Pattern Recognition Vol. 17, No. 1, pp. 13–27, 1984.
Printed in Great Britain.

0031-3203/84 $3.00 + .00
Pergamon Press Ltd.
© 1984 Pattern Recognition Society

KNOWLEDGE AND THE VISUAL PROCESS: CONTENT, FORM AND USE

JOHN K. TSOTSOS
Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S 1A4

(*Received* 10 *February* 1983; *received for publication* 6 *May* 1983)

Abstract—A knowledge based system for the analysis of imagery clearly requires large amounts of domain specific knowledge and also requires a recognition control scheme that will manipulate this knowledge in order to interpret the imagery that represents various scenes of the domain. Many current systems indeed satisfy this statement. In addition, however, they all contain modules that access the actual image data and process this data. Typically, the methodologies for the image specific aspects and the domain specific aspects are separate yet interact, and the representational formalisms and control schemes for these two tasks are not related.

This paper will attempt, by overviewing a current hypothesis of the kinds of knowledge required for general purpose vision and the current representational tools available, to reconcile the "low" and "high" levels of knowledge based vision systems and to propose a set of uniform representational tools. The discussion will be at the conceptual level and not at the implementational level. Pointers to current computer vision schemes that are relevant to the discussion will be given. Several good surveys and discussions of requirements of vision systems can be found in Nevatia,[1] Nagel,[2] Hanson and Riseman,[3] Barrow,[4] Weszka,[5] Reddy,[6] and Kanade.[7]

Artificial intelligence
Biological visual perception
Image understanding
Representation of knowledge

I. INTRODUCTION

What does the term "knowledge" mean? Philosophers have struggled with this concept for ages. A recent discussion on the meaning of knowledge for computer systems is presented in Newell.[8] Newell considers "the knowledge level" of computer systems and defines knowledge as "whatever can be ascribed to an agent, such that its behaviour can be computed according to the principle of rationality". The principle of rationality states that actions are selected to attain goals: "if an agent has knowledge that one of its actions will lead to one of its goals then the agent will select that action". These are purely functional characterizations, not structural. A symbol system is required to manipulate knowledge, thus a representation scheme provides an access mechanism to knowledge. Considerations of knowledge *content* are distinguished from knowledge *form* and knowledge *use*. For the vision domain, since an image underconstrains the scene that it represents, it may be that the principle of rationality is the reduction of ambiguity and that actions are taken by the vision system in order to move towards an unambiguous interpretation of the image. This statement is, however, rather vague and requires much elaboration. This elaboration will only be briefly touched upon in this presentation.

If indeed the driving principle behind computer vision systems is the reduction of ambiguity, the appropriate actions that could be taken by the system to satisfy this requirement must be identified, quantified and represented within the system, and structures for their manipulation must be formulated. It is currently not clear what such actions could be. An expert systems framework with which to experiment with various methodologies for visual information processing may be appropriate at the current level of understanding of the visual process. Expert systems are characterized by the use of significant amounts of knowledge to assist in interpretation of the data of some domain. I will not address some common issues in expert systems research, such as user community acceptance of the system, performance, etc., since for the current state-of-the-art of vision research, these are not important. It is important, however, to emphasize that research in expert systems is not at all complete and that there are many open issues remaining.

The expert systems approach has led to many interesting and useful computer systems in a variety of domains. Examples of vision systems that have such a flavour are VISIONS,[9] IGS,[10] Levine,[11] ARGOS[12 and others. The success of these systems has not, however, been as remarkable as that of medical diagnosis systems for example.[13] I attribute this to several factors: explicit representations of knowledge have been reserved mainly for application domain knowledge, while general visual knowledge has been buried away in procedures, i.e. there is no uniform formalism within which to define, code and manipulate all of the knowledge of the system; there is no representational formalism for combining concept definition with the process for extracting instances of that concept; there exists no well-defined mechanism for the generation and use of a focus of attention, or expectations, in vision systems, even though its importance has been emphasized by Mackworth's "cycle of perception",[14] and Kanade's model.[7] At this point of research, tools are required that are flexible enough for experimentation with possible mechanisms for interaction, communication, integration, etc., of subprocesses, and expressive enough so that both domain and general knowledge is easily included and used within the same structures. In addition, the tools must be semantically well-founded so that analysis of system performance can be accomplished and so that the underlying foundations are reliable and sound—it is hard enough to debug the knowledge without worrying about debugging the representational formalism at the same time.

A variety of representational and control tools have been employed in expert or knowledge-based systems. This presentation will not deal with all of them. Rather, I will try to motivate the need for some of the more common and well-understood ones for computer vision systems. A relatively complete list of the kinds of tools can be found in Brachman and Smith.[15]

When confronted with a large, complex task, in this case vision, "divide and conquer", or at least try to conquer, is an obvious tactic. Arbitrary task subdivision will yield structures that are unwieldy, unnecessarily complex or inappropriately simple, have poorly defined semantics, lead to inefficient processing and lack clarity and perspicuity. Within the existing representational repertoire, there exist two common tools for domain sub-division and organization, namely the *is-a* relationship (or generalization/specialization axis), and the *part-of* relationship (or the part/whole axis). Brachman,[16] Levesque and Mylopoulos[17] and Brachman[18] provide discussions on their properties, semantics and use. A third, but less well understood and used relationship is that of *projection*.[7] This is a relationship between domains, such as those between two levels of processing, and usually connects more abstract notions to more detailed ones.

The IS-A axis provides for economy of representation by representing constraints only once, inheritance of constraints along the IS-A relationship, a natural concept organizational scheme, and a partial ordering of knowledge concepts that is convenient for top-down search strategies. The PART-OF axis allows control of the level of detail or resolution represented in knowledge packages and thus the knowledge granularity of the knowledge base, i.e. the size of the knowledge packages. It provides for the implementation of a divide-and-conquer representational strategy and it forms a partial ordering of knowledge concepts that is useful for bottom-up search strategies. In addition, grouping processes which are so important in vision can be represented via this axis. Finally, the PROJECTION dimension allows for the realization of expectation biases and the enforcement of top-down grouping constraints.

These three representational tools described above are organizational axes that connect pieces of knowledge. Frames, classes and prototypes are common names for knowledge packages that include both assertional and procedural knowledge. In this presentation, prototypes will be considered as active computing units that are organized along the is-a and part-of axes. Packaging up knowledge leads to a modular representation, with all the advantages of modularity, particularly the enhancement of clarity and flexibility. Package size is referred to as the knowledge granularity of the representation. Most knowledge package representation schemes borrow strongly from Minsky.[19]

In the many vision systems that employ domain knowledge, there seems to be agreement that at the "high level", many of the usual representational tools are useful. For example, generalized cylinder representations, such as presented in Nishihara[20] and the ACRONYM system,[21] make heavy use of specialization in defining classes of cylinders and also of aggregation in combining several into an object (for example, a human body or an airplane). Land and water mass concepts are organized in this way in the MAPSEE 2 system of Mackworth & Havens[22] and this organization is exploited for recognition purposes. Finally, motion concepts are organized along the is-a hierarchy in the ALVEN system,[23] where recognition proceeds from the more general to the more specific concepts in this hierarchy in a constrained manner. However, the image specific portions of the system remain distinct. Why? Perhaps there are many reasons, and I will suggest three. Representation of knowledge research has concentrated on the problem of common sense reasoning or natural language understanding. In the latter case, for example, in a sentence there are a small number of possible combinations of words when compared to the number of combinations of pixels in an image. The extraction of a word is a trivial matter, the extraction of meaningful picture elements is not. Thus, for vision, much more so than for natural language, the representation for concepts that are extracted from an image must be intimately tied to the process that actually does the extraction. This process will be referred to as the aggregation or grouping process and more will be said about this later on. Secondly, throughout processing, there is a need for optimization or enhancement of the output and this plays a crucial role in the reduction of local ambiguity. In natural vision this is the process of lateral inhibition (see Zucker[24] for an introduction to this). This has led to the cooperative algorithms common in vision systems, perhaps the best understood being relaxation labelling processes.[25,26] A prerequisite for such processes is parallel communication among concepts and processes. Finally, the notion of processing hierarchies is useful and also has a counterpart in natural vision. Bottom-up and top-down communication is required. These are three representational issues that have received little attention from the representation-of-knowledge community, yet are crucial for computer vision systems.

