Multi-Sensor Data fusion Approach for Terrain aided Navigation of Autonomous Underwater Vehicles

Bharath Kalyan, Arjuna P. Balasuriya
School of Electrical and Electronic Engineering
Nanyang Technological University, Nanyang Avenue, Singapore 639798
Email: bharathkalyan@pmail.ntu.edu.sg, earjuna@ntu.edu.sg

***Abstract*— Autonomous navigation of underwater vehicles is a 'holy grail' of underwater robotics in many ways. For an AUV to continually estimate its pose (position and orientation) and operate within an unknown environment solely based on its onboard sensors is a very challenging problem. In this paper, commonly used AUV sensors such as Inertial Navigation System (INS), Forward Looking Sonar (FLS) and monocular CCD camera are considered. Paper also discusses a robust feature tracking technique from image sequences and thereby estimating the ego-motion. It also discusses about the trajectory recovery from the FLS and fusion of INS and FLS. Experimental results, conducted using the test-bed AUV 'NTU-UAV', are presented to highlight the performance of the proposed navigation scheme.**

I. Introduction

Significant interest exists in applying autonomous underwater vehicles to perform useful missions in the harsh underwater environments. Real-time dynamic state update of an Autonomous Underwater Vehicle (AUV) and its hostile environment is crucial for navigation and control. At present there aren't any global positioning systems (GPS) available in underwater environments. This causes a fundamental problem of knowing where the vehicle is during the mission. Currently, acoustic transponder based triangulation schemes are used for global positioning. However deployment of transponders is a tedious as well as expensive [1].

In this paper an alternative approach is considered. Here a navigation scheme is proposed solely using on-board sensors for state update including global position of the vehicle. The problem is formulated as a stationary land mark tracking problem enabling vehicle localization with respect to these land marks. Underwater terrain features are selected as natural land marks. Due to the highly complex nature of the environment and poor performance by available on-board sensors, a multi-sensor fusion scheme is vital for a reliable state estimate. The idea of combining information not new. An extensive coverage on various techniques used for combining information from multiple sensors has been reported in [2]. The challenge is therefore to extract these features from the unstructured terrain using acoustic and optical sensors available on-board the vehicle. While a variety of optical sensors are available for underwater use, sonar is a natural choice for investigating the turbid marine environment due to its superior propagating characteristics [3].

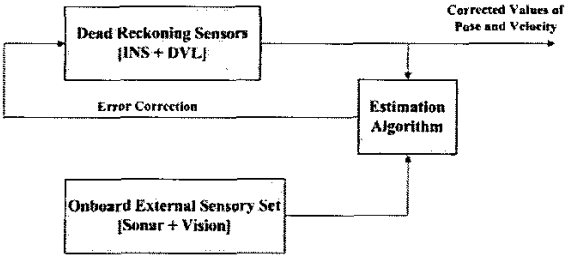

Fig. 1. Sensor Fusion and Motion Estimation Module

The first step for fusion of data from multiple sensors for accurate position estimation is to study in detail the characteristics of individual sensors. As shown in Fig. 1, dead reckoning sensors are used to obtain estimates of pose and velocity. This information is fed into an estimation algorithm which takes measurements from one or more external sensors and uses this information to provide corrections to the dead reckoning. The general principle is measurements from high bandwidth Dead Reckoning (DR) sensors are used for the prediction of the vehicle states. Subsequent observations from external sensors are used to update the state estimates.

We here present the ongoing work developing real-time sector scan and video processing algorithms. Sonar based algorithms have been designed for mapping, navigation and obstacle avoidance. Optical based algorithms are being evaluated for 3D-reconstruction, panoramic mosaicking and ego motion estimation. In this paper we have focussed on the following issues:

1) Development of a robust tracking mechanism, which tracks optimal features from the optical images and implements an automatic mechanism for rejection of outliers. Ego-motion estimation from the video sequences are dealt in section II.
2) Section III deals with data interpretation and trajectory recovery from the sonar.
3) Section IV deals with the fusion of Sonar-INS module.
4) Experimental setup, implementation issues and results are discussed in section V.
5) Finally section VI will provide a brief conclusion and outline our future work.

0-7803-8669-8/04/$20.00 ©2004 IEEE.

