MAP-DRIVEN IMAGE INTERPRETATION BY ASSOCIATIVE MODEL INDEXING

Gianluca Foresti, Vittorio Murino, Carlo S. Regazzoni, Rodolfo Zunino

Dept. of Biophysical and Electronic Engineering
University of Genoa
Via all'Opera Pia 11A, I-16145 Genoa, Italy

ABSTRACT

The problem of integrating territorial information within a multisensor vision system for autonomous-vehicle control is addressed. Environmental information is used to improve recognition results and to locate a vehicle's position in the coordinate reference frame of a map. To this end, a hypothesis-and-test search mechanism has been developed, which is based on an associative phase and a symbolic. In particular, an associative memory is first used to address the possible territorial area where the scene under examination may have been acquired. This guess is then verified by a symbolic recognition system using a model-driven strategy.

1.INTRODUCTION

The integration of multiple information sources is basic to obtain an accurate recognition of 3D outdoor scenes, especially when controlling an autonomous vehicle. In this paper, we address the problem of integrating territorial information into a multisensor vision system for autonomous driving. A set of synthetic images, representing significant viewframes reconstructed from an a-priori fixed route on a territorial map are first stored in an associative memory [9]. This process represents the training phase of the associative memory. Images acquired by a multisensor set-up are then processed by the associative memory in order to produce an estimation of the vehicle position inside the map reference frame. This strategy makes it possible to arrange the search space, in such a way as to avoid the search in the whole model space, thus obtaining a better computational performance. This initial guess gives a position estimation which is then verified by the recognition system by looking for objects associated with the viewframe. This process is performed at a high abstraction level, and consists in an expectation-driven search starting from symbolic object descriptions and using a version of a distributed blackboard system for recognition [4], where a module devoted to scene analysis has been inserted.

The paper is organized into in four sections. Section II deals with a general formulation of the problem, pointing out the characteristic of the sensors employed . Section III contains a brief review of associative memory techniques, and Section IV contains a description of the model here employed and it reports preliminary results obtained on a set of real images and on the related territorial map.

2.THE INTERPRETATION PROBLEM

This section deals with the general formulation of the recognition process. It is explained How data provided by terrain map are transformed so that they can be fused with data acquired with a TV-camera, . Then, the recognition process performed at the symbolic level is described.

2.1 Cartographic virtual sensor

A topologic map (TM) representing a scenario through which an autonomous vehicle can ride provides useful information to be used by a multisensor recognition system.

Two intermediate steps have to be performed to obtain a representation of the information contained in the map that it can be compared directly with data provided by visual sensors: first, a 3D model of the environment must be obtained; then an observation model must be provided allowing the system to simulate the acquisition of data as similar as possible to those coming from the visual sensor .

2.2 3D Model

Starting from a digitized image, like the one in Fig. 2, it is possible to detect two kinds of basic primitives which can be used to describe the environment model: lines located at equal height, (i.e. the so called contour lines) and landmarks (i.e., significant patterns which can be recognized by the system (see Fig. 3)). Using processing techniques and reconstruction algorithms (whose descriptions go beyond the scope of this paper), it is possible to obtain from contour lines a 3D map in the form of a matrix, $F(x,y)=Z$, where Z is the height computed at point (x,y) of TM. The next step is to place landmarks on the 3D ground map. Landmarks are usually characterized by regular shapes (e.g., as a first approximation level, houses can be represented as parallelepipeds). A generic landmark Li is associated with the matrix $L_i(x,y) = Z'$. Then, a complete a-priori environment model, F^*, can be obtained through an appropriate transformation of F. This operation is called landmark positioning, and can be modeled as a transformation over the original ground map, considering its landmarks:

$$F^*(x,y)=T\,(F(x,y),\, L_i(x,y)) \qquad i=1...K$$

In this way, one can obtain a representation of the information contained in the map by indicating the contour lines and some characteristic objects which can be observed during the mission.

2.3 The observation model

The environment model, F^*, can be observed with a camera emulator C, which takes as input the coordinates of the viewpoint (x_0,y_0), the axis of the visual direction (Z_V) and a vector of the camera parameters P, (e.g., depth of field, focus, etc.). The camera emulator performs a perspective transformation which allows one to obtain a 2D view (F2D).

$$C\,((x_0,y_0),\, Z_V,\, P) \text{ ------> } F2D(x',y')$$

F2D is called a viewframe of the environment F^*, and it represents a synthetic visual image with coordinates (x', y'), to be compared with the images provided by a sensor during the mission. In our system we have used routines of the HP STARBASE package to implement the camera emulator C.

