A Knowledge-Based Approach To Computer Vision Systems

Martin D. Levine
Wade Hong

Computer Vision and Robotics Laboratory
Department of Electrical Engineering
McGill University
Montréal, Québec, Canada

Abstract

In designing and constructing computer vision systems, many crucial issues need to be addressed. Foremost of these are the control and organization of the visual information processing tasks involved, and the representation and usage of both knowledge and data. As computer vision systems have evolved, growing in complexity and size, these issues have become increasingly important to their overall success. In this paper, a recent and increasingly popular approach to image understanding, the knowledge-based system, is presented as a framework in which to deal with these issues. The engineering of a computer vision system as a knowledge-based system and these issues, in the context of our evolving system is discussed.

Résumé

Lors de la conception et de la mise en oeuvre d'un système de vision par ordinateur, plusieurs questions critiques doivent être considérées. Principalement, il s'agit du contrôle et de l'organisation des tâches de traitement d'information visuelle ainsi que de la représentation et de l'usage des données et des connaissances. Parce que les systèmes de vision par ordinateur ont évolué en grandeur et en complexité, leur succès dépend de plus en plus de ces questions. Dans cet article, une approche nouvelle et de plus en plus populaire à la compréhension d'images, le système basé sur les connaissances, est présentée en tant que cadre de travail pour traiter ces questions. La réalisation d'un système de vision par ordinateur par le biais d'un système basé sur les connaissances ainsi que ces questions sont traitées dans le contexte de notre système en évolution.

1. Introduction

A visual technology capable of replicating human vision is the ultimate achievement for computer vision. To be able to accomplish such a feat would require a far superior understanding of the functioning of the human visual system. Moreover, this would require the embodiment of intelligence in a machine. Undaunted by these severe limitations in understanding, computer vision has developed over the past twenty-five years in a somewhat ad hoc fashion. The growth of this infant technology in conjunction with its maternal science of artificial intelligence has led to the emergence of computer vision systems. Albeit they are far from being general vision systems [+] they are at present the best and only available artificial approximation.

The earliest computer vision system, pioneered in the mid 1960's by Roberts [Roberts65], was capable of analyzing simple polyhedral scenes and matching the located polyhedra to stored models. Since then, computer vision systems have attained greater complexity due to the increasingly complex scenes being analyzed, as witnessed in the prominent systems of today. (See [Binford82] and [Shapiro83] for surveys on some of these systems.) In association with this increase in complexity, the control and organization of these systems have evolved from simple sequential bottom-up or top-down mechanisms into complex structures involving many levels of cooperative processes, as the amount of knowledge required to reason about the analysis increases. As these complex visual information processing systems become more ambitious, it is clearly evident that the organization and control aspects will also become increasingly more significant to their overall success.

Control of vision systems have tended to be heavily embedded within the organization of the visual processes. Such procedural methods are reliable and fast, but are very rigid in that they are application specific. Subject to variations in the goal description or the task domain, the appropriate alterations to the procedural knowledge may become a major task. Also, if the images to be analyzed consist of complex structures and great intra-class variations, a sequence of analysis cannot be reliably predetermined. Thus the analysis is necessarily data-driven, implying the need for a flexible and adaptive control struc-

+ By general it is meant in the same sense as the human visual system, capable of multiple objectives in a dynamic, unconstrained and complex visual environment ture.

This paper presents a recent and increasingly popular approach to the organization of a computer vision system, permitting a greater degree of control flexibility and subsequently, functional generality. The paradigm presented is that of a knowledge-based system.

2. The Knowledge-Based Approach

A significant result in the first twenty years of artificial intelligence research is the fact that the principal requirement for intelligence is *knowledge*. By the mid-1970's AI began shifting from a power-based strategy towards a knowledge-based approach in an attempt to achieve intelligence. The power strategy looked towards a generalized increase in computational power in resolving the problems that the current techniques faced, whereas the knowledge strategy viewed progress being achieved from better ways of recognizing, representing and utilizing diverse and specific forms of knowledge. The fundamental problem of understanding intelligence is no longer the identification of power-based techniques, but rather a question of how to represent vast amounts of knowledge in a manner which permits their effective use and interaction.

