Exploiting Natural Contours for Automatic Sonar-to-Video Calibration

Christian Barat and Maria-João Rendas
Laboratoire I3S, CNRS-UNSA
Sophia Antipolis, France
Email: {barat}{rendas}@i3s.unice.fr

***Abstract*— The paper addresses the problem of co-registration of data acquired by two distinct ego-centric exteroceptive sensors mounted in an autonomous underwater vehicle: a video camera and a mechanically scanning profiler sonar. As it is the case for the http://www.laas.fr/reforme-CNRS.html experimental platform used in this study[1], these two perception sensors can often be installed in flexible geometric configurations on the robot's crash frame, and thus their relative orientation and displacement may change from mission to mission, even if during each mission they are kept fixed. The algorithm exploits the presence in the sea bottom of a boundary between two regions that can be discriminated both in the video frames and on the sonar profiles, and requires no specially tuned artificial settings. The registration is solved in a completely autonomous manner, using the ability of the robot to track the natural contour based on the video data. Our algorithm finds the mapping from the sonar coordinates to the image coordinates that minimizes the average entropy of a probabilitic model that is fit to the inverse mapping of the segmented images. It does not require reconstruction of the robot's trajectory during data acquisition, and is thus insensitive to positioning problems.**

I. Introduction and Problem Formulation

Most underwater platforms are equipped with several kinds of sensing modalities, the most common being acoustic (sonars) and optical (video) sensors. Frequently, the sensors are mechanically attached to the vehicle's crash-frame, and their position and orientation are manually adjusted at the beginning of each mission, being thus only approximately known. To be able to fuse the data provided by the two sensors – defined in coordinate systems attached to each equipment – it is necessary to know the (fixed but unknown) rigid motion that maps the coordinate system of one sensor to the coordinate system of the other sensor. This paper addresses the problem of learning this map by exploiting natural geometric structures present in the platform's environment.

More precisely, the criterion presented here exploits the presence in the robot environment of a boundary between two (or more) distinct sea bed regions, requiring no artificial settings of simple geometric structure, like man-made objects. We assume that the two regions induce distinctive signatures on both the video images and on the profiles returned by the sonar. In previous work, we have already demonstrated the feasibility of discriminating different types of sea bottom for these two sensing modalities [2], [3]. Assuming that both sensors are able to acquire data incoming from both regions (i.e., that the contour crosses the footprint of the video camera on the sea bottom, and that the scanned (vertical) conical sector intersects the boundary between the regions), our criterion is based on finding the co-registration map for which the partition of the sonar data induced by the image segmentation is the best, in a sense to be precised in a subsequent section.

The problem addressed in this paper presents some analogies with the problem of findind extrinsic camera parameters in robotic vision, extensively studied in the literature. The problem addressed here is different in several ways. First, the two sensing modalities are different in our case (a video camera and a sonar), while the majority of references consider the co-registration of two distinct images taken with the same camera from different spatial positions and orientations. The main innovative aspects of the approach proposed here is that we do not require the establishment of an explicit correspondence between geometric features extracted from the sensor data, like points, lines or quadratic curves. We exploit only the fact that the received data correspond to perceived regions of distinct characteristics (probability distributions). No explicit description of the boundary betwen regions is required, meaning that the method can accomodate the complex geometries of natural scenes.

A. Problem formulation

Consider Figure 1, where a schematic view of the problem addressed is given.

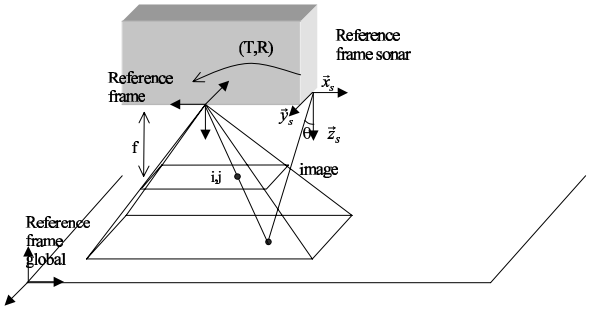

Fig. 1. Geometry of the registration problem.