2.4 Dynamic Environment

Given a sensor model C and an environment model F^*, we can define a mission M as a sequence of points in the reference system of the map $N_{Vi}=(x_{Vi}, y_{Vi})$

$$M = \{ N_{Vi} \}, \quad i=1....n$$

We can associate a set of a-priori viewframes with the mission M, provided that we take $Zv_i=(N_{Vi+1} - N_{Vi})$, i.e., at each point, the camera axis is directed to the next point to be reached, and provided that we mantain fixed the camera parameters vector P^* during the mission. Consequently we define

$$C(N_{Vi}, Z_{Vi}, P^*) = F2D^{Vi} \qquad \text{where} \quad N_{Vi} \quad \text{belongs to M}$$

$$V = \{ F2D^{V1}, ... ,F2D^{Vi}, ... ,F2D^{Vn} \}$$

where V denotes the set of a-priori viewframes representing the a-priori knowledge about the observations that a sensor can make during the mission M. The problem of identifying the position of the vehicle in the environment can now be formulated as the problem of associating with an image Sj at time t, (i.e., $S_j(t)$), a viewframe $F2D^{Vi}$ belonging to V, where j indicates the j th sensor of the autonomous vehicle. Small shifts from the position N_{Vi} should be tolerated by the system.

We solve this problem by using signal processing and data fusion techniques. We can consider TM and S_j as two information sources whose data must be fused in order to obtain a single description of the current scene considered. We have to select the viewframe $F2D^{Vi}$, which exhibits the greatest similarity to the sensors' data.

Two representation levels can be considered as possible candidates for data-fusion: the image level and the symbolic description level. In the following subsection, we shall discuss the symbolic level; the techniques employed for the image level will be described in more detail in the next section.

2.5 Symbolic description level

Each a-priori viewframe $F2D^{Vi}$ can be described in a symbolic way as a list of landmarks ($S2D^{Vi}$) that are visible inside $F2D^{Vi}$ and of relationshipsamong such landmarks:

$$S2D^{Vi}=\{L_1,...,L_k,...,R(L_k,L_j),...\} \quad k,j=1,....,K, \; i=1..n$$

Each object L_k is described in a propositional way by defining its intrinsic and relational attributes.

We suppose that a recognition system is available that is able to answer about the presence of a certain object inside a given scene, starting from $S_j(t^*)$ data. Then the problem of identifying which viewframe $F2D^{Vi}$ has generated $S_j(t^*)$ can be solved at the symbolic level by considering the list $S2D^{Vi}$ associated to each viewframe Vi, and by progressively discarding those viewframes which do not contain the searched landmark. This can be obtained by asking the recognition system yes/no queries about the presence of a landmark. This procedure is very heavy, especially in terms of response time of the recognition system. A statistical approachto the landmark choice is used. Each landmark is associated with an a-priori probability $P(L_i)$ computed on the basis of the number of occurrences of the object, in the set of stored viewframes. Therefore, the object with the highest probability is searched for in the scene. We have used the system DOORS [4] to implement this strategy; even though results are good from the point of view of recognition, it is necessary to speed up the system in order to improve response time related to search operations.

Therefore, an associative indexing technique is currently under development, which allows the system to rank, in an efficient way, the starting set of viewframes and, consequently, the set of landmarks contained in them. According to the landmarks ranking, propositional descriptions to be searched for are selected by the recognition system, in order to check, in a symbolic way, whether the chosen viewframe is supported by the actual image Sj(t*). At this point, if the system replies that a certain landmark associated to the selected $S2D^{Vi}$ is absent from the scene, a backtracking strategy is activated. This can be performed at the symbolic level by taking from the list the viewframes that contain that landmark, and by continuing the search process on the remaining viewframes. The application of the associative technique increases the conditional probabilities $P(L_i/A)$'s of those objects which are supported by associative matching. These probabilities are conditioned by the goodness of the associative matching scheme A. The average response time is expected to be reduced. In the following we describe how we can limit the search space by using an associative memory model.

3.THE ASSOCIATIVE DEVICE

In this section, we explain the employed associative memory model.

Several models have been proposed in the literature to simulate associative behaviour. The classic definition of Kohonen's correlation-matrix memories [6] implies a matrix-vector multiplication for storing and retrieving information. A few attemps have been made to apply such model to pattern analysis; the most recent work in this field is presented in [7], where emphasis is placed on scale and rotation- invariance features. The model adopted in the present paper is the associative noise-like coding memory described in [1]. The related mathematical framework is derived from the holographic model of associativity proposed by Gabor [5] and based on the complementary operations of convolution and correlation.

