[8] Kelly, M., "Visual Identification of People by Computer", Memo AI-130, Comp. Sci. Dept., Stanford University, Stanford, CA, July 1970.
[9] Fischler, M. A. and Elschlager, R. A., "The Representation and Matching of Pictorial Structures", *IEEE Trans. on Computers*, Vol. C-22, No. 1, 67-92, Jan. 1973.
[10] Garvey, T. D. and Tenenbaum, J. M., "On the Automatic Generation of Programs for Locating Objects in Office Scenes", *Proc. 2d Intl. Joint Conf. Pattern Recognition*, 162-168, Aug. 1974.
[11] Bajcsy, R. and Lieberman, L. I., "Computer Description of Real Outdoor Scenes", *Proc. 2d Intl. Joint Conf. Pattern Recognition*, 174-179, Aug. 1974.
[12] Yakimovsky, Y. and Feldman, J. A., "A Semantics-Based Decision Theory Region Analyzer", *Proc. 3d Intl. Conf. on Artificial Intelligence*, Stanford Univ., Stanford, CA, Aug. 1973.
[13] Ohlander, R., "Analysis of Natural Scenes", (Ph.D. Thesis), Comp. Sci. Dept., Carnegie-Mellon Univ., Pittsburgh, Pa, April, 1975.
[14] Winston, P. H., "The MIT Robot", *Machine Intelligence*, Vol. 7, 431-463, B. Meltzer and D. Michie, eds., Edinburgh University Press, Edinburgh and American Elsevier Publishing Company, 1972.
[15] Tenenbaum, J. M., "On Locating Objects by their Distinguishing Features in Multi-Sensory Images", Computer Graphics and Image Processing, Dec. 1973.
[16] Newell, A. and Reddy, D. R., "Image Understanding: Some Notes", Minutes ARPA Image Understanding Workshop, Science Applications Inc., March 1975.
[17] Winograd, T., "Five Lectures on Artificial Intelligence", AIM-246, Stanford Univ. Artificial Intelligence Laboratory, Stanford, CA, Sept. 1974.
[18] Sussman, G. J., "A Computational Model of Skill Acquisition", Tech. Note AI TR-297, Artificial Intelligence Laboratory, MIT, Cambridge, MA, Aug. 1973.
[19] Marr, D., "On the Purpose of Low-Level Vision", AI Memo 324, Artificial Intelligence Laboratory, MIT, Cambridge, MA, Dec. 1974.
[20] Norman, D. A. and Rumelhart, D. E., *Explorations in Cognition*, Freeman and Company, San Francisco, April 1975.
[21] Minsky, M., "A Framework for Representing Knowledge", in *The Psychology of Computer Vision*, P. H. Winston, ed., McGraw-Hill Book Company, New York, 1975.
[22] Tenenbaum, J. M. and Barrow, H. G., "MSYS: A System for Reasoning about Scenes", forthcoming SRI Artificial Intelligence Center Tech. Report, Stanford Research Institute, Menlo Park, CA.
[23] Tenenbaum, J. M. et al., "An Interactive Facility for Scene Analysis Research", SRI Artificial Intelligence Center Tech. Note 84, Stanford Research Institute, Menlo Park, CA, Sept. 1973.
[24] Freuder, E. C., "Active Knowledge", Vision Flash 53, Artificial Intelligence Laboratory, MIT, Cambridge, MA, Oct. 1973.
[25] Fahlman, S. E., "A Hypothesis-Frame System for Recognition Problems", Working Paper 57, Artificial Intelligence Laboratory, MIT, Cambridge, MA, Dec. 1973.
[26] Erman, L. and Lesser, V., "A Multi-Level Organization for Problem Solving Using Many, Diverse, Cooperating Sources of Knowledge", *Preprints 4th Intl. Joint Conf. Artificial Intelligence*.

