TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

PROCESSING AND INTERPRETATION OF SUBMARINE ACOUSTIC IMAGES

Francesco G.B. DE NATALE, Stefano FIORAVANTI,
and Daniele D. GIUSTO

Dept. of Biophysical and Electronic Engineering, University of Genoa
Via Opera Pia 11a, 16145 Genoa, ITALY

RÉSUMÉ

Cet article présente une système pour la détection et la reconnaissance de objets simple en images de trois dimensions reçu par un détecteur acoustique sous-marine.
Le méthode développé est basée sur la description de modèles des objets (forme essentielle, niveau de gris, courbature, ...) et sur la recherche des correspondances entre ces modèles et les informationes trouvé dans la scène. La choix du modèle est validée par l'estimation d'un paramètre de confiance 'fuzzy': si cette valeur est inferieur d'un seuil, le modèle est rejetée, and le suivant peut être tentée.
Le système a été testé sur des simple images réelles: le résultats obtenus sont prometteurs, néanmoin, de gros efforts de recherche et développement restent à effectuer puor attaindre la realisation d'un système de reconnaissance pour images naturelles complexes.

ABSTRACT

This paper presents a system for the processing and interpretation of underwater images acquired with an acoustic sensor. As a first objective, the system aims to recognize simple objects with known characteristics located at different positions so as to create a scene.
The proposed recognition system exhibits the features of a knowledge-based system, as it is based on ideal models (provided by an expert) of the possible objects present in the scene, and on the matching of such models with the real objects detected by processing data coming from the acquisition system.
The numerical data provided by different types of sensors are first processed by some low-level modules. The high-level modules operate a matching of the objects detected with the models, according to a criterion of maximum fuzzy reliability.
Results are promising and encourage further system developments.

1 Introduction

The goals of a system for the acquisition, processing and interpretation of underwater acoustic images are manifold. First of all, one should discriminate between navigation-oriented systems and identification systems. The former are used for autonomous driving of underwater vehicles in environments containing obstacles (i.e., to solve the problem of "obstacle avoidance"). The latter are utilized in a large number of applications, among which the most important are: recognition of prefixed targets; inspection of particular environments (e.g., bases of submarine platforms, immersed instruments, etc.); surveillance of high-risk areas (e.g., survey of the conditions of a dam); and various military applications.

Among the several instruments provided by current technology for image acquisition in underwater environments, the acoustic sensor is one of the most promising. In particular, it offers many advantages in the case of cloudy water (which strongly reduces optical visibility) and in remote-sensing applications. An additional interesting characteristic of this sensor lies in its ability to provide 3-D images acquired from a single viewpoint; this peculiarity is very useful whenever the exact distance to a target has to be known.

Due to this nature, recognition problems require, at least in part, the application of Artificial Intelligence: the more complex and the less known the environment where a recognition system must operate, the more mandatory the utilization of such methodology.

Specific technologies, like Expert Systems (to make an automated system "reason") or Knowledge Engineering (to represent and utilize efficiently all necessary information), play a fundamental role. In addition, the use of different sensors, or of sensors of the same type but located at different positions, and the fusion of data provided by such sensors enhance the interaction with the environment, and make it possible to utilize heterogeneous, though complementary, information. Obviously, different sensors are sensitive to different object properties, which could not be acquired by any single sensor.

Moreover, the information useful in solving a recognition problem includes not only the data coming directly from the sensors but also the results of intermediate processing, a-priori knowledge, previously acquired data, etc. All these data can be regarded as "virtual sensors" providing information that can be integrated with that obtained by the physical sensors.

The multisensory system presented in this paper consists (like all systems of this type) of many parts that can easily be connected in parallel, without need for solving complex synchronization problems, as many activities are entirely independent. Such a system is suitable for a multiprocessor implementation, which allows each sensor to be fully exploited, without involving considerable computational load.

The recognizer is made up of a network of expert systems, which require an appropriate knowledge-base related to the problem considered. The system performs the recognition process by following production rules managed by an inference engine. This component finds the rules whose condition parts have been verified, and selects (according to particular, generally heuristic, criteria) the rules whose action parts are to be activated. Another salient feature of the system is the use of different sensors: to accomplish its task, the recognizer performs the fusion of data acquired with various sensors, or the fusion of multiple "virtual" data obtained with a single sensor.

