The Role of Vision for Underwater Vehicles

J. Santos-Victor J. Sentieiro

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisboa – Portugal

Abstract – Different sensing techniques have been used for a long time in the development of remote operated vehicles (ROVs) and, in the recent years, autonomous underwater vehicles, (AUVs). Sonar has been the most popular choice in what concerns depth information or obstacle detection in most of the existing systems. Vision, however is a powerful sensing modality and could be used in many tasks where accurate measurements at a short range, are needed.

This paper discusses the possible use of vision in the context of autonomous underwater vehicles. Tasks, such as object avoidance or recognition, grasping, docking, sea bed reconstruction, underwater surveillance, inspection, cable maintenance, are among the set of those where computer vision may have an important role.

As an example, we present a vision system designed for recursive depth estimation based on a underwater vehicle equipped with a camera. Results obtained with underwater images are presented.

I. VISION FOR UNDERWATER VEHICLES

The technology of autonomous or semi-autonomous underwater vehicles has come to a point where very performing systems have been developed. Sonar has been the most widely used sensing modality for tasks such as depth extraction or obstacle detection.

However, vision is a sensing modality which provides very high resolution, yields accurate depth measurements specially at short range and, therefore, could be used for a set of tasks in autonomous underwater systems.

On the other hand, Computer Vision has now grown to a mature point where robust, sophisticated systems have been developed and are able to operate at fast speeds. Besides depth recovery, which could be a goal on itself, interesting tasks that could be addressed by using vision techniques are surveillance, recognition, visual-guided docking or manipulation, obstacle detection [5], trajectography, low altitude navigation [11], and so on.

For docking or manipulation, vision can directly provide feedback information to drive the motor actions of the underwater vehicle. The coordination of vision and action is, in fact, one of the core ideas of the recent research in the so called active vision [1]. Bathymetric information can be used for different purposes but in a similar way, one could use vision to estimate the vehicle position in a known environment. In the last few years, computer vision researchers have designed different obstacle avoidance techniques which tend to be robust, fast and are suitable for underwater application.

Naturally that the underwater environment offers some new challenges when compared to land robotic applications. For this reason, some specific solutions must be found in order to cope with the specificity of the working environment. However, the key message is that visual sensing may be an important improvement for some of the present autonomous underwater vehicles.

In the following sections we will describe thoroughly a computer vision system for sea bed reconstruction. We assume that an underwater vehicle is equipped with a camera while performing a given mission. During this mission, the images are used to recursively reconstruct the 3D structure of the surrounding environment.

II. DEPTH ESTIMATION

This paper describes research done during the MOBIUS project funded the European Community MAST (Marine Science Technology) program [12]. We will focus on the problem of extracting depth information by using an image sequence acquired by a camera installed on a AUV. As a first approach, we assume that the AUV trajectory is known, even though it is possible to recover the vehicle motion using some a priori known landmarks in the sea bed.

A stochastic approach is adopted to model the existence of uncertainty in every depth estimate, and well established techniques, like Kalman filtering, are used to reduce the uncertainty of the estimated depth maps, over time. The system comprises three major processing modules: the matching process, the regularization procedure and the Kalman filtering stage.

The matching process establishes correspondences between points in images acquired at different time instants. Stereo is one of the major cues for depth perception in many biological vision systems. To compute these correspondences (disparities), we use a correlation-like method together with prior knowledge of the camera motion, serving as a geometric constraint (epipolar line) and interpolation to achieve sub pixel resolution.

Since the matching process is ill-posed, a regularization stage is formulated, to constrain the desired disparity field to smooth solutions. In this way, the noise levels of the disparity estimates can be reduced, and new estimates can be calculated to fill in the empty areas.

Kalman filtering is used in the final stage of this work, to integrate multiple disparity measurements over time, in order to produce a more reliable depth estimate. Therefore, one can benefit from the advantages of having closely spaced image view points (which simplifies the matching problem, but usually degrades the depth estimates precision) without jeopardizing the precision of the depth estimates computation.

The system has been tested using both synthetic images and real underwater images acquired in an experimental pool. The results obtained are accurate and the vision system is able to recover the depth structure of the surrounding scene. Several applications are suggested, for AUVs, and improvements to the system are finally pointed out and discussed.

