Sensor fusion and map building for underwater navigation

Somajyoti Majumder
Australian Centre for Field Robotics(ACFR)
University of Sydney
Rose Street Building (J04)
Sydney, NSW 2006
Australia

Steve Scheding
Australian Centre for Field Robotics(ACFR)
University of Sydney
Rose Street Building (J04)
Sydney, NSW 2006
Australia

Hugh Durrant-Whyte
Australian Centre for Field Robotics(ACFR)
University of Sydney
Rose Street Building (J04)
Sydney, NSW 2006
Australia

Abstract

This paper describes a framework for underwater navigation by combining information from different sensors into a single scene description called "sensor map". It is shown that features extracted from such a description are more robust than those extracted from an individual sensor. Also presented is a new method of map building and representation, where the map is considered as a mapping of the real world into a multilevel feature space using probability theory. Gaussian distributions have been used for the purpose of representation for its excellent transformation properties and compactness. Experimental results from offshore trials are also presented

1 Introduction

The development of a fully autonomous underwater vehicle is an extremely challenging issue. A key problem is being able to navigate in a generally unknown and unstructured terrain given the limited sensing opportunities available in the underwater domain [Kinsle, 1950], [Kennish, 1994], [Lane et al., 1994], [Chantler et al., 1996], [Bahl, 1990]. In principle, by successively integrating information from a number of sensors, a map of landmarks can be constructed and be used to navigate a sub-sea vehicle [Leonard, 1991], [Dissanayake et al., 1999]. In practice, no single sensor in the sub-sea environment can provide the level of reliability and the coverage of information necessary to perform this type of navigation. It is therefore necessary to use a number of sensors and combine their information to provide the necessary navigation capability.

Representation of natural outdoor terrain is another important issue, on which very little work is reported in the literature. In general most of the schemes used for terrain modeling and reconstruction involve either meshing or recursive synthesis, which are more suitable for off-line map generation. These methods also suffer from scaling and interpolation errors. The computational cost and complexity for extracting information from an arbitrary map location is quite high. As a result such terrain maps are often inaccurate and cannot be used for navigation.

In this paper, a method is described using "blobs" and blob-like patches as scene descriptors to segregate feature information from background noise and other errors. A multi-layered data fusion scheme is then used to combine information from the sensors. The general principle is that all sensor information is projected into a common state-space before the extraction of features. Once projection has occurred, feature extraction and subsequent processing is based on a combined (multi-sensor) description of the environment described here as the "sensor map". This paper also presents a new method of map building and representation, where the map is considered as a mapping of the real world into a multilevel feature space. Then, using probability theory and random variables it can be shown that the "map" is essentially a conditional probability estimation or maximum likelihood estimation of detectable features. Therefore, the resulting feature map is a multi modal distribution. Joint Gaussian distributions are used for representation because of their excellent mathematical properties. However any other distribution can also be applied.

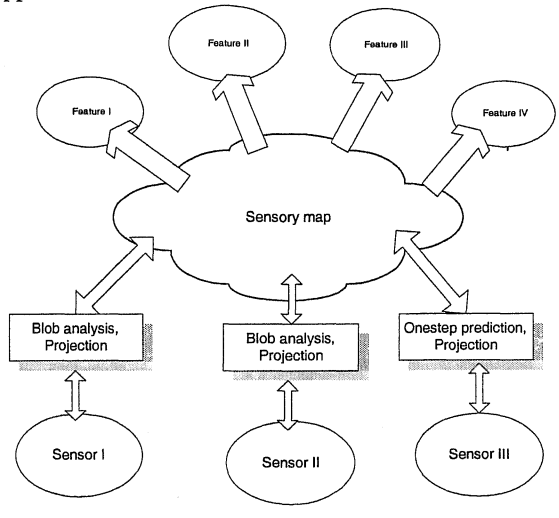

Figure 1 Sensor map : a representation

2 Sensor fusion

At the core of the present sensor fusion model lies the concept of a single composite scene description, its representation and projection to the sensor space. This concept is best defined as a sensor map as depicted in Figure 1. The sensor map is no more than a composite description of all sensor information. However, because all sensor information is projected into one representation, it is also possible to project (at least partially) one sensor's information directly to the information representation of the other sensor(s).

