillance. Advanced Concepts for

Intelligent Vision Systems, Lecture Notes in Computer Science, Volume 3708/2005, 340-347

Statheros T, Howells G, McDonald-Maier K (2008). Autonomous Ship Collision Avoidance Navigation Concepts, Technologies and Techniques. The Journal of Navigation, 61,129-142.

Sumimoto T, Kuramoto K, Okada S, Miyauchi H, Imade M, Yamamoto H, Arvelyna Y (2000) Detection of a particular object from environmental images under various conditions. Proceedings of the IEEE International Symposium on Industrial Electronics ISIE, Vol.2, On Page(s): 590 - 595

Szpak LZ, Tapamo JR (2011) Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set. Expert Systems with Applications 38 (2011) 6669–6680

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

Todorovic S, Nechyba MC (2004) A Vision System for Intelligent Mission profiles of Micro Air Vehicles. IEEE Transactions on Vehicular Technology, vol. 53, no. 6, pp. 1713-1725

Tsai RY (1987) 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, pp. 323-344

Tomasi C, Manduchi R (1998). Bilateral filtering for gray and color images. In: Proceedings of International Conference on Computer Vision (ICCV), pp. 839–846.

Trucco E, Verri, A (1998) Introductory Techniques for 3-D Computer Vision. Prentice Hall

Voles P, Teal M, Sanderson J (1999) Target identification in a complex maritime scene. IEE Colloquium on Motion Analysis and Tracking, IEE Digest / Volume 1999 / Issue 103

Voles P, Smith AAW, Teal MK (2000) Nautical Scene Segmentation Using Variable Size Image Windows and Feature Space Reclustering. 6th European Conference on Computer Vision ECCV 2000, Lecture Notes in Computer Science, Volume 1843/2000, 324-335

Wang A, Su W, Wang C, Dong Y (2010) Adaptive segmentation algorithm for ship target under complex background. 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), Volume**:** 2, On Page(s): 219-223

Wang D, Chen W, Chen T, Chen TY (2008) 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, Page(s):655-658

Yamamoto S, Win MT (2006) A Feasibility study on ship detection using stereo vision. In: Proceedings of Techno-Ocean 2006 / 19th JASNAOE Ocean Engineering Symposium, Paper No. 38

Yamamoto S, Win T (2007) Basic Study on Detection of Ships from Stereo Images. Journal of the Japan Institution of Marine Engineering, Vol.42, No.4, pp.713-718

Yang Q, Tan K, Ahuja N (2009) Real-Time O(1) Bilateral Filtering. IEEE Conference on Computer Vision and Pattern Recognition, pp. 557-564

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

Zhang W, Zhuang P, Elkins L, Simon R, Gore D, Cogar J, Hildebrand K, Crawford S, Fuller J (2009) A stereo camera system for the Autonomous Maritime Navigation. In: Proceedings of SPIE, Unmanned Systems Technology XI. Volume 7332 (2009)., pp. 733216-733216-8

Zhang Z (2000) A Flexible New Technique for Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, pp 1330-1334