PSYCHOLOGY OF VISION — A COMPUTER MAN'S POINT OF VIEW

T. Kasvand

National Research Council of Canada, Ottawa

ABSTRACT

The computer-programmer or engineer reading the theories of vision expounded in psychology-orientated literature, besides the unfamiliar terminology, tends to find that some 95 ± 5% of these studies are fruitless. Yet only the biologically developed visual systems manage to cope with the complexities of our everyday visual surroundings, while our computer programs make nothing out of such complicated scenes.

The author will attempt to "translate" some of the findings into a more familiar terminology and demonstrate how some of the image processing principles found in the visual systems can be usefully employed in computer programs for scene analysis.

RÉSUMÉ

L'ingénieur ou le programmeur informaticien qui se renseigne sur les théories de la vue exposées dans la documentation en psychologie a tendance à trouver, mise à part une terminologie inusitée, que quelque 95 ± 5% de ces études ne riment à rien. Jusqu'ici, seuls les systèmes biologiques perfectionnés de la vue sont capables d'assimiler les complexités de notre environnement visuel quotidien, tandis que nos programmes informatiques ne peuvent rien tirer de ces scènes compliquées.

L'auteur tente de "traduire" certains résultats en un langage plus simple et de démontrer comment certains principes de traitement des images intervenant dans les systèmes visuels peuvent être exploités dans des programmes informatiques d'analyse des scènes.

Introduction.

In both design and operation the human visual system, when looked upon in engineering and information processing terms, is a very elegant system. Unfortunately, however, the "design" as well as the terminology used to describe the visual system is highly unfamiliar to the engineer trained in, say, electronics and information processing. An attempt is made in this paper to look upon the entire visual system as one integrated and operational unit.

Objections to trying to describe biological vision systems in engineering terminology may be based on the fact that their components differ radically from our electronic, optical and mechanical hardware. Consequently, questions may be raised, such as: Which and how much of the observed information processing strategies of the visual system are dictated by the nature of the "hardware" layout and characteristics? Which aspects are reflecting the "theoretical" solution to the visual image processing problem? Is the seen solution the only possible one?

At the present level of understanding of human vision, these questions remain unanswered and introduce additional uncertainties to the interpretations of any findings. Despite these doubts, however, the following descriptions may be adequate to demonstrate that "hardware differences" need not obscure the operating principles.

It may also be argued that the human visual information processing system is too complex and the studies should start with simpler systems. The difficulty with this approach is that each species has a highly perfected visual system adapted to its own needs and no real continuity of development has been observed from lower to higher biological organisms. To study the natural (i.e., undisturbed) development of a visual system from its infancy to adulthood does not simplify the problem. Consequently one is left with the problem of studying very complex systems.

The author claims no particular expertise in the fields from which material has been obtained, except some formal courses and much reading. By simply interpreting the findings in terms of engineering and information processing principles, a certain insight has been obtained which is briefly summarized. Such mixing of disciplines, however, is for a variety of reasons highly disliked by most parties.

However, to quote from a knowledgeable source (Watanabe): "Yet an amateur has a fresh sense of "amazement", a balanced birds-eye view of tremendous scope, and a direct contact with the world of common sense which is the mother earth of all knowledge".

Functional Design

When we observe a scene with two eyes, the left eye sees the left and the center part of the scene, while the right eye sees the center and the right part of the scene. This is just a matter of optics. The images falling onto the retinas (light sensitive surfaces in the eyes), besides being partially processed, are assembled into a topologically correct single "image" in the visual cortex (brain surface at the back of the head).

The eye can only distinguish some 5 to 10 gray or monochromatic light intensity levels on an absolute scale, while hundreds of intensity level differences can be distinguished. Since light intensities vary so much, absolute judgement is apparently of little value. For colour constancies see Land. Those interested in compound eyes may enjoy reading Autrum.

