Performance Evaluation of a Low-Cost Stereo Vision System for Underwater Object Detection ⋆

Oleari, Fabio * **Kallasi, Fabjan** * **Lodi Rizzini, Dario** *
Aleotti, Jacopo * **Caselli, Stefano** *

* *Dipartimento di Ingegneria dell'Informazione, University of Parma, Italy (e-mail: {oleari,kallasi,dlr,aleotti,caselli}@ce.unipr.it).*

Abstract: This paper describes the development and performance assessment of a low-cost stereo vision system for underwater object detection. The system has been conceived as a prototype to investigate the performance, power consumption, and thermal dissipation tradeoffs involved in designing an embedded stereo vision unit for underwater operation. The embedded system has been experimentally assessed in underwater object detection tasks. The system has proven thermally stable and capable of guaranteeing a level of autonomy of at least two hours of video acquisition. Several algorithms for mono and stereo image processing have been evaluated to assess their effectiveness in the underwater environment along with their suitability in presence of constrained computational and energy resources. Evaluation of the stereo vision system in detecting simple objects has shown strong limitations of commodity off-the-shelf sensors when used in underwater perception. Nevertheless, the prototype described in this work provides insights for development of more advanced vision systems suitable for underwater vehicles.

Keywords: Underwater imaging, Stereo vision, Object detection.

1. INTRODUCTION

Vision technologies have gained increasing consideration as a major sensing modality in underwater environments, thanks to high sensor resolution, comparatively low cost, and rich suite of algorithms made available by mainstream research. There is indeed a high potential for exploitation of advanced sensors in underwater operation. However, in underwater environments the acquisition of 3D information through visual perception is subject to severe limitations. In particular, the propagation of light through water suffers from phenomena such as absorption and scattering. Despite the work carried out so far, there is a need for additional experimental investigation to assess the actual potential of visual perception in underwater environments.

This paper presents the development and the initial evaluation of a prototype embedded system for underwater object detection using stereo vision. The system has been developed within the MARIS project (Marine Autonomous Robotics for InterventionS, Italian National Project). The general goal of MARIS is the development of technologies useful for underwater intervention in the offshore industry, in search-and-rescue tasks, and in scientific exploration.

In particular, this work describes a low-cost box developed for underwater image acquisition, which includes a 3-sensor multi-baseline stereo camera, and the algorithmic pipeline developed for stereo image processing. Experiments of object detection have been performed to assess visual perception performance in real underwater scenarios. Experiments have also evaluated the embedded system by measuring physical parameters relevant for underwater autonomous operation such as power consumption, humidity level, and thermal characteristics. The goal of the proposed processing pipeline is to detect a target object of known shape and to estimate its pose with respect to a reference frame on the sensor. At each iteration, frames are preprocessed for underwater image enhancement. Then, the method extracts a region of interest where the target object is located and performs computation of the disparity map for 3D representation. Finally, the pose of the object is estimated through 3D model alignment.

The paper is organized as follows. Section 2 reviews the state of the art regarding underwater sensing methods for object detection. Section 3 describes the low-cost embedded vision system. Section 4 illustrates the developed image processing algorithms and section 5 reports results obtained in the initial experimental evaluation of the system. Section 6 concludes the paper with some directions for future research.

2. RELATED WORK

Although vision is a major sensing modality in robotics, it is not widely used in underwater perception due to the problems of light transmission in water. Ultrasonic sensing instead is a commonly used and robust underwater perception modality. However, acoustic sensing is not suitable when an accurate and detailed reconstruction of the

⋆ This work was supported by the Italian National Project *MARIS: Marine Autonomous Robotics for InterventionS*, PRIN call years 2010-11, N. 2010FBLHRJ-007.

object shape is required. Sonar array cameras have been developed which exploit the emission of multi-frequency acoustic signals for detection and recognition of objects. These systems allow 3D sonar imaging (Yu et al. (2006)) and extend their application to recognition tasks, but their high cost, limited resolution, and operational complexity restrict their application domain. A rather extensive survey and comparison of state-of-art ultrasonic technologies with vision in underwater scenarios is presented in Jonsson et al. (2009). Underwater laser scanners do exist and exploit an accurate modeling of light propagation in water means Gordon (1992). Such sensors can provide high performance in term of resolution and accuracy of acquired 3D images. However, underwater laser scanners are very expensive and affected by the same operating problems of vision systems. A system for underwater object manipulation has been described by Sanz et al. (2013) where object perception is achieved using a structured light laser attached to the forearm of the manipulator and unknown objects are successfully grasped in a water tank environment.

