Robust vision-based underwater homing using self-similar landmarks

Amaury Nègre, Cédric Pradalier, Matthew Dunbabin

Robust Vision-based Underwater Homing Using Self-Similar Landmarks

Amaury Negre,
INRIA Rhone Alpes, eMotion
Av. de l'Europe, Montbonnot, France
Email: firstname.lastname@inrialpes.fr

Cedric Pradalier and Matthew Dunbabin
CSIRO ICT Centre, Autonomous Systems Lab,
PO Box 883, Kenmore QLD 4069, Australia,
Email: firstname.lastname@csiro.au

Abstract

Next generation Autonomous Underwater Vehicles (AUVs) will be required to robustly identify underwater targets for tasks such as inspection, localization and docking. Given their often unstructured operating environments, vision offers enormous potential in underwater navigation over more traditional methods, however, reliable target segmentation often plagues these systems. This paper addresses robust vision-based target recognition by presenting a novel scale and rotationally invariant target design and recognition routine based on Self-Similar Landmarks (SSL) that enables robust target pose estimation with respect to a single camera. These algorithms are applied to an AUV with controllers developed for vision-based docking with the target. Experimental results show that system performs exceptionally on limited processing power and demonstrates how the combined vision and controller systems enables robust target identification and docking in a variety of operating conditions.

1 Introduction

Target identification and homing are of particular interest in our research as we desire an Autonomous Underwater Vehicle (AUV) fitted with a camera and vision processing capabilities to accurately locate and home onto points of interest in relatively cluttered environments such as coral reefs, or in close proximity to subsea oil and gas structures. Due to issues such as multi-pathing and variable lighting, the performance of typical identification and homing methods such as acoustics or feature and color-based vision systems can degrade significantly. The ability to robustly identify objects in these environments can allow tasks such as reliable localization, inspection, object collection and docking to complex structures.

Underwater targets for docking/homing can be either active or passive and are typically identified using either acoustics or vision systems. Active (energy emitting) targets are most common in literature due to their generally larger detection range, whereas passive targets can be further classified as artificial or natural and are typically detected using vision systems at closer distances.

Acoustics and vision are the most common ways of identifying and homing onto targets. Acoustics are advantageous in that they can operate over long ranges and in a variety of water conditions (visibility and lighting). However, for docking, acoustics degrade in cluttered environments such as reefs or close to the seafloor and subsea structures. Vision-based target identification has complimentary properties. Vision is suitable for close range tracking and can accommodate changing and multiple targets, however, its performance degrades in turbid water and poor lighting conditions.

1.1 Related Work

There are many examples of target identification and homing in the literature. A dominant motivation has been to enable an AUV to dock with a specially designed mooring, such as that proposed by Singh (Singh et al., 1998), allowing remote data upload and recharging for extended operations. The simplest approach for autonomous docking is to home to an acoustic target, possibly located on a mooring, using range and heading information from the target (Stokey et al., 1997). Although a reliable method at larger distances, at close ranges the performance becomes impractical for docking due to the high update rates required. Alternatively, Feezor (Feezor et al., 2001) extend this concept by successfully demonstrating homing and docking using electromagnetic guidance in which an AUV was fitted with coils to sense field strength and orientation that enabled docking from a distance of 30m.

Improvements in the docking performance of AUVs have been considered by combining acoustics with vision (Evans et al., 2003). Here, the acoustics/sonar provides the longer range homing direction with the vision providing guidance in the final stages of docking. A survey of vision-based target identification and tracking literature is provided by Dalgleish (Dalgleish et al., 2005) describing the optic flow and feature based techniques commonly used for underwater tracking applications.

Initial work in vision-based target identification schemes relied on active targets to provide a distinct point of interest in the image for tracking. For example, Cowen (Cowen et al., 1997) demonstrated an active vision-based target tracking system for terminal guidance, that is the final 10-100m of the vehicle's docking manoeuvre. Cowen (Cowen et al., 1997) proposed a method whereby the image plane was segmented into quarters and the controller designed to force the received light signal from the target to the centre of the image. In later works, Lee (Lee et al., 2003) used a monocular vision system to identify a large circular target with an LED ring and 5 large lights, whereas Wang (Wang et al., 1995) describes a stereo vision system to locate and retrieve an underwater object which has a light emitting beacon on its surface.

A particular challenge is the tracking and homing onto natural or non-specific targets (e.g. a particular pylon on a pier). This has been addressed in a number of ways. Early research by Marks (Marks et al., 1993) used image filtering and correlation to track non-specific items in both monocular and stereo images. These items to be considered targets were based on their estimated closeness to the camera. Later work such as Fan (Fan and Balasuriya, 2000) describes a monocular vision system in which the targets of interest are extracted based on their dynamic and optical features. Wettergreen (Wettergreen et al., 1998) describes a

stereo vision system for the robust target identification and vehicle control based on feature template matching. Furthermore, Cufi (Cufi et al., 2002) described a system based on natural feature selection and textural correlation to build a mosaic of the AUV's images to allow hovering over a particular point on the seafloor.

The use of color for identification of a passive target has been investigated by Yu (Yu et al., 2001). The use of color underwater has inherent problems with color consistency and appearance changes due to preferential absorption of red wavelengths. Yu addresses color changes by the use of lookup tables in UV space and a predominantly yellow target. Similarly, (Dunbabin et al., 2006) considers color-based identification of underwater network nodes based on lookup tables in normalized RG space.

1.2 Motivation

Each of the methods described above (acoustics, electromagmetics, active and passive vision) has relative merits and disadvantages depending on the operational circumstances. Our application context is environment monitoring, and especially the biological monitoring of Australia's Great Barrier Reef. As a consequence we are looking for a solution suitable for clear water (3 to 10+ meters visibility) with low current. We are also assuming that a complementary long range homing is available – either GPS or acoustics – and we are only considering the short range homing task.

Considering our requirements and all these methods of target identification and homing, it was decided that our system would require passive targets due to energy requirements on behalf of the AUV and target itself for long duration operations. Additionally, we decided to evaluate whether visible targets would be suitable for our task and environment.

When repeated visitations of a target over extended periods of time are required, natural landmarks can change significantly. As a consequence, we chose artificial landmarks. However, detection is required at varying orientation – within a conic region at least 120 degree wide – and for relatively large distances – up to 5 meters, and changing light conditions. Colour-based methods are unreliable because of this latter requirement associated with the fact that water absorbs different colours non-equally. Therefore, the target and identification routine must be robust, scale and rotationally invariant, as well as capable of running in real-time on an AUV's limited processing power.

1.3 Contribution

This paper evaluates experimentally the use of Self-Similar Landmarks (SSL) (Briggs et al., 2000) as a robust, color, scale and rotational invariant means of target identification from which the information can be used to guide an AUV for short-range homing and docking operations. SSL Landmarks has been shown to be efficient landmarks for indoor ground robot operations. To the author's knowledge, this is the first time they are evaluated in the context of an outdoor underwater vehicle, combining difficult lighting conditions, a 6 degrees-of-freedom (DOF) platform and unreliable colour perception.

1.4 Paper outline

The remainder of this paper is structured as follows: Section 2 describes the properties of Self-Similar Landmarks and our modification to make them rotationally invariant and robust to various transformations. Section 3 presents the target identification routine and method of estimating target pose with respect to the AUV and Section 4 describes some practical limitations of using SSL in field robotics. Finally, Section 5.1 presents experimental results showing the performance of the target identification and AUV homing.

2 Self-Similar Landmarks

2.1 Definition of Self-Similar Function

The notion of self-similar landmarks was first used in a robotic context by Briggs et al.(Briggs et al., 2000). Their objective was to develop planar targets that would be detected easily with a standard perspective camera on a mobile indoor robot. To this end, the targets were designed to be invariant to change of scale. This is where the self-similarity is essential.

A p-similar function for $0 < p < 1$ is a function $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ such that $\forall x > 0, f(x) = f(p \cdot x)$. Even if not strictly required, we will only consider in the following p-similar functions from $\mathbb{R}^+$ to $[0 - 1]$.

It is essential to note that the p-similarity is invariant to change of scale ($\forall x > 0, \forall k > 0, f(k \cdot x) = f(p \cdot k \cdot x)$). In the context of a computer vision application, this property is interesting since, if a method to detect self-similarity is available, it will be possible to do so from any distance – as long as the resolution provides enough information on the observed object – because a change of range results in a change of scale in an image acquired by a perspective camera.

When designing a self-similarity detector, the fact that all constant functions are p-similar for all p is problematic since any uniform region in an image will give a strong response on the self-similarity detector. To solve this, (Briggs et al., 2000) introduces the notion of anti-similarity: a function f is p-antisimilar if $\forall x > 0, |f(x) - f(p \cdot x)| = 1$ (this assumes that, $\forall x, f(x) \in [0 - 1]$). This constraint means that a high contrast is required between the point x and $p \cdot x$.

