CHAPTER 9

Regularities of Nature: The Interpretation of Visual Motion

S. R. REUMAN
D. D. HOFFMAN
School of Social Sciences
University of California
Irvine, CA

1. Introduction

The human visual system can recover the three-dimensional structure of objects from their changing two-dimensional images formed on the retina. Cinematography provides a familiar example of this ability; each frame of a movie is two-dimensional, yet we have clear impressions of scene structure and depth. A particularly informative demonstration depicts a rotating sphere represented only by dots of light randomly positioned on its surface (Figure 1). To create the first frame of such a movie sequence, three-dimensional positions on the surface of the sphere are specified at random and then projected orthographically onto a two-dimensional image plane. (Orthographic projection is "parallel" projection; the z or depth coordinate is lost entirely while the (x,y) coordinates of each point remain unaltered.) Each subsequent frame, N, of the movie is then computed by rotating the dots from their initial three-dimensional positions $N*\theta$ degrees about some fixed axis, and then projecting onto the image plane. Given a fixed frame display rate for the sequence, the choice of θ determines the rotation speed of the sphere in three dimensions. When the frames are shown in sequence, subjects have little difficulty perceiving the correct three-dimensional structure and motion of the sphere over a wide range of dot densities and film rates, even though single frames give no impression of structure or depth. Demonstrations such as this make clear that our visual systems can recover three-dimensional structure from image motion in a highly reliable and flexible manner.

How is it possible that, in general, we can open our eyes to a novel moving scene and instantly understand the three-dimensional relationships of the objects we view? The inference from two dimensions to three is clearly underconstrained. In the dot displays there are an infinite number of possible interpretations since each image dot could, a priori, be assigned any z coordinate at all. What additional information then

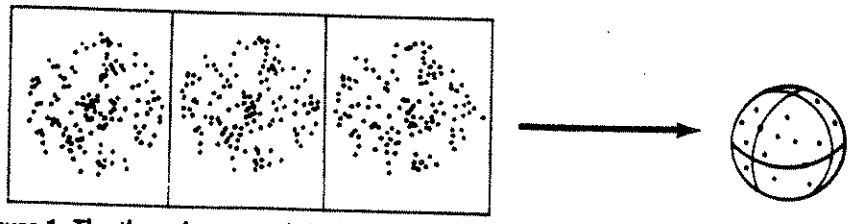


Figure 1. The three frames at left show the successive positions of dots when they are orthographically projected from the surface of a rotating sphere onto an image plane. When the frames are shown in rapid succession in their natural order, we recover the correct shape of the generating sphere, and perceive it in three-dimensions, despite the loss of information during projection.

might be brought to bear to reduce the number of possibilities to a unique solution, and more importantly, to the correct solution?

2. *Natural Regularities*

The additional constraining information needed to infer the three-dimensional state of the world from changing images is gained by exploiting regularities of nature—those patterns of the physical world which are highly stable over space and time. The number of possible visual interpretations of a scene can be reduced, often to a unique choice, by assuming that fundamental natural patterns such as the laws of physics hold for the objects being imaged. For this reason we expect that human visual machinery fully exploits such environmental regularities in solving the structure from motion problem.

2.1. *The Rigidity Regularity*

Ullman (1979a) noted the value of natural regularities in reducing or eliminating the inherent ambiguity of retinal scene analysis, calling them "reflective constraints" since they reflect properties of the physical world. Following several earlier researchers (Wallach & O'Connell, 1953; Gibson & Gibson, 1957; Green, 1961; Hay, 1966; Johansson, 1975) he suggested that rigidity is one regularity used by the human visual system to recover three-dimensional structure. The principle, as he formulated it, states

> *Any set of elements undergoing a two-dimensional transformation which has a unique interpretation as a rigid body moving in space, should be interpreted as such a body in motion.*

Ullman showed that by incorporating this rigidity principle, the visual system could indeed uniquely recover object structure from a sequence of images projected on the retina. His approach was to determine the minimum amount of visual information necessary to achieve a unique solution to three-dimensional object struc-

ture.[1] In fact, the amount of information required is quite small. Ullman's structure from motion theorem states

> *Given three distinct orthographic views of four non-coplanar points in a rigid configuration, the structure and motion compatible with the three views is uniquely determined up to a reflection about the image plane.*

The rigidity constraint as Ullman formulated it makes no explicit reference to spatiotemporal limits on its applicability. The assertion that objects in the environment can often be treated as locally rigid has some intuitive force. Yet, if an object must be perfectly rigid over a large region or extended period the plausibility and usefulness of the assumption is clearly reduced. Experiments confirm these expectations. If we construct a rotating dot display using dots widely separated in space or time, it will be found that the perception of depth and structure, so obvious in more dense displays, is entirely absent. If we then view a smaller twin of the same display sequence or step back further from the image plane, thereby lessening image (and retinal) inter-dot distances, the impression of depth returns. As yet there is no good definition of the locality within which rigidity must hold. Or, if it holds at all distances, there is no adequate explanation of why the recovery of structure from motion breaks down when the visual information is too widely separated in space or time. If normal viewing conditions and the spatio-temporal sampling rate of the human visual system are assumed, three views of four points is a quite weak requirement, and makes dependence on the rigidity principle more plausible.

A simple algorithm based on Ullman's structure from motion theorem involves testing local sets of four points for a unique interpretation as a rigid body moving in space, and combining successes. For all local groups interior to the image of an object, i.e., those whose elements belong to the same physical object, the scheme will be successful. Boundary points belonging to different objects are classified by testing them within a local group whose other members are known to lie within a single object. Local tests are independent and can be done in parallel. No backtracking is required during the process of combining successes since, by the structure from motion theorem, any successful interpretation of four points as a locally rigid body is unique.

Although Ullman's approach is quite powerful, there are displays which in principle it cannot account for, such as the "biological motion" displays of Johansson (1973). Johansson's displays are similar in concept to the dot displays already described, but focus on the more complicated patterns evident in human locomotion. Subjects are filmed in the dark with small light bulbs attached to joints

---

[1] This approach provides a methodological link between research on the perception of motion displays and research on the theoretical foundation of such abilities. By finding lower bounds on the spatio-temporal information that must be involved in early structure from motion computations, visual perception, as well as the computational theory requisite to building a successful simulation of the system, can be clarified.

(Figure 2). As in the case of the random-dot sphere, single frames of a biological motion display are completely unrecognizable and yet when shown in sequence, yield strong impression of three-dimensional structure and depth. However, Johansson's displays do not contain any group of four non-coplanar points moving in a rigid configuration. Here, the only rigid links lie between individual dot pairs such as ankle-knee, knee-hip, and so on. Thus the rigidity principle alone is inadequate and we are forced to conclude that additional (or alternative) regularities of nature must be brought to bear in this situation.

2.2. The Planarity Regularity

What is a reasonable candidate for an additional natural constraint that would explain the recovery of structure in biological motion displays? We would like to choose a constraint that provides high predictive power when applied to the image data.[2] The key observation here is that limbs in motion move in highly regular paths. Casual observation reveals that the limbs of an ambulating animal do not move about arbitrarily. During normal gait, limbs typically swing in a plane for extended periods of time. This constraint on limb motion can be formalized as a "planarity principle" (Hoffman & Flinchbaugh, 1982), and tested for its predictive power in explaining the recovery of structure in biological motion displays. The planarity principle claims that the visual system exploits planar limb motion by implementing the following:

> *Any set of elements undergoing a two-dimensional transformation which has a unique interpretation as a rigid structure moving in one plane, should be interpreted as such a body in motion.*[3]

Hoffman and Flinchbaugh (1982) show that the planarity regularity can be used to prove the following two "structure from planar motion" propositions:

Proposition 1. *Given three distinct orthographic projections of two points constrained to rotate (and translate) rigidly in a plane, the structure and motion compatible with the three views are uniquely determined up to a reflection about the image plane.*

---

[2] Note that the issue is more complicated than simply choosing the most widely applicable regularity we can think of. While the intuitive plausibility of a regularity is closely tied to how widely it applies in the physical world, its generality is not logically equivalent to its predictive power. This subtle difference is critical to the principled search for regularities actually incorporated in the visual system. Later in the chapter, this point is explained in detail by formalizing the inference process in terms of Bayes' Theorem.

[3] The planarity principle is actually a consequence of more fundamental natural regularities stemming from Newton's laws of motion and the properties of mass. While we do not yet have a satisfactory form for the underlying principle, the idea can be illustrated by noting that to a first approximation, limbs act like simple, rigid pendula doing work under the force of gravity. In the simplest case, obedience to Newton's laws implies that the pendulum will travel in a planar path. The point is that physical principles act as the fundamental natural regularities enhancing the strength of visual inferences. In the text, the regularity is stated in terms of planarity in order to restrict attention to particular displays of concern.

