THE VIEWER'S PLACE IN THEORIES OF VISION

R.I.D. COWIE

PSYCHOLOGY DEPARTMENT, QUEEN'S UNIVERSITY OF BELFAST

ABSTRACT

Analyses of vision have tended to focus on what may be under observation rather than considering the larger configuration including observer and observed. It is, however, possible to think systematically about the "missing link" of observer/observed relationships. The paper aims to clarify the logic associated with these relationships. At least three main types of observer/observed relationship are interesting: when viewpoint is general, when viewpoint is representative, and when viewpoint is privileged. It is shown that knowledge about such relationships can support interesting conclusions - though the exact forms of these conclusions will reflect other empirical and a priori knowledge. The various viewpoint analyses may be regarded as aspects of a single issue - when and how an observer may legitimately assume that properties of stimulation *are* simply and directly indicative of features of the world. Also, expressing observer/observed relationships calls for appropriate ways of classifying observables: so understanding observer/observed relationships feeds back into understanding what may be observed.

(a) Introduction

The theory of vision deals with what can be called viewing configurations - configurations including an observer and things observed. Relationships between the observer and things observed are an integral part of a viewing configuration, and so presumably we should in principle be concerned with understanding how these relationships can usefully be described, and with establishing what a viewer can profitably assume or arrange about them. An example will suggest why it seen, worthwhile in practice to consider these relationships carefully. Why should we accept that the lines in Fig. 1(a) represent wires or edges which meet, not (say) a corner and a quite separate wire as in Fig. 1(b)?

FIGURE 1

A strong reason is that to accept the interpretation shown in (b) entails accepting that the observer was in a very specific position, i.e., somewhere on the line through X and Y. We can usually sensibly assume that this kind of positioning is most unlikely, invoking what Huffman (1971) and Cowie (1982) called the assumption that viewing position is general and Binford (1983) called the principle of indifference to the observer. Here a reasonable and useful conclusion can be motivated by invoking a very weak assumption about observer/observed relationships. We should clearly like to know more about what this kind of reasoning can buy.

The aim of this paper is to provide an overview of interesting types of observer/observed relationship. Three main forms of observer/observed relationship have been explored in some ways. The paper proceeds by setting them out in turn, trying to make their underlying logic and relationships as clear as possible.

(b) General Viewpoint

The first type of relationship is associated with the idea of a general viewpoint, which has already been mentioned.

The rationale of general viewpoint arguments can be clarified considerably by starting with the issue of classification. This approach avoids problems with more intuitive approaches which were exposed by Draper (1980). To invoke the general viewpoint assumption, one needs a system of three inter-related types of class. This kind of system is illustrated in Fig. 2. First, there is

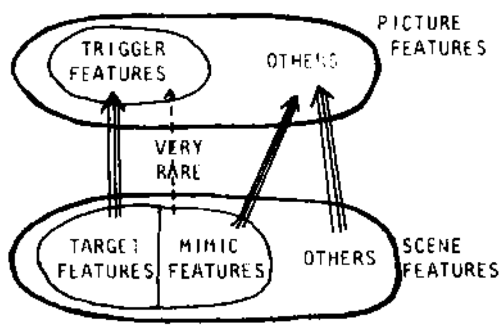


FIGURE 2

a class of picture features which will be called trigger features. Secondly, there is a class of scene fragments which will be called target features. Thirdly, there is another class of scene features which will be called mimics. Target features and mimics share the property that projections of them may be trigger features, and they must exhaust all the scene features which could give rise to trigger features. But target and mimic features must contrast in the ways in which they give rise to trigger features. Target features must give rise to trigger features from any of a wide range of viewpoints. By contrast, there will be associated with any mimic feature a locus whose volume is zero. The defining characteristic of this locus is that only when the viewpoint is on that locus will an image of that feature be a trigger feature. This kind of locus will be called a locus of deception.

The point of a classification system like this is that one can strongly expect any trigger feature to represent a target feature provided that two conditions are met.

