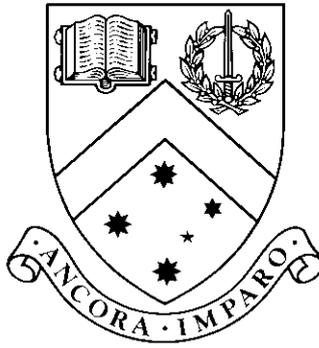


Scene Analysis for Robotic Watercraft

by

Rahul Walia,
B.Tech (Mechanical),
M.Sc (Computer Integrated Manufacturing)



Thesis

Submitted by Rahul Walia

for fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Supervisor: Prof. Raymond.A. Jarvis

Associate Supervisor: Prof. David Suter

**Electrical and Computer Science Engineering
Monash University**

June, 2012

© Copyright

by

Rahul Walia

2012

Contents

List of Figures	vii
Abstract	ix
1 Introduction	1
1.1 Navigation comparison : Land versus Water Robots	2
1.1.1 Environmental Modeling and Localization	3
1.1.2 Path planning	4
1.1.3 Motion control	5
1.2 Vision comparison: Land versus Water Robots	6
1.3 Applications	8
2 Literature Review: Water Surface	10
2.1 Photonics	11
2.2 Water Color	12
2.3 Hydrologic Optics	14
2.4 Water Dynamics	16
2.5 Inferences	18
3 Literature Review: Image Processing	20
3.1 Background Segmentation	20
3.1.1 Temporal Segmentation	21
3.1.2 Spatial Segmentation	25
3.1.3 Conclusion	28
3.2 Water-Based Navigation	28
3.3 Impact of literature review on research pathway	30

4	Sky and water: Pseudo Spectra Images	32
4.1	Background	32
4.2	Generating Pseudo Spectra Images from RGB	35
4.2.1	Definition and Derivation: PSI	37
4.2.2	Mapping (M) from RGB to XYZ	38
4.2.3	Orthogonal Wavelength Basis	38
4.2.4	Algorithm: Sampling Spectrum for Generating PSI	39
4.2.5	Application of PSI: Sky Detection	40
4.2.6	PSI: Characteristics and Discussion	42
4.3	Horizon Detection by Fitting Ellipses	43
4.4	Algorithm: Water Horizon from Derivative of PSI	46
4.5	Conclusion	48
5	Locating Objects in Water: Theory	49
5.1	Introduction	50
5.2	Statistical Distributions of a Gaussian Function	52
5.2.1	PDF of a Gaussian Function	54
5.2.2	Bimodality of the PDF of a Gaussian Function	55
5.2.3	Unbalance in modes of the PDF of a Gaussian Function	56
5.2.4	Scale Life of the GMSS of a Discontinuity	57
5.2.5	Comments	58
5.3	OT: Unbalanced Histograms	59
5.4	OT for a Gaussian Function	64
5.4.1	Definition and statistics: IC and NC	65
5.4.2	Applicability of Theorem 7	66
5.5	OT for continuous functions	69
5.6	Conclusion	69
6	Locating Objects in Water: Application	71
6.1	Heuristic	72
6.2	Algorithm	74
6.3	Algorithm Scope	75
6.3.1	Background Intensity variation	75

6.3.2	Background Frequency variation	76
6.3.3	Foreground Frequency variation	76
6.3.4	Foreground Intensity variation	77
6.3.5	Foreground Size variation	78
6.3.6	Foreground against Background of water	80
6.4	Algorithm Performance	81
6.5	Scale Detection	81
6.6	Conclusion	84
7	Conclusion and Future Research	90
7.1	Conclusion	90
7.2	Future Research	91
7.2.1	Discontinuity Detection	91
7.2.2	Pseudo Spectra Images	93
7.2.3	Horizon Detection	94
	Appendix A Watercraft Design	95
A.1	Propulsion and Steering	97
A.2	Controls	99
A.3	Power Generation and distribution	102
A.4	Conclusion	103
	Vita	105

List of Algorithms

1	Steps to generate PSI in the visible spectrum	40
2	Steps for horizon detection	46
3	Simultaneous detection of scale, discontinuity and threshold in images	74

List of Figures

1	Placement of the research components in the context of application.	x
2.1	Photon Paths prior to Capture by Camera	12
2.2	Photon Absorption by water and heavy water at different wavelengths.	13
2.3	Water constituents effecting back scatter ((Stramski et al.; 2004)).	14
2.4	Visual Realization: Water surface	17
4.1	Inadequacies of horizon assumptions in (YCrCb) color space	33
4.2	Comparison of horizon detection.	34
4.3	Plot of x,y,z weights at different wavelengths [9].	36
4.4	Pseudo Spectra Images.	41
4.5	PSI: Intensity variation at different wavelengths	42
4.6	PSI: Intensity variation at different wavelengths	45
4.7	Steps of Algorithm 2	47
4.8	A continuous horizon from a fractured horizon	48
5.1	OT Characteristics: Absence and Presence of a foreground.	52
5.2	Left to Right: Unit step function, Dirac Delta and PDF of Dirac Delta	52
5.3	PDF, CDF and derivatives (w.r.t. g) of the PDF of a Gaussian Function.	55
5.4	Schema: Otsu's Threshold in a histogram.	60
5.5	OT Schema: Symmetric Unimodal and Unbalanced Bimodal PDFs	62
5.6	Schema: Classes (C_0 and C_1) and (NC, IC) in an unbalanced PDF.	63
5.7	Graph of $\delta(t)$ for various ratios of (ϵ/A) for the GMSS of a Discontinuity.	65
5.8	Schema: Statistical parameters of IC and NC in a Gaussian Function.	66
5.9	Graphs of OT $[k * (t)]$ and $[\delta(t)]$ against Scale at various ratios of ϵ/A	68
6.1	Illustration of Heuristic presented on Synthetic Images.	74

6.2	Algorithm Scope : Background Intensity variation.	75
6.3	Algorithm Scope: Background Frequency variation.	76
6.4	Algorithm Scope: Foreground Frequency variation.	77
6.5	Algorithm Scope: Foreground Intensity variation.	78
6.6	Algorithm Scope: Background Frequency and foreground size variation. . .	79
6.7	Algorithm Scope: Water Scene Analysis.	80
6.8	Illustration : Algorithm comparison.	82
6.9	Experiment: Violation of maximum principle.	84
6.10	Segmentation comparison.	86
6.11	Algorithm in calm conditions.	87
6.12	Algorithm in moderate conditions.	88
6.13	Algorithm in rough conditions.	89
A.1	Remarks of the Head of Department	96
A.2	Steering Comparison (Left) Conventional. (Right) Differential.	97
A.3	Propulsion and Steering Components	98
A.4	Control System: Schematic illustration.	99
A.5	Control system : photographs	100
A.6	Coupling	101
A.7	Locating midpoint of overlap range of rotary switch positions.	102
A.8	Graphic User Interface (GUI) for the software to control the boat.	102
A.9	Screen Capture of code for serial communication with coontrol motors. . . .	103
A.10	Schematic Illustration: Power Distribution.	104

Scene Analysis for Robotic Watercraft

Rahul Walia,
B.Tech (Mechanical),
M.Sc (Computer Integrated Manufacturing)

████████████████████
Monash University, 2012

Supervisor: Prof. Raymond.A. Jarvis

████████████████████
Associate Supervisor: Prof. David Suter

████████████████████

Abstract

The current thesis describes image processing techniques, developed for analysis of visual scenes encountered during robotic navigation on water. The research was characterized by a lack of relevant literature which could provide an initiating platform for approaching vision guided navigation in water. It was an interesting challenge to decide the best discipline for conducting the research in, as the problem (of water robot navigation) could be approached from marine engineering, robotics, oceanography, mathematics, pattern recognition and image processing prospectives. Consequently the literature reviewed in this thesis, extends to eclectic but related (in application) scientific disciplines. A variety of these scientific techniques were researched, refined, attempted and abandoned with varying degree of success. The thesis, documents some of these techniques (successful or otherwise) along with theory and analysis. The parlance of the thesis attempts to adhere to robotics and vision terminology and conventions.

For a robotic watercraft to navigate successfully using computer vision, it should be able to integrate two major components viz. vision and navigation. Computer vision can be viewed as means to achieve the objective (end) of navigation. The uniqueness of the operating environment simultaneously challenges and assists both the components differently. Generally, absence of pathways makes water based navigation easier than its land based counterpart. The scene analysis for the purpose of vision guided water navigation is characterized by following contrasting features:

1. Unreliable Photometry: The dynamics of water and (consequently) the camera makes it difficult to prepare a valid mathematical model or extract relevant features to describe water.
2. Reliable and Sparse Scene Composition: Unlike land scenes, water scenes are less cluttered and are mainly comprised of water, sky, clouds and occasional foreground object.

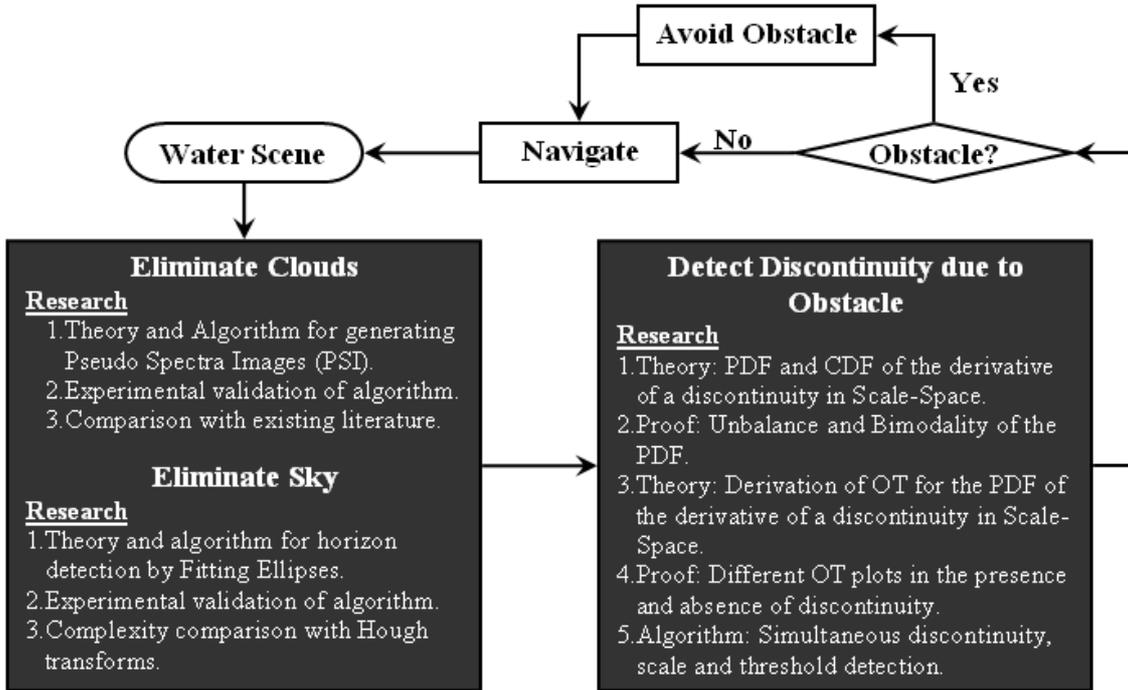


Figure 1: Placement of the research components in the context of application.

Flow chart depicts sequential analysis of various components of a water scene, with a view to locate obstacles during navigation by a water robot. The associated research is highlighted in gray boxes. Abbreviations (PDF: Probability Density Function, CDF: Cumulative Distribution Function and OT: Otsu’s Threshold (Otsu; 1979))

Navigation on water can be achieved by obstacle detection and avoidance. Consequently the phrase *Scene-analysis*, has an objective interpretation of obstacle detection, which is one of the key areas of research in this thesis. Scene analysis described herein is predominantly a sequential attrition of scene components (clouds, sky and water) in a water scene, with a view to detect and locate a foreground object in water. E.g. clouds are not identified as a separate component, but are eliminated in the scene by the use of a mathematical technique which makes clouds transparent in the gray scale-images. The

sequential steps (along with associated problems and solutions) in locating the obstacles in a water-scene are (Figure 1):

1. *Generating homogeneous Sky*: It eliminates false positive identification of clouds as obstacles. Gray scale Pseudo Spectra Images (PSI) were generated from the tri-color images at a fixed wavelengths. It was experimentally established that PSI results in similar response for sky and clouds, thereby preventing the clouds from appearing in gray scale images. In addition, PSIs increase the contrast between sky and water, enabling easy detection of the horizon. The main contribution of the research into PSI, is in developing a mathematical basis for generating images at various discrete wavelengths. The PSIs are envisaged to have applications beyond cloud elimination.
2. *Identifying horizon*: This is done with an objective to define spatial spread of water in the image. Enclosing ellipses are used to identify the horizon from the other edges. The method is better suited on water being:
 - (a) Faster: As compared to conventional Hough Transform.
 - (b) Robust: Able to detect both straight and curved horizons.
 - (c) Simple: Maximizes a mathematical criterion derived from the skew and the zeroth moment of ellipses.
3. *Identifying Foreground (Obstacles)*: The literature review and pilot studies reveal that images or videos captured by cameras are subjected to excessive variability to develop any reliable obstacle recognition algorithm. To achieve robust obstacle detection, the obstacle was identified by its boundary with the water. It was found that the obstacle boundary (with water) edge in a gray scale image has following invariable characteristics:
 - (a) *Spatial Scarcity*: The number of edges in a gray scale image that are created due to boundary of foreground (obstacle) with the background (water) edge is very small as compared to the total edges number of edges in the image.
 - (b) *High Derivative Magnitude*: The magnitude of the derivative of image intensity of the obstacle-boundary edge is higher than that of edges not due to boundary.

By using these invariable characteristics of the boundary of foreground with the background, a theoretical framework of image statistics in Scale-Space is prepared.

This framework can identify the presence or absence of the obstacle boundary / discontinuity, and locate the boundary of the obstacle if it is present. Specifically, the magnitude of the Sobel derivative of a gray scale image is subjected to Scale-Space i.e. convolved with a gaussian kernel of increasing standard deviation. At each scale, the statistical parameter Otsu's Threshold (OT)(Otsu; 1979) is calculated. The plot of Otsu's Threshold against increasing scale enables identification and location of the foreground boundary. Mathematical proofs are provided, that the OT has differing plots in the presence and absence of foreground-boundary (and therefore obstacle). Theoretical research (enumerated in Figure 1) has yielded following results proved via theorems and experimentation:

- (a) Expression for PDF and CDF of the derivative of discontinuity in Scale-Space.
- (b) Bimodality of the PDF.
- (c) Unbalance of the PDF.
- (d) Scale-Life: The duration of scales for which the discontinuity can be statistically identified as a separate mode in the PDF. Scale-Life is a function of the magnitude of the discontinuity and the upper bound of error.
- (e) Analytical expression of the OT for the derivative of a discontinuity in Scale-Space.
- (f) Different plots of OT in the presence and absence of a discontinuity.
- (g) Algorithm for simultaneous detection of discontinuity, threshold and scale appropriate to discontinuity.
- (h) Validation of algorithm on synthetic and natural images.

The results of research into discontinuity can be generalized to a variety of scientific and engineering problems which involve detecting and locating discontinuities.

Scene Analysis for Robotic Watercraft

Declaration

Notice 1

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

Notice 2

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Notice 3

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Rahul Walia
June 2, 2012

Chapter 1

Introduction

The Oxford Dictionary defines a robot as an "apparently human automaton, intelligent and obedient but impersonal machine". However advancements in robotics have rendered this definition obsolete by contradicting some of the adjectives therein. Intuitively, a robot is a machine designed to do human jobs that are tedious, slow, unpleasant or hazardous. Various application areas include but are not limited to Industry (Manufacturing, assembly, welding, painting, machining etc), Remote operations (Undersea, nuclear environment, bomb disposal, outer space, surveillance etc), Service (Hospital helpmates, handicapped assistance, retail, household servants, lawnmowers etc).

Ceaseless advancements in computing technologies and processing speeds have empowered robots with artificial intelligence. This is facilitating a transition to autonomy earmarked by robot's capability of understanding the environment and reacting to it. Intelligent robots are increasingly operating with varying degrees of independence from human intervention to complete tasks with the support of sensory intelligence, planning and mechanism control strategies. Autonomous navigation is a cardinal milestone of artificial intelligence and liberating the robot from human intervention.

The mobility of the robots is affected by a host of factors like the medium of transport (air, water, land, underwater, human fluids, pipes, space etc), terrain (outdoor and indoor), type of locomotion (propellers, legs, wheels, fins, wings, screws etc), shape (anthropoid, insect, boat, animal, snake, vehicle, tool etc), level of autonomy etc. The navigation realized by a robot is an outcome of these interdependent factors. Given these robot characteristics, the foundation blocks of mobile robot navigation are environmental modeling, localization, obstacle-free path planning/following, motion control and communications.

A truly autonomous robot should be able to execute a premeditated task in a hitherto unknown environments, by execution of key components in robotic navigation.

Water based robots usually fall into 2 categories, sub-surface and surface. This thesis focuses on surface robotics. For the realization of autonomous navigation by water based robotic platforms the research reported in this dissertation, uses a visual camera to assist a water robot to understand its environment with an objective of safe navigation. Intuitive perception for a robot's conception is application specific minimization of human intervention. This expectation is more pronounced and justified with the increasing intelligence on a robot. For a water based robot to qualify as an intelligent robot there are three basic pre-requisites:

1. Recognition of water in the scene.
2. Recognition of objects in water.
3. Navigation to or around the object of interest.

Apparently simple, these fundamental necessities in water, translate into complex and widely researched areas of obstacle detection and navigation. The obstacle detection and navigation on water differs from that on land. These differences are analyzed in more depth to enable a holistic grasp of the research and its key areas.

1.1 Navigation comparison : Land versus Water Robots

Water robotics is not a well published research area and comparing it with land robotics would highlight the key areas of research and the accompanying reasons for it. Primary differences in land and water robotics from a navigational perspective arise from:

1. *Constraining Forces:* Whilst land based robots are predominantly working against forces of gravity and friction, these are marginalized for water robots. Operation in a hydrodynamic realm exposes the water entity to forces of buoyancy, surface tension, viscosity, water state (degree of turbulence), strength of water currents, loading of the water-entity, underwater profile etc.
2. *Operational arena:* The land based robots usually need to operate in locomotion-conducive environments or paths as compared to their aqueous counterparts which

can potentially operate anywhere given adequate buoyancy and depth. Broadly this implies simpler navigation.

These macro differences obviate a variety of pre-requisites from land based navigation and simultaneously introduce changes in terms of sensors, propulsion modes, steering, controls etc. This translates into different systems at all levels from conception, to design, operation and maintenance. Under this amended topography, key components for robotic navigation are analyzed in the succeeding subsections.

1.1.1 Environmental Modeling and Localization

Robot navigation, unlike human navigation, relies on its ability to accurately define its position and pose in relation to the environment. This task carried out with seemingly little deliberate effort by humans needs to be executed recursively and deliberately by robots. Environmental modeling is a precursor to the actual process of localization unless special landmarks at known locations have been installed in the environment. The environmental modeling or map building (Thrun et al.; 1998) determines the location of entities of interest (such as: natural landmarks, obstacles) in a global frame of reference (such as a Cartesian coordinate frame). To build an environmental map, a robot must simultaneously know where it is (localization). Methods for information assimilation are broadly classified (Krse; 2000) into two classes:

1. *Model matching methods*: Herein the localization is done with an optimal fit between the sensor measurements and the global model. This method requires extensive prior environment information before a practical model can be constructed.
2. *Appearance based methods*: Which do not build a geometric model of the environment but rely on the model of sensor values (appearances) as a function of the robot position. However such a method requires an extensive data set of the sensor values and the relative robot position. Using these methods various kinds of models (probabilistic, neural networks, radial basis functions, or look up tables) are then constructed.

As discussed earlier watercrafts operate in a different (uncluttered and spatially vast) arena and under different constraining forces. This results in attrition of mapping and localization requirements and usually an elaborate map is not required. Sensor-value

models are adequate to localize the robot. Reduced localization and mapping requirements could be attributed to:

1. *Scarcity of absolute landmarks for localization:* Whenever present the interest in these landmarks is limited to that of an obstacle or rarer still, the target.
2. *Adequacy of geographical / topological maps in tandem with sensors (RADAR / SONAR / GPS) to define the position of robot.*
3. *Simple navigational requirements:* Owing to operational arena, water navigation usually does not require complicated maneuvers like turning 90 or reversing.
4. *Sufficiency of global mapping:* Global mapping is adequate for navigational requirement, eliminating the need for local mapping. E.g. for a land-robot to navigate corridors for surveillance, an absolute position in reference to a detailed floor plan is required, but a water-robot can easily patrol coast lines based on geographical maps and a GPS.
5. *Reliability / accuracy and tolerances of sensors:* Since the scales of distance involved are fairly high in water navigation, the accuracy expected from sensors proportionately reduces (higher tolerances). E.g. for navigation in a room, a laser or ultrasound is required with accuracies in terms of centimeters, but a water robot can function in a fairly reliable manner even with inaccuracy of a few meters. Additionally the outdoor and exposed (to satellites) arena increases the reliability of Global Positioning System (GPS) readings which can be used in isolation without corroboration from other sensors (like odometer).

1.1.2 Path planning

Path planning in mobile robots is in response to the following objectives (Alonzo; 1996):

1. *Obstacle avoidance:* It can take the forms of stopping or steering as implemented in the guiding strategy of the robot's path planning.
2. *Trajectory planning:* Ability to plan a trajectory to goal.
3. *Route planning:* Ability to plan a sequence of trajectories to goal.
4. *Re-plan in the face of new information.*

5. *Replenish consumables.*

Land based path planning literature exists in abundance (Khatib; 1986),(Zhiye et al.; 2004), (Setalaphruk et al.; 2003) ,(Vasudevan and Ganesan; 1994) and (Chang et al.; 2005); however not all are relevant for water. Detailed discussion for contextually relevant (water navigation) is presented in a subsequent section. The transition of some of land based path planning strategies to water is subject to fundamental differences (Constraining forces and operational arena) as discussed above. Exclusivity of some water based strategies is due to the following cardinal differences between land and water based path planning:

1. *Vehicle Dynamics:* The navigational-dynamics of water-robots are subject to less friction and negligible (in comparison to buoyancy) gravity; hence navigation to accuracy levels comparable to land robots is neither possible nor required. E.g. turning circles are wider and stopping may require reversing the propellers and the associated time delays are acceptable.
2. *Flexible Navigation:* Usually the water surface is less cluttered than land, there are no predetermined routes (e.g. roads or corridors), and traveling distances are larger. This allows for flexible path plans. The obstacles are less frequent and not as chaotic as on land, allowing ample time to arrive at an optimum plan and also to dynamically update the path. Classification and motion estimates of obstacles can be done fairly quickly and robustly by a variety of sensors (Chang et al.; 2005).
3. *Simpler and more effective localization and mapping:* Discussed in previous subsection.
4. *Universal set of navigational rules:* Although not proven to be technically optimal or otherwise, these provide a ready made legal and operational framework supporting navigation all over the world for centuries. Given the uniqueness of navigational arena these can be adapted for path planning.

1.1.3 Motion control

Robot mobility is computer controlled and on land it is subject to gravitation, inertial, centrifugal and frictional forces. These forces are sensed or estimated by the computer to generate appropriate signals for the motion. Motion control is inherently dependent on

kinematics and dynamics. There are a host of mathematical models (Kalman Filters, Particle Filters, Bayesian networks, Markov Random Fields etc) for controlling the robot motion. The motion control has also been attempted using fuzzy logic (Zhou and Raju; 1993) and (Mohan and Deb; 2002), hybrids of fuzzy logic (Da Silva et al.; 1998) and Neural networks (Krishnaswamy et al.; 1991). Seemingly endless, these combinatorial possibilities of achieving motion control have to consider kinematic forces, environmental model, physical characteristics of the robot and application. The transition to water based robotics alters all of these factors and most of them have been discussed with respect to environmental modeling and localization and path planning. The properties of water robots (surface robots and not for sub-surface robots) affecting the motion control not discussed so far are summarized below:

1. *Control Forces:* Control forces on a water-robot can include effects of rudders, propulsion units, thrusters, and pods, roll stabilizing fins, T-foils, interceptors and so on. In addition, internal design factors such as meta-centric height, roll damping tanks, buoyancy tanks, free surface effects are important forces which affects the motion control of the water-robots.
2. *Robot Parameters:* These refer to the operational parameters that the robot is endowed with like speed (propeller response), turning ability (rudder response), stabilizing ability, list, underwater hydrodynamic profile, exposed area.
3. *Environmental parameters:* These are design parameters for control system as well as the parameters encountered for the operational arena. A water based motion control system is required to factor in more environmental parameters as compared to its land-based counterpart. The parametric range is also greater in water than that encountered for land motion.

1.2 Vision comparison: Land versus Water Robots

The fundamental differences between land and water vision systems for object recognition arise from the fluid nature of water and uncertainty in terms of occurrence and description of foreground objects. Hence the interaction of scene and camera is subject to:

1. *Dynamic background:* Background dynamism is an outcome of a plethora of underlying natural and man-made phenomenon like gravity (earth and moon), surface viscosity, surface tension, density, atmospheric temperature and pressure, water depth, wind speed, geophysical occurrences (earthquakes, tectonic plate shifts, underwater volcanoes), presence of other water vehicles etc. In addition, variant natural illumination (sunrise, sunset, reflection, angle of observation, shadows, range etc) compounds the inaccuracy of the visual readings. Hence the visual images of water obtained in a natural environment are expected to be stochastic at best and chaotic at worst. This makes it hard to establish a mathematical, probabilistic or statistical model describing the temporal evolution of water. A more comprehensive treatment is given in next chapter.
2. *Dynamic foreground:* The objects in water will demonstrate unique dynamism depending on their physical and dynamic features and their interaction with the water. E.g. different visual dynamisms will be observed from a drowning man, a surfer, a fish, a speed boat, floating debris etc.
3. *Dynamic frame of reference:* Unlike its land based counterpart a water robot exhibits dynamics not only due to its own motion but due to that of the underlying fluid in which it is floating. Hence the videos inadvertently capture roll, pitch and yaw of the platform which is not the case for land robots. Since the visual sensor's platform is subjected to non-deterministic dynamics, any useful model created must allow for or compensate for this anomaly in the resulting video to obtain a reliable estimate of scene components. This estimation is itself hard owing to the factors given in point (1) and consequently model construction is also not easy. The movement of the sensor can in fact be amplified in comparison to the platform e.g. if camera is mounted on a mast the video frames will show much more variation than if it were mounted closer to the water surface.
4. *Limited foreground modeling:* Owing to physical dimensions and simplicity of the environment the model for water is not an elaborate 3 or 2 dimensional model but a sensor model. This implies that prior information about the object is also limited. E.g. a small surveillance robot in water will encounter floating debris, vegetation, garbage, marine life, coastal landmarks and so on. However a land based counterpart

in a corridor would be equipped with parametric details (dimensions / symmetry / edges / geometry of objects etc) about obstacles expected to be encountered.

5. *Scarce object population:* Objects encountered in the contextual environment are far and few. Hence the problem of object recognition withers down to background modeling and subtraction. Unlike land based vision guided systems, water based recognition of an object as a target or obstacle is usually adequate. E.g. it is adequate to recognize a floating object as a human form without any need to classify higher features like gait, facial expressions, face recognition etc. Exceptions to the scarce object population do exist e.g. boat markets or crowded marinas and are excluded from the subject matter of this thesis. Hence object recognition in water as compared to land can :

- (a) Operate effectively at a comparatively lower level of classification and with low level features.
- (b) Deploy different choice of sensors. E.g. classification relaxation can be exploited by the choice of visual sensor. Instead of using visual images a Thermal Imaging Camera (TIC) would prove to be more reliable in water as the thermal signatures between classes of objects varies and can be expressed as absolute (or a unique distribution) parameters.

1.3 Applications

Unlike land based robots the applications for water robots are not that apparent. These plausible applications can be:

1. Life saving by human form by distress gesture identification.
2. Coastal surveillance and mapping.
3. Oil spills / industrial chemical plume detection and mapping.
4. Autonomous patrolling / surveillance for defense purposes.
5. Autonomous navigation for fish protection from predatory birds.
6. Swimming pool /water bodies / sewerage hygiene monitoring and maintenance.

7. Mine sweeping and degaussing.
8. Internal inspection of large tanks (chemical, ballast, oil etc) for both state of tank and the process if it exists.
9. Mobile platforms for underwater inspection and repairs of ships (keel deflection, marine growth) and other underwater structures like oil rigs.

This chapter has highlighted cardinal differences in environmental modeling, localization, path planning and motion control with respect to land robots and water entities. These differences allow water based navigation to have a greater error tolerance. Uncluttered environment allows for inherently optimized paths (straight line) and path planning is more of an obstacle avoidance exercise. However obstacle recognition in water assumes more significance as compared to land given the operational arena, its associated characteristics. The lack of information on foreground object's visual characteristics as well its rare and erratic occurrences transfers the onus of segmentation to background model. Modeling background of water becomes a hard but unavoidable problem given the almost unpredictable visual image that water presents and the unpredictability of its interface with the unstable frame of reference of the sensor. A detailed literature review is conducted in next two chapters to expose the inadequacies of existing image processing research for obstacle avoidance in water robotics. The literature review has been extended over two chapters to analyze water images from the perspectives of

1. Causal phenomenon: These provide water with its unique visual properties like color, intensity, texture and dynamics, which in turn effect both the process of image capture and the content of the image..
2. Image processing.

Chapter 2

Literature Review: Water Surface

The literature reviewed in this chapter belongs predominantly to domains other than image processing. The text presented herein, can be considered ancillary to the main research reported in the thesis, but has been included, as it enhances a holistic understanding of the visual representation of water. This literature review provides an insight into the underlying causal phenomena viz. chemical, spectral, geological, geographical and atmospheric factors which influence the water as viewed by vision and cameras. The other reasons for inclusion of this literature are:

1. *Pragmatic and historic considerations:* The study (Blair; 1965) of oceans, seas and other water bodies far precedes, the advent of computers and therefore image processing. Therefore, it is necessary to investigate the existence of techniques, models and representations in domains other than image processing, which might assist or make redundant the current research. This factor assumed significance because of the failure of an image processing technique applied earlier in the course of research by the author. The waves appeared fairly regular, and therefore an attempt was made to characterize video of water in the frequency domain using Fast Fourier Transforms. No dominant frequency was observed. The reason for absence of a dominant frequency, was provided by a well known (to oceanographers) causal phenomenon discussed in (Blair; 1965) wherein the wavelength and frequency of the water waves are intrinsically (mathematically) linked to the depth of the floor of water body and consequently the frequency will change with depth. It became imperative to understand the causal phenomena underlying the visual representation of water to avoid further setback in the research.