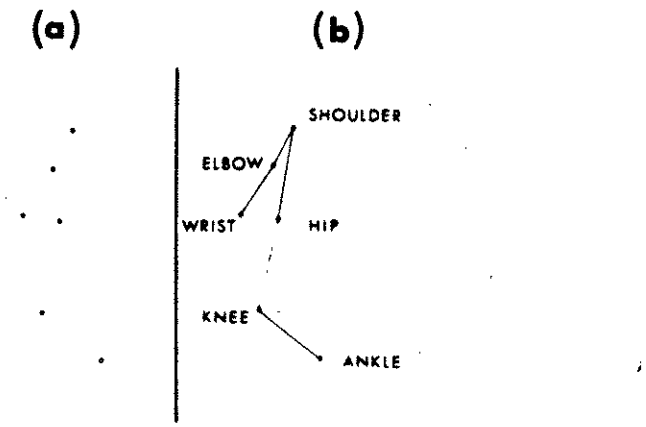


Figure 2. An example of a motion display in which the rigidity principle alone is insufficient to explain our perception of depth and structure. The display does not contain any group of four non-coplanar dots moving in a rigid configuration, the minimum condition required by the structure from motion theorem. (a) A single frame of a typical motion sequence. (b) The correct links between dots which must be discovered to interpret the display correctly.

Proposition 2. *Given two distinct orthographic projections of the three endpoints of two rigid rods linked in a hinge joint to form a pairwise-rigid structure which is constrained to move in one plane, the structure and motion compatible with the two views are uniquely determined. (A pairwise-rigid structure is a set of points moving in space so that each point remains at a constant distance from at least one other point, and no three points are in a rigid configuration. Arms and legs are obvious examples.)*

We can sketch the proof of proposition 1 as follows. Let O_i and A_i be the two points in frame i, for $i=1,2,3$. In 3-space, the vectors $\mathbf{a}_i$ from O_i to A_i have equal length and are coplanar under the rigidity and planarity constraints. Note that translation of O_i in depth or in the image plane can be ignored without loss of generality. For translation in depth, recovery is impossible under orthographic projection,[4] while for translation parallel to the image plane, recovery is trivial. Therefore we can set $O_1 = O_2 = O_3 = O$ (figure 3a). In 3-space, the vectors $\mathbf{a}_i$ from O to A_i have equal length and are coplanar under the rigidity and planarity constraints. Thus we have

$$\mathbf{a}_1 \cdot \mathbf{a}_1 = \mathbf{a}_2 \cdot \mathbf{a}_2 \tag{1}$$

$$\mathbf{a}_1 \cdot \mathbf{a}_1 = \mathbf{a}_3 \cdot \mathbf{a}_3 \tag{2}$$

[4] The use of orthographic projection (as opposed to perspective projection) has been justified elsewhere on the basis of locality arguments. See Ullman (1979a), p. 158.

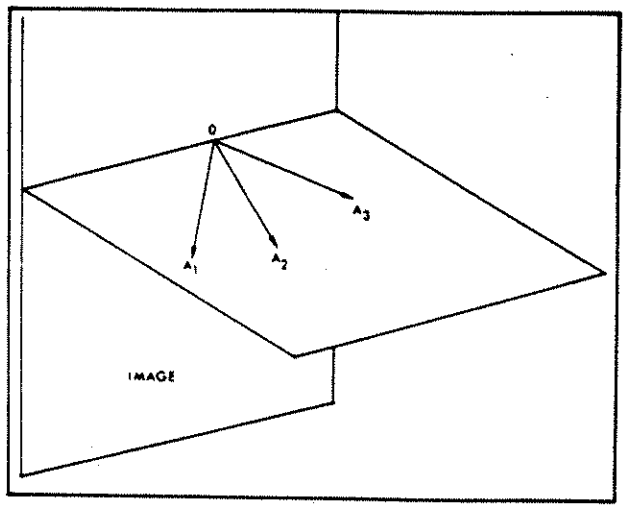


Figure 3a. The geometry underlying three views of two rigidly linked points constrained to move in a plane.

by the rigidity principle and

$$\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3) = 0 \tag{3}$$

by the planarity principle. As in the case of the rigidity principle, the planarity principle need only hold locally under normal viewing conditions, since the number of views and points required to obtain a unique three-dimensional structure is small. Recall that under orthographic projection, x and y coordinates are preserved while z coordinates are lost entirely. In this manner, the problem is reduced to recovery of the relative z coordinates of one point in each of the three frames, given the corresponding x_i and y_i. So, by eqns. (1), (2) and (3) we have

$$\begin{aligned} z_1^2 - z_2^2 + k_1 &= 0 \\ z_1^2 - z_3^2 + k_2 &= 0 \\ k_3 z_1 + k_4 z_2 + k_5 z_3 &= 0 \end{aligned} \tag{4}$$

where

$$\begin{aligned} k_1 &= x_1^2 + y_1^2 - x_2^2 - y_2^2 \\ k_2 &= x_1^2 + y_1^2 - x_3^2 - y_3^2 \\ k_3 &= x_2 y_3 - x_3 y_2 \\ k_4 &= x_3 y_1 - x_1 y_3 \\ k_5 &= x_1 y_2 - x_2 y_1 \end{aligned} \tag{5}$$

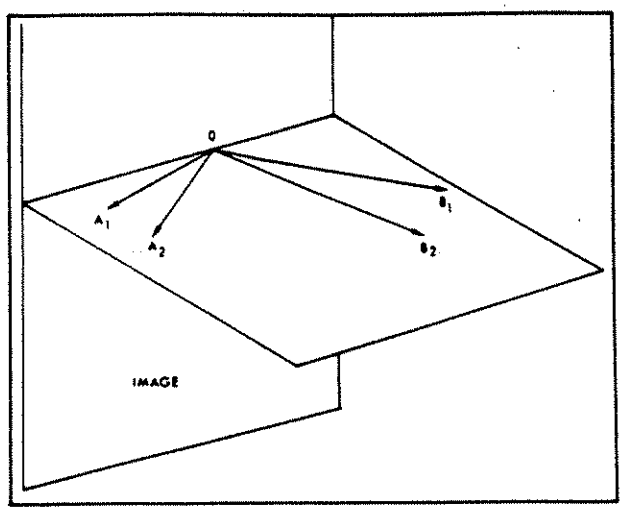


Figure 3b. The geometry underlying two views of three points linked in pairwise-rigid manner (see Proposition 2) and constrained to move in a plane.

The details of the proof that there is a unique solution to the above system can be found in Hoffman and Flinchbaugh (1982). The proof of proposition 2 is similar.

What are the potential difficulties with the above scheme? First, we must be sure of uniqueness. That is, if the object of concern is actually rigid and in planar motion, then an analysis based on these properties should deliver only one three-dimensional interpretation.[5] This is the problem of "phantom structures" (Ullman, 1979a). The proof that there can be no phantom structures follows from the proof of propositions 1 and 2 above. Second, one can ask, "What is the probability that the analysis will return a rigid, planar interpretation of a set of points if in fact the points are not moving in such a configuration?" This is the "false targets" problem. The probability of false targets is low and in fact is zero given images of infinite resolution.[6] Recall that three views of two points moving rigidly in a plane in space map to six points on an ellipse in the image. (Again, as in the uniqueness proof of Proposition 1, we can ignore translation in depth and in the image plane.) Only five points are needed to specify an ellipse. The probability that a randomly chosen sixth point lies on a given ellipse is zero and thus the probability of false targets is zero.

2.3. *The Fixed Axis Assumption*

By exploiting certain natural regularities, a visual system operating mechanistically, that is, without benefit of higher-level knowledge, can solve the problem

---

[5] Under orthographic projection, we are limited to uniqueness up to reflection about the image plane.

[6] The probability of false targets given infinite image resolution is also zero when the rigidity principle alone is used to recover depth from three views of four points.

of recovering three-dimensional object structure in an interesting and practical class of scenes. If the rigidity principle alone is assumed, three views of four non-coplanar points are required. If the planarity principle is invoked in addition, three views of two rigidly moving planar points, or two views of three such points linked in pairwise-rigid fashion will suffice. But what of scenes in which even these small requirements are not met? Given a display of fewer than four points moving rigidly in a non-planar path (thereby bypassing both the structure from motion and planarity theorems) can the human observer yet recover the relative depth of the points?