(i) There must be no relationship between viewer and viewed which makes the point of observation particularly likely to lie in a locus of deception. It is strictly the assumption that this condition is met which has been called the assumption that viewing position is general or that the world is indifferent to the observer.

(ii) The frequency with which target features occur must not be negligible by comparison with the frequency of mimic features. This will be called the target frequency condition.

Clearly, if both of these conditions are met then in any sample of pictures, trigger features will arise overwhelmingly more often as pictures of target features than as pictures of mimic features. Correspondingly one will be right in the overwhelming majority of cases if one assumes that a trigger feature always represents a target feature.

It is apparent from this that by using the assumption of a general viewing position one can get information with remarkably little a priori knowledge about the world. In fact there is even less need for a priori knowledge than it might seem. If trigger features of a particular type occur with some frequency in pictures of a given world, then one can infer that either viewing position is tending to fall on loci of deception, or the target frequency assumption is met. This is because if viewing position were general, and the frequency of target features were negligible, then the frequency of trigger features would also be negligible. There are two consequences of this observation. The minor one is that even if the target frequency assumption is not met, there is little risk of error associated with interpreting triggers as targets. One will simply not again anything, because one will rarely come across triggers. The major point is, as has already been noted, that one really needs minimal a priori knowledge to get information on the basis of the general viewpoint assumption.

It remains to be shown what kinds of information can be obtained by using the general viewpoint assumption. This hinges on what systems of classes actually meet the abstract criteria which have been set out so far.

There are two extreme ways of developing appropriate class systems. To illustrate one extreme, one could take as the target class the class of cubes. Though no proof will be offered, it is fairly easy to see that at least when orthographic projection is involved, all mimics have loci of deception associated with them. So if viewing position is general, it is a safe strategy to accept that any picture which could represent a cube does represent a cube.

This system can be extended to include any finite set of rectangular parallelopipeds, each with specified proportions, in the target set. Adding a new rectangular parallelopiped to the target set will add a new component to the locus of deception associated with many (or all) mimics. But so long as the number of rectangular parallelopipeds in the target set stays finite, the locus of deception associated with any mimic will remain at zero volume. What one cannot do is increase the target set to contain all rectangular parallelopipeds whose dimensions lie between some given limits. To do that would associate with each mimic a locus of deception with an infinite number of components which, added together, would occupy a non-zero volume.

Two points should be made about this example.

Firstly, the example exposes a tempting misunderstanding of the General Viewpoint assumption. It is obviously absurd to say, "it follows from the assumption that viewing position is general that this represents a cube". The root of the absurdity is the assumption that specific interpretations of any sort follow from the General Viewpoint assumption. In fact, the General Viewpoint assumption is rather like free credit which nature extends to perceivers. Only the amount available is fixed by the assumption: it is up to the perceiver whether he buys, so to speak, cubes; or rhomboids with angles of 60 ; or whatever.

Secondly, the example shows that the General Viewpoint assumption can support strong conclusions about a limited range of objects. This is the extreme which was mentioned in introducing the example.

The second extreme, naturally, involves weak conclusions about a large range of objects. This extreme is most simply illustrated in the case of interpreting drawings composed of straight lines. (Binford (1981) discusses a parallel kind of analysis on scenes involving curves.)

Analysis involves two levels: picking out a repertoire of basic trigger features, and building from these by inference. Table 1 illustrates some basic trigger features. Inference rules can group edges into sets whose members are coplanar with one another. (Coplanarity is a key relationship because the triggers provide information primarily about intersections. The knowledge that lines intersect implies that they lie in the same plane, so considering coplanarity provides a natural way of integrating information from triggers.)