II. KNOWLEDGE OF THE VISUAL PROCESS

The approach that I will take is the following. The definition of a concept and the process that extracts instances of that concept from an image will be packaged together. These knowledge packages, as active computing units, must accept input, must produce an output and have side-effects. Rather than introducing any new mechanisms, I will attempt to show, using current knowledge of biological visual systems as an example, that most of the concept communication can be accomplished by exploiting the is-a, part-of and projection organizational axes to the fullest. Note that the discussion is at Newell's "knowledge level" and is not a statement about the implementation of this knowledge.

At this point several distinctions must be drawn. The first is that between the information that is in an image and the information about the scene that is represented by the image. In psychology, this is the distinction between sensation and perception. From James,[27] "Sensation, then, ... differs from perception only in the extreme simplicity of its object or content. Its function is that of mere acquaintance with a fact. Perception's function, on the other hand, is knowledge about a fact. ... In perception, there are voluminous associative processes in the cortex, while in sensation alone, there are a minimum of processes. Pure sensation without some accompanying perception in an adult is impossible. Perception, thus, differs from sensation by the consciousness of further facts associated with the object of sensation. ... Sensational and associative brain processes combined, then, are what give us the content of our perceptions". This distinction will be maintained throughout the discussion below. The use of common vision terms such as "edge" when in the context of sensations will mean simply some pattern of intensities in an image and will not have a physical correlate, while in the context of perceptions it will refer to the internal representation of a portion of a physical entity.

The second distinction to be made is that although during the discussion examples of visual processing will be taken from the current understanding of biological vision, it should be clear that the intent is to uncover basic principles and basic kinds of information. It should also be clear that there are several competing theories of biological visual information processing, although I will only be describing portions of some of them. The main concern is with the "knowledge content" of visual processing. The presentation at this level is not concerned with any form of implementation of these principles. The principles will be developed by example, using the only complete vision systems that are available to us for study—namely, biological ones. General references for further reading on the characteristics of biological visual systems are Kandel and Schwartz,[28] Davson,[29] Cornsweet,[30] Caelli[31] and Wickelgren.[32]

II.1. *Grouping and response optimization processes*

Perhaps the most important and common phenomenon in visual information processing is that of the grouping of features into more abstract features and the enhancement of the result of this process. Psychologists have described grouping principles for many years, yet they have not been adequatly defined and quantified for use in computer vision systems. These principles play a role in the grouping of features into higher order ones, whether of form, motion, colour or depth information. It should be noted that no one of these alone can be guaranteed to solve a given task and that an interaction among several of them should be considered. The mechanism of interaction is an area that requires further research.

In addition, because of ambiguities in the image data, an optimization process is required to enforce the grouping principles in a local context. This is the process of lateral inhibition in the visual system. Lateral inhibition processes motivated much of the relaxation labelling process research (see Zucker *et al.*,[33] Bridgeman[34] and Zucker[24] for discussions on this topic). Lateral inhibition requires parallel computing unit communication over a local neighbourhood of units, with each unit sending signals of inhibition to other units as appropriate. Groupings and lateral inhibition always are co-operating mechanisms. Following Wertheimer[35] and others, the grouping principles are:

Proximity. The closer the features are, the stronger their tendency to be grouped together.

Similarity. The more similar features are, the stronger their tendency to be grouped together. This would include form components such as slopes of lines, colour and texture.

Continuity. Features that are enclosed within a single contour tend to be grouped together.

Smooth continuation. Features are grouped together so that they form contours that have as few abrupt changes of direction as possible.

Symmetry. Features tend to be grouped so that they form symmetrical shapes.

Familiarity. Familiar objects or concepts are favoured.

Common fate. Objects that move with similar motion parameters (velocity, trajectory) tend to be grouped together.

Examples of schemes that integrate some of these (proximity, smooth continuation, similarity) are, for the motion correspondence problem, the work of Ullman,[36] and for spatio-temporal aggregation employing the common fate principle, the work of Flinchbaugh and Chandrasekaran.[37] Prazdny[38] investigated the creation of subjective contours from random dot stereograms in which all elements in an enclosed area are displaced together as a whole, using discontinuity in image displacement vector field. Smooth edge continuity is the basic principle behind applications of relaxation processes to edge enhancement,[25] while smooth temporal continuity was the driving force behind the relaxation process for motion recognition of Tsotsos *et al.*[39] Finally, Zucker[40] describes a scheme based on quantifications of several of the above grouping principles for the grouping of dot patterns to form subjective contours.

II.2. *The visual sensations*

There seem to be, according to current understanding, at least four separate groups of sensations: form, change, colour and depth. A parallel is drawn to the intrinsic image scheme of Barrow and Tenenbaum,[41] as well as to the multiple pyramid approach of Levine.[11] It will be seen that the scheme that biological systems seem to employ is a combination of these two approaches. The sensations provide input for a perception mechanism and perceptual computing units monitor the different sensations and may guide their processing. The sensations will be briefly described, with computing units being characterized by their sensitivity to their inputs, their specificity for input and their dependencies on other information.

II.2.1. *The abstraction of simple forms: feature aggregation.* It has been proposed by Hubel and Wiesel[42] that certain portions of the visual system can be viewed as representing a hierarchy of abstraction. The type of abstraction is really a combination of aggregation (along the part-of axis) and generalization (along the is-a axis). At each level, each cell sees a greater perspective than at an earlier level and its ability to abstract is increased.

The first major level of visual information processing is in the concentric center-surround receptive fields of the retinal ganglion cells. These spot contrast units are sensitive to the contrast and size of the stimulus and are specific for location of the stimulus. These can be of two types: cells that respond with the onset of stimulation (on-center); those that respond to the cessation of stimulation (off-center). An implementation of the on-center variety is described in Marr and Hildreth[43] and the information computed here forms the basis for the raw Primal Sketch of Marr.[44] However, the temporal aspects were not considered. It is probably true that the off-center cells play a significant role as well, and thus temporal information should be included in the computation of a primal sketch-like representation of information at this level. Some cells have an output that is steady over a relatively long period of time (X-cells) and some have a transient output (Y-cells). Models for these are found in Richter and Ullman.[45] An important characteristic of retinal processing is the spatial organization and extent of the receptive fields. In the fovea, the small central region of the retina, there is a one-to-one correspondence between photoreceptors and ganglion cells. This is the region with the highest visual acuity. This relationship changes as one moves towards the periphery of the visual field. Visual information processing is necessarily less precise the farther away the object of interest is from the fovea, since increasingly larger numbers of photoreceptors form the receptive fields of ganglions in the periphery. This distorted image mapping is preserved throughout the visual system, in that there is proportionately much more neural machinery devoted to processing the small central areas than the much larger peripheral areas.

Of note is the "channel hypothesis", that states that there exist several different spatio-temporal frequency channels (or band-pass filter mechanisms) that operate over the retina. Wilson and Bergen[46] contend that there are four channels, corresponding to four receptive field types at each retinal position—two sustained and two transient—and that this is not homogeneous across the visual field. That is, the same field types are not present for each retinal position, but vary in extent, magnitude of response or other properties. Four separate channels were modelled in Marr and Hildreth,[43] however, the non-homogeneity and the temporal aspects of the response were not included. Evidence for a possible fifth channel was presented in Marr *et al.*[47] There may indeed be a continuum of sizes of channels—the integration of information from these different sized operators is an open problem.

The next processing station is the lateral geniculate nucleus (LGN). It is composed of 6 layers that are inter-posed between each other in specific ways. This laminated structure allows for the later processing of disparity, crucial to depth perception. The cells here are also of the center-surround variety, but have larger receptive fields than those in the retina. This body not only receives input from the retina, but also from higher levels of the visual cortex. It is believed that this input acts as a negative feedback influence, much the same as can be found in the LINES system in Zucker.[48]

Moving up the hierarchy to the simple and complex cells of the primary visual cortex, cells begin to respond to line segments and boundaries. Simple cells receive input from the LGN. In some cases they provide input for the complex cells, but it is also true that simple and complex cells process information in parallel. This area projects onto one or more higher order areas. Simple cells, or linear contrast units, aggregate the LGN output into edges or lines of varying sizes and orientations. The receptive fields of these cells are composed of appropriately shaped inhibitory and excitatory zones. They are sensitive to the size and contrast of stimulus and are specific for the location and orientation of the stimulus. The directional derivative operator of Zucker[40] models such linear contrast units. Complex cells, on the other hand, are larger, are orientation critical, but are not position critical, i.e. they generalize position information.