The logic behind this representation is to "fuse" sensor information before feature extraction so as to provide robustness and to base subsequent decisions on the maximum amount of sensor information. This should be compared with the normal data fusion approach of extracting features from a single sensor's information and then fusing the features extracted from many sensors. The reason behind this approach is that in a sub-sea environment individual sensor information is so poor it is difficult to reliably extract features directly without some closely correlated information from a second sensor. Sometimes it is necessary to do some pre-processing or smoothing of the data before fusion. This is an important issue that may directly influence the quality of the sensor map. In case of a very large number of sensors, the dimension or states of the sensor map will still be finite due to the fact that many of these sensors will have common states.

3 Problems in terrain modeling

The task of natural terrain modeling involves a number of issues and the quality of the final terrain map is a function of what level these issues are addressed. Most important among them are:

- lack of features,
- feature and terrain representation,
- arbitrary or unknown transform,
- large data sets with varying resolution,
- occlusion or shadows.

Most of these issues except the first two are mainly due to the fact that the platform used for data collection is situated within the environment rather than far outside. Therefore, as the observer or platform moves closer to the object, the main features decompose into multiple sub-features. Although these sub-features may provide more insight into the surface characteristics and the contents of the object or terrain, they also create anomalies in data association and tracking. Another important aspect in this respect is resolution or scale. Generally, a high scale represents the global view, disregarding the minute local features whereas at low scale finer details or sub-features become dominant. Also a common source of error in natural terrain sensing is the inability of the sensor to see the whole terrain within its viewing field due to a natural phenomenon called occlusion or shadowing. This means that some terrain areas are invisible to the sensor. The traditional Cartesian space based approaches are unsuitable to discern the occluded areas due to sparseness and interpolation error. This further creates problems in identification on upper or lower limit of the terrain. Besides, each sensor has its own inherent inaccuracies in measurements, each being most useful under different sets of operating conditions. A detailed analysis of various errors are reported in [Kweon et al., 1992]. These factors (e.g. shadowing, sparseness, measurement error) limit the completeness and hence the utility of the terrain map. Therefore a single sensor based Cartesian terrain map often contains inadequate information regarding the terrain and falls short of the requirement for a representation which can be utilised for navigation.

4 Feature Map

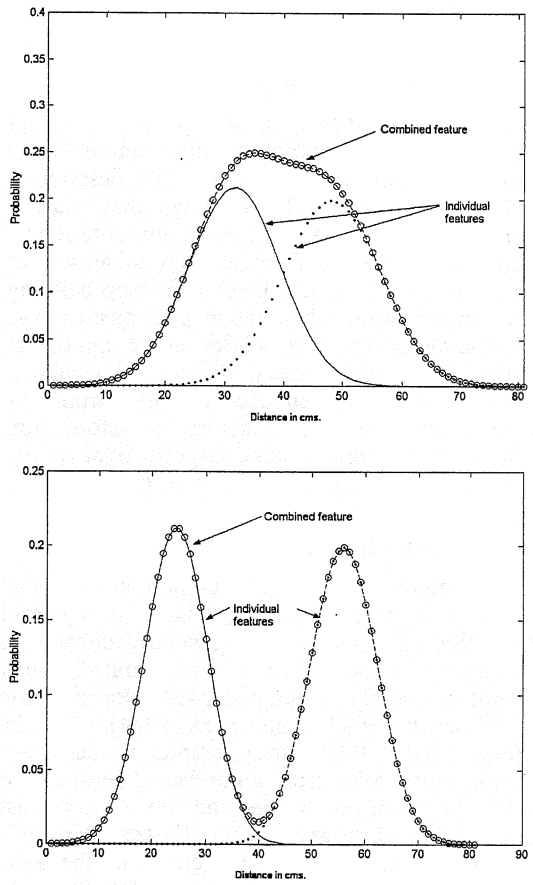

Figure 2 Probabilistic representation scheme as sums of Gaussian distribution

In this work, a feature is defined as a sensor detectable entity represented by a two dimensional spatial region or its transformation which satisfies some well defined mathematical constraints. Therefore any sensor detectable entity can be considered a feature. A few important features that can be used for map building are size, shape, color, texture etc. Some are more reliable than the others under some specific operating conditions, but a combination of them will always result in more precision/robustness than any individual one.

