A BAYESIAN APPROACH FOR IMAGE-BASED UNDERWATER TARGET TRACKING AND NAVIGATION

MUHAMMAD ASIF

UNIVERSITI SAINS MALAYSIA

2007

A BAYESIAN APPROACH FOR IMAGE-BASED UNDERWATER TARGET TRACKING AND NAVIGATION

by

MUHAMMAD ASIF

Thesis submitted in fulfilment of the requirements for the degree of Master of Science

FEBRUARY 2007

ACKNOWLEDGEMENTS

I would like to thank those who helped during my thesis work and my stay in Malaysia. Without their support, I could have never accomplished this work.

I take this special occasion to thank my parents. I dedicate this work to my parents. It would have been simply impossible to start, continue and complete without the support of my parents who, unconditionally provided the resources to me. I really missed them during my masters. Words cannot truly express my deepest gratitude and appreciation to my father and mother, who always gave me their love, blessings, and emotional support all the time. I am also indebted to my sisters, and brother, for emotional support, encouragements and prayers.

I am eternally indebted to my supervisor Dr. Mohd Rizal Arshad for all the help, invaluable guidance and generous support throughout my thesis project. His formative influence on my way of thinking about research will continue well beyond the completion of this thesis. I have been very fortunate to be associated with such a kind and good person and it would take more than a few words to express my sincere gratitude. His professionalism, guidance, energy, humour, thoroughness, dedication and inspiration will always serve to me as an example of the perfect supervisor-cheers.

There are too many people to mention individually, but some names stand out. I want to extend special thanks to my friends, Mohsin, Fahad, Husnain, and Abid for being such a good friends.

I wish to thank my lab mates, Salam, Azwan, Nadira, Shariha, Sofwan and Zulkifli at the USM Robotic Research Group for their help and friendship. I have really enjoyed working with them, and I have learned a lot from them also. I especially want to thank Prof. Farid Ghani and Dr. Shahrel Azmin for their enlightening suggestions and advices. I would also like to thank all my teachers and friends from the early days.

Finally, I would like to thanks Oceaneering International for providing us the real underwater pipeline inspection images.

Muhammad Asif
February 2007

TABLE OF CONTENTS

CHAPTER THREE : RESEARCH METHODOLOGY: THEORY AND IMPLEMENTATIONS

CHAPTER FOUR : EXPERIMENTAL RESULTS AND DISCUSSION

CHAPTER FIVE : CONCLUSION AND FUTURE WORK

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

1-D	One – dimensional
2-D	Two dimensional
3-D	Three dimensional
ACM	Active contour model
AUV	Autonomous Underwater Vehicle
BLA	Bresenham Line Algorithm
CamShift	Continuously Adaptive Mean Shift
CCD	Charge Couple Device
d.o.f.	Degree of Freedom
DSP	Digital signal processing/processor
EKF	Extended Kalman Filter
HSV	Hue, Saturation, Value
IIR	Infinite Impulse Response
IKF	Iterative Kalman Filter
LoG	Laplacian of Gaussian
PCA	Principle Component Analysis
PM	Perona-Malik
PVS	PISCIS Vision System
RGB	Red, Green, Blue
ROI	Region of Interest
ROV	Remotely Operated Vehicle
SMC	Sequential Monte Carlo
SSD	Sum of Squared Difference
UKF	Unscented Kalman Filter
UUV	Unmanned Underwater Vehicle

LIST OF PUBLICATIONS & SEMINARS

Conference Papers

Asif, M., Nasirudin, M.A. and Arshad, M.R. (2005). Active Contour for Intelligent Road Tracking System. *1st National Conference on Electronic Design (NCED 2005),* 18 - 19 May 2005, Perlis.

Asif, M., Arshad, M.R. and Wilson, P.A. (2005). AGV Guidance System: An Application of Simple Active Contour for Visual Tracking. *WEC'05 - The Fourth World Enformatika Conference,* June 24-26, 2005, Istanbul, Turkey.

Asif, M., Arshad, M.R. and Yahya, A. (2006). Visual Tracking System for Underwater Pipeline Inspection and Maintenance Application. *International Conference on Underwater System Technology: Theory And Applications 2006 (USYS'06)*, 18-20 July 2006, Penang, Malaysia.

Asif, M. and Arshad, M.R. (2006). An Active Contour for Underwater Object Tracking and Navigation, *International Conference on Man-Machine Systems (ICoMMS 2006)*, 15-16 September 2006, Langkawi Islands, Malaysia.

Yahya, A., Sidek, O., Saleh, J.M. and **Asif, M.** (2006). Frequency Hopping Spread Spectrum for Underwater Acoustic Communication and Doppler Frequency Effects on BER. *International Conference on Underwater System Technology: Theory And Applications 2006 (USYS'06)*, 18-20 July 2006, Penang, Malaysia.

Yahya, A., Sidek, O., Saleh, J.M. and **Asif, M.** (2006). Underwater Acoustic Channels and Diversity Techniques. *International Conference on Underwater System Technology: Theory And Applications 2006 (USYS'06)*, 18-20 July 2006, Penang, Malaysia.

Yahya, A., Sidek, O., Saleh, J.M. and **Asif, M.** (2006). Slow Frequency Hopping Using Different Values of M-ary FSK System in Underwater Acoustic Media. *International Conference on Underwater System Technology: Theory And Applications 2006 (USYS'06)*, 18-20 July 2006, Penang, Malaysia.