This means that given only general viewpoint and target frequency assumptions, one can obtain much of the kind of information which Huffman (1971), Clowes (1971) and their successors obtained by using a good deal of a priori knowledge. The assertion that edges are coplanar carries a good deal of the information which the labelling programs provide by associating edges with a single surface. Decisions about convexity/concavity and occlusion are always strongly constrained by coplanarity relations, as POLY showed clearly (Mackworth, 1974, and where TEE junctions are present these relationships may emerge directly from an analysis based purely on general viewpoint and target frequency assumptions. These assumptions also provide a characteristic indicator that drawings may be problematic, an issue at the heart of the Huffman-Clowes tradition. The indicator is that all of the lines representing an object must be coplanar if viewing position is general. This giveaway is associated directly with the first two pictures in Fig. 3, one of which we see as impossible and the other of which involves an accidental alignment. It is associated in an indirect form with 3(c), a "missing line" example discussed extensively by Mackworth (1974). The edges of 3(c) do not fall into a single coplanar set, but they would if one assumed that the edges around A belonged to a single plane surface. This provides a motivation for splitting A which is a good deal simpler than the one proposed by Mackworth. Having mentioned similarities, it is worth pointing a contrast between the Huffman-Clowes tradition and analysis based on a general viewpoint assumption. The a priori assumption that surfaces bounded by straight lines will be plane is central to the Huffman-Clowes tradition. But it is not necessarily true, and the visual system does not always make it. Fig. 3(d) illustrates.

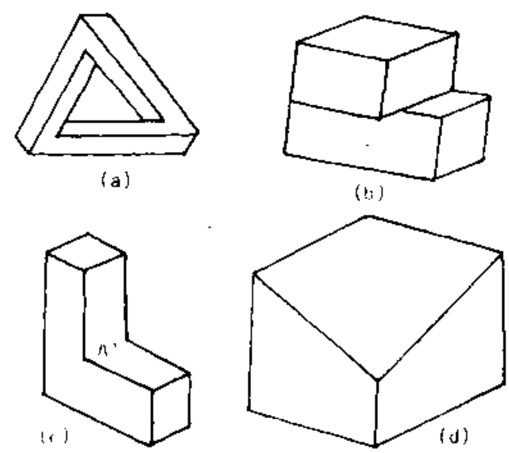
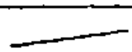
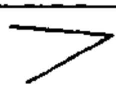


FIGURE 3

TABLE 1

	Trigger	Target
	straight line	straight edge or wire
	2 lines which meet and stop at a point	2 edges or wires which meet and stop at a point
	2 lines, one of which stops where it meets the other	2 wires one of which stops where it meets the other <u>or</u> an edge which either occludes <u>or</u> meets a second edge or wire
	2 parallel lines	2 parallel wires or edges (assuming projection is orthographic)
	3 or more convergent lines ("convergent" meaning they meet at a single point, actually or when produced)	3 or more convergent wires or edges

Most people see the top face as curved. This is in fact implied by the general viewpoint-based interpretations of Table 1.

These examples may serve to illustrate the point that the General Viewpoint assumption can sanction weak conclusions about a large range of objects, as well as strong conclusions about a limited range of objects. They should also serve to reinforce the point that although one needs minimal a priori knowledge to reach conclusions based on the general viewpoint assumption, the conclusions which one can base on it are very far from trivial.

A final illustration may clinch this point. Consider an observer moving from A to B towards a curved rigid configuration C, which will be taken as a stationary frame of reference. (See Fig. 4). His prospective point of impact will be called P. Longuet-Higgins and Prazdny (1980) showed that even if the observer's orientation changes between A and B, P has a distinctive feature which makes it identifiable: if two points appear to lie on a line through P when the observer is at A, then they will also appear to lie on a line through P when he is at B. This continued alignment simply means that the points lie in a plane through ABP. Now if a moving point is to mislead the observer into thinking that it is part of C, it must clearly be moving in a plane through ABP. This constraint can easily be inverted into a constraint on the observer's motion. If we take A as fixed, then the moving point cannot be mistaken for part of C unless B lies in the plane containing A and the rogue point's initial and final positions. This means that a non-general viewpoint is needed for a non-rigid configuration to mimic a rigid one. (Notice that if the observer moved to a third point, a new kind of constraint on viewpoint would come in. The third point would have to be at a particular distance from B, otherwise the observer would derive different estimates of the moving point's distance from the points in the rigid configuration.)