In the following sections, the models involved in the depth from motion algorithms, are described. We then show the results obtained underwater images acquired within the context of MOBIUS. Finally, we draw some conclusions and some future directions of work are pointed out.

III. MODEL

This section is devoted to the study and description of all the models involved in the 3D dynamic vision system. First, the dynamics associated to the position of a 3D point relatively to a moving camera referential is derived. Then, the camera model is introduced, and used to obtain the equations that describe the apparent motion induced in the image plane. Finally a model for the uncertainty affecting the system, is established.

Consider a camera moving with relation to a fixed point in the space. Let {C} be a Cartesian coordinate system (frame) attached to the camera. Let $\vec{\omega}$ and $\vec{T}$, be the angular and translational velocities of the camera with relation to a fixed referential, and let $\vec{P}$ be the position vector of a fixed point in the space. The velocity of P with relation to {C} (rigid body motion) is described by the following differential equation [8]:

$$\frac{d\vec{P}}{dt} = -\vec{T} - \vec{\omega} \times \vec{P} \cdot \tag{1}$$

To determine how this motion is perceived in the image plane, we have used a *pinhole* camera model [8, 3]. According to this model, the location of a given image pixel, $(x,\ y)$, is determined by the perspective projection in the image plane, of the corresponding 3D point, $[X\ Y\ Z]^T$:

$$x = \frac{X}{Z}, \qquad\qquad y = \frac{Y}{Z} \cdot \tag{2}$$

By using the camera model in (1), and eliminating Z, we finally obtain a new equation set, that expresses the apparent motion (velocity field) induced in the image plane by the actual camera motion [6, 7, 3][1]:

$$\begin{aligned} \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} &= \frac{1}{Z}\begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \end{bmatrix}\vec{T} \\ &+ \begin{bmatrix} xy & -(1+x^2) & y \\ (1+y^2) & -xy & -x \end{bmatrix}\vec{\omega}, \end{aligned} \tag{3}$$

$$\dot{Z}_{(t)} = (\omega_y x - \omega_x y) Z_{(t)} - T_z \cdot \tag{4}$$

Uncertainty (the simplified camera model, errors associated to the camera motion and parameters, the image acquisition process, etc) is modeled by additive white Gaussian noise in (3,4), and the final discrete model is obtained by approximating derivatives in (3, 4) with forward differences (with sampling period τ): [2]

State equation :

$$Z_{[t+\tau]} = a_{[t]}\ Z_{[t]} + b_{[t]} + \eta \cdot \tag{5}$$

Observer equation (measurements):

$$\vec{\mathbf{d}} = \begin{bmatrix} x_{[t+\tau]} - x_{[t]} \\ y_{[t+\tau]} - y_{[t]} \end{bmatrix} = \mathbf{C}_{[t]} \frac{1}{Z_{[t]}} + \mathbf{D}_{[t]} + \mu, \tag{6}$$

[1]For simplicity, we have not explicitly written the different variables as time functions (e.g. $x(t)$ instead of x), as it should be clear from the context.

[2]Note that the image coordinates in (6) correspond to a virtual camera with unitary focal length. The actual image pixel coordinates are obtained using the *intrinsic* camera model parameters [8]. See [13] for details.

where η is a zero mean Gaussian random variable with variance r, μ is a zero mean Gaussian random vector with covariance matrix Q and $\{\mu, \eta\}$ are independent. The terms $a_{[t]}$, $b_{[t]}$, $\mathbf{C}_{[t]}$ and $\mathbf{D}_{[t]}$ depend on the motion parameters and the image plane coordinates of a given pixel [13].

IV. SYSTEM DESCRIPTION

The complete system structure is shown in Fig. 1. At every sample instant, a new image is acquired by the moving camera and a new depth map is computed.