II. Robust tracking and ego-motion estimation from Optical Images

A considerable amount of research interest has been directed towards providing autonomy to underwater vehicles using vision, namely in self location and ego motion estimation. Based on underwater image formation models using surface irradiance and light attenuation, Yu *et al.* proposed estimation methods for motion recovery and surface orientation [4]. Marks *et al.* [5] have developed a system for ocean floor mosaic creation where they have employed a four parameter semi rigid model and small zooming and rotation of frames are estimated. In this section, a feature based approach to estimate motion from a sequence of images gathered by a single camera is discussed. A mathematical formulation and variations of this formulation is presented. Robust tracking means detecting the unreliable matches i.e.,outliers, over a sequence of images. The discussion focuses on the estimation of both motion and structure. No distinction is made between the situation where

- the camera is moving and image scene is stationary
- the camera is stationary while the imaged objects are in motion

A. Pre-Filtering

The application of underwater vision is greatly limited by the poor image quality due to forward scattering and backscattering and marine snow. The unwanted features mentioned above fall into the low and high spatial frequency domains [6]. Therefore a band pass filter has to be designed to filter out the unwanted components. The Gaussian operator acts as a low pass and the Laplacian as a high pass filter. Laplacian of Gaussian (LoG) operator in 2D is given by

$$\nabla^2 h(x,y) = \frac{1}{\Pi\sigma^4}\left[1 - \frac{x^2+y^2}{2\sigma^2}\right]\mathbf{e}^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \qquad (1)$$

The image is convolved with this LoG mask, which serves as a band pass spatial filter.

B. Feature Selection

The feature selection process is highly dependent on the kind of application for which the vision is employed. The work presented here evolves from the analysis of point projections and their correspondence between image frames. In order to improve the correspondence finding, a number of points were selected corresponding to image corners or highly textured patches. The selection of image points is based on the well-known corner detector proposed by [7]. This detector finds corners in step edges by using only first order image derivative approximations.

Given an image $I(x,y)$, following steps are used to detect whether a given pixel (x,y) is a corner feature:

- set a window w of fixed size, and compute the image gradient (I_x, I_y)
- at all pixels in the window w around (x,y) compute the matrix

$$G(x,y) = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \qquad (2)$$

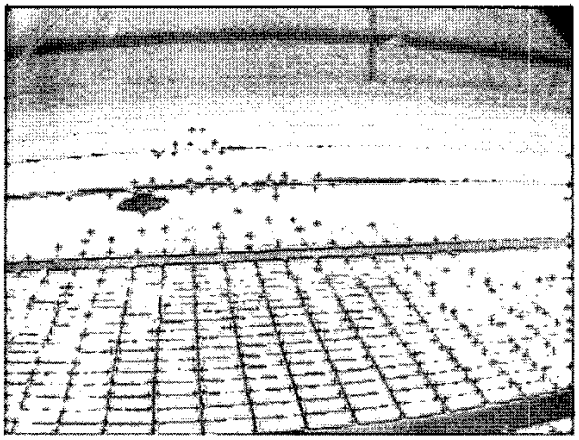

Fig. 2. Corners Detected using Harris Corner Detector superimposed on the Original Image

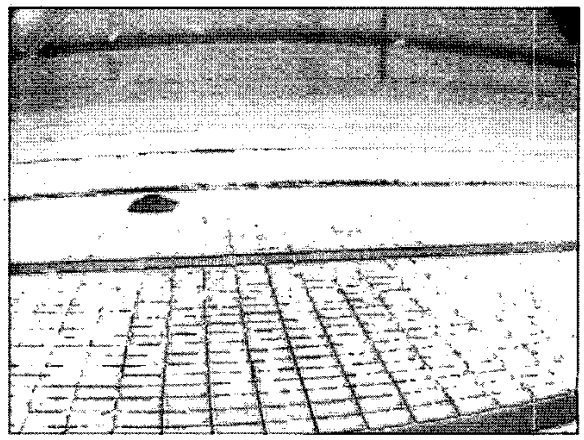

Fig. 3. Corners Detected using Harris Corner Detector superimposed on the Consecutive Image

- A corner strength signal is given by

$$\Phi(x,y) = |G(x,y)| + \kappa Trace^2(G(x,y)) \qquad (3)$$

where $\kappa \epsilon \mathbb{R}$ is a scalar, and different choices of κ may result in favoring gradient variation in one or more than one direction, or maybe both. We also extract the local maxima by performing by performing a grey scale morphological dilation and then finding points in the corner strength image that match the dilated image and are also greater than the threshold. Thereby we are doing a non maximal suppression to determine the strongest corners. The corners detected using the above mentioned scheme are superimposed on the two consecutive frames as shown in Fig. 2 and Fig. 3.