As far as implementation is concerned, low-level processing modules are realized in C language, while, for the system part that is most involved in the interpretation of a scene, LISP has been utilized, in particular, the Flavors library, which provides very powerful software tools for the management of complex data- structures.

2 System Structure

As mentioned earlier, the implemented system has the features of an expert system, and the recognition process is performed by matching data coming from the various sensors with the previously defined object models. The knowledge-base is embedded in such models, which are described and stored, during processing, in standard data-structures (called "frames") provided by the LISP interpreter. Such frames can be regarded as classes through which it is possible to access different data and procedures contained in slots. In particular, slots contain:

- attributes (e.g., feature values);
- links to other frames: they are used to connect a group of network frames to some topology (e.g., tree-structures are widely used, as their various branches provide the hierarchical descriptions of the subparts of an object model);
- computation procedures (i.e., function for the processing of data related to a class).

All knowledge necessary to the system for the recognition process is described through frames. In the following, it will be shown that frames can be utilized, also during processing, as software tools to keep track of progressive interpretation results, to manage and verify recognition hypotheses, to make the various system modules interact with one another, and so on.

2.1 Low-Level Processing

For the analysis of images acquired with an optical TV camera, geometric parameters (e.g., those derived from segmentation and edge detection) are of basic importance, as the shapes of the regions obtained are very significant for the recognition of an object and for the estimation of the object dimensions.

Instead, in the analysis of acoustic images, different parameters are considered, which provide the expert with more interesting and more meaningful information as compared with classical features.

Acoustic images exhibit the peculiarity of being three-dimensional. Two coordinates, Px and Py, are spatial and correspond to the dimensions of the detection sensor (less a multiplicative constant that takes into account the angle formed by the receiver with the emitter), and the third coordinate, $P\tau$, is temporal and takes into account the time required by an incident beam (sent by a source) to reach the receiver after being reflected by an object.

As an acoustic image is three-dimensional, the possible viewpoints are practically infinite; this corresponds to an infinite number of two-dimensional sensors. This consideration is based on the fact that, in general, it is preferable to operate in a 2-D domain, in part for the simpler implementation and the variety of available algorithms, in part for the faster execution of such algorithms. Therefore, one must face the problem of choosing some viewpoints only.

The primary task of low-level algorithms implemented is the detection of the most interesting areas in a scene, so as to restrict the range of areas to be processed by the following, more time-consuming, recognition steps. Then the algorithms extract the 2-D sections of the areas that are most likely to be associated with the objects present in a scene, and evaluate some features of such areas.

The features that have been regarded as the most significant for the recognition process are the following:

- centroid coordinates of the various regions contained in the section under examination;
- dimensions (in number of pixels);
- elongation (ratio between the dimensions x and y of the minimum bounding rectangle);
- radius of curvature in a given direction;
- average angle of orientation of a region with respect to an incident beam;
- peak grey level obtained through the convolution of response data with the signal wave (an adaptive filtering step is performed, and the maximum response amplitude is evaluated);
- signal phase (the received signal may be of the same sign as the sent one, or of opposite sign).

Centroid coordinates are used to verify the positions of the objects by following specific rules that make it possible to evaluate the fuzzy memberships of several regions in the same object. Dimensions, together with evaluation of the radius of curvature, allow one to establish if a spot is more likely to belong to an object with a curved surface than with a flat one; they also lead to a vague evaluation of the object size. The use of the adjective "vague" stresses the fact that the spots detected in an image represents only a very small part of the object, in particular, the one that reflects the acoustic beam in a sensor's reception field. Grey level is necessary for the detection of a spot and for the identification of the material the object is made of.

As can be noticed, geometric features are of particular importance, for they allow one to establish, at least in a probabilistic way, the type of surface of an object. Of major importance are the grey levels of the regions of the object surface; the various grey levels give the response value of the object, thus providing a useful parameter for identifying the material of the object. However, such a value strongly depends on the orientation of the object with respect to the signal source and on the depth at which the object is located. The greater the depth, the higher the signal dispersion and damping.

2.2 Module Structure

The system modules have been structured according to the block diagram shown in Figure 1.