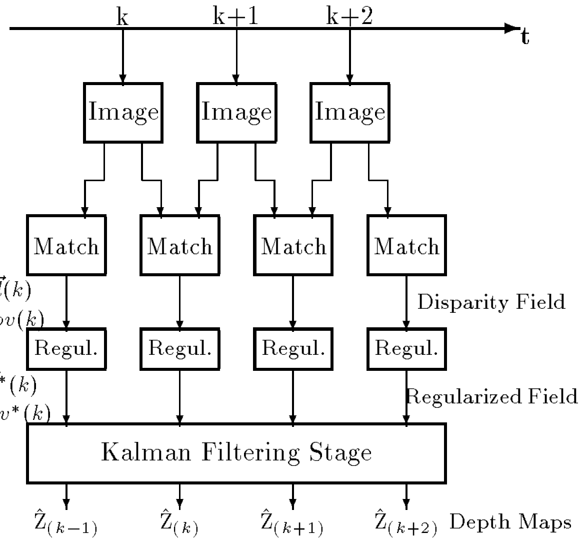

Figure 1: Block Diagram of the Vision System

A matching algorithm is applied to each pair of successive images, to compute the disparity vector field and uncertainty estimates. This vector field is the input to a regularization stage, used to decrease the noise levels and fill in unavailable estimates. Every new observation (regularized disparity vector) is finally used in a Kalman filtering module to update an estimated depth map resultant from previous measurements, thus reducing the uncertainty over time.

A. Matcher

To determine the displacement of each image pixel, induced by the camera motion in the static environment, we have used a correlation based technique, derived from the Sum of Squared Differences (SSD) method [2, 6].

To evaluate potential match candidates, p'_t and $p'_{t+\tau}$, from images acquired at time t and $t+\tau$, it is assumed that homologous points have similar gray-levels. Hence, the sum of the squared gray level differences, for pixels inside windows centered in p'_t and $p'_{t+\tau}$, can be used to quantify the "likelihood" of a matching decision [10].

Moreover, the location of a pixel $p'_{t+\tau}$, homologous to p'_t, is constrained to a line, *the epipolar line*, in the image at $t+\tau$, determined by the camera motion and camera parameters [13]. By introducing the prior knowledge of the epipolar line, we define the *Extended Sum of Squared Differences* (ESSD) criterion that measures gray level compatibility and simultaneously penalizes deviations from the epipolar line:

$$ESSD(u,v,x,y) = \sum_{\alpha,\beta} \phi_{(\alpha,\beta)} \left[I_{(t,\alpha,\beta)} - I_{(t+\tau,\alpha+u,\beta+v)}\right]^2 + \lambda_{ep}\, d_{ep}^2(x,y,u,v), \tag{7}$$

where $I_{(t,x,y)}$ designates the pixel (x,y) of the image acquired at time t, $d_{ep}(x,y,u,v)$ is the distance of the matching candidate pixel, $I_{(t+\tau,x+u,y+v)}$, to the epipolar line, and $\phi_{(\alpha,\beta)}$ is a weighting function. The contribution of the prior knowledge to the final cost functional[3] is quantified λ_{ep}.

Even though the use of the epipolar constraint significantly improves the matching accuracy, there are often slight changes in illumination which disturb the matching cost functional and, as a result, there are large image areas, where the ESSD does not exhibit a clear minimum, and matching cannot be performed [14].

The proposed method is based on the following idea. We basically partition the image in two sets of pixels:

- $\mathcal{A}$ - A pixel is part of set $\mathcal{A}$ if in the neighborhood of the pixel, the image presents significant gray level changes.
- $\mathcal{B}$ - The pixel is included in region $\mathcal{B}$ if the gray level is approximately constant in a neighborhood of the pixel.

In the situation $\mathcal{A}$, the pixel will be matched based on the changing component of the gray level, even though there might be some differences in the average gray level. This is accomplished by using the following cost functional which performs a sort of *local equalization* procedure:

$$ESSD_{(eq)}(u,v,x,y) = \sum_{\alpha,\beta} \phi(\alpha,\beta)\left[I_{(t,\alpha,\beta)} - \frac{I_t^{dc}}{I_{(t+\tau)}^{dc}} I_{(t+\tau,\alpha+u,\beta+v)}\right]^2 + \lambda_{ep} d_{ep}^2(x,y,u,v), \tag{8}$$

where I_t^{dc} is the average gray level in the matching window in image $I(t)$. In situation $\mathcal{B}$, where the image has

[3] Since the computation of the epipolar line is based on the camera and motion parameters, the value of λ_{ep} should be related to the uncertainty associated to these parameters.

a flat brightness pattern, the matching is done using the ESSD cost functional described by (7).