One eye contains some 100,000,000 light receptors. The density distribution of the receptors resembles a Gaussian curve, with clearest vision where the density is highest. It is conjectured that such an arrangement allows an "electronic" zoom effect since the number of signal paths leaving the eye (fibers in the optic nerve) is about 1,000,000. Recently it has also been argued that this distribution is related to size constancies (Wilson).

The various biologically developed (evolved) visual information processing systems go to great lengths to improve the dynamic range, sensitivity and operation of their sensors by a variety of means and finesse information extraction to an amazing degree. The example given below uses the eye as an illustration to demonstrate the variety of engineering and information processing principles involved.

The dynamic range of the eye is very great, being achieved by a combination of optical principles, chemistry and information processing. Thus starting from very bright light towards the very dim photon-limited case, one finds the following mechanisms:

1) Eye lid (which may be partially transparent in some species and acts as a filter).

2) Iris.

3) Cone cells. (Photoreceptors for colour vision.)

4) Rods. (Photoreceptors for gray level vision at lower light levels.) The concentration of the photo-chemicals plays an important role both at steps 3 and 4, to vary the sensitivity of the photo-receptors.

5) Spatial integration. (The cells in the eye operate in groups, acting like single receptors, thus we only see rough outlines but no details in dim light.)

6) Temporal integration. (The signal is integrated for about 0.1 sec. Operations 5 and 6 are partially co-occurrent.)

7) Reflectors. (Some species have reflectors in the eye to return photons that missed the photo-receptors in the first pass through the retina.)

Given the present level of technological knowhow, but the constraints of the "hardware" that biology has to "work with", could one have designed a much better system? Such an elaboration seems to be carried out at all levels and in all aspects of visual image processing.

Information Processing.

To try to discover the operating principles used by a visual system is not at all simple. A not entirely accurate analogy may explain the problem: Assume a digital computer busily working at some problem. Since no program listing is available, it is desired to find out what the computer is doing, i.e., what program is running. Given suitable oscillographs, recorders, etc., can one discover the algorithm by electrical measurements anywhere inside the computer? In the biological case the probe is rather big and is likely to damage or alter the operation of the system.

As stated, this analogy is not exactly correct, but it illustrates very vividly how difficult it is to determine the nature of an algorithm by physical measurements alone. Consequently, much of the knowledge has to be obtained by indirect means and is thus very much open to misinterpretation. Furthermore, one experiment gives only some data about a very limited aspect of the whole system. The excitation of a part of a system may not necessarily produce the same effects as when this part is operating in unison with the rest of the system, nor is there any way to isolate a function or a mechanism and measure it independently of its embedding.

The operation of the visual system of all humans is very likely to be similar, excluding visual defects. It is also well known that we can learn to discriminate very subtle visual cues. For example, Eskimos can recognize many different types of snow. The true native of a jungle can read the signs in his environment as well as or maybe better than the modern man can read traffic signs. The medical specialist can see much in a microscope slide of a specimen or in an x-ray picture which is not noticeable to the "untrained" eye, etc. The basic question is whether the visual system in all cases processes the input in the same way, and a "specialist" just learns to select what is vital for his field, or whether the input is actually processed differently in each case. The first alternative is more likely and the author has not encountered any material to the contrary.

Despite the likely correctness of the assumption that human visual systems operate in the same way, it has been observed that an adequate visual environment is needed in early life in order to develop or refine the visual system (von Senden). One of the sources of visual defects is a degenerate environment, resulting in the actual absence of some of the "measureable" visual information processing mechanisms (Blakemore).

A very large number of phenomena have been noticed and experimented with and may, of course, be found in the appropriate literature. Despite the multitude of phenomena, it might be fruitful to separate the human visual information processing system into four stages, namely, preliminary processing methods where the processing and the results are observable by physical measurements, secondary processing where the observed phenomena cannot or have not yet been measured directly but the observations are both suggestive and logical, and "high level processing" of which we can observe a variety of phenomena but an immediate algorithmic interpretation may be premature. In the subsequent sections these processing stages will be called "measureable", "likely", "observable" and "conjectured".