Our goal is to be able to associate to each sonar profile, correspondent to a sea bed point described in known sonar-centric polar coordinates (ρ, θ), the location (i, j) of the same

[1]The ROV Phantom 500XP, produced by Deep Ocean Eng. IUSA.

point in an image taken at the same instant. We denote by $T_{S\to C}$ this application:

$$T_{S\to C}: \begin{array}{ccc} \mathbf{R}\times[-\Theta_{max},\Theta_{max}] & \to & \mathbf{R}^2 \\ (\rho,\theta) & \to & (i,j) \end{array}$$

As we will see in a subsequent section, this application is parameterized, besides the intrinsic parameters of the camera, by the rigid motion – rotation matrix R and translation vector t – that maps the coordinate system of the sonar to the coordinate system of the video camera. Note that our primary goal here is not to estimate this rigid motion, but rather to learn the application $T_{S\to C}$.

We assume the following experimental conditions (see Figure 1:

- The video camera is *oriented vertically*, looking approximately in a direction orthogonal to the (locally planar) sea bottom;
- The plane scanned by the sonar is *almost vertical*, and transversely to the robot direction of motion;
- The robot is *able to track a contour* between two distinct benthic regions using the video data, keeping at constant altitude from the sea bottom. We addressed this problem before, in [4], [3];
- During tracking, the tracked contour is persistent in the image sequence, and the directions sensed by the sonar hit the bottom in both adjacent regions;
- We have access to a segmented version of the acquired video sequence, that we denote by S_k, $k = 1, 2, \ldots$.

The paper is organized in the following way. The next section establishes the parametric equations of the application $T_{S\to C}$, showing how the intrinsic and extrinsic (R and t) parameters determine its geometry, and establishes the probabilistic model that is assumed for the sonar data and on which our estimation criterion is based. Section III presents the criterion that determines our estimate of the pixel coordinates of the sonar data. Finally, Section IV presents simulation results that illustrate the adequacy of the technique proposed, and Section V summarizes our main results and suggests directions for future work.

II. Geometric Constrains and Statistical Data Models

We consider that a fixed coordinate system attached to the vehicle body is specified, with respect to which all the other reference frames are defined, and that we designate by the "body-fixed" reference frame.

For simplicity we assume in this presentation that only two distinct regions are present in the sea bottom, and below we will indicate dependency on the region by sub-indices 0 and 1. However, our methodology can be extended with no additional difficulty to the more general case of a finite number of distinct regions.

A. Sonar Model

Let $\overrightarrow{x}_s, \overrightarrow{y}_s, \overrightarrow{z}_s$ be the coordinate system attached to the sonar, with origin $O^s = (O^s_x, O^s_y, O^s_z)$ and orientation $R_s \in SO(3)$ with respect to the body-fixed frame, such that the sensor scans the plane $(\overrightarrow{y}_s, \overrightarrow{z}_s)$,see Figure 1. The sonar profile corresponding to scanning direction θ receives a return from the intersection of the line $L_\theta(\ell) = \ell\,(\sin(\theta)\overrightarrow{y}_s + \cos(\theta)\overrightarrow{z}_s)$, $\ell > 0$ with the sea bottom.

Let (ρ_k, θ_k, p_k) denote the sonar data acquired at time instant t_k: profile p_k has been acquired in direction θ_k, and for a point in the sea floor at distance ρ_k.Its sonar-centric coordinates are

$$\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = \begin{bmatrix} 0 \\ \rho_k \sin\theta_k \\ \rho_k \cos\theta_k \end{bmatrix}$$

In the body-fixed reference frame, the coordinates of the sea bed point are given by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R_s^T \begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} + O^s. \tag{1}$$

We assume that the sonar profiles p_k are statistically independent samples of *unknown* probability distributions μ_0 and μ_1 that depend on the region of the sea-bottom that returned the sonar profile. Let $s_s(k), k = 1, 2, \ldots$ denote the binary sequence that indicates which region has been hit by p_k. As we said before, we assume that this sequence visits often enough each of its two possible values, 0 or 1.