2. *Basis of research* The literature review of causal phenomenon, provides a scientific basis for accommodating rather obvious **variability** of water in research. The variability of water makes any model of water difficult to prepare, unreliable and limited in spatial scope.

Water images are a photo-optical representation in the visible spectrum, of water's composition and natural forces that impact it. These causal phenomenon effect the optical properties of water. From an image processing perspective, back scattering is the optical property that is predominantly captured by the camera and will be studied in some detail here. These factors can exist in infinite configurations which may or may not have interdependencies. An enriching body of literature related to the causal phenomenon, exists in the domain of (still actively researched) oceanography. In the last few decades, satellite imaging has also contributed to the field of hydrologic optics. It is neither feasible nor intended to delve into detailed discussions regarding the causal phenomenon, but the review presented herein highlights the salient features relevant to the current research into visual representation of water.

2.1 Photonics

A camera captures photons within the visible range. These photons originate from solar radiation, travel through the atmosphere and impinge on the water surface where they are either absorbed, scattered, reflected or transmitted. Figure 2.1 adapted from (Robinson and Mitchelson; 1983) illustrates many of these possible paths. From the various paths taken by the photons, it is evident that the water (color and intensity) as captured by the camera comprises primarily of photons radiated by the water surface. Only a small percentage of the light that enters the water (*the downwelling irradiance*) is redirected back toward the surface (*the upwelling radiance*). The upwelling radiance that actually leaves the water surface is termed the water-leaving radiance. The water-leaving radiance is what the cameras acquire as an image. The downwelling irradiance is subjected to selective (wavelength) absorption by the water molecules and dissolved impurities. Therefore the water-leaving radiance is a function of the spectral absorption by water and its constituents particularly chlorophyll present in phytoplankton cells. Even without a detailed quantitative analysis, it is obvious that the image captured by camera

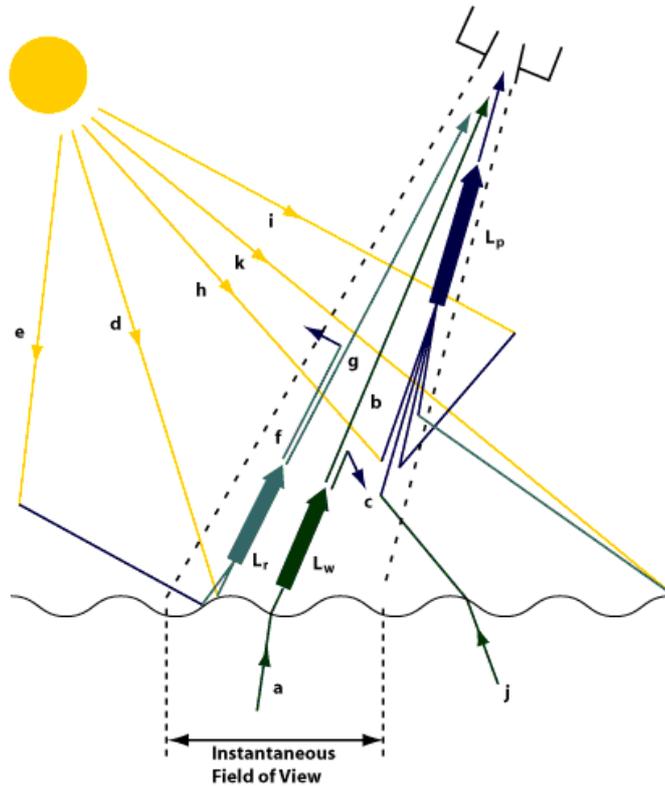


Figure 2.1: Photon Paths prior to Capture by Camera

(a) Water-leaving radiance, (b) Attenuation of the water-leaving radiance, (c) Scattering of the water-leaving radiance out of the sensor's Field of View (FOV), (d) Sun glint (reflection from the water surface), (e) Sky glint (scattered light reflecting from the surface), (f) Scattering of reflected light out of the sensor's FOV, (g) Reflected light is also attenuated towards the sensor, (h) Scattered light from the sun which is directed toward the sensor, (i) Light which has already been scattered by the atmosphere which is then scattered toward the sensor, (j) Water-leaving radiance originating out of the sensor FOV, but scattered toward the sensor, (k) Surface reflection out of the sensor FOV which is then scattered toward the sensor, (L_w) Total water-leaving radiance, (L_r) Radiance above the sea surface due to all surface reflection effects within the IFOV, (L_p) Atmospheric path radiance. This figure is available at (Aeronautics and Administration.; 2010) and has been adapted from (Robinson and Mitchelson; 1983)

will vary with the composition and temperature of water, time of day, location of camera with respect to sun, clarity of skies and particulate absorption characteristics and density.

2.2 Water Color

Adverse to conventional belief, water is not color-less but has an intrinsic blue color. Water's color is due to selective absorption in the red spectrum from the white light. The

chemical composition of water endows it with a high concentration of OH bonds. These OH bonds are very strong and hydrogen atoms are very light. Consequently the molecular stretching vibrations of water are anharmonic and occur at very high energy as compared to molecules without OH bond (Braun and Smirnov; 1993). When these water molecules with their unique vibrations are exposed to photons in the visible spectrum, they absorb light at the red wavelengths to transit to a higher vibration harmonic. To illustrate that the color is due to nuclear composition (Braun and Smirnov; 1993) demonstrated (Figure 2.2) that the color absorption by the isotope of water (Deuterium Oxide) is similar but shifted to higher spectral wavelengths (beyond visible range) making heavy water appear colorless. Since the only difference in water and heavy water is the number of neutrons, it was concluded that the color of water is due to the nuclear forces within the molecule. Water is unique in the sense that it is the only known compound that owes its color to molecular vibrations. In addition to the OH bonds, water molecules also have a weaker

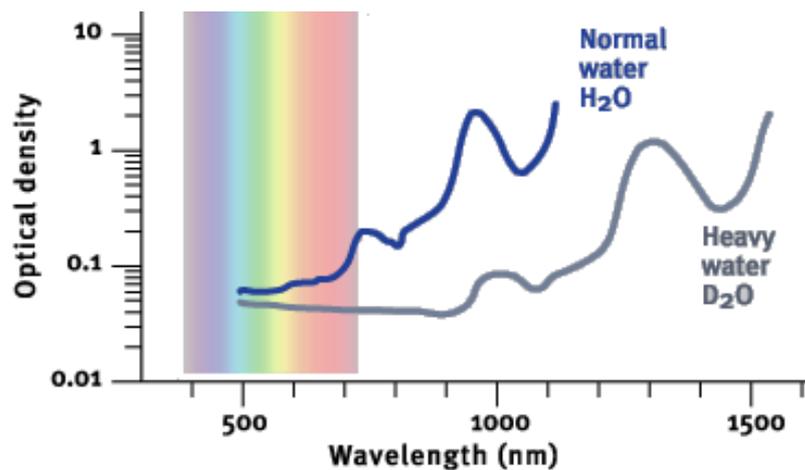


Figure 2.2: Photon Absorption by water and heavy water at different wavelengths.

hydrogen bond. The strength of this bond varies with the temperature of water and consequently the absorption of red wavelength also changes with the temperature. If pure water were available under uniform illumination, it would be possible to construct mathematical models based on photon absorption illustrated in (Figure 2.2). In fact algorithms exist in satellite imaging wherein sophisticated sensors measure the radiance at different frequencies and the ratios of these radiances at different wavelengths provide the concentration of certain phytoplankton or algae in water. However these algorithms are geographically confined and base their premise on a dominant water contaminant. Given the variety of chemical, microbiological, and particulate contamination that the

water can potentially contain, the absorption profile of (Figure 2.2) cannot be relied on for most of the water images. The second problem that arises is that the cameras do not provide wavelength segregated intensities or radiance levels, but rather integrated (or cumulative for discrete cases) intensities with a view to achieve a tri-stimulus response similar to those of human retinal neurons. In the absence of frequency specific data from the camera, algorithms based on radiance at different frequencies are impossible to conceive and construct.

2.3 Hydrologic Optics

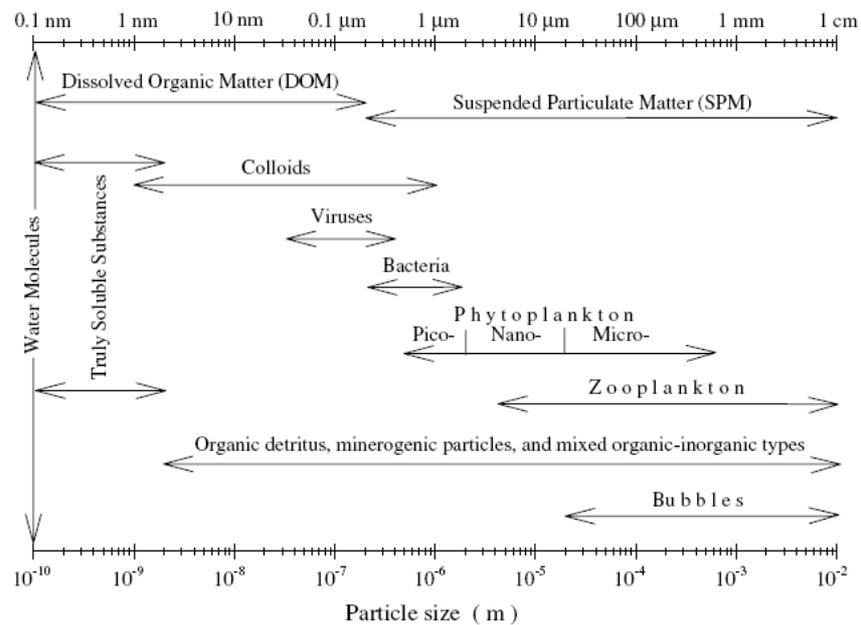


Figure 2.3: Water constituents effecting back scatter ((Stramski et al.; 2004)).

Water is an optical medium whose color and intensity is determined by the ambient light incident on the water surface and two mutually exclusive optical properties viz. inherent and apparent. The *inherent optical properties (IOPs)* are those properties that depend only on the medium and are independent of the light field within the optical medium. Examples of IOPs are absorption coefficient, the volume scattering function, the index of refraction, the beam attenuation coefficient and the single-scattering albedo. In contrast the *apparent optical properties (AOPs)* of the water depend on both the IOPs and the geometric (directional) structure of the ambient light field, and that display enough regular features and stability to be useful descriptors of the water body. Commonly used

AOP's are the irradiance reflectance, the average cosines, and the various diffuse attenuation coefficients. By using the IOPs and AOPs, the backward scattering can be established (Stramski et al.; 2004) and (Mobley; 1994). The **backward scattering** determines the color of water that is measured by optical devices (camera in current research). The back scattering in turn depends on the types and concentrations of various constituents present in the water. The role of sea water constituents has been extensively (more than 180 references) reviewed from contemporary research in (Stramski et al.; 2004). The various water constituents effecting the light back scatter are identified in Figure 2.3 taken from (Stramski et al.; 2004). The observations from the review (Stramski et al.; 2004) in respect of backscattering of light, that are relevant from the perspective of this thesis are:

1. *Theoretical modeling:* of backscattering due to marine constituents is impossible because complete knowledge of the distributions of particle size, shape, and refractive index (including internal inhomogeneities) is never available.
2. *Simplifying assumptions:* are often contradicted.
3. *Direct measurements:* of back scattering over an entire optical spectral range have not been achieved till today. Insufficient data which does exist, is fragmented because it is limited to a specific wavelength or confined to a geographical location, which in turn restricts generalization of the experimental observations.
4. *Geographical variability:* There is consensus amongst researchers on the variability of light backscattering due to geographical locations.
5. *Optical instruments:* for measuring the back scattering have stringent design and calibration requirements and in-situ measurements often show variability, which must be factored in and corrected for during experiments. Some of the issues in accurate measurements identified by (Stramski et al.; 2004) are radiometric calibration, angle and scattering volume calibrations, sensor-response function and optical geometry (involving the scattering volume, illumination beam, detection of scattered light, and path length in the water), proper angular resolution, temperature and pressure effects, as well as optical and mechanical imperfections of the instrument.
6. *Salinity:* Sea water and pure water have different back scattering coefficients at the same temperature and pressure.

7. *Water molecules:* The back scattering due to water molecules expressed as a percentage of total back scattering can vary from ($< 1\%$) to ($> 80\%$) in the blue spectrums.
8. *Bubbles:* can constitute up to 10% of back scatter and the presence of bubbles in a water patch can make it appear greener or yellower as compared to water patch without the bubbles.
9. *Organic and Inorganic particles:* Organic particles include two major categories of particulate matter: living plankton microorganisms (viruses, bacteria, and various phytoplankton and zooplankton species) and non-living organic detritus (breakdown products of microorganisms). Inorganic particles include the various mineral species (clay minerals, quartz, feldspars, calcite and many others). Tens of thousands of species of marine plankton and a few thousand mineral species are known, which include a few dozen of the minerals that are essential constituents of rocks. Many marine particles would probably be best described as a mixture of organic and inorganic types. Conflicting researches have been reported on the effect of organic particles on backscattering due to different species of organic organism constituting the bulk of marine particles. Similarly scattering due to non-organic minerals depends on the concentration and the type of mineral present.
10. *Colloids:* Colloids lie at the boundary between truly soluble chemical species and sinking particles, and are generally defined as tiny particles, macromolecules, and molecular assemblies in the size range between about 1 nm and 1 μm . These occur in high density in water and are responsible for a significant amount of back scattering owing to their shape. Colloids tend to absorb strongly in the blue and ultra-violet portion of the spectrum, giving the water a greenish or yellowish color.

2.4 Water Dynamics

Water images are subjected not only to the composition and consequent hydrologic optics, but also to the dynamics exhibited by water. Blair provides a good introduction to the waves in (Blair; 1965). Based on (Blair; 1965), Figure 2.4 has been prepared to summarize some of the natural phenomenon acting on water. These phenomena are active

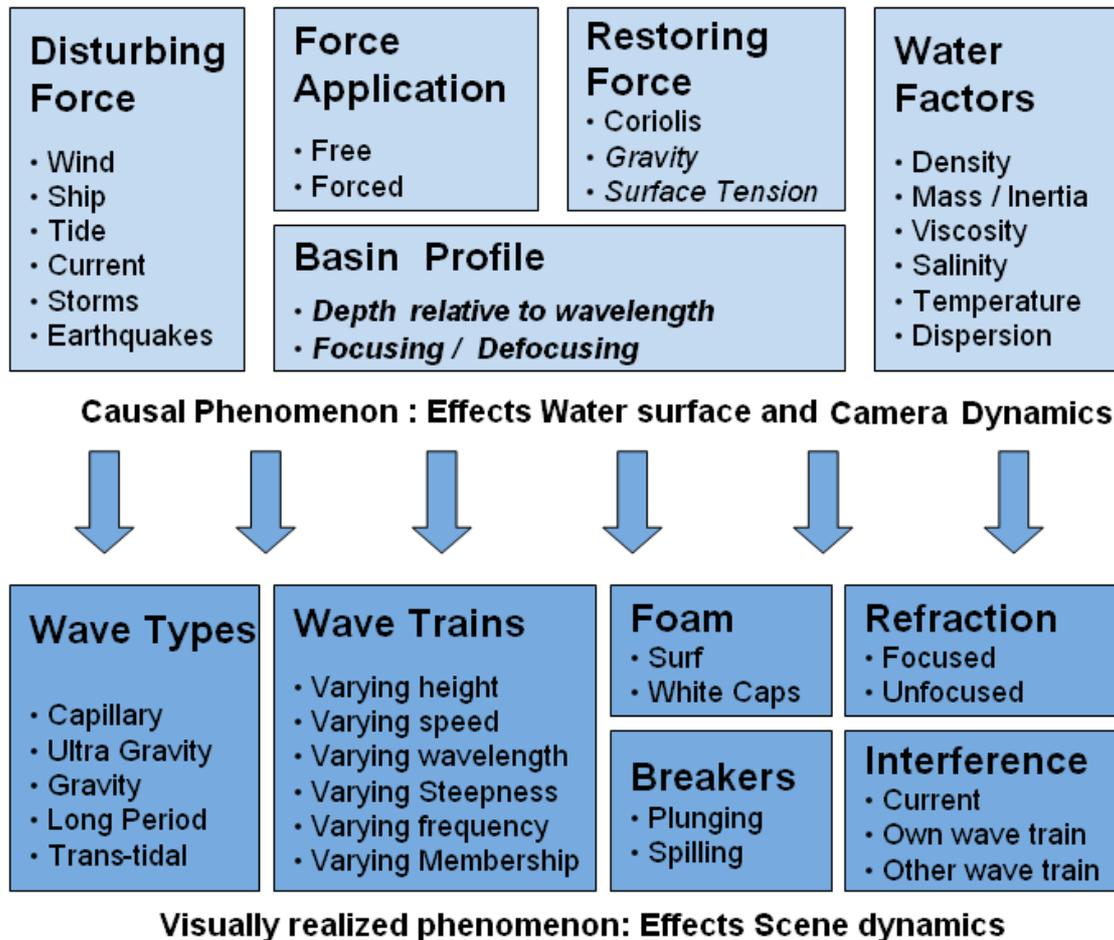


Figure 2.4: Visual Realization: Water surface

research topics in oceanography, however they also influence the water images. From the perspective of image acquisition and analysis, there are following areas of interest:

1. Various natural phenomenon contribute to the form of water surface as perceived by humans or camera.
2. Mathematical formulations are possible to associate individual causes with individual effect in oceanography. However the image captured by cameras rarely reflect these individual associations. Images are the visual realization of multile factors shown in Figure 2.4. For example, by measuring wind speeds it is possible to reliably predict the type of wave that will be formed, its approximate height, duration of existence, distance traveled and mode of dissipation. However these predictions may not be realized due to the bathymetry, tidal, surface tension and interference considerations.

3. The dynamics of water are transferred to the camera as it is mounted on the watercraft. The unpredictable movement of the camera does not allow reliable modeling of the water surface.

2.5 Inferences

The literature review in this chapter has been from diverse and still active research studies. The main inferences that can be drawn from the literature review that are relevant for this thesis are:

1. *Variability*: Composition of water is variable that effects its optical properties and consequently the images. It is impossible to determine the exact composition and consequently the effects that composites will have on the images.
2. *Instrumental inadequacy*: Instrumental inadequacies flow from the previous point. Even assuming that composition of water was an insignificant issue, the sophisticated engineering and elaborate calibration make it very difficult to record the in-situ optical properties of water. Researchers often make informed decisions on choice of the wavelengths and the optical properties fine tuned to the specific purpose like measuring density of phytoplankton or sedimentation. Camera operating in visible spectrum is a rather blunt instrument to deploy in water description. The handicap of the camera arises from its inability to
 - (a) Measure beyond the visible spectrum.
 - (b) Segregate visible spectrum to desired wavelengths.
 - (c) Measure any optical property other than cumulative intensity of Red, Green and Blue channels.
 - (d) Measure any parameter below the surface of water to obtain differential readings.
 - (e) Measure any oceanographic parameter like wind speed, current, wave height, white caps etc.
3. *Modeling Inadequacy*: A model can be based on a theoretical mathematical expressions or be data driven. Both these modes of modeling are inadequate to form a

universal model for representing water. Since the constitution and environmental factors of water are variable, optical properties and their characteristic curves are also variable. Thus models can be prepared only on the basis of assumptions and for a specific purpose. For example (Stramski et al.; 2004) back scattering differs by 67 orders of magnitude at visible wavelengths depending on the size of organic particles and in the absence of prior knowledge on the size of organic particles any model based on back scattering values will be inaccurate. The modeling inadequacy spills over to the realm of oceanography as well, where the variation, unpredictability and interdependence of causative phenomenon constitute a prohibitive barrier to formation of a reliable model.

This chapter has highlighted that it is virtually impossible to model water because of the natural dynamism and variability that it possesses. The problem is compounded by the limitations of the primary sensor (camera). In addition to the design limitations, camera is mounted on the watercraft which has its own hydrodynamic profile, which makes observation of any temporal characteristics of water subjective to camera movement. This adds another layer of complexity to the problem because camera movements have to be modeled separately and compensated for in the water model.

Chapter 3

Literature Review: Image Processing

Previous chapter shows that modeling water on the basis of hydrologic optics or oceanography is extremely difficult. In this chapter the relevant literature from the image segmenting domain will be reviewed with an objective of establishing the state of art for recognizing objects in water. Water images can be analyzed using a plethora of diverse image/video processing techniques.

3.1 Background Segmentation

Identifying objects in water can be perceived as, subtracting the background (water) from the image. Hence background definition or model assumes significance for any effective object recognition. The water robot's environment is not easy to model as it exhibits seemingly inconsistent spatial variations over time. Before delving into detailed technical literature it is imperative to mention that a comprehensive literature addressing background segmentation from water based navigational standpoint does not exist. The required segmentation can be attempted in spatial, temporal or spatio-temporal dimensions. The temporal dimension will be analyzed first with the motivation of being able to extract some high level features like foreground object's visual dynamics for the purposes of object's classification after segmentation

3.1.1 Temporal Segmentation

The following analysis draws from the approaches suggested in surveillance systems, image synthesis, animation, oceanography and dynamic textures. Dynamic textures (Doretto et al.; 2003) have been defined as a sequence of images of moving scenes that exhibit certain stationary properties in time. More concisely (Yuan and yeung Shum; 2004) defines the dynamic texture as temporally continuous and infinitely varying stream of images that exhibit certain temporal statistics.

Dynamic texture synthesis studies can be broadly classified as parametric or non-parametric. Non-parametric techniques include generating dynamic textures by sampling pixels (Wei and Levoy; 2000), frames [23], and wavelet-structures (Bar-Joseph et al.; 2001). These techniques focus on ensuring continuity of the video sequence while simultaneously trying to emulate the original video. Whilst efficient for synthesis of dynamic textures, the non parametric approaches do not provide a model generalization. The absence of parameters inhibits a machine based analysis for the purposes of segregation into foreground and / or background. Dynamic textures have been modeled using various parametric approaches. Spatio-Temporal Autoregressive model (STAR) has been proposed to model dynamic systems (Szummer and Picard; 1996). It is a 3 dimensional extension of an autoregressive model which expresses each pixel as a linear combination of the surrounding pixels lagged both in time and space. A sketch model is proposed (Wang and Zhu; 2003) by replacing the dictionary of Gabor and Fourier bases with sketches (symbolic tokens), thus changing the photometric model to a sketch model. However such a model is limited in applicability to synthesis, and does not lend itself to segmentation as the parameters are descriptive and not mathematical.

A parametric model has been proposed by Doretto (Doretto et al.; 2003) which is a noise driven Linear Dynamic System (LDS) representation. This is the most widely acceptable work and subsequent variations have emerged. Doretto et al postulate that Euclidean reconstruction of images is not possible in any machine system without prior knowledge owing to the ill posed problem of fixing either the photometry or the geometry and assuming the other. This model is designed for maximum-likelihood or minimum error variance. Doretto et al define the dynamic texture as a realization of the output of a dynamical system driven by an independent and identically distributed (IID) process. Mathematically, the linear dynamic texture is associated with an Auto-regressive moving

average process (ARMA) with an unknown input distribution. This mathematical definition allows for a choice of filters which can be Principal components, a wavelet filter bank, Fourier descriptors etc. This provides an inherent flexibility to the model.

A variant of Doretto (Doretto et al.; 2003) postulates a Markov assumption for state representation (Zhong and Sclaroff; 2003). This state inference is then converted to standard Kalman filter solution by assuming a Gaussian distribution of the observation density and state transition density. Although this method manages to exploit the inter-pixel relation it is limited in temporal representation owing to Markov assumption. Also modeling each pixel by a kalman filter has computational overheads and the computational complexity of the algorithm is $O(m^2n + n^3)$ where m is the dimension of the observation vector (image size) and n is the dimension of the ARMA model parameters. Another variant of model from Doretto et al analyses the model from a control-theory viewpoint (Yuan and yeung Shum; 2004) and argue that the noise driven LDS model is essentially an open loop control system that is contaminated by noise. The stability of the open loop LDS and the problems of pole placement and model fitting error prevent the model (Doretto et al.; 2003) from generating satisfactory dynamic textures. A feedback controller is proposed and the model is amended to a closed loop LDS which has reduced model fitting error.

Another application of model by Doretto has proved the concept of finite -horizon model reduction for a class of neutrally stable discrete systems (Sznaier et al.; 2004). The algorithm by Mario Sznaier et al (Sznaier et al.; 2004) used a deterministic model to treat the problem of static texture synthesis as that of identifying a state space realization from the impulse response data under the assumption that the image dynamics (in spatial domain) are neutrally stable. Thus they were able to extract textons (texture descriptors) for the synthesis of static textures. The same approach (Sznaier et al.; 2004) was modified by Abraham et al (Abraham et al.; 2005) to synthesize the dynamic textures. However the Fourier descriptors were used to describe the evolution of curves as applied to the ARMA model (Doretto et al.; 2003). The contribution of this study was a migration to extract and utilize n -most significant Fourier frequencies using model reduction techniques (Sznaier et al.; 2004) instead of n -first frequencies. ARMA model proposed by Doretto (Doretto et al.; 2003) has been validated by subsequent studies (Yuan and yeung Shum; 2004), (Zhong and Sclaroff; 2003), (Sznaier et al.; 2004) and (Abraham et al.; 2005) in

that it provides a general model flexible enough to accommodate system specific variations. Assessment from the current perspective highlights handicaps due to :

1. Theoretical assumptions of an IID process having an ARMA realization. The assumption of stationarity is violated due to the non rigid motion of the camera in trying to capture a non rigid texture; hence the appearance model is no longer time invariant (Vidal and Ravichandran; 2005).
2. Practical limitations due to lack of any study using the ARMA model which involves three dimensional dynamics as required in the current research. In all the works discussed the camera was mounted on fixed platform.

A sub optimal solution has been proposed by Vidal (Vidal and Ravichandran; 2005) in which the moving dynamic texture has been modeled using a time varying LDS model with a 2-D translational model to allow for rigid camera motion. The sub optimality arises from their (Vidal and Ravichandran; 2005) assumption of a constant parameter matrix when observed for a predefined time window, however for the current research the time window is difficult to define. Also the camera motion is represented as a rigid 2- dimensional translational motion whereas for the contextual application the camera motion is non-rigid and 3-Dimensional. Further the dynamic texture which validated the study comprised of moving flowers for which the optical flow had to be computed. The flowers texture comprises of well defined features in the form of consistent edges which do not exist in the case of water. Hence both the application as well as the parameter (optical flow) to be extracted also excludes the approach (Vidal and Ravichandran; 2005) as a plausible candidate.

Frequency domain based work emanates from (Spencer and Shah; 2004) wherein water has been studied separately and not under the unifying umbrella of dynamic textures. Low level features of the video such as spatial and temporal frequency spectra are used in conjunction with the physical properties of water like wavelength and amplitude to develop a higher model of environment. This has applications like estimating the wave height or even the object size in water from the pixel scales observed in videos. This approach appeared promising and the feasibility study was conducted to investigate the possibility of modeling water in frequency domain. The approach's candidature was rejected due to following inadequacies:

1. The highest frequencies are due to capillary forces and do not represent the gravity waves and the model must eliminate them.
2. Forecasting is based on the assumption of stationary ergodic process which is violated by the fluid camera motion.
3. The Discrete Fourier Transform is limited in capturing the wavelengths less than the image size or greater than 2 pixels only.
4. Geographical scope is limited to that of deep water ($\text{depth} \gg \text{wavelength}/2$); as in shallow water ($\text{depth} \approx \text{wavelength}/20$) and in transition depths the speed of the waves is a function of the depth. Hence the wavelengths estimated are inherently suspect in the absence of depth and wavelength information.

Takahashi et al (Takahashi et al.; 2005) provide a basic framework for segmenting objects in water. The application is similar to the current research being undertaken in that it attempts to offer a more pragmatic solution rather than mathematical treatment with simplifying assumptions. (Takahashi et al.; 2005) lists the problems associated with false positives in intrusion detection in a bay as repetitive waves, water splashes, light reflection, surf formation, and image reflection in water. The study seeks a set of complimentary solutions to these problems wherein the problem of reflected image is solved by hardware (polarizing filter); repetitive waves are modeled by a quasi spatio-temporal model which utilizes the scatter diagram of image blocks to create a reference image (Flapping Reference). This reference image is updated for each frame and the deviation from this reference image is used to detect intrusion. Since this approach invariably misses slower intruders, temporal averaging is used for detection wherein a differential operator is applied to the averaged image which cancels the waves and only the intruder edges are left. A third approach comprising of standard deviation calculation of spatial windows along with the presence of low brightness to classify the scene as that of being a back-light. Then the areas with higher darkness and movement in temporal domain are thresholded out as intruders. This (Takahashi et al.; 2005) approach is different from the current problem due to following considerations by

1. Unlike the current research (Takahashi et al.; 2005) have cameras mounted on a rigid platform

2. The range of the scene is much higher than would be expected on a robotic platform. This implies that the texture coarseness expected in the current research is expected to be much higher than in (Takahashi et al.; 2005).
3. Supervised scene classification system with elaborate feedback mechanism is required by (Takahashi et al.; 2005) before segmentation can be attempted. This is not a viable option for current research

Nevertheless one feature of the research by Takahashi et al (Takahashi et al.; 2005) which was found useful was the application of a polarizer filter. This filter helped in eliminating some glare from the water surface. The polarizer filter did not completely eliminate glare because of the variation in reflection angles due to spatial spread of the water surface and the uneven surface of water.

Given the current arena of operation, the models operating in the temporal domain clearly do not accommodate visual dynamism exhibited by non rigid frame of reference for the camera. A practical difficulty of estimating the motion of the camera renders void any subsequent creation of a compensation mechanism. This difficulty eliminates most of the temporal methodologies considered.

3.1.2 Spatial Segmentation

Analogous to temporal domain, spatial domain also does not offer a literary body like fingerprint analysis, face recognition or bio-medical imagery, dedicated exclusively to water. Hence no concrete information exists to describe water in terms of color, spatial frequencies, statistics, orientation, physical characteristics, texture, shape or geometry. Intuitive perception about water exists temporally and these have been sidelined to allow for camera motion. In light of this void, the literature survey will borrow on relevant researches emanating from general texture and color based image segmentation (Mojsilovic et al.; 2000),(Haralick; 1979),(Aksoy and Haralick; 1998) and (Ojala et al.; 2002), oceanography, and satellite imagery. For want of a better descriptor, water will be treated as a color texture in spatial domain and studies attempting to segment textures will be surveyed. The problem is hard as there is no universally acceptable definition of texture, rather only an essential property: texture is translation invariant and leaves the same impression on the visual system whichever part of the texture is observed.