Vision can provide, when feasible, information at lower costs and with higher resolution and acquisition rate, compared to acoustic perception. Applications of underwater computer vision include detection and tracking of cables and pipelines for surveying (Narimani et al. (2009)), image mosaicing to map seabeds (Nicosevici et al. (2009)), monitoring of underwater plants and artifacts, recognition and observation of interesting objects like artificial structures or archaeological sites (Eustice et al. (2005)), and localization and mapping in limited regions (Horgan et al. (2009); Schattschneider et al. (2011)). In Kim et al. (2012) a vision-based object detection method is presented based on template matching and tracking for underwater robots using artificial objects. In Garcia and Gracias (2011) a comparison of the performance of popular salient keypoint detectors on underwater images, degraded by turbidity, is performed. It is shown that Hessian-based approaches outperform Laplacian and Harris counterparts.

Stereo vision systems have not been extensively used until recently, due to the difficulty of homologous point matching in underwater conditions and relatively high computational requirements. The development of advanced algorithms for preprocessing and image enhancement (Bazeille et al. (2006); Corchs and Schettini (2010)) and of dedicated or high performance architectures for image processing has increased the interest for underwater application of stereo vision (Ishibashi (2009)). Disparity of stereo images can be exploited to generate 3D models (Brandou et al. (2007); Campos et al. (2011)) through the interpolation, filtering and segmentation of measurements. The resulting topological representation allows object recognition even with partial data as well as integration of multiple view images. In Wang et al. (2011) a stereo matching algorithm is proposed based on median and homomorphic filtering to minimize noise and enhance image contrast. The system also adopts Harris corner detection to obtain the characteristic information of underwater targets.

Low-cost image and 3D sensors are currently widely used in different applications of shape recognition and processing. However, 3D active devices like Microsoft Kinect (Andersen et al. (2012)) are not suitable for underwater environments. In this work we investigate the suitability for underwater perception of a stereo vision system built using inexpensive webcams, similar to the work in (Oleari et al. (2013)).

3. EMBEDDED STEREO VISION SYSTEM

The design of the embedded system has taken into account the constraints and requirements of the underwater application in terms of computational capacity, power consumption and thermal dissipation, waterproofing, and implementation time. The system has been conceived as a low-cost prototype to be assembled in a short time and to be eventually adapted during the development. For these reasons, a general purpose plastic box has been chosen as the container. This canister (260x330x92mm) has a flat transparent plate and a certified protection rating IP-68. The key design problem is the achievement of a trade-off between computational power requirements and electrical power consumption and heat dissipation through the plastic enclosure. Computer vision tasks require high computational capabilities and compatibility of the platform with common software frameworks and libraries. On the other hand, the CPU power consumption and *thermal design power* (TDP) should be as low as possible, since the battery storage is limited and, above all, the cooling down inside a waterproof sealed canister is performed through conduction. The embedded system mounts a Mini-ITX Intel Desktop Board DN2800MT with an Intel Atom processor N2800 (TDP 6.5 W) and 2 GB RAM, which is a trade-off between the power-efficient ARM architecture processors (TDP 5.0 W for ARM Cortex A9) and the powerful commodity processors in the x86 architecture (low consumption embedded Intel Core i7-3517UE has TDP 17 W).

Figure 1 shows the embedded system and its architecture. The internal structure of the canister has been organized in three vertical layers. The bottom layer contains 4 batteries ($12V$, $2Ah$) and a DC UPS (10A). The middle layer contains the processing hardware described above and a $12V$ to $5V$ DC-DC step-down unit to supply the sensors. The system also features a SSD hard drive. The top layer contains the 3-sensor multi-baseline and an Arduino Uno board to control the system (internal temperature and humidity monitoring, power on/off using a remote controller, LCD display to log information). The system comprises 3 cameras to test different baselines and fields of view. The vision sensors are three Logitech C270 webcams (1280x960 @$7.5fps$). The choice of low cost cameras is motivated by the prototype nature and testing aim of the system.

In order to monitor the temperature and humidity inside the canister these sensors have been integrated in the system: three analog temperature sensors (LM 35) for CPU, hard drive and batteries, and a digital sensor for temperature and humidity (DHT11) located in the top layer. Both sensors provide fully calibrated outputs. The LM35 sensor maintains an accuracy of $\pm 0.8^oC$ over a range from 0^oC to 100^oC. The LM35 sensor draws $60\mu A$ and possesses a low self-heating capability. The DHT11 sensor operates from $3.5V$ to $5.5V$. DHT11 can measure temperature from 0^oC to 50^oC with an accuracy of $\pm 2^oC$,