This is not mere pedantry. Our goal is essentially to find the smallest possible complete set of regularities—complete in the sense that the combined predictive power of the members should rival the structure recovery abilities of the human visual system. Johansson has shown (1975) that a display of two points moving opposite each other through rectangular or elliptical paths (Figure 4) is sufficient to create the impression of depth. However, he did not attempt to restrict the number of views to the minimum necessary to make such a judgement. We have constructed a display in our laboratory using only three frames of three dots rotating about a vertical fixed axis. The duration of the display can be made arbitrarily long by repeating the sequence (1-2-3-2-1-2-3 . . .), making judgement about shape more

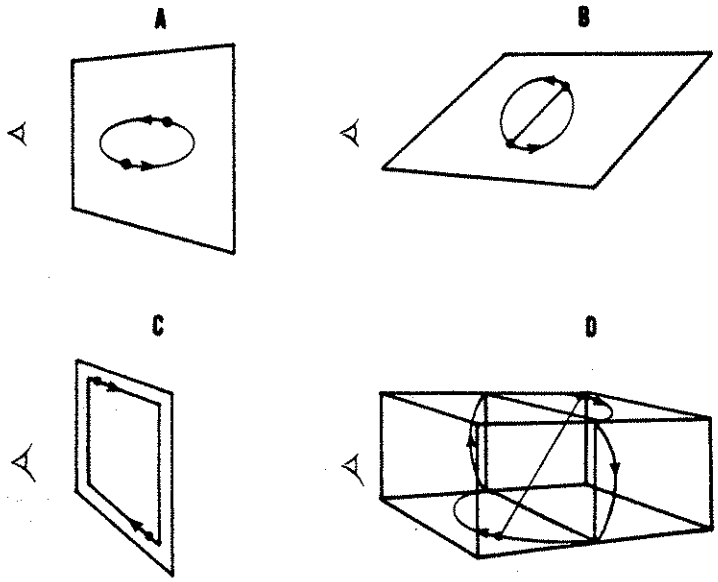


Figure 4. Johansson's displays showing depth-from-motion with only two dots. Two dots moving in an elliptical path in the image (a) are perceived as maintaining a constant distance from each other and thus seem to be moving in a circle in depth (b). When the two-dimensional path is a square (c), the result is the perception in (d). Adapted from G. Johansson, *Visual Motion Perception;* in "Recent Progress in Perception", p. 71, Freeman Press, 1976.

feasible while still restricting information input to three frames. Depth and structure can still be seen clearly when such a sequence is viewed by a naive observer. Thus again we are led to the conclusion that our model is inadequate; rigidity and planarity alone cannot explain the general recovery of depth and structure from image data.

A few researchers have sought explanations for our ability to perceive depth in these very sparse displays through the use of a fixed axis assumption. Webb and Aggarwal (1981) studied this idea but did not address the question of the minimal information necessary to arrive at a unique interpretation. Bobick (1983) also examined the fixed axis assumption. He showed that three views of two points, when coupled with knowledge of the instantaneous direction vectors of the points and the fixed axis assumption, are sufficient to specify a unique structure. Bennett and Hoffman (in preparation) have recently shown that three views of three points are sufficient to recover object structure and motion uniquely, for rigid, fixed-axis motion. Bennett and Hoffman (1985) have also found that three views of four points or four views of two points in non-rigid, fixed-axis motion are sufficient to recover object structure uniquely and with zero probability of false targets. Thus under the fixed axis assumption, it is possible to compute the three-dimensional structure of objects without relying on rigidity at all.

The work of Green (1961), using rotating dot displays, has been cited as evidence that the fixed axis assumption is justified from a psychophysical point of view. Green examined various parameters influencing perceived rigidity of rotating objects, among these the direction and motion of the axis of rotation. He found that a (projected) random-dot figure tumbling about a point[7] was perceived as less rigid than one moving about a vertical fixed axis, indicating that such an axis does at least make it easier for the visual system to recover the structure of the object in question. However he also found that rotation about various fixed but non-vertical axes was perceived as less rigid than rotation about a vertical fixed axis (Green 1961, experiments 5 and 6). In fact, there were some instances (objects consisting of random unconnected line segments) in which perceived rigidity of non-vertical fixed axis rotation was *lower* than that associated with the tumbling display. It thus seems that the situation is more complicated than the fixed axis assumption can easily explain. Orientation of the axis of rotation, particularly vertical vs. non-vertical, as well as the structure of the display (lines vs. dots in Green's experiments) are both confounding variables. Indeed, Green's work could be interpreted as providing evidence against the fixed axis assumption, since the perception of depth and three-dimensional structure still occurred in displays with moving axes, albeit with less clarity. We have duplicated some of Green's displays in our laboratory and have found similarly that complex axis motion reduces but does not destroy the ability to recover depth and structure.

[7] To tumble a figure about a fixed point, Green used Euler's formulas to specify rotation speed around each of the three coordinates axes.

2.4. *A Regularity Based on Mass*

As an alternative to the fixed axis assumption, we are currently examining the use of a natural regularity based on the fact that objects in the real world have mass and therefore momentum. By assuming that image data stems from real, physical, massive objects, the visual system can at once make use of the enormous predictive power of the laws of physics, in particular, the laws of motion.

Our choice of this approach has been motivated by the fact that objects in the real world always have mass and thus their images can be expected to be interpreted using that knowledge. The goal of the visual system is to make useful inferences about the physical world. Utility implies manipulation and prediction of position as well as recovery of three-dimensional structure. Therefore endowment of the imaged object with mass and momentum, should enhance the ability to make accurate and useful inferences about the underlying physical reality.

Assuming that objects have mass may explain the observation that rotational motion in dot displays must be smooth for the structure recovery process to operate successfully. We have constructed in our laboratory a moving dot display in which the arc distance of axis movement per frame is constant while rotational direction of the axis is altered 45 degrees every third frame. In such displays where dot motions are jerky, structure from motion falls apart, and yet the fixed axis assumption would predict correct interpretation of such sequences. Interpretation based on the assumption that the object has momentum handles this situation naturally.

We suggest that the visual system exploits knowledge of mass and momentum by implementing the following statement:

> *Any set of elements that has a unique interpretation as a moving, rigid body with nonzero mass should be interpreted as such a body.*

The mass regularity implies obedience to Newton's laws for the motion of a rigid body in space

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}$$

$$\frac{d\mathbf{L}}{dt} = \mathbf{N}$$

where the linear momentum $\mathbf{P}$ is defined by $\mathbf{P} = M\mathbf{V}$, the angular momentum $\mathbf{L}$ is defined by $\mathbf{L} = \mathbf{I}\cdot\mathbf{W}$, $\mathbf{F}$ and $\mathbf{N}$ are the total external force on the body and the total torque about a suitable point O, $\mathbf{V}$ is the velocity of the center of mass, and $\mathbf{I}$ and $\mathbf{W}$ are the inertia tensor and the angular velocity about the point O (Symon, 1971). $\mathbf{L}$ is usually measured with O at rest or at the center of mass for unconstrained bodies moving in space. In the case of rotational motion, conservation of angular momentum as expressed above states that when the resultant external torque acting on the body is zero, the total vector angular momentum of the body remains constant. We are currently engaged in determining the implications of the exploitation of this principle by the visual system in the interpretation of rotational displays.

One obvious question arises from dependence on the mass regularity. How can perceptual reversal of rotational direction be explained? That is, how is it possible that a clearly perceived, rotating random-dot sphere could instantaneously reverse direction if it were assumed to have momentum during the structure recovery process? This may or may not be a serious objection. A great deal of research has concentrated on proving that no change in the retinal image is necessary for perceptual reversal to occur, but the precise cause of perceptual reversal is still obscure. It is clear at least that the existence of a preference for smooth motion caused by momentum is not necessarily excluded. One possibility stems from noting that an initial assumption of momentum necessarily leads to unidirectional smooth motion only if the path of each dot is predicted, computed, and tested continuously with each frame. If this continuous predict-compute-test cycle were interrupted for any reason, a bottom-up system would "forget" the previous direction of motion and initiate again, in an arbitrary direction, using the new incoming data. Another possibility is that such ambiguous demonstrations lead to two interpretations being computed simultaneously, each one consistently "seeing" motion in one direction, but alternately admitted to consciousness. Switching one or the other interpretation into consciousness could happen for a variety of reasons having little or nothing to do with the method of structure recovery. Rigorous establishment of the meaning of reversal in these displays requires strict control of a number of factors (e.g., eye movements) which may be confounding the effect of a regularity underlying system interpretation of the display. It will be necessary to design experiments with this in mind to settle the issue.

3. *Inference: The Link Between Regularities and Physical Structure*

We have considered the importance of seeking plausible regularities in the natural world that might assist the visual system in solving the structure recovery problem. The key to their plausibility lies in the idea that the goal of vision is to infer useful properties of the physical world. We want to use regularities which convey the greatest "predictive power" possible. How can we determine precisely what this predictive power stems from? How do we know if a particular regularity will provide the power to make our inferences correct? Before answering this question it will be helpful to discuss the categories of reasoning—inductive and deductive—into which inference can be divided, and to clarify the difference between them.