Scene Segmentation

Jerry Potter
Box 27 Webster, N.Y. 14580

Introduction

The purpose of this communication is to present a slight restructuring of the segmentation and interpretation tasks. This restructuring enables the segmentation of a scene with a limited amount of general information. The interpretation aspect of a scene segmented by the method proposed here is very dynamic and flexible.

Scene Segmentation

Traditionally, scene segmentation is defined as dividing a scene into its basic components or features. The implicit assumption in this definition is that the segmentation (including the feature identification and location process) and interpretation processes are independent and sequential. This situation virtually forces a "bottom-up" approach to scene interpretation. Once all of the atomic features of a scene have been identified, the obvious next step is to apply transformations to these features producing a second level of features which in turn are similarly processed. The processing continues until the top level is reached and the interpretation complete.

In fact, whenever all of the atomic features of a scene (and their locations) have been identified, essentially the scene has been described. Any additional processing produces no additional information about the scene but only reworks it into a form which is more comprehensible (to a human) and hopefully more useful. Consequently, all (or at least much) of the knowledge required to process scenes must be embodied in the scene segmentation algorithms, not the scene interpretation processes. Such an approach suffers from the integration of interpretation information with the segmentation task. The definition of segmentation given below allows an easy division of segmentation and interpretation tasks.

For the purpose of the remainder of this communication, segmentation will be defined as the division of a scene into subscenes which can be characterized by a single common visual clue. For example, a scene can be segmented into subscenes on the basis of digitized gray scale values. Several procedures have been written which produce this type of segmentation (Brice and Fennema, 1970, for example). A single segment in this case consists of all of the points which have a common gray scale value.

Clues other than gray scale value can be used for this type of segmentation. Potter (1974) showed how sequences of scenes can be processed to determine velocity measurements for each picture point. He showed that the motion measurements so obtained can be used to segment a picture in the sense used in this communication. That is, the different segments of the scene contain points which all have the same motion values. Similarly, texture, color, three-dimensional spatial clues, and other features can be used to segment a scene.

The Segmentation Process

Obviously, the different visual clues may produce different segmentations of the same scene. For example, two objects of the same color may have different velocities. In a color segmentation, they would be included in the same segment. In a motion segmentation, they would be in different segments. Consequently, a method must be found which combines the individual segmentations into a single master segmentation.

An obvious approach would be to produce a master segmentation from a linear combination of the individual segmentations. For example, if the output of each segmentation routine is a list of the line segments (hereafter referred to simply as *lines*, to avoid confusion with scene segments) which segment the scene, then the lines in each segmentation can be weighted according to the clue's relative importance (i.e., motion most important, color next, and texture least) and the weights of identical lines summed. The resultant set of lines would produce a complete segmentation of the scene with the most significant scene segments bounded by the most heavily weighted lines.

Scene Segmentation Utilization

The approach to scene segmentation proposed in this communication is advantageous only if a process of scene interpretation can be designed to work with it. A brief proposal of how such a process could be designed follows.

The result of the segmentation process is application independent and consists not of "features," but of "areas of interest." In general, a single area or segment may contain several features. Conversely, several segments may be included in a single feature. Task dependent information would be required to process the individual scene segments and determine the features they contain.

The feature extraction process would not however be done in a single step, but instead would be an interactive process, integrated with interpretation. To begin the process, a procedure (such as picking the segment surrounded by the most heavily weighted lines) would determine which segment should be processed first. Application and environmental factors would determine which set of feature extraction routines should be run. Application of these routines would result in an information base which is mapped by an evaluation function onto a "state" tree.[1] The terminal nodes[2] of a "state" tree would contain semantic output or descriptive information as well as information on how to pick the next segment to process (such as a weight adjustment to the lines and which features to look for. The information obtained from this step would be added to the information base and the process repeated until the entire scene had been interpreted.

References

Brice, C. R. & C. L. Fennema, "Scene Analysis Using Regions," *Artificial Intelligence*, *1*, 1970, 205.