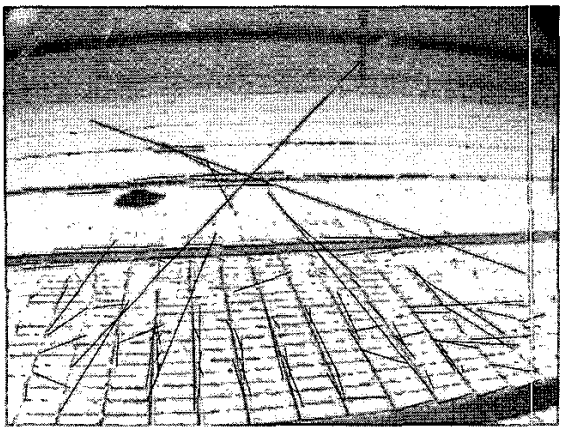

Fig. 4. The Putative matches obtained using the correlation measure

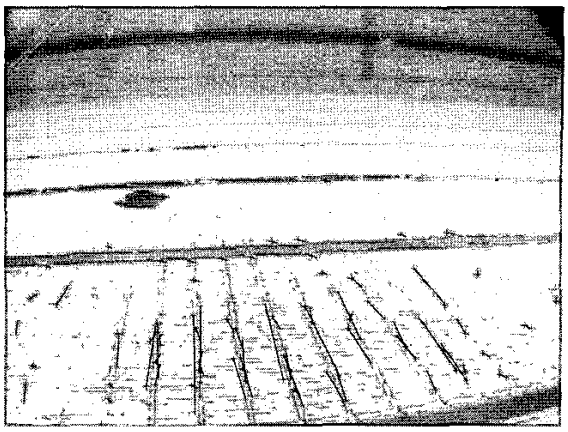

Fig. 5. The inlying matches obtained using RANSAC is superimposed on the 2 images

C. Feature Matching

Putative matches between the previously detected feature points in the consecutive image frames are obtained as shown in Fig. 4. This is done by looking for points that are maximally correlated with each other within windows surrounding each point. Only points that correlate most strongly in both the directions are considered as putative matches. If $I_k(x,y)$ and $I_{k+1}(x,y)$ are the two consecutive image frames with X_k representing the detected features in the image I_k and X_{k+1} in image I_{k+1}, then the correlation measure is given by [8]

$$C_m(X_{k,k+1}) = \frac{1}{\sqrt{\sigma^2_{(X_k)}\sigma^2_{(X_{k+1})}}} \sum_{i=-w}^{w}\sum_{j=-w}^{w} [I_k(x+i,y+j) - \bar{I}_k][I_{k+1}(x+i,y+j) - \bar{I}_{k+1}] \tag{4}$$

where σ^2 is the variance of patch intensity and $\bar{I}$ is the average intensity of the patch. To reduce the effects of ambient lighting, we normalize each pixels intensity by the average intensity of the window. It has been found that normalized correlation measure gives significantly better matches than the un-normalized ones. The assumption of small disparity between the two image frames can be used to significantly reduce the computational burden. We subtract the consecutive image frames with a averaging filter. This compensates for the brightness differences in each frame and allows faster correlation calculation.

D. Outlier Rejection and Ego-motion estimation

After the putative matches have been found, a set of displacement vectors relating the features between the image frames of the sequence is obtained. Due to the error prone nature of the matching process, it is expected that a proportion of these putative correspondences obtained from the previous section will infact be mismatches. Thus we need to have a robust estimation technique for outlier rejection. Here RANSAC (RANdom SAmple Consensus)algorithm is applied to the putative correspondence set to estimate the homography and the inlier correspondences which are consistent with this estimate [9]. The algorithm is described below [10]:

- *Normalization: Compute the similarity transformation T, that takes points x_i to a new set of points $\tilde{x}_i$ such that the centroid of the points of $\tilde{x}_i$ is the coordinate origin $(0,0)^T$, and their average distance from the origin is $\sqrt{2}$ Similarly compute the transformation T' for the points in the second image.*
- *Select a random sample of 4 correspondences and compute the homography $\tilde{H}$.*
- *Calculate the distance $d_\perp$ for each putative correspondence. This distance measure is the symmetric transfer error,*

$$d^2_\perp = d(x, \tilde{H}^{-1}x')^2 + d(x', \tilde{H}x)^2 \tag{5}$$

  where $x \leftrightarrow x'$ is the point correspondences.
- *Compute the number of inliers consistent with $\tilde{H}$ by the number of correspondences for which $d_\perp < t = \sqrt{5.99}\sigma$ pixels.Choose $\tilde{H}$ with largest number of inliers*
- *Do a final least square fit on the data points considered to be the inliers using Direct Linear Transformation (DLT)to obtain the Homography $\tilde{H}$*
- *Denormalize: $H = T'\tilde{H}T$*

Repeat for N samples, where N can be determined adaptively [10](in our case $N = 1000$)

It can be clearly seen from Fig. 5 that the outliers present in Fig. 4 has been effectively removed using RANSAC.

As soon as the best transformation H between two image frames has been found, the two image frames can

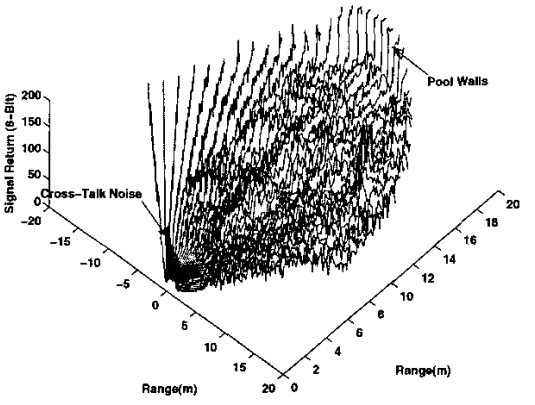

Fig. 6. Range Bin Map: It shows a 135^o sweep, with each sonar return representing range bins scaled to a preset distance; In this case it is 20m

be warped together to form a mosaic. The 2D motion of the camera is known in pixels from one frame to the next, as an affine measure of translation, rotation and scaling. With the knowledge of the intrinsic parameters of the camera, the metric information about the vehicle motion can be recovered.

III. Sonar Based Tracking and Navigation

Amplitude returns contain the summed effect of reflections from a wide variety of objects in the beam together with a high level of "acoustic noise" [3]. Data from underwater environments, the gray scale intensity data obtained in successive scans, is often changed even when the relative location of sonar and the surrounding objects are slightly changed. This is due to the fact that at any given instant, the echo received is a collection of returns from both the target and the surrounding environment. Main elements of the return are multi-path reflections, or returns from side lobe pickups. A range bin map indicating the sonar returns is shown in Fig. 6. These factors complicate the task of segmentation, classification and feature detection. Often, a priori knowledge of the working environment is essential for interpretation of the sonar images. There has been a great deal of research undertaken in developing techniques for automatic feature detection and classification of sonar data, generally with limited success [11] [12] [13]. A motion estimation technique based on correlation measure between sonar image frames is presented. The images are processed as shown in the flowchart of Fig. 7

A. Filtering and Segmentation

The data returned by the imaging sonar consists of the complete time history of each sonar ping in a discrete set of bins scaled over the desired range. The large amplitude returns at low range as shown in Fig. 6 is a result of cross-talk effect. The examination of the returns generated by the target shows that, they typically have a large magnitude return concentrated over a short section of the ping (sonar return). The speckle noise is generally present in the sonar data from a variety of sources like bottom scattering or multi-path effects. It is for this reason that