Simple and complex cells provide input for the hypercomplex cells of the visual association areas. Hypercomplex cells respond to changes in contrast boundaries. One end of a bar, or both ends of a bar, must be within the excitatory zone of those cells. These units are sensitive to the size and contrast of a stimulus and are specific for stimulus orientation and the existence of a boundary perpendicular to the orientation of the stimulus.

Connections are not quite so simple as the above may imply and the story is far from over. An overview of competing theories can be found in Stone *et al.*[49] However, from the above, several conclusions can be drawn. The level of complexity of form processing may be accounted for by a hierarchical aggregation of lower level inputs, occurring in parallel for each location of the visual field, with appropriate contrast enhancement via lateral inhibition at each level. It is not only aggregation and lateral inhibition that come into play here. The stimuli that are important at each processing level also vary. At the retina and LGN, position is important. In the simple cells, position and orientation are used. In the complex cells, the axis of orientation is important but the position has been generalized. In hypercomplex cells, edges and corners are important. It may be true that other properties are generalized farther along in processing. Thus, the concept of specialization of information plays a role.

The representation implied by these several levels of abstraction, just for form information, leads to a richer description than even the representations put forward by Nishihara.[20] The work by Marr and his colleagues is in the right direction, and requires further elaboration.

II.2.2. *The abstraction of change: temporal feature aggregation.* In very much the same fashion as the form abstraction hierarchy, there exists a change abstraction hierarchy, although it appears as if this has not been as extensively studied. Change abstraction begins with the amacrine cells of the retina that detect changes in incident illumination. Their response tends to be transient and many cells respond both to the beginning and end of a stimulus. They have large receptive fields, resulting in a certain amount of spatial blurring, and some are of the center-surround variety and are sensitive to moving spots of light. Note that the spatial blurring effect necessitates the need for integration of form information with change information for precise motion–position computation. Ganglion cells of the retina that are change sensitive receive input from the amacrine cells preferentially, and are sensitive to spots of light or dark that move into or out of the center. Amacrine cells are chained together, thus implementing a directionally selective mechanism that provides input to the ganglion cells. These spot change units are sensitive to velocity and amount of contrast change, and are specific for direction of change.

At higher processing levels, other constraints are added in order to make the responses more and more specific. In the first layer of cortical motion processing, spot or bar change detectors are found that are specific for both velocity and direction. The next layer presents spot or bar correspondence units adding displacement to the specificity list. At this layer, therefore, the units are specific for velocity, direction and displacement. Compare this with the correspondence process of Ullman,[50] where only displacement is considered as the key factor in correspondence. In fact, it seems that in biological vision systems, a great deal of processing must be done before displacement is even considered. The hierarchy of change detecting units and their characteristics are described in Orban[51] and is more involved than the simplified view presented here.

A model of directionally selective motion cells is presented in Marr and Ullman.[52] Their model proposes the measurement of the time derivative of their Laplacian operator output at the zero-crossings. This constrains the local motion direction to within 180 degrees. The true velocity is then measured by combining these local constraints. Optic flow schemes on the other hand,[53] do not capture the totality of change information that the hierarchy just described does. Changes in brightness due to light source motion, moving shadows or motion of features produced by occlusion are not within the optic flow definition. Optic flow computations depend on image projections of objects in motion. On the other hand, since the visual system does not distinguish these different forms of change, it is reasonable to wonder where and how the distinction comes about.

There is much less known about higher order motion abstraction. It seems that the abstraction of motion concepts takes place in much the same fashion as for form concepts, with more complex information being coded the higher the processing level. Generalization of such abstracting structures into higher order units may prove useful for computer vision experimentation. For example, the motion frames of the ALVEN system,[23,39] are defined in this way, creating motion concept frames by specializing along motion semantic components, such as direction, velocity, etc., as well as aggregating more and more motion concepts into the specialized concepts. This scheme provides an interaction between form (object description) and motion systems.

II.2.3. *The abstraction of colour information.* Human colour vision theory was first proposed by Thomas Young in 1802. He suggested a trichromatic theory of colour vision based on the perception of three colour sensations: blue, green and red. His hypothesis was confirmed when the colour absorption characteristics of the cones of the retina were measured. Individual cones detect only one of three colours: blue, maximally absorbing at a wavelength of 445 nm, green, at 535 nm, or red at 570 nm. Note that the absorption spectra for these colours overlap. Colour vision requires that neurons compare these 3 inputs.

Colour experience depends on three semi-independent impressions: hue, saturation and brightness. Hue is the strongest effect and is a measure of the proportion of the three cone mechanisms activated. About 200 varieties can be defined. Saturation reflects how much a hue has been diluted by gray and is determined by the degree to which all three cone mechanisms are stimulated by object and background. There are 20 steps for each hue. Brightness, finally, is a measure of the total effect on the cones of an object relative to its background. About 500 gradations are possible for each hue–saturation combination.

Neural mechanisms are hierarchical for colour vision and are separate from the other abstraction hierarchies. The cells described for form aggregation used a spatially opponent mechanisms. Two sets of cones (or rods) containing the same visual pigment are connected to a single cell in a spatially separate center-surround manner, so that contrast between the regions is the determinant of cell response. Colour cells have a similar opponent mechanism. However, colour cells receive information from two types of cones; one wavelength excites the cell, the other inhibits it. Ganglion cells of the retina and of the LGN have a concentric center-surround structure, with the center being connected to one type of cone and the surround to the other. Work by Rubin and Richards(54) presents a similarly structured operator for detecting spectral crosspoints across edges as material changes. There are red–green cells (red exciting, green inhibiting), green–red cells and blue–yellow cells (inhibition by both red and green leads to optimum inhibition by yellow).

Higher order processing in the primary visual cortex involves a different sort of concentric receptive field. These are colour contrast cells. The center has as input a red–green system while the surround has an opponent colour system, for example. The best stimulation for such a cell would be red in the center and green in the surround. In further processing, simple cells respond to bars and rectangles of one colour opponent system with flanking regions of reverse opponent colour system, providing input to complex and hypercomplex cells in which orientation is important but exact position is less critical. Rather little has been done on computational models of such processing.

II.2.4. *The abstraction of depth information.* Stereopsis allows us to perceive objects in spatial depth and occurs when the same object stimulates appropriate retinal regions in each eye. There are at least two separate aspects to stereopsis: the information available from binocular parallax; the information available from the fusion of two images into one. This is the vergence effect or the degree of accommodation that must be achieved in order to bring an object into focus when shifting the eyes to objects at different depths. Different single cells receive input from both right and left retinal areas that are exactly in register or from those that are disparate or contain binocular parallax information. The layered construction of the LGN allows such a mixing of left and right signals. This is the level of processing that the theory of steropsis of Marr and Poggio(55,56) addresses. When fixation shifts to a different depth in space, all the corresponding retinal and cortical points must be re-arranged. The same retinal ganglion cells are used to detect all form patterns, but different groups of cortical cells are used to detect the same form pattern at different depths in space. As an object moves away, the eyes diverge to maintain binocular fixation, new disparity cells are activated and the amount of divergence, which is controlled by higher processes, is used for interpretation of displacement in depth. Disparity detecting units are therefore sensitive to the form of the stimulus, contrast and size, are specific for stimulus orientation and disparity and are dependent on the point of fixation.

In addition to the binocular form units, there are also stereoscopic motion detecting units, described by Regan *et al.*(57) These units are based on the change abstraction hierarchy described earlier in that their input consists of corresponded spots and bars from left and right registered images. Because of this, they can not be used for the representation of the exact position in space of the moving object and, thus, somewhere in further processing there must be interaction with the form hierarchy and disparity output. These stereo motion detectors are specific for velocity ratio and for direction of change. They are independent of actual velocity and actual depth.

The computation of three dimensional models has received much attention from a computer vision point of view. Examples of such models include the scheme of Nagel and Neumann(58) that incorporates a version of photometric stereo, while Grimson(59) incorporates the stereopsis theory of Marr and Poggio.(55)

II.3. *The visual percepts*

Perception is characterized in this discussion by the requirement that several sensations must interact in the creation of a percept. I hypothesize (and do not develop further herein) that the computing units involved in perception are organized in a network, connected to computing units of the sensations, at varying levels of abstraction, and that the information is integrated via an aggregation mechanism. Perceptual units can be organized using the is-a relationship as well. These units monitor the output of units in the sensation hierarchies and may guide their processing when appropriate. Connections between perceptual and sensation units implicitly require a translation mechanism between image-specific information and object-specific information. The "projection" representational axis is relevant here. The interactions between sensations and percepts form the focus of the following discussion.