The p-similarity and $\sqrt{p}$-antisimilarity are then used by Scharstein and Briggs to define the "Self-similar square wave" function which is both p-similar and $\sqrt{p}$-antisimilar:

$$s_p(x) = \lfloor 2(\log_p x - \lfloor \log_p x \rfloor) \rfloor \tag{1}$$

Where "$\lfloor x \rfloor$" is the function "$floor(x)$" and "$\log_p x$" is the logarithmic function in the base p. This function maps $\mathbb{R}^+$ to $\{0, 1\}$. It is used to print black and white landmarks, as shown in Figure 1(a).

To recognize the self-similar landmarks, (Briggs et al., 2000) defines a matching function that will respond strongly to a sequence of pixels looking like a p-similar and $\sqrt{p}$-antisimilar

function. For a one dimensional function I, the matching function at point x is computed by integrating the p-similarity and $\sqrt{p}$-antisimilarity conditions on a window of length w starting at x :

$$\begin{aligned} m^w(x) &= \frac{1}{w}\int_0^w |I(x+\xi) - I(x+\sqrt{p}\cdot\xi)|\, d\xi \\ &- \frac{1}{w}\int_0^w |I(x+\xi) - I(x+p\cdot\xi)|\, d\xi \end{aligned} \tag{2}$$

In this equation, the window is made of the w pixels after x. In practical implementations, w takes values in the $[20, 100]$ pixel range. As in (Briggs et al., 2000), we use $w = 40$ as a good balance between robustness and computation cost. It should be noted that in discrete images, the integral is approximated by a summation over the pixels.

Given a direction $\vec{u}$, the matching function can be extended to a two dimensional function like an image. The 2D matching function for the direction $\vec{u}$ and for the pixel $X = (x, y)$ is equal to the 1D matching function over the line defined by the point X and the vector $\vec{u}$:

$$\begin{aligned} m^w_{\vec{u}}(X) &= \frac{1}{w}\int_0^w |I(X+\vec{u}\cdot\xi) - I(X+\sqrt{p}\cdot\vec{u}\cdot\xi)|\, d\xi \\ &- \frac{1}{w}\int_0^w |I(X+\vec{u}\cdot\xi) - I(X+p\cdot\vec{u}\cdot\xi)|\, d\xi \end{aligned} \tag{3}$$

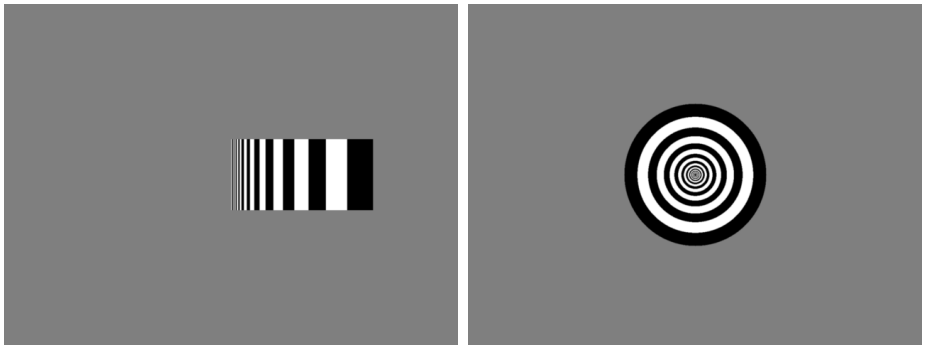
In the following sections we denote m^w_{+x} m^w_{-x}, m^w_{-y} and m^w_{+y} the matching function in the direction $(1, 0)$, $(-1, 0)$, $(0, 1)$ and $(0, -1)$ respectively. These four functions are easily computed in 2D images as we only have to scan image lines or columns.

2.2 A Modified SSL for Pose & Range Estimation

2.2.1 Circular landmark

The two-dimensional landmark used by (Briggs et al., 2000) is self-similar in the horizontal direction and constant in the vertical direction as shown in Figure 1(a). This landmark is not optimal for many practical applications, particularly underwater on vehicles capable of 6DOF movement:

- This is not a point landmark and consequently it is hard to localize precisely in the image. In (Briggs et al., 2000), the border of the landmark is detected, and a binary coding is used to identify the landmark. Our practical experience in outdoor experiments found this to be the weakest part of the algorithms. It was found that the SSL was always detected correctly, however, small blurs on the edges often prevented correct identification of the binary code, especially in the presence of rotation or slant.
- This brings us to our second point, this landmark is not enough robust to various transformations: a motion blur or a perspective transformation in the Self-Similar



(a) (b)

Figure 1: Self-Similar Landmarks (SSL): (a) Original linear SSL as used in (Briggs et al., 2000), and (b) our proposed modified circular landmark.

pattern direction affect considerably the detection. Thanks to the multi-directional matching function, the rotation invariance can be performed but can still be improved.

In order to solve those problems, we designed a circular landmark, shown in Figure 1(b), where the intensity I is self-similar and anti-similar in all directions such that

$$I(\rho, \theta) = s_p(\rho) \tag{4}$$

where ρ, θ are the polar coordinate and s_p the self-similar square wave function defined in Equation(1).

This landmark is interesting because it is rotationally invariant thanks to the circular geometry and the matching function exhibits a single maximum point in the center. Moreover, to improve robustness to noise and motion blur, we can apply the matching function in several directions like top, bottom, left and right. In this case the resulting matching function is then:

$$M^w(x, y) = \max\{m^w_{+x}(x, y), m^w_{-x}(x, y), m^w_{+y}(x, y), m^w_{-y}(x, y)\} \tag{5}$$

With the aim of studying the stability of landmark detection to various transformations, we realized the following experiments. For each transformation, we artificially applied the transformation to an image of the landmark and evaluated the resulting matching function, plotted as a function of the transformation parameter with the results shown in Figure 2.

Four kind of transformations have been considered: (1) a rotation in the image plane, (2) a perspective transformation (characterized by the angle between the camera projection axis and the normal vector of the landmark's plane), (3) a directional Gaussian blur in the horizontal direction and in a diagonal direction (equivalent to a motion blur), and (4) a contrast change. As can be seen in Figure 2, our modified landmark is – as expected – rotationally invariant. The original landmark's response fall down around 50 degrees with a monodirectional matching function, but give good response with the four-directional function. The

original landmark gives good results to perspective transforms under 50 degrees, whereas the modified target due to the vertical symmetry is robust to larger viewing angles (see top two rows of Figure 2). For an directional blur in the direction of the original landmark (third row of Figure 2), the matching function decreases rapidly for the original landmark. However, for the circular landmark, the matching functions orthogonal to the blur are not affected by this blur and consequently the matching score keeps a high value. Nevertheless, for a 45 degrees directional blur, the two landmarks give the same results, the matching function decrease moderately. Finally, as can be noticed in Equation (3), the matching function is proportional to the image intensity. This means that the image intensity must be normalized before evaluating the matching function (see Section 5). The weakness of the circular landmark is the sensitivity to occlusion just at center, the linear one can handle better occlusion on the edge of the pattern. However, Figure 3 shows that the circular landmark can handle a significant amount of occlusion while still being identifiable.

2.2.2 Three patterns landmark

In our application we also need to estimate the range and the pose of the target to allow the AUV to track and home onto it. Range could be estimated using a single circular landmark and stereo camera system. However, detection of distant objects requires large camera baselines and this also doubles the image processing requirements. In this research, a monocular vision was chosen to reduce computation burden and allow portability to other platforms. As we can detect only the center of the pattern we need at least 3 patterns to estimate the pose (it is well known that at most 8 poses will be consistent with the observation (Faugeras, 1993; Hartley and Zisserman, 2000), but it is in general possible to eliminate the ambiguities). Our final target, shown in Figure 3(a), consists of 3 circular self-similar patterns on an equilateral triangle: one large pattern for detection at greater distances and two small patterns for the pose estimation at short range.

3 SSL Target Identification & Tracking

3.1 Target Detection

The first step in the detection algorithm consists of applying the matching function on every pixel of the image in four perpendicular directions (see Equation (5)). Next, the location of the local maxima of the matching function is determined with a threshold function applied to remove small outlier peaks.