B. Camera

Let $\overrightarrow{x}_c, \overrightarrow{y}_c, \overrightarrow{z}_c$ be the coordinate system attached to the video camera, with origin $O^c = (O^c_x, O^c_y, O^c_z)$ and orientation $R_c \in SO(3)$ with respect to the robot's body-fixed reference frame.

As we said before, we assume that a segmented version of the video sequence is available (we addressed this problem in [3]). Let I_k be the raw image acquired at t_k, and denote by S_k its segmented version, such that $S_k(i,j) \in \{0,1\}$, for all image pixels (i,j).

We introduce now some notation that will be useful in the sequel. For each segmented image S_k, denote by $\{\mathcal{R}^k_i\}_{i=0,1}, k = 1, 2, \ldots$ its binary partition, such that $S_k = \mathcal{R}^k_0 \cup \mathcal{R}^k_1$. The sequence of segmented images defines an increasing partition of the image coordinate space (in pixels), that we denote by $\mathcal{P}^k$. Each member of the partition $\mathcal{P}_k$ collects the pixels that have the same classification *for all frames* in the sequence $S_1, \ldots, S_k$. Since we assume that only two regions are present in the observed scene, the size of the partition is, at most, doubled for each new image. For each region P^k_n in $\mathcal{P}^k$, we denote by $s^{(k)}_n$ the binary sequence that indicates the classification of the pixels belonging to P^k_n:

$$s^{(k)}_n = s_n(1), s_n(2), \ldots, s_n(k), s^{(k)}_n \in \{0,1\}^k.$$

Consider a point in the robot environment with body-fixed coordinates (X, Y, Z). Its coordinates in the camera frame are

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = R_c \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - O^c. \tag{2}$$

The projective image-formation transformation, denoting by f the camera focal distance, yields the following expression for the point coordinates in the image plane (see Figure 1):

$$\left[\begin{array}{c} u \\ v \end{array} \right] = \left[\begin{array}{c} \frac{x_c f}{z_c} \\ \frac{y_c f}{z_c} \end{array} \right].$$

Finally, considering the image coordinates of the optical center, u_0, v_0) and the horizontal and vertical scaling factors of the video camera, k_u and k_v, we obtain the image coordinates of the original point as

$$\left[\begin{array}{c} i \\ j \end{array} \right] = \left[\begin{array}{c} k_u \frac{x_c f}{z_c} + u_0 \\ k_v \frac{y_c f}{z_c} + v_0 \end{array} \right]. \quad (3)$$

Together, equations (2) and (3) define the map from the body-fixed reference frame to image coordinates.

C. Sonar-to-Image mapping

Combining equation (1) with the camera model, we obtain the complete model of the mapping to be learnt, by letting

$$\left[\begin{array}{c} x_c \\ y_c \\ z_c \end{array} \right] = R_c R_s^T \left[\begin{array}{c} x_s \\ y_s \\ z_s \end{array} \right] + R_c O^s - O^c = R \left[\begin{array}{c} x_s \\ y_s \\ z_s \end{array} \right] + t, \quad (4)$$

where the rigid motion parameters that relate the two coordinate systems are

$$R = R_c R_s^T \qquad t = R_c O^s - O_s. \quad (5)$$

Denote by α the parameters characterizing this rigid motion (3 rotation angles defining the rotation R and the 3 entries of vector t, in a total of 6 real parameters): $R = R(\alpha)$ and $t = t(\alpha)$. Combing equations (4) and (3) we can express the image coordinates of detected sea-bottom points at all instants, i.e., we have expressed $T_{S \to C}$ in terms of α. To emphasize the dependency of this application on the rigid motion parameters collected in α, we denote it by $T_{S \to C}(\alpha)$:

$$\begin{array}{rll} T_{S \to C}(\alpha): & \mathbf{R} \times [-\Theta_{max}, \Theta_{max}] & \to \mathbf{R}^2 \\ & (\rho, \theta) & \to (i^\alpha, j^\alpha) \end{array}$$