II.3.1. *Monocular depth cues.* In addition to stereopsis, a set of different properties of the visual system also contribute to three-dimensional perception. According to Wyburn *et al.*,(60) these are:

Apparent relative size. A comparison of the relative sizes of various objects in the visual field with our knowledge of their actual sizes.

Occlusion. If an object of known distance occludes another, then it provides a strong constraint on that object's distance from the observer. Also. much information can be gained from the form of shadows due to occlusion and from the shape of an object's contour that occludes the remainder of the scene behind it. The interaction of shape from shading and occluding contours has been examined in Ikeuchi,[61] and Marr[62] defines methods for extracting axes of generalized cone representations of smooth surface objects given their occluding contours.

Relative appearance. Objects that are close tend to have sharper outlines and more clearly defined detail than objects that are farther away.

Colour. In natural imagery, objects that have cold, less saturated colours (blue, green) seem to be farther away than those with warmer colours (red, orange, brown).

Shadowing cues and intensity graduation. This includes self-shadowing, considerations of surface reflectance and highlights, as considered by Horn[63] and Woodham,[64] and shadows cast on other objects.

Linear perspective. Parallel lines receding into the distance appear to converge.

Motion parallax. If an observer moves, near objects appear to move in the opposite direction while those in the background move in the same direction. Two objects moving at the same speed but at different depths appear to be moving at different speeds, the farther one seeming to move slower.

It is not clear how the interactions of each of these principles contribute to three-dimensional perception.

II.3.2. *Form, colour and change interactions.* There have been many good surveys of motion analysis systems—see Nagel[2] for a complete account of this topic, as well as Snyder,[65] Aggarwal and Badler[66] and Ullman.[36] This section will concentrate on the interactions of the motion perceptual system with the other sensation systems.

An important factor in motion interpretation by humans is that in the face of ambiguity, the visual system prefers to achieve as high a degree of object constancy or rigidity as possible, according to Johansson.[67] The rigidity assumption was key in the structure from motion work described in Ullman.[50] Two examples of such ambiguity are: uniform change in size vs motion in depth, and change in size in one dimension vs rotation in depth.

The change aggregation hierarchy described earlier simply cannot handle such interpretations. It was noted earlier that even at the first level of change information detection, a certain amount of spatial blurring had occurred, thus implying that in order to compute accurate motion–position information there must be some interaction between the change and form abstraction mechanisms. In addition, interaction between form and motion information is necessitated because of the motion aperture problem (Marr and Ullman[52]). If the motion of an oriented element is detected by a unit that is small compared to the size of the moving element, the only information that can be obtained is the component of motion perpendicular to the local orientation of the element. This means that for higher order interpretations, there is a need for a motion combination stage, perhaps guided by short range motion correspondence information Ullman.[36]

This motion combination stage is the component that interacts with the depth, form and colour systems. The experiments described by Kolers and his colleagues reveal some of these interactions. Their main result is that the coding of form, depth, colour and motion information cannot be completed in all cases independently of each other. These four systems constantly monitor each others results and compute a consistent interpretation in parallel.

The interactions between form changes and rigid rotation in depth were investigated in Kolers and Pomerantz.[68] They discovered that if the visual system is given enough time, a rigid interpretation is the preferred interpretation. In contrast, if the amount of time allotted to interpretation is decreased below a certain point, the interpretation will involve a form change, thus implying that computations involving rotations are more demanding computationally. If more than one kind of change is presented, there is no difference in processing efficiency, implying that computations are performed in parallel between these two systems. Also, in apparent motion with collision, the preferred interpretation is motion in depth to avoid collision. This implies that the spatial characteristics of the traversed path are being monitored and have an effect on the interpretation. In Kolers and von Grunau,[69] results were obtained providing evidence for the interactions between form and motion and form and colour systems.

On the other hand, work in Hay[70] relates image characteristics to object motions in 3D. This includes image-specific changes, such as translation, stretch, shear, foreshortening and magnification, related to object-specific changes, such as motion in depth, rotations and translations in all directions. These results provide a starting point for an important aspect of expert vision systems, that of translating between image specific and object specific characteristics—the projection between the two domains. It is a crucial component of the model proposed by Kanade[7] for vision system design.

The interactions between form and motion fields are explored in the work of Hoffman[71] and those between depth and optical flow for rigid body motion derivation in Ballard and Kimball.[72] Finally, the positive influence on interpretation of motion by expectations is discussed in Sekuler and Levinson.[73]

II.3.3. *Generation and use of expectations.* The use of expectations or predictions as aids in the processing of visual information, as mentioned previously, is not a new idea. In this discussion, the purpose of expectations will be to direct the attention of the system to particular sets of concepts and events. Expectations were used in the SEER system of Freuder[74] to guide region growing and identification of specific portions of a hammer. A thorough understanding of human body motions and a model of the allowed joint configurations enabled the design of a constraint propagation network that integrated current motions and known body positions with hypothesized ones, producing expected locations in 3D for given body joints (O'Rourke and Badler[75]). The work presented in Browse[76] addressed the problem of how a computer system can use the non-uniformity of retinal processing detail in conjunction with a knowledge base of the domain in order to generate new fixation points. Finally, Down[77] explored the generation of expectations from a knowledge base of motion concepts in a manner relating work described earlier[70] to motion understanding. However, much research is still needed.

Evidence from psychophysical experimentation for both the use of generalization as a concept organizational tool and for the use of expectations and their relationship to the generalization relationship comes from the following. The experiments described in Cooper and Shepard[78] show the strong positive effect of *a priori* expectations on time for interpretation, while those of Bugelski and Alampay[79] and Palmer[80] show the effects of generalization of expectation classes. Cooper and Shepard reported that in the identification of letters presented at varying orientations, the time taken to identify the letter varied with the amount of rotation (to a maximum value at 180 degrees), implying that mental rotation and matching was being performed by the visual system and that if identity and orientation were given previous to the stimulus, the response time was flat across all orientations, as long as sufficient time was allowed before the stimulus was presented for expectation formation.

Bugelski and Alampay showed that if a subject is conditioned to expect a given category (or generalization) of stimulus, then the identification time of the stimulus is reduced. They presented stimuli all belonging to the same class of concept (animals) and when non-animal stimuli were presented, the response time increased. This was further examined by Palmer, who also noted the impairment of identification if the context is mis-leading. It should be pointed out here that the mechanisms that produce such behaviour are not understood.

At this point two separate notions must be distinguished, namely those of image-specific focus of attention and semantic focus of attention. Following James,[26] attention is defined as: 'It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others". The process of visual attention, the selection of important objects in the visual world, has a response, that can take many forms—reaching for it or visual fixation are but two. The response of changes of image attention is specific eye movement, while the response of changes of semantic attention does not necessarily involve eye movement. Considerations as to external goals or motion of the head, etc., are beyond the scope of this discussion.

Changing system attention and determining the focus of attention are important components of expert systems, as mechanisms for reducing computational load and allowing the system to behave in a more "intelligent" manner. In medical diagnosis expert systems, such as those described in Szolovits and Pauker,[13] hypotheses are generated from the data. Signs and symptoms are input to the system and the appropriate hypotheses are activated. Data is continually being added as a result of questions asked by the system, so that new hypotheses are introduced throughout the diagnostic process. The eye movements that must be performed in order to "see" all of the scene can be likened to the need for a medical diagnosis system to ask questions, because it is only given a subset of all the data it requires. The system then directs the information gathering so that a diagnosis or interpretation of the data can be achieved. This should likewise be an important component of vision systems—however, there has been very little work on the topic.

With respect to semantic attention, Cornsweet[30] defines semantic attentional sets as biases on nodes in semantic memory that are activated by appropriate input regardless of position. That is, position in space is generalized. This definition seems to apply to "higher order" concepts only, and in order for attention to be useful in practical systems, a definition encompassing biases at all levels of processing must be formulated.