A powerful tool that has emerged from this shift of focus in AI is the knowledge-based system which is a problem solving system that applies knowledge about a specific domain to solve practical problems [Sowa84]. A class of knowledge-based systems known as expert systems has recently received much attention [Waterman, Hayes-Roth&Lenat83].

Knowledge-based systems have either adopted or developed programming styles where there exists a clear distinction between knowledge and its use (for an introduction to and survey of a few of existing tools, see [Waterman,Hayes-Roth&Lenat83], pp. 169-215). This separation of control flow from its knowledge permits modular extensions to a system's capabilities. The knowledge engineering tools that have emerged employ principles of knowledge representation and a related inference mechanism for bringing knowledge to bear on a problem. Knowledge about the problem domain and self-knowledge are stored in a knowledge base using a representational framework. Current representational frameworks include rule-based, frame-based and logic-based schemes [Buchanan&Duda83]. Facts or data about the particular problem and processing are stored in a global database. The system retrieves pertinent knowledge to the problem and utilizes symbolic reasoning to make inferences about the facts in the global database to solve the problem at hand.

Although one of the first domains of research in AI to incorporate knowledge was computer vision, the extent of improvement in this application has been slow and limited. The application of knowledge has been restricted to domain specific knowledge of the scene analyzed in model-based vision. However, the use of world knowledge has been weak [Binford82]. There is now interest in the computer vision community to apply knowledge-based system techniques to improve this level of processing [Matsuyama 84,Nagao 82].

As complex and large as current computer vision systems are, they are very limited in their abilities [Binford82,Matsuyama84]. Much effort, of late has been directed towards improving and understanding specific vision tasks, particularly, in low level vision [Brady82]. A major emphasis in this work has been focused on the use of physical knowledge – knowledge about the physical world and the laws that govern it. Shape from shading and stereo vision, for example, use knowledge about the imaging process to recover 3D shape from projected 2D image features. More recently, another level of knowledge has been introduced in computer vision systems, perceptual knowledge – knowledge used to group image features into aggregates. The basis for this knowledge comes from Gestalt laws of visual grouping [Zucker,Rosenfeld&-Davis75]. Such knowledge has been successfully applied in refining low level segmentations [Nazif83] and forming perceptual groupings from 2D image features as the basis for 3D object recognition [Lowe84].

Apart from the need to improve every facet of the image analysis process, there is also a need to increase the overall intelligence of these image understanding systems [Rosenfeld82, Matsuyama84]. The capacity of intelligence implies the ability to reason about the image analysis and the scene. Rosenfeld identifies a lack of a general theory of control in image analysis, i.e. there exists no general principles describing how vision processes should interact in performing a particular task. He also identifies a lack of a general theory of how to combine evidence from multiple sources of information available in performing a particular task. Such general purpose knowledge is imperative if hopes of achieving a general vision system are to be satisfied.

To achieve functional generality, a computer vision system must be capable of performing a variety of tasks. Upon specification of a particular task, the system must be able to determine the necessary processing modules, parameters and control strategy for performing the task. Given the requirement of being able to analyze a wide variety of complex images, this cannot be rigidly specified a priori. The system should possess the ability to evaluate its performance at various stages of processing and be capable of adaptively improving it (whether it be by modifying parameters, modifying the control flow, integrating information, augmenting processes, or other mechanisms). Thus, it is necessary that the image being analyzed and its many abstractions dictate the processing flow and consequently, how the vision processes should interact. The control of the image analysis is therefore necessarily data-driven. This type of flexible control is easily realizable in a knowledge-based methodology.

Ultimately, computer vision must address the important issue of integration of evidence from multiple sources, especially in view of the increased sophistication in applications and the need for improved performance. This is especially desirable since descriptions produced by computer vision techniques are incomplete and often imprecise, stemming from the inherent ambiguities that arise in an image. For example, consider the problem of image segmentation where partitions may be obtained from several measurable or extractable properties such as colour, luminance, texture or edges. In typical computer vision applications the "best" technique for segmenting the image, based on a single property, is often used to build an intermediate representation ‡ for the higher level processes. This "best" technique is often arrived at by trying a set of techniques and deciding on the best. However, it is necessary in a general system, where the "best" technique is not definable, to have a larger number of techniques available, and in some way be capable of integrating the results of these techniques into a "best" possible usable intermediate representation. Integration of this nature can be viewed as a refinement process which operates on local extracted features. Nazif [Nazif83] has demonstrated the refinement of low level segmentations using a rule-based mechanism to represent processing knowledge for integrating information from a line-based and a region-based segmentation. Note that the integration of information can also be useful in the refinement of the interpretation or recognition processes.

