

# A Vision Based Target Detection System for Docking of an Autonomous Underwater Vehicle

**Frederic Maire**  
School of IT, QUT,  
GPO Box 2434,  
Brisbane QLD 4001  
f.maire@qut.edu.au

**David Prasser, Matthew Dunbabin**  
CSIRO ICT Centre  
P.O. Box 883,  
Kenmore QLD 4069  
Firstname.Lastname@csiro.au

**Megan Dawson**  
Faculty of BEE, QUT  
GPO Box 2434,  
Brisbane QLD 4001  
m5.dawson@student.qut.edu.au

## Abstract

This paper describes the development and preliminary experimental evaluation of a vision-based docking system to allow an Autonomous Underwater Vehicle (AUV) to identify and attach itself to a set of uniquely identifiable targets. These targets, docking poles, are detected using Haar rectangular features and rotation of integral images. A non-holonomic controller allows the Starbug AUV to orient itself with respect to the target whilst maintaining visual contact during the manoeuvre. Experimental results show the proposed vision system is capable of robustly identifying a pair of docking poles simultaneously in a variety of orientations and lighting conditions. Experiments in an outdoor pool show that this vision system enables the AUV to dock autonomously from a distance of up to  $4m$  with relatively low visibility.

## 1 Introduction

This work has been motivated by the need to allow Autonomous Underwater Vehicles (AUVs) to perform long duration missions in challenging dynamic environments with minimum human operational support. Two primary docking scenarios have been considered. The first scenario requires the AUV to remain subsea in conditions where the water currents can exceed the vehicle's maximum speed capabilities for all but two hours either side of a changing tide. In the situation we propose a deployment and docking system where the AUV fitted with a gripper is attached to a heavily weighted target and lowered to the seafloor as shown in Figure 1.

This docking station has an acoustic beacon attached and provides a non-moving reference for the AUV to localise from. When the water currents are favourable, the AUV releases from the target and conducts its mission until such time that it determines it unsafe to continue. At that stage it uses its localisation system, aided by

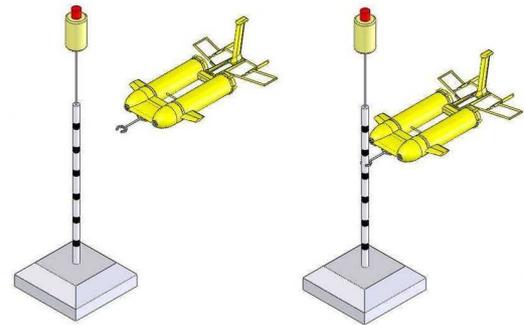


Figure 1: Proposed docking system for use in high-current and/or deep sea operations when it is desired not to return to the surface (on approach (left), and after docked(right)).

the acoustic beacon on the target to home in until visual range is detected. Then using the vision system, it attaches to the target and powers down to a sleep state until tides are again favourable.

The second scenario relates to the automated deployment and collection of the AUV by other robotic vehicles. In previous work [Dunbabin *et al.*, 2008], it was proposed to use a vision system attached to an Autonomous Surface Vehicle (ASV) and move the boat to collect an AUV at the surface. However, in certain sea-states and other operating conditions, it is conducive to have the AUV dock to a system attached to the surface vessel whilst still underwater. In this work, we consider the situation where a cage is suspended underneath a surface vessel as shown in Figure 2. Here the AUV is required to align itself within the cage using the docking poles and once within the cage it is lifted along with the AUV out of the water.

The platform considered in this study is the Starbug AUV [Dunbabin *et al.*, 2005] as shown in Figure 3. It has two stereo camera pairs, one facing downward for odometry, and the other forward for obstacle avoidance and docking. The vehicle is capable of powering down to



Figure 2: Conceptual diagram of AUV docking system slung underneath the CSIRO developed Nikki Autonomous Surface Vehicle.