An argument is deductively valid if and only if it is logically impossible that its conclusion is false while its premises are true. It is characteristic of all deductively valid arguments that the conclusion must be contained in the premises. Most people are familiar with simple examples such as

All mammals are animals.
All skunks are mammals.

Therefore, all skunks are animals

An inductive argument differs from a deductive argument in degree of strength only, and should not be thought of as a different "type" of argument. An inductively strong argument is defined as one which is not deductively valid, and whose conclusion is probable given that the premises are true. The degree of inductive strength depends on the degree of probability involved. Note that because the argument is probabilistic, the conclusion, unlike in the deductive case, could still be incorrect even when inductive strength equals 1.

As another illustration of the difference between inductive and deductive reasoning, observe the fallacies in the following two common misconceptions. First, it is sometimes thought that deductive arguments proceed from general to specific while inductive arguments proceed from specific to general. Actually, deductively valid arguments can proceed in all four possible ways, from general to general, general to specific, specific to general, and specific to specific (Skyrms, 1975). Inductively strong arguments have similar versatility. The difference between deductively valid and inductively strong arguments lies not in the domain of their potential application but in the different standards by which they judge validity. Only in the deductive argument must the facts comprising the conclusion be contained in the premises. A second misconception is the idea that inductive arguments necessarily imply the presence of consciousness, i.e., they cannot be implemented mechanically as is possible for deductive arguments. This is simply incorrect. To implement inductive reasoning on a computer, it is only necessary 1) to provide a means of learning from experience, or 2) to program in our own experience by basing the computer's decisions on natural regularities which will increase the inductive strength of its inferences.

The logical framework within which the visual system must operate is inherently inductive. We are given (x,y) coordinates and must recover the corresponding z coordinates of each point. Since the facts comprising the conclusion of a deductively valid argument must be stated explicitly in the argument premises, and the z coordinates are indeed not given, it is impossible to recover the z coordinates via a deductively valid argument. Hence we are constrained to maximize the degree of inductive strength of arguments used in attempting the recovery.

If it were possible to use a deductively valid argument and ensure the success of the recovery process, we might hope to decode a conclusion (three-dimensional structure) already somewhere available in the "premises" provided by the image information and the various constraining assumptions. If this indeed were possible, it would suggest an entirely different approach to the recovery (and the research) process. In fact it would be far more satisfactory to use such an approach, since so much more is known about the rules and structure of deduction than induction. However, this is simply not possible and we must study the inductive strength of possible reconstruction schemes.

Now we are ready to analyze the roots of predictive power in inductive reasoning. What is the key to inductive strength? An example using Bayes' Theorem will clarify the answer. Let w (world) stand for the following assertion about four points in the world: "are in rigid motion in three-dimensions". Let i (image) stand for the following assertion about the retinal images of the same four points: "have two-

dimensional positions and motions consistent with being the projections of rigid motion in three-dimensions". Then what is the probability of w given i? The existence of rigid objects does not in itself make this conditional probability high. Using Bayes' Theorem we find that

$$P(w|i) = \frac{P(w) \cdot P(i|w)}{P(w) \cdot P(i|w) + P(\sim w) \cdot P(i|\sim w)}$$

Since the numerator and the first term of the denominator are identical, $P(w|i)$ is near one only if

$$P(w) \cdot P(i|w) >> P(\neg w) \cdot P(i|\neg w)$$

And since $P(\neg w)$, though unknown is certainly much greater than zero, $P(w|i)$ is near one only if $P(i|\neg w)$ is near zero. That is, only if $P(i|\neg w)$ is near zero can the inference from the image to the world be inductively strong. But what is $P(i|\neg w)$? It is precisely the "false targets" probability (FTP) encountered in the proof of Ullman's structure from motion theorem earlier. Ullman found that with three views of three points the FTP is one, but with three views of four points it is near zero. Similarly, a major goal of other "structure from motion" proofs (Hoffman & Flinchbaugh, 1982; Bobick, 1983; Longuet-Higgins & Prazdny, 1981; Richards, Rubin, & Hoffman, 1983) is to determine under what conditions the FTP is near zero.

The answer to our question about the root of predictive power in visual inference is the FTP. Is there any more that can be said about this probability? In particular, can we obtain guidance in the search for plausible regularities? That is, what can the FTP tell us about selecting a regularity for incorporation in a model visual system? One observation is that a necessary precondition to obtaining a near-zero FTP is an overconstrained system of equations. There must be at least one more equation than unknowns (an inconsistent set). However, this is not much help when the regularity providing the equations has not yet been chosen! Another moment of thought tells us that $P(w)$ should not be zero, since this will leave $P(w|i)$ equal to zero. But this does not narrow our choices much, since phenomenologically at least, there are many regularities that seem general in space and time. Unfortunately, it turns out that the problem of choosing regularities presents deep problems for logic and the philosophy of science that have never been adequately solved. While the FTP provides a post hoc means of evaluating a chosen regularity, it can tell us little about why we should choose one regularity over another, and we must rely on intuition for guidance.

4. *A Model: Rigidity and Planarity as Regularities*

The basic problem with the formulation of the rigidity and planarity principles as presented so far is that they make no allowance for minor real world violations of their mathematical idealizations. There are at least three sources of such violations.

First, noise is introduced during transmission of the visual waveform and during detection at the cornea, lens, and retinal levels. Second, an animal must deal with minor deviations from the principles. For example, it is necessary to overlook the minor deviations from planarity in limb motion in order to use the planarity principle (though it is of course possible that an animal searches for and knowingly puts such deviations to use in object recognition). The third source of inaccuracy stems from the difficulty of identifying real world feature points. Corners are rarely sharp under close inspection. When we look at a biological motion display that replaces the knee with a single well-defined point of light, consistent identification of that point over time is relatively simple. In an actual lighted scene of a human walking however, the knee has an extended surface which could be tagged at several places in attempting to establish links of constant length between thigh and ankle.

Assume that the visual system receives a stimulus and reconstructs the three-dimensional object using only the rigidity and planarity constraints. How can it best deliver output given its inaccurate inputs? Or, as Marr (1982) has put it, how can it maintain the principle of "graceful degradation"? Surely it cannot afford to fail completely in making a judgement with respect to its goals, but rather must in some sense provide the most rigid or the most planar interpretation possible. That is, the system could (1) assume strict rigidity but relax its planarity constraint to fit the input data, (2) assume strict planarity but relax its rigidity constraint to fit the data, or (3) allow the minimum relaxation of both rigidity and planarity necessary to fit the data.

We have constructed a computer model implementing the first two alternatives within a constrained optimization paradigm. For each alternative, we examined the two cases noted in Proposition 1 and 2 earlier (three views of two rigidly linked points or two views of three pairwise rigidly linked points) given different levels of added gaussian noise. The model has two goals. First, it was designed to enable comparison of structure recovery strategies in the face of noise for the two cases above—relaxing planarity while maintaining rigidity, and relaxing rigidity while maintaining planarity. Second, we were interested in evaluating a particular algorithm (see below) for the practical computation of structure from motion.

Addition of varying degrees of gaussian noise was simulated as follows. On each trial, image data ((x,y) coordinate pairs computed from an exact parallel projection of a three-dimensional dot configuration) were radially perturbed by gaussian noise, truncated at three standard deviations from the original points. The standard deviation was a parameterized percentage of the two-dimensional distance between accurate data values. Gaussian noise was generated by a simple algorithm based on the central limit theorem:

$$y = \frac{\Sigma_i^K X_i - \frac{K}{2}}{\sqrt{K/12}} \tag{6}$$

where the X_i are uniformly distributed pseudo-random numbers between 0.0 and 1.0. As K approaches infinity, the random variable y approaches a perfect normal

distribution. For purposes of the model, we set $K = 100$, and varied the standard deviation as necessary. Each perturbed data point was computed independently from (6). The perturbed data points were then used as input to our algorithm, with one hundred trials per perturbed data set computed for each of eleven standard deviations of noise. The mean absolute value of the triple product and the mean difference in vector length between successive views of individual points were used as measures of relative planarity and rigidity respectively. The implementation recovers the optimal relative z coordinates from the perturbed (x,y) coordinates by minimizing rigidity (planarity) subject to the condition of maintaining strict planarity (rigidity).

It is important to use an algorithm that could in fact be accomplished by neural computations, even though we do not claim that the chosen algorithm resembles that underlying the neural implementation of depth recovery. Computation by neural networks is usually taken to have at least two characteristics: (1) a high degree of parallelism, and (2) locality of processing. Locality here refers to the general principle that computation is distributed among a large number of relatively simple "processors", each of which performs some relatively independent sub-computation. We have chosen an algorithm which supports these principles and provides a means of implementing the optimization in parallel by a network of simple processors, each connected only locally to its N nearest neighbors.