Given the importance of knowledge in image analysis, the engineering of a computer vision system as a knowledge-based system is very appealing. However, to have successful systems, the knowledge levels (physical, perceptual, domain and processing) must be further enhanced and the use of this knowledge be more effectively applied. Also, an appropriate knowledge engineering tool for vision applications must be formalized.

3. Our System

The aim of our system is to build a general purpose tool for experimenting with various approaches to image analysis. Constructing the system as a knowledge-based system permits us the flexibility to do so. In such a system where there is a distinct separation between its knowledge and the mechanisms that apply it, the task domain or its goals may be changed easily and as the system evolves, the modular extensibility of its capabilities by simply augmenting its knowledge is attractive. Equipped with a large set of visual processing algorithms and modules, by setting up the task domain and selecting the appropriate analysis strategy, this computer vision system can attain a greater degree of functional generality and utility. Also, due to its data-driven nature, this system can be attentive to the processing requirements as dictated by the image, demonstrating the capacity of dynamic control [Levine&Nazif85b].

The basic computer vision system is identified as consisting of two major processing modules performing the low level or early processing and the high level or cognitive processing. Low level processing is concerned with extracting image features and structures to build an intermediate representation. The principal task of the high level process is to match object models with structures described in the intermediate representation. Achieving object recognition or scene interpretation is the product of both of these levels of processing. A meta supervisor coordinates the interaction and flow of information and processing between both processes. This simple organization is depicted in Figure 1. We follow the doctrine of separating the domain independent knowledge from the domain dependent knowledge in this form of dichotomy.

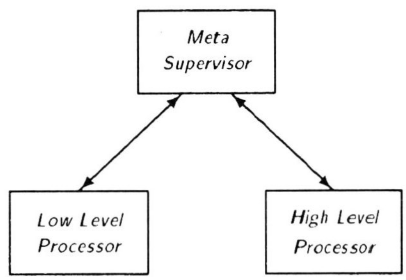

Figure 1 Basic System Structure

The organization of this system is presented in this fashion to express flexibility and generality which is permitted by the knowledge-based system paradigm. Although the interaction between the low level processor and the high level processor may be simply a one pass sequential flow or be governed by a hypothesis-verification paradigm, this arrangement permits either explanation. The point is not to mask the control structure but to emphasize that a knowledge-based approach permits greater flexibility in control. Changes in control strategy require only alterations in the meta control knowledge embedded in the meta supervisor or its usage as opposed to major reorganization necessary in a more conventional procedural control structure. Conceptually, the low level and high level processors and their respective subtasks are viewed in the same manner. For example, the low level processor has its meta supervisor controlling its subprocesses and similarly these subprocesses have supervisors controlling their respective subprocesses. Each of these respective processes are themselves self-contained knowledge-based systems. Organizing the system in this fashion suggests a natural pyramid or tree hierarchy for the control of the entire system.

‡ Intermediate representation is the general term used to describe the representations produced at various stages of processing between the signal (image) and the semantic (scene) levels. For our purposes by intermediate representation, we mean the principal representation that is used by the interpretation (high level) process.

4. Our Current Work

A system of the nature described above is currently evolving at the Computer Vision and Robotics Laboratory at McGill University. The knowledge representation framework chosen for the system implementation was a rule-based methodology and OPS5, a production system language [Forgy81, Dill&Hong84] was selected as the knowledge engineering tool. This latter choice was based primarily on availability.

A low level processor based on Nazif's low level segmentation expert [Nazif83] has been implemented and is currently being tested. Extensions to the capabilities of this system are currently being implemented. Work will be initiated soon on the high level processor.