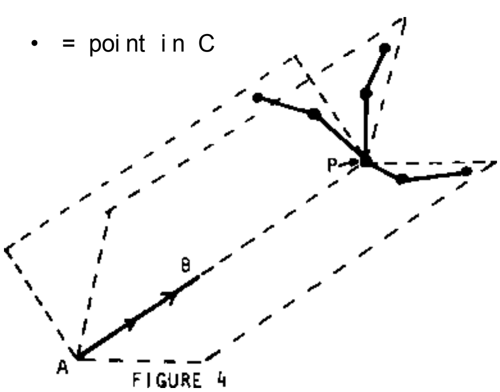


FIGURE 4

This analysis is, of course, closely related to the one which lets Ullman (1979) conclude that "the probability that three views of four points not moving rigidly together will admit a rigid interpretation ... is zero" (p. 149). Ullman's conclusion seems extraordinary: given that stimulation is in principle profoundly ambiguous, how can we give a guarantee which implies, as this does, that a system may derive full interpretations with negligible probability of error? The point of the sketch above is to show that this property of "structure from motion" schemes is in fact yet another illustration of how much can be gained, and at what low risk, by trading on the assumption of a General Viewpoint. We can be confident of interpretations involving rigid structures because to be fooled, the observer would have to be maintaining a strange and exact relationship to the things observed: and in our world, that will almost never happen unless the observer arranges it.

(c) <u>Representative Viewpoint</u>

Attention is now turned to relationships which are very like those which have just been discussed, but less exact. These allow one to use what Cowie (1982) called the assumption that viewing position is representative.

Research into this area was suggested by some strange facts about human vision. Consider, for instance, Fig. 5. We tend to see (a) and (b)

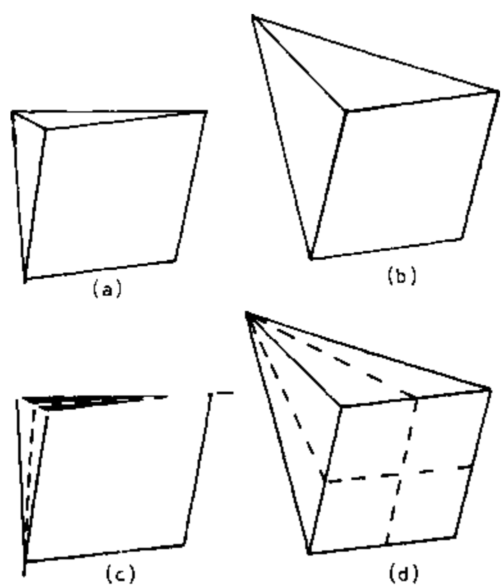


FIGURE 5

as having different axes of symmetry, roughly those indicated by the dotted lines in Figs. 5(c) and (d) respectively. Yet both types of symmetry could be imposed on both drawings.

It is not easy to explain our preferences here in terms of viewed configurations, and intuition suggests that viewpoint might be an important consideration. Roughly, if 5(a) represents a pyramid like the natural interpretation of 5(b),it is seen from an extreme and unhelpful angle. This intuition suggests that our preference might be justified by something akin to the general viewpoint assumption but relating to large and small volumes of space rather than zeroand non-zero volumes.

Two kinds of relationship can be established which *are* relevant to this example. The numerical estimates *are* appropriate to orthographic projection, though they *are* probably not far out for perspective.

(i) From most viewpoints, edges of similar length appear as lines of similar length. The strength of the relationship depends on the angle between the edges. If the angle is 15 , the shorter line will be over 70% of the longer from around 95% of viewing angles. If the angle is 90%, about half of viewing angles lead to a shorter line more than 70% of the longer.

(ii) From most viewpoints, edges which meet at a particular angle appear as lines which meet at a similar angle. Edges which meet at 10 will appear as lines meeting at 15 or less from about 90° of viewing angles. Edges which meet at 90 will appear as lines meeting at 15 or less from around 5% of viewing angles.

These results can be applied to Fig. 5(a) as follows.