III. Estimation Criterion

As it was briefly mentioned in the Introduction, our estimation criterion is based on finding the image mapping of the sonar data that induces the best partition of the set of received sonar profiles. We present now the criterion that measures the quality of these partitions.

Consider a fixed value of the rigid motion parameters α, and that the intrinsic camera parameters are known. We can thus compute the image coordinates of all acquired sonar profiles:

$$(\rho_k, \theta_k, p_k) \Rightarrow (i_k^\alpha, j_k^\alpha, p_k), \quad k = 1, 2, \ldots.$$

We can now associate to each sonar profile p_k the label l_k^α of the pixel in the segmented image S_k that corresponds to it under the mapping $T_{S \to C}(\alpha)$:

$$(\rho_k, \theta_k, p_k) \to (i_k^\alpha, j_k^\alpha, p_k, l_k^\alpha = S_k(i_k^\alpha, j_k^\alpha)).$$

This association defines a binary partition of the complete set of received sonar profiles according to the labels l_k^α:

$$Q_n^\alpha = \{p_k \ : \ l_k^\alpha = n\}, n = 0, 1. \quad (6)$$

We remember now that we assume that sonar profiles corresponding to different regions follow distinct (and *unknown*) probability distributions μ_0 and μ_1. As presented in detail in [3], the optimal segmentation $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1$ of a set of data items generated by two distinct (but unknown) probability distribution is (in the sense of the generalized Maximum Likelihood) the one that minimizes the following *average entropy* criterion:

$$C(\mathcal{P}) = N_0 H(\hat{\mu}_0) + N_1 H(\hat{\mu}_1),$$

where $\hat{\mu}_n$, $n = 0, 1$ are the *empirical distributions* of data in each element of the partition $\mathcal{P}_n$, $n = 0, 1$, respectively, $N_n, n = 0, 1$ denote the cardinal of the sets $\mathcal{P}_n$, and $H(\mu)$ denotes the Shannon entropy of the probability distribution μ [1]:

$$H(\mu) = -\mathrm{E}_\mu [\ln \mu(x)].$$

where $E_\mu[\cdot]$ denotes statistical expectation under μ.

Our proposal is to use as the estimate of α the value that minimizes this criterion, for the partition $\{Q_0^\alpha, Q_1^\alpha\}$ defined in equation (6):

$$\hat{\alpha} = \arg\min_\alpha C(\alpha),$$

where

$$C(\alpha) = N_0(\alpha) H(\hat{\mu}_0(\alpha)) + N_1(\alpha) H(\hat{\mu}_1(\alpha)), \quad (7)$$

where $\hat{\mu}_n(\alpha)$ are the empirical data distributions for the sets Q_n^α, $n = 0, 1$.

We show below that $C(\alpha)$ is indeed minimum for the correct value of the rigid motion parameters (assuming that all other parameters are perfectly known). Our analysis considers a large sample limit, i.e., that the number of acquired sonar profiles and video images, k, is very large. Indeed, if k is large, the empirical distributions of the partition elements are, for general values of α, a mixture of the distributions associated to the two regions:

$$\hat{\mu}_n(\alpha) = \frac{N_{n0}(\alpha)}{N_n(\alpha)} \mu_0 + \frac{N_{n1}(\alpha)}{N_n(\alpha)} \mu_1, n = 0, 1, \quad (8)$$

where

$$N_n(\alpha) = N_{n0}(\alpha) + N_{n1}(\alpha), n = 0, 1, \quad (9)$$

and $N_{nm}(\alpha), m = 0, 1$, denotes the number of sonar profiles p_k incoming from region m and that receive a label $l_k^\alpha = n$ under the mapping $T_{S \to C}(\alpha)$.