Statistically, a feature map can be best explained as a conditional probability or probability density function for a given location of any observable feature over entire period of time. Joint Gaussian distributions are used for representing the feature where volume or area under the distribution is proportional to feature likelihood. If the same features are observed in successive scans, then they are combined to form a multimodal distribution. This is further depicted in Figure 2, which shows how features are combined to form a feature map. The top image shows that two adjacent specially localized features combine to form a multimodal distribution where as the bottom one shows that even after combining they remain isolated. In this sense this approach is more logical extension of "probabilistic map" [Thrun et al., 1998] and "occupancy grid" [Elfes, 1987] methods. One of the major advantage of this is that it can handle any number of features resulting a multi level map yet its representation is compact (due to the fact that only two terms mean and variance can represent a joint Gaussian distribution).

5 Experimental setup

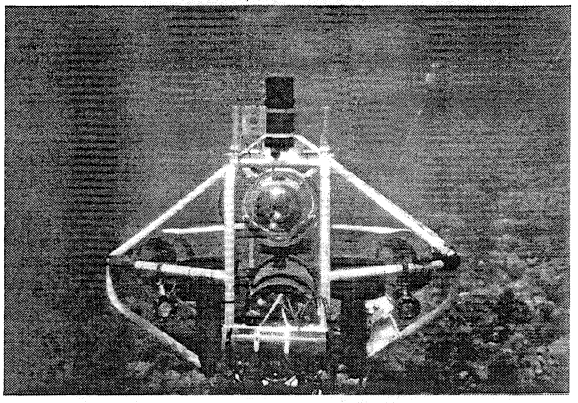

Figure 3 Oberon during offshore trial

For experimental verification of the concept described in Sections 2 and 3, test data has been collected from several offshore trials using the Remotely Operated Vehicle "Oberon", developed at the Australian Centre For Field Robotics of the University of Sydney. Oberon is shown in Figure 3. Further details of the experimental platform are available in [Williams et al., 1999], [Williams et al.,2000], [Newman et al., 1997]. Sensors used in the work described here are a 585 kHz or 1210 kHz (user selectable) pencil-beam sonar, a 675 kHz fan beam sonar, a specially constructed color camera, a fiber optic gyro and a pressure transducer for depth.

An offshore test site with natural terrain features has been selected for the experiments. During the trial, Oberon' was allowed to follow a course parallel to the concretised shore line facing the sea followed by a similar course in opposite direction. The length of each run is approximately 20 - 25metres. The experimental data consists of gyro, depth, sonar and camera data Two forward looking sector scan sonars used here are mounted at an angle of 45°, pointing towards the sea floor. The upper one has a narrow pencil beam whereas the other one is a fan beam sonar. Therefore the sonar objects belong to a narrow window within the respective camera image.

6 Underwater application

To combine information form Sonar to Vision and vice versa, the method of perspective transformation has been applied. The location of forward looking sonar and uncalibrated camera is also shown in Figure 3. Camera calibration was not essential considering the intended accuracy of the system. It is evident that the sonar window is a part of the visual window separated by the distance between the camera and sonar center. In order to build the sensor map from offshore data, either all or a useful part of it needs to be projected into the common state space.

A "blob" based approach is adopted here to minimise the effect of outlier data and to limit the required computation. For this purpose a structure was introduced to segregate important scene information from the background termed the scene descriptor. Blobs or blob-like structures are suitable for this purpose. Blobs are defined (in computer vision) as soft irregular shaped objects, the main purpose of which is to segregate the useful foreground information with all its characteristics from the background, such that any feature extraction algorithm can extract appropriate features such as color, texture, or shape from them.

For the purpose of feature representation joint Gaussian distribution is used using Bayesian update which estimates the posterior distribution conditioned on all given previous information. The main reason for the selection of Gaussian distributions is in their compactness in representation and other useful properties in the transformation mathematics.