Figure 3: The Starbug MkIII AUV used for experimental studies.

a sleep state to allow minimum power consumption during periods of non-activity. Therefore, using the vision system onboard the vehicle, a docking target that allows both scenarios is a binary encoded pole which serves as a localisation target as well as structural support for the AUV to grasp. The requirements for the docking system are that it must be detectable from any angle and reasonable roll/pitch orientation. Also the system should at a minimum operate effectively with monocular vision allowing a level of redundancy on the AUV but can also work with stereo cameras.

### 1.1 Related work

Target identification and homing has been widely published in the literature with a dominant motivation being the ability to dock with specifically designed moorings for recharging, data upload and inspection [Singh *et al.*, 1998]. Some examples of docking systems use acoustic

homing to obtain a range and heading estimate to the target [Stokey *et al.*, 1997; McEwen *et al.*, 2008]. A similar approach was applied by Freezor [Feezor *et al.*, 2001] using electromagnetic guidance.

Combining vision and acoustics to improve docking performance has been proposed by Evans *et al.* [Evans *et al.*, 2003] where acoustic homing is used until within visual range of the target and using vision-based target identification for the final stages of docking. Initial vision-based target identification schemes relied on active targets to provide distinct points of interest in the scene [Cowen *et al.*, 1997; Lee *et al.*, 2003; Wang *et al.*, 1995]. In the applications proposed here, it was decided that active targets would be too expensive and require power and less effective in regions where significant lighting variations occur.

In the domain of passive underwater targets, the use of colour based identification has been demonstrated [Yu *et al.*, 2001; Dunbabin *et al.*, 2006], however, depending on lighting conditions, their effectiveness can deteriorate with range to the target. Therefore, it was proposed to use other forms of black and white targets. Negre *et al.* [2008] described a method of monocular vision target identification that allowed range estimation using self-similar landmarks (SSL). This method was rotationally invariant and provided robustness to variations in camera model, distortion and observation range, as well as precision docking capabilities when close into the target. However, using SSL is computationally too costly for Starbug.

In this work it is proposed to use another form of passive black and white target, an encoded pole (see Figure 5). This allows uniform detection from any approach angle as well as a structural component to grasp.

Section 2 introduces the pole detection algorithm. Section 3 describes the proposed docking control module. Experimental results are presented in Section 4.

## 2 A Fast Pole Detection Algorithm Based on Haar Rectangular Features

### 2.1 Target Design

To design the targets, several options were contemplated as the candidate targets could potentially utilise any pattern, shape or colour. Colour was initially considered, but dismissed as in water the longer wavelengths of sunlight are attenuated rapidly with depth. Moreover, glare effects and lighting variations are likely to be common in practice and make colour identification unreliable.

The target was required to be detectable from a tilted orientation. Starbug would approach the target with a relative depth difference of less than 2 metres and its pose would not necessarily be perfectly horizontal.

After investigating various candidate computer vision

algorithms and taking into account the navigation constraints, striped poles were selected as the targets.

## 2.2 Initial Attempts

One of the first pole detection solutions that was developed and tested consisted of a cascade of classical computer vision functions, followed by an application specific search function. More precisely, after applying a Gaussian filter and contrast stretching using histogram equalisation, a Hough transform of the input image was performed, then candidate scan lines for the poles were identified by looking of the ratio of dark and light segments. Multiple classes of targets could be distinguished by using distinct ratios.

One of the drawbacks of this approach was having to tune several parameters. Unfortunately, changes in lighting conditions require adapting these parameters. However, the main drawback was that the Hough transform was computationally too expensive for the embedded computer to run in near real time. The AUV's Nano-ITX embedded computer on which the computer vision module runs is roughly five times slower than a standard desktop PC.

A second approach (computationally more efficient) based on rectangular patch segmentation and clustering by alignment was investigated. The various segmentation methods implemented in the OpenCV library were tested in turns. This second approach led to a program that could run at a few frames per second, and could handle more noise in the image and more changes in lighting conditions than the first Hough transform based approach. However, this second approach was still judged not robust enough to changes in lighting conditions.