Note that the lines which we see as roughly equal edges are themselves roughly equal. To assume that the edges were very far from equal would imply a viewpoint within a restricted range - unless the edges were at a fairly large angle to each other. But the lines *are* at a fairly small angle to each other, so to assume that the edges were at a large angle to each other would also imply a viewpoint within a restricted range.

This argument is not in itself a justification for our interpretation of Fig. 5(a), because it lacks the analogue of the target frequency assumption. If edges which met were highly unlikely to be similar in length, then the proper conclusion would have to be that lines of a similar length meeting at a small angle were probably due to edges of dis-similar length seen at an odd angle. The point is rather that the viewpoint argument provides a pressure towards interpreting the pictures as we do.

Note that this example involves a point which is familiar from, for instance, Marr and Nishihara (1978): certain features associated with symmetry are preserved to a reasonable approximation over a relatively wide range of viewpoints. Analysing out the assumption that viewing position is representative is in part, therefore, simply formalising and clarifying a kind of argument which we do already use.

The kind of reasoning used above also applies to a very different, and very important context. The interpretation of curves has not yet been considered. The general viewpoint assumption can clearly be applied in some cases. For instance, conic sections could be used as triggers whose targets would be spheres or (in orthographic projection) circles and elipsoids. But it is difficult to image any comparable analysis being extended to very much of the huge range of curves, many very irregular, which we find in everyday life.

There are nevertheless very general senses in which reasoning about viewpoints does seem to bear on interpreting curves. For instance, it does seem that only in rather rare alignments would a curve like Fig. 6(a) turn out to look like 6(b) from the side, with a totally unsignalled kink. One can make some progress towards a principled rejection of interpretations like 6(b) by considering a curve as a series of straight line segments, and invoking the principle that the angle between lines tends to be like the angle between the edges which they represent. This is obviously not a good analysis. The important point is that viewpoint arguments do seem relevant to interpreting curves, and that representative viewpoint arguments may well be more relevant than general viewpoint arguments. (Binford (1981) shows that a more extreme case is handled by a general viewpoint argument.)

(d) Privileged Viewpoint

The discussion of viewing configuration so far has invoked the idea that the viewer's position is likely to be unrelated to the relevant features of the world. The third kind of relationship is where the viewpoint is definitely not a random choice, but is specially and usefully related to the features in question. Such viewpoints can aptly be described as privileged.

(a) (b)

FIGURE 6

Some instances of privileged viewpoint arise for very simple reasons. The obvious example involves verticals. We expect these to appear vertical. This is because a main axis of our viewing apparatus tends to be vertical. This is a restriction which is formally very like being in a locus of deception: we deny ourselves, so to speak, a degree of freedom In positioning our viewing apparatus. But we turn the restriction to advantage.

Other instances of privileged viewpoint arise because we arrange them. Many kinds of visual judgement we make by positioning ourselves to get 'a good view", for instance on a perpendicular to the centre of a surface or exactly edge-on to it (if we want, for instance, to judge whether a surface is truly flat). When we cannot control our own viewpoint, we expect our picture-makers to take account of our preferences - for instance by keeping the picture plane parallel to the facade of a building, and the station point in a comfortable position (see Hagen, 198O on this point).

A last class of privileged viewpoints is more hypothetical. We accept without qualms the idea that a perceiver may select an interpretation on the basis of the "goodness" of the objects perceived. It is surely unwarranted prejudice to accept this and deny that an interpretation may be chosen because it allows one *to* assume "good" viewer-object relationships. An example which suggests that we do this arises when we interpret an el ipse in a picture as a solid. An el ipse could be a perspective projection of a sphere at a distance from the principal point, but our interpretation tends to involve a less regular shape - an elipsoid - viewed in a more friendly way. In addition, we tend to adopt an interpretation where the el ipse presented is a medial section of the elipsoid, assuming that our view informs us very directly of the object's shape: we do not readily accept that the elipsoid's long axis might be at an angle to us. It will be noted that this is a special case of the situation examined by Marr (1977). His analysis adds the general point that assuming a viewpoint which was friendly in certain ways would explain our tendency to interpret silhouettes as something akin to generalised cylinders.