7 Data registration

Registration means correct alignment of two image frames or data sets by rotating and translating one over the other so that corresponding features of the two images match. For the purpose of registration, two methods were applied, pair wise matching and global matching with respect to the first image. The former one generates a reliable local estimate where as the later one is more suitable for a global estimate. Most of these methods are neither robust nor accurate and also suffer from several inaccuracies. In case of global matching, (since the processing window is small and the environment does not contain many distinctive features) after a few frames the matched entity disappears from the processing window Therefore pair wise matching is used and instead of one, eight small windows (25pixelX25pixel) are taken from the first image as template. The second image is then searched until the best match for the template is found. To reduce the orientation search space, rotational matching is limited by the angular rotation obtained from the integration of gyro data within 15° orientation change.

For a slow speed linear operation, instantaneous orientation of the "Oberon" has been obtained by integrating gyro data.

8 Implementation.

A three level computational framework has been developed to generate the feature map is discussed here.

Level I: This is basically an alignment and registration level that generates the necessary position and orientation estimates from the sensor measurements.

Level II: Sensor data from both sonar and vision is fused to create a composite image from successively aligned scans using blob detection, clustering and projection method.

Level III: This is the actual map building and representation level that generates 3D Gaussian distribution map and updates the posterior distribution using Bayes rule for each identified feature parameter.

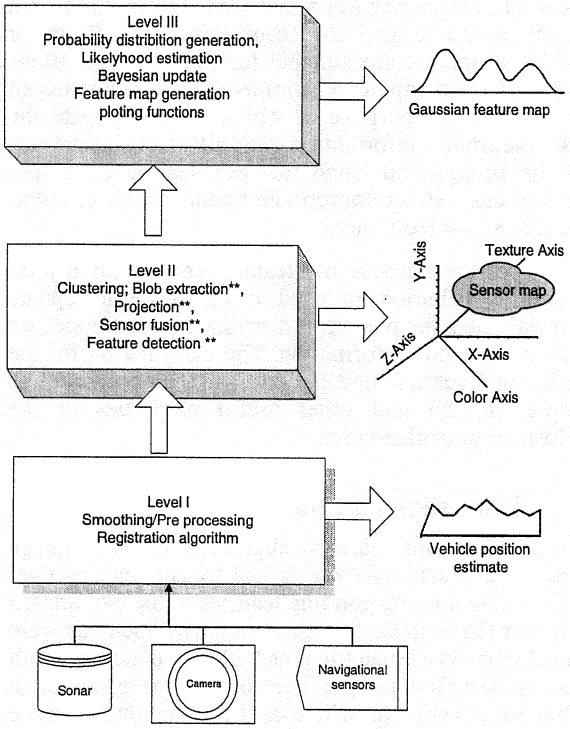

Figure 4 Computational framework

The computational representation of the present concept is given in Figure 4 (Further information on items with ** are available in [Majumder et al.]) which clearly depicts the distinguishable nature of various data processing level to achieve the terrain map from the raw data. This cycle can be repeated as many times based on the number of required feature maps. The final terrain is the combination of this individual feature maps.