The inspiration of the solution we finally adopted comes from the Viola-Jones face detection algorithm [Viola and Jones, 2001].

## 2.3 Pole Detection with Haar-like Features

The simple rectangular Haar-like feature that we use is the difference of sums of pixel values of rectangular areas, which can be at any position and scale within the original image [Viola and Jones, 2001]. One of the contributions of Viola and Jones was to use *Integral Images* for the fast computation of Haar features. Integral images can be defined as 2-dimensional lookup tables in the form of a matrix with the same size as the original image. Each element  $s(x, y)$  of the integral image contains the sum  $s(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$  of the values  $i(x', y')$  of the pixels located in the axis parallel rectangle with top-left corner  $(0, 0)$  and bottom-right corner  $(x, y)$ . The integral image allows computing any sum of a rectangular area in the image, at any position or scale, using only 4 lookups. For example, in Figure 4, the sum of the pixel values in the grey rectangle is  $s(D) - s(C) - s(B) + s(A)$ .

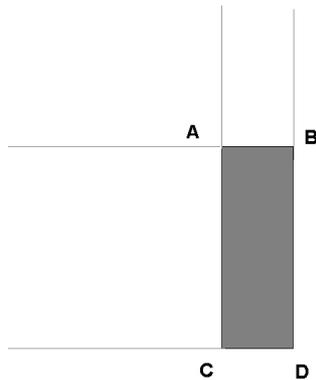


Figure 4: Integral Image.

The stripe pattern that we selected is of the form *dash dot dash dot dash dot*, where a dash band is twice as long as a dot band. The two types of poles that we use can be seen in Figure 5. The left pole is of type white because it has white dashes and black dots, whereas the pole on the right is of type black because it has black dashes and white dots. The benefit of using these patterns is that the two poles can be detected with the same template. Consider a rectangular template *dash dot*, and let  $s_{\text{dash}}$  and  $s_{\text{dot}}$  be the sums of the pixel values respectively in the areas *dash* and *dot* of the template. The weighted sum  $s_{\text{dash}} - 2 \times s_{\text{dot}}$  will be zero on a uniform background, will take a large positive value on a pattern *white-dash black-dot*, and will take a large negative value on a pattern *black-dash white-dot*.

The white poles generate a large positive response whereas the black poles generate a large negative response.

Our core function scores the signal strength of a *dash dot dash dot dash dot* pattern in a vertical position at a given scale and position. This test requires 14 accesses to the integral image. To accommodate a relative tilt of the poles with respect to Starbug, we perform rotations of the input image.

The processing of an input image requires the execution of 3 nested loops. The outer loop varies the rotation angle of the image, the next loop varies the scale of the template, and the most inner loop varies the position of the top left corner of the template. Figure 6 shows the output of the scoring function applied to Figure 5, using the appropriately scaled template and after rotating the image so the targets are vertically orientated. Bright areas are locations with a strong *white-dash black-dot* pattern while dark areas have a strong *black-dash white-dot* pattern. The strongest values correspond to the detected

target locations shown in Figure 7.

The green vertical bars in Figure 7 represent the different scales searched, whereas the fan-like bundle of green rays on the bottom right of the image indicate the rotation angles.

Rotating an image is a computationally expensive operation. A significant speed up was obtained by caching in auxiliary images the destination of each pixel. This is a trade-off between memory and speed.