A generalized definition of attentional sets is: the attentional set for a given computing unit of the visual information processing system is the set of all other active input signals from computing units that provide efferent or top-down (stimuli that travel away from a processing center and towards a sensory unit) input to that computing unit. This input is viewed as a bias placed on the response of that unit to afferent or bottom-up stimuli (stimuli that travel towards a processing center and away from sensory units) and may take the form of excitatory or inhibitory biases of varying degrees. These may also be termed priming signals. Thus, each computing unit at any level of processing that has top-down connections has its own attentional set. Note the distinction between expectations and attentional sets. Expectations exist only at the perceptual level amongst concepts at similar levels of detail, while attentional sets identify those computing units at either the sensation or perceptual levels that prime less abstract units.

Although the use of expectations and attentional sets within computer vision has been discussed previously, no clear mechanism exists. It is clear, however, that a sufficiently rich, hierarchical form of sensation information is required for the expectations to be useful. In the systems that employ processing cones, VISIONS[9] for example, vertical communication is crucial, yet its mechanism is unclear. It may be true that too much is abstracted in the early levels for the effective use of a focus of attention.

II.4. *Discussion*

It should be clear, even from this abbreviated presentation, that no existing vision system can cope with all of the visual knowledge presented above. Moreover, not all topics that are relevant have been touched upon. The effect of language on perception, object identification and the forms of visual memory are but examples of topics that have been omitted. Those who believe that representation schemes are not necessary in vision, because they only have relevance for application domains, would be hard pressed to include all of the above in an integrated, uniform manner in a single system without a representation scheme. I submit that it is crucial that representation of knowledge research be influenced by vision research in that current formalisms are lacking in their ability to handle many of the concepts presented above. Briefly summarizing the important aspects of visual knowledge that were discussed, the visual system incorporates at least the following:

concept aggregation or grouping for form, change, colour and depth information;
concept generalization and specialization for organization of different types of concepts;
lateral inhibition for enhancement of contrast or difference among concepts in a given processing layer that are obtained via grouping processes;
attentional sets and expectations that guide processing;
perception as an integration of and interaction among sensations.

III. REPRESENTATIONAL ISSUES

A large amount of information has been compiled as a result of a survey of research groups involved in representation of knowledge.[15] Many of the concepts discussed in that survey are also applicable to visual information representation. However, I will also point out directions for further research motivated by some unsolved problems in visual information processing. I will begin with a brief excursion into some common representational schemes.

III.1. *Common representational schemes*

There are three major types of representational schemes according to Mylopoulos[81]: logical, network and procedural. Logical schemes define a knowledge base as a collection of logical formulae (first order predicate calculus, for example), which represent a partial description of the state of the world. Modifications to the knowledge base are accomplished via the addition or deletion of formulae, so that the formula is the atomic unit of KB manipulation. An example of such a scheme is that of Kowalski.[82] Of interest here is also the work on incompleteness in knowledge bases (Levesque[83]). These formalisms provide well understood semantics, simplicity of notation, theorem proving as the only inference mechanism and conceptual economy. On the other hand, they lack any sort of organizational principles and it is difficult to represent procedural or heuristic knowledge in such a notation.

Network schemes model the world using objects and relations for constructing a knowledge base. Changes are done with addition and deletion operations on objects and by manipulating relationships, as described in Quillian.[84] Historically, semantic networks have favoured binary relationships. They have obvious graphical representations using labelled nodes for objects and labelled arcs for relationships among objects. The proliferation of relation types, however, has been criticized by Woods[85] and Schubert.[88] Organizational principles such as class-token (instance-of), aggregation (part-of) and generalization (is-a), are in wide use, in addition to contexts and partitions (Hendrix[87]). Retrieval issues are addressed directly. Many semantic network schemes, however, lack a formal semantics and the area suffers from a lack of standard terminology.

Procedural schemes allow specification of direct interactions between entities at the expense of ease of understanding and modification. Examples of these are production systems, characterized by rule-action units,[88] the actor systems of Hewitt *et al.*[89] and the theorems or demons of Hewitt.[90] It should be noted that the ACRONYM vision system[21] utilizes a rule-based reasoning component.

Knowledge representation formalisms, such as FRL,[91] KRL,[92] OWL,[93] KLONE[16] and PSN,[17] are examples of frame-based schemes—representations that combine ideas from semantic networks, procedural schemes and other sources. According to Minsky's original proposal,[19] a frame is a complex data structure that represents some prototypical situation or object. A frame has slots that play a role in the definition of the prototype as well as relations between slots. These relations are constraints on possible slot values and their relationships to other slot values. In addition, each frame has attached to it information on how it is to be used, default values for its slots, and what to do if something unexpected happens. Exceptions and exception-handling schemes have been proposed to handle such occurrances and are described in Lesperance.[94] The idea of similarity between frames is of importance as a form of organization over frames. The information associated with a frame on its use is usually represented procedurally, namely via procedural attachment, and at least specifies how to add instances of frames to the knowledge base, how to test if a given object is an instance, how to retrieve all the instances of a given frame and how to delete an instance.(17) Frames can be organized using the is-a and part-of relationships, providing that the semantics of these relationships are carefully defined and enforced by the representational system.(16)

III.2. *Representing a concept*

Following Brachman and Smith,(15) a brief overview of representational issues is presented. A formalism for representing concepts must be expressive enough to handle in a semantically well defined way at least the following.

Prototypical concepts and instances. A prototype provides a generalized definition of the components, attributes and relationships that must be confirmed of a particular concept under consideration in order to be able to make the deduction that the particular concept is an instance of the prototypical concept. The prototype for a car, for example, would be a complex structure spanning many levels of description in order to adequately capture the structuring of discrete objects into more complex ones, define spatial and functional relationships for each object and assert constraints that must be satisfied in order for a particular object in a scene to be identified as a car. It also involves the differentiation between sensation and perceptual concepts and their relationship.

Discrete and structured concepts. Concept structure can be represented using slots in a prototype definition. The slots form an implicit part-of relationship with the concept. The ALVEN system(39) uses frames with this characteristic for the definition of motion concepts. The leaves of the part-of hierarchy are discrete concepts. Structured concepts representing physical objects were represented in the human body models of O'Rourke and Badler(75) using generalized spheres, while Marr and Vaina(95) used generalized cylinders for the same purpose.

Ordinality and measure. In the prototype of a car above, one must include the facts that there are 4 wheels, 2 in the front, 2 in the back, and what the sizes of the wheels are in relation to the remainder of the car. Prototypical car motions must include constraints on velocity, acceleration and direction of travel. Ordinality refers to quantities of parts, while measure refers to concepts that require units for proper representation.

Properties, qualities, attributes. All physical objects have some physical attributes such as colour, size, shape, surface texture, etc., and more generally, all concepts are defined partially in terms of their properties. These are assertions about the nature of the concept and may be considered as constraints that aid in the discrimination of one concept from another.

Quantification and its scope. It is often valuable to be able to represent existential and universal quantifiers. For example, the creation of a prototype is a universal statement that all instances of the prototype have the attributes defined in the definition of the prototype. A definition of the structure of some concept is an existential statement for the concept being considered, stating that each component must exist for that concept.

Causality. Causal relationships are rather difficult to deal with, yet in many cases it is causality which can disambiguate between two interpretations. A causal relationship has both existential and temporal implications. It implies that some future condition must be present given the current state. Currently, causal relationships are not used in vision systems. Causal relationships were studied in Reiger and Grinsberg(96) and an application of those relationships, as well as others, for signal analysis is described in Shibahara *et al.*(97)

Spatial knowledge. This is perhaps the main type of knowledge that most vision systems employ. This includes spatial relationships (above, between, to the left of), shape information (curvature, shape type such as cone-like), location in space, and continuity constraints, again with the sensation/perception distinction. Comparison of object location and the representation of the corresponding relations is considered in Freeman.(98) Kuipers(99) describes his TOUR model for route solving problems and discusses the spatial knowledge relevant to that task. The VISIONS system employs a representation of 3D complex surfaces and 2D curves based on B-splines and surface patches and also makes use of the part-of and instance-of relationships in building complex structures.(100)

Temporal knowledge. Information about temporal constraints. Time provides a context in which events can be interpreted. Allen(101) describes an interval based representation along with an inferencing scheme for a wide variety of temporal constraints. Tsotsos(102) employs a point based representation that is used for representing the temporal constraints in a motion-understanding system, where recognition is driven by the inertia of good temporal continuity of motion concepts.