The results and a typical response of the matching function with the three-pattern landmark are shown in Figure 3(b). In our testing, this target appeared robust to occlusion (as long as the center point is visible), deformation resulting from bending or shaking the paper landmark, camera model (the same detection software worked unmodified on various focal length pin-hole cameras in air or water, and also on fish-eye and catadioptric cameras) and lighting change. Figure 3(c) shows a series of images from experiments illustrating the performance of the algorithm to these changing parameters. For underwater detection, the most common problematic visual effects are water turbidity, and when in smooth shallow water, the inter-

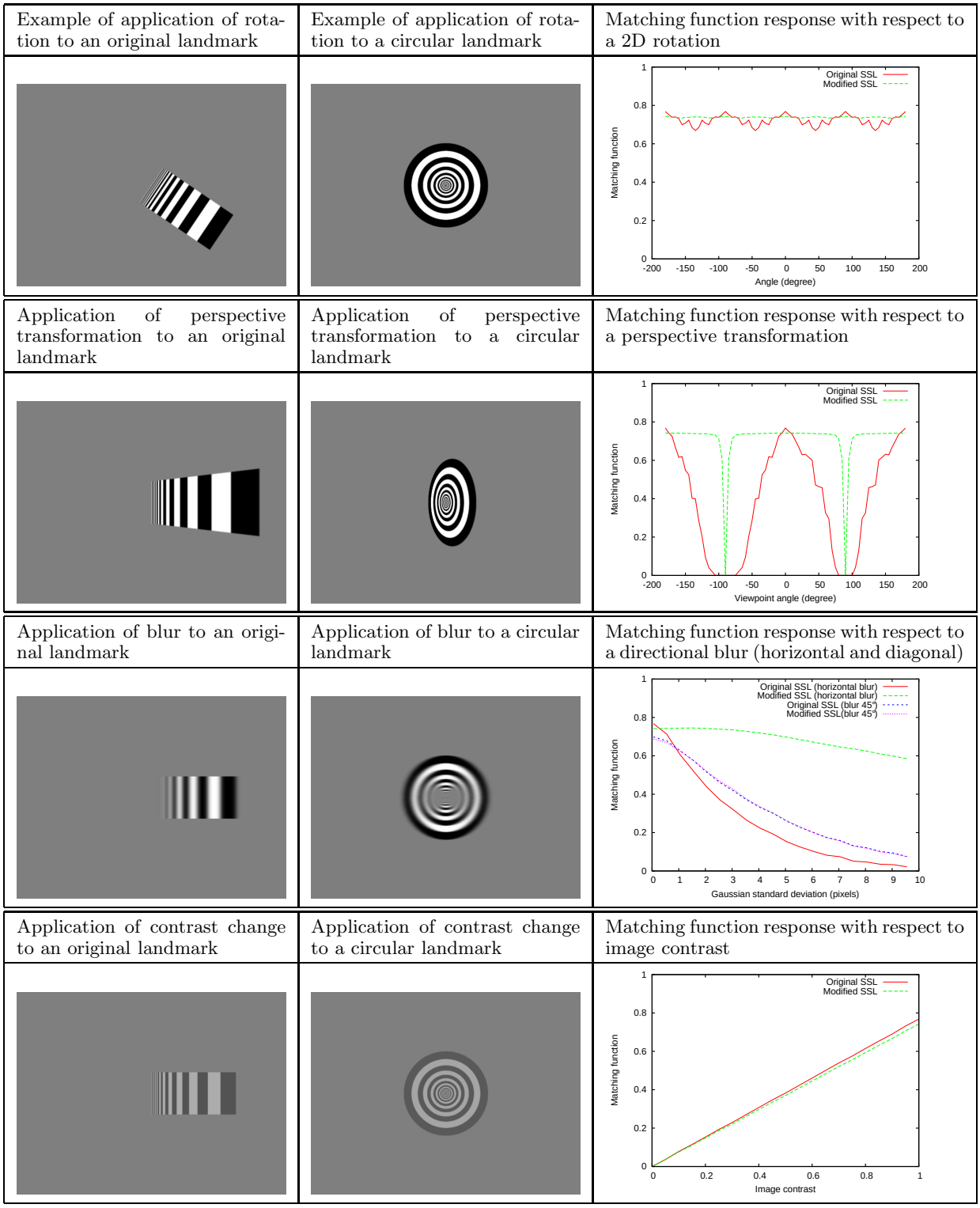
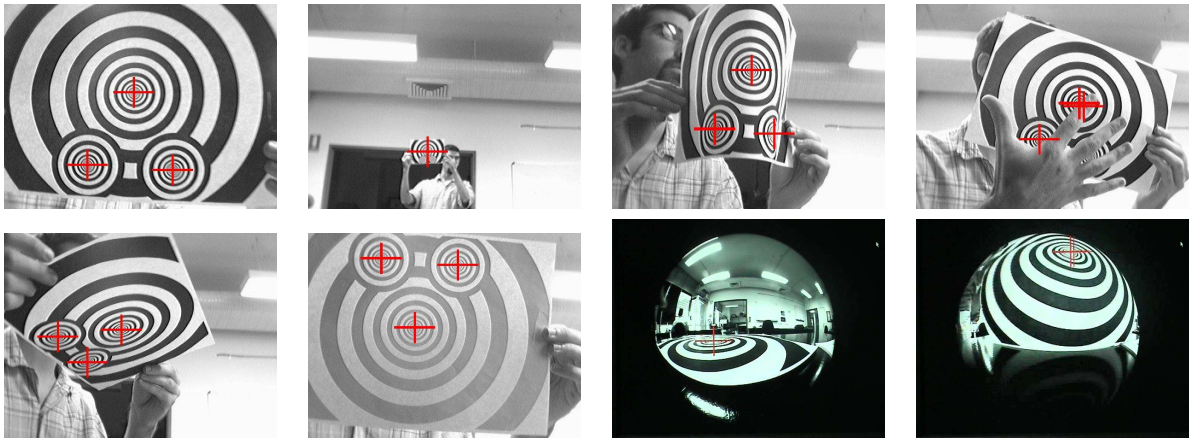


Figure 2: These plots show the matching function response for original Self-Similar Landmark and our modified circular landmark while applying a 2D rotation (top row), a perspective transformation (second row), a directional blur (third row), and a contrast change (bottom row). We can see that – thanks to circular symmetry – the proposed circular landmark gives stronger response under these different transformations and is not affected by 2D rotation. As can be noticed in the last plot, both landmarks are directly related to the contrast by a linear relation.

(a) Circular 3-pattern SSL target showing results of landmark identification.

(b) SSL detector output from (a)



(c) SSL detection result with difficult conditions: (from top left) close range, long range, target deformation, occlusion, perspective change, inversion, and two images from fish-eye lenses.

Figure 3: The circular 3-patterns target (a) used in this investigation, the output of the matching function (b) and typical detection results with hard deformation (c) including large scale changes, bending, occlusions, perspective effect, loss of contrast and distortions using a fish-eye lens. Note that the crosses represent raw maxima of the matching functions.

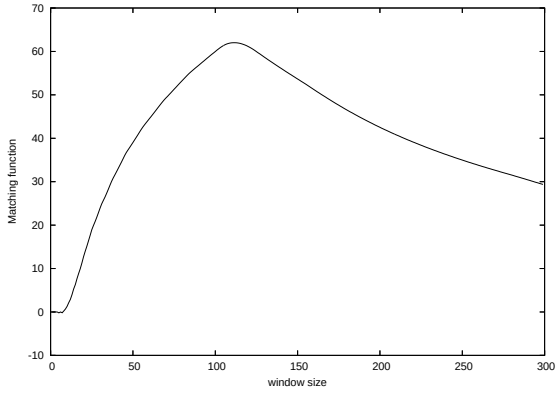


Figure 4: Value of the matching function with respect to the integral window size. The matching function exhibits a maximum at its apparent radius.

nal reflection of the landmark by the water-air interface resulting two complete landmarks in the image.

3.2 Pose & Range Estimation

Once the position of the landmarks in the image has been found, we then need to evaluate the pose of the target with respect to the camera. In this section, we use a perspective camera model and assume that the camera is appropriately calibrated (in our experiments, the camera has been calibrated using the OpenCV toolkit).

Three cases for range and pose estimation are possible according to the number of visible landmarks:

Case 1: only one visible landmark When the landmark is at distance, the two little patterns cannot be detected, nevertheless, if we assume that the target plane is roughly perpendicular to the focal axis, we can estimate the range of the target based on the size of landmark in the image. Actually, with a pin-hole camera model, it's easy to show that the range Z of the target is proportional to the inverse of the landmark size in the image:

$$Z = \frac{F \cdot S_r}{S_i} \tag{6}$$

where F is the camera focal length, S_r the real size of the landmark and S_i the size of the landmark in the image.