Unification of previous work has been attempted by Harlick (Haralick; 1979). Harlick's (Haralick; 1979) summary includes statistical approaches, optical transforms, textural edges, structural elements, gray tone cooccurrence, run lengths and autoregressive models. Discussing all would be beyond the current scope; however the most relevant contribution appears to be the co-occurrence matrix. In the cooccurrence matrix the relationship between the various gray levels is expressed in polar coordinates i.e. relative frequencies are calculated for transition from gray level $G(a)$ to $G(b)$ in direction X calculate over a distance d . Having obtained these frequencies higher level texture descriptors are obtained like homogeneity, contrast, entropy, correlation, complexity, run length and nature of gray level transitions. These perceptual parameters can potentially be used in the segmentation of water. The volume of the difference space equals $(2G-1)d$, where $d=2,4,8$, corresponds to the distance and G is the gray levels. A straightforward description of the difference space with a d dimensional histogram, would generate, very large histograms that are computationally expensive and suspect to statistical unreliability (Ojala et al.; 2002). Contextual experiments need to be conducted to identify the optimum parameter / parameter vector for segmenting the image although the computational memory also remains a concern.

The cooccurrence matrices have been successfully demonstrated (Aksoy and Haralick; 1998) and (Celebi and Alpkocak; 2000) for content based image retrieval and automated inspection (Kyllönen and Pietikäinen; 2000) of wood. However cooccurrence matrices lead to exponential increase in the number of matrices. Ojala et al (Ojala et al.; 2002) introduced Local Binary Patterns (LBP) wherein the texture is defined in the local circular neighborhood by the difference of the centre pixel with the neighboring pixels. The texture is then assumed independent of the center pixel value and the sign of difference is used to describe the texture giving a 2^p LBP number where p is the number of neighborhood. LBP is by definition, invariant to monotonic transformation of the gray scale and rotation invariance is achieved by the use of minimization function applied to LBP operator. This ensures that LBP will always have the minimum value irrespective of the order in which the neighbors are accessed. Using this LBP as primitive a histogram is formulated from the accumulations of the labels over a texture sample. Variance (VAR) is introduced as a measure of contrast where required. LBP (Ojala et al.; 2002) were tested on a set of strong man-made patterns like canvas and carpet which had strong underlying texture

pattern, whereas the water texture does not exhibit equally consistent patterns. The definition and derivation of the LBP do not accommodate pattern's temporal evolution, as the underlying LBP would change. To conclude the method works well for illumination and rotational invariance however it is too rigid to classify an evolving texture owing to inherent definitions and lack of relevant training samples.

Ojala et al (Ojala et al.; 2002) combine the LBP and Harlick's cooccurrence matrix (Haralick; 1979) to generate signed gray level differences instead of absolute difference in calculation of the cooccurrence. The new textural description is more compact and efficient, impervious to luminance change and provides more information than the original cooccurrence calculations. The k-dimensional (corresponding to k gray levels) difference space is reduced by partitioning the k-dimensional difference space using vector quantization. A code book is generated using LVQ algorithm by using training samples. In the current context elaborate training algorithms are undesirable owing to the need to process video frames. Also the water texture is evolving hence continual training and consequential computational overheads are expected in addition to the actual segmentation itself.

Although (Kyllönen and Pietikäinen; 2000) uses color along with texture, this application is similar to the research being undertaken. This methodology utilizes signed gray level differences (Ojala et al.; 2002) to compare the segmentation with the original cooccurrence approach and the results indicated that the incumbent intensity loss from the gray level differences does not augur well for the segmentation. However visual inspection is undertaken in optimum illumination, which is not so in natural scenes and usually will add to chromatic inconsistency in natural scenes. Hence the approach cannot be directly adapted to water based segmentation.

Devaux et al (Devaux et al.; 2001) present an interesting approach to segmenting color aerial images by utilization of Karhunen-Loeve transform (KLT) applied to region based segmentation. The work is appealing in terms of representation of the image color space by its Eigen vectors. Application of KLT results in transformation of the energy to its Eigen values with the property that the greatest possible differences in variances are distributed along these Eigen values or axis. Devaux (Devaux et al.; 2001) also go on to prove that that energy along the first 2 axis is greater than 98.5 % hence a 2-dimensional KLT space is adequate for representing the 3 dimensional color space. The image is segmented separately along these axis and results merged. However this produces over segmentation,

which is a concern because the aerial images have far more color consistency than the water owing to smoothing effect resulting from the large distance between the camera and the scene. Hence the employment of KLT for removal of water background must dedicate preprocessing for color smoothing to avoid over segmenting.

3.1.3 Conclusion

This section has given a relevant glimpse and a quick summary should reveal that various methodologies in the temporal domain are constrained to model non linear and non rigid camera motion thereby corrupting any subsequent model to capture water's dynamic texture. In the spatial domain various studies exist but do not give a convincing and universally acceptable definition to texture in general and water in particular. Structural, statistical and transformation descriptions discussed above are successful but limited to the application domain for which conceived. The invariant parameter/water-descriptor sought is difficult to find given the variable physical, chemical, environmental and optical construct of water. The implications of individual variability (of camera, obstacle and the water) are compounded in images because all the components involved in the image capture have individual dynamics.

3.2 Water-Based Navigation

Sea navigation has been in existence for many millenia. Given the generations that it has spanned, navigation and its in-attendance tools have evolved unceasingly and interdependently. Chronological evolution can be categorized as coastal navigation which was guided by coastlines; chart based navigation with the advent of navigational charts; astro-navigation which is guided by celestial bodies; compass navigation which is guided by magnetic compass, to modern navigation which is assisted by sophisticated localizing, charting, stabilizing and navigational tools. Although the tools have evolved considerably in the last century the rules for navigation have changed little. Even today the point to point navigation practiced for centuries is still the optimum path. The succeeding paragraphs will explore some of the navigation methods for water based platforms.

There is a plethora of navigation methods for robot navigation on land and some of the algorithms deployed for water robots are extensions. A few path planning approaches for

Autonomous Underwater Vehicles (AUV) are summarized in (Vasudevan and Ganesan; 1994). These include APF (Warren; 1990), the path planner developed by Carroll et al (Carroll et al.; 1992) that employs information regarding bathymetry, exclusion zones, obstacles, and ocean current stored in separate databases. A quadtree organization is used to represent depth and obstacle information. The inputs are start task point, a goal task point, a set of constraints such as minimum and maximum depth, different desired speeds, and fuel resources. The planner attempts to generate a minimum cost collision free path using A* search algorithm.

Case based reasoning CBR had been used by (Vasudevan and Ganesan; 1994) wherein a hierarchical database was maintained. The data base consisted of routes and situations stored and assessed by indices. These indices are derived based on several considerations such as the salient features of the case, the type of problem solving, the utility and uniqueness of the case, failures encountered, and other context-specific information. When a new problem is given, relevant indices are extracted from its features by applying a set of indexing rules. The approach adapts to synthesize new routes when historic equivalents are not located. However such an approach does not give a unified solution rather relies on past plans. It has other constraints in terms of memory requirement, feature definitions, path retrieval, fuzzification of indices, and new path synthesis (depends on old path segments).

Jarvis (Jarvis; 1984) has advocated the use Distance Transforms (DT) for path planning and navigation. It is a simple algorithm which generates the path based on the cost involved. This cost can be varied depending on the requirement of the navigational goal. E.g. increasing the cost of navigating close to the obstacle will result in a path through the close to center of two obstacles. It is a fast algorithm which propagates in a raster scan and can be easily extended to multiple dimensions (not necessarily Euclidean). The speed advantage is built into the construction of the DT itself as the weights or costs can be included at the time of computing the DT. Although a general approach has been presented (Jarvis; 1984), *prima facie* DT appears suitable for application in marine environment due to its ability to include angular data from a line of sight perspective for dynamic acquisition of target / obstacle position. Since marine environment is a fairly simple environment and exact localization information (of starting point and obstacle) is available, the DT could in principle exploit this minimal but adequate information to chalk out the optimal path for dynamic obstacle avoidance.

3.3 Impact of literature review on research pathway

The literature review in this and previous chapter, emphasizes the following **push factors** for directing the course of research:

1. *Inadequacy of related disciplines to model water:* Studies of hydrologic optics and oceanography are inadequate to provide a suitable model, parameter or property for detection of objects in water by camera.
2. *Difficulty in modeling Background:* The inherent variability of water also inhibits various algorithms in computer vision to reliably detect obstacles in water.
3. *Futility in modeling Foreground:* Sparse presence of different foreground objects, makes it futile to model foreground for the purposes of recognizing foreground objects.
4. *Inadequacy of Camera:* Camera has inherent design and operational limitations that are amplified owing to non-stationary platform.

Due to the above mentioned push-factors, it was decided to use the interface of foreground and background as the feature for locating an obstacle in water. Usage of the interface boundary as the feature avoids all the problems related with modeling either a foreground or background. It was decided to utilize the edges which represent the boundary of water as the feature that would demarcate the water in an image from other artifacts. To illustrate the ideology behind this approach, the mechanism of image capture illustrated in Figure 2.1 is used. An image is the result of the real world (3-dimensional) projected onto a 2-dimensional plane of the image. Consequent to this projection, spatial edges exist in 2-dimensional images, which pertain to perceived spatial-boundaries of the real world. The word '*perceived*' is important, because while some edges (in an image) are due to physical contact between two real world entities (e.g. plate on table), other edges may be created due to factors like projection of 3D geometry (e.g. horizon), illumination (e.g. shadows), surface of entity (e.g. texture or patterns) etc. In the context of obstacle recognition in water, the entities expected in water images are obstacle(s), sky and water. The desolate and consequently uncluttered water scape has limited types of edges in images. The different kinds of edges as they relate to real world entities can be due to:

1. Textured water surface.

2. Sky-Water boundary edge (horizon) due to 3 dimensional projection on 2D image plane
3. Artifacts in the sky (clouds).
4. Water-Obstacle boundary edge.
5. Water-Shadows (of obstacles) boundary edge.

A research to segregate the above edges from water scenes, should hypothetically achieve obstacle detection. For the purpose of obstacle detection in water robot navigation, water-shadow edges can be considered a subset of water-obstacle boundary edges as:

- Shadows will not exist without the obstacle in the water.
- Spatial projection of shadow will be in physical vicinity of obstacle, and avoiding obstacle shadows would at best cause navigation to err on the side of caution.

Chapter 4

Sky and water: Pseudo Spectra

Images

Based on the limited classes of entities and consequently edges present in the water images, it is possible to subdivide the obstacle recognition into Water-sky segregation and obstacle-water segregation. The former will be solved in this chapter as a more general problem of horizon detection and the latter in the next chapter. This thesis makes following three contributions to horizon detection which have also been reported in a conference paper (Walia and Jarvis; 2010):

1. A theoretical framework for generating pseudo spectra images (PSI), from spectrum analysis of XYZ color-space is presented.
2. Wavelengths in the visible spectrum are identified, at which the PSI has similar intensities for sky and clouds. Generating PSI at these wavelengths minimizes artifacts due to clouds in the sky, resulting in well defined horizon.
3. Fitting ellipses are presented as an alternate to Hough Transform for horizon detection. Ellipses have lower computational complexity than Hough Transform and can accommodate curved edges as candidates for horizon.

4.1 Background

Conventional research (Zafarifar et al.; 2008; Zafarifar and Peter; 2006; Cornall and Egan; 2004) in horizon detection comprises of heuristic and/or probabilistic assumptions about

horizon and the sky. Work by Zafarifar et al (Zafarifar et al.; 2008) and (Zafarifar and Peter; 2006) assumes that:

1. Luminance increases and chrominance decreases from zenith to horizon.
2. Rate of change of chrominance and luminance is low.
3. Sky is earmarked by lack of texture.

It is easy to show Figure 4.1 that the presence of clouds in sky can contradict all the above assumptions, making detection of horizon difficult and thereby warranting an explicit mechanism for compensating the presence of clouds.



Figure 4.1: Inadequacies of horizon assumptions in (YCrCb) color space

Clockwise from Top Left: Horizon with clouds, Luminance, Chrominance 1 and Chrominance 2. Images highlight the absence of luminance and chrominance vertical trends i.e. from zenith to horizon, abrupt luminance and chrominance changes and strongly textured areas.

Other approaches (Thakoor et al.; 2002), (Chahl et al.; 2003) inspired from physiology of dragon fly's Ocelli (small eyes on the forward and lateral region on the head) deploy opponent color processing in Ultra Violet (UV) and green spectra to remove artifacts in sky. UV and green channels have similar response to artifacts in the sky part of the images; therefore integrating the output from these two channels with suitable weights removes variations in the sky due to clouds and sun. However processing beyond the

visible spectrum (UV) requires expensive and dedicated hardware. Method proposed by Cornall and Egan (Cornall and Egan; 2004) occasionally and unintentionally alleviates the presence of clouds by calculating the intensity image exclusively from the blue component of the RGB image 4.2, but does not locate the horizon accurately (Zafarifar and Peter; 2006).



Figure 4.2: Comparison of horizon detection.

Left: Intensity image from Cornall and Egan (Cornall and Egan; 2004). Right: Pseudo Spectra Image (PSI) at 415 nm.

In the next section, a method (inspired from image rendering) to address the presence of clouds in images will be presented. Glassner’s method (Glassner; 1989) for image rendering, which derives the intensities at different wavelengths of the spectrum from the Red Green Blue (RGB) triplet, is modified to derive and define Pseudo Spectra Images (PSI) in the visible spectrum and identify the wavelength at which the response of sky and clouds is similar. PSI at this wavelength reduces the artifacts in the sky and therefore assists in horizon detection.

Horizon is usually identified by the Hough Transform as the most prominent Sobel/Canny straight edge (Zafarifar et al.; 2008), (Zafarifar and Peter; 2006) and (Mcgee et al.; 2005). Hough Transform projects points from the image to Hough space which has a high computational complexity. Additionally Hough Transform can identify either straight or curved edges depending on the mapping to Hough space. However the horizon can be curved (Figures 4.5, 4.6 and 4.7) due to the curvature of the earth, therefore a more robust algorithm is required which allows for both straight and curved edge representation of horizon. In the subsequent sections, the methods for extraction of PSI from RGB images and utilizing fitting ellipses for segregating horizon from multiple edges will be derived. Results from section 2 and 3 will be combined to form an algorithm for horizon detection in section 4.

4.2 Generating Pseudo Spectra Images from RGB

The motivation of this section is to generate a set of images at multiple wavelengths in the visible spectrum. Color is usually represented as a set of three intensities (Red, Green and Blue) obtained from the color filter arrays (CFAs) of the cameras/sensor. The camera capture is a many to one process, wherein the CFAs sample the reflectance from the scene at different wavelengths in the visible spectrum to give cumulative RGB values. The process of rendering is the opposite of camera capture, wherein the RGB (or equivalent 3D) value of pixel is given and the motivation is to have a color display on the hardware (monitor/printer) which produces a tri-stimulus response in the human visual system as close as possible to the actual scene. Glassner (Glassner; 1989) proposes a method to re-fabricate the reflectance from original scene, by transferring the RGB triplets to CIE XYZ color space, and then uses Color Matching Functions (CMFs) (Wyszecki and Stiles; 2000) to regenerate the discrete wavelength spectra.

If \mathbf{R} and \mathbf{X} represent the row vectors $[r, g, b]$ and $[x, y, z]$ respectively and \mathbf{M} the transformation matrix which maps \mathbf{R} to \mathbf{X} then:

$$\mathbf{X} = \mathbf{R}\mathbf{M} \quad (4.1)$$

Spectrum $A(\lambda)$ conversion to a point in XYZ space $\mathbf{X} [x, y, z]$ involves integrating $A(\lambda)$ over the visible spectrum with the CIE color matching functions (CMFs) $[\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)]$:

$$x = \int_{380}^{780} A(\lambda)\bar{x}(\lambda)d\lambda; \quad y = \int_{380}^{780} A(\lambda)\bar{y}(\lambda)d\lambda; \quad z = \int_{380}^{780} A(\lambda)\bar{z}(\lambda)d\lambda \quad (4.2)$$

Defining monochromatic wavelengths as:

$$(\lambda_u, \lambda_v, \lambda_w) \in [380, 780] \quad \forall u \neq v \neq w \quad (4.3)$$

Based on 3 monochromatic wavelengths, a point $\mathbf{S} = [S(\lambda_u), S(\lambda_v), S(\lambda_w)]$ in 3D space of monochromatic wavelengths in visible spectrum can be defined. \mathbf{S} provides an orthonormal basis for spectrum generation.

Define \mathbf{C} as a 3X3 subset of discrete CMFs (Wyszecki and Stiles; 2000), with amplitudes $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ with $\lambda \in (\lambda_u, \lambda_v, \lambda_w)$. Amplitudes of CMFs from CIE XYZ 64 (Wyszecki and Stiles; 2000) are plotted in Figure 4.3.

$$\mathbf{C} = \begin{bmatrix} \bar{x}(\lambda_u) & \bar{y}(\lambda_u) & \bar{z}(\lambda_u) \\ \bar{x}(\lambda_v) & \bar{y}(\lambda_v) & \bar{z}(\lambda_v) \\ \bar{x}(\lambda_w) & \bar{y}(\lambda_w) & \bar{z}(\lambda_w) \end{bmatrix} \quad (4.4)$$

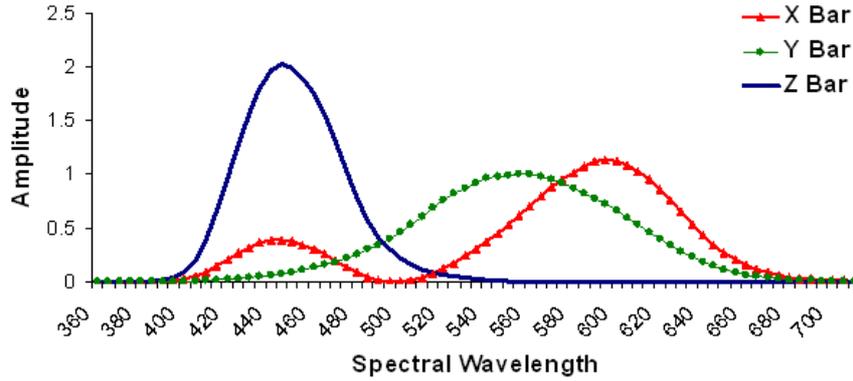


Figure 4.3: Plot of x,y,z weights at different wavelengths [9].

\mathbf{X} can be calculated from \mathbf{S} using \mathbf{C}

$$\mathbf{X} = \mathbf{S}\mathbf{C} \quad (4.5)$$

Equation (4.5) is a discrete version of (4.2) at 3 discrete orthonormal wavelengths. Using (4.5) and (4.1):

$$\mathbf{S} = \mathbf{X}\mathbf{C}^{-1} = \mathbf{R}\mathbf{M}\mathbf{C}^{-1} \quad (4.6)$$

CMFs \mathbf{C} are discrete representations of the tri-stimulus response of three cones of human eye to wavelengths in visible spectrum (Wyszecki and Stiles; 2000). Tristimulus response is continuous in the visible spectrum; therefore CMFs (\mathbf{C}) can assume infinite discrete values. Consequently \mathbf{S} computed from \mathbf{C} (4.5) and (4.6) can have infinite solutions (metamers). Equation (4.6) designed for rendering, can be adapted for analysis of images by ascertaining \mathbf{M} and \mathbf{C} . \mathbf{C} requires a mechanism for selecting orthonormal wavelengths $[\lambda_u, \lambda_v, \lambda_w]$. Subsequent sub sections will define PSI, \mathbf{C} , select $[\lambda_u, \lambda_v, \lambda_w]$ and present an algorithm for sampling the visible spectrum.

4.2.1 Definition and Derivation: PSI

A pseudo spectra image (PSI) is a monochromatic, gray-scale image, generated at a discrete wavelength. \mathbf{S} (4.6) is a 3D vector with the intensities $[S(\lambda_u), S(\lambda_v), S(\lambda_w)]$ corresponding to wavelengths $[\lambda_u, \lambda_v, \lambda_w]$. All the 3 dimensions (intensities) of \mathbf{S} are mathematically equivalent and can individually generate monochromatic images. The consideration in sampling the spectrum range is not the choice of dimension of \mathbf{S} , but consistency in the choice of dimension. In this thesis $S(\lambda_v)$ has been used to generate gray scale PSIs. Formally PSI is defined as a gray scale image $P(x, y)$ generated from the color image $\mathbf{R}(x, y)$ corresponding to wavelength $S(\lambda_v)$ and given by:

$$\begin{aligned} P(x, y) &= S(\lambda_v; x, y) \quad \text{where} \\ S(\lambda_v; x, y) &: \text{Intensity of } \mathbf{S}(x, y) \text{ at} \\ \lambda_v &: \text{Wavelength, } (x, y) : \text{Spatial coordinates} \end{aligned} \quad (4.7)$$

From (4.6) and (4.7):

$$\mathbf{S}(x, y) = \mathbf{R}(x, y)\mathbf{MC}^{-1} \quad (4.8)$$

$P(x, y)$ is the 2nd dimension of the vector $\mathbf{S}(x, y)$, therefore $P(x, y)$ can be computed from (4.7) and (4.8) as:

$$\begin{aligned} \mathbf{S}(x, y) &= \mathbf{R}(x, y)\mathbf{T} \quad \text{where} \\ \mathbf{T} &: \text{Column matrix from second column } \mathbf{MC}^{-1} \end{aligned} \quad (4.9)$$

Writing \mathbf{T} from the second column of \mathbf{MC}^{-1} :

$$\mathbf{T} = \begin{bmatrix} m_{11}c_{12} + m_{12}c_{22} + m_{13}c_{32} \\ m_{21}c_{12} + m_{22}c_{22} + m_{23}c_{32} \\ m_{31}c_{12} + m_{32}c_{22} + m_{33}c_{32} \end{bmatrix} \quad \text{where} \quad (4.10)$$

m_{ij} : Value of i row and j column in \mathbf{M}

c_{ij} : Value of i row and j column in \mathbf{C}

PSI computation (4.9) is reduced to the multiplication of the RGB values of pixels with the weights of the column matrix \mathbf{T} . The use of three dimensions in \mathbf{T} can be questioned,

because only one dimension is used in (4.10). Three dimensions are due to matrix inversion of \mathbf{C} (4.6) which requires a square matrix.

4.2.2 Mapping (\mathbf{M}) from \mathbf{RGB} to \mathbf{XYZ}

In the original work (Glassner; 1989) \mathbf{M} is fixed and computed from the chromatic response of the monitor. In the current application a fixed mapping \mathbf{M} for \mathbf{R} to \mathbf{X} is used. Image processing library openCV (Bradski; 2000) defines \mathbf{R} to \mathbf{X} mapping as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} \quad (4.11)$$

Taking the transpose of (4.11), \mathbf{M} and (4.1) can be obtained :

$$\begin{aligned} [\mathbf{X}] &= [\mathbf{R}][\mathbf{M}] \text{ where} \\ \mathbf{M} &= \begin{bmatrix} 0.412453 & 0.212671 & 0.019334 \\ 0.357580 & 0.715160 & 0.119193 \\ 0.180423 & 0.072169 & 0.950227 \end{bmatrix} \end{aligned} \quad (4.12)$$

4.2.3 Orthogonal Wavelength Basis

There are no rules for choosing the basis wavelengths (other than considerations for numerical stability) as infinite spectra $A(\lambda)$ (metamers) can give the same values of \mathbf{X} and \mathbf{R} . The choice of basis wavelengths is application specific. For example Glassner (Glassner; 1989) chooses $[\lambda_u, \lambda_v, \lambda_w]$ as the wavelengths at which the sampled CMFs $[\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)]$ attain their maxima. However for the current research the orthonormal wavelengths are defined as follows: Let the wavelength range of the spectrum being sampled be defined by λ_{range} :

$$\begin{aligned} \lambda_{range} &= \lambda_{max} - \lambda_{min} \text{ where} \\ \lambda_{max} &: \text{ The maximum wavelength in the spectrum} \\ \lambda_{min} &: \text{ The minimum wavelength in the spectrum} \end{aligned} \quad (4.13)$$

The set of orthogonal wavelengths $[\lambda_u, \lambda_v, \lambda_w] \in (\lambda_{min}, \lambda_{max})$ are constrained to the defined spectrum range by:

$$\begin{aligned}
 & \text{if } \lambda_n > \lambda_{max} \text{ then } \lambda_n = \lambda_{min} + (\lambda_n - \lambda_{max}) \\
 & \text{if } \lambda_n < \lambda_{min} \text{ then } \lambda_n = \lambda_{max} - (\lambda_{min} - \lambda_n) \\
 & \text{where } n \in (u, v, w)
 \end{aligned} \tag{4.14}$$

Since an unbiased (towards any wavelength) set of images is being generated; therefore an orthonormal wavelength basis is defined as a set of 3-wavelengths which are confined to the spectrum range (4.14) and equidistant from each other (4.15):

$$\begin{aligned}
 & \text{For a given } \lambda_v \in \lambda_{range} \\
 & \lambda_u = \begin{pmatrix} \lambda_v - \frac{\lambda_{range}}{3} & \text{if } \lambda_v \geq \frac{\lambda_{range}}{3} + \lambda_{min} \\ \lambda_v - \frac{2\lambda_{range}}{3} & \text{if } \lambda_v < \frac{\lambda_{range}}{3} + \lambda_{min} \end{pmatrix} \\
 & \lambda_w = \begin{pmatrix} \lambda_v + \frac{\lambda_{range}}{3} & \text{if } \lambda_v \leq \lambda_{max} - \frac{\lambda_{range}}{3} \\ \lambda_v - \frac{2\lambda_{range}}{3} & \text{if } \lambda_v > \lambda_{max} - \frac{\lambda_{range}}{3} \end{pmatrix}
 \end{aligned} \tag{4.15}$$

Thus one value λ_v of the orthogonal wavelength set can discretely sample the visible spectrum. The other two values (λ_u, λ_w) of the orthogonal wavelength set can be generated from (4.14) and (4.15). Hence an orthogonal basis can be computed at each wavelength wherein the wavelengths are confined to the visible spectrum and equidistant from each other.

4.2.4 Algorithm: Sampling Spectrum for Generating PSI

Sampling the spectrum at discrete wavelengths involves two concerns.

- *Determining the spectrum range:* From CMFs (Wyszecki and Stiles; 2000) and Figure 4.3, it follows that the weights are approximately zero ($\bar{x} \approx 0, \bar{y} \approx 0, \bar{z} \approx 0$) for wavelengths ($700 \leq \lambda \leq 400$). Therefore the spectrum range from 400 to 700 nm is sampled in wavelength increments of 5nm. Increment of 5nm is inherent in CMF (Wyszecki and Stiles; 2000) therefore it was chosen as the sampling interval.
- *Orthonormality of the basis wavelengths $(\lambda_u, \lambda_v, \lambda_w)$:* Sampling in λ_{range} is being done at an interval of 5nm, i.e. 60 wavelengths are sampled. The orthonormality

((4.15)) is maintained by keeping the basis wavelengths 100 nm apart. Algorithm 1 generates PSIs after orthogonal wavelengths' sampling and selection subject to ((4.14)) and ((4.15)) and CMF (Wyszecki and Stiles; 2000):

Algorithm 1: Steps to generate PSI in the visible spectrum

for i: 0 to 59

1. $\lambda_v = 400 + i \times 5$
2. (If $i \geq 20$), $\lambda_u = 400 + (i - 20) \times 5$ else $\lambda_u = 400 + (40 + i) \times 5$
3. (If $i \leq 40$), $\lambda_w = 400 + (i + 20) \times 5$ else $\lambda_w = 400 + (i - 40) \times 5$
4. Populate \mathbf{C} from (4.4) and $(\lambda_u, \lambda_v, \lambda_w)$
5. Compute Transformation matrix \mathbf{MC}^{-1} from (4.6)
6. Compute Column matrix \mathbf{T} from (4.10)
7. Compute PSI by applying (4.9) to all pixels.

Algorithm 1 generates a set of 60 PSIs corresponding to wavelengths 400 to 700nm in increments of 5nm. These 60 PSIs are metamers as they have been generated from the same set of \mathbf{R} , albeit at different wavelengths. These metamers are plausible spectra of the actual scene reflectance. The PSI(S) is a pseudo wavelength image; because it is created from mathematical considerations [(4.4) to (4.15)] of image capture/representation (\mathbf{R}, \mathbf{X}) and mapping (\mathbf{M}) at a particular wavelength λ_v with the view to mimic the human visual system (\mathbf{C}). The ability of the PSI (S) to replicate reflectance of original scene at λ_v is not mandatory and should not be expected. It should be re-emphasized that PSI is not a unique solution of (4.6). It is well known (Wyszecki and Stiles; 2000), (Glassner; 1989) that (4.6) is an under-constrained and therefore ill-posed problem as infinite solutions of (4.6) exist. The reason for designing PSI is its ability to elicit similar intensity response for different colors at particular wavelengths.

4.2.5 Application of PSI: Sky Detection

Multiple PSIs generated using Algorithm 1 revealed that the sky and clouds had similar intensity at wavelengths from 405 to 425nm with the optimum being 415 nm. Some of the PSIs (with cloudy sky) from Berkley Database (Martin et al.; 2001) processed at an orthogonal wavelength basis centered at 415nm are shown in Figure 4.4. PSIs in Figure 4.4 highlight the similar intensity of the sky and its common artifacts at 415 nm.