States, events, actions, change. Motions of all types require representation for a vision system that deals with motion. A representation formalism for motion and event concepts is presented in Tsotsos.(102) Representation of states is an important aspect of the representation of causality and was considered in Reiger and Grinsberg.(96) A framework for research in action perception was presented in Sridharan *et al.*(103)

Procedural knowledge. The "how-to" knowledge of a system. The PSN(17) representation scheme handles such knowledge in two ways. Programs can be included in class definitions, as well as four standard programs that define the semantics of the test, instantiate, delete and generate instance operations on classes. Procedural attachment is also important in the representation of knowledge in the system of

Ballard *et al.*,[104] as well as many other representational schemes, such as KRL.[92] The packaging of definitional knowledge and procedural knowledge together allows for the construction of computing units that contain both the knowledge that describes a given concept as well as the process that extracts that concept from the data.

Situations and contexts. Hypotheses that are used during an interpretation exist only in a given context. They are constructed from prototypes that are anchored in some way to the data currently being considered. The contexts and partitions of Hendrix[87] are relevant here.

Description by comparison and differentiation. Similarity measures that can be used to assist in the determination of other relevant hypotheses on hypothesis matching failure are useful in the control of growth of the hypothesis space. These measures usually relate mutually exclusive categories. Similarity links are components of the frame scheme of Minsky[19] and a realization of similarity links as an exception-handling mechanism is presented by Tsotsos,[102] based on a representation of the common and differing portions between two frames.

Conjunction, disjunction, negation. Most systems represent constraints as Boolean conditions and clearly these three concepts are required. In addition, co-existence of components, the fact that existence of one component implies non-existence of another, etc., must be handled.

Inheritance, instantiation and reasoning with defaults. Inheritance along the is-a and instance-of axes is not at all straightforward. Discussions of the topic are presented by Brachman[18] and Woods,[85] while the definition of the PSN language[17] provides a set of rules for inheritance along these two different axes. Default reasoning is discussed by Reiter.[105]

Certainty, strength of belief. One mechanism for ranking hypotheses is to attach certainty measures to each. Many schemes exist for this, ranging from Bayesian techniques as described by Duda *et al.*[106] to relaxation schemes,[107] to more domain dependent approaches such as for medical diagnosis.[108] In vision, the lateral inhibition process is a major component and can be viewed as a process that decreases the certainty of concepts that are inconsistent in local neighbourhoods. This can be modelled using relaxation processes.

Expectations. These are beliefs that are held as to what exists in the context of the scene, beliefs in time of future events, beliefs in space of related objects. The concept of plan-directed vision appeared in the work of Kelly[109] and has appeared in many other systems since then. Line-finding using a hypothesize-and-test scheme that utilized expectations derived from previous results is described by Shirai.[110] Temporal expectations were employed by O'Rourke and Badler[75] as well as by Tsotsos *et al.*[39] The issue of sensation to perception translation and vice versa and the use of expectations at the perceptual levels, creating priming signals to sensation computing units, has not been adequately studied.

III.3. *Concept organization*

Several means of concept organization were discussed earlier. These organizational schemes really define a great deal of the control structure of the vision system, for they provide axes along which inferencing takes place and along which concept activation occurs. Common organizational methods, from Brachman and Smith,[15] are as follows.

Plurality (sets, sequences, membership, partial orders). Common definitions of regions are constructed by referring to a set of pixels that is maximal and that satisfies some predicate and for which all elements of the set are connected. A "walking" motion, for example, can be defined as a sequence of simpler motions.[75] The is-a and part-of hierarchies form partial orderings of concepts, while the instance-of relationship implies membership.

Projection. This relationship forms one of the important links in the model of Kanade.[7] It is a transformational link relating representations of the same concept in differing domains. It is, for example, the relationship between a prototypical object and its actual appearance in an image, given lighting conditions and viewpoint. Thus, a mechanism is required that takes lighting, observer motion, temporal continuity and viewpoint into account to create an internal representation of an object's appearance in an image. Some work on this has been done by Hay[70] and is further explored by Down.[77] In Down's work information contained in an is-a hierarchy of motion concepts is exploited for the generation, verification and modification of expectations of actual object appearance in a sequence of images.

Abstraction. Two kinds of abstraction were discussed previously, namely, feature aggregation and concept specialization. The part-of hierarchy can be traversed bottom-up in aggregation mode or top-down in decomposition mode. Top-down traversal implies existence of components and thus constrains the lower level computing units. Bottom-up traversal implies a form of hypothesize-and-test, where computing units activate other computing units that may have them as components. Top-down traversal of an is-a hierarchy, moving downward when concepts are verified, implies a constrained form of hypothesize-and-test for more specialized concepts. In this case the constraints differ from the part-of traversal—an is-a parent implies that perhaps one of its is-a siblings applies, while the confirmation of an is-a sibling implies that its parents must also be true. Issues of inheritance and exceptions, as mentioned earlier, must also be addressed. Both is-a and part-of are used in the MAPSEE2 system of Mackworth and Havens,[22] as well as in the ALVEN system[102] and in the VISIONS system.[9]

Multiple viewpoints. In vision, the concept of multiple views has an obvious relevance. The problem, however, is how is the information obtained from each view correlated with the information from all the other views in order to derive a consistent interpretation. Nagel and Neumann[58] construct 3D models from two perspective views. The structure from motion theorem of Ullman[50] required three orthographic views. However, physical viewpoint is only one of the aspects that are relevant here. Different views can result from employing differing beliefs about the context of the scene. Belief is treated from an AI ɩoint of view by Moore.[111]

Meta-knowledge. Self-knowledge, so that the system can examine itself, know its limitations, can chose from amongst different strategies. The MYCIN system[108] employed a set of rules that implemented meta-knowledge and guided the system by specifying what rules to try given the problem at hand or the context, as well as other forms of guidance. The system of Ballard *et al.*[104] employed a scheme for choosing from among a set of procedures to determine which would prove most useful. Explicit representation of meta-knowledge can be accomplished using the PSN version of meta-classes.[17] Classes are related to meta-classes by the instance-of relationship and the meta-classes contain information that is derived by considering all its instance classes as a whole. Another important aspect is the handling of knowledge gaps or incompleteness (Levesque[83]).

Control issues are difficult to deal with and no representation scheme exists that can explicitly represent a variety of control schemes. An overview of a variety of control schemes can be found in Weszka.[5] The inclusion of is-a, part-of and similarity relationships allows a great deal of flexibility for top-down, bottom-up and lateral search, hypothesize-and-test and ranking of hypotheses via relaxation processes. Connections between sensations and perceptions are "projections" and mechanisms for the translation from one to the other are unclear. It should be noted that there does not currently exist an appropriate well-founded representation that makes the distinction between parallel and sequential processing, although steps in that direction are being explored by Fahlman,[112] Sabbah,[113] Hinton[114] and Feldman and Ballard.[115] Also, communication among concepts has not been examined extensively.

CONCLUSIONS

An abbreviated overview of visual knowledge was presented, with the distinctions drawn among content, form and use. Content was determined using biological visual systems as examples and several basic principles were shown to play a role in the many aspects of vision. A strong correspondence between knowledge organization and knowledge use was demonstrated. The key principles are those of prototypes as the basic representational building block, generalization and aggregation as interacting abstraction mechanisms, lateral inhibition as an optimization scheme (particularly useful for local disambiguation), generation and use of expectations at the perceptual level, the use of attentional sets and priming as a form of communication of expectations from perceptual to sensation levels, and the existence of at least four separate sensation processing hierarchies (form, motion, colour and depth). These four processing hierarchies must interact and several examples of interactions were described. Finally, the need for both parallel and sequential processing was demonstrated. Current representational research can provide well-defined tools for only some of these concepts.

Acknowledgements—I am grateful for the helpful comments on earlier drafts of this paper from STEVE ZUCKER, JOHN MYLOPOULOS, HECTOR LEVESQUE and PAUL KOLERS. The preparation of this paper was supported in part by the Defense Research Establishment Atlantic under contract 8SC81-00316 and by the Natural Science and Engineering Research Council of Canada.