Asif, M., Arshad, M.R. and Yahya, A. (2007). AGV Guidance System: An Application of Active Countor and Kalman Filter for Road Tracking. 4th International Symposium on Mechatronics and its Applications, 2007 Sharjah, UAE.

Journal Paper

Asif, M., Arshad, M.R. and Wilson, P.A. (2005). AGV Guidance System: An Application of Simple Active Contour for Visual Tracking, *A Transactions on Engineering, Computing and Technology*. Vol. 6, June 2005, 74-77.

Book

Asif, M. and Arshad, M.R. (2006). Chapter 18: An Active Contour and Kalman Filter for Underwater Target Tracking and Navigation, Cutting Edge Robotics, Mammendorf, Germany, *Pro Literatur Verlog,* ISBN 3-86611-198-3, December 2006.

Seminar

Asif, M. (2006). A Bayesian Approach for Image-Based Autonomous Underwater Target Tracking and Navigation. *School of Electrical and Electronic Engineering, Universiti Sains Malaysia.* 12th July, Pulau Pinang, Malaysia.

SATU PENDEKATAN BAYESIAN BAGI PENJEJAKAN DAN PENGEMUDIAN SASARAN DALAM AIR BERDASARKAN IMEJ

ABSTRAK

Operasi pemeriksaan dan pemantauan di dasar laut merupakan aktiviti penting untuk industri di luar persisiran pantai terutamanya bagi tujuan pembangunan dan pemasangan infrastruktur. Sejak kebelakangan ini, pemasangan struktur di dasar laut seperti saluran paip gas atau petroleum dan kabel telekomunikasi telah meningkat. Pemeriksaan rutin adalah sangat mustahak untuk mencegah kerosakan. Kaedah pemeriksaan dan pemeliharaan struktur di dasar laut ketika ini menggunakan kamera video atau penderia penglihatan yang dipasang pada kenderaan dasar laut berautonomi. Pelbagai algoritma penglihatan bagi pemeriksaan di dasar laut telah dicadangkan di seluruh dunia. Walau bagaimanapun, kebanyakannya tidak memberikan prestasi yang mencukupi bagi keadaan laut yang kompleks. Usahan penyelidikan ini mengkhususkan isu penjejakan saluran paip di dalam air menggunakan penglihatan kamera dalam situasi yang kompleks. Objektif utama penyelidikan ini adalah untuk implimentasikan sistem penglihatan kamera untuk memandu arah sasaran AUV dan menyediakan sistem yang penting untuk tujuan penjejakan saluran paip di dalam air.

Terdapat dua aspek penting untuk membangunkan sistem ini. Pertama, mengesan saluran paip dalam turutan imej. Pada mulanya, pra pemprosesan imej dilakukan dengan menggunakan kaedah tidak konvensional iaitu skala klabu dan Perona-Malik Menapis dan diikuti dengan Pengubah Hough digunakan untuk mengesan sempadan objek. Setelah saluran paip itu dikenalpasti, lengkung diparameter pula digunakan untuk menggambarkan objek tersebut dan untuk penyarian sifat. Berdasarkan penyarian sifat ini, penyuaian lengkung telah digunakan untuk mengukur kedudukan dan orentasi saluran paip tersebut. Aspek kedua adalah penjejakan saluran paip tersebut dalam turutan imej. Dalam usaha penyelidikan ini, masalah penjejakan saluran paip di dalam air telah diformulasikan dalam istilah model bentuk ruang. Penapis Kalman dan Algoritma Kondensasi digunakan untuk menganggar kedudukan objek di dalam air ke atas masa menggunakan pemprograman dinamik. Penapis Kalman dan Algoritma Kondensasi merupakan satu pendekatan Bayesian, prestasi kedua-dua algoritma ini telah diterokai bagi penjejakan dan pandu-arah sasaran dalam air. Melihat secara khusus pada setiap bahagian dalam sistem penjejakan, telah terbukti secara ujikaji bahawa Algoritma Kondensasi lebih teguh keatas sebarang latarbelakang yang berselerak berbanding Sistem Penapis Kalman dan ia merupakan kaedah yang paling sesuai untuk aplikasi penjejakan saluran paip di dalam air.

A BAYESIAN APPROACH FOR IMAGE-BASED UNDERWATER TARGET TRACKING AND NAVIGATION

ABSTRACT

Undersea inspections and surveys are important requirements for offshore industry and mining organisation for various infra-structures installations. During the last decade, the use of underwater structure installations, such as oil or gas pipeline and telecommunication cables has increased many folds. Routine inspections are essential for preventive measures. Current method for the inspection and maintenance of underwater structures adopt video camera or vision sensor mounted on an autonomous underwater vehicle. Various vision based underwater inspection algorithm have been proposed worldwide. However, most of them have inadequate performance on complex marine environments. The present research effort addresses the issues of autonomous underwater pipeline or cable tracking for routine inspection in complex marine environments using vision. The main objective of this research work is to implement a vision system capable of carrying out visually guided task using an AUV, and provide the necessary functionality for tracking underwater pipeline or cables in an image sequences.

