A Fast Object Orientation Estimation and Recognition Technique for Underwater Acoustic Imaging

Alessandra TESEI, Andrea TRUCCO, and Daniele ZAMBONINI

Department of Biophysical and Electronic Engineering (DIBE) - University of Genoa
Via All'Opera Pia 11/A, I - 16145 Genova (Italy)

Abstract. This paper describes a technique for automatic fast orientation estimation and recognition of man-made objects present in underwater acoustic images. A method able (a) to estimate the orientation of the object principal axis, and (b) to recognize the object itself by using information about boundary segments and their angular relations, is presented. This method is based on a simple voting approach applied directly to the edge discontinuities of an image. Often, in underwater acoustic images some undesired effects are present, such as halos around objects, speckle that creates discontinuities, and distortion. Voting approach is very robust with respect to these effects, so allowing good results also when images have a very poor quality. A sequence of real acoustic images is shown for testing the proposed technique validity.

1 Introduction

In modern underwater missions, the use of automatic or autonomous devices is in constant increase, in order to avoid that human operators are directly involved. The analysis of the acoustic images, acquired at the mission site, is a very heavy task for human eyes both for the long duration of the mission and for the poor quality of such images. So, algorithms able to recognize a wanted object on the basis of *a priori* known model and to estimate its orientation are very useful.

We address the problem of fast recognition and orientation estimation of man-made objects present in 2D images particularly acquired by acoustic camera for a wide number of different tasks. One can mention inspection, survey, manipulation and positioning in offshore structures or mineral mining sites, along pipelines or communications cables, relocation of lost objects, identification and removal of toxic wastes, mine counter measures, and so on. For man-made objects we consider objects having a regular and well-defined structure. The term "regular" means the possibility to describe the object by means of geometric surfaces bounded by straight lines.

In literature, some recognizer were suggested for other kind of acoustic images (i.e. side-scan images) [1] but their complexity is high and real-time functioning is very difficult. Other authors [2,3,4] proposed real-time operations on underwater acoustic images, but they limited their goal to object detection.

We present a method able to estimate the orientation of an object by searching for the direction of its principal axis, and to recognize the object itself by using information about boundary segments and their angular relations. This method is based on a simple voting approach [5] applied directly to the edge discontinuities of an image. Two different levels of increasing complexity are used to detect important

image features (e.g., straight lines) with different accuracy. The first level is very easy and allows to estimate the object orientation quickly. The second level is computationally more complex as allows to obtain both more accurate orientation estimation and fast object recognition [6]. The proposed method is able to discriminate between different sets of man-made objects, thus making the recognition process invariant to translation and scale, and the orientation estimation process also invariant to rotation.

2 The Voting Approach

In general, voting methods are very robust to noise and are not affected by shape partial occlusions [7]. In underwater images some undesired effects are often present, such as halos around objects (due to poor camera resolution), speckle creating discontinuities in the objects (due to coherent interference among waves back-scattered from the scene), and distortion (due, above all, to inhomogeneities of the medium). The proposed approach is very robust with respect to these effects.

The voting method maps each feature space point $e \in E$ into a set of parameter space points $\pi \in \Pi$ by means of a generating equation $G(e, \pi) = 0$, where E and Π represent the feature and the parameter spaces, respectively [8]. Let us consider the edge image, $E(\mathbf{x}) = E(x, y)$ as a 2D feature space of x and y coordinates (i.e., $\mathbf{e} = \mathbf{x} = (x, y)$), obtained by applying a Canny filter [9] to the original image $I(\mathbf{x}) = I(x, y)$. An image reference system (X, Y) with origin O in the bottom left corner of the image is chosen. Each edge point $\mathbf{x}$ in this space votes for all 2D parameter space points which satisfy the voting equation G . The voting algorithm consists of two processing levels: (a) object orientation estimation, and (b) object recognition.

2.1 Object orientation estimation

The voting process is performed directly from the edge-image space E into a 1D parameter space $\Pi = \{\alpha\}$, composed by all possible angle values α associated with object axes. As shown in Figure 1, α is the angle between the normal to the object axis and the X -axis of the image reference system.

Figure 1. 2D reference system for object orientation estimation.

Let Θ the upper limit of the parameter space (e.g., $\Theta = 180^\circ$) and $\Delta\theta$ the quantization step, thus the parameter space dimensions are given by $\Theta \cdot \Delta\theta$. In order to obtain an orientation accuracy equal to 1° it is necessary to choose $\Delta\theta = 1$. The relation G is 1D and considers rectilinearity and gradient similarity. Two edge

points, x_1 and x_2 , belonging the same straight line and with similar gradient orientation $\gamma(x)$ ($|\gamma(x_1) - \gamma(x_2)| \leq \gamma_{th}$) vote for the same orientation value α [10]:

$$G(x, \alpha) = \begin{cases} 1 & \forall \alpha \in \Delta\gamma(x) \\ 0 & otherwise \end{cases} \quad (1)$$

where $\Delta\gamma(x) = [\gamma(x) - \gamma_{th}, \gamma(x) + \gamma_{th}]$, $\gamma(x)$ is the gradient orientation at the point x and γ_{th} is a local threshold. For each edge pixel x in the image domain, which satisfies the relation $G(x, \alpha)$, an accumulator array $C(\alpha)$ is incremented. At the end of the voting phase, an histogram reports, for each α value, the number of votes of the correspondent cell $C(\alpha)$. The orientations α^* , common to several edge points into the image domain (e.g., object borders), generate multiple peaks in the parameter space. Each peak will be as tall as great the edge point number will be.

