FROM EARLY PROCESSING TO CONCEPTUAL REASONING: AN ATTEMPT TO FILL THE GAP

Giovanni Adorni, Lina Massone, Giulio Sandini, Mauhzio Immovilli
Department of Communication, Computer and System Sciences - University of Genoa
Via airOpera Pia 11a - 16145 Genoa - Italy
Tel. +39-10-310223 - Netmail: bambi©ugdist.UUCP

ABSTRACT

This paper shows how multiple low level sources of information can be integrated in a Iconic Data-Base, and how this data-base can be suitably organized for symbolic manipulation. The heuristics used to segment images of a simplified real world are described. Experimental results are presented.

I INTRODUCTION

A main task in artificial vision is the ability to analyze images obtained from real scenes to understand their content. This may be done by assigning appropriate interpretations to objects within the scene. Accordingly, an image must firstly be segmented into regions that roughly correspond to objects, surfaces, or parts of objects in the scene represented by that image. Two factors combine to make the task of segmentation difficult: i) noisy data, ii) object occlusions. To solve the problem, at least three requirements are necessary, according to the computational model proposed by Marr [1]: (a) full exploitation of physical and geometrical constraints of the environment, (b) integration of multiple low level sources of information, (c) use of global knowledge (i.e. knowledge about basic physical properties of the objects).

Some of these concepts have been implemented in several systems (see, for example [2,3,4,5,6,7,8,9,10]). In particular, Nazif and Levine [11] propose an architecture for the integration, in a rule-based environment, of evidences derived from several low level processes (edge detection, region analysis ...); the aggregation of elements - edges into lines, lines into regions — is performed on the basis of Gestalt principles [12]. A similar approach is followed also by Flinchbaugh and Chandrasekaran [13].

This paper, stemming from some of the ideas and intuitions of the mentioned systems, deals mainly with iconic segmentation of images, with the purpose of providing a description of images entities, like, for example, closed contours; some principles derived from the psychology of perception are used, many of which have been developed on an experimental ground. In particular it is shown how data derived from early processing stages can be integrated in a Iconic Data-Base, and how this database can be suitably organized for symbolic manipulation. The heuristics used to segment images of a simplified real world (a low scale model of an office-like environment with planar surfaces) are described, together with their actions on a Symbolic Data-base. Experimental results are presented.

II BUILDING A ICONIC DATA-BASE

The edge detection procedure is based on the so-called Marr-Hildreth approach, i.e. on the detection of the zero crossings of the second derivative of a Gaussian filtered image [1]. The filtering operator is a rotationally symmetric Laplacian of Gaussian producing contours that are closed or closed at the image boundary. During the extraction of the zero crossings, local orientation and slope are also computed to provide additional information about the local discontinuity. Unfortunately these measures are not sufficient to evaluate the perceptual relevance of a given edge because edge detection is a local operation while the perceptual evaluation of a scene is driven also by global features and by some kind of knowledge about the scene. For this reason, removing the irrelevant edges at this early stage of the visual process may be a hopeless procedure unless suitable hypotheses are made about the global properties of the scene. For the class of scenes used in our experiments (a low scale model of an office-like environment with planar surfaces, see Figure 1) the information about local orientation is used to extract linear segments as primitive elements to code boundary information (see below).

Along with the information about contours, also regional segmentation is performed. In spite of the fact that boundary and regional representations are two different descriptions of the same entities, designing a unique framework to include both boundary and regional information is no simple matter. The problem has been faced stemming from the observation that the image convolved with the Laplacian of Gaussian operator is naturally segmented into regions of either positive or negative sign, and that the borders of these regions are the zero—crossings contours. This regional segmentation has the advantage of being exactly dual to the boundary representation, making easy the process of switching from one description to the other. Moreover, due to the fact that all the contours are closed, each contour is the boundary between only two regions (one of negative sign, the other of positive sign).

At the present stage of implementation for each image a stereo pair is acquired in order to extract rough depth information [14] and, from a single view three images are acquired using red, green and blue filters.

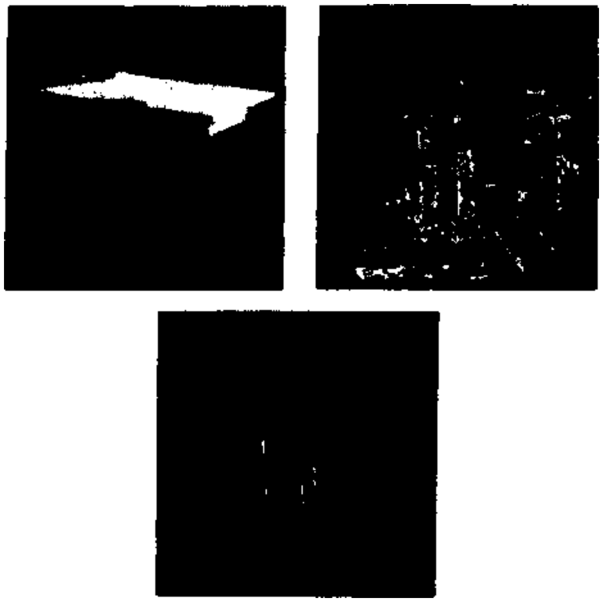

Figure I - Original image acquired by a TV camera, zero crossings extraction, and segments reconstruction.