There are two aspects of the developed vision system. First, is the detection of underwater pipeline in an image sequences. Initially, image preprocessing is performed for image enhancement, and then Hough transform is used to detect the object boundary. After detecting the pipeline, parameterised curve is used to represent the underwater pipeline and for feature extraction. Based on the extracted feature, curve fitting is used to measure the current pose and orientation of underwater pipeline. The second aspect is the tracking of pipeline in an image sequences. In this research effort, the underwater pipeline tracking problem is formulated in terms of shape-space models. The Kalman filter and Condensation algorithm are used to estimate the state of the underwater object over time using a linear dynamic model. Though the Kalman filter and the Condensation algorithm are both based on the Bayesian framework, the performance of both algorithms are explored for underwater pipeline tracking and navigation. Looking specifically on individual parts of the tracking systems, the experimentation proved that the Condensation tracking algorithm is more robust to background clutter and occlusion then Kalman tracking system and most suitable for underwater pipeline tracking application.

CHAPTER ONE

INTRODUCTION

1.0 Overview

The applications of unmanned underwater vehicles or UUV's have extensively grown in last twenty year (Yoerger *et al.* 2000). They typically enter areas that present conditions impossible for humans to endure, that pose a risk to human life greater than their possible benefit, or that are simply too expensive to reach with a similarly equipped manned-vehicle. Technological enhancements in software and hardware have considerably improved the performance of these vehicles in many areas. The potential uses of these vehicles included but are not limited to: scientific (oceanography, geology, and geophysics), environmental (waste disposal monitoring and wetland surveillance), military (mine warfare, tactile information gathering, and smart weapons) and commercial (oil and gas pipeline inspection, harbors, and dam inspection).

Unmanned underwater vehicles employed in commercial application are usually classified into two groups (Kumar *et al.* 2005): Remotely operated vehicles or ROV's and Autonomous underwater vehicle or AUV's.

1.1 Remotely Operated Vehicles

The Remotely operated vehicles receive continuous control input, or piloting, from a train operator who makes decision based on output from a video camera. Unlike the air and land remotely operated vehicles, ROV's are linked to a host ship by cables or tethers that allow two way communications between the vehicle and operator. These tethers provide ample power supplies and large communication bandwidths. The effective use of ROV's required relatively large mother vessel that increase the cost of operations and not suitable for frequent inspections. Moreover, tethering the vehicle limits both the operation range and the vehicle movements (Ortiz *et al.* 2002).

1.2 Autonomous Underwater Vehicles

The autonomous underwater vehicle's do not have such limitation and essentially present opposing capabilities to those of ROV's. AUV's have a wider range of operations as there is no physical link between the control station on the surface and the vehicle, as they carry their power supply onboard. The small sized AUV's are able to be operated with small sized ships, so their operation costs are reduced significantly and can be use frequently which makes it better choice for surveying and inspection tasks (Wick and Stilwell 2002).

AUV is a self contained unit that run control programs stored in onboard memory and execute pre-programmed mission. It does not require any continued human intervention in decision-making (the operator may intervene for emergency surfacing or emergency stop) and work without interruption over any distance or duration allowed by onboard power supplies. The vehicle usually extracts information about its environment using a variety of sensor, and then uses this information to make navigational decisions. The recent development in sensor and autonomous control technology have made AUV's more flexible. Hence, there has been a definite trend toward more robust methods of autonomous navigation such as vision guided control (Lots *et al.* 2000).

1.3 Underwater Vision

Current method for the inspection, surveying and maintenance of underwater structures adopt video camera mounted on an autonomous underwater vehicle. Video camera provides lots of information that can be examined by onboard vision processing unit. These data are used to navigate and control the autonomous underwater vehicle in complex and hazardous underwater environments. Over the last decade, lots of efforts have been made to design and develop vision based control system for vehicle guidance and navigations. This is due to the fact that computers are capable of processing several frames per second and the real time image processing can be realized (Meribout *et al.* 2002).

There are various application where vision system can considerably improve the vehicle performance such as, obstacle avoidance, station keeping, surveying and inspection applications (Lots *et al.* 2000, and Zwaan *et al.* 2002).

Nevertheless, the application of vision system in complex marine environment presents several challenges. Due to the properties of water, optical waves are rapidly attenuated. Back scattering caused by marine snow, which are the presence of floating organic or inorganic particles in water reflect light and degrades visibility conditions. These anomalies must be addressed and accounted for when information is extracted from the images in order to improve accuracy (Ortiz *et al.* 2002).

1.4 Problem Formation

The underwater inspections are mandatory step for offshore industry and for mining organization from onshore-offshore structures installations to operations (Whitcomb 2000). There are two main areas where underwater target tracking are presently employed for offshore and mining industry: (1) sea floor survey and inspection (2) subsea installations, inspection and maintenance.

In this research effort, an AUV vision system is developed that can track underwater installation such as oil or gas pipeline, and power or telecommunication cables for inspection and maintenance application. The usage of underwater installations has increased many folds, and it is desirable to do routine inspection and maintenance to protect them from marine traffic such as fishery and anchoring (Asakawa *et al.* 2000). However, detecting and tracking the underwater pipeline are fairly difficult tasks to achieve. Especially in the complex marine environment, due to the frequent presence of noise in a sub-surface system. Noise is commonly introduced in underwater images by sporadic marine growth and dynamic lighting condition.

Traditionally, inspections and maintenances of underwater man-made structures are carried out by using the remotely operated vehicle (ROV) controlled from a mother ship by a trained operator (Whitcomb 2000). The use of ROV's for underwater inspections are expensive and time consuming job. Furthermore, controlling an ROV from the surface, by a trained operator, requires continuous attention and concentration to keep the vehicle in the desired position and orientation. During long mission, this becomes a tedious task , and is highly prone to errors due to lack of attention and weariness.