As a result, we are able to match a large number of areas in the image provided that there is some signal variation in the neighborhood of a pixel, even though there might be some illumination changes. To determine whether a pixel is in situation $\mathcal{A}$ or $\mathcal{B}$, we compare the gray-level variance to a specified threshold.

In [10] the disparity vector is estimated by fitting a quadratic surface to a neighborhood of the minimum SSD point, while in [6] two one dimensional quadratic curves are fit, both in the x and y directions. Due to the increased robustness to noise and simplicity, we adjust two one dimensional quadratic curves in each direction, in a neighborhood of the minimum ESSD point:

$$q(u) = au^2 + bu + c, \tag{9}$$

where a, b and c are estimated from the data. The minimum point can then be determined, with sub pixel resolution, yielding the optimal disparity estimate:

$$\hat{u}_{opt} = -\frac{b}{2\,a}, \tag{10}$$

whenever the coefficient a is different from 0.

The uncertainty of the estimate is related to the shape of the ESSD error surface [10, 2, 6]. In [2], the uncertainty is a function of the SSD surface curvature along the principal axis, and in [10] error propagation techniques in the SSD cost functional, were used. In each direction, we will approximate the estimate variance by [6]:

$$\sigma_u^2 = \left(\frac{d^2 q(u_{opt})}{du^2}\right)^{-2} q(u_{opt}) \cdot \tag{11}$$

The first term of (11) expresses the decrease of the uncertainty with the increase of the error surface curvature, while the second term is a normalizing factor dependent of the minimum ESSD surface value.

B. Regularization

Many visual reconstruction processes, aimed at recovering 3D information by processing 2D image data, are inverse, ill-posed problems (e.g. the estimation of disparity fields between successive images) [17, 4].

Using the framework of regularization, ill-posed problems can be reformulated as variational principles, by introducing *a priori* knowledge about the solution. Standard Tikhonov regularization [18], uses stabilizing functionals to restrict the space of admissible solutions to smooth functions.

To determine a regularized function $\mathcal{U}$, using a set of data $\mathcal{D}$, we define an error function $\Psi_d(\mathcal{D},\mathcal{U})$, to measure the compatibility of the proposed solution and the data, a stabilizing functional $\Psi_p(\mathcal{U})$ that quantifies the smoothness conditions on the desired solution, and search for the $\mathcal{U}^*$ that minimizes the following cost criterion [15].

$$\Psi(\mathcal{U},\mathcal{D}) = \Psi_d(\mathcal{U},\mathcal{D}) + \lambda\Psi_p(\mathcal{U}) \cdot \tag{12}$$

The choice for both functionals Ψ_p and Ψ_d guarantees that, under certain weak conditions, a solution for the optimization problem exists [2]. For the regularization of the displacement/disparity vector field, a *thin membrane* [2, 17, 15] stabilizing functional was used:

$$\begin{aligned}\Psi(\mathcal{U},\mathcal{D}) &= \lambda \sum_{x,y} (\vec{u}-\vec{d})^T Q^{-1} (\vec{u}-\vec{d}) \\ &+ \int\!\!\int trace\ \{(\nabla\vec{u})(\nabla\vec{u}^T)\} dxdy, \qquad (13)\\ Q &= \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}, \qquad (14)\end{aligned}$$

where $\vec{u}(x,y) = [u(x,y) \;\; v(x,y)]^T$ is the regularized solution, σ_u^2 and σ_v^2 are the variances of the x and y components of the measured disparity vectors, ∇ is the gradient operator and λ quantifies the relative weight of the fitness to data term, in the global cost functional. [4]

The domain of the surface $\vec{u}(x,y)$ is usually discretized using either the finite differences method or the finite element method [2, 8, 17]. Applying finite element analysis, as proposed by Terzopoulos [17], to the functionals involved in the minimization problem we obtain:

$$\begin{aligned}\Upsilon(\mathcal{U},\mathcal{D}) = \sum_{x,y} [\lambda(\vec{u}-\vec{d})^T Q^{-1}(\vec{u}-\vec{d}) + \\ \| \vec{u}_{(x+1,y)} - \vec{u}_{(x,y)} \|^2 + \| \vec{u}_{(x,y+1)} - \vec{u}_{(x,y)} \|^2] \cdot \qquad (15)\end{aligned}$$