The low level processor possesses the ability of non-purposive segmentation. A final partitioning of an image is obtained from the integration of initial region- and line-based segmentations. This integration is facilitated by the three knowledge sources which comprise the segmentation module: the line, region and area analyzers (see Figure 2). Each of these analyzers consists of rules which reason about the entities extracted from the image, i.e. lines, regions and areas of attention. These heuristics are domain independent, being based on the principles of visual grouping [Nazif83, Zucker, Rosenfeld&Davis75]. As well as the need for these heuristics to achieve the segmentation, some knowledge about how to apply them is also required. Hence the control problem.

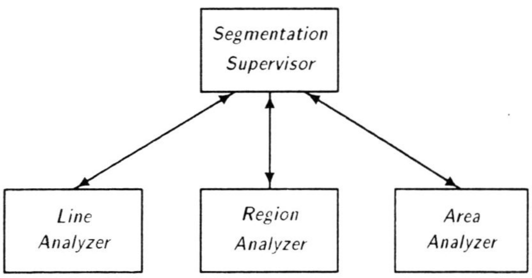

Figure 2 The Segmentation Module

Control is effected by dynamically setting strategies for the processing of areas, regions and lines. The selection of the strategies is based on a fuzzy concept of a region's or line's "need for further processing". A measure of this fuzzy notion is discernable from a set of performance parameters [Levine&Nazif85a, Nazif83] reflecting the quality of the segmentation at that instant in processing. Such a control strategy is very appealing in that it attends to the needs of the current segmentation and also by nature is domain independent.

The resulting intermediate description obtained from this segmentation module is a region-based representation of the image. However, the low level processor that is envisioned would combine many functional modules to provide a rich intermediate representation of the scene, of which the segmentation module is one (see Figure 3). A second module now being implemented, which transcends the picture domain, is concerned with the extraction of scene domain cues. Such three-dimensional cues as occlusion, cast shadows, and skewness, extractable from the image contour, gives rise to some depth and orientation information. Exploiting this information, the shapes of objects may be inferred. This would yield an object-based segmentation of the scene. Similar to the segmentation module, the resulting partition of the scene would be obtained from the integration of the refined, region-based segmentation and this initial object-based segmentation. With the addition of other modules (perhaps a segmentation based on texture or a surface recovery module based on laser vision), the required integration would certainly be of greater complexity.

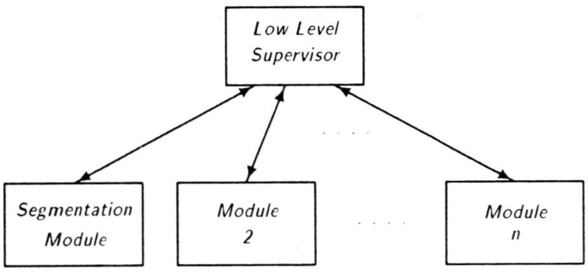

Figure 3 The Low Level Processor

The described low level processor, a general purpose subsystem by design, is oblivious to the task domain. It is in the high level processor where interaction with world knowledge is a necessity to achieve recognition or interpretation tasks. To accomplish this, the high level processor must possess the ability of matching object models to the intermediate representation supplied by the low level processor. More specifically, it must be able to resolve ambiguity (which is inherent in both the image data and world knowledge) and to identify instances of the object models by examining the consistency amoung local image features.

Some common paradigms that have been employed in image analysis include constraint propagation, template matching and hypothesis-verification [Matsuyama84, Binford82]. In these methods, initially some match or inference is made of image features to object models. Then these initial inferences are verified for local consistency whether in a sequential manner as is the case for template matching and hypothesis-verification, or in parallel for constraint propagation. Local consistency at some level is sufficient for object recognition tasks, but for interpretation, the inferences must be propagated to attain global consistency. These paradigms may be viewed as consisting of two characteristic mechanisms, one to make the initial inferences or matches and the other to propagate them (see Figure 4).