Autonomous underwater vehicles offer cost effective alternative to the ROV's. The practice of using an AUV for underwater pipeline or cable inspection and maintenance becomes a very popular area of research for mining and offshore industries (Griffiths and Birch 2000). During the last decade, lots of efforts have been done in the design and development of different AUV tracking systems, especially in conducting routine inspection and maintenance for underwater installation (Asif and Arshad 2006).

Nevertheless, most of them are focus mainly on the robustness of tracking technique, which may have a poor performance on real underwater environments. The object appearance in complex marine environments changes frequently, and this makes the tracking systems non-robust. Also, they may fail to detect and track the underwater installation in occasions where the underwater pipeline is occluded due to the background cluttering, sub surface noise or subsea mud. Hence, a more reliable tracking system is required for enhancing the performance of AUV vision system for underwater surveying, inspection and maintenance application.

1.5 Research Objective

This thesis addresses the issues of underwater target tracking utilising the recent developments in the field of image processing and computer vision. The main objective of present work is to implement a vision guidance system using underwater vision for AUV's that can track underwater pipeline in an image sequences. This research work also try to solve the issue of detection, pose and orientation measurement of underwater pipeline in an image sequences. This research work will be conducted on real underwater image sequences provided by the Oceaneering International (Oceaneering 2003) where background cluttering and partial occlusions are frequent. It is noted that, this thesis does not address the issue of real time hardware implementation of the developed vision tracking algorithm.

1.6 Thesis Outlines

Chapter one has provided an overview of the presented work in this thesis. The remainder of the thesis will be organised as follows.

In **chapter two**, a review of modern tracking systems will be presented with emphasis on underwater tracking methods. The review of various computer and vision processing algorithms suitable for the tracking applications will be covered. Previous efforts employed so far for underwater pipeline and cable detection and tracking will also be presented.

Chapter three will describe all the methodologies that are utilised for underwater pipeline tracking. There are six main section of this chapter. The first is the image processing and image analysis. The second section is on underwater pipeline modeling using parameterised curve. The third section discusses the feature extraction and visual measurement methods. The fourth section explains the Bayesian approach and developed dynamic model for underwater pipeline tracking. Section five and section six are on Kalman filter and Condensation algorithm for underwater pipeline tracking respectively.

In **chapter four**, experimental results of various steps of both Kalman and Condensation tracking system will be presented. This chapter also summarised the contributions and analyse strengths and weaknesses of both tracking algorithms.

Finally, **chapter five** will present the overall conclusion of underwater pipeline tracking system developed, and subsequently, some possible future works will be discussed.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

This chapter provides an extensive review of literature relevant to the research that will be conducted. Initially, a general overview of object tracking, feature extraction and image processing techniques is presented. In recent times, these three fields of research have been studied extensively. Subsequently, more focus will be given on underwater object tracking with emphasis on underwater pipeline or cable tracking. Figure 2.1 outlines the general object tracking methods and also shows the image processing techniques used in object tracking application.

The first section focuses on detail overview of related work in object tracking. The area of object tracking in computer vision is vast, and it should be pointed out that this overview is by no mean claim to be exhaustive. Nevertheless, it tries to capture the principle techniques and algorithms for object tracking. The second section will discuss the various feature extraction and image processing methods. The feature extraction or image processing is an integrated and the most important part of any object tracking method, and for this reason a separate review is presented. Finally, the third section of this literature review will be on underwater object tracking. The main focus on this section is on underwater pipeline or cable detection and tracking.