A more robust object orientation procedure can be obtained by considering the angles between the object axes. A set of angle relations corresponding to two or more model object axes can be easily computed by:

$$|\alpha_i - \alpha_j| \equiv \delta_{ij} \quad \forall i, j \in [1, J] \quad (2)$$

where α_i represents the orientation of the i -th model object axis and J is the number of model object axes. A set of mobile windows $W = \{W_j, j=1, \dots, J\}$ (one for each axis of the object model) of inter-distance $\delta_{ik} = \delta_{ik} = \|\omega_i - \omega_k\|$, where ω_i is the medium point of W_i , and size δ_{th} , are shifted on the histogram to detect candidate peaks, whose position satisfies the angle relation among the 2D object model axes. Candidate peaks related to the same object are grouped into different sets $P_k = \{\alpha_{i,k} : i=1, \dots, J, k=1, \dots, K\}$. In this way, the probability of detecting false maxima because of both false peaks and noise data is reduced. The object orientation will be computed as higher maximum $\alpha_{i,k}^*$ belonging to the group P_k^* whose candidate maxima received the highest number of votes. P_k^* is:

$$P_k^* = \left\{ P_k : \sum_{i=1}^J C(\alpha_{i,k}) > \sum_{j=1}^J C(\alpha_{j,h}) \quad k, h = 1, \dots, K \quad k \neq h \right\} \quad (3a)$$

Then, the global maximum α^* (i.e., object orientation) is given by:

$$\alpha^* = \{\alpha_i : C(\alpha_i) > C(\alpha_j) \quad \alpha_i \in P_k^*, \alpha_j \in P_k^*, i, j = 1, \dots, J, i \neq j\} \quad (3b)$$

2.2 Object recognition

The object recognition process is based on the shape information contained in the object boundaries. Our approach is restricted to regular (man-made) objects, whose boundaries are straight lines. To this end, a complex process able first to extract from the edge image a set $S = \{s_m : m=1, \dots, M\}$ of straight lines and to match them with object model segments contained in a data-base is applied. In order to reduce the process time-consume, a preliminary phase able to focus attention on the most probable object boundaries is performed: a 1D voting step like that used in section

2.1 is applied. A label $l(\alpha)$ is used to mark the edge points corresponding to the histogram peaks. The angles between the boundary segments of the 2D model object are used for selecting the histogram peaks α_p^* , $p=1,..P$, containing votes coming from segments. P is the number of boundary segments. Then, a 2D voting step, restricted to the edge points characterized by a label $l(\alpha_p^*)$, is performed.

Let (X', Y') a new reference system with origin O' in the center of the image. Each straight line can be parametrically described by polar coordinates (ρ, θ) [8] where ρ is the normal distance of the line from the origin O' and θ is the angle between the line and the x axis. The mapping process is performed from the edge image space E into a 2D a coupled space $\Lambda = \{[\pi, p_j(\pi)]: \pi \in \Pi, j \in [1..J(\pi)]\}$, where $\pi=(\rho, \theta)$ is a parameter-space point representing a straight line $s(\pi)$ on the image plane, and $p_j(\pi)=[x_j^{\min}(\pi), x_j^{\max}(\pi)]$ is a pair of edge points representing the endpoint coordinates of a straight segment of $s(\pi)$. $J(\pi)$ is the number of straight segments belonging to $s(\pi)$. The 2D parameter space is quantized into cells of size $\Delta\rho$ and $\Delta\theta$ chosen as explained in [11] for reducing the spreading vote effects. This is accomplished by defining a 1D reference system $S(\pi)=S(\rho, \theta)$ for each straight line $s(\pi)$, whose origin O_s is placed at the foot point of the normal.

The voting method proposed is an improvement over classical voting representation [7], as it allows edge points to be associated with different, adaptively chosen straight segments $s_j(\pi)$. Each collinear segment $s_j(\pi)$ is defined as a set of edge points x on $S(\pi)$ satisfying the generating equation and the two constrained endpoints. Each edge point $x=(x, y)$ votes for an image line according with a specific relation which favour collinear segments $s_j(\pi)$ having similar orientation. For each pixel in the image domain, satisfying such relation, the corresponding cell $C(\rho, \theta)$ in the 2D parameter space is incremented and the segment endpoints are updated.

At the end of the voting step, the parameter space Π is examined to select higher peaks (ρ^*, θ^*) and recover from these peaks the segments $s_j(\pi^*)$ representing the object boundaries on the image. A set of segments $S=\{s_m: m=1,..M\}$, extracted from the input image is created. Given a set G of model segment groups $GS_q=\{m_{r,q}: r=1,..R, q=1,..Q\}$ representing the 2D-projection of the searched object model from different viewpoints, the recognition step must verify whether a group of image segments can match with that set. The matching method proposed follows some basic ideas shown in [6], and uses geometric constraints to reduce the search.