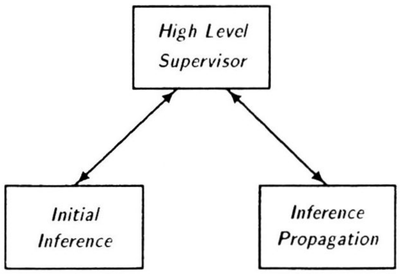

Figure 4 The High Level Processor

The objective of this high level processor is to achieve a scene description or object recognition given an object-based intermediate representation. But because the high level process is inherently limited by the quality of low level segmentations, ambiguity may not be easily resolved. Therefore, the high level process should have the ability to integrate evidence from other intermediate representations (region-based, line-based, etc.) in the inference forming and propagation processes. As a final recourse in the face of unresolvable ambiguity, the high level process should be able to request that the low level process either further refine or re-construct, a part or the whole of the intermediate description.

Work is now being initiated on the development of such a high level processor.

5. Discussion

Though the construction of a computer vision system as a knowledge-based system is very attractive, problems do however present themselves. They stem from the limitations and deficiences of the representational framework, the knowledge engineering tool, its data-driven nature and knowledge itself. These shortcomings are not unique to this application, they are apparent in knowledge-based systems in general.

A major part of the effort in building a knowledge-based system is the identification and acquisition of pertinent knowledge applicable to the problem. Such knowledge is limited in its scope, incomplete and inexact, because we lack complete laws and theories about the problem. This is representative of the various knowledge levels (physical, perceptual, domain and processing) present in computer vision systems. Often the knowledge is ill-specified because it is not clear what exactly is known about the problem or how to apply it. To improve the performance of knowledge-based visual information processing systems a greater amount of knowledge must be identified and applied to the problem. Unlike the domain of expert systems, where there exist experts from which knowledge is accessible through interaction, knowledge useful to computer vision systems must be determined from the slow process of understanding human vision.

Control flow in a computer vision system such as ours is governed by the data, but this data is often unreliable and incomplete. As a consequence, such a system could easily run astray. Coupled with ill-specified knowledge, the possibility is even greater. To cope with this problem, either the integrity of the data must be substantiated in some manner, by for example, incorporating redundancy (confirmation or combination of evidence from multiple sources) or the ability to reason with uncertainty must be established.

Although OPS5 is a general purpose production system programming language, our experience has shown that as a tool for constructing computer vision systems it suffers from several deficiencies. The principal one is that is inadequate for representing the diverse knowledge and data that must be embodied. The predominant nature of knowledge that must be encapsulated, especially at the low level is procedural; that is, it prescribes a set of operations. However, OPS5 does not facilitate procedural mechanisms nor complex computations on the right hand side of a rule. To capture a "chunk" of knowledge often requires the chaining of a set of productions. As well, there exist no generic control mechanisms that permit the accessing of a set of data in an orderly fashion, that is, the application of a rule (or a set of rules) sequentially on a set of data. Nevertheless, it is actually possible to accomplish this, but it requires the construction of specific control rules and the generation of control state data to ensure the proper processing flow. Finally, the data representation capabilities of OPS5 do not facilitate the representation of the lowest forms of visual data. There are no data structures for maintaining images, nor are there constructs to manipulate them.

These inadequacies and others using this knowledge engineering tool, though not insurmountable, suggest that perhaps some of our future work should be directed towards developing a more suitable knowledge engineering tool for constructing knowledge-based computer vision systems. An adequate tool would make the system more efficient and manageable. However, the specification of such a tool would require one to first identify the requirements necessary for building a knowledge-based computer vision system.

The rule-based methodology is a very general and flexible framework for representing knowledge and data, as is evident by its prevalant use in expert systems, covering a wide scope of problem domains. Even so, it is found to be not entirely adequate for our purposes. Subject to the nature of certain representations and processing requirements in our system, our experiences with OPS5 as discussed above, have shown that a classic pure production system model has its deficiencies. This suggests that a purely rule-based representational framework is not appropriate. A blend of the rule-based model and the imperative model would be more suitable.

The work that we have described here is only in its formative stages. Though we cannot yet conceive of all the many problems that will face us, we are however beginning to understand some of the major issues involved in attempting to build such a massive system. This knowledge will become invaluable in the future evolution of this system.

Acknowledgements

This work was partially supported by the Natural Sciences and Engineering Research Council and the Ministry of Education of the Province of Québec. Martin D. Levine would like to thank the Canadian Institute for Advanced Research for its support. The authors would also like to thank Kamal K. Gupta for his constructive comments.