The basis of the algorithm is the following (see Ullman, 1979b). Given a differentiable function of n variables $f(\mathbf{x})$, it can be shown that motion of x_i in the direction of $f'(x_i) = \delta f/\delta x_i$ is convergent to the local unconstrained maxima of f. Such a process can be realized in a network of processors by assigning the computation of $f'(x_i)$ to the i^{th} processor p_i, where p_i has access to the minimal subset of arguments necessary to compute $f'(x_i)$. This idea is made useful in the constrained optimization case by a theorem due to Kuhn and Tucker (1951) that established equivalence between the constrained maxima of a function and saddle points of its associated Lagrangian. Saddle points of a Lagrangian can be computed by the gradient method noted above and thus by a network of simple processors similar to the unconstrained case. The minimization case is included since minimizing $f(\mathbf{x})$ is equivalent to maximizing $-f(\mathbf{x})$. Thus (for the minimization case), given a differentiable objective function $f(\mathbf{x})$ of n variables subject to m constraints of the form $g_i(\mathbf{x}) \geq 0$, the Lagrangian associated with the problem is a function of $n+m$ variables defined as

$$L(\mathbf{x},\mathbf{u}) = f(\mathbf{x}) - \Sigma u_i g_i(\mathbf{x}) \tag{7}$$

where $\mathbf{u}$ is an m-vector of chosen multipliers. A non-negative saddle point of the Lagrangian is a point $(\mathbf{x}',\mathbf{u}')$ such that

$$L(\mathbf{x}',\mathbf{u}) \leq L(\mathbf{x}',\mathbf{u}') \leq L(\mathbf{x},\mathbf{u}') \tag{8}$$

for every $\mathbf{x} \geq 0, \mathbf{u} \geq 0$.

Arrow, Hurwicz, and Uzawa (1958) investigated the application of Kuhn and Tucker's (1951) theorem to constrained optimization. They defined a set of differential equations describing motion with respect to local gradients toward the saddle point optimum which can be implemented iteratively by the recurrence relation

$$x_i^{n+1} = max[0, x_i^n - pL_{x_i}]$$
$$u_i^{n+1} = max[0, u_i^n + pL_{x_i}] \tag{9}$$

where p is a selected step size and L_{x_i} and L_{u_i} are partial derivatives of the Lagrangian. Arrow et al. proved convergence theorems with relevance to linearity, continuity, and convexity. Briefly, the gradient method described above can be proved to converge to a global minimum for the continuous case if $f(\mathbf{x})$ and all the $g_i(\mathbf{x})$ are convex. Where f is not convex, convergence can be proven within a sufficiently local neighborhood of the minimum point. A modified version of the method can be used to convert linear objective functions to non-linear convex counterparts to ensure convergence in the continuous linear case. An implementation by Marschak (chap. 9, Arrow, 1958) showed convergence in the discrete linear case using the modified method. His work uncovered useful practical lessons which have been reiterated in our experience with the non-linear case. These lessons are discussed in detail in the next section.

4.1. *Results: The Model*

Listed below are the Lagrangians associated with the four cases we have considered. The k_i's are constant expressions in (known) x and y coordinates. For cases 2 and 4, (two views of three points) z_{ij} refers to the z coordinate of vector i in view j (Figure 3b).

CASE 1: Allow relaxation of planarity in the face of noise, but require perfect rigidity. Three views of two points.

$$L_1(\mathbf{z},\mathbf{u}) = (k_3z_1 + k_4z_2 + k_5z_3)^2 - u_1(z_1^2 - z_2^2 + k_1) - u_2(z_1^2 - z_3^2 + k_2) \tag{10}$$

CASE 2: Allow relaxation of planarity in the face of noise, but require perfect rigidity. Two views of three points.

$$L_2(\mathbf{z},\mathbf{u}) = (k_3z_{a_1} + k_4z_{b_1} + k_5z_{a_2})^2 + (k_6z_{a_1} + k_7z_{b_1} + k_8z_{b_2})^2 - u_1(z_{a_1}^2 - z_{a_2}^2 + k_1) - u_2(z_{b_1}^2 - z_{b_2}^2 + k_2) \tag{11}$$

CASE 3: Allow relaxation of rigidity in the face of noise, but require perfect planarity. Three views of two points.

$$L_3(\mathbf{z},\mathbf{u}) = (z_1^2 - z_2^2 + k_1)^2 + (z_1^2 - z_3^2 + k_2)^2 - u_1(k_3z_1 + k_4z_2 + k_5z_3) \tag{12}$$

CASE 4: Allow relaxation of rigidity in the face of noise, but require perfect planarity. Two views of three points.

$$L_4(\mathbf{z},\mathbf{u})=(z_{a_1}^2-z_{a_2}^2+k_1)^2+(z_{b_1}^2-z_{b_2}^2+k_2)^2-u_1(k_3z_{a_1}+k_4z_{b_1}+k_5z_{a_2})-u_2(k_6z_{a_1}+k_7z_{b_1}+k_8z_{b_2}) \tag{13}$$

These four cases could have been implemented as is by the gradient method we have described. However, we chose to modify the basic algorithm slightly. Note that in the appropriate cases, maintenance of perfect planarity or rigidity allows us to set all the Lagrange multiplier terms to zero. Since the constraint equations from which these terms are derived are to be maintained at zero by our model "visual system", there is no need to iteratively approximate the u_i's. This simplification reduces processing time. The same idea suggests solving for as many z_i's as possible in the constraint equations and substituting into the corresponding objective function. In this way, n variable objective functions each subject to m constraint equations were reduced to functions of $n-m$ variables. This lessens further the time necessary to compute and follow gradients. (We are indebted to T. Indow for discussions on the above points.)

These modifications do not alter the Arrow-Hurwicz gradient method fundamentally. If planarity (say) is to be maintained perfectly, there is no need to iterate with respect to the planarity "dimension". The system need only have the $n-m$ processors necessary to minimize our objective function with respect to rigidity. In effect, we have made those n-m processors more powerful by incorporating the planarity requirement into them.

Practical considerations also suggest employing the modified algorithm. In addition to reducing processing time, the most important practical benefit of the modified algorithm is the reduction in information necessary to begin the iteration process. In the naive approach using all z_i and u_i, an initial value is necessary for each of the $n+m$ variables. If little prior knowledge is available to motivate this choice, it becomes a serious obstacle to ensuring convergence in the non-convex case. Even in the convex case, initial values too far from the convergence point can lead to erroneous results because of range or accuracy restrictions on the particular machine being used. With the modified method, we reduced the information requirement by $2m$ initial values, thereby reducing startup difficulty.

The objective functions to be minimized now become

CASE 1:

$$L(z_1)=(k_3z_1+k_4(z_1^2+k_1)^{.5}+k_5(z_1^2+k_2)^{.5})^2 \tag{14}$$

CASE 2:

$$L(z_{a_1},z_{b_1})=(k_3z_{a_1}+k_4z_{b_1}+k_5(z_{a_1}^2+k_1)^{.5})^2+(k_6z_{a_1}+k_7z_{b_1}+k_8(z_{b_1}^2+k_2)^{.5})^2 \tag{15}$$

CASE 3:

$$L(z_1,z_2)=(z_1^2-z_2^2+k_1)^2+(z_1^2-((k_3z_1+k_4z_2)/k_5)^2+k_2)^2 \tag{16}$$

CASE 4:

$$L(z_{a_l},z_{b_l})=(z_{a_l}^2-((k_3z_{a_l}+k_4z_{b_l})/k_5)^2+k_1)^2+(z_{b_l}^2-((k_6z_{a_l}+k_7z_{b_l})/k_8)^2+k_2)^2 \tag{17}$$

Table 1 lists all the statistics collected on 100 trials for each of 11 levels of noise for each of the four cases. A single set of three 3-space vectors was used as unperturbed data throughout all trials reported for cases 1 and 3. Another set consisting of the above set with an added fourth vector was used as unperturbed input data for cases 2 and 4. Figures 5–9 plot each statistic graphically against increasing noise. Figure 5 shows the proportion of trials converging to a solution as noise increases. When the standard deviation of noise is below 5–8% of the average two-dimensional distance between unperturbed points, results are quite stable, with all trials converging as expected. Figures 6, 7, and 8 compare mean triple product magnitudes (a measure over all trials at a given level of noise, of the relative deviation from planarity), mean vector lengths (a measure of the relative depth from the origin assigned to the two-dimensional data points by the algorithm), and mean differences in length between successive views of individual point vectors (a measure of relative deviation from rigidity). When planarity is relaxed (cases 1 and 2), there is an increase in both mean vector length and mean triple product magnitude as noise increases over 5%. When rigidity is relaxed (cases 3 and 4), up to 5% noise can again be handled, but the increase in mean vector length is no longer observed. The mean triple product magnitude stays lower as expected, while the deviation from rigidity is higher and increases with noise.