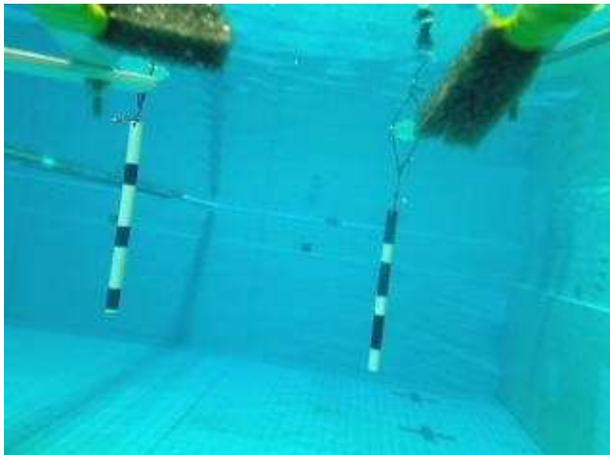


Figure 5: The docking poles as viewed by the AUV. A passive gripper mechanism is visible at the top of the image.



Figure 6: Output of the detection function. Bright patches are possible locations of the white pole while the dark areas correspond to the black pole.

## 2.4 Pole Tracking at Close Range

As the AUV gets closer to the target, we have to switch templates first from *dash dot dash dot dash dot* to a *dash dot dash dot*, then from *dash dot dash dot* to *dash dot*, and finally from *dash dot* to *dash*. The timing of the switching is decided by the Control System Module described in the next section.

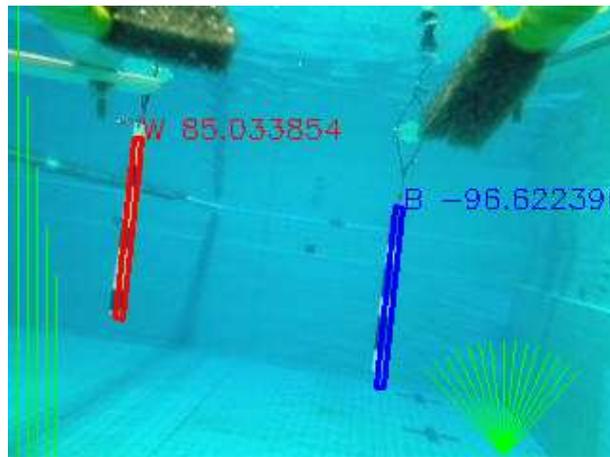


Figure 7: Detected Poles. Colours denote the type of pole.

## 3 Docking Control Module

The docking controller connects the vision system with the vehicle's low level controller and is responsible for searching for the target, manoeuvring the vehicle to the docking target and detecting that the docking has been completed.

There are a number of assumptions that can be made when designing the control module. Firstly, a higher level navigation system, such as an acoustic homing behaviour, must be responsible for bringing the AUV within visual range of the docking target.

The AUV is non-holonomic and is able to control its roll, pitch, heading, depth and velocity, it lacks lateral control authority. For most docking applications it is reasonable to assume that the docking target's depth is known (either preprogrammed or provided dynamically via an acoustic modem), especially since the docking mechanism can accommodate a large error in depth. Similarly the target roll and pitch for docking to a vertical pole should be zero, assuming that the docking mechanism on the AUV has no special requirements. From these assumptions, three of the vehicle's degrees of freedom can be controlled to constant values and it is only necessary to use the vision system to control the AUV's velocity and heading. Further, it is the bearing to the target that is the main input to the control system.

The control system must take into account the latency of the vision system. The current global heading of the target relative the AUV is calculated using the delayed, vehicle relative bearing from the vision system and a time-stamped history of the vehicle's heading. All of the docking control operates in a global heading space measured by the AUV's compass. It is not required that the compass is necessarily correctly aligned with North, merely that there is a constant reference point.

There are three stages of docking: search; alignment; and approach. Transition between these stages is governed by a state machine. The search stage slowly rotates the AUV and transitions to the alignment state when the target is positively identified. In the alignment state the AUV's heading is controlled to equal the target's heading. Once the heading has stabilised the controller transitions to the approach state. The approach state implements the same control rule as the alignment state while traveling forwards at 0.2 m/s. A short range infra-red obstacle detector is used to signal the completion of the docking process. If the target is lost for more than two seconds the latter two states revert to the search state.