Measureable Processes.

It has been verified by actual physical measurements that in the retina and in the visual cortex the nerve cells are interacting to produce responses to flickering light, moving edges, edge directions, and so on (Hubel, Dodwell). The spatial extent or region, or the time duration, over which the light receptors are grouped or combined for these operations is relatively small or local, i.e., such groupings may be called "local operators". Groupings over large areas of the retina may exist, but the author is unaware of any experimental proof. Furthermore, it has been shown that at least some of the results of such local processing are systematically ordered in the visual cortex. Beyond these observations nothing much seems to be either accessible or interpretable by direct measurements.

Programming of equivalent effects is rather common in computer vision, at least for gray level gradients, Laplacians, edges, etc. When realistic images are processed, the results are usually rather noisy and not very easy to use in subsequent processing. Furthermore, a profusion of local operators may be defined which generate a combinatorial "explosion" if all the possible combinations and permutations of the results of these local operators are contemplated or used. The minimal set of local operators should be defined, for example, edge preserving smoothing, Laplacian type enhancement, gradient in magnitude and direction, motion in direction and velocity, time derivative of intensity, etc.

Likely Processes.

Since the results from local processing only refer to small areas in the image, besides being rather noisy, it is logical to group these results into larger units so as to reduce noise and to refine the results. In computer vision some such attempts have been given the unfortunate name of "relaxation labelling". A better name may be filtering for similarity or filtering for enhancement of similarity to achieve a consensus within a group. Such "consensus filtering" processes, besides reducing noise also serve to organize the local features into larger units or "wholes", thereby reducing the number of individual entities to be handled by subsequent "higher level" processes.

In the visual system, organization of local features is carried to a very high level of sophistication, most likely in order to minimize the number of possible combinations. In fact, the visual input image is organized into a unique combination of structures before it is presented to the "higher level". These organizations may be over-ridden by the "higher levels" if the input "makes no sense" or when there are other reasons to organize the inputs differently. The organization "laws" in the visual system have been called (elementary) Gestalt laws by the original discoverer, Max Wertheimer. Gestalt means shape or form in German. Subsequently the Gestalt idea was generalized into absurdity, but this is of no concern here. A list of such "laws" (Koffka) is given at the end of this section.

The exact number as well as the exact behaviour of these laws is not certain at present since no physical verification procedures exist. However, subjectively their operation is easy to demonstrate and, as argued earlier, the existence of such laws reduces the computational complexity of the image processing problem. These laws can also organize the data into incorrect groupings, which is the basic principle of successful camouflage.

All these laws appear to be operating when an image is processed, but they do not operate entirely in parallel. Some series operations occur also, as for example in being able to see two crossing dotted lines as two individual (dotted) lines (that cross). In this example both the similarity (or proximity) and good continuation laws are involved.

The researchers in computer vision do not seem to be aware of the existence of most of these laws, or at least no serious attempts have been made (as far as the author knows) to systematically develop algorithms for them. In a paper (Kasvand, 1979) it is demonstrated that the application of just one of these laws, namely the law of good continuation, results in a remarkably good segmentation of handwritten Chinese characters into their underlying lines.

Gestalt Laws (Koffka).

Law of assimilation.
Law of association.
Law of closure.
Law of contiguity.
Law of contrast.
Law of effect.
Law of equality.
Law of exercise.
Law of fittingness.
Law of frequency
Law of good continuation.
Law of good shape.
Law of pragnanz.
Law of proximity.
Law of recency.
Law of similarity.
Law of the simplest path.
Law of size.
Law of stability.
Law of substitution.
Law of success.
Law of transposition.
Law of organization.

Observable Processes

The observable processes are even larger groupings of visual information which are appropriately integrated with other senses. These effects we can see and frequently name, but we are unlikely to be able to verify their existence by physical measurements on the visual system. A very good example is distance determination by image processing. The entirely obvious fact that the human visual system has two eyes and is mobile has a significant bearing on depth or distance detection.