To estimate the size of the target in the image, we compute the matching function in the center of the landmark while varying the size of the window w (refer to Equation (3)). As we can see in Figure 4, the matching function exhibit a maximum at its apparent radius. The peak is strong enough to be detected robustly and we can then estimate the size of the landmark S_i with the following function :

$$S_i = arg\max_w \left(\frac{m^w_{+x}(x,y)}{w} \right) \tag{7}$$

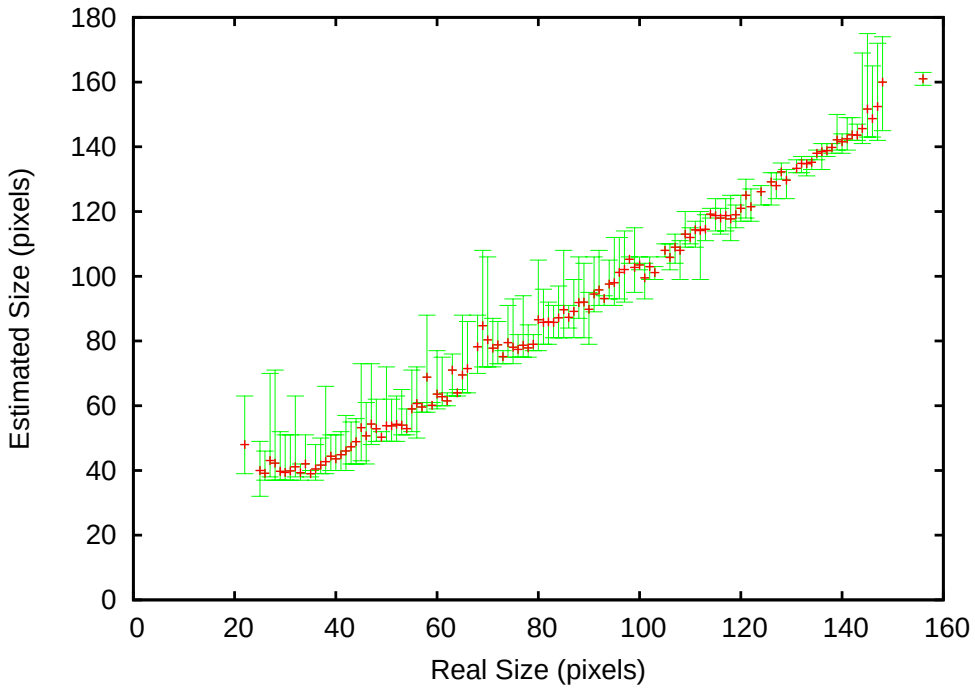


Figure 5: Estimated size of the circular target with respect to the actual measured target size, the errorbars represent the minimal and maximal estimated values. It should be noted that estimated size is a lightly above the actual size due to the background color that can extend the Self-Similar pattern.

To test this, we recorded a sequence of images and measured by hand the radius of the target in the image, and calculated the radius using the matching function. These results are plotted in Figure 5 and show a reasonably linear correlation validating the method.

Case 2: two visible landmarks In the case where only two of the three landmarks are visible, there is no reliable way to detect which have been detected. Consequently, we treat this case as if only the larger landmark was visible.

Case 3: three visible landmarks When the target is close enough, all landmarks are visible and it is possible (with some ambiguities) to compute the pose from three points. With a calibrated pin-hole camera, the problem can be expressed as a system of three quadratic equations. In the worse case we can have eight solutions (Hartley and Zisserman, 2000), however, we can easily remove half of the solutions which are behind the camera but the four remaining incorrect solutions cannot be separated immediately. However, the incorrect solutions will be eliminated later by the target tracking process.

3.3 Target Tracking

The underwater environment and the motion of the submarine make the target detection subject to noise and misdetection. In order to be robust to these perturbations, the target

localization is integrated into a particle filter (Doucet et al., 2001) and fused with movement information coming from the vehicle's navigation sensors.

Such a particle filter uses a set of samples (particles) to represent a probability distribution resulting for a Bayesian filtering process. Formally, if we let X_t be the state of the target at time t, Z_t be the observation of the target at time t and Δ_t^u be the vehicle displacement between times t and u, then the goal of the particle filter is to estimate :

$$P(X_t|Z_{t_1}...Z_{t_n}\Delta_{\tau_1}^{\tau_2}...\Delta_{\tau_{m-1}}^{\tau_m})$$

To this end, the filter relies on two models: (1) the observation model $P(Z_t|X_t)$ that predicts a target observation, and (2) a displacement model $P(X_{t_k}|X_{t_{k1}}\Delta_{t_{k1}}^{t_k})$ that predicts the motion of the particles. It must be noted that any recursive state estimation technique would be suitable for this task, especially a Kalman filter. The choice of a particle filter was motivated by the non-linearity of the system and the multiple hypothesis of the pose estimation.

The target localization is implemented as follows; we use a set of one thousand particles to track the target position and orientation coded by a 7-dimensions vector

$$X = \begin{pmatrix} r \\ q \end{pmatrix} = (x, y, z, q_0, q_1, q_2, q_3)^\top$$

The first three coordinates represent the 3D position r of the centroid of the target and the four last coordinates represent the rotation as a quaternion q. All the coordinates are expressed in the vehicle frame.

When a target is identified in an image for the first time (or after a long period), its possible positions are estimated as seen in the previous section and all particles are initialized around these positions. When a target is identified again, the image location of the landmarks in the image (Z_t) is used as an observation to update the particle filter. As we can detect one to three landmarks, we need to associate each detection to a landmark, it is done by projecting the particle's virtual landmark in the image plane and associate each detection to the nearest projected landmark. We then compute a distance d by summing all associated distance. The observation model used is then

$$P(Z_t|X_t)) = P(d(Z_t, X_t)) \sim \mathcal{G}(0, \Sigma_1)$$

where $\mathcal{G}(\mu, \sigma)$ is the Gaussian distribution of mean μ and covariance σ, and $d(Z_t, X_t)$ is the distance explained above.

In between visual identifications, the vehicle odometry (obtained from accelerometers, compass and depthmeter) is used to infer the motion of the target in the AUV's frame. Between t_1 and t_2, the vehicle displacement can be expressed as a translation T_v and a rotation R_v. By changing the reference frame, the displacement of the target is the inverse displacement of the vehicle and can be expressed by a new rotation $R_{t_1}^{t_2} = R_v^\top$ and a translation $T_{t_1}^{t_2} = -R_v^\top t$

The resulting displacement model is

$$P(X_{t_k}|X_{t_{k-1}}R^{t_k}_{t_{k-1}}T^{t_k}_{t_{k-1}}) \sim \mathcal{G}\left(\begin{pmatrix} r + T^{t_k}_{t_{k-1}} \\ q \times quat(R^{t_k}_{t_{k-1}}) \end{pmatrix}, \Sigma_2\right)$$

where $quat(R)$ denotes the quaternion defined by the rotation matrix R.

The output of the particle filter is a probabilistic estimation of the target pose. In practice, to obtain a single pose estimation of the target it is possible to compute the average pose or to get the best particle. We can also notice that using this tracking, the incorrect initial poses (see section3.2, case 3) are not stables and are automatically removed.

4 Practical Limitations of SSL in Field Robotics

Using SSL is computationally costly. The detection requires $O(n^2w)$ non-sequential image accesses, where n is the image size and w the size of the integration window. The non-sequential pixel indexing aspect is critical for an optimized implementation because this prevents an efficient implementation using specialized processor instruction such as *MMX*, *SSE* or *SSE2* and also prevents efficient caching of image data. Our comparison of implementation using floating point operations, integer operations, MMX, SSE and SSE2 showed compiler optimized floating point implementation to be the most effective on our platform.

4.1 Improving Performance

As mentioned previously, the processor requirement is important to compute the self-similar matching function in real-time. Table 1 lists the measured performance of the proposed target identification technique on different computing platforms and image sizes.

To improve performance, a simple way is to only compute this function in a reduced Region Of Interest (ROI) in the image. Using the particle filter described in Section 3.3, we can predict the position of the 3 landmarks in the image and reduce the search region around these 3 positions. Typically, in our application the ROI is a 100x100 pixel rectangle around each predicted point, dividing the computation time by 3 for an image 640x480 (see Table 1). Nevertheless, to avoid the risk of losing the target, we maintain a search in the whole image at least every 10 frames.

Processor	image size	FPS for whole image	FPS with tracking (ROI 100x100 pixels)
Pentium IV 3GHz	640x480	2.8	8.7
	320x240	11.1	18.7
Pentium M 1.4GHz	640x480	2.1	6.8
	320x240	8.9	15.4

Table 1: Measured computation time for different processors and image sizes (FPS: frame per second, ROI: region of interest).

5 Experimental Results

The vision-based target identification algorithm presented above was evaluated on an AUV in semi-controlled (test tank) and uncontrolled ocean conditions. The purpose of these experiments was to evaluate the robustness and performance of the circular SSL tracking system, as well as the AUV's target identification and homing/docking capabilities under different operating scenarios.