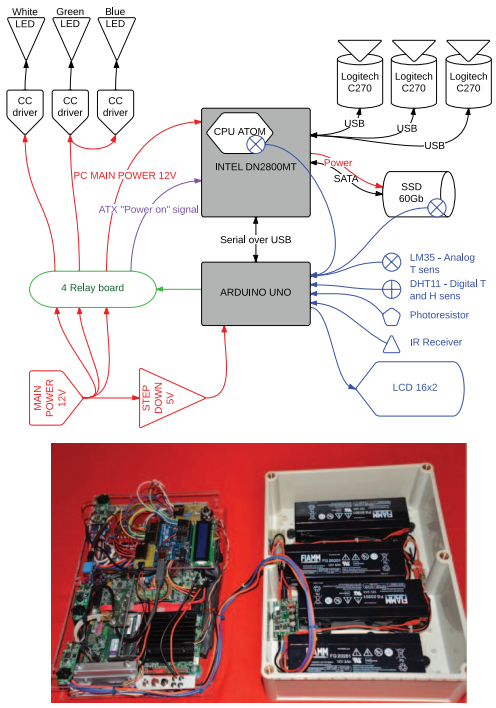

Fig. 1. Architecture of the embedded vision system (top) and image of the opened canister (bottom).

and relative humidity ranging from 20% to 95% with an accuracy of ±5%. The Arduino Uno board checks the temperature and humidity measurements and is in charge of the shutdown of the system, if the measurements exceed their respective safety thresholds. The embedded system can also be supplied by an external power source. An Ethernet cable enters the canister to connect the system to an external computer for monitoring operations. The system does not use underwater illuminators.

4. ALGORITHMS

The embedded system described in the previous section has been designed to detect objects. The system can acquire color images from each of the three cameras and compute the 3D representation of the observed environment by stereo-processing the data provided by any camera pair. Thus, object detection may be performed using algorithms based on the processing of images acquired from a single camera or with three-dimensional shape recognition algorithms based on stereo processing. In our evaluation we have assumed that the objects to be detected have cylindrical shape and can be conveniently represented by a geometric parametric model. However, this assumption is exploited only at the end of the processing pipeline, to identify the object and estimate its pose. The other vision modules operate on general hypotheses about the environment and the targets.

A suitable underwater vision system must be designed to cope with the difficult underwater light conditions. In particular, light attenuation produces blurred images with limited contrast, and light back-scattering results into artifacts in acquired images. Object detection becomes even more difficult in presence of suspended particles or with an irregular and variable background. Hence,

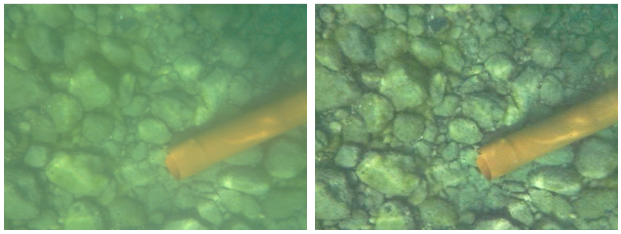

Fig. 2. An underwater image before (left) and after (right) the application of contrast mask and CLAHE.

with underwater images special attention must be paid to algorithmic solutions improving their quality in the early processing stages. Object detection, therefore, has been decomposed into several sub-tasks including image enhancement, image segmentation, 3D model matching and object pose estimation. Detection is performed by processing both a single colored image and the point cloud achieved by stereo vision 3D reconstruction. The mono-camera and stereo-camera processing algorithms are discussed in the following.

4.1 Mono-Camera Processing

Processing of individual images is performed on the image stream produced by one of the cameras and aims at detecting the region of the image that contains the target object. The identification of a region of interest (ROI) restricts the search region of the target object in later processing stages and, therefore, prevents possible detection errors. Since object recognition on a 3D point cloud is computationally expensive, mono-camera processing helps in decreasing the requested overall computation time.

The first step is the image pre-processing performed in order to improve image quality in the underwater environment. A *contrast mask* method based on component L of CIELAB color space is applied to the input image. In particular, the component $L_{in,i}$ of each pixel i is extracted, a median filter is applied to the L-channel of the image to obtain a new blurred value $L_{blur,i}$, and the new value is computed as $L_{out,i} = 1.5\ L_{in,i} - 0.5\ L_{blur,i}$. The effect of the contrast mask is a sharpened image with increased contrast. Next, a contrast-limited adaptive histogram equalization (CLAHE) is performed in order to re-distribute luminance. The combined application of contrast mask and CLAHE lessens light attenuation and reduces the effect of light artifacts on the objects, as shown in Figure 2. It was observed that the image enhanced by CLAHE alone is not discernible from the one achieved after applying both filters. Hence, the contrast mask may not be required.