To minimize (15), the Gauss-Seidel relaxation algorithm is used, and $\vec{u}$ is obtained iteratively at each point as a function of the values of its neighbors (similar equations are obtained for u and v):

$$u_{(x,y)}^{n+1} = \bar{u}_{(x,y)}^n + \frac{\lambda\sigma_u^{-2}}{1+\lambda\sigma_u^{-2}}(\, u_{(x,y)}^0 - \bar{u}_{(x,y)}^n \,), \tag{16}$$

where $u_{(x,y)}^0$ is the measured x disparity at pixel (x,y) and $\bar{u}_{(x,y)}$ is the local mean value, given by:

$$\bar{u}_{(x,y)} = \frac{u_{(x+1,y)} + u_{(x-1,y)} + u_{(x,y+1)} + u_{(x,y-1)}}{4}. \tag{17}$$

To determine the uncertainty associated to the regularized disparity field, the uncertainty values must be updated, as the regularization procedure imposes changes in the original field. Unfortunately, as the regularization iterations proceed, $\bar{u}^n$ will depend on an increasing number

[4] Whenever the measured variable $\vec{d}$ is unavailable, λ will be set to zero.

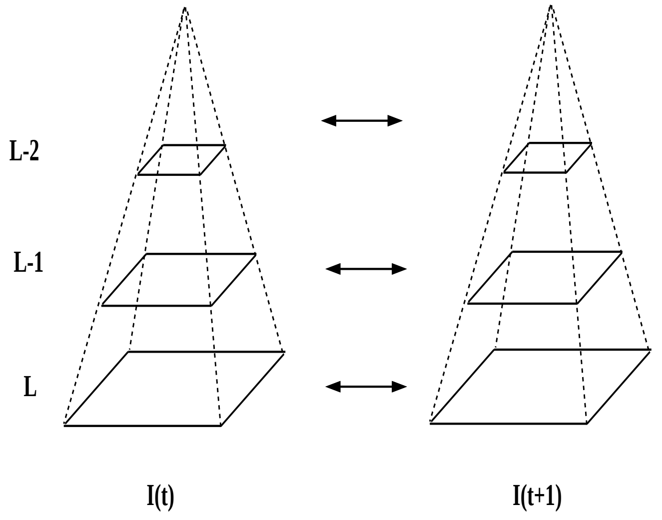

Figure 2: Coarse-to-fine control strategy based on a Gaussian Pyramid.

of samples, thus hardening the calculation of the uncertainty. For computation efficiency, we have adopted the following simpler expression:

$$var[u^{n+1}] = \frac{\alpha_m^2 var[\bar{u}^n] + \alpha_0^2 var[u^0]}{\alpha_m^2 + \alpha_0^2}. \qquad (18)$$

C. Coarse to fine control strategy

To improve efficiency and accuracy, the complete algorithm runs in a coarse-to-fine control strategy, based on a Gaussian pyramid. The pyramid is obtained by using the acquired image at the base level, and successively filtering and subsampling the image to obtain the coarser levels of resolution. At each level the image dimension is half of the precedent level.

The matching is initiated in the coarsest level of resolution. At this level, estimates of the disparity vectors are computed, and the resulting vector field is regularized. This process is computationally efficient due to the reduced image size and the small search area for the matching process. The efficiency of the information propagation with the regularization is also increased [16].

Once solved the matching/regularization at the coarsest level, the disparity vector field is projected to the subsequent (higher) resolution level, thus being a disparity estimate for the new level. The algorithm then proceeds by performing matching in a small neighborhood of the predicted disparity. The same regularization procedure is then applied. This process is repeated until the final estimates are obtained at the highest level of resolution. Fig. 2 illustrates the pyramid structure and the information flow among the different levels of resolution.

The coarse-to-fine control strategy leads to large computational gains, since the most costly computations are performed at the coarsest level of resolution. Moreover this strategy improves the speed of the spatial propagation of the regularization constraints. The results are often better than the single resolution algorithms, since the search-space and spatial support of the matching are extended [16].

D. Kalman Filtering - Recursive Estimation of Depth Maps

We will now see how to use Kalman filtering theory, to combine different disparity measurements over time, yielding a more reliable single depth estimate.

Using the state space model framework, we can formulate an estimation problem, to determine the value of $Z_{[t]}$, based on the noisy observations $\vec{d}_{[t]}$. Depth is estimated independently at each pixel, while spatial dependencies are embodied in the regularization process.