5.1 AUV Platform

The Starbug MkI AUV (Dunbabin et al., 2005a) was the platform used for these experiments. Starbug, shown in Figure 6, was developed at CSIRO and uses vision as its primary sensing modality for navigation and task execution. The vehicle design is a compromise between endurance, maneuverability and functionality. Endurance is best achieved with a streamlined torpedo style vehicle, however, this requires the vehicle to have longitudinal motion to obtain any control authority. Maneuverability is best achieved with the well actuated "crate" style vehicles typical of most research platforms. The "Starbug" vehicle is a hybrid of these two concepts and was chosen for these experiments due to the ability to hold position without the need for longitudinal motion.

The primary performance specifications for the Starbug Mk I AUV are: mass 26kg, length 1.2m, maximum speed 1.5m/s with an endurance of 3.5 hours with current lead-acid battery technology.

The vehicle, as utilized in these experiments, is under-actuated with five thrusters providing forward and vertical translations as well as yaw (rotation about the z-axis), roll and pitch rotations. Although there is capacity for lateral (sideways) motion, this is only marginal compared to the other axes due to the thrusters chosen in this direction. In order to minimize longitudinal drag on the vehicle, three custom developed "flat thrusters" were installed (two in the rear tail and one close to the center of gravity) which provide depth, roll and pitch control authority, with commercially available ROV thrusters proving longitudinal and rotational (yaw) authority.

Vehicle control is implemented with each of the motion axis decoupled and the control inputs determined using proportional plus derivative controllers. This control axis decoupling has proven an effective technique with satisfactory control performance through extensive experimental evaluation. The desired controlling forces are sent to the individual thrusters via a CANBus network which is serially linked to all the thrusters via the single connector. This performs well for this vehicle and has proven reliable in operation.

A 1.4GHz Pentium M PC/104 stack running the Linux operating system provides the software interface to record and process all sensor information for navigation, control and mission software in addition to the SSL tracking software in real-time. The final integration of all hardware used on the vehicle is packaged into two trays which are placed within each of the two interconnected pressure hull components of the vehicle.

The vehicle's onboard sensing systems include cameras, IMU, GPS and pressure sensor for pose estimation, as well as infra-red distance sensors for detecting close range objects (<60cm) located in the forward and downward camera housings. The vision system consists of two stereo camera pairs, one looking forward along the x-axis for obstacle avoidance, and the other looking downwards along the z-axis for visual odometry and terrain following (Dunbabin et al., 2005b). The CMOS color cameras from Point Grey Research (www.ptgrey.com) have a resolution of 640x480 pixels with a firewire interface to the on-board CPU.

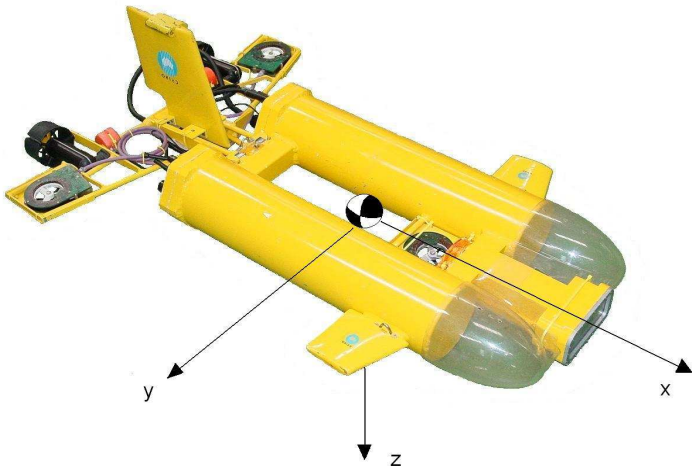


Figure 6: The Starbug MkI AUV used for docking experiments showing its local coordinate system.

As the SSL algorithm requires only monocular vision, only one of the forward (x-axis) cameras was used in these experiments. The camera was calibrated using the radial distortion model (Cucchiara et al., 2003) with the distortion coefficients determined using the OpenCV toolkit. Additionally, the roll, pitch and yaw measurements from the on-board IMU were made available to the target tracking algorithm. All visual and inertial measurements are shared amongst all the AUV's sub-systems using the DDX middleware (Corke et al., 2004).

The target consisted of the circular SSL geometry (Figure 3(a)) printed on A3 paper which was laminated and glued to a rigid backing. Anchors and floats attached to the target enabled the distance from the sea-floor to be set, however, it could translate and rotate slightly with water currents.

5.2 Target homing and stand-off experiments

The experiments consisted of developing a target searching routine for the AUV and when detected, use the vision-based SSL tracking algorithm to guide (home) the AUV towards the target and maintain a stand-off distance from it.

The AUV target identification and docking behavior consisted of three states; *State(1)* is the target search mode whereby the AUV performs a constant rate spin about its z-axis at

varying depths. The AUV enters *State(2)* when the target has been identified and remains until the AUV's x-axis is pointing at the identified target's center. Finally, *State(3)* is the homing/docking mode where the AUV moves towards the target and maintains a prespecified distance from the target.

Figure 7 shows results of a typical autonomous target identification and homing experiment in which it is desired to maintain a distance of 2m from the target. Here the target was placed at one end of the test tank and the AUV at the other. Figure 7 shows the target tracking state as well as demanded (dashed) and actual (solid) AUV yaw angle and depth. Additionally, the number of SSL patterns found ($nbPoints$) in the image is shown along with the estimated range from the target identification algorithm (see Section 3).

From Figure 7, it can be seen that the target is first identified at t=75s and as the AUV moves towards the target, the 2m standoff distance is maintained. At t=140s, the target was removed from the water and the AUV re-entered the search mode until t=225s where the target was replaced and the AUV reacquired and tracked it. This experiment was repeated many times with the system consistently able to identify and home onto the target. The system was found capable of detecting the target from a distance of at least 8m during the tests.

To demonstrate the robustness of the target tracking, during maintaining the 2m standoff distance shown in Figure 7, the target was moved left then right before being returned to its original position. Figure 8 shows the AUV yaw angle as well as the estimated target position relative to the AUV coordinate frame. Here the lateral (y_{rel}) position of the target moves as the target moves left and right and the yaw angle varies to maintain the target directly in front of the AUV.

Finally, the ability of the system to dock with the target is demonstrated in Figure 9 by setting the stand-off distance to zero. The figure shows the actual and demanded AUV yaw and depth, as well as the estimated range to the target and number of SSL circles tracked. Additionally, the ADC value of the frontal collision infra-red sensor is shown. A collision is detected when the sensors ADC value exceeds 300. In this case, the AUV detects a collision with the target at t=511s. Again this experiment was repeated many times with consistent performance demonstrated.

5.3 Field Experiments

The target identification system was also evaluated in a series of outdoor experiments at Peel Island in Brisbane's Moreton Bay. In these experiments, the target consisted of three circular SSL geometry targets arranged in an equilateral triangle with buoyancy and an anchor attached to allow it to float at a fixed distance above the sea-floor. This arrangement of three targets was chosen to allow detection from any approach angle and improve detection rate.

The purpose of this ocean test is to simulate the revisitation of a known landmark to allow relocalization and perform a transect survey. To initiate the search, the AUV travels to the