The second step of mono image processing performs image segmentation, i.e. identification of a ROI that contains the target object. The ROI may be searched according to different criteria based on a specific feature of the object to be found. One criterion is the uniformity of the target object color w.r.t. the roughness and variety of the background. Two algorithms have been investigated. The first algorithm is a variant of *Eigen transform* (Targhi et al. (2006)), a texture descriptor that allows identification of regions with similar patterns. The algorithm is applied to the luminance component of the color image and it processes the matrix containing the grey-scale values of

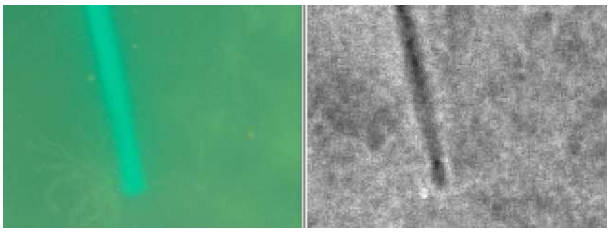

Fig. 3. An example of LU transform applied to an underwater image (left). The area corresponding to the object is represented by a black blob (right).

a small square window with size $w \times w$ centered on each pixel p_{ij}. The smaller eigenvalues or the singular values of such matrix provide a measure of the roughness of the area around each pixel: the rougher is the texture in the $w \times w$ window, the greater are the smaller eigenvalues of the corresponding matrix. Since computation of eigenvalues or singular values is time-consuming, a more efficient algorithm is the *LU transform* that evaluates roughness using the LU decomposition. The diagonal of U matrix captures the same information of eigenvalues. In particular, the output of such method is a scalar computed as the average absolute values of the l smaller diagonal elements u_{kk} of matrix U

$$\Gamma(l, w) = \frac{1}{w - l + 1} \sum_{k=l}^{w} |u_{kk}| \tag{1}$$

where $|u_{11}|, \ldots, |u_{ww}|$ are sorted in decreasing order. This technique can be applied to detect artificial objects with uniform color and pattern that are placed in a natural irregular background: if the value of eq. (1) is used as the color of a mask, the uniform colored objects appear as darker blobs on a brighter background (see Figure 3). The ROI of objects can be obtained by threshold classification or by region growing. The main drawbacks of this technique are its computational cost (still high even using LU matrix decomposition) and its effectiveness in underwater scenarios. Indeed, light absorption tends to smooth the roughness of natural background and, in some cases, it is difficult to distinguish between objects and background.

Another approach based on color level segmentation has been tested to detect ROIs in the image. This method assumes that the object to be detected has a uniform color. In this approach the image is converted to HSV (Hue Saturation Value) color space and segmented according to 16 intervals of channel H values. Hence, the image can be partitioned into 16 subsets of (possibly not connected) pixels with the same hue level. The rough level quantization is not affected by the patterns generated by light back-scattering. The region corresponding to a given hue level is estimated as the convex hull of the pixels. Only regions whose area is less than 50% of the image are selected as part of the ROI. This heuristic rule rests on the hypothesis that the object is observed from a distance such that only the background occupies a large portion of the image. ROI estimation only exploits the relative color uniformity of a texture-less object, but it does not identify a specific object. When the object color is known, a more specific *color mask* (CMask) can be applied to detect the object with an accurate estimation of object contour. In our experiments, both ROI and CMask always contain the target objects. CMask can be used either as a groundtruth to estimate ROI precision or to detect the object in the image.

In general, object pose estimation cannot be performed on a single image, and requires 3D perception as shown in the next section. However, if the object shape is known, such as with the cylindrical pipes to be recognized in our experiments, pose estimation is possible also with a monocular camera. In particular, the cylinder is defined once are given the cylinder radius c_r and its axis, a line with equation $c(t) = c_p + c_d\, t$. The contour of a cylinder in the image plane is delimited by two lines with equations $l_i^T\ u = 0$ with $i = 1, 2$, where $u = [u_x, u_y, 1]^T$ is the pixel coordinate vector and l_1, l_2 are the coefficients. Let l_0 be the parameters of the line representing the projection of the cylinder axis in the image. The two lines with parameters l_1 and l_2 are the projections on the image plane of the two planes, which are tangent to the cylinder and contain the camera origin. The line with parameter l_0 is the projection of the plane passing through the cylinder axis and the camera origin. The equations of these three planes in the 3D space are

$$l_i^T\ (Kp) = (K^T\ l_i)^T\ p = n_i^T\ p = 0 \tag{2}$$

where K is the camera matrix obtained from the intrinsic calibration, $n_i = K^T\ l_i$ the normal vectors of the planes corresponding to the lines l_i with $i = 0, 1, 2$ (in the following, the normalized normals $\hat{n}_i = n_i / \|n_i\|$ are used), and p a generic vector in camera reference frame coordinates. The direction of the cylinder axis is given by direction vector $c_d = \hat{n}_1 \times \hat{n}_2$. If the cylinder radius c_r is known, then the distance of the cylinder axis from the camera center is equal to

$$d = \frac{c_r}{\sin\left(\frac{1}{2}\operatorname{acos}\left(|\hat{n}_1 \cdot \hat{n}_2|\right)\right)} \tag{3}$$

The projection of the camera origin on the cylinder axis is equal to $c_p = d(c_d \times \hat{n}_0)$ (if $c_{p,z} < 0$, then substitute $c_p \leftarrow -c_p$). These geometric constraints allow estimation of the object pose in space using only a single image. The accuracy of such estimation depends on the image resolution and on the extraction of the two lines. It can be used as an initial estimation that can be refined by processing the 3D point cloud computed using stereo vision, or to validate the results.