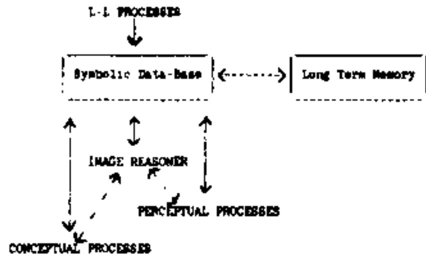

Figure 2 — Architecture of the system

The RGB color coding is then transformed into a HSV (hue, intensity, saturation) color base. The zero crossings are extracted from the "intensity" image only but the information about local slope is computed also for hue and saturation. Each contour is described as a chain of segments (see Figure 1), which seems to be a reasonable hypothesis for the class of objects under consideration; each segment codes, besides its position in the image domain, the local properties measured from the edge detection algorithm (averaged along the segment) and a pointer to the two regional descriptors. The inclusion of exploration strategies to provide 3D information by means of active movements of the observer within the environment is under consideration; occluding contours [15,16] and motion [17,18] are used as volumetric sources of information.

III <u>BUILDING</u> A <u>SYMBOLIC DATA-BASE</u>

This section shows how numeric data derived from early processing can be organized in a data-base suitable for symbolic manipulation (Symbolic Data-Base, SDB).

SDB has been hierachically organized into three linked levels:
- *segment* level (SEGM)
- *region* level (REGION)
- *simple object* level (OBJ)

It is worth pointing out that SDB is not just an interface to allow symbolic manipulation but it is a data structure shared by different processes which interact through it; these processes are (see Figure 2):
(1) *Image Reasoner,* described in the next section;
(2) *Perceptual Processes,* devoted to scene segmentation;
(3) *Conceptual Processes,* devoted to scene interpretation.

IV REASONING ABOUT <u>SYMBOLIC DATA</u>

To locate significant entities in images it is necessary to organize segments already coded in SDB into closed contours which do not, by the way, necessarily correspond to contours in the scene domain. This task is performed by the Image Reasoner (IR).

At this purpose, information about segments is considered selectively on the basis of their length (that is, longer segments are analyzed first). All the information is hierarchically explored; information at a higher detail is used to confirm or reject the inferred structures.

IR is organized as a production system; general heuristic knowledge about image properties is embodied into rules of the type *precondition* — > *action.* Actions can modify segments and regions as follows:
- create a new segment on the basis of features of

Figure 3

existing neighbour segments;

- delete segments which are discovered to be generated by noise or unsignificant details;
- merge more segments into one segment;
- create new regions and compute their features;
- delete existing regions;
- modify features of existing regions;
- merge two regions into one region;
- subdivide one existing region into subregions.

The output of IR is a line—drawing representation of the original image on which perceptual processes (PP) can perform occlusion analysis.

It is worth noting that IR and PP are interacting processes in the sense that, in case PP find some inconsistencies, IR is recalled on a better knowledge basis. At the actual stage of the implementation of the system IR is composed of about 320 rules logically divided into Segment Rules, which operate on segments and hypothesize new segments and segment clustering, and Region Rules, which hypothesize directly the presence of a region.

Let's consider the situation shown in Figure 3. By applying the rule:

```
R REPLACE_SEGM_12.

IF   (two segments are on the SAME_ALIGNMENT) AND
     (the DISTANCE between LINE_END and LINE_IN_FRONT
                                of the segments is LOW) AND
     (the INFO are SIMILAR)
THEN (there is an EVIDENCE[value] that the two segments
                           can be REPLACED by a NEW_SEGMENT)
```

the two segments can be merged into a new one. Predicates in the rule preconditions return a degree of evidence to take into account errors which may occur at the early processing level; the evidence associated to the rule action is the wheighted average. A threshold criterion is used to decide when backtracking is needed. This rule can be regarded as an application of the Gestalt principle of "good continuation\ It can handle those situations where region contours present partially occluded lines which are perceived by a human observer as converging into a common vertex. The term occlusion refers in this context, not only to lines hidden by physical objects, but also to lack of information due to early processing errors.