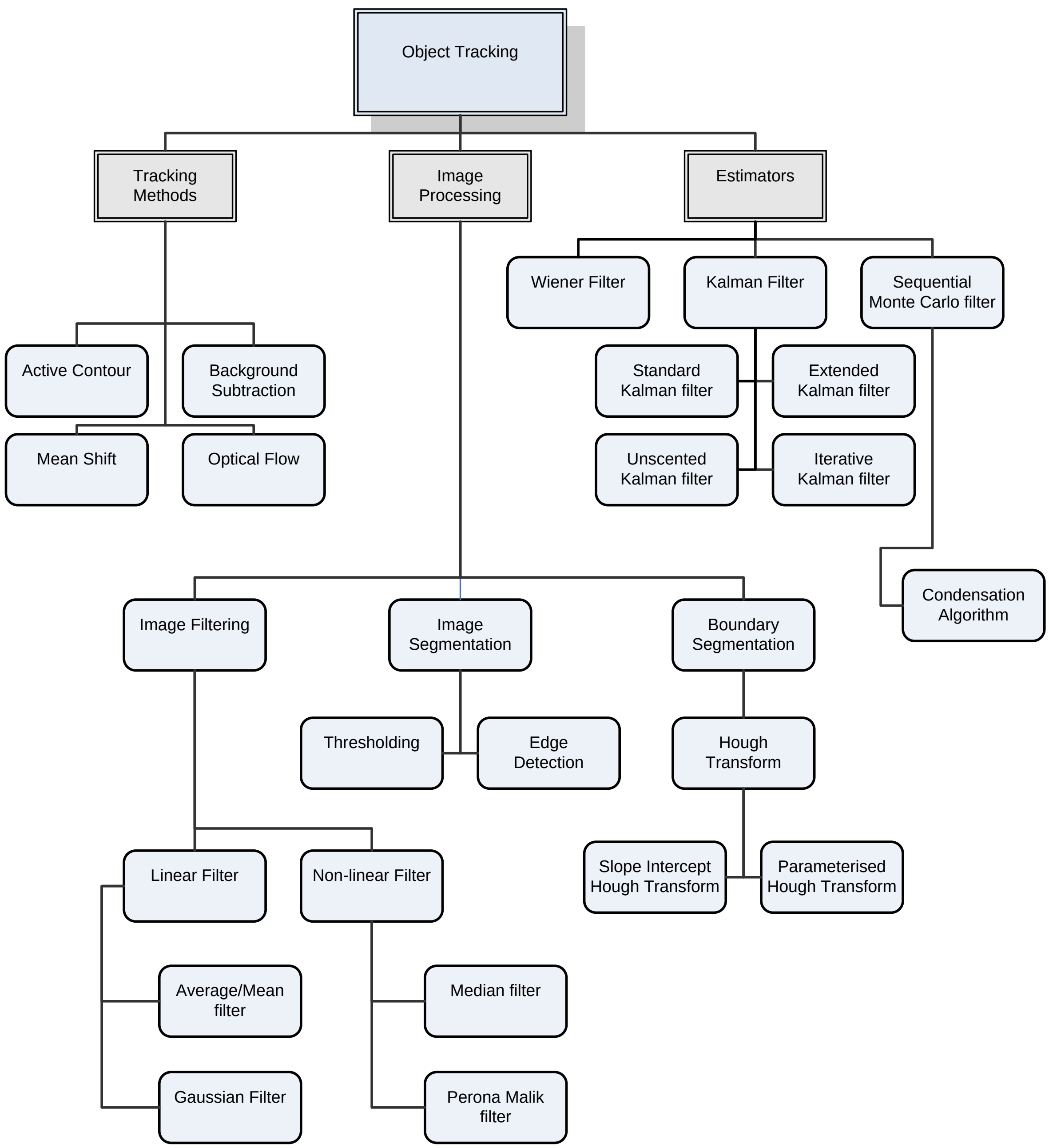

Figure 2.1: Commonly use object tracking techniques

2.1 Object Tracking

A central thread of computer vision research is the development of algorithm or system to track the position and orientation of a target object or objects within images or image sequences. Object tracking, while a simple task for humans is monumentally more challenging for computer vision systems. Over the years, a vast number of algorithms have been proposed for object tracking, and there are large numbers of applications that require such algorithms to track different target in different conditions (Maurin *et al.* 2005). For example, to guide an autonomous vehicle in a simple or complex environments (Kia and Arshad 2005, and Asif *et al.* 2005) or it may be used to track vehicle for collecting the traffic data from highway scenes (Kastrinaki *et al.* 2003) or even to detect human in a surveillance system (Collins *et al.* 2000a). Tracking may also be used in robot arm applications either to provide guidance to surgical robot (Ginhoux *et al.* 2003, and Zhang and Payandeh 2002) or to select an optimal grasp for picking-up object (Han and Kuc 1998). General techniques for tracking are independent from any particular application. More detail of some of these techniques and algorithms are as follows.

2.1.1 Tracking with Background Subtraction

Background subtraction is a conventional and effective technique for finding non-stationary objects in an image sequence (Toyama *et al.* 1999, and Wren *et al.* 1997). When the background is uniform or stationary, detection of moving object can be done by subtracting two frames.

In Haritaoglu *et al.* (2000), and Collins *et al.* (2000b) the background of the sequence of images was defined as the combination of all stationary objects, while the foreground consists of moving objects. The background image was constructed by averaging all the past frames. This simple approach neglects the effect of moving object in the long run with the assumption that the camera is stationary.

In Collins *et al.* (2000a), and Karmann and Brandt (1990), the current background of the image sequences was recursively estimated from past image frames using recursive first order infinite duration Impulse Response (IIR) filters. The IIR filter acts on each pixel of the image sequences, and updates slow and gradual changes in the background. By using two IIR filter with different update parameters in parallel, two different background images can be estimated as well. The proposed method is applicable to backgrounds consisting of stationary objects or slow-moving objects, and may fall short to the background variation caused by imaging noise, illumination changes, and the motion of non-stationary objects.

Statistical background modeling makes the foreground detection more robust to illumination changes, shadow and other artifacts. Several researchers suggest background estimation and up-gradation based on statistical functions on a sequence of most recent frames such as mean, mode or median.

Stauffer and Grimson (1999) proposed a method of statistical background estimation. In this method each pixel was modeled as a mixture of Gaussian and the model was updated in on iterative manner. This system can deal with small and frequent illumination changes, and slow-moving objects.

A similar framework proposed by Francois and Medioni (1999), in which background pixels values are modeled as mixture of Gaussian distributions in HSV colour space. The value observed for each pixel in a new frame is compared to the current corresponding distribution. The pixels on the moving object in the image then are grouped into connected components. The distribution is updated using the latest observation. The assumption is that the object will not appear in the first few frames, which are used for constructing the background distribution.

Elgammal *et al.* (2000) proposed a nonparametric model for background modeling, where a kernel-based function was employed to represent the colour distribution of each background pixel. The kernel-based distribution is a generalisation of mixture of Gaussian which does not require parameter estimation. The proposed approach handled the situations where the background of the scene is cluttered and not completely static but contains small motions and illumination changes. The model estimated the probability of observing pixel intensity values based on a sample of intensity values for each pixel. The model adapt quickly to changes in the background scene which enables very sensitive detection of moving targets. The computation was high for this method. A variant model was used in Haritaoglu *et al.* (2000), where the distribution of temporal variations in colour at each pixel is used to model the spectral feature of the background. Mixture of Gaussian performs better in a time varying environment where the background is not completely stationary. However, the method can lead to misclassification of foreground if the background scenes are complex.