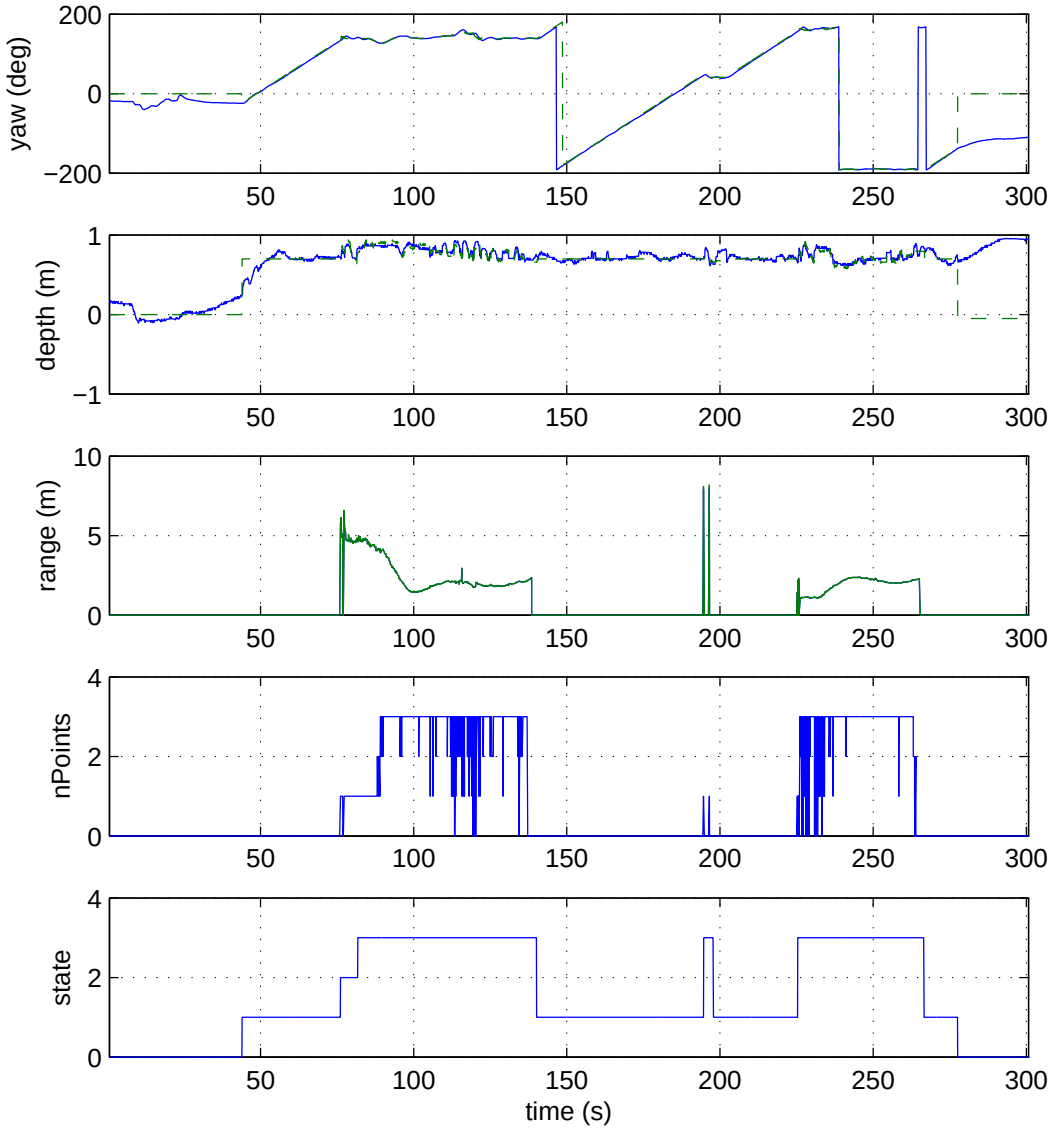


Figure 7: AUV pose and SSL detection properties during a typical target identification and homing experiment. Here the AUV was required to identify the target and maintain a stand-off distance (range) of 2m.

expected target location using the mission primitives and navigation strategies described in (Dunbabin and Allen, 2007). Once within a prespecified radius of the expected target location (typically 1.5m GPS error), it begins a search routine as described above to locate the SSL target using the proposed method.

The experiments were conducted in approximately 3m of water where the sea floor consisted of rock and coral reef. Visibility during the experiments was determined in adjacent deep

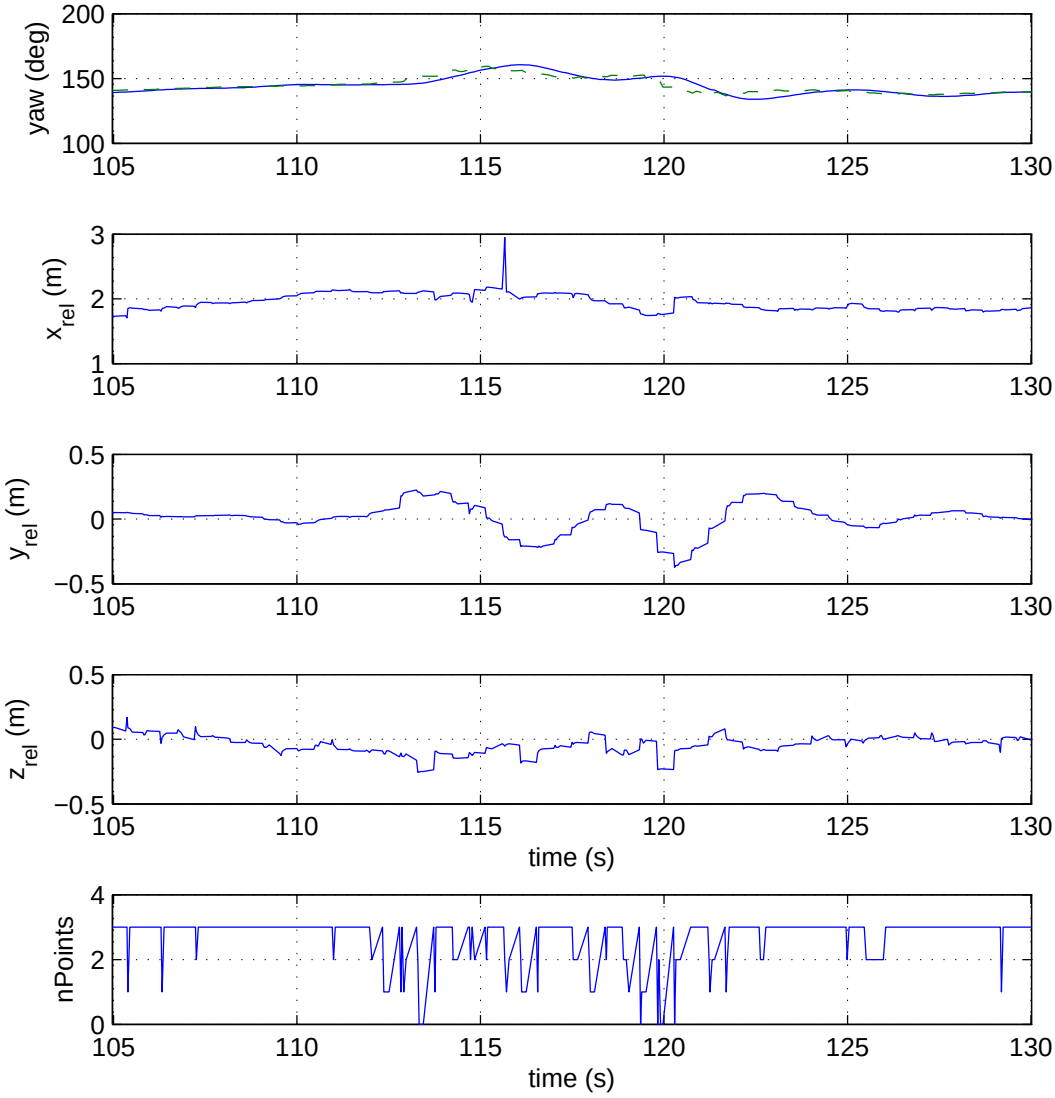


Figure 8: Target tracking performance showing target position relative to AUV coordinate frame moving the target left and right and orienting it relative to the AUV.

water to be limited to approximately 4m based on a Secchi measurement (Kirk, 1994). The target floated approximately mid-water and again whilst anchored, was still free to rotate and sway with the currents.

Although visibility was poor, the proposed method still proved reliable in detecting the target (at distances up to 3m), allowing the AUV to manoeuvre to the desired standoff distance of 1.5m. Figure 10 shows the recorded relative target position with respect to the vehicle during an autonomous mission.