The methods of depth detection may be grouped in a variety of ways, for example

a) The one-eyed observer.
b) The two-eyed observer.
c) The stationary observer.
d) The passively moved observer.
e) The actively moving observer.
f) Effects dependent on observer.
g) Effects independent of observer.
h) Etc.

The following list has been compiled from Gibson, according to his classification and terminology. In addition to these, depth is also detected from "lens setting for clear vision" by some species.

Cues to Depth (from Gibson).

Perspectives of position: (The observer is stationary and the effect is visible with one eye.)

1. Texture perspective. (Texture gradient)
2. Size perspective. (Object size versus distance)
3. Linear perspective. ("Parallel lines meet at infinity")

Perspectives of parallaxLaw of reproduction.

4. Binocular perspective. (The skew of the image in one eye with respect to that in the other, stationary observer)
5. Motion perspective. (The change in binocular perspective when eyes are moved and the change in relative displacement of objects due tothe motion of the observer). Several aspects of motion perspective are also observable by the moving one-eyed observer.

Perspectives independent of observer's motion or position.

6. Aerial perspective. (The haziness, blueness and desaturation of colours as a function of distance)

7. The perspective of blur. (Variation in the quality of blur as a function of displacement from the center of clear vision)
8. Relative upward location in the visual field. (The angular extent of background between the lower margin of the visual field and the object)

Depth at contour.

9. Shift of texture density or linear spacing. (Sudden changes of texture or texture gradient)
10. Shift in the amount of double imagery. (Sudden changes in the skew of texture over distance)
11. Shift in the rate of motion. (Changes in the displacement of texture elements on one side of a contour with respect to the other due to motion of the head)

Depth due to object shape.

12. Completeness or continuity of outline. (Complete objects appear closer)

Depth effects due to lighting.

13. Transition between light and shade.

As seen, the methods of depth detection are extremely elaborate. Nothing of similar sophistication has yet been attempted in computer vision. However, since the eye is a passive light detector and we cannot generate light ourselves, the biological equivalent to a laser range finder is missing! It is intriguing to speculate whether organisms in deep oceans, who do carry their own sources of light, might have developed even that.

Conjectured Processes.

Some aspects of the behaviour of the visual system may become more clearly observable under suitable experimental or circumstantial conditions. The literature on the psychology of vision is full of examples. However, much interesting information is also obtainable from studies of the effects of brain damage and even from how children structure their drawings (Arnheim). Only a few examples will be given

i) Eye motion studies (Buswell, Yarbus).

When we observe a picture or a scene, our eyes move in a "stop and go" fashion. The eyes stop at ("fixate") a point in the picture and process ("take in") the neighbourhood of that point. The size of the neighbourhood is variable, but has not been measured experimentally. Then our eyes fixate another region in the picture, and the processing is repeated. During these (large) eye motions no input image is accepted. Eye motions have been recorded and studied for at least 50 years.

It is observed that the scene is only sampled, where the sampling process is driven both by picture content and by the problem to which an answer is sought in (the picture of) the scene. Neither colours nor gray levels seem to affect this process significantly. Thin line drawings representing the contours alone are adequate in normal circumstances. If this were not the case, we ourselves would not be able to understand line drawings and the entire concept of "a line drawing" would most likely not exist. In a line drawing, however, the information has been reduced to a minimum and a far greater global integration of the data is required to assign a meaning to each line configuration, than in the case of more information-rich pictures.

ii) Visual illusions (Beeler, Coren).

Visual illusions, i.e. situations where the scene or something in it is misinterpreted, occur seldom in natural environment. Camouflage, however, is very common, which is implying that the equivalents to Gestalt rules are used by most species. Whether imitations and protective colouring etc., are to be classified as illusions or not is immaterial here. Normally visual illusions are created by information-deficient (two dimensional) pictures, deliberate combinations of contradictory information, restricted views of three dimensional scenes, etc. Such experiments can demonstrate where and how certain information in a picture is used. Interpretations of these experiments imply that vision operates in a very logical and elegant manner.

iii) Brain damage (Luria).