4.2 Stereo-Camera Processing and Pose Estimation

Object detection and pose estimation are performed on the 3D point cloud computed using stereo vision techniques. The ROI obtained from single camera processing is used to restrict the region where stereo point matching is performed and the object is searched. The benefit of restricting the region size where stereo processing is performed is limited when the disparity image is performed using incremental *block-matching SAD* (sum of absolute differences) algorithm. Since the SAD of a block is computed using the SAD values of adjacent blocks, the advantage of computing the disparity image only on the ROI is not significant. However, estimation of noisy point cloud limited to the ROI saves about 15% of the time for each frame.

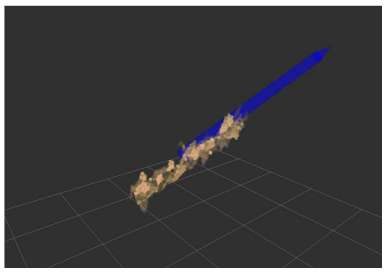

Fig. 4. An example of pose estimation by matching the raw point cloud and a cylinder model (blue).

The importance of a ROI is more apparent in object recognition, since this step requires computationally expensive operations on point clouds. In particular, the ROI can be used to select the point cloud $\mathcal{C}$ where to search objects. The objects to be recognized have a cylindrical shape and can be represented by a parametric model. In particular, we represent cylinders using 7 parameters: the three coordinates of a cylinder axis point $c_p = [c_{p,x}, c_{p,y}, c_{p,z}]^T$, the axis direction vector $c_d = [c_{d,x}, c_{d,y}, c_{d,z}]^T$, and the radius c_r. The model matching algorithm simultaneously searches for a subset of the point cloud that better fits a cylindrical shape and computes the value of the cylinder parameters $c = [c_p^T, c_d^T, c_r]^T$. Three algorithms have been applied to the problem: RANSAC (RANdom SAmple Consensus) Fischler and Bolles (1981), PSO (Particle Swarm Optimization), and DE (Differential Evolution) (for PSO and DE see Ugolotti et al. (2013)). These algorithms require a fitness function that measures the consensus of a subset of the point cloud $\mathcal{C}$ over a candidate model c. A natural fitness function is the percentage of points $p_i \in \mathcal{C}$ such that their distance to the cylinder c is less than a given threshold d_{thr}. The more obvious measure of the displacement between a point p_i and a cylinder c is the Euclidean distance

$$\mathrm{d}_E(p_i, c) = \left| \frac{\|c_p \times (c_p - p_i)\|}{\|l_d\|} - r \right| \tag{4}$$

However, the Euclidean distance may not take into account some orientation inconsistencies. If the normal vector n_i on point p_i can be estimated, the angular displacement between the normal and the projection vector of the point p_i on the cylinder c (called $\mathrm{proj}(p_i, c)$ hence after) provides

$$\mathrm{d}_N(p_i, n_i, c) = \min(\alpha_i, \pi - \alpha_i) \tag{5}$$

$$\alpha_i = \arccos\left(\frac{n_i \cdot \mathrm{proj}(p_i, c)}{\|n_i\| \, \|\mathrm{proj}(p_i, c)\|} \right)$$

$$\mathrm{proj}(p_i, c) = p_i - c_p - \left(\frac{p_i \cdot c_d - c_p \cdot c_d}{\|c_d\|^2} \right) c_d$$

The chosen distance function is a weighted sum of two distances

$$\mathrm{d}(p_i, n_i, c) = w \cdot \mathrm{d}_E(p_i, c) + (1 - w) \cdot \mathrm{d}_N(p_i, n_i, c) \tag{6}$$

Figure 4 shows an example where the cylinder pose is approximately recovered from the point cloud. It should be observed that the cylinder model parameters and the point-to-model distance are the only parts of the algorithm depending on the specific object shape.

5. EXPERIMENTAL EVALUATION

Two experimental sessions were conducted at the Lake of Garda (Italy) to assess the performance of the system. The prototype was fixed to a floating unit, as shown in figure 5. Cylindrical objects were submerged at a depth that ranged from $1.8m$ to $3m$. PVC tubes of different colors were used ($10cm$ diameter, $1m$ length). Table 1 summarizes the experimental conditions of the two sessions. In both sessions the average depth of the camera was about $40cm$ below water level. In the two sessions the frame rate was set at different rates resulting in the effective frame rates in Table 1.