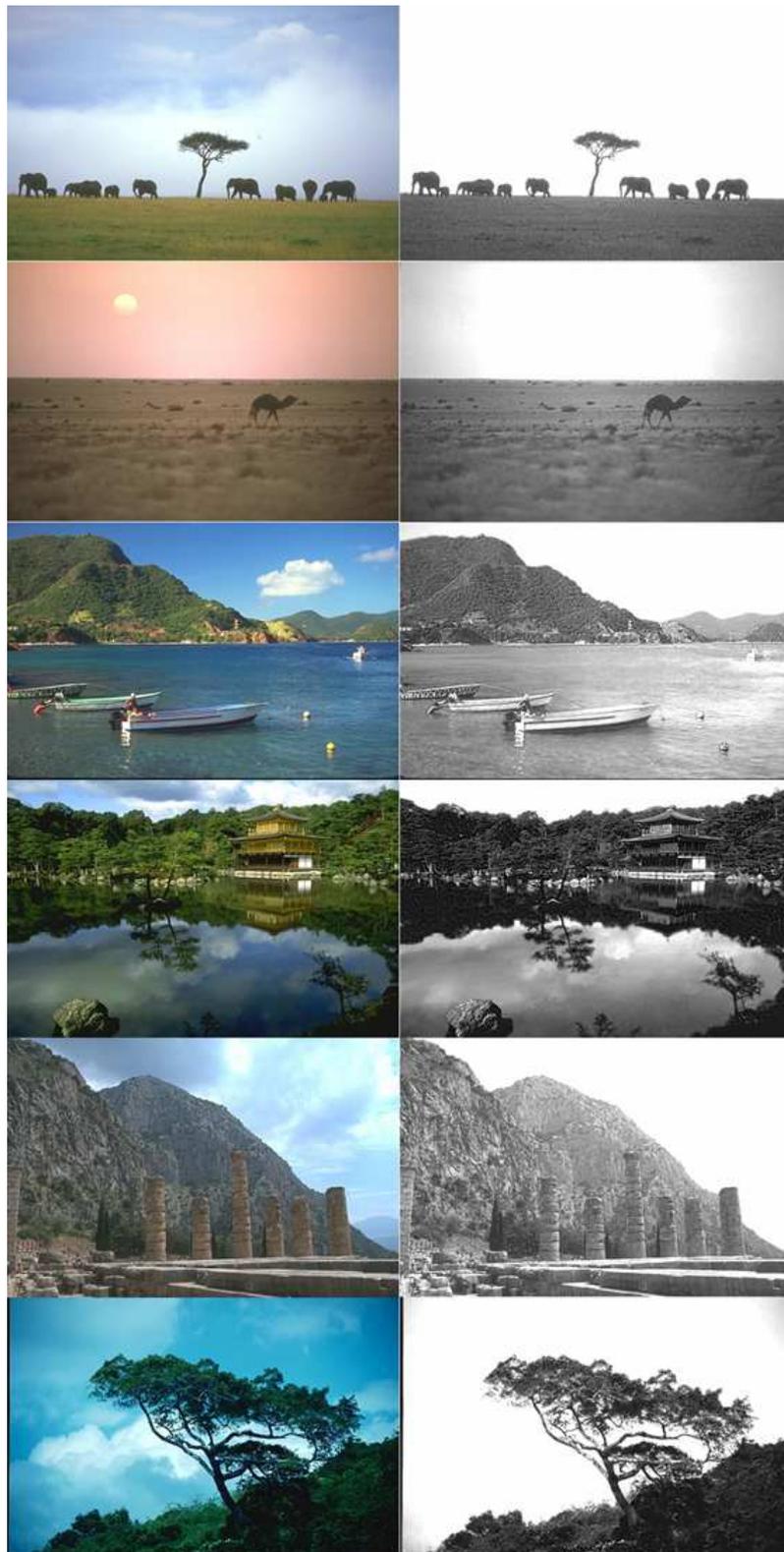


Figure 4.4: Pseudo Spectra Images.

Sky portion of PSI usually has high intensity, with occasional loss of detail in images having chromatic similarity with the sky. This problem assumes significance when locating the horizon on water bodies with bright sky (Figure 4.5 Top Left). The characteristic high

luminance near the horizon arising out of distance to horizon is compounded by reflectance from water (Figure 4.5 Top Right). This problem can be resolved by generating PSI at an orthogonal wavelength basis centered at 475 nm. At this wavelength (475 nm) the similarity of intensity of the sky and clouds is reduced as compared to 415 nm (Figure 4.5 Bottom Left). For some applications PSI at 475 nm subtracted from PSI at 415 nm might offer more details in the image below horizon (Figure 4.5 Bottom Right).



Figure 4.5: PSI: Intensity variation at different wavelengths

Top Left: Bright sky with clouds. *Top Right:* PSI(415 nm) with high saturation near horizon. *Bottom Left:* PSI(475 nm) without artifacts and low saturation near horizon. *Bottom Right:* PSI (475 nm) subtracted from PSI(415 nm) showing more detail below horizon.

4.2.6 PSI: Characteristics and Discussion

The PSI induces disparate (or similar) response to different chromatic stimuli, which when suitably integrated can augment the scene understanding/classification by chromatic denoising. Figures 4.4 and 4.5 illustrate two methods of exploiting PSI:

- *Similar Response:* In Figure 4.4 varying (inter and intra images) chromaticity of sky and clouds have resulted in similar intensities of the sky portion of the PSI which illustrates the ability of PSI to smoothen chromatic variations using one wavelength.

- *Similar/Dissimilar Response:* In Figure 4.5 different responses of two components in an image at different wavelengths are illustrated. Sky and cloud have similar intensities at two different wavelengths (415 and 475), but sea has differing intensities at these wavelengths. Subtracting PSI at different wavelengths, segments the image (Figure 4.5. Bottom Right). The usage of the similar/dissimilar response at different wavelengths, to remove artifacts in sky is comparable to the biologically inspired research (Thakoor et al.; 2002) and (Chahl et al.; 2003). However the method presented herein uses mathematical techniques in visible spectrum, unlike (Thakoor et al.; 2002) and (Chahl et al.; 2003) which uses UV and dedicated hardware.

Numerical stability is a concern in generating PSIs, as highlighted in (Glassner; 1989) and experienced by the author. Matrix inversion (4.6) results in negative values and the overflow of intensity, which must be handled explicitly. E.g. 8 bit image representation in current work was exceeded, with calculations resulting in values more than 255 and less than zero. Owing to variability in image acquisition, storage, representation and processing some experimentation is recommended to obtain the optimum wavelength before utilizing the results presented in this section. Sky detection using PSI, removes artifacts in the sky without any segmentation or heuristics.

4.3 Horizon Detection by Fitting Ellipses

This section derives an expression for horizon detection from the parameters of a fitting ellipse. Edges are treated as eight-connected components and their fitting ellipse is obtained from the method by Chaudhari and Samanta (Chaudhuri and Samanta; 1991) which equates the zeroth order moment of the ellipse to that of the connected component. For a *digital edge binary image* $P(x, y)$ with connected components the $(p + q)^{th}$ order moment of the k^{th} connected component is defined as:

$$mom_k(p, q) = \sum x^p y^q \quad \text{where} \quad (4.16)$$

$$P(x, y) = (0 \quad \text{or} \quad 1) \quad \text{and} \quad \in k$$

The coordinates of Centroid (x_{ck}, y_{ck}) of the k^{th} connected component are:

$$x_{ck} = \frac{mom_k(1, 0)}{mom_k(0, 0)}, \quad y_{ck} = \frac{mom_k(0, 1)}{mom_k(0, 0)} \quad (4.17)$$

The alignment (θ_k) of the longer axis of the fitting-ellipse for the k^{th} connected component with the X-axis is given by:

$$\begin{aligned} \theta_k &= 0.5 \tan^{-1} \left[\frac{2\delta_{k(xy)}}{\delta_{k(xx)} - \delta_{k(yy)}} \right] \quad \text{where} \\ \delta_{k(xx)} &= mom_k(2, 0) - \frac{[mom_k(1, 0)]^2}{mom_k(0, 0)} \\ \delta_{k(xy)} &= mom_k(1, 1) - \frac{mom_k(1, 0) \times mom_k(0, 1)}{mom_k(0, 0)} \\ \delta_{k(yy)} &= mom_k(0, 2) - \frac{[mom_k(1, 1)]^2}{mom_k(0, 0)} \end{aligned} \quad (4.18)$$

The major $(2a_k)$ and minor $(2b_k)$ axis of the k^{th} fitting ellipse ¹ can be calculated from the moment of inertia $I_k(x_{ck}, y_{ck}, \theta_k)$ by:

$$I_k(x_{ck}, y_{ck}, \theta_k) = \delta_{k(xx)} \cos^2(\theta_k) + \delta_{k(yy)} \sin^2(\theta_k) - 2\delta_{k(xy)} \cos(\theta_k) \sin(\theta_k) \quad (4.19)$$

$$b_k = 2 \sqrt{\frac{I_k(x_{ck}, y_{ck}, \theta_k)}{mom_k(0, 0)}}; \quad a_k = \frac{mom_k(0, 0)}{\pi b_k} \quad (4.20)$$

Since the horizon is the largest edge (either straight or with slight curvature) therefore the following parameters of the fitting ellipse of the horizon should be largest for all the edges present in image:

- The skew: $\frac{b_k}{a_k}$
- The zeroth moment (area): $mom_k(0, 0)$

Using the product of the zeroth moment and the skew of the fitting ellipse a new parameter termed as *Horizon Measure (HM)*, can be defines a for edges:

$$HM_k = \frac{b_k}{a_k} mom_k(0, 0) = \pi(b_k)^2 \quad (4.21)$$

¹The major and minor axis referred in this text are those computed from moments. It should be pointed out that an ellipse also has its own Geometric axes which are opposite to that of Moment axes. E.g. major $(2a_k)$ (moment) axis of the fitting ellipses is the minor geometric axis and vice versa. The reader is cautioned against confusing the two.

The edge which maximizes the minor axis of the fitting ellipse (b_k) is the horizon (Figure 4.6 Top Right).

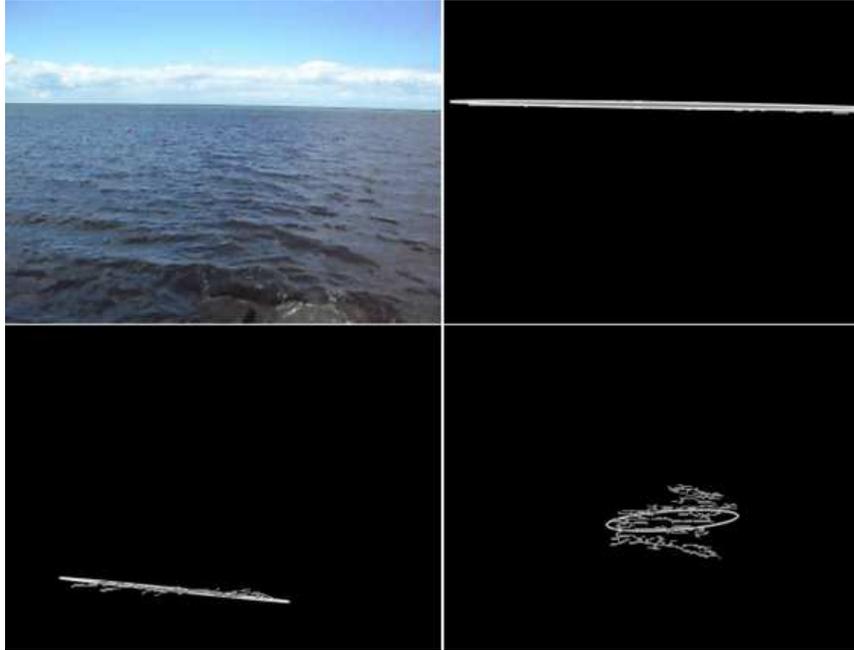


Figure 4.6: PSI: Intensity variation at different wavelengths

Top Left: Image. *Top Right:* Horizon Measure-(Skew X Area) of ellipse. *Bottom Left:* Horizon Measure-(Skew). *Bottom Right:* Horizon Measure-(Area). In Gray scale images: The fitted ellipse and enclosed edge which maximize the horizon measure are shown in white and gray respectively.

The problems, in using the skew and the zeroth moment, in isolation are eliminated by taking their product (Figure 4.6 Top Right). By taking the product of the skew and the zeroth moment as a HM, individual false positive probabilities (of skew and zeroth moment) are multiplied, resulting in reduced total probability of a false positive. Qualitatively it can be stated that it is more difficult to find an edge in water that is as straight (skew) and long (zeroth moment) as the horizon, than finding edges that are as straight or as long as the horizon.

Hough transform is a popular choice for detecting the horizon, but requires projection of pixels to Hough space which usually has a complexity of $O(\frac{N\pi}{\rho})$ where N is the number of pixels evaluated and ρ is the angular bin size. The ellipse is calculated from the first two moments of connected components (Chaudhuri and Samanta; 1991), hence the total complexity in horizon detection from image edge map is $O(3N)$; N for calculation of connected components (Chang et al.; 2004) and $2N$ for the first two moments of connected

components. Therefore ellipses are faster than Hough transform and can identify straight or curved edges.

4.4 Algorithm: Water Horizon from Derivative of PSI

Previous sections have presented two theoretical aspects of horizon detection. These are combined to generate an algorithm (2) for horizon detection..

Algorithm 2: Steps for horizon detection

1. Identify the appropriate wavelength and the transformation matrix at which the intensity of the sky and clouds is similar.
2. Generate the PSI of the image/video using (4.9).
3. Convolve the PSI with a 2-D gaussian of standard deviation 1.
4. Compute the Sobel edges of the smoothed PSI
5. Threshold the Sobel edge from PSI using the Otsu's Threshold (Otsu; 1979).
6. Perform the 8-connectivity (Chaudhuri and Samanta; 1991) on the thresholded image and calculate the parameters of enclosing ellipses.
7. Calculate the HM (4.21) of the 10 largest connected components to identify the horizon.

Step 3 of Algorithm 2 reduces noise and produces a continuous edge from a fractured horizon. Figure 4.8 illustrates horizon candidates with and without a gaussian operator. Step 5 reduces the search space by removing edges with lower intensity values. The horizon edge is computed from heterogeneous backgrounds (sky and water), unlike other edges in the image which are generated from the homogeneous backgrounds; therefore the intensity of the horizon edge is usually greater than the other edges (Walia and Jarvis; 2009). The higher intensity of the horizon edge enables the horizon to belong to the set of edges left remaining after thresholding by Otsu's Threshold (Otsu; 1979),(Walia and Jarvis; 2009). Step 6 computes the connected components and the minor axes of the enclosing ellipses. Step 7 maximizes the HM of the edges, thereby identifying the horizon.

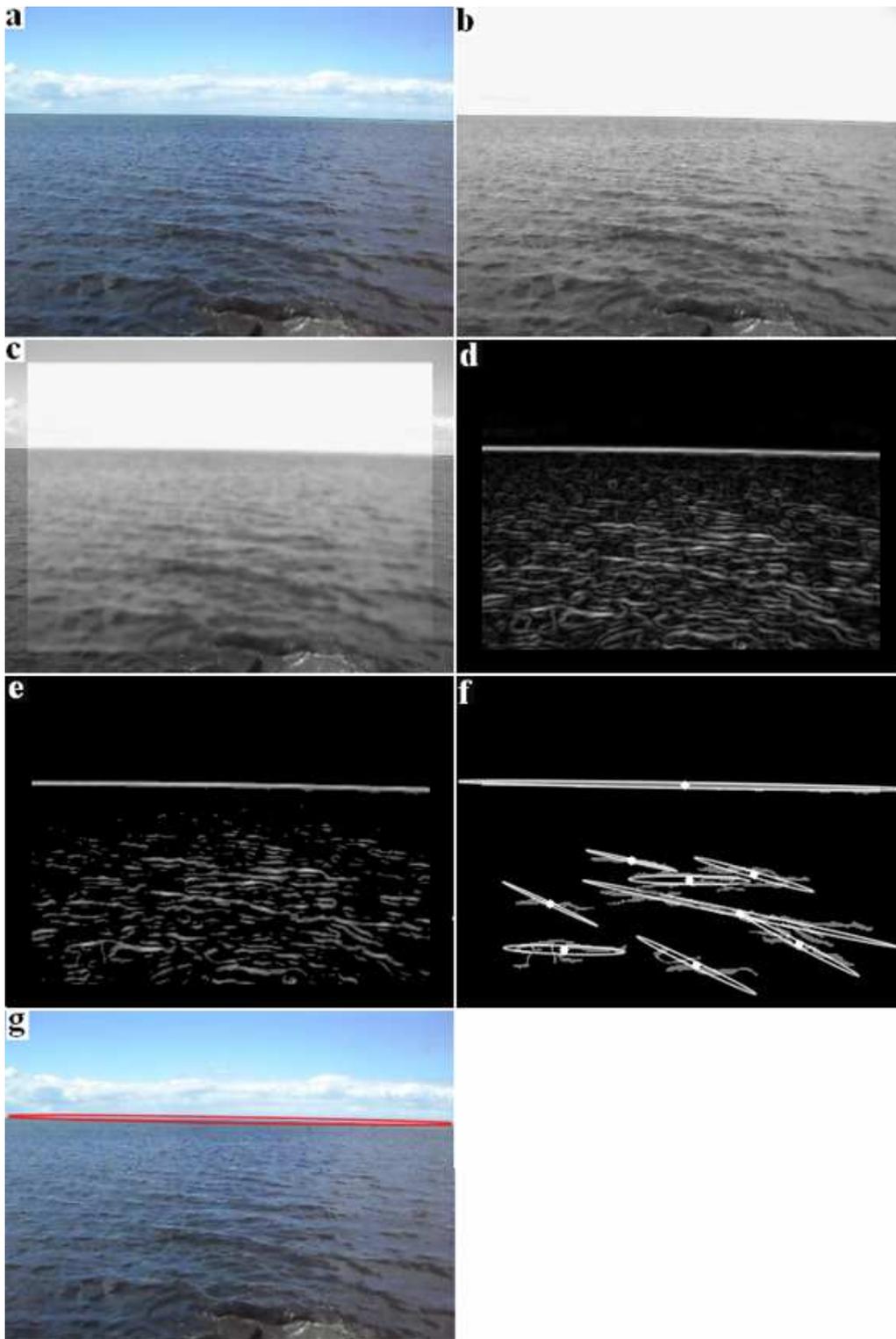


Figure 4.7: Steps of Algorithm 2

a: Image. *b*: PSI (Step 2). *c*: Smoothed PSI (Step 3) *d*: Sobel of Smoothed PSI (Step 4). *e*: Sobel edges of Smoothed PSI thresholded at Otsu's Threshold (Step 5). *f*: 10 largest edges with enclosing ellipses (Step 6). *g*: Horizon (red fitting ellipse) identified by maximizing HM (Step 7)

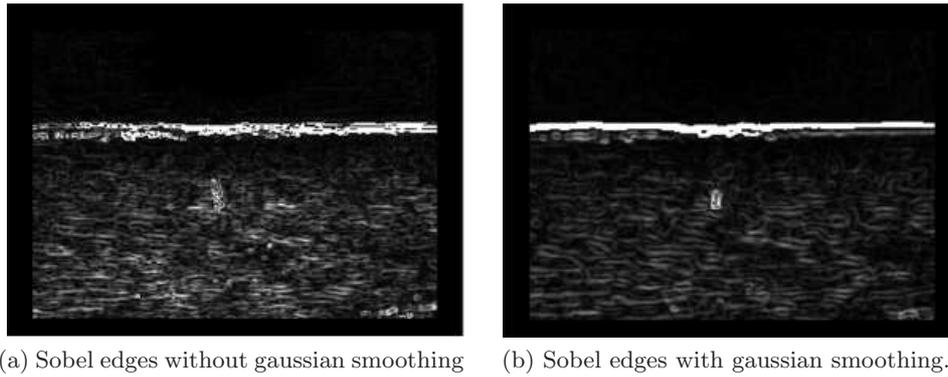


Figure 4.8: A continuous horizon from a fractured horizon

4.5 Conclusion

In this chapter, theoretical framework for generating and analyzing PSI has been presented. The simplicity of PSI generation and uniqueness of results should allow its adaptation for other vision based applications. The theoretical framework was augmented with application viz. horizon detection. Edge classification in horizon detection has also been discussed and fitting ellipses have been presented as a faster alternate to Hough transform. The ellipses can replace Hough Transforms for autonomous navigation in water and air crafts i.e. where occlusion of horizon is not a frequent occurrence. It is envisaged that, PSI can be used as a pre-processing step for eliminating chromatic aberrations in natural textures like grass, water, skin, fruit/vegetable etc, with a strong predisposition to certain colors. Specific applications of PSI for sky detection can be in horizon detection, video processing (Zafarifar and Peter; 2006), scene labeling and robotics i.e. in autonomous watercrafts/aircrafts (Cornall and Egan; 2004), (Chahl et al.; 2003) and (Thakoor et al.; 2002). Sophisticated patterns based on intensities of PSIs at multiple wavelengths can potentially be described to identify specific entities in images. From the perspective of obstacle avoidance in water, this chapter achieves the segregation of water from sky. The contrast between the water and sky is enhanced using PSI and the interface between water and sky viz horizon is used to demarcate the two regions. By utilizing the interface edge rather than water, the problems associated with variability are avoided. In the succeeding chapters the obstacle will be detected from edge it creates with the water.

Chapter 5

Locating Objects in Water: Theory

As previously stated, this chapter discusses obstacle detection. While engineering and science is challenged to detect obstacles in water, the apparent ease with which human visual system is able to detect the obstacles in water is remarkable. This ability of biological vision, is not restricted to humans, but most of the animals are also inherently capable of recognizing obstacles in the environment that they inhabit. The front end of human vision system could therefore provide a solution to obstacle detection. Human vision system utilizes edges and neurons have been reported (Young; 1987) to have response similar to gaussian derivatives (up to fourth order). The human vision formed the '**pull factor**' in deciding the methodology of research. Scale-Space was chosen as the framework for detecting obstacles owing to documented similarity to human vision (Hubel and Wiesel; 1987; Young; 1987; Koenderink and van Doorn; 1987; Koenderink and Doorn; 1992).

To compensate for variability in water, the edge(s) that water makes with the obstacle is used as the obstacle's identifying feature. In the physical 3 dimensional world, the edge due to the interface of water and obstacle is characterized by a physical discontinuity between the surface of water and the surface of the obstacle. This physical discontinuity manifests as an image edge when the 3 dimensional world is projected onto the image plane (Figure 2.1). In the analysis presented in the remaining chapter, the interface edge is modeled as a theoretical discontinuity in functions.

The qualitative ideology underlying this chapter is that the interface edges will not exist unless there is an obstacle in water. In the absence of an obstacle, the water is a continuous

surface that can be represented mathematically as a continuous function. Therefore if a discontinuity were detected in the images, it can be justifiably and heuristically attributed to the presence of an obstacle. In addition the location of the discontinuity demarcates the physical boundary of the obstacle. Therefore, in this chapter, the problem of obstacle detection is mathematically analyzed as that of discontinuity detection. Specifically a statistical solution to detect discontinuity is provided by analyzing the statistical properties of a discontinuity in its Scale-Space representation. The research reported here resulted in a conference publication (Walia and Jarvis; 2009) and a book chapter (Walia et al.; 2012).

5.1 Introduction

Discontinuity detection is studied across disciplines of thermodynamics, chemistry, geology, manufacturing, equipment maintenance, signal processing, computer architecture (bit recognition), finance (jump processes to model markets) and image processing. In image processing an edge is often modeled as a discontinuity (Lindeberg; 1998). Hence discontinuity detection, can provide edge information for image analysis with applications in robotic vision, medical imaging, tomography and surveillance etc. Scale-Space theory (Lindeberg; 1994; Witkin; 1983; Koenderink; 1984), is a framework for multi-scale analysis of function/image. While there are non-Gaussian Scale-Space representations (Duits et al.; 2003), this chapter is confined to widely accepted Gaussian Scale-Space (Lindeberg; 1994; Babaud et al.; 1986). Existing Scale-Space literature is focused mainly on developing Scale-Space theory with a view to:

1. Study impact on underlying signals/images (Witkin; 1983; Koenderink; 1984; Romeny; 1994; Babaud et al.; 1986; Lindeberg; 1998)
2. Determine appropriate scale(s) relevant to the image/signal(Lindeberg; 1994).
3. Extract information, knowledge and develop applications like feature detection, feature classification, image segmentation, image matching, motion estimation, shape computation and object recognition etc.
4. Correlate the Scale-Space framework with biological vision (Hubel and Wiesel; 1987; Young; 1987; Koenderink and van Doorn; 1987; Koenderink and Doorn; 1992).

Current Scale-Space literature, does not adequately explore the statistical component of Scale-Space. There are contextual applications of various statistical parameters (Zagal et al.; 2000; Rodriguez; 2006; Sakai and Imiya; 2009), in contemporary Scale-Space research, but they are limited in scope to specific applications or/and statistics of image features like blob volume, clusters, thresholds etc. Researchers would be well-assisted if some theoretical basis were available for statistical assumptions in Scale-Space. In this chapter, theoretical foundation for some statistical assumptions with regard to the derivative of a discontinuity in Scale Space are presented. A discontinuity has an infinitesimal existence in Scale-Space, which leads to the assumption of continuity of underlying image/function in any conventional Scale-Space analysis (Lindeberg; 1994; Koenderink; 1984). This chapter reveals that even though Scale-Space eliminates discontinuity at an infinitesimal scale, the Probability Density Function (PDF) of the derivative of a discontinuity retains its unbalanced bimodality in Scale-Space. This chapter makes following theoretical contributions:

1. Derivation of Probability Density Function (PDF) and Cumulative Distribution Function (CDF) for the derivative of a discontinuity in Scale-Space (Theorem 2).
2. Proof of bimodality (Theorem 3) and unbalance (Theorem 4) of the PDF of the derivative of a discontinuity in Scale-Space.
3. Proof that the Otsu's Threshold (OT) (Otsu; 1979) owing to its sensitivity to unbalanced and bimodal PDFs has different patterns in Scale-Space based on the presence / absence of a discontinuity:
 - (a) Transient Increase: Discontinuity present.
 - (b) Monotone Decrease: Discontinuity absent.

The above mentioned theoretical results, are then applied for a simultaneous solution of following problems in image processing (Figure 5.1):

1. Scale appropriate to the discontinuity.
2. Threshold appropriate to the discontinuity.
3. Boundaries of entities in images.

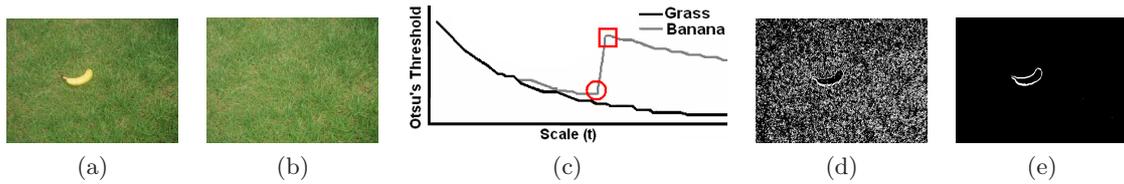


Figure 5.1: OT Characteristics: Absence and Presence of a foreground.

(a) “Banana”: Foreground and Background Image. (b) Segmentation at lower point of inflection. (c) Segmentation at Upper point of inflection. (d) “Grass”: Background Only. (e) OT Plots against Scale for “Grass” and “Banana”. □ and ○ : Upper and Lower points of inflection.

5.2 Statistical Distributions of a Gaussian Function

The term Gradient Magnitude in Scale-Space (GMSS) will be used hereon to represent

1. *In 1-D Non-Discrete Functions*: The derivative of the Scale-Space representation of the functions.
2. *In 2-D Discrete Images*: The magnitude of the gradient (computed by Sobel operator) of Scale-Space representation of the images.

In this section, the statistical distributions of the GMSS of a discontinuity will be derived. The reason for doing so is to show that the PDF of the GMSS of a discontinuity is bimodal and unbalanced i.e. the probability of one mode far exceeds the probability of the other mode. A discontinuity is mathematically represented as a step function. Consequently the derivative of the discontinuity is a Dirac’s Delta (Khuri; 2004) as shown in Figure 5.2. Convolution of the Dirac delta (δ)¹ with a Gaussian function will result in GMSS of a discontinuity.

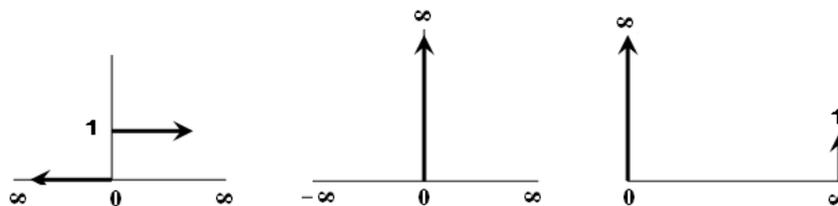


Figure 5.2: Left to Right: Unit step function, Dirac Delta and PDF of Dirac Delta

¹Since a Dirac’s Delta is zero everywhere except at one point, therefore its PDF will be of the type shown in Figure 5.2. The PDF of a Dirac’s Delta exists only at two points i.e. at ($x = 0$ and $x = \infty$) and is bimodal and unbalanced.

Theorem 1. The GMSS $L'(x; t)$ of a Step function is given by a Gaussian Kernel $G(x; t)$:

$$\begin{aligned}
 L'(x; t) &= G(x; t) \quad \text{where} \\
 G(x; t) &= \frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}} : \text{Gaussian Kernel at Scale (t)} \\
 &(-N < x < N : N \rightarrow \infty) : \text{Domain of the Function}
 \end{aligned} \tag{5.1}$$

Proof. For a Step function $H(x)$, the Scale-Space (SS) representation $L(x; t)$ is given by

$$H(x) = \begin{cases} 1 & \text{if } N > x > 0 \\ 0 & \text{if } -N < x < 0 \end{cases} \tag{5.2}$$

$$L(x; t) = H(x) \otimes G(x; t) \tag{5.3}$$

Convolution commutates with differentiation and the derivative of a Step function is a Dirac Delta ($\hat{\delta}(x)$) (Khuri; 2004) (Figure 5.2). Consequently:

$$\begin{aligned}
 L'(x; t) &= \frac{d}{dx}[H(x) \otimes G(x; t)] \\
 &= \frac{d}{dx}[H(x)] \otimes G(x; t) \\
 &= \hat{\delta}(x) \otimes G(x; t) \\
 &= G(x; 0) \otimes G(x; t)
 \end{aligned} \tag{5.4}$$

$$\text{where} \quad \hat{\delta}(x) = H'(x) = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

A Gaussian Kernel reduces to Dirac Delta at zero Scale i.e ($\hat{\delta}(x) = G(x; 0)$), therefore (5.4) is equivalent to the convolution of two Gaussian Functions with scales ($t_1 = 0, t_2 = t$). Since $[G(x; t_1 + t_2) = G(x; t_1) \otimes G(x; t_2)]$ (Lindeberg; 1994), therefore (5.4) simplifies to:

$$L'(x; t) = G(x; t) \quad \text{where } x \in 2N \quad \square$$

Theorem 1 simplifies the GMSS of a discontinuity to a Gaussian Function, which in turn allows formulation of the statistical characteristics of the GMSS of a discontinuity.