Using the concavity of the Shannon entropy [1],

$$H(\lambda \mu + (1 - \lambda)\nu) \geq \lambda H(\mu) + (1 - \lambda) H(\nu)$$

and the asymptotic expression of the empirical distributions in terms of the regions' probability laws, equation (8), we can easily verify that

$$H(\hat{\mu}_n(\alpha)) \geq \frac{N_{n0}(\alpha)}{N_n(\alpha)} H(\mu_0) + \frac{N_{n1}(\alpha)}{N_n(\alpha)} H(\mu_1), n = 0, 1,$$

which leads to the following bound on the value of $C(\alpha)$:

$$C(\alpha) \geq \sum_{n=0}^{1} \left(N_{0n}(\alpha) + N_{1n}(\alpha)\right) H(\mu_n).$$

Recognizing $N^n = N_{0n}(\alpha) + N_{1n}(\alpha)$ as the *true* number of profiles coming from region n, which is independent of α, we finally obtain the bound

$$C(\alpha) \geq N^0 H(\mu_1) + N^1 H(\mu_1),$$

which is the value of C for the value of α that generates true image locations of all received profiles, and for which all the image-induced classifications are correct ($N_{01} = N_{10} = 0$).

To illustrate the discriminating properties of the criterion, we consider a fixed scanning direction θ_0 and a fixed altitude above (a locally planar) sea-bottom. In these conditions, the application $T_{S \to C}$ simply identifies the image coordinates of the fixed scanning direction θ_0: $(i(\theta_0), j(\theta_0))$, and we will directly denote the dependency on the pixel location by writing $C(i, j)$. Consider the evolution of the criterion $C(\cdot)$ as the number of images grows. As we argued above, the value of $C(\cdot)$ is dictated by the number of false classifications of the sonar profiles induced by the labels of the segmented images. Remember now the increasing partitions $\mathcal{P}_k$ of the image plane defined in Section II. It can easily be shown that the value of $C(i, j)$ is constant inside each element of the partition, since all pixels belonging to the same set have the same label sequence $s_n^{(k)}$, and thus induce the same partition of the sonar data. This shows that it is not necessary to update the value of $C(\cdot)$ for all possible pixel locations (i, j), but only for each partition in $\mathcal{P}_{\|}$, which will be in general much less than the total number of image pixels. This analysis shows that the minimization problem presented above for estimation of α may not have a unique solution. However, analysis of the geometry of the partition element for which $C(\cdot)$ is minimal directly indicates the residual uncertainty, and can be used to define "active observation" strategies with the goal of decreasing the size of the ambiguity region: the robot should try to drive the contour inside the ambiguity region, refining the partition of the image plane where more discrimination is required. This subject will be pursued in future work.

IV. Results

The following figure illustrates the simulation scenario used to test the estimation of the image coordinates corresponding to a single sonar direction, as explained at the end of the previous section. The robot is autonomously tracking a sea-bed contour, at constant altitude. We assumed perfect synchronization of the image and sonar acquisitions, and that the altitude above sea-bottom is held constant during the entire observation. Note that the linear contour assumed has the least informative geometry possible, and that co-registration information is obtained by the oscillating motion of the robot around the contour, during the simulated contour tracking. However, even if we explore the ability of the platform to move, observing the contour from distinct points of view, our estimation criterion does not require spatial reconstruction of the robot's trajectory, increasing its robustness with respect to positioning errors, and enabling a solution of the registration problem with low equipment requirements (for instance, we can co-register data for a platform with no global positioning information, as it is the case for many manually operated ROV's).