As an example of Region Rules let's consider the concept of subjective contours [22,23], i.e. those illusory contours which are perceived in spite of the absence, in the visual field, of discontinuities in brightness, or surface orientation, or color and so on. In [26] a computational model for treating subjective contours has been presented, based on the concept of *induced in homogeaeity* [24,25]. Some rules implementing that model have been introduced in this system to handle situations like the one shown in figure 4a. An example of such rules is the following:

```
R_SUBJ_CONT 279:

IT    (there is a SUBJECTIVE CONTOUR) AND
      (INDUCED_INHOMOGENITY Is HIGH)
THEN  (there Is an EVIDENCE!value) that
        (COMPUTE the TYPE of the subjective contour) AND
        (CREATE_REGION according to the type) )
```

Figure 4b shows possible reconstructions of the subjective contour of figure 4a.

By using the information about the X,Y,Z coordinates of each segment vertex (see appendix 1), IR is also able to infer qualitative spatial orientation of regions and spatial relationships among them. These qualitative attributes are coded as follows:

```
ON(REQION_I,REGION_D.
HORIZONTAL(REOION_I.REaiON_J,...).
VERTICAL(REOION_H.RBOION_K....).
```

Such attributes are used by perceptual and cognitive processes.

As far as the control strategy is concerned, it has been already pointed out that the analysis starts from longer segments. Actually, four length thresholds are considered, which have been experimentally fixed. Processing an image with a lower threshold means adding new segments; their consistency with manipulations operated at a higher threshold is checked and the additional constraints introduced by the presence of new segments are, possibly, used for backtracking.

V CONCLUSIONS

In this paper a Iconic Data-Base and a Symbolic Data-Base have been described as shared dynamic structures for low level image segmentation and high level perceptual reasoning. The main characteristics of the approach are: i) the use of multiple sources of information (including edges, color, stereo and multiple views), ii) the combined extraction of regional and boundary information, iii) the use of symbolic reasoning to extract perceptually important entities, iv) the possibility of reasoning over segments or groups of segments and not over single edges, which constrains the combinatorial explosion of rules applicability, v) the use of perceptual models with a strong psychological experimental background;

Low level processing is implemented in C on a MICROVAX II connected to a VDS 7001-Eidobrain image processing system. Symbolic Data-Dase and Image Reasoner are implemented in Franz Lisp and OPS5 on a VAX-Station. Data transfer is performed through ETHERNET network.

Future developments of this work are concerned with

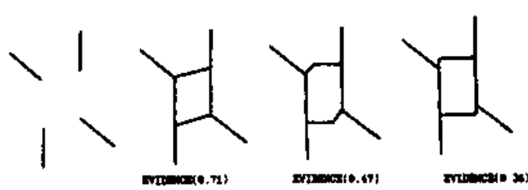

Figure 4. (a) example of subjective contour; (b) possible reconstructions using the concept of induced in homogeneity.

the introduction of volumetric information derived from occluding contours and motion; an interface with a geometric modeling environment based on octrees is also envisaged.

APPENDIX 1

Figure 5 shows the general structure of segment, region and object descriptor in SDB.

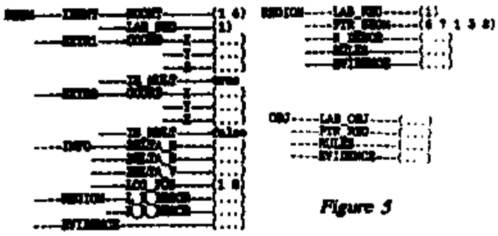

Figure 5

The meaning of the various subfield is explained as follows. Every SEOM structure contains a symbolic label (LAB_SEOM) and an order numbering (LOG_POS) along a possible segment chain (NCONT). Every extremal point of a segment (EXTRi) can be a fork point (in this case IS_MULT=TRUE) or not. The numerical information (INFO) is related to HSV color base (DELTA_H. DELTA_S. DELTA_V). RfiOIONs information is related to pointers to the two regional descriptors. The EVIDENCE field contains information about the degree of evidence of the presence of the hypothesized segment. The initial hypothesis is that data derived from L - L stages arc TRUE; in this way the EVIDENCE field of a SEOM derived from iconic data is intially set to 1.0. Every REGION structure contains a symbolic label (LAB_REG), pointers to the segments composing the hypothesized region (PTR_SEGM), a regional descriptor (RDfiSCR) which is actually only the average light intensity, computed from contours segments. A region is found by applying some rules which are stored in the RULES field and a value is assigned measuring the EVIDENCE of the hypotheses which have triggered the clustering operations.

Every OBJ structure contains a symbolic label (LAB_OBJ), pointers to segments composing the hypothesized regions (PTR_REG), the RULES used to assume the presence of a simple object and a i ure of the EVIDENCE of the simple object.