The docking controller also controls how many elements of the target's repeating *dash-dot* pattern should be searched for by the vision system. While in the approach state the docking controller continuously adjusts the number of patterns to search for as the pixel size of the target changes.

There are several improvements still to be made to the docking controller. At the moment there is no explicit accounting for external disturbances such as those caused by currents. In addition, for the surface vessel docking concept, there may be a requirement to approach the dock at a particular orientation to avoid some other part of the surface vessel. This would require a quite different control strategy, either blind travel to a better position or alternatively crabbing the AUV around the target, while keeping it in view.

## 4 Experimental Results

We carried out two testing sessions in the water. The first session took place in a large dive pool ( $25m \times 25m \times 5m$  deep). During this first water session, a visual data set was collected by moving the AUV in the dive pool. Another data set was collected on dry land to examine the computer vision system performance in a laboratory room where the lighting conditions could be varied.

The second testing session in the water took place in an outdoor pool with natural sunlight and  $< 8m$  visibility. The whole system was tested in this session, and the AUV successfully performed docking operations from a distance of up to  $4m$ .

### 4.1 Dive Pool Experiment

The True Positive Rate (TPR) reflects the detection performance of the pole detection algorithm. The TPR is the ratio of correctly detected poles among all images with poles in them. The False Positive Rate (FPR) is the frequency of how many times the algorithm incorrectly detects a pole among all images with no poles in them.

The response to a template is a real value (continuous output). The boundary between the two classes (pole, no-pole) must be determined by a threshold value. A ROC space is defined by FPR and TPR as the  $x$  and  $y$  axes respectively. The ROC plot depicts relative trade-offs between true positive and false positive rates.

The video frames collected in a dive pool (800 images) were annotated manually with the presence or not of a pole in the proper distance range and tilt range.

Figure 8 shows the ROC curve for the detection of the black pole over this data set. For the threshold value of  $\pm 72.07$ , the TPR is 0.92 and the FPR is 0.04. This threshold value corresponds to the point closest to the point (0,1) in the ROC figure (the point closest on the curve to the top left corner). This point is in some sense an optimal trade-off value for the threshold with both false positives and false negatives being equally weighted, as the point (0, 1) would be ideal.

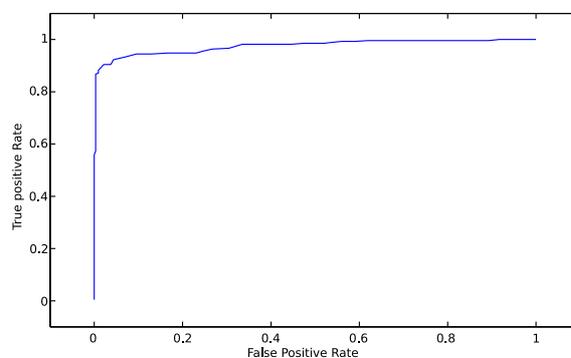


Figure 8: Receiver Operating Characteristic (ROC) curve.

Whenever a pole is present in an image collected in the dive pool, the pole is detected. Even if the image is blurred by motion.

### 4.2 Docking Experiments

A series of docking trials were successfully conducted using both the vision system and the docking controller in an outdoor pool with natural lighting. An example docking run from these experiments is shown in Figure 9. This figure shows the AUV's heading and the heading to the target when it is detected by the vision system. Also shown is the largest response of the target detector over the course of the run.