This state estimation problem can be conveniently dealt with, using Kalman filtering techniques. Since the observation equation is non-linear in the state variable, the discrete-time Extended Kalman Filter (EKF) [9] must be used. The filtering equations comprise a prediction step and an update/filtering step. In the prediction stage, the future values of depth and associated uncertainty are predicted, based on past information and on the dynamic model:

Prediction :

$$\hat{Z}(t/t-1) = a_{[t-1]}\hat{Z}(t-1/t-1) + b_{[t-1]}, \qquad (19)$$

$$\sigma^2_{Z(t/t-1)} = a^2_{[t-1]}\sigma^2_{Z(t-1/t-1)} + r, \qquad (20)$$

where $\hat{Z}(t/t-1)$ is the predicted depth value for time t, based on the data available up to time $t-1$, $\sigma^2_{Z(t/t-1)}$ is the corresponding variance and r is the variance of the motion equation noise (see Section III.).

At time t, a new disparity measurement is available, and the predicted depth can be updated. This is the filtering step procedure and the EKF uses a linearized version of the observation equation in a neighborhood of the predicted value.

$$\vec{\mathbf{d}}_{[t]} \approx \vec{\mathbf{d}}^L_{[t]} = \mathcal{C}_{[t]} Z_{[t]} + \mathcal{D}_{[t]} + \mu, \qquad (21)$$

where $\mathcal{C}_{[t]}$, $\mathcal{D}_{[t]}$ are the coefficients of the linearized model [13]. The filtering step is given by

Filtering :

$$\mathbf{K}_t = \sigma^2_{Z(t/t-1)}\mathcal{C}^T_{[t]}\left[\,\mathcal{C}_{[t]}\sigma^2_{Z(t/t-1)}\mathcal{C}^T_{[t]} + \mathbf{Q}_t\,\right]^{-1}, \quad (22)$$

$$\sigma^2_{Z(t/t)} = (1 - \mathbf{K}_t\mathcal{C}_{[t]})\sigma^2_{Z(t/t-1)}, \qquad (23)$$

$$\hat{Z}(t/t) = \hat{Z}(t/t-1) + \mathbf{K}_t\,(\,\vec{d}_{[t]} - \mathbf{C}^L_{[t]}\hat{Z}(t/t-1) - \mathcal{D}_{[t]}\,), \quad (24)$$

with $\hat{Z}(0)$ and $\sigma^2_{Z(0)}$ being the initial depth and depth uncertainty estimates, and $\mathbf{K}_t$ being the Kalman gain.

Warping the Depth Map

To complete the EKF analysis, there is still an additional problem to be solved, concerning the prediction step.

The predicted depth value $\hat{Z}(t/t-1)$ is obtained by applying the motion equation. However, this predicted depth value, no longer corresponds to the original pixel (x, y), since the x and y coordinates have also changed. As the predicted depth values does not coincide with image pixel positions, the depth at the grid point (x, y) must be inferred from these values. This problem can be approached in several ways like bi-linear or bi-cubic interpolation [10, 6].

We compute the depth estimate at (x, y) as a weighted average of the estimates falling within a 3x3 window centered in (x, y). The warped depth uncertainty is estimated using error propagation techniques:

$$Z_{(x,y)} = \frac{\sum_{i=1}^{n} d_i^{-2} Z_{(x'_i, y'_i)}}{\sum_{i=1}^{n} d_i^{-2}}$$

$$\sigma^2_{Z(x,y)} = \frac{\sum_{i=1}^{n} d_i^{-4} \sigma^2_{Z(x'_i, y'_i)}}{(\sum_{i=1}^{n} d_i^{-2})^2} \quad (25)$$

where $Z_{(x'_i, y'_i)}$ stands for the predicted depth values, and d_i is the Euclidean distance from (x'_i, y'_i) to (x, y).

V. RESULTS

The 3D vision system was tested using a wide variety of synthetic and real images (see report [13] for details). The initial tests allowed to improve the matching and regularization stages and led to ideas to overcome the different problems.

In this section, we show some results using a set of real underwater images acquired within the MOBIUS project. A camera is moving vertically in a special test pool with a downwards speed of 1 m/s, and the images are acquired every 0.4 seconds, yielding a separation of 40 cm between successive images. The camera was specially designed for underwater operation without need for external lighting.