In the case of brain damage a large variety of defects in visual information processing may be observed. Many of the defects are very suggestive of a failure in a particular well recognizable processing step. For example, it may be conjectured that the defect is the result of a missing feature detector or a Gestalt rule or a transformation (such as rotation), or that the descriptive interrelationships of parts belonging to an object or an object class have somehow been "lost", etc.

Remarks.

This extremely sketchy survey of the human visual system, it is hoped, has at least demonstrated that the system is operating on entirely logical principles. The references should be consulted for illustrations and for better and more precise descriptions. Even though our knowledge is rather limited, much is understandable but not necessarily verifiable by impartial or scientific measurements. It is also rather apparent that our visual information processing system is orders of magnitude more computation intensive than our present computer vision systems. The information processing on lower levels operates "a mass", i.e., massive parallel and serial processing steps take place, while at the higher levels very elegant and efficient methods are used to minimize the processing time to a few seconds or less.

Acknowledgements

The author wishes to express his sincere thanks to the Electrical Engineering Division and the Computer Technology Program of the National Research Council of Canada, and the Electrotechnical Laboratory in Japan, for making it possible to study and think about these problems.

References.

Arnheim R., Art and Visual Perception, Univ. of California Press, 1954.

Autrum H., Comparative Physiology and Evolution of Vision in Invertebrates, Springer-Verlag, 1979.

Beeler N.F., and F.M. Branley, Experiments in Optical Illusions, Thomas Y. Crowell Co., 1951.

Blakemore C. and G.F. Cooper, Development of the Brain depends on the Visual Environment, Nature, vol. 228, Oct. 31, 1970.

Buswell G.T., How People look at Pictures, Univ. of Chicago Press, 1935.

Coren S. and J.S. Girgus, Seeing is Deceiving, John Wiley, 1978

Dawson H., ed., The Eye, Academic Press, 1962, (four volumes).

Dodwell P.C., Visual Pattern Recognition, Holt, Rinehart and Winston Inc., 1970.

Gibson J.J.,The Perception of the Visual World, Houghton Mifflin Co., 1950.

Graham C.H., ed., Vision and Visual Perception, John Wiley, 1965.

Ellis W., ed., A Sourcebook of Gestalt Psychology, Humanities Press Inc., 1967.

Hubel D.H. and T.N. Wiesel, Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex, Journal of Physiology, vol. 160, 1962.

Kasvand T., Ideogram Segmentation and Recognition, International Conference on Cybernetics and Society, Oct. 1979, pp. 674-678.

Kasvand T., Some Observations and Experiments on the Elements of for a Picture Language, in: Pattern Recognition Theory and Applications, eds., K.S. Fu, A.B. Whinston, Noordhoff, 1977. (There is also an NRC internal memo with same name and author, Aug.28,1975, from which the paper was abstracted.)

Koffka K., Principles of Gestalt Psychology, Harcourt, Brace and World Inc., 1963.

Land E.H., The Six Eyes of Man, 3D Imaging, SPIE, vol. 120., Aug. 25-27, 1977, San Diego, Calif., p. 43.

Luria A.R., Higher Cortical Functions in Man, Basic Books Inc., 1966.

Metzger W., Gesetze des Sehen, Verlag von Waldemar Kramer, 1953.

Polyak S.L., The Vertebrate Visual System, University of Chicago Press, 1957.

Senden von M., Space and Sight, Methuen and Co. Ltd., 1960.

Watanabe S., Knowing and Guessing,

Wilson S.W., On the Retino-Cortical Mapping, research report, Polaroid Corp.

Zusne L., Visual Perception of Form, Academic Press, 1970.