The drop in mean triple products and vector lengths at the highest levels of noise is not due to a real effect on these variables but to the increase in trials with no solution—only those trials with relatively unperturbed data survived to converge and be included in the statistics. (This conclusion was made by comparing the average perturbation of input data from converging trials with that from degenerate trials.)

4.2. *Results: The Algorithm*

Our experience with the speed of the gradient algorithm parallels that of Marschak (Arrow, 1958, chapter 9). Even the simplified method we have used can result in uselessly slow convergence at higher noise levels. Figure 9 shows that the number of iterations[8] required increases 3–4 orders of magnitude as noise increases

[8] This is the number of iterations per *set* of new z iterates. Thus, parallel implementation of these computations with $n - m$ processors would not eliminate the problem.

Table 1. Statistical Comparison of Algorithmic Recovery of Structure from Motion for 11 Levels of Gaussian Noise. The Four Cases Are Defined and Described in the Text. Measures Used Were MTP (Mean Triple Product Magnitude), MVL (Mean Vector Length), MR (Mean Rigidity = Mean Absolute Difference in Frame-Specific Vector Lengths), N (Number of Trials Converging to a Solution), and I (Mean Number of Iterations to Convergence). For MTP, MVL, and MR, Standard Errors Are Listed with the Statistic (Stat ± S.E.).

% Noise	Stat.	CASE 1	CASE 2	CASE 3	CASE 4
0	MTP	.207±5.49e-5	.202±5.49e-5	.207±5.49e-5	.202±6.94e-5
	MVL	1.00±0	1.00±5.11e-4	1.00±3.48e-4	1.00±4.38e-4
	MR	0	7.45e-9	1.18e-4±3.59e-8	1.89e-5±4.74e-9
	N,I	100, 1	100, 138	100, 14	100, 78
.125	MTP	.207±1.03e-3	.202±8.50e-4	.207±7.91e-4	.203±8.50e-4
	MVL	1.00±6.13e-3	1.00±5.42e-3	1.00±4.87e-3	1.00±5.54e-3
	MR	5.14e-9±6.92e-9	6.85e-9±7.08e-9	5.04e-4±1.66e-4	2.16e-5±7.70e-6
	N,I	100, 694	100, 674	100, 1047	100, 220
.250	MTP	.207±2.30e-3	.203±1.96e-3	.207±2.07e-3	.203±1.96e-3
	MVL	.999±1.44e-2	1.00±1.05e-2	.999±1.32e-2	1.00±1.06e-2
	MR	5.29e-9±7.44e-9	7.75e-9±8.06e-9	5.41e-4±1.30e-4	1.99e-5±5.57e-6
	N,I	100, 1624	100, 1313	100, 3070	100, 381
.50	MTP	.207±5.37e-3	.202±2.99e-3	.207±5.13e-3	.202±3.00e-3
	MVL	1.00±3.09e-2	1.00±1.98e-2	1.00±2.96e-2	1.00±1.98e-2
	MR	7.00e-9±8.53e-9	7.90e-9±7.55e-9	5.37e-4±7.72e-5	1.94e-5±4.16e-6
	N,I	100, 2851	100, 2131	100, 6453	100, 569
1.0	MTP	.207±1.00e-2	.203±6.73e-3	.207±9.80e-3	.203±6.73e-3
	MVL	1.00±6.37e-2	1.01±4.28e-2	1.00±6.26e-2	1.00±4.29e-2
	MR	6.48e-9±7.16e-9	8.57e-9±8.63e-9	5.42e-4±7.02e-5	1.98e-5±5.00e-6
	N,I	100, 4338	100, 3047	100, 10822	100, 776
2.0	MTP	.213±2.17e-2	.204±1.36e-2	.213±2.15e-2	.204±1.36e-2
	MVL	1.04±.139	1.01±8.65e-2	1.04±.138	1.01±8.66e-2
	MR	6.03e-9±7.61e-9	8.05e-9±7.76e-9	5.36e-4±9.30e-5	1.96e-5±4.80e-6
	N,I	100, 6394	100, 4186	100, 16624	100, 1012
4.0	MTP	.219±4.84e-2	.206±2.93e-2	.219±4.57e-2	.206±2.93e-2
	MVL	1.11±.398	1.04±.192	1.10±.358	1.04±.191
	MR	7.30e-9±7.71e-9	9.24e-9±8.07e-9	9.07e-4±2.79e-3	1.95e-5±4.33e-6
	N,I	100, 9434	100, 5513	100, 23565	100, 1276
8.0	MTP	.253±.135	.219±7.12e-2	.220±6.42e-2	.216±5.54e-2
	MVL	1.52±1.56	1.12±.668	1.15±.545	1.06±.308
	MR	1.63e-8±3.65e-8	1.11e-8±2.50e-8	1.60e-2±5.71e-2	4.43e-3±4.34e-2
	N,I	93, 22432	98, 7732	95, 32663	97, 1625
16.0	MTP	.363±.315	.288±.235	.236±.108	.221±.103
	MVL	3.30±4.95	2.61±6.36	1.44±.946	1.27±.865
	MR	3.96e-9±8.38e-8	3.18e-8±1.37e-7	7.00e-2±.124	6.55e-2±.186
	N,I	88, 76552	88, 31039	79, 52267	80, 2869
32.0	MTP	.323±.376	.288±.212	.163±.136	.145±.144
	MVL	2.68±5.32	2.69±4.97	1.09±.755	1.13±.961
	MR	4.13e-8±1.44e-7	2.20e-8±2.83e-8	.185±.258	.238±.295
	N,I	63, 146517	72, 63319	62, 82307	62, 5159
64.0	MTP	.314±.339	.506±.578	.191±.171	.193±.177
	MVL	1.81±3.35	6.83±27.4	1.16±1.12	1.61±1.46
	MR	1.37e-8±3.35e-8	4.05e-8±8.99e-8	.387±.426	.471±.595
	N,I	54, 199969	46, 117896	50, 115372	29, 11671

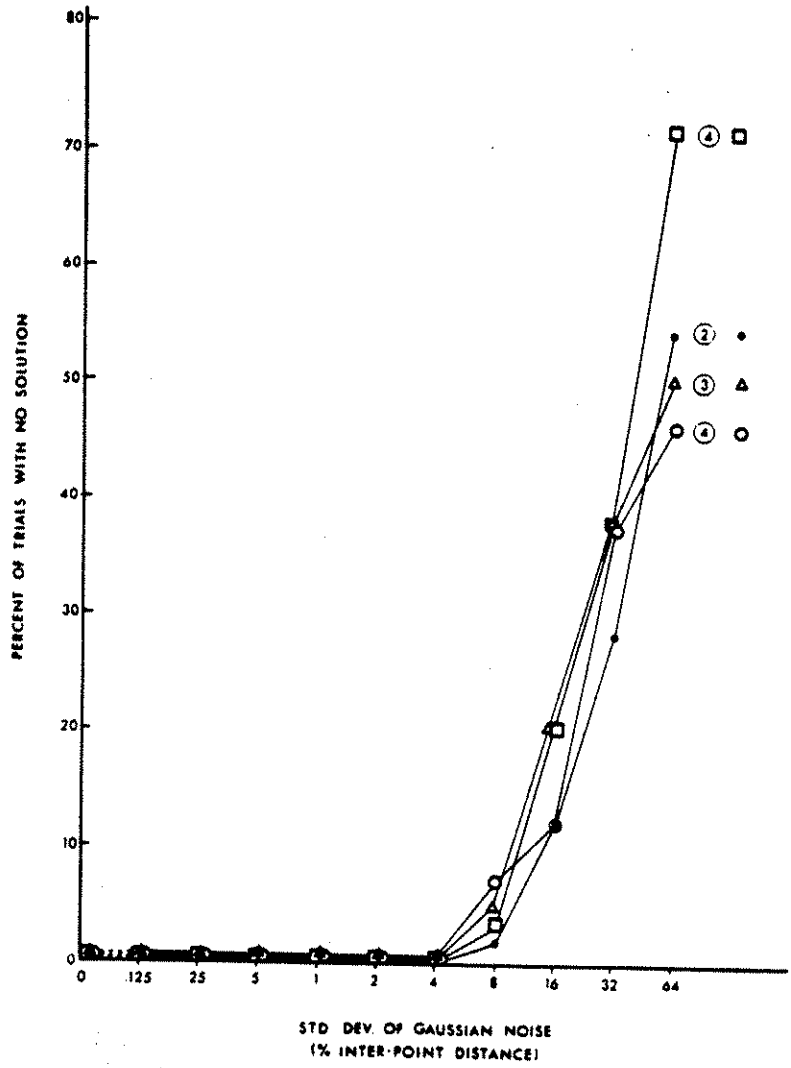


Figure 5. Percent of trials with no solution versus different levels of gaussian noise added to two-dimensional input data. The standard deviation of the noise was measured as a percentage of the average two-dimensional (image) distance between unperturbed dot positions. Cases 1–4 are defined and explained in the text.