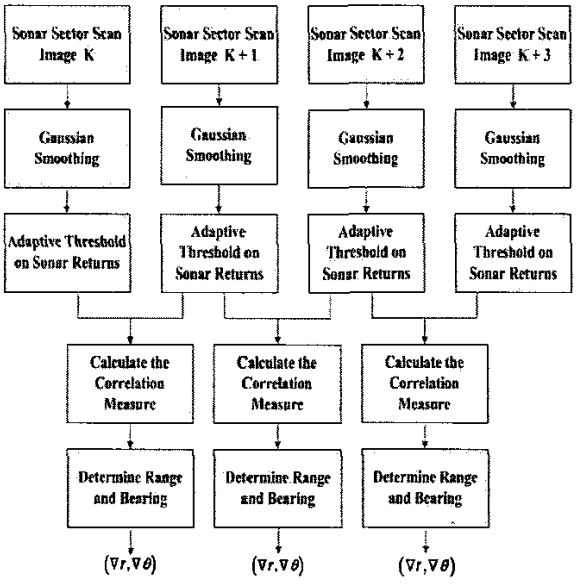

Fig. 7. Sonar Motion Estimation using Correlation measure

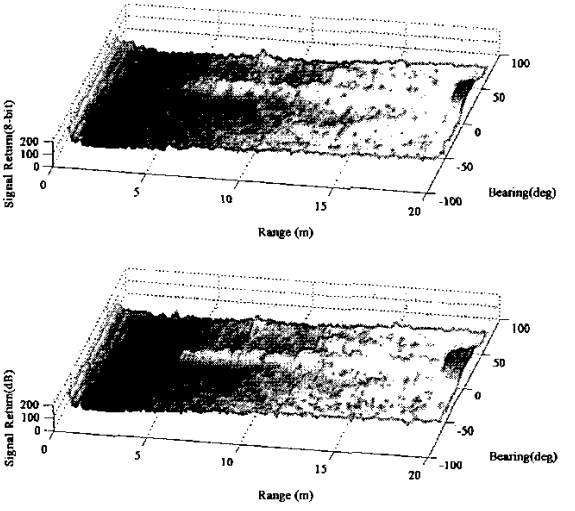

Fig. 8. Consecutive Sonar Frames: These consecutive frames are used to perform correlation

the data is low pass filtered. It is assumed that the noise is of higher frequency and uncorrelated. A Gaussian filter is employed here, which is comparable in performance with other with mean and median filters, but, at a reduced computational cost.

Adaptive threshold technique is employed, which sets a threshold for each ping based on its maximum value. Each ping is then passed through a module which determines if the intensity of the sonar returns is below a predetermined threshold (based on sonar characteristics). Such returns are discarded thereby ensuring that only the highest intensity returns will be considered. It should be noted that the threshold calculation is done before the Gaussian smoothing of the image as indicated in Fig. 7.

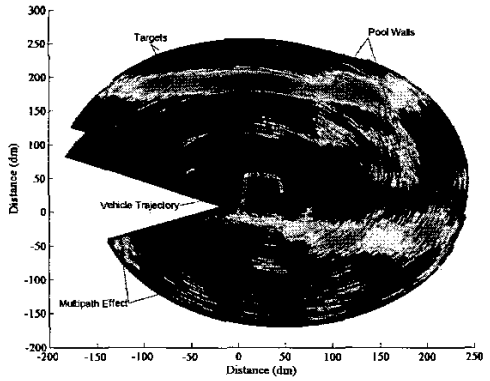

Fig. 9. Environment Map: The map indicating the vehicle trajectory obtained from the algorithm described along with the map of the environment

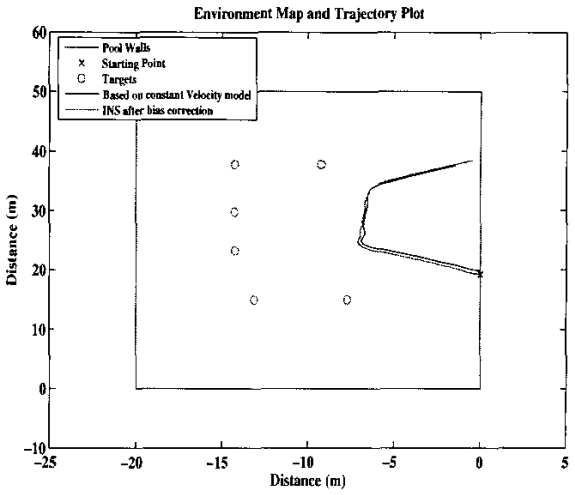

Fig. 10. The Vehicle Trajectory obtained from the INS with a map of the environment. It also depicts a trajectory obtained based on a constant velocity vehicle model

B. Motion estimation from Sonar Scans

For a given pair of consecutive sector scans as in Fig. 8, correlation measure is obatined as follows

$$y(n_1,n_2) = \sum_{i=0}^{N_1-1}\sum_{j=0}^{N_2-1} I_{k+1}[n_1,n_2]I_k[n_1+i,n_2+j] = \sum_{i=0}^{L_1-1}\sum_{j=0}^{L_2-1} I_k[n_1,n_2]I_{k+1}[n_1+i,n_2+j] \quad (6)$$

for $0 \leq n_1 \leq M_1 - 1$; $0 \leq n_2 \leq M_2 - 1$, where L_1, L_2 are the dimensions of $I_k[n_1,n_2]$, N_1, N_2 are the dimensions of $I_{k+1}[n_1,n_2]$ and M_1, M_2 are the dimensions of the result $y[n_1,n_2]$.