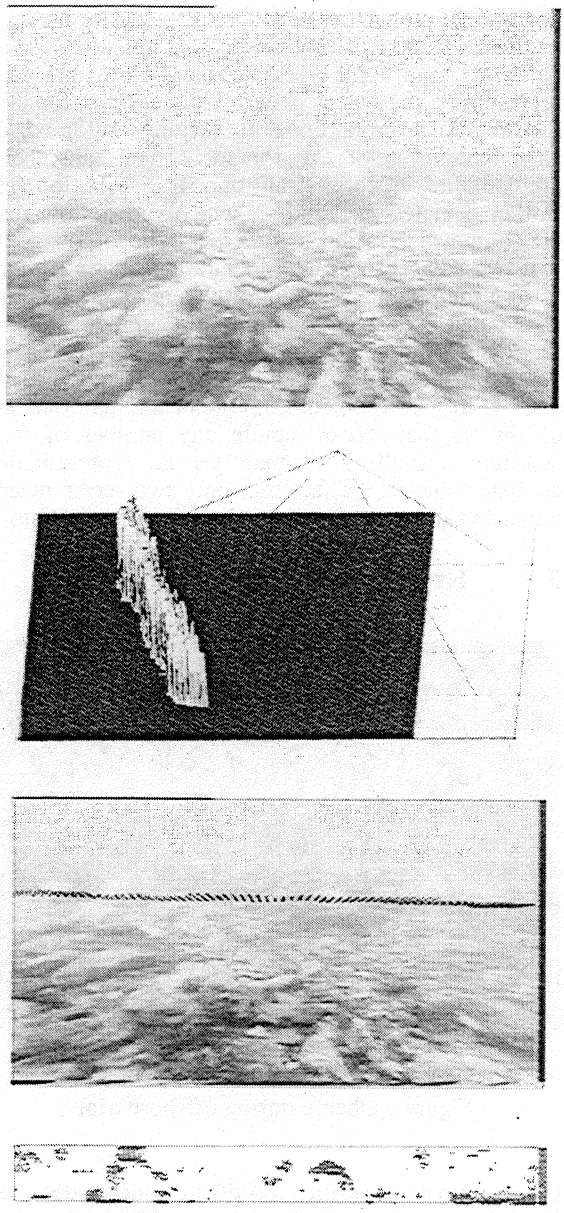

Figure 5 Various stages of sensor fusion ,Reading from top camera image, sonar image, sensor map, detected objects.

9 Results

Offshore data after blob analysis and subsequent image plane fusion using the projection method described is shown in Figure 5 for a single scan where both of the sensors are looking ahead and are tilted to an angle of 45° downwards to view the terrain. It also shows the various stages of processing. The uppermost one is the sub-sea image of natural terrain devoid of tractable features The second image from top shows the perspective view of the actual terrain, seen by a co-located sonar. It further shows the camera viewing field within the sonar domain. The third figure is the sensor map which combines both sonar

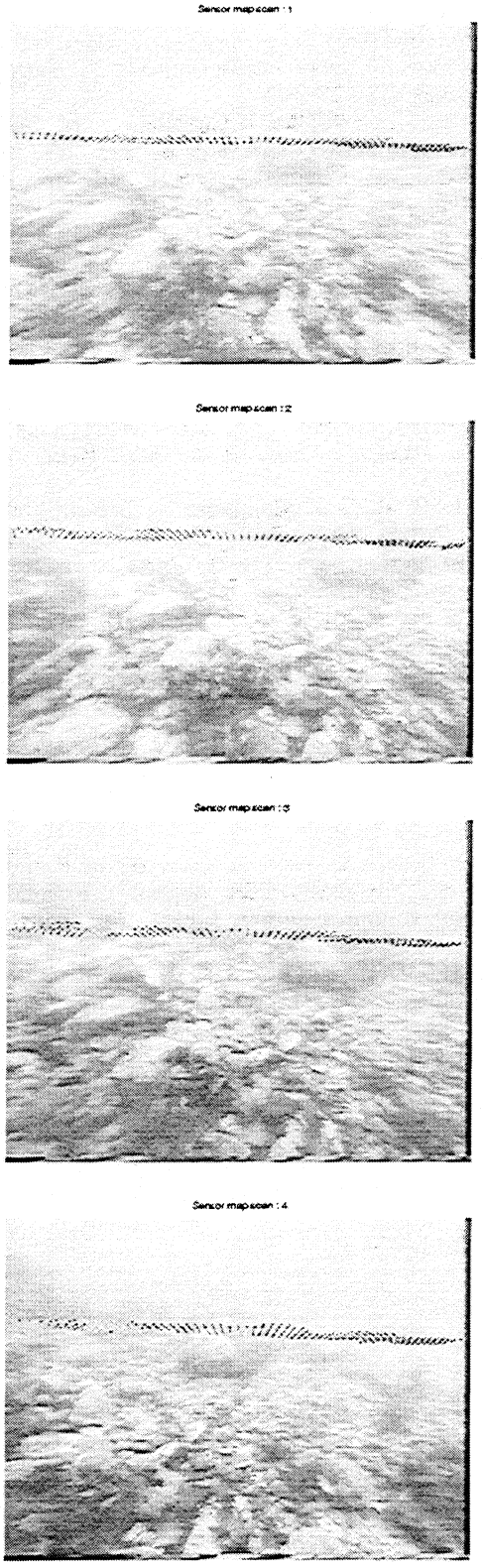

Figure 6 A set of successive sensor maps (reading from top) at 2 second interval.