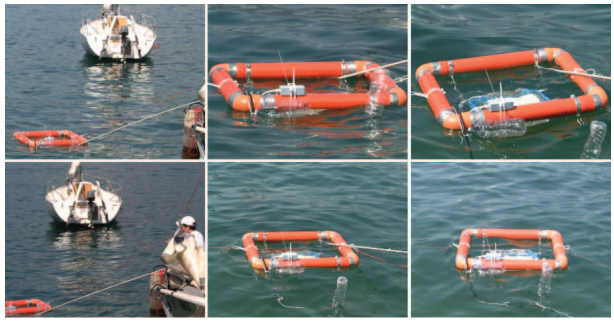

Fig. 5. Images of the experimental sessions.

location	session 1 Bardolino	session 2 Malcesine
time	$10:00 - 12:00$	$10:00 - 12:00$
weather	clouds	sun
floor	stones and algae	stones and algae
object depth	$[1.8m, 2.3m]$	$[2m, 3m]$
camera depth	$\sim 40cm$	$\sim 40cm$
max canister temp	68^oC	63^oC
max fps	$3.31Hz$	$6.62Hz$

Table 1. Experimental session data.

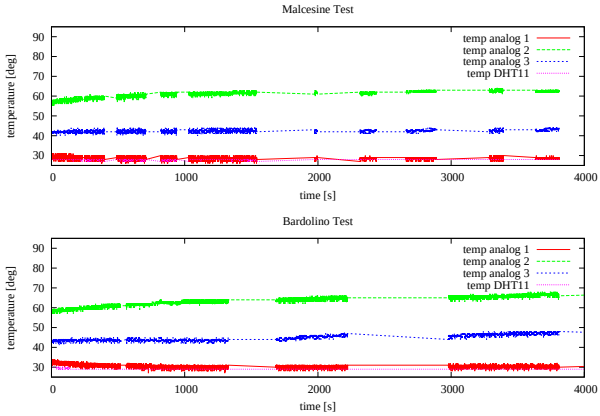

Fig. 6. Temperature trends over time for the experimental sessions in Bardolino and Malcesine.

One of the goals of the experiments was to test the physical properties of the embedded system, such as the water-proof endurance of the low-cost canister and the thermal balance of the electronic devices inside it. The sealed canister has proven to be able to transfer heat through thermal conduction and without any active device like a fan. Figure 6 illustrates the temperature values measured during the two sessions, each lasting more than one hour. Sensors *temp analog 1* and *DHT11* measure the environmental temperature inside the canister, *temp analog 2* is placed on the heat sink of the CPU and *temp analog 3* is placed on the SSD hard drive. The

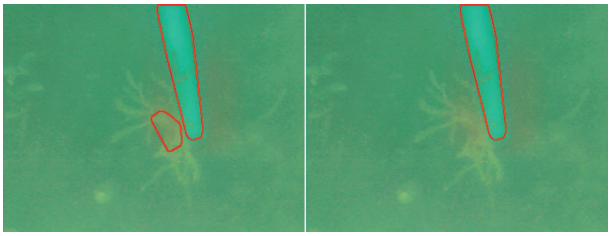

Fig. 7. Example of ROI (left) and CMask (right) computed on the same input frame.

Pre-processing	Frames	CMask	ROI	ROI/CMask
no	304	9.32%	33.18%	3.72
yes	304	9.07%	11.98%	1.31

Table 2. ROI and CMask computation w.r.t. image pre-processing.

Num. Frames	Distance [mm]	
	avg	std.dev.
302	1441	169

Table 3. Mono-camera estimated distance.

maximum temperature value of 68° C has been measured by the sensor on the CPU, as it might be expected. All measurements seem to converge to stable and safe values. The approximate power consumption measured in the laboratory for the whole system is about 17 W. Thus, the thermal dissipation behavior of the embedded system has proven adequate for its correct operation.

The image pre-processing algorithms discussed in section 4.1 significantly influence underwater object detection performance. In order to assess the effectiveness of the pre-processing algorithms, the CMask and the ROI have been computed on a set of 304 sample images. Results have been computed on both the raw and the pre-processed images. The average percentage of CMask and ROI pixels over the whole image and the ratio between the two quantities are reported in Table 2. The region found by the CMask only slightly depends upon the quality of the input image (since it exploits the information about the color of the object), whereas the computed ROI is more affected by the image quality. The ROI in the pre-processed image is on average only one third of the ROI computed in the raw image. Thus, assuming that the CMask reasonably approximates the groundtruth, the ROI provides an adequate estimate of the object for underwater detection, as long as appropriate pre-processing is performed. Figure 7 shows an example of ROI and CMask computed on the same input frame. The complete mono-camera processing is performed on average in 74.82 ms, with a standard deviation of 3.20 ms.