The result of the matching process is object recognition. Recognition can be performed also in the case of incomplete data, due to occlusion or high noise level, if the pair of data-model segments is greater than a fixed threshold. Recognition results allow to give a more accurate orientation estimation. On the basis of the model, the orientation of some segments is used to determine the principal axis direction.

3 Results

Results aim to estimate the validity of the method by testing it on image sequences acquired by an acoustic camera. About orientation estimation, a mean error ε_m is

computed as absolute difference between the actual and the estimated object orientation, normalized to the number of images M contained in the sequence:

$$\epsilon_m = \frac{1}{M} \sum_{i=1}^M |\alpha_i - \alpha_i^a| \quad (4)$$

where α_i is the estimated and α_i^a the actual object orientation at the i -th image.

Figure 2. Sequence of actual images acquired by an acoustic camera.

Figure 3. Behaviour of the real and estimated object orientation for the sequence in Fig. 2.

An acoustic camera is able to form both a 3D map of the scene and some interesting 2D projections visualized as images. We show a sequence of orthoscopic images extracted from 3D maps and analogous to optical images. The object that should be recognized is a cylinder floating in the water and having principal axis parallel to the image plane, so, its orthoscopic projection is a rectangle. Figure 2 shows the sequence: the actual orientation of the cylinder is, respectively, $\alpha = 0^\circ$, $\alpha = 0^\circ$, $\alpha = 43^\circ$, $\alpha = 58^\circ$, $\alpha = 110^\circ$, and $\alpha = 145^\circ$.

The cylinder had a diameter equal to 60 mm, a length equal to 150 mm and it was placed about 1 m far from the camera. The carrier frequency used for the scene insonification was equal to 500 kHz and the high distortion of the images is due to low resolution achieved by the acoustic camera. Nevertheless, the algorithm was able to estimate the orientation with a satisfactory mean error value: $\epsilon_m = 4.1^\circ$ for the coarse voting function described in sub-section 2.1 and $\epsilon_m = 2.3^\circ$ for the accurate

voting function described in sub-section 2.2. The orientation estimation α'' is obtained after object recognition which has been performed successfully on the images in Fig. 2 by matching data with 9 different object models. Figure 3 shows the behaviour of the actual (angle α) and estimate (angle α' for coarse method, angle α'' for accurate method) object orientation. Computation times of about 0.2 sec and 0.8 sec (on a SUN Sparc 10) are required by the 1D and 2D voting functions, respectively. The matching step requires 0.2 sec to obtain a complete object recognition and 0.3 sec for the object reconstruction.

4 Conclusions

A voting approach devoted to automatic fast recognition and orientation estimation of man-made objects present in underwater acoustic images acquired mainly by acoustic cameras was proposed. Despite distortion and noise present in the acoustic images, the proposed technique (based on a simple voting approach applied directly to the edge of the image) shown encouraging results. A mean error measure was used to test the good accordance between actual and estimate object orientation.

References

1. S.G. Johnson, M.A. Deaett: The Application of Automated Recognition Techniques to Side-Scan Sonar Imagery. *IEEE Journal of Oceanic Engineering*, 19, 138-141 (1994)
2. L. Linnett, D. Carmichael: The Analysis of Sidescan Simages for Seabed Types and Objects. *Proc. 2nd Conference on Underwater Acoustics*, 733-738 (1994)
3. L. Henriksen: Real-time Underwater Object Detection Based on an Electrically scanned High-Resolution Sonar. *Proc. 1994 Symposium on Autonomous Underwater Vehicle Technology* (1994)
4. M. Ashraf, J. Lucas: Underwater Object Recognition Techniques Using Ultrasonics. *Proc. OCEANS 94*, I, 170-175 (1994)
5. G. Foresti, V. Murino, C.S. Regazzoni, G. Vernazza: Grouping of Straight Segment by the Labelled Hough Transform. *Computer Vision, Graphics and Image Processing: Image Understanding* 58, 22-42 (1994)
6. W.E.L. Grimson: *Object Recognition by Computer, The Role of Geometric Constraints*, The MIT Press (1990)
7. H. Illingworth and J. Kittler: A survey on the Hough transform. *Computer Vision, Graphics and Image Processing*, 44, 87-116 (1988)
8. R.O. Duda, P.E. Hart: Use of the HT to detect lines and curves in pictures. *Comm ACM* 15, 1, 11-15 (1972)
9. J.F. Canny: Finding edges and lines in images, Artificial Intelligence Laboratory. Technical Report TR-720, MIT (1983)
10. C.M. Brown: Inherent bias and noise in the Hough transform. *IEEE Trans. on PAMI* 5, 493-505
11. T.M. Van Veen and F.C.A. Groen: Discretization errors in the Hough transform. *Pattern Recognition*, 14, 137-145 (1981)