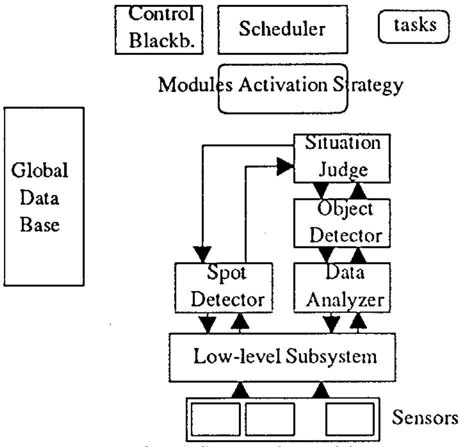

Fig. 1 System's architecture

The scheduler is a particular module devoted to managing the system architecture; it selects the modules and coordinates the activations of both the modules and the tasks associated with them, on the basis of the system status and of the strategy applied. It does not take part in the actual recognition process and is application-independent; it contains the knowledge about the system architecture. Using a set of rules, it selects, from among the activable modules, the one that is most useful in pursuing the system's goal, and then launches it. The goal may be modified in a dynamic way; in this case, the list of activable modules and the operating strategy are modified, too.

The first module that is scheduled in a top-down way is the Situation-Judge: it launches the module Spot-Detector, which, by suitably utilizing the predefined frames, performs a search for possible spots (objects) present in the scene, without need for identifying the nature of such objects. Such a module can access the property-tables created a priori, containing the numerical data extracted from an image of the scene.

The Spot-Detector searches for spots by simply scanning the regions present in the property-table related to a given image section. The main evaluation concerns the response value, coded in a particular gray level: such a value is the maximum gray level in a region, and is obtained through a convolution of the whole image with respect to the source signal. Finally, it is verified if the response value is included in a fixed range of data dependent on the nature of the predefined modules.

At this point, the Spot-Detector calls the Situation-Judge and sends it the response about the possible presence of spots in the scene. If no spots have been detected, the Situation-Judge informs the operator about the probable absence of obstacles in the scene, otherwise it sends the operator an alarm message and sends the next module (i.e. the Object-Detector) a 'go' message.

In order to recognize an object, the Object-Detector must use a first tentative hypothesis about a model; if this hypothesis is not verified, it generates another hypothesis, and so on, until all objects are recognized, or the interpretation process fails. The first model hypothesis required by the Object-Detector is provided by the Situation-Judge, which, through its internal functions, select the objects to be searched for in the scene on the basis of probabilistic evaluations: in this way, it tries to minimize the recognition steps.

The predefined models constitute the basis for the creation of a hypothesis-tree by the Object-Detector, which is assigned the task of verifying the presence of all model subparts. These may be more or less necessary for the correct recognition of an object; therefore, some fuzzy values are associated with the subparts to express the necessity for searching for each of them. The hypothesis-tree consists of frames called 'nodes', each of which maintains some links to 'father' nodes and 'son' nodes, and contains information about the recognition process at its level. The root node of the tree includes some particular slots, among which:

- a list of still expansible hypothesis;
- a list of nodes that cannot be expanded any more; such nodes are regarded as terminal leaves of the tree;
- the number of hypotheses so far created;
- a list containing some key-words to be sent to the Situation- Judge module to inform it about the final result of high-level processing.

Hypotheses are managed by utilizing a technique consisting in maintaining a tree, each node of which is characterized by a label, which may be OPEN (expansible node) or CLOSE ('dead' node). However, a node is expanded by instantiating not all possible sons, but only the best one. In case a failure should impose a backtracking step, new sons would be instantiated. The direct sons of the source node are the model hypotheses, which copy the information provided by the models defined a priori by the user, and instantiate some slots useful to all their successors.

In the tree, the successors of a model hypothesis are more and more detailed hypotheses about the object to be recognized, starting from the hypothesis containing only one recognized subpart up to the hypothesis in which, for each model subpart defined by the user, the system has found a satisfactory set of data.

If the Object-Detector verifies the presence of a sufficient number of subparts, and concludes, on the basis of available data, that the object has been recognized with good reliability, the result of this module is evaluated after modifying the viewpoint (utilizing other perspectives of the same 3-D image, or acquiring additional information from the sensors).

At the end of the above processing steps, the Object-Detector, on the basis of a weighted average over the results obtained from the various viewpoints, informs the Situation-Judge module about the type of object detected, and sends it the related data required (e.g., the depth at which the object is located, the object dimensions, the kind of material the object is made of, etc.). Subsequently, the control returns to the Situation-Judge.