5.2.1 PDF of a Gaussian Function

Theorem 2. A continuous random variable \mathbf{g} which takes the values $g \in (\frac{1}{\sqrt{2\pi t}}, \frac{1}{\sqrt{2\pi t}} e^{-\frac{N^2}{2t}} : N \rightarrow \infty)$, given by a Gaussian function $G(x; t)$ has following statistical distributions:

$$PDF : \quad f_g(g) = \frac{t}{Ng\sqrt{-t\log_e(2\pi tg^2)}} \quad (5.5)$$

$$CDF : \quad F_g(g) = \frac{\sqrt{-t\log_e(2\pi tg^2)}}{N} \quad (5.6)$$

Proof. Gaussian function ($G(x; t)$) is symmetric and provides a one to one, monotonic and inverse mapping between x and g , in each half of the Cartesian plane. Therefore a *uniformly distributed* random variable (\mathbf{X}) which takes the values $x \in (0, N : N \rightarrow \infty)$ in the positive spatial domain of the Gaussian function can derive the PDF and CDF for \mathbf{g} . The uniformly distributed PDF of (\mathbf{X}) is given by:

$$f_x(x) = \frac{1}{N} \text{ where } x \in \{0, N : N \rightarrow \infty\} \quad (5.7)$$

The equivalent PDF and the domain for the gaussian variable (\mathbf{g}) is given by:

$$f_g(g) \text{ where } g \in \left\{ \frac{1}{\sqrt{2\pi t}}, \frac{1}{\sqrt{2\pi t}} e^{-\frac{N^2}{2t}} : N \rightarrow \infty \right\} \quad (5.8)$$

The probabilities for both the random variables are equal in the mapped ranges :

$$\int_{\frac{1}{\sqrt{2\pi t}}}^g f_g(g) dg = \int_0^x f_x(x) dx \quad \text{where} \quad (5.9)$$

$$x = G^{-1}(g; t) = \sqrt{-t \log_e(2\pi tg^2)}$$

Introducing a change of variable from x to g in the right hand side of (5.9) and solving:

$$\int_{\frac{1}{\sqrt{2\pi t}}}^g f_g(g) dg = \int_{\frac{1}{\sqrt{2\pi t}}}^g f_x(x) \left\| \frac{d[G^{-1}(g; t)]}{dg} \right\| dg \quad (5.10)$$

$$f_g(g) = \frac{t}{Ng\sqrt{-t\log_e(2\pi tg^2)}} \quad (5.11)$$

The CDF ² can be computed by integrating the PDF (5.11):

$$P[g \leq \mathbf{g} \leq \sqrt{2\pi t}] = F_g(g) = \frac{\sqrt{-t \log_e(2\pi t g^2)}}{N} \quad \square$$

Alternate proof of PDF($F_g(g)$), can be provided by replacing the value of (x) from (5.9) in the CDF ($F_x(x) = x/N$) of uniformly distributed variable \mathbf{X} and then differentiating it w.r.t (g), which would provide expression (5.11).

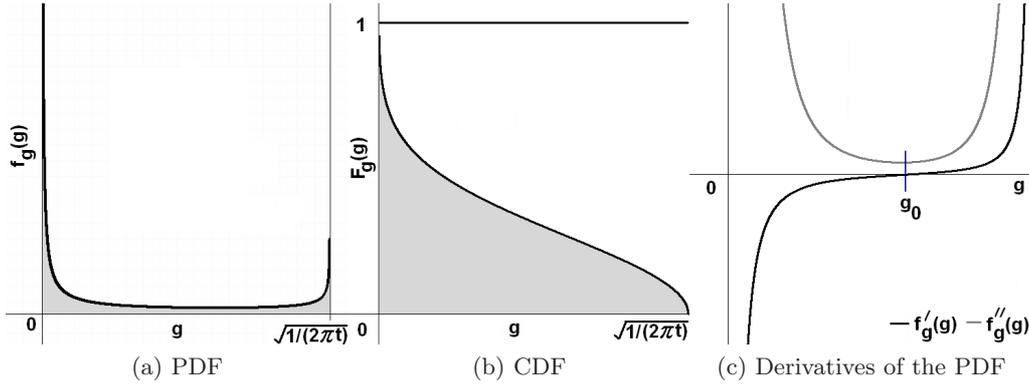


Figure 5.3: PDF, CDF and derivatives (w.r.t. g) of the PDF of a Gaussian Function.

5.2.2 Bimodality of the PDF of a Gaussian Function

Theorem 3. The PDF $f_g(g)$ of a Gaussian function is bimodal.

Proof. The bimodality of the PDF can be proved by the existence of exactly one minima in the PDF ((Eisenberger; 1964; Kemperman; 1991; Schilling et al.; 2002)). For a point (g_0) belonging to the domain of the PDF to be a minima, it's first derivative should be zero ($f'_g(g_0) = 0$) and it's second derivative should be positive ($f''_g(g_0) > 0$). The first and the second derivatives (with respect to g) of the PDF are given by:

$$f'_g(g) = \frac{t^2}{Ng^2(G^{-1}(g;t))^3} - \frac{t}{Ng^2G^{-1}(g;t)} \quad (5.12)$$

$$f''_g(g) = \frac{3t^3}{Ng^3(G^{-1}(g;t))^5} - \frac{3t^2}{Ng^3(G^{-1}(g;t))^3} + \frac{2t}{Ng^3(G^{-1}(g;t))} \quad (5.13)$$

²The correctness of the expression can easily be verified by replacing the term ($\sqrt{-t \log_e(2\pi t g^2)}$) in CDF with x from (5.9). This gives ($P[0 \geq \mathbf{X} \geq x] = F_x(x) = x/N$) which is the expression of a CDF of a uniformly distributed variable (\mathbf{X})

The minima g_0 in the PDF can be located from the root(s) of (5.12)

$$\begin{aligned} \frac{t^2}{g_0^2(G^{-1}(g_0; t))^3} &= \frac{t}{g_0^2 G^{-1}(g_0; t)} \\ \Rightarrow g_0 &= \frac{1}{\sqrt{2e\pi t}} \quad \forall \quad t > 0 \end{aligned} \quad (5.14)$$

Existence of only one root for the first derivative of the PDF, implies that only one extrema exists in the PDF. Substituting the value of g_0 from (5.14) into (5.13) and solving

$$f_g''(g_0) = 4\sqrt{2}t^2(e\pi)^{\frac{3}{2}} > 0 \quad (5.15)$$

From (5.14) and (5.15), g_0 is the (only) minima in the PDF, therefore it establishes the bimodality of the PDF (Figure 5.3). \square

5.2.3 Unbalance in modes of the PDF of a Gaussian Function

Theorem 4. The bimodal PDF $f_g(g)$ of the GMSS of a Step Discontinuity is unbalanced, i.e the probability of one mode is much greater than the other.

$$F_g(g_0) \ll 1 - F_g(g_0) \quad (5.16)$$

Proof. The probabilities of the two modes separated at g_0 can be computed from CDF (5.6)

$$\begin{aligned} P(g \geq g_0) &= F_g(g_0) \\ &= \frac{g_0}{N} \sqrt{-t \log_e(2\pi t g_0^2)} = (N\sqrt{2e\pi})^{-1} \end{aligned} \quad (5.17)$$

$$P(g < g_0) = 1 - F_g(g_0) \quad (5.18)$$

Dividing (5.17) by (5.18) gives the ratio of the probabilities of the two modes:

$$\frac{F_g(g_0)}{1 - F_g(g_0)} = \underset{N \rightarrow \infty}{\text{Limit}} \frac{1}{N\sqrt{2e\pi} - 1} \ll 1 \quad \square \quad (5.19)$$

5.2.4 Scale Life of the GMSS of a Discontinuity

Theorem 5. The Scale-Life (SL) of a discontinuity (with a magnitude A), i.e the interval of scales ($t \in (0, SL)$) within which the discontinuity can be statistically identified by the unbalanced bimodality of the PDF of the GMSS of a discontinuity is given by:

$$SL = \frac{A^2}{2\pi\epsilon^2} \quad \text{where } \epsilon = \text{Upper Bound of error}^3 \quad (5.20)$$

Proof. From Theorem 1 for a discontinuity with a magnitude A , the GMSS will be given by:

$$L'(x; t) = \frac{Ae^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}} \quad (5.21)$$

In a manner similar to proof of Theorem 2, it can be shown that the PDF ($f_g(g)$) of the GMSS (5.21), will be defined in the interval $g \in (0, A/\sqrt{2\pi t})$ with the second mode existing at $(A/\sqrt{2\pi t})$. For this mode to be identifiable as a separate mode it should be greater than or equal to (ϵ) i.e.

$$\frac{A}{\sqrt{2\pi t}} \geq \epsilon \Rightarrow t \leq \frac{A^2}{2\pi\epsilon^2} \quad \square \quad (5.22)$$

The concept of an infinitesimal existence of a discontinuity in Scale-Space/Heat Equation is acknowledged by research community (Lindeberg; 1994; Gonzalez-Velasco; 1995; Widder; 1975), but seldom defined. Theorem 5 provides one (amongst plausibly many) rigorous definition of the life of a discontinuity, derived from (and therefore limited to) the statistics of the GMSS of a discontinuity.

Implication of (ϵ) : Any discrete application of the theoretical results would invoke the upper bound of error(ϵ), and therefore needs to be understood in the context of discretization in general, and selection of histogram bin size in specific (in images/signals). The selection of bin size inadvertently defines (ϵ) and is dependent on:

1. Physical limitations of the sensor/hardware: As an example, a camera might be able to distinguish 8, 64 or 256 intensity levels depending on 3,6 or 8 bit representation. The upper bound of error (as measured with respect to absolute ambient intensity) for a 8 bit representation will be much lower than that of 3 bit representation.

³Formal definition is provided in Section 5.4

2. Accuracy desired by the user: Even though the sensor is capable of higher precision (or lower error), an algorithm/user might require a lower precision, wherein the upper bound of error is artificially set at a higher value.

Hence the ability of algorithms to capture the unbalance and bimodality of the PDF within the Scale-Life will depend on the precision of the hardware as well as the bin size of the histogram.

5.2.5 Comments

The theoretical results of this section can be perceived to be at slight variance with the assumptions and models of a discontinuity in conventional scale-space, and the reasons for this variance will be discussed in this subsection. The widely accepted norm of *ignoring a discontinuity in Scale-Space and analyzing the underlying signal/image as if it were continuous*, can be attributed to the following factors:

1. *Requirement of Scale-Space framework*, to comply with the principles of homogeneity and isotropy, necessitates the framework to remain uncommitted to a gaussian scale. Consequently modeling a discontinuity as done in heat equation, would result in violation of the fundamental requirements of Scale-Space.
2. *Inadequate Information*: Most of the problems of Computer Vision, are related to identifying the presence/absence of a discontinuity followed by a contextual analysis of the discontinuity. In the absence of this basic information about the presence of a discontinuity, much less its properties like the magnitude and location of the discontinuity, it is difficult to model the transient presence of a discontinuity in Scale-Space.
3. *Absence of appropriate model*: Even if the location and magnitude of the discontinuity were available, a model to represent the discontinuity is difficult to prepare, because it leads to a lot of unanswered questions like how long does the discontinuity last? and, how to model the transfer from a discontinuous state to a continuous state?
4. *Mathematical simplification*: Theoretically a discontinuity disappears at an infinitesimal scale, therefore by ignoring this infinitesimal scale, a continuous model of a discontinuity can be mathematically justified.

The text (Theorems 1, 2 and 5) so far, is not meant to contradict or discredit existing conventions of Scale-Space, but to present a mathematically valid alternate representation of the Scale-Space. As an illustration of alternate (to Scale-Space) representations of a discontinuity, consider the heat equation (Gonzalez-Velasco; 1995; Widder; 1975). In heat equation a discontinuity may be explicitly modeled in following conventions:

1. As a Neumann Boundary Condition which specifies the rate of temperature (equivalent to intensity) change at the boundary.
2. As a heat source in space delimited by boundaries.

These alternate models adopted by a broader theoretical framework of heat equations, illustrate the need to represent a discontinuity in forms other than the one adopted in conventional Scale-Space theory. The PDF of the GMSS of a discontinuity (Theorems 1, 2 and 5) is an alternate representation of a discontinuity, which is unaltered in the Scale-Life (Theorem 5) including the zeroth scale. The unbalanced bimodality of the PDF can be applied homogeneously across the scales for the Scale-Life of a discontinuity, without the ambiguity of modeling or ignoring a transition from a discontinuous state to a continuous state.

This section provides a mathematical expression for the the PDF of a Gaussian Function, with a universal applicability for disciplines employing Gaussian Functions e.g. inverse problems of heat equation, chemical diffusion and Scale-Space Theory. The unbalanced bimodality of the PDF of a Gaussian Function facilitates interpretation of the second mode of the PDF as a statistical outlier. Consequently any problem of a discontinuity detection can be reformulated as a statistical problem of outlier detection. The sections hereon can be viewed as one application of the general results of this section, wherein a statistical parameter (OT) sensitive to outlier data, is used to detect a discontinuity in images.

5.3 OT: Unbalanced Histograms

In this section a general review of OT will be presented and an expression of OT for unbalanced bimodal PDF will be developed for 1-Dimensional function, with a view to accommodate the GMSS of a discontinuity. OT is statistically generated from a normalized histogram with M bins corresponding to M gray levels in an image (Figure 5.4). Each bin

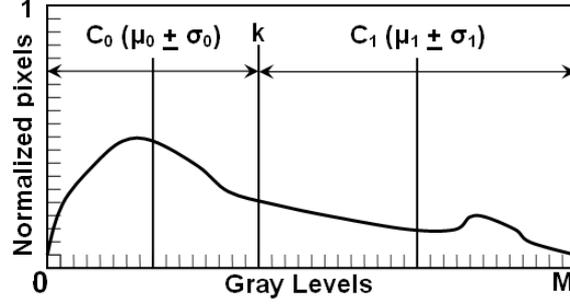


Figure 5.4: Schema: Otsu's Threshold in a histogram.

represents the percentage of the pixels in the image with corresponding gray level. This normalized histogram is bifurcated into two classes C_0 and C_1 at a hypothetical threshold (k). The hypothetical threshold (k), Means (μ_0, μ_1) and Standard Deviations (σ_0, σ_1) of two classes are shown in Fig 5.4. The maximum of Between Class Variance (BCV) determines the appropriate threshold (OT). BCV ν_B is defined by 5.23:

$$\nu_B = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \quad (5.23)$$

where

$$\omega_0 = \sum_{i=0}^k p_i, \quad \omega_1 = \sum_{i=k+1}^M p_i \quad (5.24)$$

0th order Cumulative Moment for C_0 and C_1

$$\mu_T = \sum_{i=0}^M ip_i, \quad \gamma_k = \sum_{i=0}^k ip_i \quad (5.25)$$

1st order Cumulative Moment up to M and k

$$\mu_0 = \frac{\gamma_k}{\mu_T}, \quad \mu_1 = \frac{\mu_T - \gamma_k}{\mu_T} \quad (5.26)$$

Mean Gray Levels for C_0 and C_1 respectively

$$p_i = \frac{n_i}{N_T} : \text{Normalized probability at gray level } i \quad (5.27)$$

n_i, N_T : No of pixels at gray level i , Total pixels

Theorem 6. OT (which maximizes BCV) is obtained at gray level (k^*) defined by:

$$k^* = \frac{\mu_0 + \mu_1}{2} \quad (5.28)$$

Proof. Differentiate ν_B (5.23) with respect to gray levels (k) and equate to zero. For details see (Lin; 2003). \square

The proof (Lin; 2003) is for histograms, but the results can easily be generalized to continuous PDFs. One solution of Theorem 6 is when the OT exists at the function/image mean.

Corollary 1. The maximum of BCV is obtained at the image mean (μ_T) if and only if the probabilities of the two classes are equal:

$$k^* = \mu_T \iff \omega_0 = \omega_1 = 0.5 \quad (5.29)$$

Proof (If). Substituting $\omega_0 = \omega_1 = 0.5$ in μ_T (5.25):

$$\begin{aligned} \mu_T &= \mu_0\omega_0 + \mu_1\omega_1 \\ \mu_T &= 0.5\mu_0 + 0.5\mu_1 \quad \square \\ \mu_T &= k^* \quad \text{from (5.28)} \end{aligned}$$

Proof (Only if). Equating μ_T (5.25) to k^* (5.28).

$$\begin{aligned} \frac{\mu_0 + \mu_1}{2} &= \mu_0\omega_0 + \mu_1\omega_1 \\ \text{substituting } \omega_0 &= 1 - \omega_1 \text{ from (5.24)} \quad \square \\ \omega_0 = \omega_1 &= 0.5 \end{aligned}$$

Corollary 1 allows analysis of the OT in terms of the function/image mean, without constructing a PDF/histogram. Some of the plausible distributions mentioned by (Lin; 2003) where Corollary 1 is applicable are *unimodal, perfectly balanced bimodal and unbalanced bimodal*. Corollary 1 can be tailored to a PDF containing two linearly separable classes with unbalanced probabilities (Figure 5.5). This is done with a view to develop an expression for OT, applicable to the PDF of a Gaussian function.

Let a random variable $\mathbf{Q} = \{q \in (0, q_{end})\}$ with PDF $f_q(q)$ be composed of two populations (NC, IC) which are linearly separable at ($q = \psi$) having distributions $f_{nc}(q)$

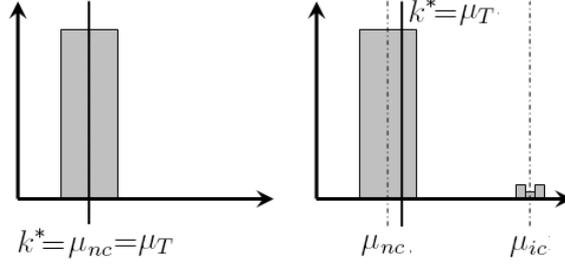


Figure 5.5: OT Schema: Symmetric Unimodal and Unbalanced Bimodal PDFs

(Left) Symmetric PDF and (Right) Unbalanced dual-class PDF (larger class is Symmetric about its mean)

and $f_{ic}(q)$ respectively, such that

$$f_q(q) = \left\{ \begin{array}{ll} m_{nc}f_{nc}(q) & \text{if } q < \psi \\ m_{ic}f_{ic}(q) & \text{else} \end{array} \right\} \quad \text{where}$$

a.	$\psi \in (0, q_{end})$ Point of linear separation of two classes
b.	$m_{ic} = \frac{q_{end} - \psi}{q_{end}}, m_{nc} = \frac{\psi}{q_{end}}$ Probabilities of the two classes
c.	μ_{ic}, μ_{nc} Averages of the two classes
d.	$m_{ic} \ll m_{nc}$ PDF is Unbalanced
e.	$\mu_{ic}m_{ic} + \mu_{nc}m_{nc} = \mu_T$ Average value of the PDF
f.	$\mu_{ic} > \mu_{nc}$ Order of classes
g.	$f_{nc}(\mu_{nc} - q) = f_{nc}(\mu_{nc} + q)$ $\forall (q < \mu_{nc}) : f_{nc}(q) \neq 0$ $f_{nc}(q)$ is symmetric about μ_{nc}

(5.30)

Equation (5.30) presents a PDF which is a super-set of the PDF of a Gaussian function with following salient features:

1. It is unbalanced.

2. It is not strictly Bimodal, but accommodates bimodal PDFs.
3. Has an additional requirement of symmetry of the first mode about its mean, which is satisfied by a Gaussian function's PDF under limiting conditions (proof follows in Corollary 2).

Theorem 7. For a PDF of the kind (5.30), the OT is given by:

$$k^* = \mu_{nc} + m_{ic}(\mu_{ic} - \mu_{nc})$$

Proof. First consider that the PDF consists of *only* NC, i.e:

$$(f_q(q) = f_{nc}) \iff (\psi = q_{end}, \mu_T = \mu_{nc} \text{ and } m_{ic} = 0)$$

The probability of two halves of the PDF separated at a mean value ($k^* = \mu_T$) is ($\omega_0 = \omega_1 = 0.5$) (Figure 5.5 Left). Therefore from Corollary 1:

$$k^* = \mu_{nc} = \mu_T \tag{5.31}$$

If class (*ic*) with a very small probability ($m_{ic} \ll m_{nc}$) is added to this distribution (Figure 5.5 Right), then the OT and the PDF mean (μ_T) will change slightly, because for a very small change in μ_T , applicability of Corollary 1 will persist (Lin; 2003):

$$\begin{aligned} k^* &= \mu_T = m_{nc}\mu_{nc} + m_{ic}\mu_{ic} \\ &= \mu_{nc} + m_{ic}(\mu_{ic} - \mu_{nc}) \end{aligned} \tag{5.32}$$

□

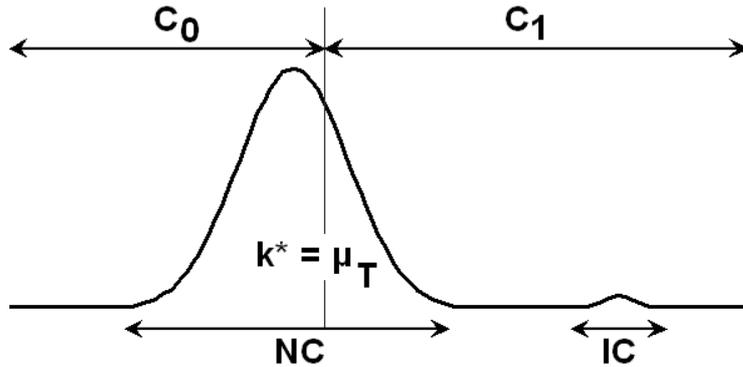


Figure 5.6: Schema: Classes (C_0 and C_1) and (NC, IC) in an unbalanced PDF.

Schematic illustration of two sets of classes as discussed so far: (C_0 and C_1) and (NC, IC) in an unbalanced PDF is shown in Fig 5.6. The First set (IC and NC), appears as two separate distributions in the PDF. The second set of classes (C_0 and C_1) exists due to the hypothetical bifurcation of the PDF at OT. Equation (5.30) imposes less rigorous conditions for application of Theorem 7 for unbalanced PDFs as compared to (Lin; 2003). The PDF need not be strictly bimodal as long as it is unbalanced and the larger class (NC) is symmetric about its mean. In the next section, Theorem 7, will be adapted specifically for the PDF of a Gaussian function. The unique inverse J Shape 5.3 of the PDF of the Gaussian function implies that the PDF is concentrated around a value of zero. Thus by providing a rigorous definition of zero ϵ and its associated spatial domain ($\delta(t)$), following simplifications of Theorem 7 can be achieved:

1. Definition of point ψ in (5.30) and consequently linear separation of the PDF of the Gaussian Function into IC and NC.
2. Expressions for average values and the probabilities of the two classes IC and NC.
3. Proof of PDF's i.e. unbalance $m_{ic} \ll m_{nc}$.
4. Elimination of need to prove symmetry of bigger class (NC). Proof is provided in the next section.

5.4 OT for a Gaussian Function

Formal definition of zero (ϵ) and the corresponding spatial domain $\delta(t)$, is obtained from Cauchy's Epsilon-Delta ($\epsilon, \delta(t)$) definition (Felscher; 2000) as applied to limit of $G(x; t)$ when $x \rightarrow \infty$.

Theorem 8. If $(\epsilon, \delta(t) \in \mathfrak{R}^+)$ represent the real and positive upper bounds of error, for the Gaussian function ($G(x; t)$) and the associated spatial variable (x) respectively, where ϵ can be made infinitesimally small and $\delta(t)$ depends continuously on ϵ and scale (t), then the limit of the Gaussian function when $x \rightarrow \infty$ is given by:

$$\begin{aligned} & \lim_{x \rightarrow \infty} G(x; t) = 0 \end{aligned} \tag{5.33}$$

Alternatively for a given ϵ and a scale (t) there exists a $\delta(t)$, such that for all x belonging to the interval $(\|\delta(t)\|, \|\infty\|)$, the Gaussian function takes a value less than ϵ :

$$\exists \delta(t) : \forall x \in (\|\infty\| > \|x\| > \|\delta(t)\|) \Rightarrow \|G(x;t)\| < \epsilon \quad (5.34)$$

Proof. It is trivial to show from the definition of $G(x;t)$ (5.1), that for a given ϵ following value of $\delta(t)$ provides the interval $(\|\delta(t)\|, \|\infty\|)$ for which $(G(x;t) < \epsilon)$:

$$\delta(t) = \sqrt{-t \log_e (2\pi t \epsilon^2)} \quad \square \quad (5.35)$$

Magnitude (A) instead of unity, will change (5.35) to:

$$\delta(t) = \sqrt{-t \log_e \left(\frac{2\pi t \epsilon^2}{A^2} \right)} \quad (5.36)$$

Graph of $\delta(t)$ from (5.36) at various ratios of (ϵ/A) for a Step Discontinuity is shown in

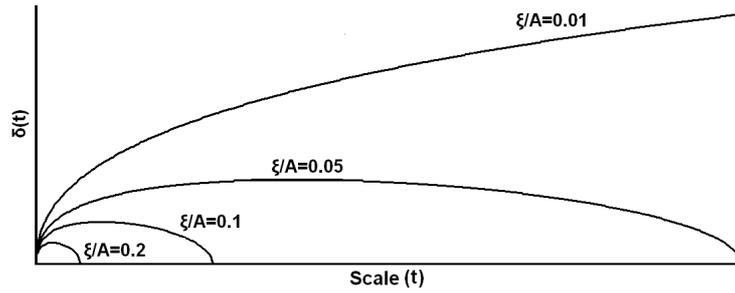


Figure 5.7: Graph of $\delta(t)$ for various ratios of (ϵ/A) for the GMSS of a Discontinuity.

Figure 5.7. The graphs show that a maximum exists in each plot. At this point IC and NC can be defined in the context of the Gaussian Function with the help of ϵ .

5.4.1 Definition and statistics: IC and NC

Definition 1. Interface Class (IC) at scale (t) is defined as the spatial domain of the Gaussian Function where the value of the Gaussian Function is greater than the upper bound of error (ϵ).

$$\begin{aligned} IC(t) &= \{x \in N : G(x;t) > \epsilon\} \Rightarrow \\ IC(t) &= \{x \in N : x < \delta(t)\} \end{aligned} \quad (5.37)$$

Definition 2. Non-interface Class (NC) at scale (t) is defined as the spatial domain of the Gaussian Function where the value of the Gaussian Function is lesser than or equal to

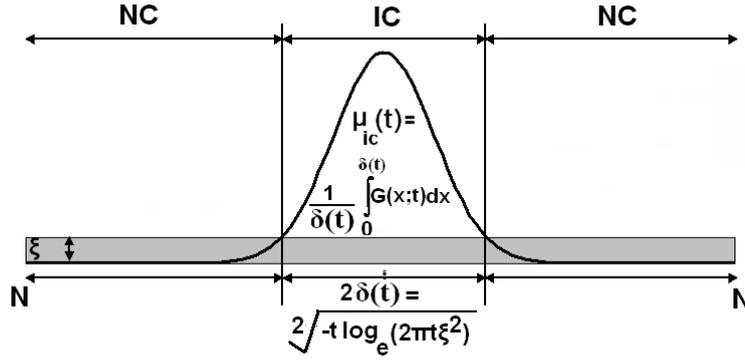


Figure 5.8: Schema: Statistical parameters of IC and NC in a Gaussian Function.

the upper bound of error (ϵ).

$$\begin{aligned}
 NC(t) &= \{x \in N : G(x;t) \leq \epsilon\} \Rightarrow \\
 NC(t) &= \{x \in N : x \geq \delta(t)\}
 \end{aligned}
 \tag{5.38}$$

Figure 5.8 depicts the two classes IC and NC in the context of a Gaussian Function. Based on the above definitions the statistics of IC and NC can be determined.

Class	Statistic
NC	Mean: $\mu_{nc}(t) = 0$ Probability: $m_{nc}(t) = \frac{N - \delta(t)}{N}$
IC	Mean: $\mu_{ic}(t) = \frac{1}{\delta(t)} \int_0^{\delta(t)} G(x;t) dx$ Probability: $m_{ic}(t) = \frac{\delta(t)}{N}$

(5.39)

5.4.2 Applicability of Theorem 7

Corollary 2. Bifurcating the Gaussian function ($G(x;t)$) or it's PDF ($f_g(g)$) at ($\psi = \epsilon$), results in two classes IC and NC such that:

1. Two classes have an unbalanced probability.

$$P[g < \epsilon] \gg P[g \geq \epsilon] \tag{5.40}$$

2. The larger class NC can be assumed to be symmetric about its mean μ_{nc}

Proof (Unbalance). Since $G(x;t)$ provides a one to one, monotone mapping from x to g and x is uniformly distributed over the interval $(0, N : N \rightarrow \infty)$, therefore the ratio of

probability of NC to IC is given by:

$$\frac{P[g < \epsilon]}{P[g \geq \epsilon]} = \frac{P[x > \delta(t)]}{P[x \leq \delta(t)]} = \frac{Lt}{N \rightarrow \infty} \frac{N - \delta(t)}{\delta(t)} \gg 1 \quad \square \quad (5.41)$$

Proof (Assumption of Symmetry of bigger class). The domain $g < \epsilon$ of the NC in the PDF is infinitesimally small, lesser than the upper bound of error and consequently immeasurable. The PDF of (NC) can be computed by applying the limits ($g \rightarrow 0^+$) to (5.5).

$$\lim_{(g \rightarrow 0^+)} f_g(g) = \lim_{(g \rightarrow 0^+)} \frac{t}{Ng\sqrt{-t \log_e(2\pi tg^2)}} = \frac{\sqrt{t}}{N} \lim_{(g \rightarrow 0^+)} \frac{1/g}{\sqrt{\log_e(2\pi tg^2)^{-1}}} \quad (5.42)$$

Equation (5.42) is of the form $(\frac{\infty}{\infty})$, therefore a simplification of (5.42) is possible by the application of *L'Hopital's Rule*, i.e. differentiating both the numerator and the denominator w.r.t (g).

$$\frac{\sqrt{t}}{N} \lim_{(g \rightarrow 0^+)} \frac{\frac{d}{dg}(\frac{1}{g})}{\frac{d}{dg}(\sqrt{\log_e(2\pi tg^2)^{-1}})} = \frac{\sqrt{t}}{N} \lim_{(g \rightarrow 0^+)} \frac{\sqrt{\log_e(2\pi tg^2)^{-1}}}{g} = \infty \quad (5.43)$$

Since the PDF ($f_g(g)$) has a value of infinity in an infinitesimal interval ($g < \epsilon$), therefore the PDF in the interval ($g < \epsilon$) can be approximated by a Dirac Delta. The Dirac Delta is the limiting case of the Symmetric Gaussian Function (with zero standard deviation), therefore assumption of symmetry of the PDF of the NC is justified. \square

Corollary 2 implies that Theorem 7 is applicable for a Gaussian function where the IC and NC are separated at a Gaussian value ($G(x; t) = \epsilon$) or at equivalent space coordinate ($x = \delta(t)$).