Potter, J., "The Extraction and Utilization of Motion in Scene Description," Ph.D. Thesis, Univ. of Wisconsin, 1974.

The VISIONS System

Thomas D. Williams
Computer and Information Science Department
University of Massachusetts Amherst, MA 01002

VISIONS (for Visual Integration by Semantic Interpretation Of Natural Scenes)[3] deals with the questions posited by the technical topic for the June *SIGART Newsletter*. E. Riseman and A. Hanson are developing the VISIONS research project with the long-range goal of analysing a wide variety of natural scenes. The problems incurred in machine interpretation of natural scenes are manifold. The three questions, however, do summarize several of the major thrusts of our effort.

[1] In this description, the knowledge needed to interpret a scene could be organized as a "forest" of trees, with each tree independent of the others. This would enable information to be easily added or deleted from the knowledge base without disturbing other data.

[2] The branches of "state" tree would represent the application and environmental factors mentioned.

[3] Allen R. Hanson and Edward M. Riseman, "The Design of a Semantically Directed Vision Processor (Revised and Updated)," Tech. Report 75C-1, Computer and Information Science, University of Massachusetts, February 1975.

1) What knowledge is necessary to segment and interpret scenes? Knowledge is present in several areas of VISIONS which is used to segment and interpret. Application of *general* knowledge about scenes is a very powerful tool for scene analysis. We propose a hierarchy of functions which first (in parallel) reduce the image to sets of features. A second set of functions capable of assembling feature sets into object likelihoods (called vision routines) utilize knowledge about consistencies of object attributes. The application of these vision routines is directed by a model-building process which uses knowledge of spatial, temporal, and functional relationships of objects in the real world. This representation level we call "semantic". This visual semantic data base includes both expected relationships between objects and also the attribute value of objects.

2) How should knowledge be represented and used? VISIONS represents knowledge in various subsystems according to its use. For example, a vision routine need only "know" the properties of the object for which it is designed to detect. This information is coded into the routine directly. However, there is variable information, such as color (of leaves, in a tree routine). This information is dependent on contextual information (e.g., seasonal) retrieved from semantic knowledge store. It can be organized as packets of dependent information so that sets of regions can be analyzed in a dependent fashion. Procedural representations of perspective, occlusion, and shadows allows (possibly redundant) heuristic information to generate and/or verify hypotheses. Simple axiomatized information can also be used by a deductive process which checks the consistency of a proposed partial model.

3) Are there methods general enough for a wide class of scene types? "Low level" routines which essentially extract features for a specific set of vision routines are of a general nature. Line-finders, region-growers, texture descriptors, size, shape, color, and other features have already been programmed in an emulation of a parallel machine [1]. These general processes have given favorable results and indicate that the methods employed might also be applicable to a wide variety of natural scenes. Of particular importance is the flexibility of interface between low and high levels, yielding possible application to other domains.

Note that VISIONS is a model-building approach to natural scene analysis. As such, it may use feedback or self-directing techniques which allow us to describe a sufficient rather than a necessary set of components to accomplish the task of computer visual perception.

Semantic Picture Recognition[2]

Michael L. Baird[3]
School of Information and Computer Science
Georgia Institute of Technology Atlanta, Ga 30332

This paper is a slightly edited version of a research progress report prepared for an internal publication of the Georgia Institute of Technology. It is reproduced here to make its contents more widely available.

The goal of this work has been the development of a systematic approach to the recognition and description of pictures by computer. At the commencement of the project, existing

[1] Allen R. Hanson and Edward M. Riseman, "Preprocessing Cones: A Computational Structure for Scene Analysis," Tech. Report 74C-1, Computer and Information Science, University of Massachusetts, September 1974.

[2] Research described here was supported in part by grant GN-665 from The National Science Foundation.

[3] Author's current address: 39654 Balboa, Sterling Hts., Mich. 48078.