REFERENCES

1. R. Nevatia, Characterization and requirements of computer vision systems, *Computer Vision Systems*, Hanson and Riseman, eds., p. 81. Academic Press (1978).
2. H. Nagel, Image sequence analysis: what can we learn from applications? *Image Sequence Analysis*, Huang, ed. p. 19. Springer-Verlag (1981).
3. A. Hanson, E. Riseman, Defining the field of computer vision, *Computer Vision Systems*, Hanson and Riseman, eds., p. 15. Academic Press (1978).
4. H. Barrow, Computational vision, *Proceedings Joint IBM/University of Newcastle upon Tyne Seminar on Artificial Intelligence*, p. 1 (1980).
5. J. Weszka, (ed.) *Workshop on Control Structure and Knowledge Representation for Image and Speech Understanding*, University of Maryland (1979).
6. R. Reddy, Pragmatic aspects of machine vision, *Computer Vision Systems*, Hanson and Riseman, eds., p. 89. Academic Press (1978).
7. T. Kanade, Survey: region segmentation: signal vs semantics, *Comput. Graphics Image Process.* **13**, 279 (1980).
8. A. Newell, The knowledge level, *Artif. Intell.* **18**, 87 (1982).
9. A. Hanson, E. Riseman, VISIONS: a computer system for interpreting scenes, *Computer Vision Systems*, Hanson and Riseman, eds., p. 303. Academic Press (1978).
10. J. Tenenbaum and H. Barrow, Experiments in interpretation guided segmentation, *Artif. Intell.* **8**, 241 (1977).
11. M. Levine, A knowledge-based computer vision system, *Computer Vision Systems*. Hanson and Riseman, eds. p. 335. Academic Press (1978).
12. S. Rubin, Natural scene recognition using LOCUS search, *Comput. Graphics Image Process.* **13**, 298 (1980).
13. P. Szolovits, S. Pauker, Categorical and probabilistic reasoning in medical diagnosis, *Artif. Intell.* **11**, 115 (1978).
14. A. K. Mackworth, Vision research strategy: black magic, metaphors, mechanisms, miniworlds, and maps, *Computer Vision Systems*, Hanson and Riseman, eds., p. 53. Academic Press (1978).
15. R. J. Brachman and B. C. Smith, (eds.) Special issue on knowledge representation, *SIGART Newslett.* **70** (1980).
16. R. J. Brachman, On the epistemological status of semantic networks, *Associative Networks*, Findler, ed., p. 3. Academic Press (1979).

17. H. Levesque and J. Mylopoulos, A procedural semantics for semantic networks, *Associative Networks*, Findler, ed., p. 93. Academic Press (1979).
18. R. Brachman, What 'ISA' is and isn't, *Proceedings Canadian Society for Computational Studies of Intelligence*, p. 212. Saskatoon (1982).
19. M. Minsky, A framework for representing knowledge, *The Psychology of Computer Vision*, Winston, ed., p. 211. McGraw-Hill (1975).
20. H. K. Nishihara, Intensity, visible surface and volumetric representations, *Artif. Intell.* **17**, 265 (1981).
21. R. Brooks, R. Cereiner and T. Binford, The ACRONYM model-based vision system, *Proceedings International Joint Conference on Artificial Intelligence*, Tokyo, p. 105 (1979).
22. A. K. Mackworth and W. S. Havens, Structuring domain knowledge for visual perception, *Proceedings International Joint Conference on Artificial Intelligence*, Vancouver, p. 625 (1981).
23. J. K. Tsotsos, Temporal event recognition: an application to left ventricular performance assessment, *Proceedings International Joint Conference on Artificial Intelligence*, Vancouver, p. 900 (1981).
24. S. W. Zucker, Computer vision and human perception—An essay on the discovery of constraints, *Proceedings International Joint Conference on Artificial Intelligence*, Vancouver, p. 1102 (1981).
25. S. Zucker, R. Hummel and A. Rosenfeld, An application of relaxation labelling to line and curve enhancement, *IEEE Trans. Comput.* **26**, 394 (1977).
26. R. Hummel and S. Zucker, On the foundations of relaxation labelling processes, TR-80-7, Dept. of Electrical Engineering, McGill University (1980).
27. W. James, *The Principles of Psychology* (1890). Republished in *Great Books of the Western World*, Encyclopaedia Britannica (1971).
28. E. Kandel and J. Schwartz (eds.), *Principles of Neural Science*. Elsevier/North-Holland, New York (1981).
29. H. Davson, (ed.), *The Eye: Visual Function in Man*, Vol. 2A. Academic Press, New York (1976).
30. T. Cornsweet, *Visual Perception*. Academic Press, New York (1976).
31. T. Caelli, *Visual Perception: Theory and Practice*. Pergamon Press (1981).
32. W. Wickelgren, *Cognitive Psychology*. Prentice-Hall (1979).
33. S. Zucker, A. Rosenfeld and L. Davis, General purpose models: expectations about the unexpected, *Proceedings International Joint Conference on Artificial Intelligence*, Tiblisi, USSR, p. 716 (1975).
34. B. Bridgeman, Distributed sensory coding applied to simultaneous iconic storage and metacontrast, *Bull. math. Biol.* **40**, 605 (1978).
35. M. Wertheimer, Untersuchung zur Lehre von der Gestalt, *Psychol. Forsch.* **4**, 301 (1923).
36. S. Ullman, Analysis of visual motion by biological and computer systems, *IEEE Trans. Comput.* **14**, 57 (1981).
37. B. Flinchbaugh and N. Chandrasekaran, A theory for spatio-temporal aggregation, *Artif. Intell.* **17**, 387 (1981).
38. K. Prazdny, Perceptual segregation in apparent motion of random dot patterns, TR-985, Computer Science Center, University of Maryland (1980).
39. J. K. Tsotsos, J. Mylopoulos, H. Covvey and S. Zucker, A framework for visual motion understanding, *IEEE Pattern Anal. Mach. Intell.* **PAMI-2**, 563 (1980).
40. S. Zucker, Early orientation selection and grouping: type I and type II processes, Computer Vision and Graphics Laboratory TR82-6, McGill University (1982).
41. H. Barrow and J. Tenenbaum, Recovering intrinsic scene characteristics from images, *Computer Vision Systems*, Hanson and Riseman, eds., p. 3. Academic Press (1978).
42. D. Hubel and T. Wiesel, Brain mechanisms of vision, *Scient. Am.* **241**, 150 (1979).
43. D. Marr and E. Hildreth, A theory of edge detection, *Proc. R. Soc.* **B207**, 187 (1980).
44. D. Marr, Early processing of visual information, *Phil. Trans. R. Soc.* **B275**, 483 (1976).
45. J. Richter and S. Ullman, A model for the spatio-temporal organization of X and Y type ganglion cells in the primate retina, MIT AI LAB memo 573 (1980).
46. H. Wilson and J. Bergen, A four mechanism model for spatial vision, *Vision Res.* **19**, (1979).
47. D. Marr, E. Hildreth and T. Poggio, Evidence for a fifth, smaller channel in early human vision, MIT AI LAB, memo 541 (1979).
48. S. W. Zucker, Vertical and horizontal processes in low level vision, *Computer Vision Systems*, Hanson and Riseman, eds., p. 187. Academic Press (1978).
49. J. Stone, B. Dreher and A. Leventhal, Hierarchical and parallel mechanisms in the organization of the visual cortex, *Brain Res. Rev.* **1**, 345 (1979).
50. S. Ullman, *The Interpretation of Visual Motion*. MIT Press (1979).
51. G. A. Orban, Visual cortical mechanisms of movement perception, Ph. D. Thesis, Dept. of Brain and Behaviour Research, Katholieke Universiteit Leuven, Belgium (1975).
52. D. Marr and S. Ullman, Directional selectivity and its use in early visual processing, MIT AI LAB, memo 524 (1979).
53. D. Lawton, Optic flow field structure and processing image motion, *Proceedings International Joint Conference on Artificial Intelligence*, Vancouver, p. 700 (1981).
54. J. Rubin and W. Richards, Colour vision and image intensities: when are changes material? MIT AI LAB memo 631 (1981).
55. D. Marr and T. Poggio, A theory of human stereo vision, MIT AI Laboratory memo 451 (1977).
56. D. Marr and T. Poggio, Some comments on a recent theory of stereopsis, MIT AI LAB memo 558 (1980).
57. D. Regan, K. Beverley and M. Cynader, The visual perception of motion in depth, *Scient. Am.* **241**, 7, 136 (1979).
58. H. Nagel and B. Neumann, On 3D reconstruction from two perspective views, *Proceedings International Joint Conference on Artificial Intelligence*, Vancouver, p. 661 (1981).
59. W. Grimson, *From Images to Surfaces*. MIT Press (1981).
60. C. Wyburn, R. Pickford and R. Hirst, *Human Senses and Perception*. University of Toronto Press (1964).
61. K. Ikeuchi, Numerical shape from shading and occluding contours in a single view, MIT AI LAB memo 566 (1980).
62. D. Marr, Analysis of occluding contour, MIT AI LAB memo 372 (1976).
63. B. Horn, Obtaining shape from shading information, *The Psychology of Computer Vision*, Winston, ed., p. 115. McGraw-Hill (1975).
64. R. Woodham, Reflectance map techniques for analyzing surface defects in metal castings, MIT AI LAB TR 457 (1978).
65. W. Snyder, (ed.), Computer analysis of time-varying imagery, *IEEE Trans. Comput.* **C-14**, 7–69 (1981).
66. J. Aggarwal and N. Badler, (eds.), Special issue on motion and time-varying imagery, *IEEE Pattern Anal Mach. Intell.* **PAMI-2**, 493–587 (1980).
67. G. Johansson, Visual motion perception, *Scient. Am.* **232**, 6, 76 (1975).
68. P. Kolers and J. Pomerantz, Figural change in apparent

motion, *J. exp. Psychol.* **87,** 99 (1971).