Theorem 9. The OT, for the GMSS of a step function is given by:

$$k^*(t) = \frac{1}{N} Erf[\delta(t)] \quad \text{where} \quad Erf[\delta(t)] = \frac{1}{\sqrt{\pi}} \int_0^{\delta(t)} e^{-p^2} dp \quad (5.44)$$

Proof. Replacing ($\mu_{nc}(t) = 0$) from (5.39) in Theorem 7:

$$k^*(t) = m_{ic}(t)\mu_{ic}(t) \quad (5.45)$$

Substituting $\mu_{ic}(t)$ and $(m_{ic}(t))$ from (5.39) and $(G(x;t))$ from (5.1) into (5.45) and solving:

$$k^*(t) = \frac{\delta(t)}{N\delta(t)} \int_0^{\delta(t)} \frac{e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t}} dx = \frac{1}{N} \text{Erf}[\delta(t)] \quad \square$$

Plots of OT and $\delta(t)$ from Theorem 9 at various ratios of (ϵ/A) are shown in Figure 5.9. Both the plots $(k^*(t), \delta(t))$ contain a maximum, which can also be verified by differentiating (5.44) and (5.36) w.r.t scale (t) .

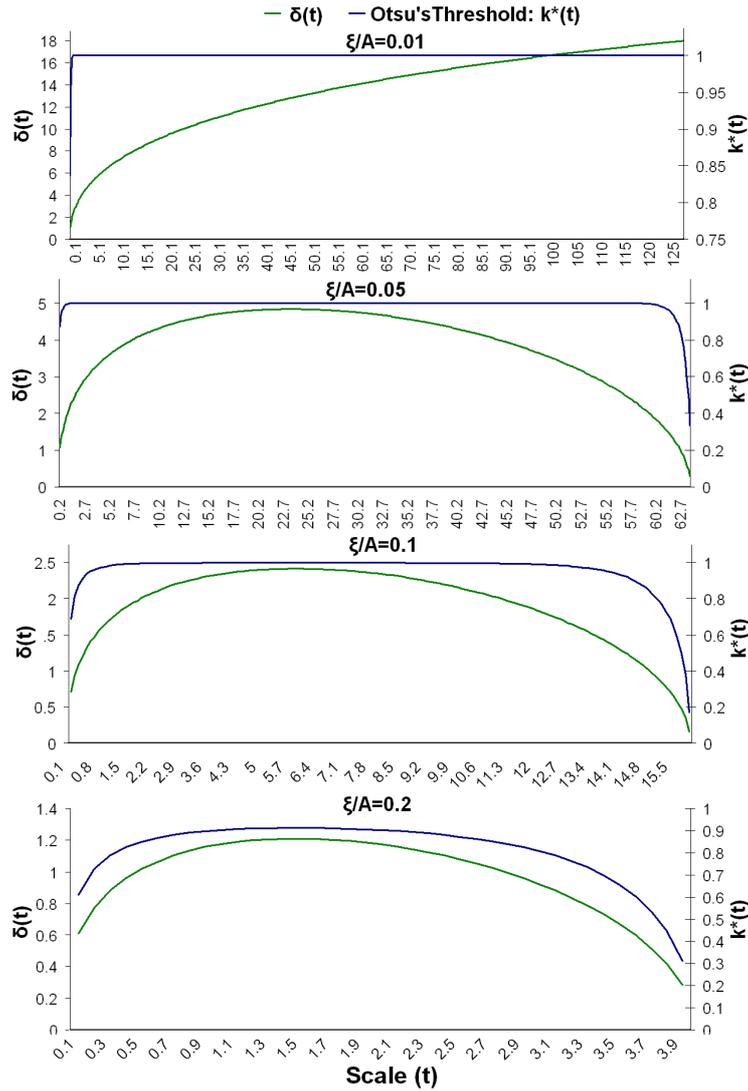


Figure 5.9: Graphs of OT $[k^*(t)]$ and $[\delta(t)]$ against Scale at various ratios of ϵ/A .

5.5 OT for continuous functions

Theorem 10. The OT for the GMSS of a continuous signal is monotonically decreasing i.e.

$$k^*(t) > k^*(t + \Delta t) \quad \forall \quad \Delta t > 0 \quad (5.46)$$

Proof. Due to the Central Limit Theorem, the PDF of the GMSS of a continuous function can be approximated by a normal (and consequently symmetric) distribution. Therefore from Corollary 1:

$$k^*(t) = \mu_T(t) \quad \text{and} \quad k^*(t + \Delta t) = \mu_T(t + \Delta t) \quad (5.47)$$

For continuous functions, the Maximum Principle is valid (Gonzalez-Velasco; 1995; Widder; 1975; Lindeberg; 1994). Due to the Maximum principle the GMSS of the continuous function (and its mean) will be monotonically decreasing with increase of scale ((Babaud et al.; 1986; Lindeberg; 1994, 1998)), i.e. :

$$\mu_T(t) > \mu_T(t + \Delta t) \quad \forall \quad \Delta t > 0 \quad (5.48)$$

Combining (5.47) and (5.48)

$$k^*(t) > k^*(t + \Delta t) \quad \forall \quad \Delta t > 0 \quad \square$$

Comparison of Theorem 9 with Theorem 10 reveals the contrasting patterns for OT when traced against scale:

1. Transient Increase: When a discontinuity exists.
2. Monotone Decrease: When a function is continuous.

5.6 Conclusion

This chapter shows that the PDF of the derivative of a discontinuity is unbalanced and bimodal in Scale-Space. By taking the derivative of functions, the discontinuities can be formulated as outliers with higher average value and low probability. Since OT is a statistical parameter sensitive to the outliers (smaller mode in bimodal distribution) in a

given data set, therefore it can detect and locate discontinuities. It is likely that many statistical parameters sensitive to outliers would exhibit similar response to discontinuities in scale-space. The research reported in this thesis is considered adequate for the purpose of current application. This chapter has been restricted to theory. The applications of these theorems will be evolved in the next chapter. Heuristics and algorithm will be presented to detect foreground in general and obstacle in water navigation in particular.

Chapter 6

Locating Objects in Water: Application

In this chapter, the theoretical results from non-discrete and 1-Dimensional functions of the previous chapter will be applied to the analysis of discrete 2-Dimensional images. Specifically the interface between the entities will be detected. The ideology underpinning this section is that the edges in images can be broadly classified as:

1. *Boundary-Edges*: These edges correspond to interface between two entities of a 3D physical world, when projected onto the 2D image surface as a consequence of the process of image capture (e.g. an edge located at the interface of foreground-background). Since Boundary-Edges exist at the interface of heterogeneous surfaces or/and processes, the discrete gradient (Sobel) of Boundary-Edges is computed from dissimilar neighborhoods leading to a high intensity gradient and also spatial Scarcity, rendering Boundary-Edges similar to IC of the 1D functions.
2. *Non-Boundary-Edges*: In contrast, the Non-Boundary-Edges owe their discrete gradient computation to homogeneous neighborhood, resulting in low gradient and high spatial probability which is similar to NC of the 1D functions.

This apparent similarity, has empirical support from contemporary literature (Rosin; 2001; Bhanu and Faugeras; 1982; Lin; 2003; Medina Carnicer and Madrid Cuevas; 2008) wherein existence of unbalanced histograms for the derivative of images have been reported. The similarity of the histograms and the associated statistics (probability and mean) of IC and NC with Boundary-Edges and Non-Boundary edges respectively, allows

for the development of a heuristic to extrapolate Theorem 9 and Theorem 10 for detection of Boundary-Edges. The problem of Boundary-Edges identification can be subdivided into:

1. Finding the optimum Scale
2. Finding the optimum Threshold at the Scale
3. Locating the interface, as the discontinuity travels in scale-space (Lindeberg; 1994).

To locate the Boundary-Edges, the following heuristic has been evolved which identifies both the scale and threshold appropriate to the interface using OT.

6.1 Heuristic

In the presence of an inflection in the plot of OT (calculated for the GMSS of an image) against incremental scale, it can be assumed that a discontinuity due to a foreground background interface exists in the image. This discontinuity can be identified by thresholding the GMSS of the image at the scale and OT corresponding to the upper point of inflection in the trace of OT.

Justification: There are three aspects of the Heuristic i.e. presence of discontinuity, appropriate scale and appropriate threshold, which need to be justified individually:

1. *Presence of a discontinuity:* The presence of an inflection only in the presence of discontinuity has been shown via Theorem 9 and Theorem 10 for functions with and without a discontinuity respectively.
2. *Scale at upper point of inflection:* There is no universally agreed definition of appropriate scale; hence the justification of appropriate scale is qualitative rather than mathematical. The upper point of inflection is the scale appropriate for the discontinuity owing to following reasons:
 - (a) Lower point of inflection wrongly classifies the pixels as belonging to IC (Figure 6.1). Hence for all scales lower than the lower point of inflection, false classification as IC is a strong possibility.
 - (b) Experimentally and theoretically (most of the graphs of Figure 5.9), it has been observed that often from the lower to the upper point of inflection there is only

small difference of scales. Hence attempting to locate the scale between the upper and lower points of inflection is mostly futile.

- (c) Scales greater than the scale at the upper point of inflection can give comparable results for the discontinuity identification as the scale at upper point of inflection, but at some scale greater than the upper point of inflection the IC will cease to exist. In the absence of a priori information of this scale where IC ceases to exist, using a scale greater than the one identified by the upper point of inflection, runs the risk of attempting to locate IC at a scale at which the IC does not exist. Therefore the upper point of inflection is the best scale for IC detection.

3. *OT at upper point of inflection as the threshold:*

- (a) OT at the upper point of inflection has been chosen as the threshold as it corresponds to the scale appropriate to the discontinuity.
- (b) Minimum False positives: The upper point of inflection represents the highest threshold intensity (utilizing OT in Scale-Space). Since the average value of the IC is greater than the rest of the image, therefore the highest threshold results in lowest false classification of the pixels as IC.

The application of the heuristics is demonstrated in Figure 6.1, wherein the Boundary-Edges have been located by thresholding at scale and OT corresponding to upper point of inflection¹. Figure 6.11d depicts a synthetic image comprising of background only. A foreground is added to the texture of Figure 6.11d as shown in Figure 6.1c resulting in Figure 6.1b. OT plotted against incremental scale for the GMSS of Figures 6.11d and 6.1b results in OT graphs shown in Figure 6.1d, wherein the plot corresponding to the background only image has a monotonic decay in contrast to the image with a foreground which shows an inflection. Thresholding the image with the foreground (Figure 6.1b) at scale and OT corresponding to the upper point of inflection identifies the foreground-background interface.

¹ The detection of the appropriate scale is not for a general discontinuity but conditional to the presence of a specific discontinuity. The discontinuity should be due to interface and therefore have an unbalanced histogram similar to the PDF in(5.30). The scale *cannot* be identified using the method presented here for discontinuities which are not due to interfaces.

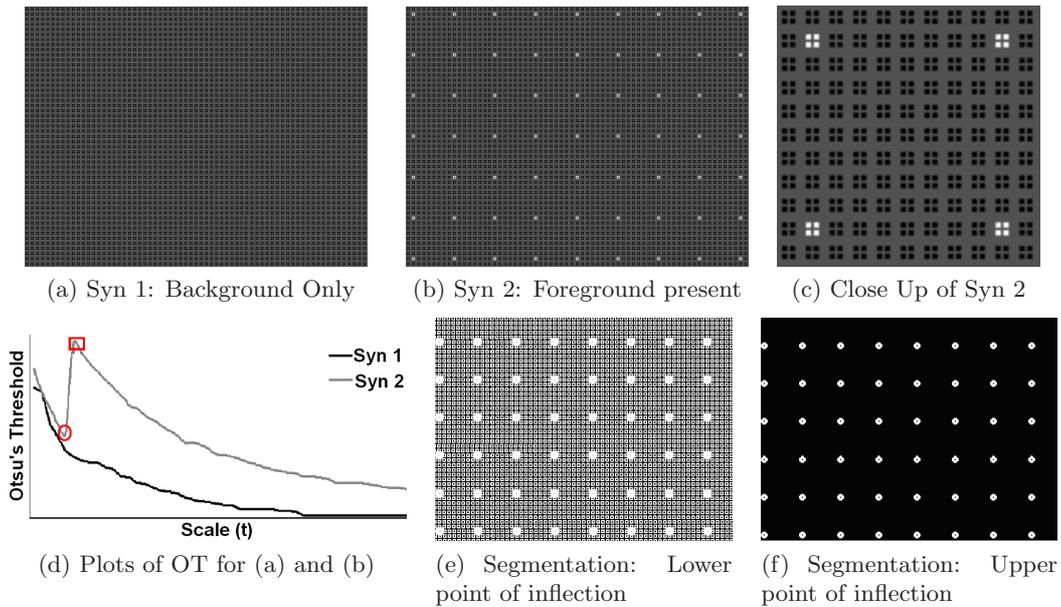


Figure 6.1: Illustration of Heuristic presented on Synthetic Images.

(d) OT plots indicate a monotonic decay and an inflection in the absence and presence of a foreground (and consequent Boundary-Edge) respectively. (f) Upper point of inflection detects Boundary-Edges.

6.2 Algorithm

Based on the Heuristic of the OT for discontinuity detection, a simple algorithm comprising of following steps can locate Boundary-Edges in images (Walia and Jarvis; 2009):

Algorithm 3: Simultaneous detection of scale, discontinuity and threshold in images

Compute the Sobel derivative of the input image;

while *Not end of Scale Range* **do**

 Convolve the Sobel derivative of the image, with a Gaussian Kernel of current scale;

 Compute histogram;

 Compute and record OT at current scale;

if *Increment of OT in plot against scale exists* **then**

 Identify the (scale, OT) pair at which the OT attains a maximum;

 Convolve the Sobel derivative of the input image with the scale identified;

 Threshold at the OT identified ;

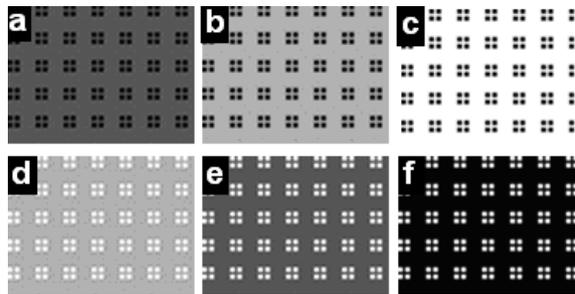
else

 Output: No Discontinuity;

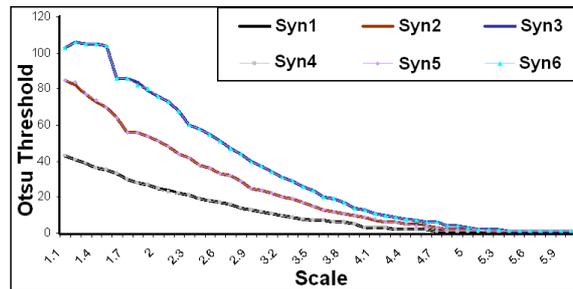
6.3 Algorithm Scope

In this section, the applicability of the algorithm will be demonstrated on a variety of synthetic images generated to test the scope of the algorithm. Various parameters of the background and foreground will be varied and OT calculated for the GMSS of the synthetic images. The OT will then be plotted against scale to locate the inflection points. The results indicate algorithm's suitability for a wide range of image processing problems.

6.3.1 Background Intensity variation



(a) a to f: Syn1 to Syn6



(b) Threshold Graphs

Image	Syn1	Syn2	Syn3	Syn4	Syn5	Syn6
Foreground Intensity	0	0	0	255	255	255
Background Intensity	85	170	255	170	85	0

Figure 6.2: Algorithm Scope : Background Intensity variation.

Tiles in Figure 6.2a are from images (Syn1 to Syn6) of dimensions 640X480 with the intensities as listed in the table in Figure 6.2. The gradient magnitude between foreground and background of three sets of complimentary images [(Syn1, Syn4);(Syn2, Syn5);(Syn3, Syn6)] is identical although the gradient direction of FG and BG is reverse. The plots of OT for the GMSS of the complimentary sets of synthetic images are shown in Figure 6.2b. Each complimentary pair has identical graph of OT which leads to following two deductions:

1. OT decay is exponential and proportional to the magnitude of gradient.
2. Decay is independent of gradient direction.

6.3.2 Background Frequency variation

Tiles and TG from synthetic images Syn10, Syn11, and Syn12 are shown in Figure 6.3. In each image pattern the intensity is same but the frequency of background is varied. Threshold graphs show that the OT decays with minor variations in decay rate due to background frequency.

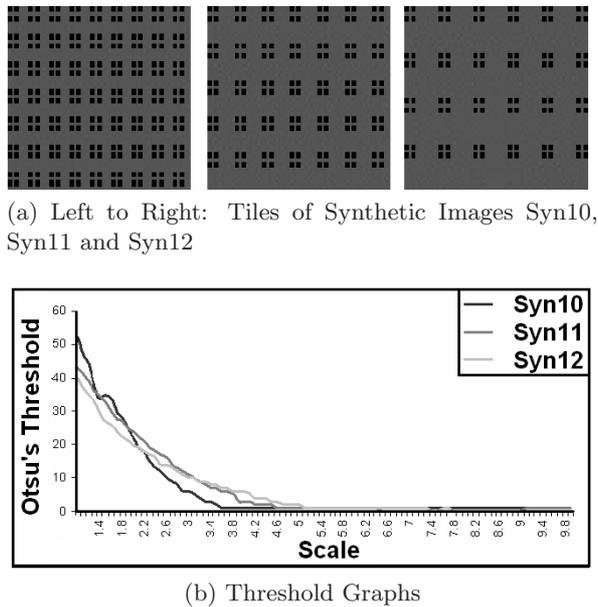
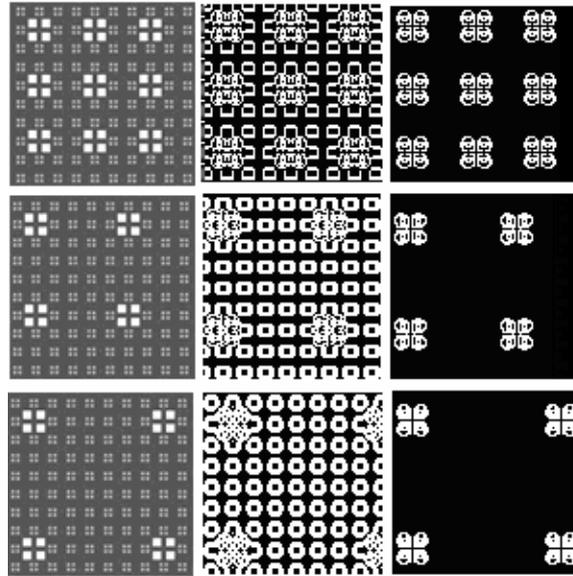


Figure 6.3: Algorithm Scope: Background Frequency variation.

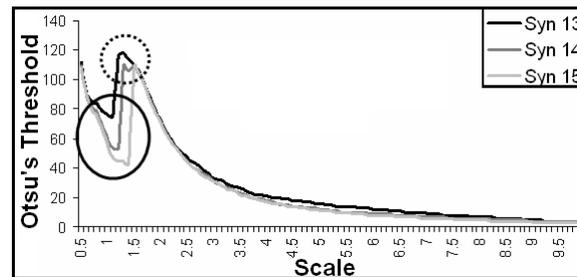
6.3.3 Foreground Frequency variation

In synthetic Images Syn13, Syn14, Syn15 the background pattern comprises of square clusters of intensity 170 against a backdrop of intensity 85. The foreground comprises of cluster of four squares of intensity 255 and a size of 4 pixels each. These foreground clusters are separated by distance of 8, 16 and 24 for Syn13, Syn14 and Syn15 respectively as illustrated by tiles in first column of 6.4a. The graphs of OT are shown in Figure 6.4b.

The results of thresholding the image with the Otsu's threshold at the lower and upper points of inflection are shown in second and third column of Figure 6.4. It is observed that a scale appropriate to the discontinuity due to the interface of background-foreground in



(a) Top, Middle and Bottom Row: Images Syn 13, Syn 14 and Syn 15. First, Second and Third Column: Tiles of Synthetic Image, Thresholded Images at Lower Inflection point and Upper Inflection point.



(b) Threshold Graphs

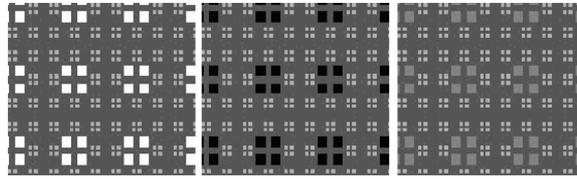
Figure 6.4: Algorithm Scope: Foreground Frequency variation.

an image is given by the upper point of inflection. Column 3 of Figure 6.4 shows that the images thresholded at upper points of inflection, identify give the important structure in the image with respect to the background. A slight shift in the local maxima is observed with the decrease of the foreground frequency, which can be attributed to attrition in the percentage of pixels due to the background-foreground interface. Nevertheless the ability of the graphs of Otsu's Threshold to adapt to the internal structure of the image background is illustrated.

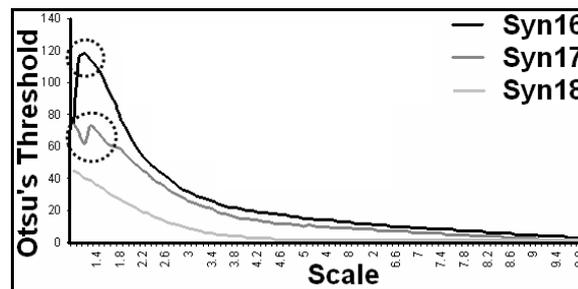
6.3.4 Foreground Intensity variation

Images Syn16, Syn17 and Syn18 shown in Figure 6.5 have the same pattern as the image Syn13, however the foreground intensity is set to 255, 0 and 127 respectively. Since the background intensity varies between 85 and 170, three scenarios are:

1. Syn16: Foreground intensity $>$ Background range.
2. Syn17: Foreground intensity $<$ Background range.
3. Syn18: Foreground intensity is within the Background range.



(a) Left to Right: Tiles of Synthetic Images Syn16, Syn17 and Syn18



(b) Threshold Graphs

Figure 6.5: Algorithm Scope: Foreground Intensity variation.

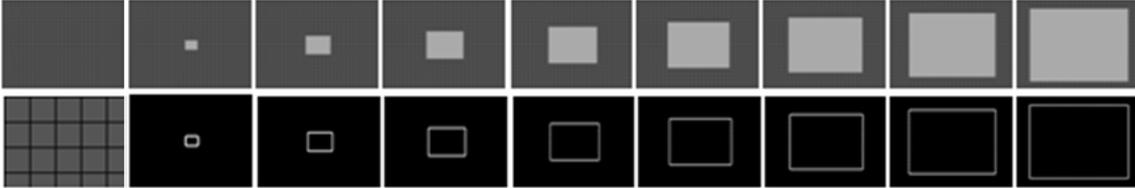
The Otsu's Threshold Graphs are depicted in Figure 6.5b. Where the foreground intensity is beyond the background intensity range, inflection exists in the graph. However when the foreground intensity range is confined to the range exhibited by the background (Syn18), the inflection does not occur; hence the foreground-background discontinuities are not detected. Absence of inflection can be attributed to similar magnitude of the interface derivatives as the non-interface derivatives. It is a unique case and does not occur frequently, especially in natural textures e.g. detecting a grass hopper in grass. Natural textures like pebbles, sand, water, hay, grass, vegetation, wood, and even man made textures like rugs, carpets usually have background chromaticity which when projected onto gray scale has a narrow range. Consequently the foreground object's intensity exists outside the background image intensity range.

6.3.5 Foreground Size variation

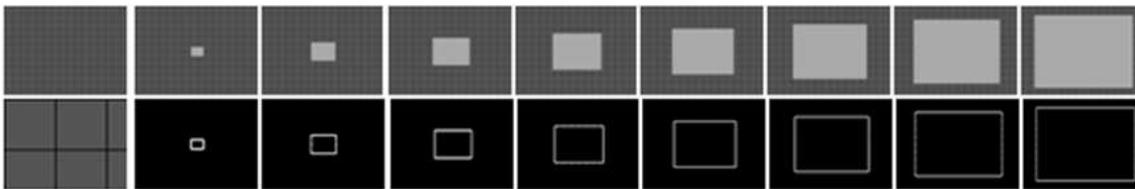
Three sets of synthetic images and the plots of OT are shown in Figure 6.6. Set 1 corresponds to Figure 6.6a, (Syn19 to Syn27) and Figure 6.6d; Set 2 corresponds to Figure 6.6b (Syn28 to Syn36) and Figure 6.6e; Set 3 corresponds to Figure 6.6c, (Syn37 to Syn45) and



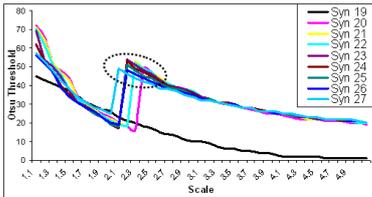
(a) Top: Syn19 to Syn27. Bottom: Results. Bottom Row 1st Column: Background Pattern for Top Row



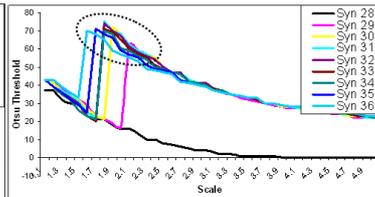
(b) Top: Syn28 to Syn36. Bottom: Results. Bottom Row 1st Column: Background Pattern for Top Row



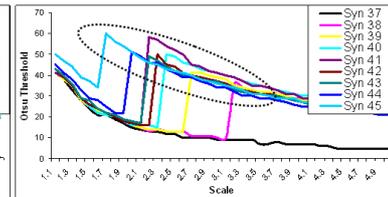
(c) Top: Syn37 to Syn45. Bottom: Results. Bottom Row 1st Column: Background Pattern for Top Row



(d) OT plots: Syn19 to Syn27



(e) OT plots: Syn28 to Syn36



(f) OT plots: Syn37 to Syn45

Figure 6.6: Algorithm Scope: Background Frequency and foreground size variation.

Figure 6.6f. Each set of Synthetic images contains a rectangular foreground of varying sizes having area 0.01, 0.04, 0.09, 0.16, 0.25, 0.36, 0.49, 0.64 times that of the synthetic image. Reduced images are shown to optimize space usage; however the background patterns are shown in the first column of second rows in (Figures 6.6a, 6.6b, 6.6c) for sets 1, 2 and 3 respectively. The thresholded GMSS at the points of inflection of OT for all the 3 sets are shown in column 2 of (Figures 6.6a, 6.6b, 6.6c) respectively. An inspection of the OT graphs for for the sets under consideration leads to the following observations:

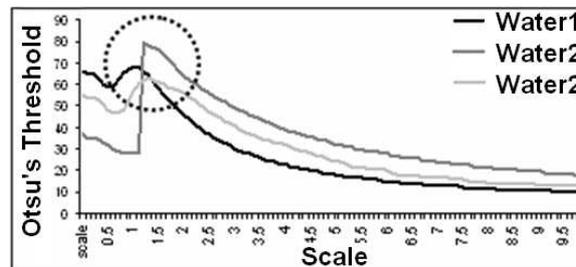
1. In 3 cases of background only i.e. Syn19, Syn28 and Syn37 represented by black lines in (Figures 6.6d,6.6e,6.6f) there is only decay of the OT without any inflection. Remaining OT plots of synthetic images however undergo an inflection, even on introduction of foreground of an area which is 1 percent of total image area.

2. Intra set points of inflection for all 3 sets reveal a decreasing trend (dotted ellipses) with increase of foreground area.
3. Inter set points of inflection exhibit an increasing range of inflection values with respect to scale as shown by the dotted ellipses in (Figures 6.6d,6.6e,6.6f), highlighting increased sensitivity to the foreground object size. However the foreground object's size increment and intensity is identical across the three sets, hence the increase range can be attributed to the decrease in density (frequency) of background pattern.

6.3.6 Foreground against Background of water



(a) Row 1: Water1, Water2 and Water3. Row 2: Thresholded Images (at upper point of inflection).



(b) Threshold Graphs

Figure 6.7: Algorithm Scope: Water Scene Analysis.

Water scenes comprise of a host of image signal distortion factors which have already been discussed in previous chapters. In spite of inherent variability in water scenes, inflection of OT is fairly robust in detecting background-foreground interface. Figure 6.7 shows the results of applying OT to scenes with a water background ². Some noise persists in thresholded images. Noise in results can be eliminated by one or combination of following:

²Figure 6.11 (Calm Conditions), figure 6.12 (Moderate Conditions) and figure 6.13 (Rough Conditions) illustrate the scope of algorithm in different water conditions. Images have been obtained from various sources (including shore) and replicate scenarios expected to be encountered by a water robot.

1. Utilizing a scale higher than that of the inflection, as the important edges will persevere across scales.
2. Hysteresis thresholding across scales.
3. Sequentially thresholding the interface edges identified by Otsu's threshold.

6.4 Algorithm Performance

It is difficult to provide a comprehensive comparison between the Algorithm 3 and other contemporary research because usage of *statistics* of Scale-Space representation of derivative of image/functions to identify discontinuity is a novel proposition. OT is a well known method for segmenting images and therefore provides a good benchmark to compare the performance of the algorithm presented here. A comparison of the segmentations based on Algorithm 3 with OT is shown in Figure 6.10. Algorithm 3 had scale increments of 0.1, and histograms comprised of 255 bins. Figure 6.10 shows images having well defined IC. When OT is applied at zeroth scale, the probability $m_{ic}(0)$ of IC is very small, therefore the OT instead of segmenting IC from NC, segments the NC at approximately the mean of NC ($\mu_{nc}(0)$). In comparison tracing the OT in the GMSS of the images, results in identifying both the scale and threshold appropriate for identifying IC.

The images set used in Figure 6.10 originate from eclectic sources, without a ground truth so a simple measure was chosen to compare Algorithm 3 with Otsu's Algorithm (Otsu; 1979). The thresholded results of the two algorithms were stored as binary images (Figure 6.8b and Figure 6.8c). The difference in the number of positives (Figure 6.8c) between the Otsu's algorithm and Algorithm 3 expressed as a percentage of the total pixels in the image determined the improvement in the boundary classification. The average improvement in classification was computed for the dataset of images shown in Figure 6.10. Algorithm 3 had on an average 31.3% better classification of the foreground-background boundary owing to a reduction in false positives as compared to Otsu's Algorithm.