2.1.2 Optical Flow

Optical flow has long been used as a way both to approximate dense motion field over the entire visible region of an image sequence, and to segment areas of consistent flow into discrete object (Hussain 1991, Beauchemin and Barron 1995, and Ju *et al.* 1996). It specifies how much each image pixel moves between successive images, so it is an approximation of the local image motion. The ultimate goal from this approximation is the recovery of the 2D motion field (i.e. the projection of the 3D velocity profile onto a 2D plane; or the apparent motion of image brightness patterns in an image).

Okada *et al.* (1996) proposed a generalised method to extract optical flow. From this optical flow motion, localisation can then be achieved. Okada *el al.* (1996) implemented a real time object system which is based on iterative flow algorithm and parallel DSP hardware. However, this system cannot track multiple objects and heavily dependent on finite object model information.

Smith (1993), and Smith and Brady (1995) have built a system to track vehicle in an image sequences. In this system, optical flow method was computed using two dimensional features such as corners and edges. The clusters of flow vectors which are spatially and temporally significant, provide the object motion information. The system was implemented on a set of PowerPC based image processing system for real time performance.

Ohnishi and Imiya (2006) developed an algorithm using optical flow technique for detecting the obstacle and dominant plane in an image. The dominant plane (plane occupies the largest domain in the image) detection is a vital task for the mobile robot navigation and path planning. The optical flow field was computed by obtaining the points on a dominant plane in a pair of successive image from an image sequences. Then affine coefficients were computed of the corresponding points in two successive images to obtain the dense planar flow from the pre-detected images. The computed optical flow field and planer flow were then used to compute the dominant plane area and obstacles.

A comprehensive survey on optical flow technique and its real time implementation can be found in Liu *at el.* 1998.

2.1.3 Mean Shift

Mean shift tracking has recently been developed for tracking object(s) in a sequence of image frames (Comaniciu *at el.* 2000, and Beleznai *et al.* 2004). The standard mean shift algorithm is a non-parametric technique that determines the location of the moving object in the next frame through an iterative process. This iterative procedure shifts each data points to the average of data points in its neighborhood. The data could be visual feature of the object such as colour, texture and gradient. Their statistical distribution characterise the object of interest, e.g. in Comaniciu *et al.* (2000) the spatial gradient of the statistical measurement is exploited. The basic mean shift algorithm is as follows:

1) Choose a search window size.
2) Choose the initial location of the search window.
3) Compute the mean location in the search window.
4) Centre the search window at the mean location computed in step 3.
5) Repeat steps 3 and 4 until convergence (or until the mean location moves less than a preset threshold).

Bradski (1998) developed a modified version of the mean shift algorithm, named Continuously Adaptive Mean Shift algorithm or CamShift algorithm to deal with dynamically changing colour probability distributions derived from sequence of image frames. This probability is created via a histogram model of the object colour or other specific colours. The tracker moves and resizes the search window until its center converges with the center of mass.

2.1.4 Active Contour Model and Deformable Model

Active contour model or ACM and deformable model are developed as useful tool for image segmentation and tracking of rigid or deformable object. Active contour model or Snakes or deformable model is first introduced by (Kass *et al.* 1987).

Deformable model or active contour is mainly used to find objects and feature in grey level images. An active contour model represents an object outline or boundary as a parametric curve that is allowed to deform from some arbitrary initial shape towards the desired final shape. This deformation is relatively insensitive to illumination changes, and imposing smoothness constraints on the curvature of the contour and the motion of the object.

Many researchers have done the modification on the ACM, and different derivatives of active contour model are proposed. In addition they have also shown how these active contour models can be used to locate and track an object in an image (Ray *et al.* 2002, Kim *et al.* 1999, and Peterfreund 1999).

Jain *el al.* (1998) presented the state of the art survey on active contour. They combined several research articles and categorised the various active contour models into two categories: freeform model and parametric active contour. According to their survey, the freeform active contour model (Terzopoulos and Metaxas 1991, Cohen 1991, and Christensen *et al.* 1996) can represent any arbitrary shape as long as some general regularisation constraints (continuity, smoothness, etc) are satisfied. They are generally called active contours. In contrast, the parametric deformable models are based on prior knowledge or information of geometrical shape and variation of object. This prior knowledge or information about object makes deformable template more robust against boundary gaps and noise in an image. They further categorised the parametric deformable models into two groups: the analytical deformable templates and the prototype deformable templates. The analytical deformable template (Chow *et al.* 1991 and Yuille *et al.* 1992) is defined by a set of analytical curves, preferably with a small number of parameters. The template deforms according to the geometrical shape of the object by using different values of the parameters. Shape variation is determined using the parameter values. The prototype based deformable templates (Staib and Duncan 1992, Sclaroff and Pentland 1995, and Zhong *et al.* 2000) are considered a more flexible approach to derive the deformable templates. In this approach the templates are defined around the standard object which describes the characteristic shape of a class of objects. Each instance of the shape class is derived from the prototype via a parametric mapping. The shape variations in an object class were achieved by imposing a probabilistic distribution on the deformation parameters.

Since this literature review is focused mainly on tracking applications, the variant of active contours are not covered here. However the detail on active contour model and its various types can be found in (Jain *el al.* 1998, and Ray *et al.* 2002).