This measure is further employed to calculate the range and bearing difference between the consecutive scans as indicated in Fig. 7. Thus distance travelled by the vehicle between consecutive frames is estimated. Based on this estimate the consecutive scans are stitched to form a mosaic as shown in Fig. 9. The figure also shows the estimated trajectory based on the algorithm described.

IV. Integration of Sonar and INS

The IMU measures body frame accelerations and rotations allowing indicated pose to be calculated by way of suitable integrations. As for all sensors, accelerometers and gyros provided noise corrupted outputs. The integration of such signals over time to give position and velocity results in drifts, which could be modeled as Brownian [14]. Clearly then the IMU cannot be used to provide a continuously reliable pose estimate. However by resetting the errors in the inertial package periodically, given true world frame alignment information from the sonar, the inertial package can be used to provide valuable 'piecewise' reliable pose estimates when fused with the vehicle model. Figure 10 indicates the trajectory recovered solely from the INS system along with the environment map.

By using INS to determine the ego motion and combining the FLS information to build the range map, we can effectively build the map as well as localize

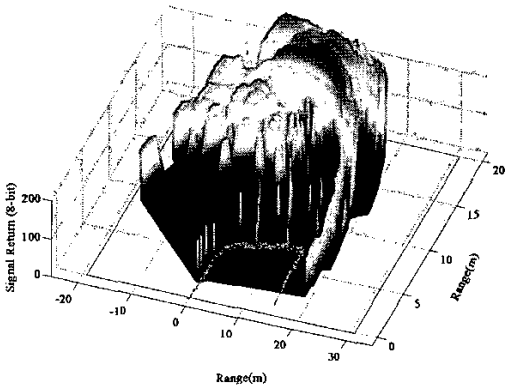

Fig. 11. Map of the environment built using the FLS and the Trajectory estimate based on INS and FLS

ourselves. Figure 11indicates the motion estimated from sonar and INS along with the sonar range map stitched together.

V. Experimental setup and implementation issues

Tests of all the algorithms are performed on a midsize underwater robotic vehicle called NTU-UAV [15] designed and built in-house as shown in Fig. 12. The experiments were mainly conducted in the pool environment. Figure 13 gives a bird's eye view of the experimental site.

A. NTU-UAV - The test bed platfrom

The vehicle is equipped with a low frequency terrain aiding forward look sonar, a color CCD camera, altimeters, depth sensor, along with a compass, motion reference unit and a Doppler velocity log. This device is intended primarily as a research platform upon which the sensing strategies and control methods are being tested. Autonomous navigation using the information provided

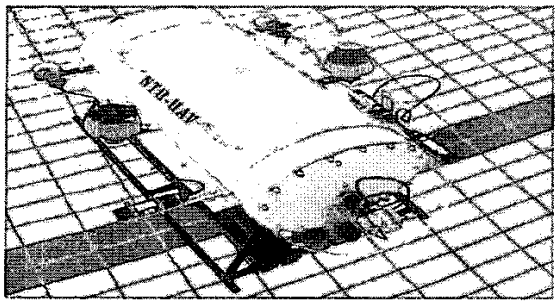

Fig. 12. Research Test Bed Platform: The NTU-UAV

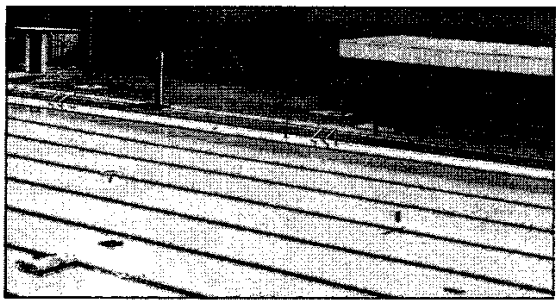

Fig. 13. Experimental Site

by the on-board sensors represents one of the main goals of the project.