6.5 Scale Detection

Before discussing scale detection, it is important to highlight that scale detection is incidental to the main research. Lindeberg (Lindeberg; 1994) describes the problem of

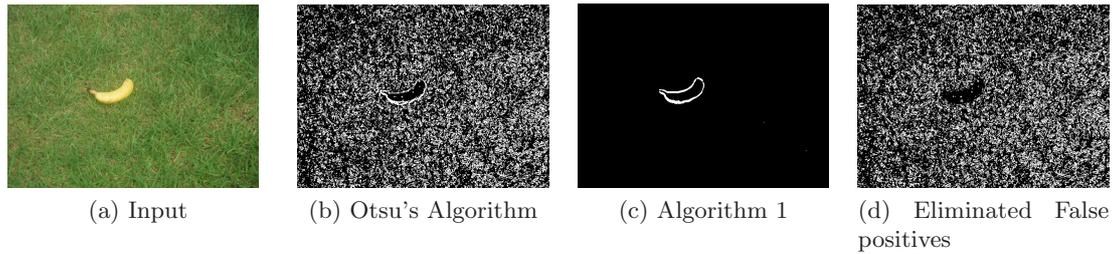


Figure 6.8: Illustration : Algorithm comparison.

Number of eliminated false positives ($d=b-c$) expressed as a percentage of the total pixels provides the improvement in boundary/discontinuity detection.

establishing an appropriate scale in the absence of prior information as intractable. Although models emulating mammalian vision take cognizance of the need to establish an appropriate scale they do not exclusively address this need, but rather avoid the issue by using large scales (Malik and Perorna; 1990) or contextual scales (Ren et al.; 2006). Lindeberg (Lindeberg; 1994, 1998) addresses scale identification in two different ways:

1. A 4-Dimensional structure composed of a scale-space-blob is generated from images. This structure is tracked over multi scales with the hypothesis that prominent structures persist across scales. Blobs are derived at different scales using monotonic gradients from local extrema and are then analyzed for their effective scale range using blob-descriptors like volume, contrast and area, and blob-events like annihilation, creation, merging and splitting.
2. The principle of non enhancement of extrema as applicable to Gaussian differential operators is applied. A normalized (with scale and consequently scale invariant) Gaussian derivative operator is applied to signals/images and then maximized over scales. The scale corresponding to the maxima is heuristically hypothesized to coincide with the characteristic length of corresponding structure in image data. For a rigorous mathematical treatment, chapter 13 of (Lindeberg; 1994) refers.

These approaches proposed by Lindeberg have three drawbacks:

1. *Usage of local properties* in the initial identification of entity which, in the case of blobs, is seeding originating from a blob event and, in the case of Gaussian derivative operator, is the edge maxima. Both these entities are dependent on local spatial properties like the intensity and nearness to another entity which often give rise to

spurious structures. In the case of the Gaussian derivative operator, all the edges (including noise) are guaranteed a maximum (Lindeberg; 1994) over some scale; hence the problem of appropriate scale identification still persists. To address this problem a ranking mechanism grades the entities based on the properties of entities like contrast, life, spatial spread, volume etc across the scales. The ranking mechanism is unreliable as the local properties like the geometry of entity will influence both the Scale-Space evolution as well as the properties over scale. For example response to a Gaussian derivative of a curved edge will vary from that of a straight edge and, without apriori information on the kind of edge being detected, the response will be unreliable and in fact can often lead to a choice of improper derivative function.

2. *Restricted spatial scope of local extrema.* Figure 6.9 Top shows a 1 dimensional signal (termed original) comprising of local maxima in the vicinity of global minima of the signal. The original signal is convolved with 3 Gaussian kernels as shown in Figure 6.9 Bottom. The standard deviations of the 3 kernels coincide with the spatial spread of local extrema (Gauss 1), neighborhood of local extrema (Gauss 2) and the global neighborhood of local extrema (Gauss 3). The results of convolving with Gauss 1, Gauss 2 and Gauss 3 are also shown in Figure 6.9 Top by Result 1, Result 2 and Result 3, respectively. The evolution of local maxima is shown inside the dotted rectangle of Figure 6.9. This evolution is consistent with the proof of violation of non-enhancement of extrema in the presence of a discontinuity (Gonzalez-Velasco; 1995) and shows that:

- (a) Local extrema violates the principle of non-enhancement of extrema, as the intensity of local extrema is first reduced and then increased with increasing scale. This violation is not due to scale increment but due to consideration of local extrema in isolation from global neighborhood.
- (b) Evolution of local maxima can have valid but conflicting classification depending on the scale. E.g. based on Result 1 the local extrema can be hypothesized as local maxima and based on Result 2 as global minima.

3. *Minimal representation of image* can lead to a plausibly flawed hypothesis. The methods (Lindberg, 1994) to identify appropriate scale omit the evolution of non-extrema neighborhood with scale. This neighborhood is quantitatively significant

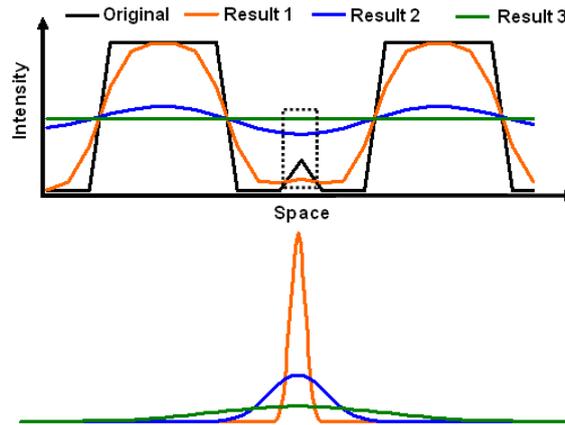


Figure 6.9: Experiment: Violation of maximum principle.

(Top) 1D signal and its Scale-Space representation at various scales. Signal has a local maxima in the vicinity of global minima. **(Bottom)** Scale of Gaussian kernels corresponds to spatial spreads of local extrema (orange), local neighborhood (blue) and global neighborhood (green) respectively.

as locating even the first cut (zeroth scale) extrema involves discarding 8 neighboring pixels. Increasing the scale also increases the discarded neighborhood, due to non-enhancement property. Hence the important features are generated from a hypothesis based on a minimal representation of the image.

The scale detection method proposed herein, disassociates itself from the above mentioned drawbacks of Lindeberg's techniques because of :

1. *Global Statistics*: Algorithm 3 utilizes image histogram, and is therefore immune to problems arising out of local properties, minimal representation and limited spatial scope.
2. *Non-assumption of continuity*: Algorithm 3, does not assume continuity of the underlying function/image. On the contrary GMSS of a discontinuity is utilized as a feature whose statistics determine the scale.

Owing to theoretical differences, the current and Lindeberg's methods have contextual applications and any one is not necessarily superior than the other.

6.6 Conclusion

In this chapter, theoretical framework of previous chapter has been adapted, to solve following problems of image processing viz:

1. Detecting the presence/absence of a discontinuity due to interface of foreground/background.
2. Identifying a scale and threshold appropriate to the discontinuity. The scale and threshold relevant to the discontinuity are a function of both:
 - (a) The magnitude of the discontinuity.
 - (b) The upper bound of error.

The heuristic has been extensively tested on synthetic images to demonstrate the broad range of foreground and background variations that can be accommodated by the heuristic/algorithm. The application under consideration is that of obstacle/foreground detection against a background of water, and is therefore adequately addressed by the algorithm presented in the chapter.

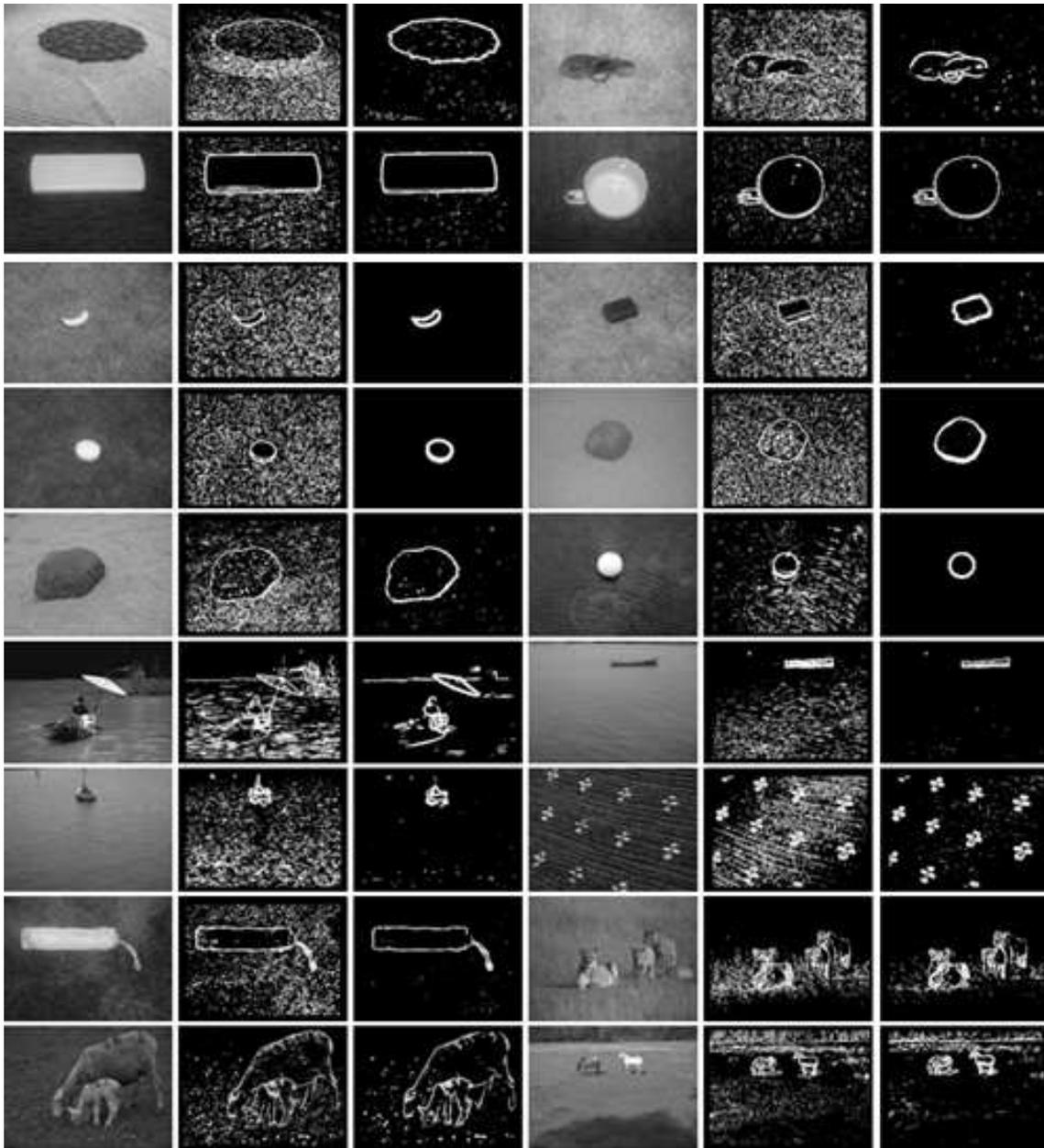


Figure 6.10: Segmentation comparison.

Column 1 and 4: Images with obvious boundaries. (some from Berkeley Dataset (Martin et al.; 2001)). *Column 2 and 5:* Thresholded by OT. *Column 3 and 6:* Thresholded at Scale and OT corresponding to inflection of OT.



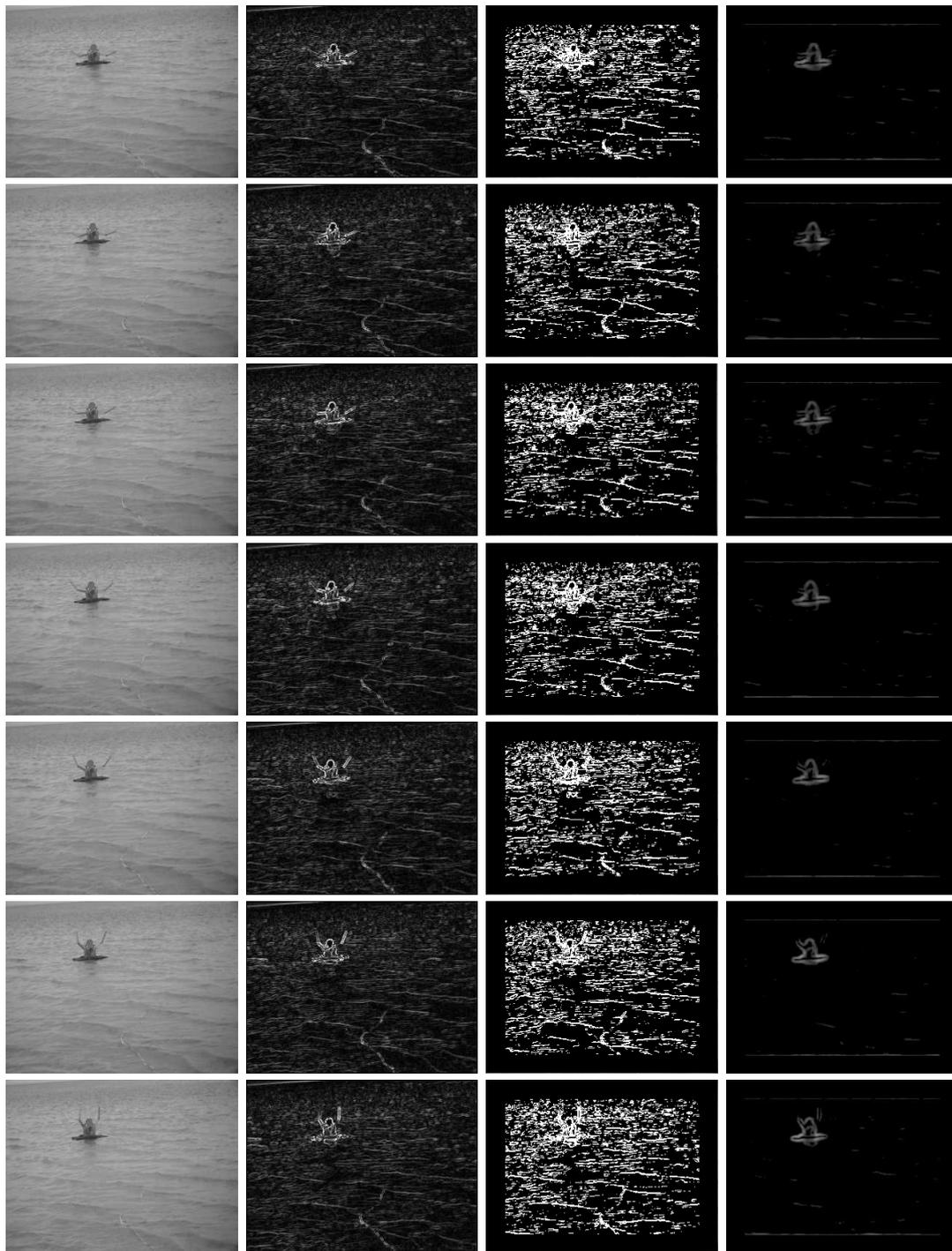
(a) Image

(b) Sobel Edges

(c) Otsu's Method

(d) Proposed Algorithm

Figure 6.11: Algorithm in calm water-conditions.



(a) Image (b) Sobel Edges (c) Otsu's Method (d) Proposed Algorithm

Figure 6.12: Algorithm in moderate water-conditions.



(a) Image (b) Sobel Edges (c) Otsu's Method (d) Proposed Algorithm

Figure 6.13: Algorithm in rough water-conditions.

Images sourced and used under permission from (www.innimage.com; 2012)

Chapter 7

Conclusion and Future Research

7.1 Conclusion

This thesis has been written with a view to understand water scenes from the perspective of water robotics. The differences in navigation (objectives and environment) of a water robot as compared to land robot have been discussed (Chapter 1). These differences along with the visual representation of water surface (Chapter 2) have lead to the identification of major components in a water scene, which are necessary (either by presence or absence) for water-robot navigation. The inadequacy or irrelevance of existing image processing literature has been discussed (Chapter 3) to establish the need for current research. The major components analyzed are sky (with clouds), horizon (Chapter 4) and obstacles (Chapters 5 and 6) in water. The contribution of this thesis extends beyond the scope of application for which it was conceived.

In pursuit of scene analysis, new theories as shown in abstract (Figure 1) with mathematical proofs and experimentation have been presented with regard to:

- *Pseudo Spectra Images*: Theory developed herein is rather unique because it:
 - Reverses the photometric process. The process of photometric capture integrates the visible spectrum, whereas PSIs differentiates an integrated RGB triad into individual frequencies.
 - Applies a new philosophy to aberration (clouds in this thesis) removal in images. By identifying a frequency, for which the gray scale intensity response of the

dominant feature (sky) is similar to that of aberration (clouds) in the feature, a homogeneous gray scale image can be generated from color images.

- *Horizon detection:* Enclosed ellipses are presented as a more robust, faster and simpler alternative to Hough Transforms.
- *Discontinuity detection:* Statistical properties of the Scale-Space representation of the derivative of a discontinuity have been evolved and utilized for detection.

7.2 Future Research

The most obvious application of the research would be the integration of the three components for deployment on a water robot. Some applications of water robot have already been discussed in the First chapter. The analysis of three components of water scene, provides opportunities for, further research in both application as well as theory as discussed below:

7.2.1 Discontinuity Detection

Discontinuity detection is the most promising candidate for extending research in both theory and application, because discontinuity is studied and applied across disciplines as discussed below:

Theory

- Scale-Space analysis of statistical parameters which:
 - Are sensitive to outlier data.
 - Can be expressed as a function of the first two moments of the PDF.
- 3-Dimensional analysis of the PDF (Theorem 2) in Scale-Space, especially with regard to:
 - Linear and non linear sampling of scales.
 - Normalized (with respect to scale) representation of PDF (Theorem 2).
- Generalization of the Step-discontinuity, to any discontinuity using Fourier transforms.

- Imposition of the CDF (Theorem 2) on histograms of images with known unbalanced histograms with a view to threshold images. The proposed process is expected to reverse the well known image processing technique of histogram equalization.

Applications

The algorithm and the underlying theory of discontinuity detection would be under appreciated without a reference to the following plausible applications in different disciplines:

1. Thermodynamics: Inverse problems of Heat Equation.
2. Chemistry: Inverse problems of Diffusion Equation.
3. Finance: Jump Events in markets.
4. Geology: Analyzing discontinuities in data.
5. Manufacturing: Locating thermal cracks.
6. Equipment maintenance: Tomography in aircrafts, ships, vehicles, pressure vessels, railway tracks, power lines, locating welds etc
7. Signal processing: Smart Card readers, passive sensors, radio trans-receivers.
8. Computer Architecture : Bit recognition
9. Marine Surveillance: Cameras, SONARS, RADARS, echo sounders etc.
10. Medical Images
11. Optical instruments automation (self-focus): Camera, Microscope, Bionic eye and Telescope
12. Pattern recognition.
13. Image Processing
 - (a) Pre and post processing.
 - (b) Detecting discontinuity in time (videos).
 - (c) Correlating dynamism of textures in space and time.

- (d) Parameter comparison. E.g. in image segmentation Hue, saturation and Value can be utilized. Comparative utility can be evaluated by tracing OT in Scale-Space for all three parameters.

7.2.2 Pseudo Spectra Images

The scope of Pseudo Spectra Images (PSI) is not as extensive as that of discontinuity detection, because the theory developed is limited by the tri-stimulus response of three cones of human eye to wavelengths in visible spectrum (Wyszecki and Stiles; 2000). Consequently applications and theory will also be limited to wavelengths in the visible (to human) spectrum. The research perspectives of PSIs originate from the novelty of the theory and has potential in image processing as discussed below:

Theory

- Automating appropriate frequency determination. The automating process would involve defining, testing and validating mathematical criteria for locating the appropriate frequency.
- Evolving multi frequency descriptors / patterns for application specific features. E.g. Predominantly chromatic features like water, grass, gravel, vegetation, roads, skin, fruits etc can be described by a set of say 8 frequencies generated from PSI. The descriptors can be binary and similar to (Ojala et al.; 2002), but the binary pattern would comprise of spectral frequencies instead of neighboring pixel intensities.
- Extrapolating intensity response beyond the frequency range of the visible spectrum, by mathematical manipulation of intensities generated by PSIs within the visible spectrum.

Applications

- *PSI*: Image pre-processing for eliminating chromatic aberrations in natural textures like grass, water, skin, fruit/vegetable etc, with a strong predisposition to certain colors.

- *Sky Detection*: Horizon detection, video processing (Zafarifar and Peter; 2006), scene labeling, image retrieval, land surveying, and robotics i.e. in autonomous watercrafts/aircrafts (Cornall and Egan; 2004), (Chahl et al.; 2003) and (Thakoor et al.; 2002).

7.2.3 Horizon Detection

Horizon detection as presented in this dissertation, is a contextual application of an available mathematical technique (enclosed ellipses). Consequently further research options are not as numerous as those of the previous two components of the scene. Limited future research can be undertaken in :

- Replacing Hough transform with enclosed ellipses, where edges are expected to have slight distortion.
- Extending the scope of enclosed ellipses to occluded horizons.
- Estimating the tilt/roll of ships/aircrafts from the angle of minor axis of the enclosed ellipse of the horizon. This can potentially feed into a control system for stabilizing the (air/water) craft.

Appendix A

Watercraft Design

This appendix has been included to record the change in the PhD topic and deliverables. The research was initiated as a vision guided water robotics project, but had to be altered to water scene analysis owing to following reasons (Figure A.1):

- Occupational Health and Safety considerations: Monash university does not have a risk assessment strategy for experiments conducted on water. Consequently the author was advised to find alternate methodology for completing the research.
- Lack of resources: Monash University was unable/unwilling to allocate adequate equipment, manpower, time and finances required for the project.

As a consequence of the institutional inadequacies and prohibitions, experiments involving deployment of watercraft in water had to be abandoned. Regardless of outcome, a substantial amount of time, effort and research was expended on the design and construction of watercraft with no academic publication. This appendix documents the research and work involved in the watercraft construction. The watercraft design was subjected to following factors:

1. *Marine Environment:* The components were either marine grade or protected against marine environment by aluminum, plastic and rubber encasing.
2. *Transport and Assembly:* The watercraft had to be dismantled for transportation which required design to allow for ease of assembly/disassembly. This lead to problems of alignment, interfaces (mechanical, electrical and electro-mechanical) and power distribution, and the design had to be inherently flexible to enable on-site assembly and alignment.

 **MONASH University**

Section C: To be completed by the head of department/head of school/director of centre or nominated representative

Please return the completed form by Friday 31 July 2009

C.1 Please comment on candidate's progress and any causes for concern. Please indicate what arrangements have been made to address any concerns raised by the candidate and/or supervisor:

The complaints about unfair OHS are unjustified. The Risk Assessment should be discussed with the supervisor. Watercraft do produce a risk of drowning, especially when operated by 1-person. I wonder whether sufficient results could be obtained using a camera and a pier, (and the correct time of day for the sun). Yes, there are always problems with experiments, but students should try to think around them, if possible. The supervisor should also try to modify the goals of the thesis if the experiments are unreasonable.

Figure A.1: Remarks of the Head of Department

3. *Launch and Retrieval:* The launch and retrieval mechanisms had to be designed to accommodate the heavy loads (of batteries and motors) concentrated in the aft section of the watercraft. This required augmenting the (rubber) base of the watercraft with a much harder and stiffer marine ply and usage of inverted skateboards for bearing the loads and reducing the friction.
4. *Remote and Autonomous Operations:* The watercraft was designed for both autonomous and remote operations. This was achieved by the use of an umbilical cord comprising of
 - Rope: To tether the watercraft.
 - Power Cable: To remotely operate the emergency cut-off switch which controlled the power to the propellers.
 - LAN Cable: To connect the on-board laptop, with the remote laptop. The remote control was achieved via Virtual Network Connection (VNC), running on both the laptops which replicated the display and functions (of mouse and keyboard) of on-board laptop on the remote laptop.

As a consequence of the constraints of the watercraft design factors, the following systems were installed on-board with varying degrees of fabrication:

1. Propulsion and Steering.
2. Controls.
3. Power distribution.
4. Vision Guidance: For Obstacle detection.
5. Navigation: Compass and Camera integration.

Of the above mentioned systems, vision guidance has already been discussed in various chapters. A brief overview of propulsion, steering, controls and power distribution for sensors would be provided in subsequent subsections.

A.1 Propulsion and Steering

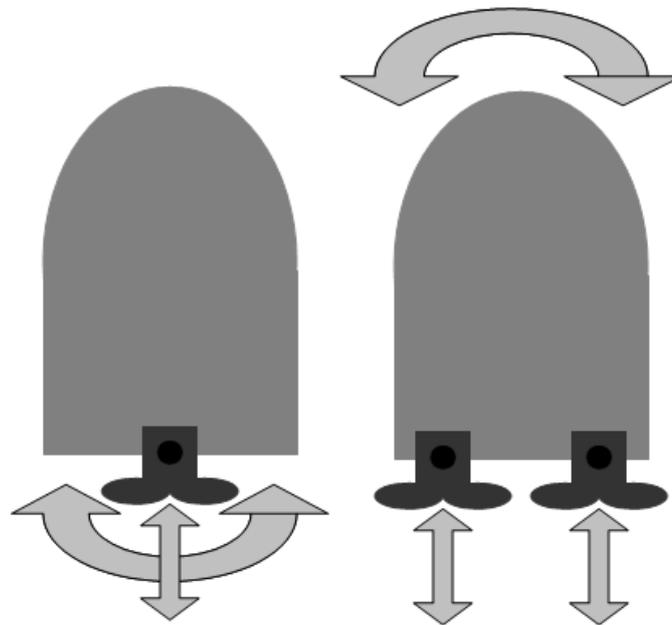
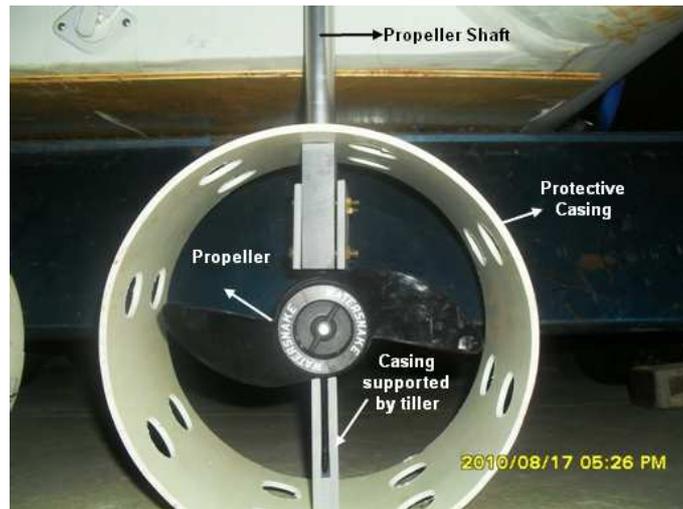
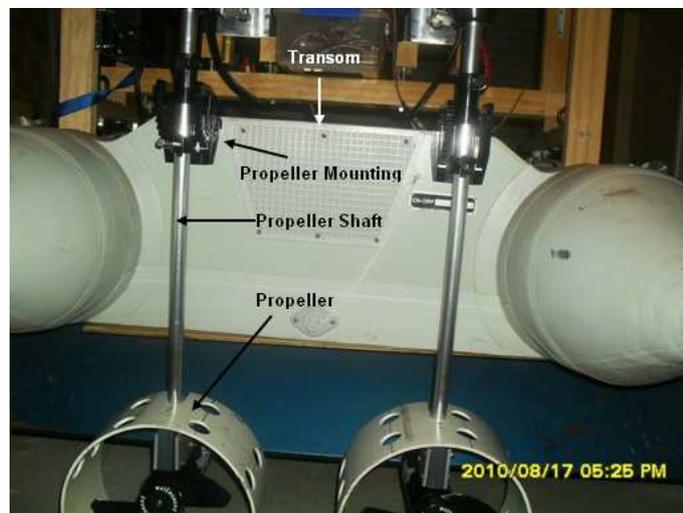


Figure A.2: Steering Comparison (Left) Conventional. (Right) Differential.

The watercraft used was an inflatable boat about 2.5 meters in length with buoyancy of 150 Kgs. It was propelled by two FWT30 (model No) Snake (Manufacturer) propellers with a maximum thrust of approx 14 Kgs. Each propeller is powered by two 12V DC motors. Each propeller is controlled by a rotary switch which can be engaged in 5 forward and 2 astern speeds. Since a wide range of speeds was available it was decided to steer the watercraft by differential thrust from the two propellers. The remainder of the appendix covers design aspects of various systems of the watercraft.



(a) Close up of propeller and its protective casing



(b) Propellers mounted on the transom.

Figure A.3: Propulsion and Steering Components

Conventionally, the propellers (Figure A.2) are designed to be used in isolation and provide dual usages of propulsion and steerage. The propellers are designed to provide steerage by their usage as manual tillers. The propeller is mounted in the center of the transom of the boat and is free to rotate about their longitudinal axis (Propeller shaft -Figure A.2). The operator can rotate the propeller to steer the boat in the requisite direction. To generate a differential thrust on the watercraft, the two propellers were rigidly mounted on the transom. Rigid mounting (against free to rotate) of the propellers enabled generation of controlled differential thrust to steer the watercraft. A schematic illustration of the conventional vs modified steerage mechanism is shown in Figure A.2. The speed of both the propellers was individually controlled to enable a differential steering.

The lake chosen for experiments was shallow, therefore to avoid damage to the propellers a protective casing was mounted around the propellers. The casing was made out of two PVC pipes and mounted on the propeller motor. To avoid slipping or turning the casing was securely bolted to the tiller and the motor portion of the propeller as shown in Figure A.3a The control system to achieve individual control of propellers is discussed in next section.