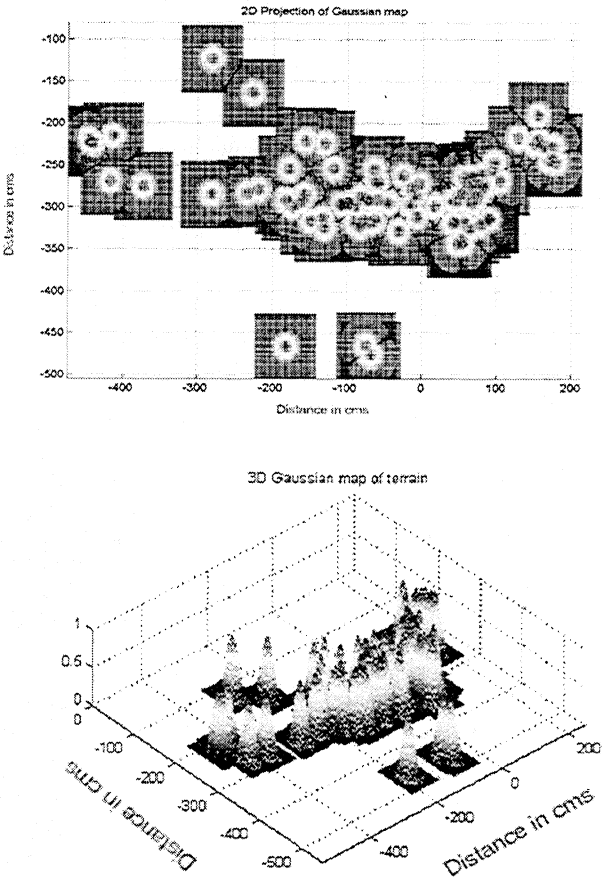

Figure 7 (top) 2D feature map, (bottom) 3D feature map

and vision where sonar objects are shown on the top part of image. The fourth one from the top shows the detected objects with combined features in the scene that are corroborated by both sonar and vision. Another four sensor maps are shown in Figure 6 which are generated from a 25 metre long offshore test run (where sonar objects are combined with camera and plotted as dark points on the top part of the image). This combined feature is the basis of our map building process, which is represented by a bivariate Gaussian distribution to generate the feature map. This is depicted in Figure 7. Where size of the rock and variance of sonar returns (a measure of texture) are used as a feature metric. This further shows that as the vehicle moves, tractable features are continually building the map, where large objects form a multi modal distribution due to their continuity over several scan of the sensor without application of any data association algorithm. However, this does not limit application potential, any number of appropriate features could also be added, effectively increasing the robustness and reliability of the map for subsequent use.

10 Discussions and Conclusion

This paper presents a new method for of combining information from two of the most common sensors used for obstacle detection and navigation of sub-sea vehicle i.e. active sonar and underwater cameras to construct complete environmental map for navigation, using all possible sensor information. The main reason for this is that sonar and vision operate on different physical principles, and provide completely different type of information; point range data and passive two-dimensional images respectively. Also, completely different processing principles are used to manipulate and extract information from these two different sensor sources. A wide variety of computer vision algorithms exist for extracting color, texture, shape from an image (although most of them are not very reliable in the under water domain), whereas sonar data processing is mainly concerned with time of flight and amplitude information to determine object range and size. In natural sub-sea environments normally robust features, such as points and lines, and their associated feature extraction methods, turn out to be very fragile. This work demonstrates a new approach to combine multiple sensor information at a very low level to create a composite scene description - the sensor map, which is essentially the opposite to conventional data fusion methods based on feature extraction followed by fusion.