Note that one could not unreasonably subsume general and representative viewpoints as special cases of privileged viewpoints. If one defines a privileged viewpoint as one from which a feature of stimulation bears a specially straightforward and useful relationship to a scene feature which it represents, then general viewpoint analysis would deal with the type of feature for which almost all of space consists of privileged viewpoints, and representative viewpoint analysis with the type of feature for which most of space is consists of privileged viewpoints. The main point of this comment is to make it clear that there is nothing incompatible about saying that a perceiver invokes the assumption that viewing position is general and also the assumption that viewing position is privileged. Each assumption reflects the underlying motivation to assume, where it is justifiable, that a picture feature has a relatively straightforward, direct significance.

(e) Conclusion

The general argument of this paper has been that viewpoint is a substantial topic, one which it is possible and worthwhile to think about systematically. Two kinds of corollary should be considered in conclusion.

The first concerns implementation. To some extent the discussion above is concerned with clarifying points which *are* taken for granted in existing systems anyway. More systematically related ideas about processing are discussed by Binford (1981). His philosophy of expecting features to have a straightforward significance in the absence of other evidence seems generally appropriate to trading on viewpoint expectations. But it is not really possible to forsee how the habit of thinking about issues of viewpoint might affect practice. This is partly because reasoning about viewpoint is not really a self-contained body of theory, a point which is now considered.

The second corollary brings us back to the outlook which was offered at the beginning. The suggestion there was that we should as a matter of routine think in terms of viewing configurations instead of restricting our analysis to viewed configurations. Much of this paper may seem to make the opposite error of considering only viewpoint. However it should be clear that careful analysis of viewpoint cannot help being concerned with viewing configuration as a totality. In particular it cannot help being concerned with the way we classify and represent objects. It was made explicit at the start of the general viewpoint section that the ability to use the assumption is a function of the way objects *are* classified. That link between object classification and viewpoint recurs up to the last example in the section on privileged viewpoints, Marr and Nishihara's demonstration that the category of generalised cones is logically linked to assumptions about friendliness of viewpoint. If we want an analysis which exposes the regularities and opportunities associated with viewpoint, then we cannot simply solder it to an arbitrary analysis of the world. And, conversely, if we want to understand how things in the world may be classified and represented, we may do worse than to start by thinking about viewpoint.

REFERENCES

Binford, T. (1981) Inferring Surfaces from Images. Artificial Intelligence, 17, 205-244.

Binford, T. (1983) Figure/Ground : Segmentation and Aggregation in Physical and Biological Processing of Images, ed. Braddick, O.J. and Sleigh, H.C. (Springer-Verlag, Berlin Heidelberg New York).

Clowes, M. (1970 On Seeing Things. Arti fi cial Intel 1 igence, vol. 2, no. 19, 79" 16.

Cowie, R. (1982) Modelling People's interpretation of Line Drawings. D. Phil. Thesis, Universi ty of Sussex.

Draper, S. (1980) Reasoning about Depth in Line-Drawing Interpretation. D. Phil. Thesis, University of Sussex.

Hagen, M. (1980) (ed.) The Perception of Pictures, vol. II : Durer's Devices, Academic Press.

Huffman, D. (1971) Impossible Objects as Nonsense Sentences in Machi ne Intelligence 6, ed. Meltzer B. 6 Michie, D., Edinburgh : Edinburgh University Press.

Longuet-Higgins & Prazdny (1980) The interpretation of a moving retinal image. Proc. Roy. Soc. Lond. B 208, 385-397.

Mackworth, A. (1974) On the interpretation of drawings as three-dimensional scenes. D.Phil. Thesis, University of Sussex.

Marr (1977) Analysis of Occluding Contour. Proc. Roy. Soc. Lond. B 197, 441-475

Marr & Nishihara (1978) Representation and Recognition of the spatial organisation of three-dimensional shapes, Proc. Roy. Soc. Lond. B, 200 269-29.

Ullman, S. (1979) The interpretation of Visual Motion, Cambridge, Mass. : M.I.T. Press.