69. P. Kolers and M. von Grunau, Shape and colour in apparent motion, *Vision Res.* **16,** 329 (1976).
70. J. Hay, Optical motions and space perception: an extension of Gibson's analysis, *Psychol. Rev.* **73,** 550 (1966).
71. D. Hoffman, Inferring shape from motion fields, MIT AI LAB memo 592 (1980).
72. D. Ballard and O. Kimball, Rigid body motion and depth from optical flow, TR 70, Dept. of Computer Science, University of Rochester (1981).
73. R. Sekuler and E. Levinson, The perception of moving targets, *Scient. Am.* **236,** 1, 60 (1977).
74. E. Freuder, A computer system for visual recognition using active knowledge, *Proceedings International Joint Conference on Artificial Intelligence,* Cambridge, p. 671 (1977).
75. J. O'Rourke and N. Badler, Model-based image analysis of human motion using constraint propagation, *IEEE Pattern Anal. Mach. Intell.* **PAMI-2,** 522 (1980).
76. R. Browse, Interpretation based interactions between levels of detail, *Proceedings Canadian Society for Computational Studies of Intelligence,* Saskatoon, p. 27 (1982).
77. B. Down, Using feedback in motion understanding, M.Sc. thesis, Dept. of Computer Science, University of Toronto (1983).
78. L. Cooper and R. Shepard, Chronometric studies of the rotation of mental images, *Mental Images and Their Transformations,* Shepard and Cooper, eds., p. 72. Bradford-MIT Press (1982).
79. B. Bugelski and D. Alampay, The role of frequency in developing perceptual sets, *Can. J. Psychol.* **15,** 205 (1961).
80. S. Palmer, The effects of contextual scenes on the identification of objects, *Memory Cognition* **3,** 309 (1975).
81. J. Mylopoulos, An overview of knowledge representation, *Proceedings Workshop on Data Abstraction, Databases and Conceptual Modelling,* Pingree Park, Colorado. (1980).
82. R. Kowalski, Predicate logic as a programming language, *Proceedings IFIP Congress* (1974).
83. H. Levesque, A formal treatment of incomplete knowledge bases, CSRG TR-139, Dept. of Computer Science, University of Toronto (1982).
84. R. Quillian, Semantic memory, *Semantic Information Processing,* Minsky, ed., p. 227. MIT Press (1968).
85. W. Woods, What's in a link: foundations for semantic networks, *Representation and Understanding,* Bobrow and Collins, eds., p. 35. Academic Press (1975).
86. L. Schubert, Extending the expressive power of semantic networks, *Artif. Intell.* **7,** 163 (1976).
87. G. Hendrix, Encoding knowledge in partitioned networks, *Associative Networks,* Findler, ed., p. 51. Academic Press (1979).
88. D. Waterman and F. Hayes-Roth, (eds.), *Pattern-Directed Inference Systems.* Academic Press (1979).
89. C. Hewitt, P. Bishop and R. Steiger, A universal modular actor formalism for artificial intelligence, *Proceedings International Joint Conference on Artificial Intelligence,* Palo Alto, CA, p. 235 (1973).
90. C. Hewitt, PLANNER: a language for proving theorems in robots, *Proceedings International Joint Conference on Artificial Intelligence,* London, p. 167 (1971).
91. I. Goldstein and R. Roberts, NUDGE: a knowledge-based scheduling program, *Proceedings International Joint Conference on Artificial Intelligence,* Cambridge, MA, p. 257 (1977).
92. D. Bobrow and T. Winograd, An overview of KRL, a knowledge representation language, *Cognitive Sci.* **1,** 3 (1977).
93. P. Szolovits, L. Hawkinson and W. Martin, An overview of OWL, a language for knowledge representation, MIT/LCS/TM-86, MIT Laboratory for Computer Science (1977).
94. Y. Lesperance, Handling exceptional conditions in PSN, *Proceedings Canadian Society for Computational Studies of Intelligence,* Victoria, BC, p. 63 (1980).
95. D. Marr and L. Vaina, Representation and recognition of the movements of shapes, MIT AI LAB memo 597 (1980).
96. C. Reiger and M. Grinsberg, The causal representation and simulation of physical mechanisms, TR 495, Computer Science Center, University of Maryland (1976).
97. T. Shibahara, J. Mylopoulos, J. Tsotsos and H. Covvey, CAA: a knowledge based system with causal knowledge to diagnose rhythm disorders in the heart, *Proceedings Canadian Society for Computational Studies of Intelligence,* Saskatoon, p. 71 (1982).
98. J. Freeman, Survey: the modelling of spatial relations, *Comput. Graphics Image Process.* **4,** 156 (1975).
99. B. Kuipers, Modelling spatial knowledge, *Cognitive Sci.* **2,** 129 (1978).
100. B. York, A. Hanson and E. Riseman, 3D object representations and matching with B-splines and surface patches, *Proceedings International Joint Conference on Artificial Intelligence,* Vancouver, p. 648 (1981).
101. J. Allen, An interval-based representation of temporal knowledge, *Proceedings International Joint Conference on Artificial Intelligence,* Vancouver, p. 221 (1981).
102. J. Tsotsos, A framework for visual motion understanding, CSRG TR-114, Dept of Computer Science, University of Toronto (1980).
103. N. Sridharan, J. Goodson and C. Schmidt, A research strategy for computational studies of event and action perception, CBM-TR-122, Lab. for Computer Science, Rutgers University (1981).
104. D. Ballard, C. Brown and J. Feldman, An approach to knowledge-directed image analysis, *Computer Vision Systems,* Hanson and Riseman, eds., p. 271. Academic Press (1978).
105. R. Reiter, A logic for default reasoning, *Artif. Intell.* **13,** 81 (1980).
106. R. Duda, P. Hart and N. Nilsson, Subjective Bayesian methods for rule-based inference systems, TN 124, SRI AI Center (1976).
107. S. Zucker, Production systems with feedback, *Pattern-Directed Inference Systems,* Waterman and Hayes-Roth, eds., p. 539 (1979).
108. E. Shortliffe, *Computer Based Medical Consultations: MYCIN.* Elsevier Press (1976).
109. M. Kelly, Edge detection in pictures by computer using planning, *Mach. Intell.* **6,** 228 (1971).
110. Y. Shirai, Analyzing intensity arrays using knowledge about scenes, *The Psychology of Computer Vision,* Winston, ed., p. 93. McGraw-Hill (1975).
111. R. Moore, *Reasoning about Knowledge and Action.* Garland Press (1981).
112. S. Fahlman, *NETL: A System for Representing and Using Real-World Knowledge.* MIT Press (1979).
113. D. Sabbah, Design of a highly parallel visual recognition system, *Proceedings International Joint Conference on Artificial Intelligence,* Vancouver, p. 722 (1981).
114. G. E. Hinton, Shape representation in parallel systems, *Proceedings International Joint Conference on Artificial Intelligence,* Vancouver, p. 1088 (1981).
115. J. Feldman and D. Ballard, Connectionist models and their properties, *Cognitive Sci.* **6,** 205 (1983).

About the Author—JOHN K. TSOTSOS was born in Windsor, Ontario, Canada. He received the B.A.Sc. degree in Engineering Science in 1974 and the M.Sc. and Ph.D. degrees in Computer Science in 1976 and 1980 respectively, all from the University of Toronto. He is currently an Assistant professor in the Departments of Computer Science and Medicine at the University of Toronto. His research interests are in artificial intelligence, computer vision, motion interpretation and medical expert systems.