This paper also presents a method of representation for non-point features using probability theory, random variables, and Bayesian statistics to show that a feature map is essentially a conditional probability or maximum likely hood estimation of (observable and detectable) features. Bivariate Gaussian distributions have been used for the purpose of representation for their excellent transformation properties and compactness. Any other distribution could also be used for the purpose. A set of results from several offshore trials is also presented here. Further extension of this method for localisation is currently being pursued.

Acknowledgement

The authors would like to acknowledge the help of Mr. Stefan B.Williams during various stages of this work.

References

[Willliams et al., 1999] S. B. Willliams, P . Newman , S. Majumder, H.F. Durrant-Whyte and J. Rosenblatt, *Autonomous transect Surv eying of the Great Barrier Reef,* Proceedings of Australian Conference on Robotics and Automation(ACRA) , G. Wyeth and J. Roberts, Eds., , v ol. 1, pages. 16 –20, 1999.

[Willliams et al., 2000] S.B.Williams , P. Newman, MWMG Dissanayake, J Rosenblatt and H.F.Durrant-Whyte , *A decoupled, distributed AUV control architecture*, in Proceedings of 31st International Symp osium on Rob otics May 14-17, Montreal,PQ, Canada, 2000.

[Newman et al., 1997] P . Newman and H.F. Durrant-Whyte, *To wards Terrain-Aided Navigation for Sub-sea Vehicles*, in FSR'97 International Conference on Field and Survice R obotics, Canberra, Austr alia , pages. 244-248, 1997.

[Bahl, 1990] R. Bahl, *Object classication using compact sector scanning sonars in turbid waters*, Proc. 2nd IARP Mobile Robots for Subse a Envir onments, Monterey, CA, vol. 23, pages 303-327, 1990.

[Chantler et al., 1996] M. J. Chantler, D. M. Lane, D. Dai,and N. Williams, *Detection and Tracking of returns in sector scan sonar image sequences*, IEE Proc. Radar, Sonar Navig. , vol. 143, pp. 157-162, 1996,.

[Lane et al., 1994] David M. Lane and J. P Stoner, *Automatic Interpretation of Sonar Imagery Using Qualitative Feature Matching*, IEEE Journal of Oceanic Engineering , v ol. 19, no. 3, 1994, pp. 391-405.

[Kinsle, 1950] La wrence E. Kinsler and Austin R. frey , *Fundamentals of Acoustics* , John Wiley and Sons, Inc, second edition, 1950.

[Kennish, 1994] Michael J. Kennish, *Practical Handbook of Marine Science* , CRC Press, Boca Raton, Fla, 2 edition, 1994.

[Majumder et al.,] S. Majumder S., Scheding and H. F.Durrant-Whyte, *Multi sensor Data Fusion for Under Water Navigation*, (in Press) Journal of Robotics Autonomous Systems.

[Kweon et al., 1992] In so Kweon and Takeo Kanade *High resolution terrain map from multiple sensor data*, IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 14, number 2, pages 278 – 292 , 1992.

[Thrun et al., 1998], S. Thrun D. Fox and W. Burgard, *A probabilistic approach to concurrent mapping and localisation for mobile robots*, Machine Learning, volume 31, pages 29-53, 1998.

[Elfes, 1987] Alberto Elfes, *A Sonar-Based Real World Mapping and Navigation*, IEEE Journal of Robotics and Automation, volume 3, number 3, pages 249-265, 1987,

[Dissanayake et al., 1999], M.W.M.G. Dissanayake P. Newman H.F. Durrant-Whyte S. Clark and M. Csorba, *An Experimental and Theoretical Investigation into Simultaneous Localisation and Map Building*, Proceedings 6th International Symposium on Experimental Robotics, Sydney, NSW, Australia, pages 171 - 180 , 1999.,

[Leonard, 1991], J. J. Leonard and H. F. Durrant-Whyte, *Simultaneous Map Building and Localization for an Anautonomous Mobile Robot*, In Proc. IEEE Int. Workshop on Intelligent Robots and Systems, New York, USA, pages 1442—1447, 1991.