To track object in an image sequence Zhong *et al.* (2000) integrated the several techniques with deformable model and designed a new deformable template model. In their approach Zhong *et al.* (2000), considered both the exterior contour information and internal information of an object to be tracked. The template which was based on the object's properties such as colour or edges was defined manually as a prior knowledge of an object shape and was executed until it converges. This allows the system to learn the shape of the object to be tracked. For all the subsequent frames, the process were initialised using the template from the previous frame and the object was tracked as it is being deformed in the image sequence. The deformable template was based on two possible transforms, spline or wavelet-based, and both model deform locally. The shape variations in each class were achieved by imposing a probabilistic distribution on the deformation parameters. The template was deformed via attraction of high intensity gradient (edges), motion boundaries and colour/grayscale similarity in future frames. The results were then evaluated using the objective function value, which takes into account both shape deviation from the prototype and the fidelity of the deformed template to the input data.

Koller *et al.* (1994) developed a multiple car tracking system in road scene using explicit occlusion reasoning based on Kalman snakes. The system provide track and shape description of the vehicles for traffic surveillance. The initialisation of the tracker was performed by background subtraction between a continuous update background image and the newly acquired image. The background update was based on motion segmentation. Differential Gaussian intensity filters are applied to obtain gradient edge set models that incorporated the motion segmentation information. One of the features of the car tracking system was its ability to deal with multiple occlusions. The exploitation of prior knowledge of object in the system allows the processing of objects from the bottom to the top of the image frame, i.e. allowing explicit reasoning about occlusion situations.

Paragios and Deriche (1998) proposed a level set snake for detection and tracking the moving objects in image sequence by defining the energy function over the image. The main difference in this technique is the independence of the topology due to the level-set representation. This allows detection of all the objects which appear in the image plane, without knowing their exact number. To overcome the computational problem, a fast algorithm was developed to track and detect contour in an image using “Narrow Band” and

"Fast Marching" methods. Finally real time tracking and detection were shown in this article.

Liu *et al.* (2005) proposed the object detection and tracking algorithm using an active contour for monocular robots in indoor environments. In the proposed system, level set active contour was used to avoid the contour re-initialisation problem. The initial contour converged precisely and quickly into the actual contour by computing the optical flow in subsequent image. The algorithm detects and track object without any prior knowledge at the beginning.

Han and Hahn (2005) developed a new visual tracking scheme for a mobile robot for detecting and tracking the moving target using a single camera mounted on the mobile robot. They proposed a shape adaptive SSD (sum of squared difference) algorithm for detecting the target whose shape may be changed in the image frame due to rotation and translation. The SSD algorithm used the extended snake algorithm to extract the contour of the target and updates the template in every step of the matching process. The 2D template of the target shape was initialised in the first stage by computing the difference of two consecutive frames and morphological closing. Subsequently, the target position in the next frame was predicted using the velocity vector of the target. The velocity of the target in the image frame had been computed as the sum of the velocity components caused by the mobile robot and the target itself. The proposed visual tracking scheme can process 12 frames per second and considered feasible for real time implementation.

2.1.5 Estimators

The methods or techniques for object tracking discussed so far, are trying to track object by finding the region or features that best match the characteristics of the object being tracked. This section describes the algorithms that go one step further, and attempt to predict the state or location of the object in the next step or in the next image frame based on the previous measurement. These algorithms are also called the estimators, because they are used to estimate the parameters or state of the system using noisy and indirect measurements. In the field of object tracking, Bayesian Sequential Estimation which is also called Bayesian filtering, is the most widely accepted framework for object state estimation (Ristic *et al.* 2004). The two major implementations of Bayesian filtering are Kalman filtering and sequential Monte Carlo (Ristic *et al.* 2004).

The Kalman filter is probably the most well known estimation algorithm for linear systems. It provides an efficient, recursive technique to minimise the least-squares error of each prediction where the system model is governed by a linear, stochastic difference equation. The Kalman filter works under two following conditions. First, the process model is represented by a linear differential equation corrupted by an additive Gaussian noise. Second, the measurements are a linear function of the (unknown) states corrupted by an additive Gaussian noise.

Many tracking systems have used Kalman filters in its variants. Harris (1992a) discusses a system that use Kalman filter for the pose estimation of the model being tracked. The system used pre-define geometrical model of the rigid object and the edges extracted by the canny edge detector are used to match with the model. The system effectively track object in an image sequence, if there have been few objects to track or limited motion between frames. Later, Harris (1992b) presented an extension of their work, in which many features were tracked simultaneously and each has its own Kalman filter. In this system, the corner features of many stationary objects were tracked, and the system attempts to derive the 3D geometry of the scene and ego motion of the camera.

Based on Kalman filtering, a lips and finger tracking system was developed by Blake and Isard (1994). The lips and finger was modeled by using the B-spline function and the Kalman filter was used to estimate the coefficients of B-spline. Measurements were made to find the minimum distance to move the spline so that it lies on a maximal gradient portion of the image. These measurements were used as the next input to the Kalman filter. However the background clutter affects the tracking result significantly.

A similar frame work was used by Tai *et al.* (2004) to design and implement an image tracking system for traffic monitoring at a road intersection. In this system, an active contour model was adopted to obtain the locations of automobiles as well as motorcycles in real-time. This active contour model used B-spline function to represent the vehicle contour in an image frame. To track the individual vehicle motion in sequence of images, Kalman filter was used. They have also developed a PCI bus image processing card using Flex10K200s FPGA from Altera™ which provides real-time edge detection. The developed image tracking system gives, real-time, online traffic parameters such as the number of vehicles, vehicle speed and traffic flow to a traffic control centre.