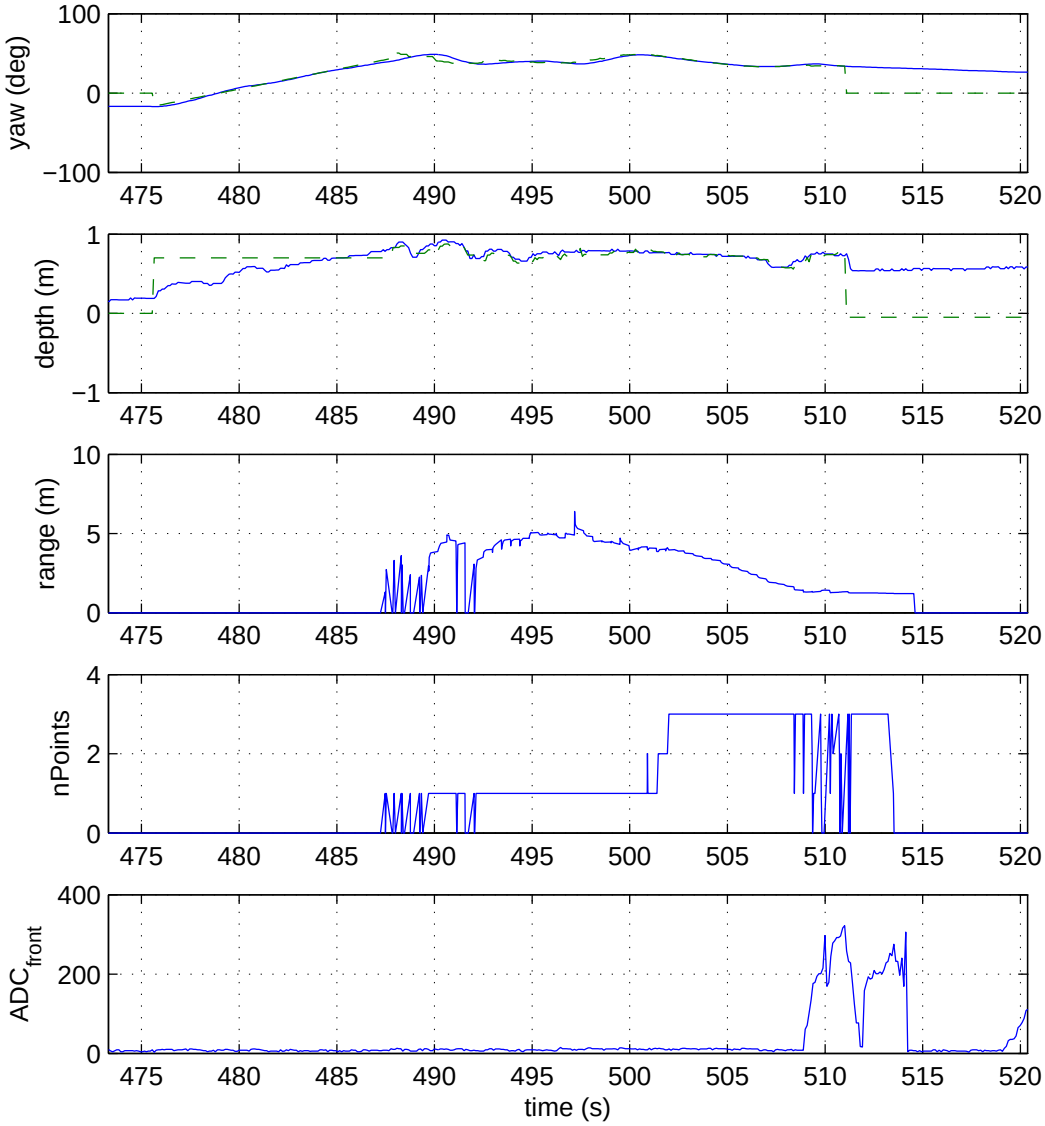


Figure 9: AUV pose and SSL detection properties during a typical docking experiment where the AUV identifies and moves towards the target until the frontal collision sensor was triggered.

As observed in Figure 10, due to visibility conditions the larger target is primarily only detected even at the closer range. Additionally, the detection stability decreases as compared to the clearer water tank experiments shown above. However, it does illustrate the ability of the target detection and tracking system to home into the target.

Figure 11 shows histograms of the number of tracked images with their relative detected target position with respect to the vehicle for a combined set of three ocean tracking experi-

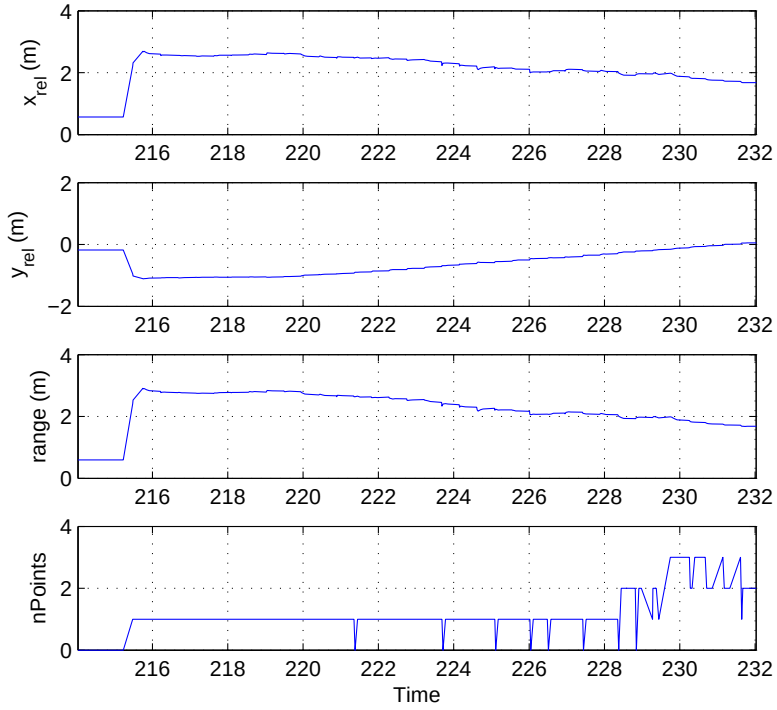


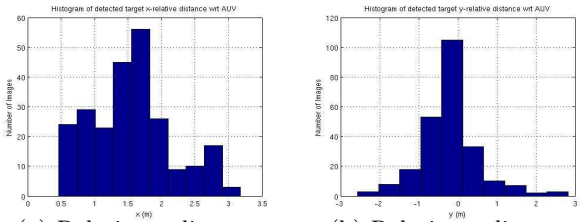
Figure 10: Relative target position and range with respect to the AUV during an autonomous tracking experiment off Peel Island in Brisbane. The lower trace shows the number of detected landmarks (nPoints).

ments. As can be seen, the AUV tracking performance for x_{rel} peaks at the desired stand-off distance of 1.5m, and the y_{rel} peaks at around zero, that is pointing directly at the target.

Visibility and lighting conditions were observed to be the dominate factor in deteriorating the tracking performance of the proposed system during field experiments. Figure 12 shows a series of images of the target from the AUV and associated SSL detection results during various stages of the mission.

The results of Figure 12 show that the algorithm is robust to perspective and lighting discontinuity (first and second columns) being able to reliably identify the target. The third column shows the minimum required contrast before the target can be identified, however, as can be seen in Figure 13, a brightness normalization can dramatically improve the detection performance (this is particularly relevant for underwater images where most of the image frame is blue). The system also proved robust to features within the environment that could falsely register as potential targets as presented in the forth column of Figure 12 which shows an image with a black and white striped fish which is not segmented out.

Another lighting problem encountered during outdoor experimentation was due to the reflectivity of the laminate coating on the target itself. Figure 14(a) shows an image recorded by the AUV during a tracking experiment where one side of the target is visible (shaded



(a) Relative x-distance (b) Relative y-distance

Figure 11: Histograms showing the number of SSL detected images with relative target position for a combined set of three ocean tracking experiments.

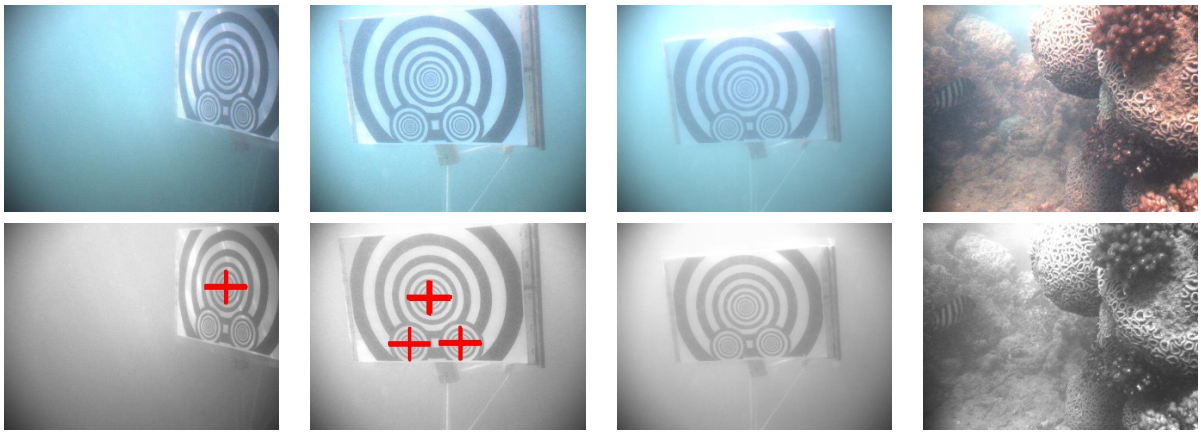


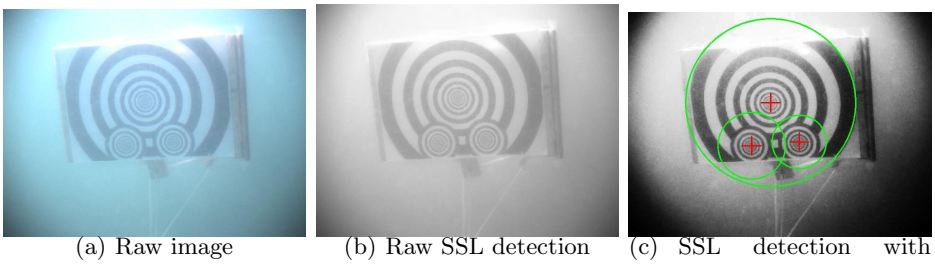
Figure 12: SSL detection in marine environment. Upper row: original images, lower row: detection. From left to right: robustness to perspective and lighting discontinuity, minimum required contrast, insufficient contrast, absence of false positives in a natural environment.

side) whereas the other side appears bright due to the shiny reflective surface. Despite this lighting problem, the visible target can be adequately identified as shown in Figure 14(b).