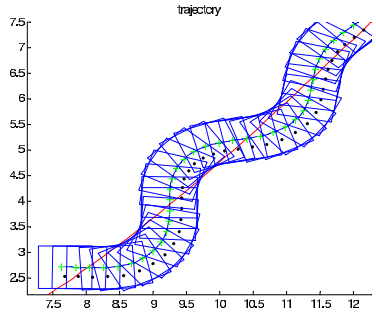

Fig. 2. Simulated calibration operation.

The next Figures show the evolution of the criterion $C(i, j)$ over the entire image plane, at iterations $k = 10, 25, 50$ and 75. As we see, there is a residual uncertainty along a diagonal direction, that could be reduced by making the robot observe the contour in the reverse direction, as we discussed in the previous section.

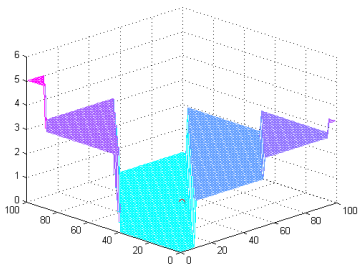

Fig. 3. $C(i, j), k = 10$.

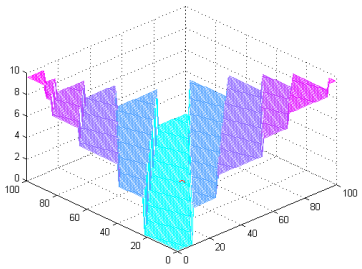

Fig. 4. $C(i, j), k = 25$.

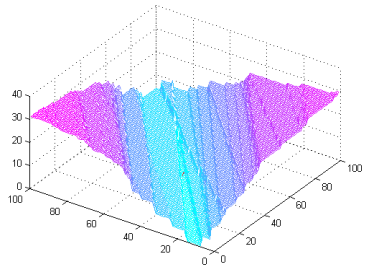

Fig. 5. $C(i,j), k = 50$.

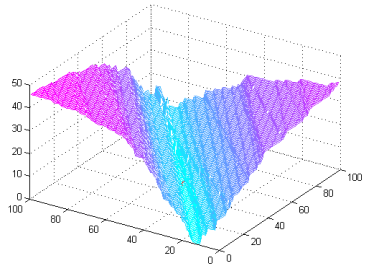

Fig. 6. $C(i,j), k = 75$.

Figure 7 shows the contour plot of $C(i,j)$ at iteration $k = 75$. The true image location of the scanning direction for this operational altitude is indicated by the red cross.

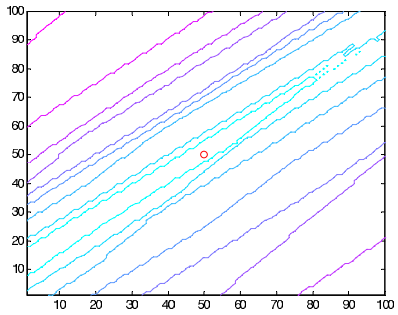

Fig. 7. Contour plot of $C(i,j), k = 75$.

V. Conclusion

We presented a statistically inspired criterion to estimate the rigid motion parameters that map sonar profiles, defined in polar sonar-centric coordinates, to image coordinates. The approach presented relies only on the ability of the vehicle to track a sea-bed contour and does not rely on reconstruction of the robot trajectory during data acquisition. Simulation results are presented that demonstrate the validity of the approach proposed.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.

[2] C. Barat and M. J. Rendas, *Using Statistical Mixture Models for Tracking Natural Underwater Boundaries*. Proc. UUST, 2003, New Hapshire, USA.

[3] A. Tenas, M. J. Rendas and J. Folcher, *Image Segmentation by Unsupervised Adaptive Clustering in the Distribution Space, for AUV Guidance Along Sea-bed Boundaries Using Vision*. Proc. Oceans 2001.

[4] S. Rolfes, M. J. Rendas, *Statistical Snakes: Robust Tracking of Benthic Contours Under Natural Variations*. Proc. IROS 2004, Sendai, Japan.