3.1 Information storage and retrieval

Any associative model involves three kind of components: a memory device, the pattern(s) to be stored, and the "keys" associated with each pattern and used for both information storage and retrieval. In detail, the data structure are the following:

* a "memory device" square matrix, say M[1..N,1..N], which holds the results of memory-writing cycles (convolutions);

* a square matrix P[1..N,1..N], which codes an image to be stored in M; the elements of P can assume values included in the gray-level range imposed by the vision system's characteristics;

* a square matrix R[1..N,1..N], used to hold the results of data retrieval from M;

* a Key-matrix, K[1..N,1..N], with the following characteristics (noise-like coding):

a. if we assume K_{ij} to be a stochastic variable associated with the matrix element K[i,j], then K_{ij} is independent of K_{ks} when i = k or j = s;

b. K[r,s]=0 or $E\{K_{ij}\}=0$ (mean value);

c. K is normalized, that is, $K[r,s]^2 = 1$;

d. different keys matrices do not correlate with one another.

In the following, the process by which a key is derived from a pattern will be denoted by **K-GEN**.

MEMORY WRITING is performed by means of the convolution operation:

$$M := K * P$$

$$M[i,j] = \Sigma_{rs} K[i - r, j - s]\, P[r,s]$$

MEMORY READING is performed by means of the correlation operation:

$$R := K \quad M$$

$$R[i,j] = \Sigma_{rs} K[i + r, j + s]\, M[r,s]$$

Results of single convolutions are summed up to build a complete memory; this implies that the order followed in information storage does not affect the system's performance. If the noise-like coding conditions imposed on keys are fulfilled, the matrix R_h (recall of pattern P_h) will be close to P_h. This can be expressed as

$$M_L := \Sigma_h m_h = (K_h * P_h)$$

for memory writing, and as

$$R_h = K_h \otimes M_L = K_h \otimes \Sigma_r (K_r * P_r) \approx K_h \otimes (K_h * P_h) \approx P_h$$

for information retrieval from the memory.

3.2 Use of the associative memory to perform image classification

The theoretical analysis is presented in [2], showing how pattern classification can be performed within the framework described in the previous subsection. The general problem is to identify which of a set of prototype patterns (images $F2D^{Vi}$ in our case) is closest to an unknown input pattern $S_i(t^*)$.

The principle of operation is that error variance is a discriminating parameter that allows one both to detect a prototype candidate for classification and to have a reliability measure of such a conclusion.

When an image to be classified is supplied, first a key is derived from it by using K-GEN; then the key is employed to perform a memory recall cycle. Finally, the system computes, for each image in the set of prototypes, the 'error matrix', defined as:

$$D_h := R - P_h$$

Each element $d_{ij}(h)$ of D_h can be viewed as a stochastic varible, and the minimum variance of $d_{ij}(h)$ will be associated with the (h-th) prototype that is closest to the input image supplied. The effectiveness of this discrimination principle is analyzed in [3]. Fig. 1 shoes the overall functional schema of the system.

4. THE CLASSIFICATION TASK

The above associative classification system operates within a vision system [4,8] for the evaluation of an autonomous land vehicle's position .

In our case, the set of viewframes $F2D^{Vi}$ are used as prototypes to be stored in the associative memory, while $S_i(t^*)$ provides the current input image to be classified,. In other words, the goal of the associative mapping system is to assess which of the synthetic prototypes is closest to the actual image supplied by the camera. This is especially useful when the positioning system 'starts blindly' and is supposed not to have any a-priori information about the vehicle's position, i.e., when there is no valid expectation about the camera signal. In this case, a rough position evaluation, corresponding to prototype addressing, may aid in using the expectation-driven recognition module correctly.

Synthetic images like the ones shown in Fig. 4 cannot be stored directly in the associative memory because the camera emulator is unable to assign correct gray levels to such images, and because the associative system could not conform to the varying degrees of brightness of TV images. Therefore, edges represent the only certain a-priori information tto be used, and an edge-extraction filter must be used before accessing the associative memory. Finally, for each image, a blurring process is performed on its edges so that information can be distributed in its whole matrix and matching probabilty is enhanced. The same procedure is used for the image provided by the TV camera to allow a coherent comparison .

The associative approach proves very efficient, and is faster than usual image-analysis techniques. Moreover, the outcome of the classification system (i.e. the set of error variances associated with the prototypes) allows one to arrange the alternative prototypes according to their reliability values, thereby facilitating the search process in case of backtraking.

Table A gives the measured error variance for each test image. These results are obtained by storing in the memory a set of four synthetic images, very similar to one another and then providing, as input to the system, a correspondent set of real images. Despite the high similarity among the images of the training set , a correct classification has been obtained.

5. CONCLUSIONS

A method exploiting cartographic information by fusing it with TV-camera data has been presented; the method can be used by a vision system for autonomous vehicle driving.

Data contained in a ground map are first transformed by detecting the contour lines; then, characteristic patterns (landmarks) that can be observed along a vehicle's route, are considered. This information is further processed, by means of a camera emulator, obtaining several views (viewframes) of the map path. A viewframe is a synthetic image with which symbolic descriptions of the objects it contains and their relations are associated .