References

[Binford82] Binford, T.O., *Survey of Model-Based Image Analysis Systems*, The International Journal of Robotics Research, Vol. 1, No. 1, MIT Press, Cambridge, Massachusetts, 1982.

[Brady82] Brady, M., *Computational Approaches to Image Understanding*, ACM Computing Surveys, Vol. 14, No. 1, March 1982.

[Buchanan & Duda 83] Buchanan, B.G., and Duda, R.O., *Principles of Rule-Based Expert Systems*, Advances In Computers 22, Academic Press, New York, 1983, 163-216.

[Dill&Hong84] Dill, A.R. and Hong, W., *Everything You Wanted To Know About OPS5 But Were Afraid To Ask (A Practical Guide To Using OPS5 at CVaRL)*, Technical Report 84-5R, Computer Vision and Robotics Laboratory, Department Of Electrical Engineering, McGill University, Montréal, August, 1984

[Forgy81] Forgy, C.L., *OPS5 User's Manual*, Technical Report CMU-CS-81-135, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 1981.

[Havens&Mackworth83] Havens, W.S. and Mackworth, A.K., *Representing Knowledge of the Visual World*, IEEE Computer, Vol. 16, No. 10, October, 1983, 90-98.

[Levine&Nazif85a] Levine, M.D. and Nazif, A.M., *Measurement of Computer Generated Image Segmentations*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-7, No. 2, March, 1985, 155-164.

[Levine&Nazif85b] Levine, M.D. and Nazif, A.M., *Rule-Based Image Segmentation: A Dynamic Control Strategy Approach*, Computer Vision, Graphics and Image Processing, Vol. 32, No. 1, October, 1985, 104-126.

[Lowe84] Lowe, D.G., *Perceptual Organization and Visual Recognition*, Ph.D Dissertation, Report No. STAN-CS-84-1020, Department of Computer Science, Stanford University, Stanford, California, September, 1984.

[Matsuyama84] Matsuyama, T., *Knowledge Organization and Control Structure in Image Understanding*, Proceedings of the Seventh International Conference on Pattern Recognition, Montreal, 1984, 1118-1127

[Nagao82] Nagao, M., *Control Strategies in Pattern Analysis*, Proceedings of the Sixth International Conference on Pattern Recognition, Munich, 1982, 996-1006.

[Nazif83] Nazif, A.M., *A Rule-Based Expert System for Image Segmentation*, Ph.D Dissertation, Department of Electrical Engineering, McGill University, Montréal, March, 1983.

[Roberts65] Roberts, L.G., *Machine Processing of Three-Dimensional Solids*, in Optical and Electro-optical Information Processing, ed. M. Tippett, MIT Press, Cambridge, Massachusetts, 1965, 159-197.

[Rosenfeld82] Rosenfeld, A., *Image Analysis: Progress, Problems and Prospects*, Proceedings of the Sixth International Conference on Pattern Recognition, Munich, 1982, 7-13.

[Shapiro83] Shapiro, L.G., *Computer Vision Systems: Past, Present, and Future*, in Pictorial Data Analysis, ed. R.M. Haralick, Springer-Verlag, Berlin, 1983, 199-237.

[Sowa84] Sowa, J.F., *Conceptual Structures, Information Processing in Mind and Machines*, Addison-Wesley, Reading, Massachusetts, 1984.

[Waterman,Hayes-Roth&Lenat83] Waterman, D.A, Hayes-Roth, F. and Lenat, D., eds., *Building Expert Systems*, Addison-Wesley, Reading, Mass., 1983.

[Zucker,Rosenfeld&Davis75] Zucker, S.W., Rosenfeld, A. and Davis, L.S., *General Purpose Models: Expectations About The Unexpected*, Advance Papers of the Fourth International Joint Conference on Artificial Intelligence, Tbilisi, Georgia, USSR, 1975, 716-721.

[Zucker81] Zucker, S.W., *Computer Vision and Human Perception, An Essay On The Discovery Of Constraints*, Proceedings of the Seventh International Joint Conference on Artificial Intelligence, Vancouver, 1981, 1102-1116.