### 4.3 Lighting Variation Experiments

The laboratory experiments test the range of lighting conditions that can be accepted. Figure 10 shows that even with little ambient light, the poles are detected and

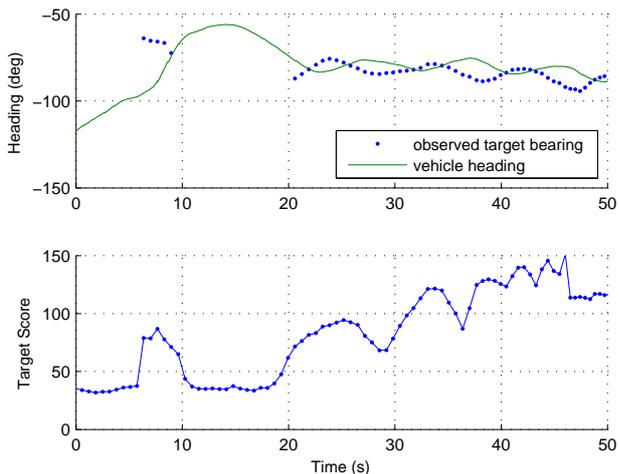


Figure 9: Top, example vehicle and target heading during a docking run. Bottom, best target score during the same run. The target score threshold was set at 65.

correctly identified. However, if there is too much reflection on the black patches, the system can fail as illustrated in Figure 11. While the targets are still the strongest feature in the image they are detected with a response level well below the threshold used in the docking experiments. This problem could be compensated in part with brightness control [Negre *et al.*, 2008].

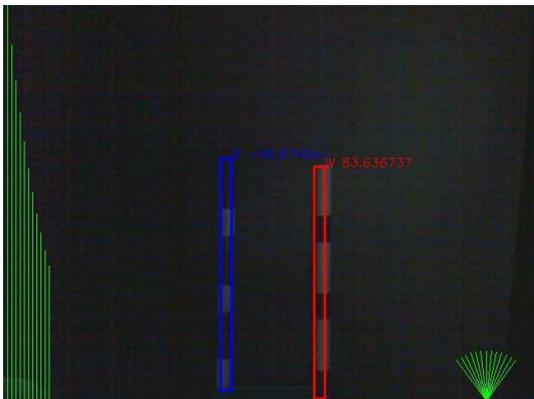


Figure 10: Dim light condition. Targets are correctly identified.

#### 4.4 Processing Time

Table 1 shows the measured time to process one frame with the vision algorithm using the AUV's onboard processor (a Via C7 processor at 1.5 GHz) and, for comparison, an Intel Core 2 Duo running at 2.8 GHz. These figures are for processing  $320 \times 240$  at 15 different orientations and 5 different template scales. The main loading

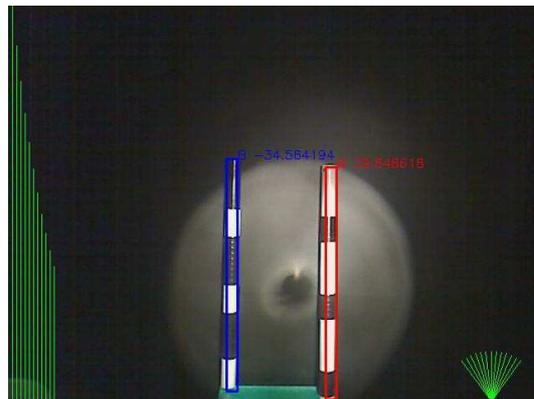


Figure 11: Single spotlight condition. Targets are identified but with low confidence.

Processing Stage	AUV	Desktop
Image Rotation	111	20
Integral Image	66	10
Feature Computation	244	48
Locate Maximum	21	4
Total Time	442	82

Table 1: Processing time in milliseconds.

of the algorithm is in looking up the integral image to compute the feature transform.

## 5 Conclusion

This paper addresses robust vision-based target recognition by presenting a novel marked pole detection algorithm based on techniques that have been very successful in face detection.

The speed of the algorithm comes from the use of integral images and cached rotation mappings. This patch based approach is more robust to noise and changes in lighting conditions than edge based approaches because it considers larger chunks of the image.

Experimental results show that vision system performs very well with limited processing power. The combined vision and controller systems enable robust pole detection and docking in an outdoor pool environment and should work in a wide range of operating conditions. Future experiments will test the integrated system in ocean conditions.