While searching for the various subparts, the Object-Detector does not communicates directly with the data-base, but asks another module, i.e. the Data-Analyzer, for the information required: to this end it provides the Data-Analyzer with various information: the area in which to search for a subpart, the name of the subpart searched for, the list of necessary parts, the guide-sensor suggested, etc. The area is deduced from the relationships (defined in the knowledge-base) between the subparts to be searched for and the already found subparts.

In a subsequent phase, such relationships are computed more accurately in order to evaluate the global relational consistency of the subparts, and hence to verify the hypothesis previously made.

The Data-Analyzer's response may inform that:

- no subpart corresponding to the description provided by the Object-Detector has been found;
- the found subpart does not meet, in a satisfactory enough way, the high-level requirements;
- the relationships among the already identified model subparts are not verified by the relationships associated with such subparts in the hypothesis made;

in all those cases the node is closed.

If such situations do not occur, the node is kept open, and the hypothesis is rated. Rating is performed in a fuzzy way.

The Data-Analyzer is also responsible for the data-fusion of the data acquired by the various sensors.

3 System testing

The objects that are currently being considered for the definition of the models making up the knowledge-base are cylindrical or flat (e.g., tubes, flat plates, etc.). The models implemented in the system have been defined according to heuristic criteria derived from accurate observations of the various sections extracted. Such implementation has been supported by theoretical consistency with the physics of acoustic wave propagation.

In the following an example is proposed of a scene to which the recognition system has been applied. Measurements have been carried out by the Institute of Applied Physics (TNO) of Delft (Holland), which is equipped with specific instruments, and has been doing research in the area of acoustic imaging for many years.

The objects contained in the scene are three; they are located on the bottom at a depth of about 40 cm. The objects are two cylinders and a flat-plate, made of metal, perspex and PVC, respectively; their shapes are regular and their surfaces are smooth. The acoustic signal receiver is located on the y axis at a distance of 23 mm. from the emitter, and consists of a matrix of 64x64 elements space 1 mm. apart; the delay time of the receiver is equal to 472 ms.

The acquired acoustic image of this scene measures 128x128x1024 voxels, and each voxel has been digitized into 8 bits. Figure 2 shows a map of the maximum values along Px in the section Py-Pτ; Figure 3 (a) and (b) refer to the maps of the maximum values along Pτ (section Px- Py) for the ranges of such Pτ values as to contain separately the transitions of the flat-plate and the transitions of the two cylinders. To better clarify the meanings of the various spots making up the original image (Fig. 2), each spot in the image has been labeled by a number, and the caption to Figs. 2 and 3 associates with each number the corresponding spot in the scene. Figures 4 and 5 (a), (b) give the results of the segmentations performed on images 2 and 3 (a), (b) respectively; a classic segmentation algorithm has been applied. Finally, a table is presented, that gives the values related to the global recognition process.

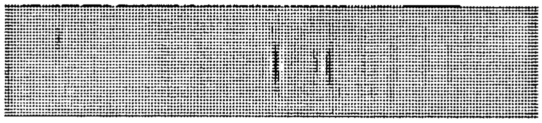

Fig. 2 Original image along Pt axis

Fig. 3 (a) Flat plate section (b) Cylinders section

Legenda: 1 Big Cyl. Upper Interface - 2 Small Cyl. Upper Interface - 3 Flat Plate Upper Interface - 4 Flat Plate Lower Int. - 5 Multiple Reflections

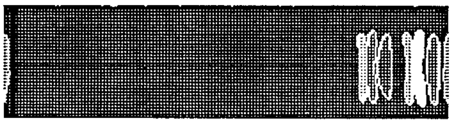

Fig. 4 Segmentation of fig. 2

Fig. 5 (a) Segment. of fig. 3 (a) (b) Segment. of fig. 3 (b)

Model	Found Subparts	Found	Hypothesis evaluation	Probable material
Big Cylinder	interf. 1-1	fuzzy .79	Cylinder found fuzzy .8	Brass with fuzzy .72
	interf. 1-2	fuzzy .75		
	interf. 1-3	fuzzy .73		
Flat Plate	interf. 1-1	fuzzy .87	Flat Plate found fuzzy .92	PVC with fuzzy .85
	interf. 1-2	fuzzy .82		
	interf. 1-3	fuzzy .75		
	interf. 2-1	fuzzy .8		
	interf. 2.2	fuzzy .9		
	interf. 2-3	not found		

Table showing results obtained by the recognition system

Acknowledgments

This work was supported by the EEC, within the framework of the program MAST: Project WAICS (Wideband Acoustic Imaging and Classification System).