In all the examples, a three-level pyramidal structure was used for the matching and regularization procedures. The images acquired show some of the typical problems of underwater imagery like gradual changes in the global illumination. The equalization techniques described in the paper proved to be essential to get some acceptable results [14].

The *a priori* depth map estimate is 12 meters for every image pixel, with 0.001 m^2 variance. A 7x7 match window was used at the coarsest level, with $\lambda = 0.1$ and $\lambda_{epip} = 0$.

Fig. 3 shows results for one image sequence. On the left we have the images being acquired by the camera during the vertical motion. On the right column, we have the corresponding estimated depth maps shown in perspective. Notice the low contrast of the acquired images and the brightness variation along the image sequence. Nevertheless, due to the improved matching mechanism, the system has succeeded in estimating the depth structure of the scene.

VI. CONCLUSIONS

In this paper, we have argued for the use of computer vision techniques for underwater vehicles. Even though sonar has been the most frequent choice in the past years, computer vision techniques can now in a number of tasks in underwater vehicles technology, particularly at short range. Examples of these applications include visual-based navigation, visual-based docking, obstacle avoidance, inspection and surveillance, recognition, and so on.

Computer vision is attractive due both the high resolution provided by visual sensing and also by the recent accomplishments in land robotics. Naturally, the specific nature of the underwater medium may require specific techniques or approaches different from those used in land robotics, as the example given in this paper where the matching methodology has been optimized for underwater imagery.

As an example, we presented a depth from motion vision system designed to compute dense depth maps from an image sequence, acquired by a camera installed on a moving underwater vehicle. The system is based on a state space description of the depth from motion problem and models the existence of several uncertainty sources. A matching procedure that takes into account possible illumination variations, often found in underwater images, and includes the epipolar constraint was defined, and regularization is used to filter the disparity vector field. These processes work on a coarse-to-fine structure based on a Gaussian pyramid. Finally, depth measurements are combined over time, using Kalman filtering, to reduce uncertainty.

The system can be applied to a large number of problems in underwater robotics, where the estimation of the scene structure may be found useful. A particular application related to underwater robotics was presented, with underwater images acquired during the experiments