Mono-camera images have been used to estimate the pose of a cylindrical pipe, as discussed in section 4.1. The algorithm computes all the parameters of the cylinder axis that allow localization of the target object. However, during experiments at the Garda lake, the embedded system swung rather fast attached to the floating support, due to the continuous waves (see Figure 5). In such experiments no groundtruth is usually available, therefore a parameter invariant to camera motion is required to assess the precision of the proposed method. The object lies on the lake floor and the camera depth remains approximately constant. Thus, the distance between the camera center

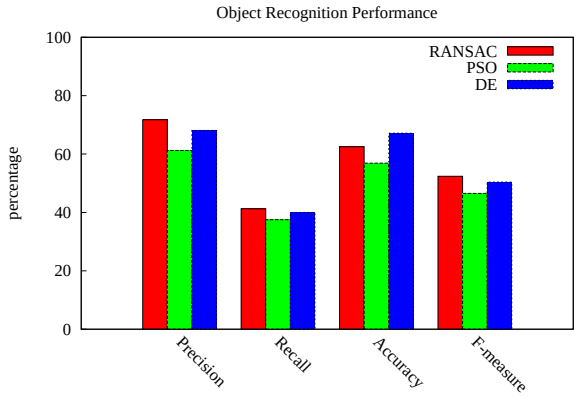

Fig. 8. Object recognition results on the point cloud.

and the cylinder axis in equation (3) approximately meets this pre-requisite. Table 3 illustrates the average distance and the standard deviation of the axis computed in a sequence of 302 frames. The standard deviation of 17 cm is due to both the estimation error and to the slight variation of distances caused by waves.

A second set of experiments has aimed at assessing the object detection and pose estimation performance on the point cloud acquired in the stereo camera configuration. Unfortunately, the point cloud obtained from the underwater dataset is rather sparse and noisy. As mentioned above, in water the embedded system was attached to a floating support, and the camera baseline swang due to waves. Since the webcams are not synchronized by a hardware trigger, the computed disparity image becomes noisy and inaccurate. The triggering delay between two cameras is on average about 65 ms. Thus, an alternative dataset of images has been acquired in air to obtain an evaluation of the full stereo-processing pipeline. In this alternative setting, the target cylindrical pipes lay in a dry river bed among sand and stones, and the embedded acquisition box is manually moved. Figure 8 summarizes the object recognition results for RANSAC, PSO, and DE recognition algorithms. The three algorithms obtain comparatively similar recognition results, but as could be expected RANSAC is at least one order of magnitude faster than the alternative algorithms. Figure 4 shows the recognition of a cylinder from 3D point cloud data.

6. CONCLUSIONS

This paper has presented the design and experimental evaluation in real underwater environment of an embedded vision system for underwater object detection. The design approach has focused on a low power budget, low cost, and inevitably low performance embedded system. The system has proven thermally stable and capable of guaranteeing a level of autonomy of at least two hours of video acquisition. Suitable preprocessing and image enhancement algorithms have proven effective in improving underwater images, thereby enabling detection of regions of interest as well as detection and localization of known objects in sequential image streams gathered from a single camera. On the contrary, the 3D point clouds obtained from stereo processing of multiple underwater camera streams have not allowed reliable object detection and localization. The

stereo processing pipeline has been eventually evaluated on a dataset obtained in outdoor, in-air conditions. The low quality of the 3D point cloud computed from underwater images is mainly due to lack of hardware synchronization between the webcams, which of course is affected by the in-motion nature of underwater perception (with possibility of strong fluctuations at low depths or in presence of water streams). Moreover, the accuracy, resolution, and acquisition rate afforded by inexpensive webcams have proven inadequate for challenging underwater perception tasks.

Although the experimental evaluation has shown its strong limitations, the prototype described in this work has provided insights for development of more advanced underwater vision systems. Indeed, low-cost stereo vision technologies seem adequate for evaluation of early-stage image processing algorithms in underwater environments. The essential requirement for the next version of our system is the adoption of more advanced vision sensors. Moreover, based on the experience gathered with this prototype, we are currently investigating different tradeoffs between power demand and performance, including better heath trasfer mechanisms from the electronics to the canister body. The final aim is to support real-time stereo vision algorithms in fully autonomous configuration.

ACKNOWLEDGEMENTS

We thank Andrea Minari for his support in the experimental evaluation of the system. We also thank the municipality of Bardolino (Garda Lake, Italy) for the permission to carry out underwater experiments at the public pier.