Acknowledgements

The authors wish to thank Manuel Leeuwenberg, Walter Gerbino, Marco Sambin and Peter Van der Helm, for their helpful comments and suggestions. The present work was partially supported by an ESPRIT grant (P419) and from the Italian Ministry of Education under a special grant on vision.

REFERENCES

1. D. Marr, *Vision.* Freeman, San Francisco (1982).

2. J.M. Tenenbaum and H.G. Barrow, "Experiments in Interpretation-Guided Segmentation," *Artificial Intelligence* S. No.2pp. 241-274(1977).

3. A.K. Mackworth and W. Havens. "Representing Knowledge of the Visual World," *IEEE Transaction on Computer,* (October 1983).

4. J. Gucksman, "Using Multilple Information Sources in a Computational Vision System," *Proc. 8th. UCAI,* (1983).

5. S.F. Venabte. DJ. Richter, and M. Wiedemann. "A Rule-Baaed System for Improving on Image Segmentation." Proc. *2nd. IEEE Goaf, on A.I. Applications, pp.* 94-99 (December 19t5).

6. W.A. Perkins, T.J. Lafley, and T.A. Nguyen, "Rule-Based Interpreting of Aerial Photographs Using LBS," *Proc. SPW - Applications of A.I.* 1754a pp. 138-146 (1985).

7. A.R. Hanson and E.M. Riseman, "VISIONS: A Computer System for Interpreting Scenes," in *Computer Vision Systems,* ed. A.R. Hanson and B.M. Riseman,Academic Press. New York (1978).

8. R. Brooks, "Symbolic Reasoning among 3-D Models and 2-D Images." AIM-343, Stanford Univ.. Stanford. CA (1981).

9. M.D. Levine and S.I. Sbaheen. "A Modular Computer Vision System for Picture Segmentation and Interpretation," *IEEE Transnetion on Pattern Analysis and Machine Intelligence* PAMI-3, No.5 pp. 540-556 (1981).

10. D.M. McKeown, W.A. Harvey, and J. McDermott, "Rule-based Interpretation of Aerial Imagery," Proc. IEEE *Workshop on Principles of Knowledge Based Systems,* pp. 145-157 (December 1984).

11. A.M. Nazif and M.D. Levine, "Low Level Image Segmentation: An Expert System," *IEEE Transaction on Pattern Analysis and Machine Intelligence* PAMI-6, No.5 pp. 553-577 (1984).

12. K. Koffka, *Principles of Oestaltpsychology,* Harcourt, Brace & World. New York (1935).

13. B.E. Flinchbaugh and B. Chandrasekaran. "A Theory of Spatio-Temporal Aggregation for Vision," *Artificial Intelligence* 17(1981).

14. P. Morasso and G. Sandini, "3D Reconstruction from Multiple Stereo Views," *Proc. III Int. Conf. on "Image Analysis and Processing",* (October 1985).

15. L. Massone, P. Morasso, and R. Zaccaria, "Shape from Occluding Countours," *S.P.I.E Symposium "Intelligent Robots and Computer Vision",* (november 1984).

16 Y. F. Wang, M. J. Magee, and J. K. Aggarwal. "Matching Three-Dimensional Objects Using Silhouettes," *IEEE Trans. PAMI* PAMI-6 No.4 pp. 513-518 (1984).

17. N. J. Bridwell and T. S. Huang. "A Discrete Spatial Representation for Lateral Motion Stereo." *CVGIP* 21 pp. 33-57 (1983).

18. D. T. Lawton, "Processing Translational Motion Sequences," CVG/P 22 pp. 116-144(1983).

19. H.K. Nishihara, "PRISM: a practical real-time imaging stereo matcher," A.I. Memo 780. MIT A.I. Laboratory. Boston, Mass. (May, 1984).

20. G. Adorni and E. Trucco, "A Quantitative Theory of Visual Perception: a Case Study." in *Advances in Image Processing and Pattern Recognition,* ed. V. Cappellini and R. Marconi .Elsevier Science Publishers B.V. (North-Holland) (1986).

21. G. Adorni, M. DiManzo, F. Giunchiglia. and L. Massone. "A Conceptual Approach to Artificial Vision," *Proc. 4th. Int. Conf. on Robot Vision and Sensory Controls,* (October 1984).

22. C.E. Osgood. *Method and Theory in Experimental Psychology, Oxford* University Press. New York (1953).

23. G. Kanisza, "Margini Quasi-Percettivi in Campi con Stimolazione Omogenea," *Rivista di Psicologia* 49 pp. 7-30 (1955).

24. M. Sambin. "Angular Margins without Gradient." *Italian Journal of Psychology* 3 pp. 355-361 (1974).

25. G. Adorni. L. Massone, and M. Sambin, "Subjective Contours: s Computational Approach," *Proc. 7th, Int. Congress of Cybernetics & Systems,* (September 1987).