from 0 to 4%. It may be possible to enhance speed with various implementation tricks. For example, p, the step size, was initially set at 0.1 for cases 1, 2, and 4, and at 0.01 for case 3 following initial testing showing that these values provided a good tradeoff between speed of convergence and the increase of degeneracy with noise. Following Marschak's example, p was halved and the iteration repeated

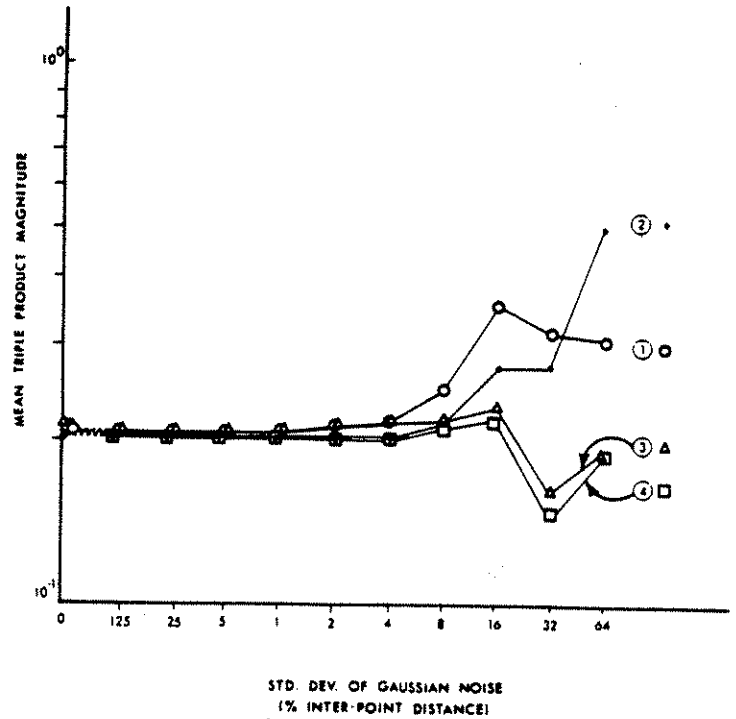


Figure 6. Deviation from planarity: Mean triple product magnitude over 100 trials versus different levels of gaussian noise added to input data. Relaxation of planarity (cases 1 and 2) differs from relaxation of rigidity (cases 3 and 4) only at noise levels above 5–8%. At the highest two noise levels, many trials involve input data too perturbed to result in convergence within implementation parameters. Thus at these levels, only those on the low end of the distribution survive to be included in the results, lowering the observed mean.

whenever a z_i iterate became negative. A fast but more sophisticated means of choosing p initially and adjusting it as a function of gradient magnitude might alleviate the speed problem to an extent. However, beyond such attempts, the iterative nature of the algorithm is an inherent limitation on its speed.

In Marschak's implementation, a Runge-Kutta averaging method was used to enhance accuracy. Such enhancement is helpful because solution by the gradient method involves extrapolation of local gradient information to obtain successive iterates. Extrapolation in following gradients generally results in more serious error accumulation as the step size between successive iterates is increased. On the other hand, we have already noted that an excessively small step size can mean uselessly slow convergence. We chose not to use Runge-Kutta averaging in order to enhance speed. The consequences of this choice would be felt in combining local solutions across an image since all slightly inaccurate local solutions would have to be

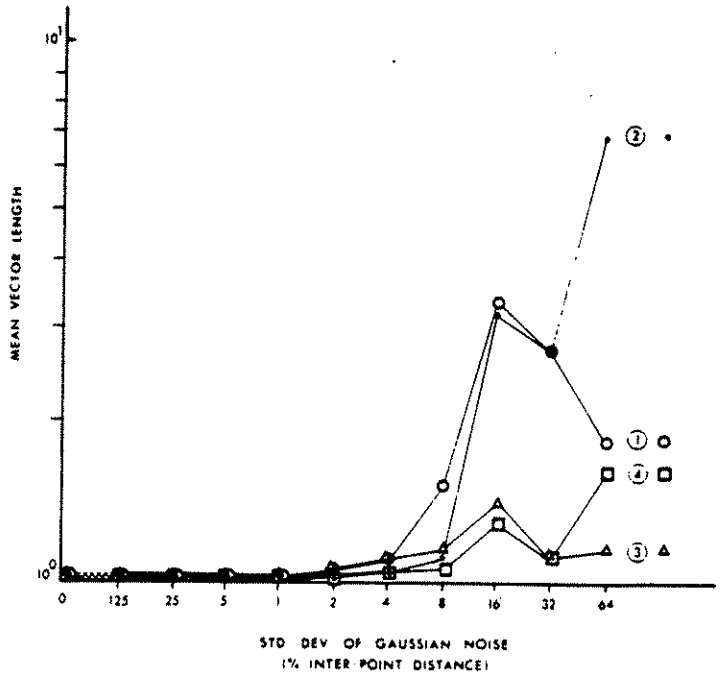
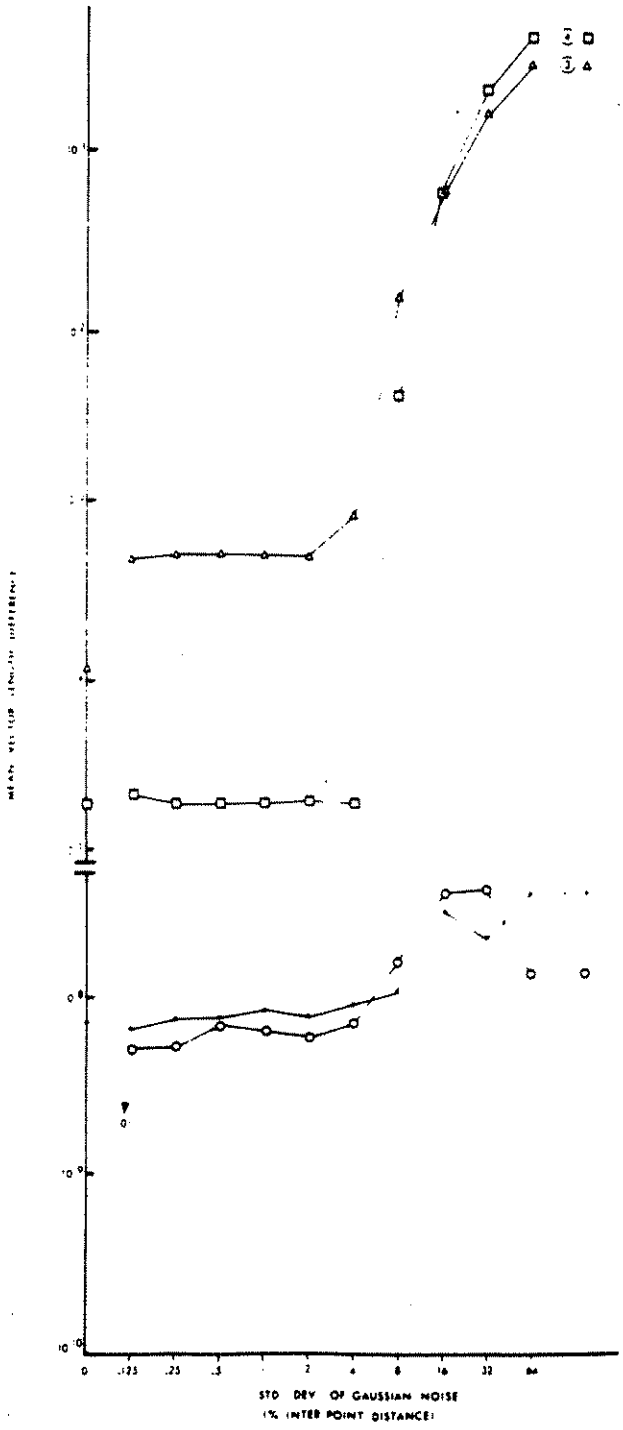


Figure 7. Mean vector length over 100 trials versus different levels of gaussian noise added to the two-dimensional input data.

reconciled to converge at a suitable global interpretation, but this stage has not yet been explored. However, it is useful to note that here in the case of accuracy, as in the case of speed, step size is critical in determining performance.