Overall, the target identification system proved effective in the marine environment with visibility being the greatest restriction in identification performance. Further enhancements such as brightness normalization have assisted to improve target detection performance in varied lighting conditions.

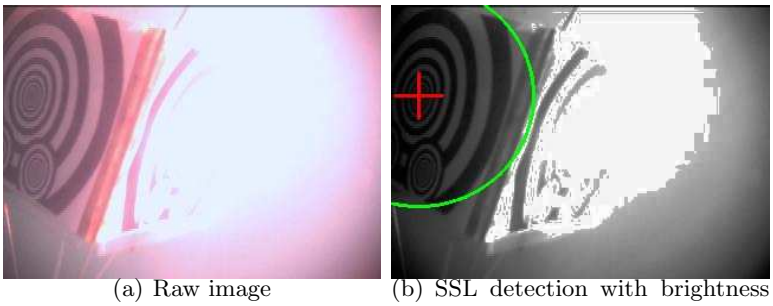
6 Conclusions

Underwater vision-based tasks are typically complex to implement due to poor lighting conditions, refractions and moving objects such as fish. In this paper we developed a rotationally invariant circular self-similar landmark and demonstrated its use for target identification and in enabling vision-based docking using the Starbug AUV. The method provides an exceptionally robust landmark with very little sensitivity to camera model, distortion and observation range. The resulting docking task was proven effective through extensive pool trials. Experiments in an uncontrolled reef environment has been conducted to demonstrate the applicability of our approach in the field. The SSL detection proved effective in robustly



(a) Raw image (b) Raw SSL detection (c) SSL detection with brightness normalization

Figure 13: A brightness normalization can dramatically improve the detection performance. This is particularly relevant for underwater images where most of the image frame is blue/green due to red-wavelength absorption.



(a) Raw image (b) SSL detection with brightness normalization

Figure 14: Example of effect of varied lighting on the ocean target with target detection results using brightness normalization.

identifying the target with poor visibility and varied lighting conditions.

Acknowledgments

This work was funded in part by the CSIRO Wealth from Oceans flagship project on Platform Free Fields. The authors would like to thank the following Autonomous Systems Laboratory team members for their assistance in vehicle preparation and gathering of experimental results: Polly Alexander, Peter Corke, Kane Usher, Les Overs, Stephen Brosnan and John Whitham.

References

Briggs, A., Scharstein, D., Braziunas, D., Dima, C., and Wall, P. (2000). Mobile robot navigation using self-similar landmarks. In *Proc. IEEE International Conference on Robotics and Automation*, volume 2, pages 1428–1434 vol.2, San Francisco, CA, USA.

Corke, P., Sikka, P., Roberts, J., and Duff, E. (2004). DDX: A distributed software architecture for robotic systems. In *Proc. Australian Conf. Robotics and Automation*, Canberra, Australia. published on CDROM.

Cowen, S., Briest, S., and Dombrowski, J. (1997). Underwater docking of autonomous undersea vehicles using optical terminal guidance. In *Proc. of MTS/IEEE OCEANS'97*, volume 2, pages 1143–1147, Canada.

Cucchiara, R., Grana, C., Prati, A., and Vezzani, R. (2003). A hough transform-based method for radial lens distortion correction. In *ICIAP '03: Proceedings of the 12th International Conference on Image Analysis and Processing*, pages 182–187, Washington, DC, USA. IEEE Computer Society.

Cufi, X., Garcia, R., and Ridao, P. (2002). An approach to vision-based station keeping for an unmanned underwater vehicle. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and System*, volume 1, pages 799–804 vol.1, Lausanne, Switzerland.

Dalgleish, F., Tetlow, S., and Allwood, R. (2005). Vision-based navigation of unmanned underwater vehicles: A survey. part 2: Vision-based station-keeping and positioning. *Journal of Marine Design and Operations*, (8):13–19.

Doucet, A., De Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo methods in practice*. Springer.

Dunbabin, M. and Allen, S. (2007). Large-scale habitat mapping using vision-based AUVs: Experiences, challenges & vehicle design. In *Proc. OCEANS 2007 Europe*, pages 1–6, Aberdeen, UK.

Dunbabin, M., Corke, P., Vasilescu, I., and Rus, D. (2006). Data muling over underwater wireless sensor networks using an autonomous underwater vehicle. In *Proc. IEEE International Conference on Robotics and Automation*, pages 2091–2098, Orlando, Florida, USA.

Dunbabin, M., Roberts, J., Usher, K., Winstanley, G., and Corke, P. (18-22 April 2005a). A hybrid AUV design for shallow water reef navigation. In *Proc. IEEE International Conference on Robotics and Automation*, pages 2105–2110, Barcelona, Spain.

Dunbabin, M., Usher, K., and Corke, P. (2005b). Visual motion estimation for an autonomous underwater reef monitoring robot. In *Proc. Int. Conf. on Field and Service Robotics*, pages 31–42, Port Douglas, Australia.

Evans, J., Redmond, P., Plakas, C., Hamilton, K., and Lane, D. (22-26 Sept. 2003). Autonomous docking for intervention-AUVs using sonar and video-based real-time 3D pose estimation. In *Proc. OCEANS 2003.*, volume 4, pages 2201–2210 Vol.4, San Diego, CA, USA.

Fan, Y. and Balasuriya, A. (2000). Autonomous target tracking by AUVs using dynamic vision. In *Proc. 2000 International Symposium on Underwater Technology*, pages 187–192, Tokyo, Japan.

Faugeras, O. (1993). *Three-Dimensional Computer Vision: a Geometric Viewpoint*. MIT Press.

Feezor, M., Yates Sorrell, F., Blankinship, P., and Bellingham, J. (Oct 2001). Autonomous underwater vehicle homing/docking via electromagnetic guidance. *IEEE Journal of Oceanic Engineering*, 26(4):515–521.

Hartley, R. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Kirk, J. (1994). *Light and photosynthesis in aquatic ecosystems*. Cambridge University Press, Cambridge.

Lee, P.-M., Jeon, B.-H., and Kim, S.-M. (22-26 Sept. 2003). Visual servoing for underwater docking of an autonomous underwater vehicle with one camera. In *Proc. OCEANS 2003.*, volume 2, pages 677–682 Vol.2, San Diego, CA, USA.

Marks, R., Rock, S., and Lee, M. (1993). Automatic object tracking for an unmanned underwater vehicle using real-time image filtering and correlation. In *Proc. of IEEE Systems, Man, and Cybernetics*, France.

Singh, H., Lerner, S., von der Heyt, K., and Moran, B. (1998). An intelligent dock for an autonomous ocean sampling network. In *Proc. of OCEANS'98*, volume 3, pages 1459–1462, France.

Stokey, R., Purcell, M., Forrester, N., Austin, T., Goldsborough, R., Allen, B., and von Alt, C. (6-9 Oct 1997). A docking system for REMUS, an autonomous underwater vehicle. In *Proc. OCEANS '97. MTS/IEEE*, volume 2, pages 1132–1136 vol.2, Halifax, Canada.

Wang, H., Rock, S., and Lees, M. (9-12 Oct 1995). Experiments in automatic retrieval of underwater objects with an AUV. In *Proc. OCEANS '95. MTS/IEEE.*, volume 1, pages 366–373 vol.1.

Wettergreen, D., Gaskett, C., and Zelinsky, A. (1998). Development of a visually-guided autonomous underwater vehicle. In *Proc. of OCEAN'98 Conference*, volume 2, pages 1200–1204, Nice, France.

Yu, S.-C., Ura, T., Fujii, T., and Kondo, H. (2001). Navigation of autonomous underwater vehicles based on artificial underwater landmarks. In *Proc. OCEANS, 2001. MTS/IEEE Conference and Exhibition*, volume 1, pages 409–416 vol.1, Honolulu, HI, USA.