## Acknowledgements

This work was completed while the first author was visiting the Autonomous Systems Laboratory at CSIRO QCAT in the first half of 2009.

## References

- [Cowen *et al.*, 1997] S. Cowen, S. Briest, and J. Dombrowski. Underwater docking of autonomous undersea vehicles using optical terminal guidance. In *Proceedings of MTS/IEEE OCEANS97*, volume 2, pages 1143–1147, 1997.
- [Dunbabin *et al.*, 2005] M. Dunbabin, J. Roberts, K. Usher, G. Winstanley, and P. Corke. A hybrid AUV design for shallow water reef navigation. In *Proc. IEEE International Conference on Robotics and Automation*, pages 2105–2110, 2005.
- [Dunbabin *et al.*, 2006] M. Dunbabin, P. Corke, I. Vasilescu, and D. Rus. Data muling over underwater wireless sensor networks using an autonomous underwater vehicle. In *Proc. IEEE International Conference on Robotics and Automation*, page 20912098, 2006.
- [Dunbabin *et al.*, 2008] M. Dunbabin, B. Lan, and B. Wood. Vision-based docking using an autonomous surface vehicle. In *In Proc. International Conference on Robotics and Automation*, pages 26–32, 2008.
- [Evans *et al.*, 2003] Jonathan Evans, Paul Redmond, Costas Plakas, Kelvin Hamilton, and David Lane. Autonomous docking for intervention-AUVs using sonar and video-based real-time 3D pose estimation. In *Proceedings of OCEANS 2003*, volume 4, pages 2201–2210, 2003.
- [Feezor *et al.*, 2001] M.D. Feezor, F. Yates Sorrell, P.R. Blankinship, and J.G. Bellingham. Autonomous underwater vehicle homing/docking via electromagnetic guidance. *IEEE Journal of Oceanic Engineering*, 26(4):515–521, 2001.
- [Lee *et al.*, 2003] Pan-Mook Lee, Bong-Hwan Jeon, and Sea-Moon Kim. Visual servoing for underwater docking of an autonomous underwater vehicle with one camera. In *Proceedings of OCEANS 2003*, volume 2, pages 677–682, 2003.
- [McEwen *et al.*, 2008] R.S. McEwen, B.W. Hobson, L. McBride, and J.G. Bellingham. Docking control system for a 54-cm-diameter (21-in) AUV. *Oceanic Engineering, IEEE Journal of*, 33(4):550–562, Oct. 2008.
- [Negre *et al.*, 2008] Amaury Negre, Cedric Pradalier, and Matthew Dunbabin. Robust vision-based underwater homing using self-similar landmarks. *Journal of Field Robotics*, 25(6-7):360–377, 2008.
- [Singh *et al.*, 1998] H. Singh, S. Lerner, K. von der Heyt, and B.A. Moran. An intelligent dock for an autonomous ocean sampling network. In *Proceedings of OCEANS98*, volume 3, pages 1459–1462, 1998.
- [Stokey *et al.*, 1997] R. Stokey, M. Purcell, N. Forrester, T. Austin, R. Goldsborough, B. Allen, and C. von Alt. A docking system for REMUS, an autonomous underwater vehicle. In *Proceedings of MTS/IEEE OCEANS 97*, volume 2, pages 1132–1136, 1997.
- [Viola and Jones, 2001] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE CVPR*, 2001.
- [Wang *et al.*, 1995] H.H. Wang, S.M. Rock, , and M.J. Lees. Experiments in automatic retrieval of underwater objects with an AUV. In *Proceedings of MTS/IEEE OCEANS 95*, volume 1, pages 366–373, 1995.
- [Yu *et al.*, 2001] Son-Cheol Yu, T. Ura, T. Fujii, and H. Kondo. Navigation of autonomous underwater vehicles based on artificial underwater landmarks. In *Proceedings of MTS/IEEE OCEANS 01*, volume 1, pages 409–416, 2001.