In order to handle the nonlinear and non-Gaussian situations, various extension or variant of the standard Kalman filter such as Extended Kalman filter (EKF), Iterated Kalman filter (IKF) and Unscented Kalman filter (UKF) are proposed and covered thoroughly by most textbooks on estimation theory, e.g. Grewal and Andrews (2001), and

Zarchan and Musoff (2005). The relevant literatures on object tracking based on Kalman filter are discussed accordingly to the following sections.

The main drawback in the Kalman filters, including EKF and UKF, is the uni-model Gaussian distribution assumption which cannot represent simultaneous alternative hypotheses about the object being tracked (Isard and Blake 1998). Moreover in active vision systems, motion of both object and camera or background clutter makes the distribution of the state more complicated and unpredictable.

One approach to deal with nonlinear non-Gaussian estimation problem is to use the sequential Monte Carlo (SMC) technique. The SMC has shown up in the vision community under several different names, including particle filtering, Monte Carlo filters, bootstrap algorithm and the Condensation algorithm or condensation filtering. Iba (2001) provided a good survey of Sequential Monte Carlo techniques, and tried to unify some of the various names for different disciplines that have developed similar algorithms. Perhaps Gordon *et al.* (1993), presented the first application of SMC in machine vision. However, the results presented were only for a synthetic scenario. Later Isard and Blake (1998) formalised the use of SMC in their Condensation (CONditional DENSity propogATION) algorithm (Isard and Blake 1998) and presented result for real tracking scenarios such as head and hand tracking.

The Condensation algorithm models the prior state variable by a set of samples. At each time-step, samples are randomly chosen, allowed to diffuse forward according to the state noise model, and then checked for support by the measurements. The ability to handle highly nonlinear and non-Gaussian models in Bayesian filtering with a clear and neat numerical approximation enable the Condensation algorithm to gain reasonable popularity.

Philomin *et al.* (2000) used a shape model and Condensation algorithm to track pedestrians from a moving vehicle. They used a point distribution model to represents a class of training shapes and then principle component analysis (PCA) was used to analyse the training set, and detect the shape in the tracking. If the shape of the object varies significantly, a large training contour should be used which leads to an increased computation in the tracking process.

Sidenbladh *et al.* (2000), presented a similar work on tracking people walking using a condensation algorithm. In addition to motion, ridges and the edges of limbs also added as features for computing likelihoods. A PCA technique was used to learn a visual model for the appearance of each limb, and to describe the likelihood that edges and ridges were being generated by the people being tracked or by the background.

Verma *et al.* (2003) proposed face detection and tracking system by using the Condensation algorithm. They developed the temporal relationships between the frames to detect and multiple human faces in a video sequence, instead of detecting them in each frame independently. They first developed a wavelet based probabilistic method for face detection. After that, the probability associated with each pixel, for different scales and two different views (frontal and profile faces) were computed. They also computed the face position, scale and pose, frame by frame. The Condensation algorithm was used to incorporate the temporal information in a video sequence.

Isard and MacCormick (2001) developed a Bayesian Multiple Blob tracker (BraMBLe) for tracking multiple objects using the Condensation algorithm when the number of objects present is unknown and varies over time. The Bayesian correlation method was used to model the background image. For the object model, a simple cylinder was used to represent a standing or walking person. The Condensation algorithm evaluated the current situation of each blob and tracks them over the time.

2.2 Feature Extraction and Image Processing

This section presents the literature review on feature extraction and image processing techniques which are the integral parts of any object tracking system. The image processing and feature extraction provide information about the object and its surrounding environments in image or image sequences. The techniques that are commonly used in tracking application are discussed.

2.2.1 Image Filtering

The conditions for tracking in underwater environment are comparatively different from those in the atmosphere (Matsumoto and Ito 1995). The images captured in underwater medium are corrupted with noise due to several factors, such as organic or inorganic particle, dynamic nature of lighting conditions in marine environment and motion blurring.

Noise in image introduces irregular intensity values of pixels in random locations which cause unwanted edges. These unwanted edges produce extra burden on object detection methods in term of its processing and accuracy. It is desirable to reduce the effect of these pixels for effective image processing in an automated environment. This is most often done by convolving an image with a structure commonly referred as mask, filtering mask or kernel. The mask is mainly consists of odd numbers of row and column in order to have specific centered cell. Some of common image filters are discussed below.

The average or mean filter (Gonzalez and Woods, 2002) computes the average (mean) of the grey-level values within a rectangular filter window surrounding each pixel. This has the effect of smoothing the image (eliminating noise). The basic idea of average filter is to make particular pixels intensity similar to its neighbours. The amount of smoothing or filtering in an image is proportional to mask size.

Like the mean filter, the median filter considers each pixel in the image in turn, and looks at its nearby neighbours to decide whether or not it is representative of its surroundings. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces with the median of those values. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value. (If the neighborhood under consideration contains an even number of pixels, the average of the two middle pixel values is used.)

The Gaussian filter is perhaps most commonly used filter, and it often used as a preprocessing step in edge detection (Basu 2002). It based on a convolution with Gaussian mask. This mask is used to ‘blur’ images and remove detail and noise. In this sense, it is similar to the mean filter, but utilise different mask for convolution operation.

Traditional image processing filters such as mean filter and Gaussian filter are acceptable to smooth the noise in the image. However they also smooth the edges, blur them and change their location. To address this problem, Perona and Malik proposed a