A.2 Controls

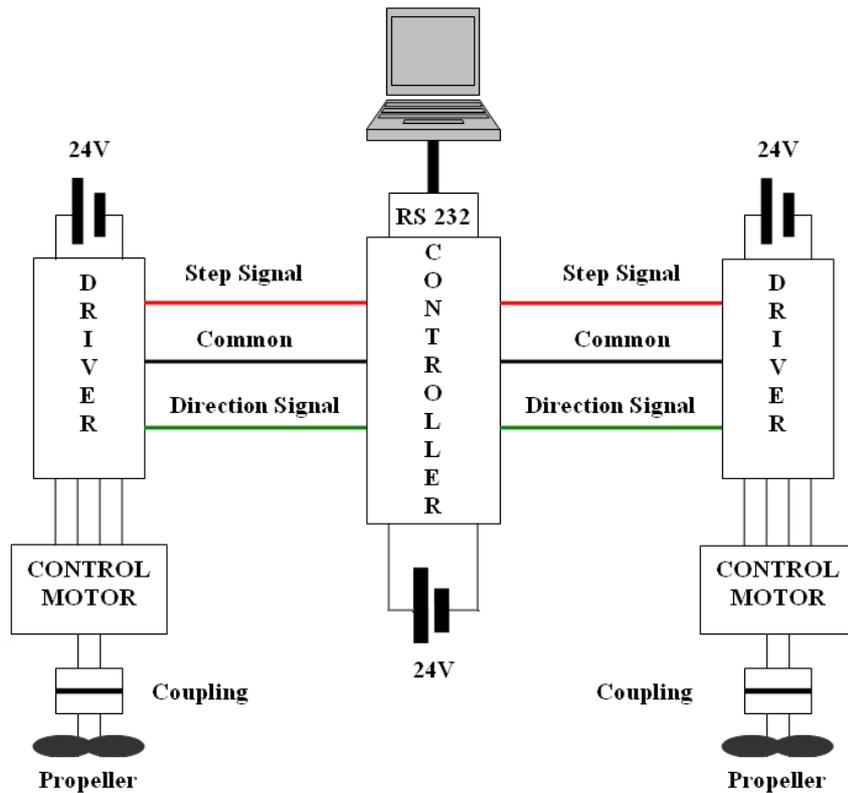
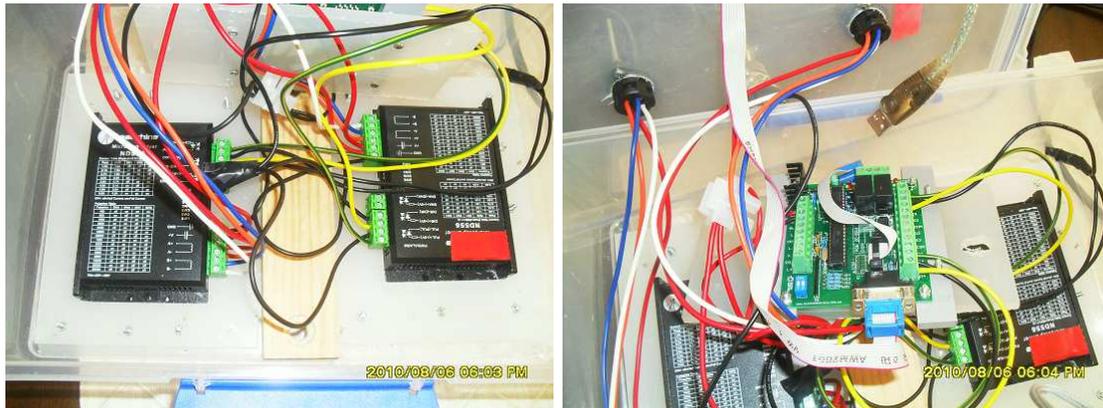


Figure A.4: Control System: Schematic illustration.

The watercraft needs to be controlled by computer to integrate results obtained from the vision component of research. A schematic illustration of the control system is depicted in Figures A.4 and A.5. To enable computer based controls two RS-232 controlled stepper motors were used. These stepper motors were mechanically coupled to the rotary switch (speed control) of the propellers. Photograph (A.6) illustrates the mechanical coupling.

The control motors are 4 coil stepper motors. The motors have the capability of positioning the shaft with a precision ranging from $1/200$ to $1/4000$ of 360° while drawing current from 1 Ampere to 5.5 Amperes. To optimize the battery life and still retain



(a) Photograph: Drivers.

(b) Photograph: Controller.

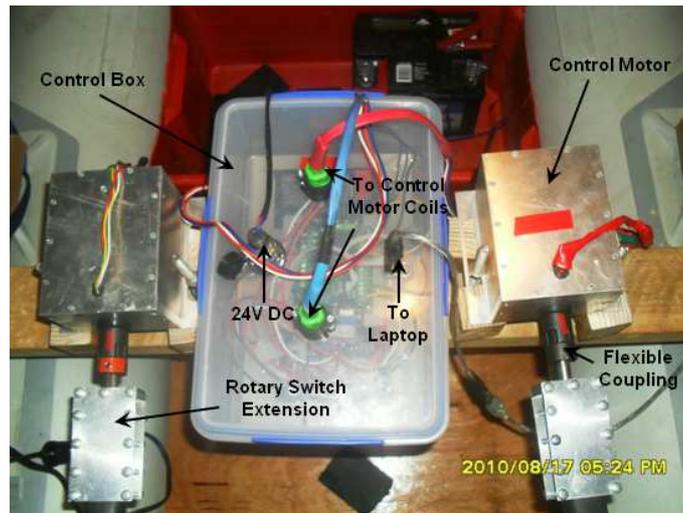
Figure A.5: Control system : photographs

adequate turning torque a current of 2.2 Ampere was selected after experimenting with the propeller coupled with the motors. It was also experimentally found that the precision of $1/400$ of 360° of stepper motor was adequate for positioning the rotary switch of the propellers to all the positions. By controlling the shaft rotation of the stepper motor, the propellers could be engaged in any desired mode.

The control of rotary switch of the propeller by the stepper motor presented a problem. The rotary switch has certain characteristics:

- Rotary switch is spring loaded.
- Engages in forward and astern modes only for certain angular range of the switch.
- This angular range is different for both the propellers.
- This angular range varies in clockwise and anticlockwise direction.

Variations exist in the angular range of rotary switch for which propeller engages. These variations are due to manufacturing, assembly and mounting (on boat). The spring loaded mechanism compensates for these variations, when operated manually, as the spring brings the rotary switch to the point of engagement. The problem arises in using stepper motor, as the holding torque of the motor is greater than the spring tension, which renders the spring mechanism ineffective. The problem was solved by locating the midpoints of spring engagement range for the rotary switches. There are angular locations between two switch positions for which the switch is disengaged. The range of engagement of the switch is not same for clockwise and anticlockwise movement of the switch. To compensate for this mismatch, the port and starboard propellers were not interchangeably



(a)



(b)

Figure A.6: Coupling

to retain consistency between control-motor and propeller. In addition the overlap (between clockwise and anticlockwise) range of rotary switch's engagement for each position was established. A schematic diagram for one position of rotary switch is illustrated in Figure A.7. The midpoint of this overlap range was calculated which provided the angular location of rotary switch's engagement in both the directions.

The angular positioning of the stepper motor shaft is controlled by SSM- which is a driver manufactured by a local supplier M/S Ocean controls. The driver operates on a 20-50V DC supply and provides requisite electrical impulses for the shaft to turn to the angular position required. The driver in turn is controlled by a controller which provides actuation signals to the driver. The actuation signals between the controller and the driver consists of 5V Step and Direction Signals. The Controller was controlled by a series of

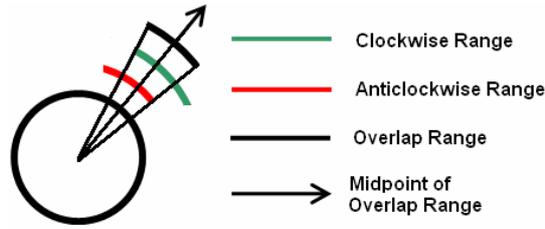


Figure A.7: Locating midpoint of overlap range of rotary switch positions.

predefined text commands which rotated the motor shaft to appropriate angular position so as to engage the propeller speed switch in required speed and mode (astern/ahead). The communication between the laptop and controller was achieved through serial communication via USB port (Figure A.8 and A.9).

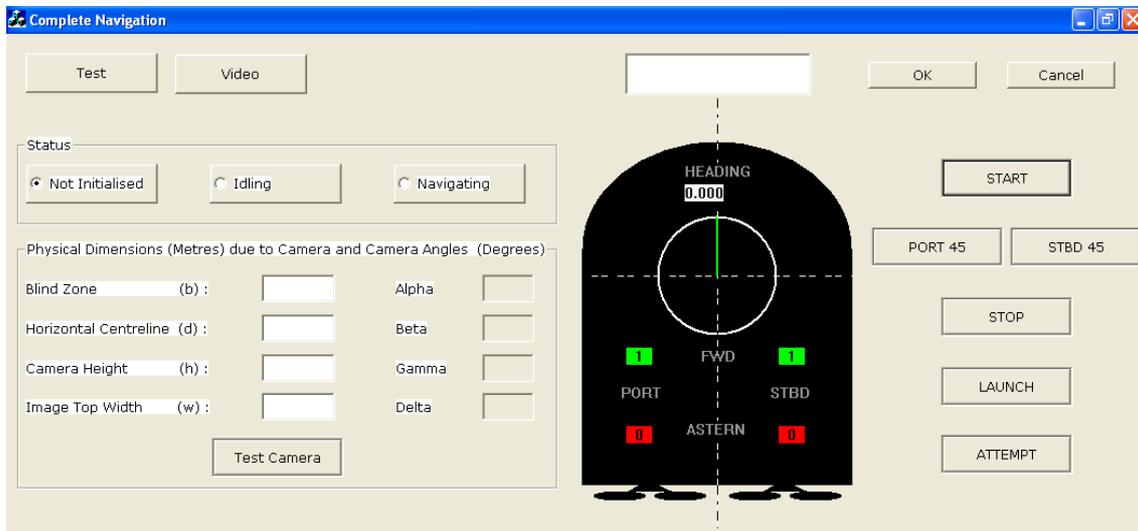


Figure A.8: Graphic User Interface (GUI) for the software to control the boat.

Figure A.8 also provides flavors of

1. Relative localization of the obstacle with respect to the camera by using 3 dimensional projective trigonometry of image capture.
2. Current heading of the boat as acquired from the fluxgate compass.

A.3 Power Generation and distribution

Power generation was by means of 12 Volt car batteries, 4 on-board and 1 ashore. Propeller, camera, compass required 12V, whereas the control motors, and their controllers and drivers required a 24V power source. This was achieved by arranging 2 batteries in series and another two in parallel (Figure A.10). Two circuits one of 12V and one of 24V

```

Complete - Microsoft Visual C++ - [CompleteDlg.cpp *]
File Edit View Insert Project Build Tools Window Help
CCompleteDlg [All class members] Write_Serial_Port
bool CCompleteDlg::Write_Serial_Port(LPTSTR port_message)
{
    CString str, str1;
    HANDLE h = CreateFile("COM2", GENERIC_READ|GENERIC_WRITE, 0, NULL, OPEN_EXISTING, 0, NULL);
    if (h==INVALID_HANDLE_VALUE){PrintError("Failed to open port"); }
    else{
        //set timeouts
        COMMTIMEOUTS cto = {10,10,10,10,10};
        if (!SetCommTimeouts(h,&cto)) {PrintError("SetCommTimeouts failed");return 0;}
        // set DCB
        memset(&dcb,0,sizeof(dcb));
        dcb.DCBlength = sizeof(dcb);
        dcb.BaudRate = 9600;
        /*dcb.fBinary = 1;
        dcb.fDtrControl = DTR_CONTROL_ENABLE;
        dcb.fRtsControl = RTS_CONTROL_ENABLE;
        dcb.fOutxCtsFlow = 1;
        dcb.fRtsControl= DTR_CONTROL_HANDSHAKE;*/
        dcb.Parity = NOPARITY;
        dcb.StopBits = ONESTOPBIT;
        dcb.ByteSize = 8;
        if(!SetCommState(h,&dcb)) {PrintError("E014_SetCommState failed"); return 0;}
        char buf[30];
        char temp_buf[30];
        str=port_message;
        DWORD read = 30;
        DWORD write= 30;
        str="";
        for (DWORD j=0;j<30;j++) temp_buf[j] = port_message[j];
        WriteFile(h,temp_buf,write,&write,NULL);
        ReadFile(h,buf,sizeof(buf),&read,NULL);
        str1.Format("%.*s", sizeof(buf),buf);
        CEdit* handle_setBaudRate = (CEdit*)GetDlgItem(IDC_BAUD_RATE);
        handle_setBaudRate->SetWindowText(str1);
    }
    CloseHandle(h);
    return 1;
}
Ready Ln 894, Col 49 REC COL OVR READ

```

Figure A.9: Screen Capture of code for serial communication with control motors.

were thus formulated which supplied power to all the on-board equipment. Remote safety cut off was achieved by positioning a 12V relay, in both the circuits, which was controlled by a battery located on shore.

A.4 Conclusion

This appendix has briefly covered various design aspects and related issues involved in construction of watercraft and subsequent amendments to the PhD goals. The text presented herein has an additional objective of documenting a methodology of designing a remotely-controlled/autonomous watercraft from Commercial Off the shelf (COTS) equipment.

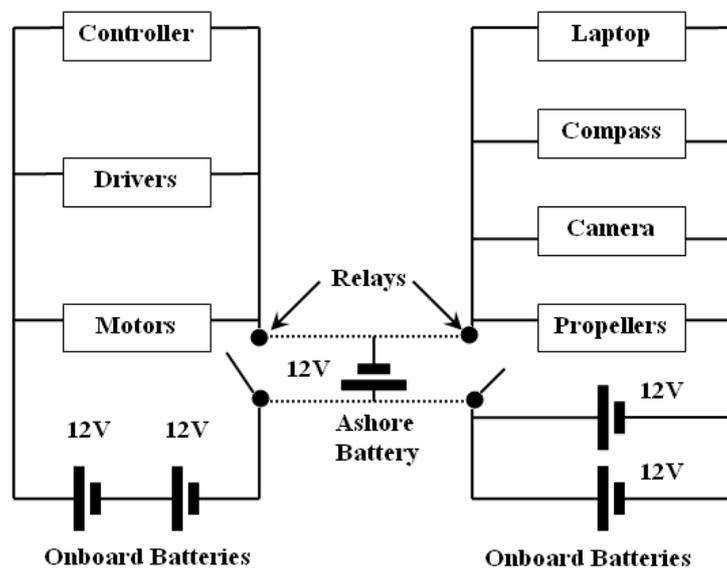


Figure A.10: Schematic Illustration: Power Distribution.

Vita

Publications arising from this thesis include:

- Walia, R. and Jarvis, Ray. (2009)**, Structure,Scale-Space and Decay of Otsu's Threshold in Images for Foreground/Background Discrimination. In *The Proc. Int'l Conf on Computer Vision Theory and Applications*, 2009 (Vol 2), pp 120-128.
- Walia, R. and Jarvis, Ray. (2010)**, Horizon detection from pseudo spectra images of water scenes. In *2010 IEEE Conference on Cybernetics and Intelligent Systems (CIS)*, 2010 , pp 138-144.
- Walia, R. Suter, D. and Jarvis, Ray. (2012)**, Discontinuity Detection From Inflection of Otsus Threshold in the Derivative Of Scale-Space. In *Machine Vision (INTECH Publications)*, 2012, pp 205-226.

References

- Abraham, B., Camps, O. I. and Sznaier, M. (2005). Dynamic texture with fourier descriptors, *Proc Of the 4th. International Workshop On Texture Analysis and Synthesis.*, IEEE, pp. 53–58.
- Aeronautics, N. and Administration., S. (2010). Chapter 11: On the level from radiation to scientific imagery.
- Aksoy, S. and Haralick, R. M. (1998). Content-based image database retrieval using variances of gray level spatial dependencies, *MINAR '98: Proceedings of the IAPR International Workshop on Multimedia Information Analysis and Retrieval*, Springer-Verlag, London, UK, pp. 3–19.
- Alonzo, K. (1996). Introduction to mobile robots. (last accessed February 17, 2010).
URL: <http://www.ctan.org/tex-archive/macros/latex/contrib/harvard/>
- Babaud, J., Witkin, A. P., Baudin, M. and Duda, R. O. (1986). Uniqueness of the gaussian kernel for scale-space filtering, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(1): 26–33.
- Bar-Joseph, Z., El-Yaniv, R., Lischinski, D. and Werman, M. (2001). Texture mixing and texture movie synthesis using statistical learning, *IEEE Transactions on Visualization and Computer Graphics* **7**(2): 120–135.
- Bhanu, B. and Faugeras, O. D. (1982). Segmentation of images having unimodal distributions, *Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-* **4**(4): 408–419.
- Blair, K. (1965). *Wind Waves their generation and propagation on the ocean surface.*, Prentice Hall Inc.
- Bradski, G. (2000). The OpenCV Library, *Dr. Dobb's Journal of Software Tools* .

- Braun, C. L. and Smirnov, S. N. (1993). Why is water blue?, *Journal of Chemical Education* **70**(8): 612.
URL: <http://pubs.acs.org/doi/abs/10.1021/ed070p612>
- Carroll, K.P. and McClaran, S. N. E., Barnett, D., Friesen, D. and Williams, G. (1992). Auv path planning: an a* approach to path planning with consideration of variable vehicle speeds and multiple, overlapping, time-dependent exclusion zones, *Proc of the 1992 Symposium on Autonomous Underwater Vehicle Technology*, pp. 79–84.
- Celebi, E. and Alpkocak, A. (2000). Clustering of texture features for content-based image retrieval, *ADVIS '00: Proceedings of the First International Conference on Advances in Information Systems*, Springer-Verlag, London, UK, pp. 216–225.
- Chahl, J., Thakoor, S., Le Bouffant, N., Stange, G., Srinivasan, M. V., Hine, B. and Zornetzer, S. (2003). Bioinspired engineering of exploration systems: A horizon sensor/attitude reference system based on the dragonfly ocelli for mars exploration applications, *Journal of Robotic Systems* **20**: 35–42.
- Chang, F., Chen, C.-J. and Lu, C.-J. (2004). A linear-time component-labeling algorithm using contour tracing technique, *Comput. Vis. Image Underst.* **93**(2): 206–220.
- Chang, Z.-H., Tang, Z.-D., gao Cai, H., cheng Shi, X. and qian Bian, X. (2005). GA path planning for AUV to avoid moving obstacles based on forward looking sonar, *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, Vol. 3, pp. 1498–1502.
- Chaudhuri, B. B. and Samanta, G. P. (1991). Elliptic fit of objects in two and three dimensions by moment of inertia optimization, *Pattern Recogn. Lett.* **12**(1): 1–7.
- Cornall, T. D. and Egan, G. K. (2004). Measuring horizon angle from video on a small unmanned air vehicle, *2nd International Conference on Autonomous Robots and Agents*.
- Da Silva, I., Gomide, F. and Do Amaral, W. (1998). Navigation of mobile robots using fuzzy logic controllers, *Proceedings of the 15th International Workshop on Advanced Motion Control*, pp. 346–349.
- Devaux, J. C., Gouton, P. and Truchetet, F. (2001). Karhunen-lo[e-grave]ve transform applied to region-based segmentation of color aerial images, *Optical Engineering*

40(7): 1302–1308.

URL: <http://link.aip.org/link/?JOE/40/1302/1>

Doretto, G., Chiuso, A., Wu, Y. N. and Soatto, S. (2003). Dynamic textures, *International Journal of Computer Vision* **51**: 91–109. 10.1023/A:1021669406132.

URL: <http://dx.doi.org/10.1023/A:1021669406132>

Duits, R., Felsberg, M., Florack, L. and Platel, B. (2003). α scale spaces on a bounded domain, *Lecture Notes in Computer Science* **2695**: 494–510.

Eisenberger, I. (1964). Genesis of bimodal distributions, *Technometrics* **6**(4): pp. 357–363.

Felscher, W. (2000). Bolzano, cauchy, epsilon, delta, *The American Mathematical Monthly* **107**(9): 844–862.

Glassner, A. S. (1989). How to derive a spectrum from an rgb triplet, *IEEE Comput. Graph. Appl.* **9**(4): 95–99.

Gonzalez-Velasco, E. (1995). *Fourier Analysis and Boundary Value Problems.*, Academic Press Limited.

Haralick, R. M. (1979). Statistical and structural approaches to texture, *IEEE Trans. on Image Processing* **67**(5): 786–804.

Hubel, D. H. and Wiesel, T. N. (1987). Brain mechanisms of vision, *Nature* .

Jarvis, R. (1984). Collision-free trajectory planning using distance transforms, *Proc. of National Conference and Exhibition on Robotics*.

Kemperman, J. H. B. (1991). Mixtures with a limited number of modal intervals, *The Annals of Statistics* **19**(4): pp. 2120–2144.

Khatib, O. (1986). Real-time obstacle avoidance for manipulators and mobile robots, *The International Journal of Robotics Research* **5**(1): 451–457.

Khuri, A. I. (2004). Applications of dirac's delta function in statistics, *International Journal of Mathematical Education in Science and Technology* **35**: 185–195.

Koenderink, J. (1984). The structure of images, *Biological Cybernetics* **50**: 363–370. 10.1007/BF00336961.

URL: <http://dx.doi.org/10.1007/BF00336961>

- Koenderink, J. J. and Doorn, A. J. V. (1992). Receptive field assembly pattern specificity, *Journal of Visual Communication and Image Representation* **3**(1): 1 – 12.
- Koenderink, J. J. and van Doorn, A. J. (1987). Representation of local geometry in the visual system, *Biol. Cybern.* **55**(6): 367–375.
- Krishnaswamy, G., Ang, M. and Andeen, G. (1991). Structured neural-network approach to robot motion control, *Proceedings of the IEEE International Joint Conference on Neural Networks*, Vol. 2, pp. 1059–1066.
- Krse, B. (2000). An efficient representation of the robot’s environment, *Proceedings Intelligent Autonomous Systems, IOS*, IOS press, pp. 589–595.
- Kyllönen, J. and Pietikäinen, M. (2000). Visual inspection of parquet slabs by combining color and texture., *Proc. IAPR Workshop on Machine Vision Applications (MVA’00), November 28-30, Tokyo, Japan.*, pp. 187–192.
- Lin, K. C. (2003). Fast image thresholding by finding the zero(s) of the first derivative of between-class variance, *Mach. Vision Appl.* **13**(5-6): 254–262.
- Lindeberg, T. (1994). *Scale Space Theory in Computer Vision.*, The Kluwer International Series in Engineering and Computer Science., Kluwer Academic Publishers, Netherlands.
- Lindeberg, T. (1998). Feature detection with automatic scale selection, *International Journal of Computer Vision* **30**: 79–116. 10.1023/A:1008045108935.
- Malik, J. and Perorna, P. (1990). Preattentive texture discrimination with early vision mechanisms., *J. Opt Soc Am A.* **7**(5): 923–32.
- Martin, D., Fowlkes, C., Tal, D. and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *Proc. 8th Int’l Conf. Computer Vision*, Vol. 2, pp. 416–423.
- Mcgee, T. G., Sengupta, R. and Hedrick, K. (2005). Obstacle detection for small autonomous aircraft using sky segmentation, *Proceedings of the 2005 IEEE International Conference on Robotics and Automation. ICRA 2005.*, pp. 4679–4684.
- URL:** http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1570842

- Medina Carnicer, R. and Madrid Cuevas, F. (2008). Unimodal thresholding for edge detection, *PR* **41**(7): 2337–2346.
- Mobley, C. D. (1994). *Light and water. Radiative transfer in natural waters.*, San Diego: Academic Press.
- Mohan, A. and Deb, K. (2002). Genetic-fuzzy approach in robot motion planning revisited: Rigorous testing and towards an implementation, *Proceedings of AFSS International Conference on Fuzzy Systems*, Vol. 4.
- Mojsilovic, A., Mojsilovic, R., Kovacevic, J., Hu, J., Safranek, R. J., Member, S., Member, S. and Ganapathy, S. K. (2000). Matching and retrieval based on the vocabulary and grammar of color patterns, *IEEE Trans. Image Processing* **9**: 38–54.
- Ojala, T., Pietikinen, M. and Menp, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions On Pattern Analysis and Machine Intelligence* **24**(7): 971–987.
- Otsu, N. (1979). A threshold selection method from grey-level histograms, *SMC* **9**(1): 62–66.
- Ren, X., Fowlkes, C. C. and Malik, J. (2006). Figure/ground assignment in natural images, *IN ECCV*, Springer, pp. 614–627.
- Robinson, I. S. and Mitchelson, E. G. (1983). Satellite observations of ocean colour [and discussion], *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* **309**(1508): 415–432.
URL: <http://rsta.royalsocietypublishing.org/content/309/1508/415.abstract>
- Rodriguez, R. (2006). A strategy for blood vessels segmentation based on the threshold which combines statistical and scale space filter: Application to the study of angiogenesis, *Computer Methods and Programs in Biomedicine* **82**(1): 1–9.
- Romeny, B. M. H. (1994). *Geometry-Driven Diffusion in Computer Vision.*, Kluwer Academic Publishers., Kluwer Academic Publishers, Netherlands.
- Rosin, P. L. (2001). Unimodal thresholding, *Pattern Recognition* pp. 2083–2096.

- Sakai, T. and Imiya, A. (2009). Unsupervised cluster discovery using statistics in scale space, *Engineering Applications of Artificial Intelligence* **22**(1): 92 – 100.
- Schilling, M. F., Watkins, A. E. and Watkins, W. (2002). Is human height bimodal?, *The American Statistician* **56**(3): 223–229.
- Setalaphruk, V., Ueno, A., Kume, I., Kono, Y. and Kidode, M. (2003). Robot navigation in corridor environments using a sketch floor map, *Proceedings of the 2003 IEEE Symposium on Computational Intelligence in Robotics and Automation*, Vol. 2, pp. 552–557.
- Spencer, L. and Shah, M. (2004). Water video analysis, *ICIP*, pp. 2705–2708.
- Stramski, D., Boss, E., Bogucki, D. and Voss, K. J. (2004). The role of seawater constituents in light backscattering in the ocean, *Progress in Oceanography* **61**: 27–56.
- Sznaier, M., Camps, O. and Mazzaro, C. (2004). Finite horizon model reduction of a class of neutrally stable systems with applications to texture synthesis and recognition, *Proceedings of the 43rd IEEE Conference On Decision and Control.*, pp. 3068–3073.
- Szummer, M. and Picard, R. W. (1996). Temporal texture modeling, *IEEE International Conference on Image Processing*, pp. 823–826.
- Takahashi, K., Kobayashi, Y., Fujii, M., Shimbo, N., Ueda, H. and Tsutsui, K. (2005). Combined detection method in a sea surveillance system, *IEICE - Trans. Inf. Syst.* **E88-D**(2): 230–238.
- Thakoor, S., Chahl, J., Srinivasan, M. V., Young, L., Werblin, F., Hine, B. and Zornetzer, S. (2002). Bioinspired engineering of exploration systems for nasa and dod, *Artificial Life* **8**(4): 357–369.
URL: <http://www.mitpressjournals.org/doi/abs/10.1162/106454602321202426>
- Thrun, S., Fox, D. and Burgard, W. (1998). Probabilistic mapping of an environment by a mobile robot, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*.
- Vasudevan, C. and Ganesan, K. (1994). Case-based path planning for autonomous underwater vehicles, *Proceedings of the 1994 IEEE International Symposium on Intelligent Control*, pp. 160–165.

- Vidal, R. and Ravichandran, A. (2005). Optical flow estimation and segmentation of multiple moving dynamic textures, *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, IEEE Computer Society, Washington, DC, USA, pp. 516–521.
- Walia, R. and Jarvis, R. (2009). Structure, scale-space and decay of otsu's threshold in images for foreground/background discrimination, *VISSAPP (2)*, pp. 120–128.
- Walia, R. and Jarvis, R. (2010). Horizon detection from pseudo spectra images of water scenes, *Cybernetics and Intelligent Systems (CIS), 2010 IEEE Conference on*, pp. 138–144.
- Walia, R., Suter, D. and Jarvis, R. (2012). Discontinuity detection from inflection of otsu's threshold in the derivative of scale-space, in F. Solari, M. Chessa and S. Sabatini (eds), *Machine Vision*, INTECH, chapter 10, pp. 205–226.
- Wang, Y. and Zhu, S.-C. (2003). Modeling textured motion: Particle, wave and sketch, *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, p. 213.
- Warren, C. (1990). Multiple path coordination using artificial potential fields, *Proc. of IEEE Conf. on Robotics and Automation*, pp. 500–505.
- Wei, L.-Y. and Levoy, M. (2000). Fast texture synthesis using tree-structured vector quantization, *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp. 479–488.
- Widder, D. (1975). *The Heat Equation.*, Academic Press inc.
- Witkin, A. P. (1983). Scale-space filtering, *IJCAI'83: Proceedings of the Eighth international joint conference on Artificial intelligence*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 1019–1022.
- www.inimage.com (2012). (last accessed May 31, 2012).
URL: <http://www.inimage.com>
- Wyszecki, G. and Stiles, W. S. (2000). *Color Science: Concepts and Methods, Quantitative Data and Formulae (Wiley Series in Pure and Applied Optics)*, 2 edn,

Wiley-Interscience.

URL: <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0471399183>

- Young, R. A. (1987). The gaussian derivative model for spatial vision: I. retinal mechanisms., *Spatial vision* **2**(4): 273–293.
- Yuan, L. and yeung Shum, H. (2004). Synthesizing dynamic texture with closed-loop linear dynamic system, *In Proc. European Conference on Computer Vision*, Springer, pp. 603–616.
- Zafarifar, B. and Peter, H. N. (2006). Adaptive modeling of sky for video processing and coding applications, *27th Symposium on Information Theory in the Benelux*, pp. 31–38.
- Zafarifar, B., Weda, H. and de With, P. H. N. (2008). Horizon detection based on sky-color and edge features, Vol. 6822, SPIE, p. 682220.
- URL:** <http://link.aip.org/link/?PSI/6822/682220/1>
- Zagal, J. C., Bjrkmán, E., Lindeberg, T. and Roland, P. E. (2000). Significance determination for the scale-space primal sketch by comparison of statistics of scale-space blob volumes computed from pet signals vs. residual noise, *NeuroImage* **11**(5, Supplement 1): S493.
- Zhiye, K. L., Dongyue, L. and Chen., C. X. (2004). Improved artificial potential field for unknown narrow environments, *Proceedings of the 2004 IEEE International Conference on Robotics and Biometrics*).
- Zhong, J. and Sclaroff, S. (2003). Segmenting foreground objects from a dynamic textured background via a robust kalman filter, *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, p. 44.
- Zhou, J. and Raju, G. (1993). Fuzzy rule-based approach for robot motion control in the presence of obstacles, *Proceedings of the 1993 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4, pp. 662–667.