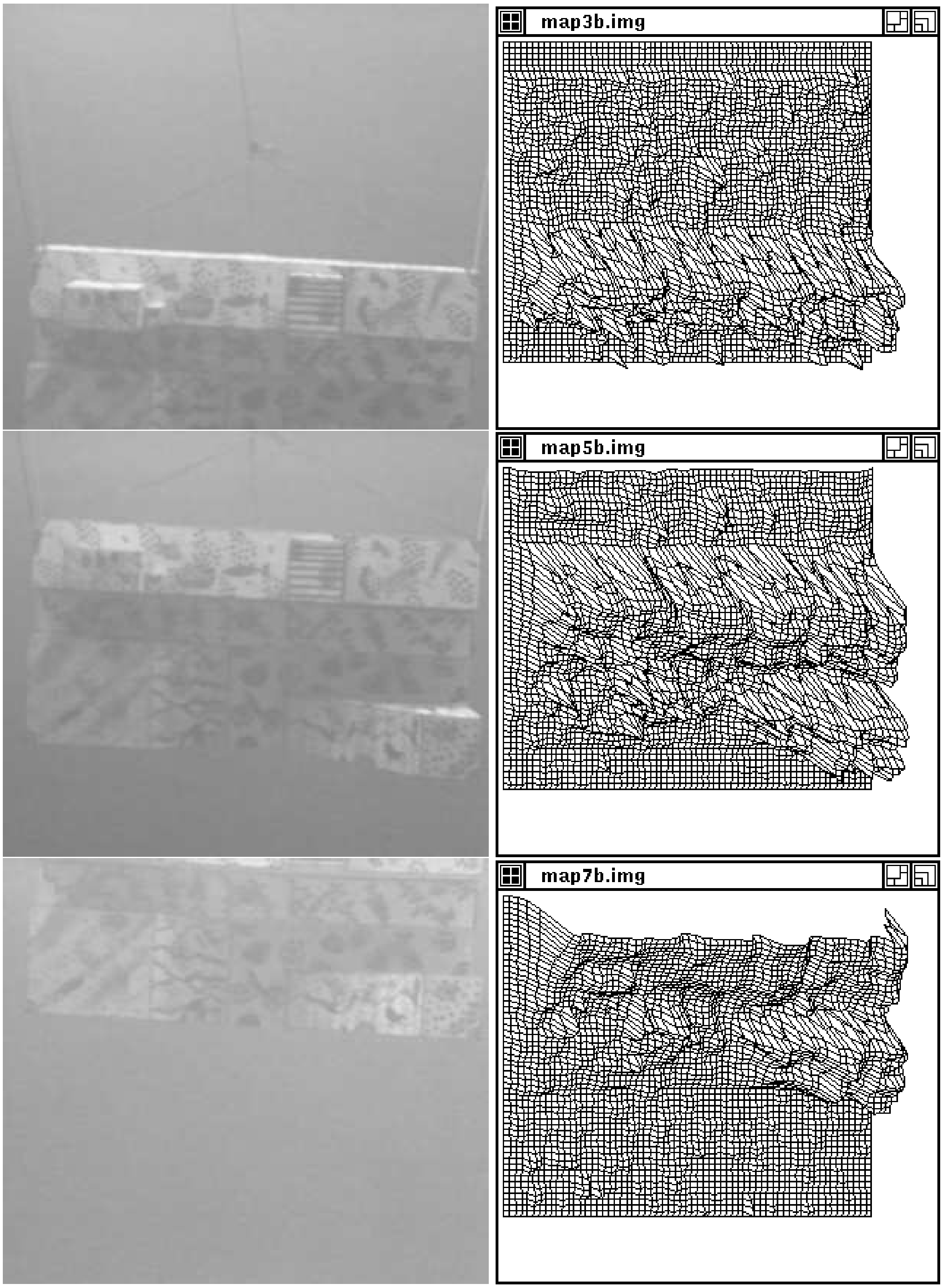

Figure 3: The left column shows the image sequence acquired during the camera movement in the experimental pool. On the right we show the perspective view of the reconstructed depth maps.

of Mobius project. The results show that the scene structure was recovered with some accuracy.

REFERENCES

[1] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1988.

[2] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(4):283–310, 1989.

[3] D. Ballard and C. Brown. *Computer vision*. Prentice-Hall, London, 1982.

[4] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, 1988.

[5] J. Gaspar, J. Santos-Victor, and J. Sentieiro. Ground plane obstacle detection with a stereo vision system. Submitted to the International Workshop on Intelligent Robotic Systems, July 1994.

[6] J. Heel. Dynamic motion vision. In *Proceedings of the DARPA Image Understanding Workshop*. Morgan-Kaufman Publishers, May 1989.

[7] B. Horn and B. Shunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[8] B.K. Horn. *Robot vision*. MIT Press, 1986.

[9] Jazwinski. *Stochastic processes and filtering theory*. Academic Press, 1970.

[10] L. Mathies, T. Kanade, and R.Szelisky. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(4):209–238, 1989.

[11] J. Santos-Victor, G.Sandini, F. Curotto, and S. Garibaldi. Divergent stereo in autonomous navigation: from bees to robots. *International Journal of Computer Vision*, 1994.

[12] J. Santos-Victor and J. Sentieiro. A 3d vision system for underwater vehicles: an extended kalman-bucy filtering approach. In *NATO Advanced Studies Institute - Acoustic Signal Processing for Ocean Exploration*, Madeira, Portugal, August 1992. Kluwer Academic Press.

[13] J. Santos-Victor and J. Sentieiro. A dynamic 3d vision system. Technical Report Mobius/rpt/02/92 JASV/JJSS, Instituto Superior Técnico, 1992.

[14] J. Santos-Victor and J. Sentieiro. Image matching for underwater 3d vision. In *International Conference on Image Processing: Theory and Applications*, San Remo, Italy, June 1993.

[15] R. Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision*, 3(5):271–301, 1990.

[16] D. Terzopoulos. Image analysis using multigrid relaxation methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):129–139, March 1986.

[17] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(4):413–424, July 1986.

[18] A.N. Tikhonov and V.A. Arsenin. *Solution of ill-posed problems*. Washington DC: Winston, 1977.