B. Forward Look Sonar

The FLS mounted on the NTU-UAV is a mechanically scanned forward looking sector scan sonar. The problems associated with these is its slower update rate. Current AUV speeds are sufficiently slow for the sonar update rate not to be a hindrance. However, predicted speeds for AUVs will surely lead the way to the use of multi-beam sonars. This consists of a single hydrophone which is mechanically scanned along the horizontal axis. It produces a 3^o wide pencil beam at $0.675MHz$ which can be rotated continuously through 360^o. The working ranges are up to 50m, but, longer range will result in a slower refresh rate. Typical working range of 20m and a sector of size 135^o was scanned during our experiments.

C. Vision

The camera is mounted in an housing on the front of the vehicle as shown in Fig. 12. The camera obtains normal color images at rates up to 20 frames per second. These images are grabbed using IP7500, which is a dedicated vision processing card. Since the vehicle is moving at a very slow rate one is 5 frames is considered for the vision processing algorithm described in section II.

VI. Conclusions

The potential of the navigational algorithms has been demonstrated by the results. It has been shown that these results can be expected to improve with multiple sensors and a detailed vehicle model is integrated into the current system. The limitation of the algorithm, which needs to be addressed, is the fact that the map grows without bounds with time. Thus a multi-sensor fusion scheme is vital for a reliable state estimate.

A close look at the properties of sonar and a monocular camera clearly shows the complementary characteristics of the sensors. A sonar has good range properties, both in terms of actually measuring distances and (dependent on the frequency) good range resolution, which is indeed important in determining slow relative motion as described in section III. An underwater (monocular) camera has a better resolution than sonar and can give much more details of the observed objects. A camera is also able to extract gradients not only from physical objects but also purely visual information such as texture, which is totally invisible for sonar, can be detected. Having looked at individual sensor performances based on experimental results, this paper concludes that, the information from multiple sensors needs to be combined for more reliable position estimation and robust underwater navigation.

References

[1] Leonard.J.J, Bennett.A, Smith.C.M, and Feder.H.J.S. Autonomous underwater vehicle navigation. In *MIT Marine Robotics Laboratory Technical Memorandum*, 1998.

[2] Hall.D.L. *Mathematical Techniques in Multisensor Data Fusion.* Artech House, 1992.

[3] Urick.R.J. *Principles of Underwater Sound.* McGraw-Hill Company, 1975.

[4] Yu.C and Negahdaripour.S. Underwater experiments for orientation and motion recovery from video images. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 93–98, Atlanta, GA, May 1993.

[5] Marks.R.L, Rock.S.M, and Lee.M.J. Real-time video mosaicking of the ocean floor. In *IEEE Journal of Oceanic Engineering*, volume 20:3, pages 229–241, 1995.

[6] Balasuriya.B.A.A.P. *Computer Vision for Autonomous Underwater Vehicle Navigation.* PhD thesis, Dept. of Naval Architecture and Oceanic Enginerring, University of Tokyo,Japan, 1998.

[7] Mike Stephens Harris Chris. A combined corner and edge detector. In Matthews.M.M, editor, *4th ALVEY vision conference*, pages 147–151. Springer-Verlag, September 1988.

[8] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. *An Invitation to 3-D Vision.* Springer-Verlag, November 14 2003.

[9] Martin A.Fischler and Robert C.Bolles. Random sample and consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *Communications of the ACM*, June 1981.

[10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[11] Brutzman.D.P, Compton.M.A, and Kanayama.Y. Autonomous sonar classification using expert systems. In *IEEE Oceans*, Newport, RI, 1992.

[12] Petillot.Y, Ruiz.I.T, and Lane.D.M. Underwater vehicle obstacle avoidance and path planning using a multi- beam forward looking sonar. In *IEEE Journal of Oceanic Engineering*, volume 26:2, pages 229–241, April 2001.

[13] Lane.D.M and Stoner.J.P. Automatic interpretation of sonar imagery using qualitative feature matching. In *IEEE Journal of Oceanic Engineering*, volume 19(3), pages 391–405, 1994.

[14] Barshalom.Y and Fortman.T.E. *Tracking and Data Association.* Academic Press, 1988.

[15] Balasuriya.A, Wijesoma.S, BharathKalyan, and ThomasLim. Development of a test-bed auv: The ntu-uav. In *Second International Conference on Computational Intelligence Robotics and Autonomous Systems(CIRAS)*, Singapore, 2003.