The aim of the vision system is to determine the vehicle position (in the ground-map reference system) by associating the actual image acquired with a TV camera with one of the viewframes. When performed in a symbolic way this operation is computationally too heavy, whereas the application of an associative methodology can yield better results in a shorter time. By using this technique, we obtain a list of the viewframes, arranged according to their degres of similarity with the test image. Then, the symbolic recognition system checks whether the candidate viewframe is adeguately supported by the descriptive primitives extracted from the actual images. The application of this method allows one to obtain an improvement in the symbolic recognition process thanks to a fast indexing of the set of prototypes. Future developments will include the extension of the combined use of associative and symbolic recognition techniques to sequences of images, and the implementation of the associative system on a parallel architecture to achieve real-time performances.

ACKNOWLEDGMENTS

This work was carried out within the framework of the italian PRO-ART section of PROMETHEUS, an EUREKA project.

REFERENCES

[1] Bottini S. (1980), An algebraic model of an associative noise-like coding memory, Biol. Cybern. 36, pp.221-228

[2] Parodi GC., Zunino R. (1990), Image classification by using associative noise-like memories, Proc. IEEE Int. Phoenix Conf. on Comp. and Comm. IPCCC '90 Phoenix AZ March 1990

[3] Parodi G.C., Zunino R. (1990), Noise-insensitivity of an associative image classification system, IEEE Int. Workshop on Robust Comp. Vision, Seattle WA Oct. 1990

[4] Merialdo P., Pecollo P.C., Regazzoni C.S., Vernazza G., Zunino R., Integration of territorial maps in the vision system of an autonomous land vehicle, Proc. Int. Conf.Intelligent Autonomous Systems, Amsterdam, Dec. 1989, pp. 694-704

[5] Gabor D. (1969), Associative holographic memory, IBM J. Res. Dev. 13, pp. 156-159

[6] Kohonen T. (1972), Correlation matrix memories, IEEE Trans. Comput., pp. 353-359

[7] Wechsler H., Zimmermann J.L. (1988), 2-D invariant object recognition using distributed associative memory, IEEE Trans. on PAMI, Vol. 10-6, pp. 811-821

[8] Giusto D.D., Regazzoni C.S., Vernazza G., (1989), Multilevel data fusion for detection of moving objects, Proc. IEEE Int. Conf on Syst. Man and Cybern., Boston MA, Dec. 1989

[9] Hinton G.E., Anderson J.A., Parallel Models of Associative Memory (updated edition), Lawrence Erlbaum, 1989

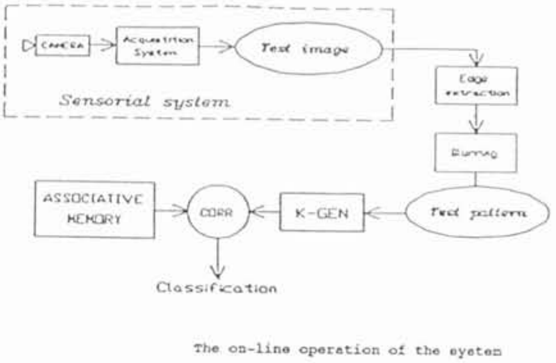

Fig. 1 Functional schema of the associative system

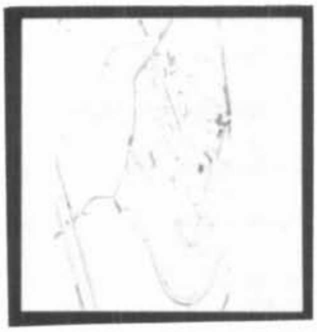

Fig. 2 Ground map of the mission environment

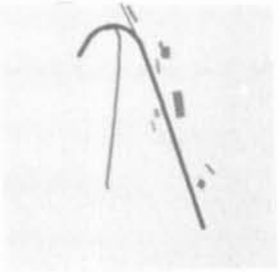

Fig. 3 Landmarks interactively extracted from the ground map

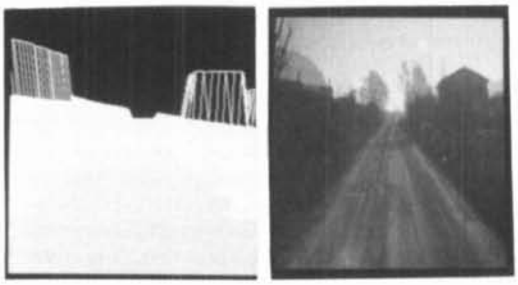

Fig. 4 Reconstructed 3D scene and corresponding b\w camera

▨	Measured error variances			
	I1	I2	I3	I4
I1	0.2613	0.5886	0.4691	0.3314
I2	0.6138	0.3077	0.4075	0.5484
I3	0.4377	0.5006	0.2015	0.3168
I4	0.3709	0.7078	0.4786	0.3615

Table A Measured error variance