The issue of accuracy in practical use of the gradient method is wider than error generation and propagation as described above. It is necessary to choose an $\epsilon_1 = |z_i^{n+1} - z_i^n|$ at which to stop, a lower bound ϵ_2 on z_i^n to distinguish degenerate cases, a lower bound ϵ_3 on p to avoid false convergence caused by halving p too many times, and an upper bound ϵ_4 on the range of partial derivatives representable on the machine. It may even be necessary to choose a lower bound ϵ_5 on the range of derivatives allowed in addition to ϵ_3 to avoid prolonged occillation of the derivative around zero near the solution. The lack of a principled means of choosing and varying these parameters with input data makes the performance of the algorithm at

Figure 8. Mean differences between successive views of individual point vectors (used as a measure of deviation from rigidity). The exact formulas used were $\Sigma(abs(\|a_1\| - \|a_2\|) + abs(\|a_1\| - \|a_3\|))\ /n$ for cases 1 and 3 and $\Sigma(abs(\|a_1\| - \|a_2\|) + abs(\|b_1\| - \|b_2\|))\ /n$ for cases 2 and 4. Thus comparison is only appropriate between cases 1 and 3 or between cases 2 and 4.


STD DEV OF GAUSSIAN NOISE
(% INTER POINT DISTANCE)

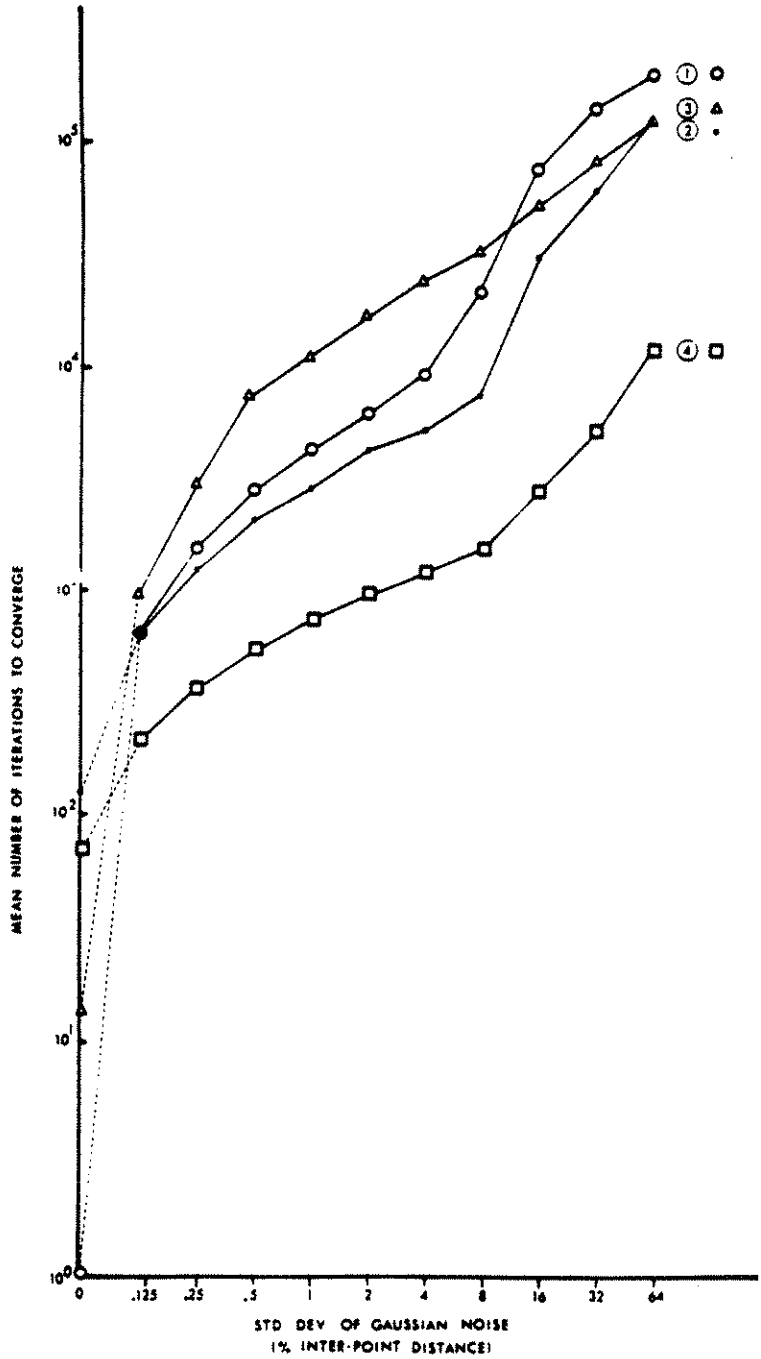


Figure 9. Mean number of iterations to convergence. See text for case definitions.

higher noise levels more variable and difficult to judge objectively. In our implementation, the choice of these parameters was made in initial test runs. They were then held constant throughout the simulation reported here as a basis for eliminating trials with no solution ($\epsilon_1=10^{-6},\epsilon_2=10^{-5},\epsilon_3=0,\epsilon_4=10^{10},\epsilon_5=0$).

Our final observation with respect to the algorithm used concerns the need for previous knowledge. The gradient method in the case of rigidity and planarity cannot be considered a startup method since arbitrary selection of z_i^1 (and p) does not guarantee convergence even when a solution exists. Thus, performance, i.e., speed and frequency of convergence, is ultimately dependent on estimation outside the ability of the algorithm.

We are led to the conclusion that the gradient method of solving the structure from motion problem is of some value but has important limitations. First, it is hampered by its iterative nature and potential for error accumulation. Additionally, while it is robust enough to deliver a solution in the face of 5% standard deviation gaussian noise, such performance does not approach that of the human system. Finally, initial parameter choices and their optimal variation with input are not yet well determined. Given adequate "external system knowledge" of this sort, it may be possible to initiate and maintain sufficient locality to lessen the significance of the speed issue. It remains to be seen whether a more sophisticated use of parameters in this manner can overcome some of the algorithm's limitations.

5. *Summary and Conclusions*

Our discussion has emphasized the importance of the search for plausible regularities of nature that can adequately constrain image interpretations and allow the visual system to infer the three-dimensional structure of the world. The driving force behind these inferences is the need to achieve goals such as manipulation and movement with respect to the scene analyzed. The strength of the inferences which can be made about the scene, and thus the successful achievement of visual goals, hinges on choosing regularities that minimize the false targets probability—the probability that image brightness patterns can be consistent with phenomena obeying the regularities when in reality those phenomena do not obey the regularities. We have examined some of the consequences of exploiting the useful regularities of local rigidity of objects and local movement in a plane. Both enable the unique recovery of three-dimensional scene structure from very little spatio-temporal information. Our simulation compared alternative structure recovery strategies in the face of noisy data by varying the relative importance of optimizing the rigidity and planarity regularities in the interpretation process. The results suggest that both strategies—relaxing planarity while maintaining rigidity and relaxing rigidity while maintaining planarity—degrade at about the same rate with noise. However, at the highest noise levels, relaxation of planarity can lead to solutions with longer mean vector lengths.

ACKNOWLEDGMENTS

We thank A. Bobick, T. Indow, W. Richards, and A. Witkin for helpful discussions. This material is based upon work supported by the National Science Foundation under Grant No. IST-8413560, and by the office of Naval Research, Psychological Sciences Division, Engineering Psychology Group.

REFERENCES

Arrow, K. J., Hurwicz, L., & Uzawa, H. (1958). *Studies in linear and non-linear programming*. Stanford, CA: Stanford University Press.

Bennett, B. M., & Hoffman, D. D. (1985). The computation of structure from nonrigid motion. *Biological Cybernetics, 51*, 293–300.

Bobick, A. A. (1983). Hybrid approach to structure from motion. *Proceedings from Siggraph Interdisciplinary Workshop, "Motion: Representation and Perception,"* 91–109.

Gibson, J., & Gibson, E. (1957). Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology, 54*(2), 129–138.

Green, B. F. (1961). Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology, 62*(3), 272–282.

Hay, C. J. (1966). Optical motions and space perception—an extension of Gibson's analysis. *Psychological Review, 73*, 550–565.

Hoffman, D., & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics, 42*, 195–204.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14*(2), 201–211.

Johansson, G. (1975). Visual motion perception. *Scientific American, 232*(6), 76–88.

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In J. Neyman (Ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (pp. 481–491). Berkeley, CA: University of California Press.

Longuet-Higgins, H. C., & Pradzny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society* (London), *B208*, 385–387.

Marr, D. (1982). *Vision*. San Francisco: Freeman Press.

Richards, W. A., Rubin, J. M., & Hoffman, D. (1982). Equation counting and the interpretation of sensory data. *Perception, 11*, 557–576.

Skyrms, B. (1975). *Choice and chance. An introduction to inductive logic*. Wadsworth. Belmont, CA: Dickenson Publ.

Symon, K. R. (1971). *Mechanics*. Reading, MA: Addison-Wesley.

Ullman, S. (1979a). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, S. (1979b). Relaxation and constrained optimization by local processes. *Computer Graphics and Image Processing, 10*(6), 115–125.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*(4), 205–217.

Webb, J., & Aggarwal, J. (1981). Visually interpreting the motion of objects in space. *Computer*, 40–46.