REFERENCES

Andersen, M., Jensen, T., Lisouski, P., Mortensen, A., Hansen, M., Gregersen, T., and Ahrendt, P. (2012). Kinect depth sensor evaluation for computer vision applications. Technical report, Technical report ECETR-6, Dep. of Engineering, Aarhus University (Denmark).

Bazeille, S., Quidu, I., Jaulin, L., and Malkasse, J. (2006). Automatic underwater image pre-processing. In *Proceedings of CMM'06*.

Brandou, V., Allais, A.G., Perrier, M., Malis, E., Rives, P., Sarrazin, J., and Sarradin, P.M. (2007). 3D reconstruction of natural underwater scenes using the stereovision system iris. In *OCEANS 2007 - Europe*, 1–6.

Campos, R., Garcia, R., and Nicosevici, T. (2011). Surface reconstruction methods for the recovery of 3D models from underwater interest areas. In *OCEANS, 2011 IEEE - Spain*, 1–10.

Corchs, S. and Schettini, R. (2010). Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 2010, 1–14.

Eustice, R., Singh, H., Leonard, J., Walter, M., and Ballard, R. (2005). Visually navigating the rms titanic with slam information filters. In *Proceedings of Robotics: Science and Systems*. Cambridge, USA.

Fischler, M.A. and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6), 381–395. doi: 10.1145/358669.358692.

Garcia, R. and Gracias, N. (2011). Detection of interest points in turbid underwater images. In *IEEE OCEANS*, 1–9.

Gordon, A. (1992). Use of laser scanning system on mobile underwater platforms. In *Autonomous Underwater Vehicle Technology, 1992. AUV'92., Proceedings of the 1992 Symposium on*, 202–205. IEEE.

Horgan, J., Flannery, F., and Toal, D. (2009). Towards real time vision based uuv navigation using gpu technology. In *OCEANS 2009 - EUROPE*, 1–6.

Ishibashi, S. (2009). The stereo vision system for an underwater vehicle. In *OCEANS 2009 - EUROPE*, 1–6.

Jonsson, P., Sillitoe, I., Dushaw, B., Nystuen, J., and Heltne, J. (2009). Observing using sound and light: a short review of underwater acoustic and video-based methods. *Ocean Science Discussions*, 6(1), 819–870.

Kim, D., Lee, D., Myung, H., and Choi, H.T. (2012). Object detection and tracking for autonomous underwater robots using weighted template matching. In *OCEANS, 2012 - Yeosu*, 1–5.

Narimani, M., Nazem, S., and Loueipour, M. (2009). Robotics vision-based system for an underwater pipeline and cable tracker. In *OCEANS 2009 - EUROPE*, 1–6.

Nicosevici, T., Gracias, N., Negahdaripour, S., and Garcia, R. (2009). Efficient three-dimensional scene modeling and mosaicing. *Journal of Field Robotics*, 26(10).

Oleari, F., Lodi Rizzini, D., and Caselli, S. (2013). A low-cost stereo system for 3d object recognition. In *Intelligent Computer Communication and Processing (ICCP), 2013 IEEE International Conference on*, 127–132. doi:10.1109/ICCP.2013.6646095.

Sanz, P.J., Penalver, A., Sales, J., Fornas, D., Fernandez, J.J., Prez, J., and Bernabe, J.A. (2013). Grasper: A multisensory based manipulation system for underwater operations. In *IEEE International Conference on Systems, Man, and Cybernetics*, 1–9.

Schattschneider, R., Maurino, G., and Wang, W. (2011). Towards stereo vision slam based pose estimation for ship hull inspection. In *OCEANS 2011*, 1–8.

Targhi, A.T., Hayman, E., Eklundh, J.O., and Shahshahani, M. (2006). The eigen-transform and applications. In *Proceedings of the 7th Asian conference on Computer Vision - Volume Part I*, ACCV'06, 70–79. Springer-Verlag, Berlin, Heidelberg.

Ugolotti, R., Nashed, Y.S., Mesejo, P., Ivekovi, S., Mussi, L., and Cagnoni, S. (2013). Particle swarm optimization and differential evolution for model-based object detection. *Applied Soft Computing*, 13(6), 3092–3105. doi:http://dx.doi.org/10.1016/j.asoc.2012.11.027.

Wang, H., Sun, H., Shen, J., and Chen, Z. (2011). A research on stereo matching algorithm for underwater image. In *4th International Congress on Image and Signal Processing (CISP)*, volume 2, 850–854.

Yu, S.C., Kim, T.W., Asada, A., Weatherwax, S., Collins, B., and Yuh, J. (2006). Development of high-resolution acoustic camera based real-time object recognition system by using autonomous underwater vehicles. In *OCEANS 2006*, 1–6.