PERGAMON

Vision Research 41 (2001) 3145–3161

Vision Research

www.elsevier.com/locate/visres

Mini-review

Perception viewed as an inverse problem

Zygmunt Pizlo *

Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364, USA

Received 28 June 1999; received in revised form 15 March 2000

Abstract

The modern study of perception began when Fechner published his 'Elements of Psychophysics' in 1860. This book has guided most perception research ever since. It has become increasingly clear that there are problems with Fechner's approach, which assumes that the percept is completely determined by the sensory input. Fechner's approach cannot explain the processes that allow our percepts to be veridical. Post-Fechnerian schools (Helmholtzian, Structural, Gestalt and Gibsonian) have tried to deal with this problem, but have not been successful. An alternative to the Fechnerian approach is required. This paper describes an alternative that has been developing over the last 20 years within the computer vision community. It treats perceptual interpretation as a solution of an inverse problem that depends critically on the operation of *a priori* constraints. Contemporary research, which adopted this approach, has concentrated on verifying the usefulness of Bayesian and standard regularization methods. This paper takes the next step; it discusses theoretical and empirical aspects of studying human perception as an inverse problem. It reviews the literature that illustrates the power of the inverse problem approach. This review leads to the suggestion that progress in the study of perception will benefit if the inverse approach were to be adopted by experimentalists, as well as by the computational modelers, who have been actively exploring its potential to date. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Inverse problems; Simplicity; Likelihood; Regularization; Bayes' rule

1. Introduction

The modern study of perception began with the theoretical framework provided by Gustav Theodor Fechner (1860/1966). The 'percept', according to Fechner, is the result of a causal chain of events that starts with a distal stimulus (e.g. an object), and proceeds through a proximal stimulus (e.g. the object's retinal image), transduction, brain processes, and ends with the percept of the distal stimulus. Fechner distinguished two special kinds of relationships that involve the percept, which he called inner and outer psychophysics. Inner psychophysics, which deals with what is often referred to as the 'mind–body' problem, refers to the relationship between the percept and its underlying brain processes. Outer psychophysics refers to the relationship between the percept and the stimulus (distal or proximal).

Fechner concentrated his psychophysical studies on outer psychophysics as the more tractable case. The assumption that the percept is a result of a causal chain of events led Fechner to a conjecture that the percept is a mental measurement of the physical stimulus. It is quite obvious how some properties of the proximal stimulus, e.g. the intensity of light incident on the retina or distance on the retina, can be measured by the perceptual system. It is less obvious, however, how properties of the distal stimulus, e.g. the shape of a solid object, can be measured by the perceptual system. It is less obvious because the perceptual system has direct access only to the proximal stimulus, and the relationship between distal and proximal stimuli is, in general, quite complex. This means that an accurate measurement of the proximal stimulus may actually have little relevance to the distal stimulus. Consider the

* Tel.: + 1-765-494-6930; fax: + 1-765-496-1264.
E-mail address: pizlo@psych.purdue.edu (Z. Pizlo).

PII: S0042-6989(01)00173-0

percept of a 3D object from a single retinal image. The 2D image projection determines an infinite number of different 3D interpretations. Despite this geometrical ambiguity, the observer usually perceives only one 3D object and the percept is often veridical, i.e. the percept accurately describes the physical object. Even if we assume that the observer could 'measure' her retinal image of the object, the result of this measurement would not be sufficient to account for a 'veridical' percept of the object.

This difficulty can be avoided by using simple stimuli that give rise to a one-to-one mapping between the distal and proximal stimulus. For example, to decide which of two line segments presented in the frontal plane at the same viewing distance is longer, the visual system can simply measure and compare the lengths of the retinal projections of these line segments. Most psychophysicists in the 19th century adopted this kind of simplification. This early research led to several important accomplishments, namely, theories of detection and discrimination, Weber's and Fechner's laws, 'classical' psychophysical methods, and mathematical theory of measurement (Fechner, 1860/1966; Stevens, 1946; Krantz, Luce, Suppes, & Tversky, 1971; Narens, 1985; Suppes, Krantz, Luce, & Tversky, 1989; Luce, Krantz, Suppes, & Tversky, 1990). Difficulty in generalizing this measurement framework to more complex stimuli that involved a many-to-one mapping from the distal to the proximal stimulus eventually resulted in the establishment of the Helmholtzian, Structural and Gestalt schools. These schools provided evidence showing that the percept is not merely caused by the proximal stimulus but involves: (i) unconscious inferences, (ii) associating boundles of sensations, or (iii) *a priori* assumptions or constraints biasing the percept towards a simple interpretation. Such concepts cannot be readily incorporated within the framework of Fechnerian psychophysics. However, despite the fact that during the last 100 years, much progress in our understanding of perceptual phenomena has been made by those who used inferences, associations, and *a priori* constraints in theories of perception, Fechnerian psychophysics has remained the most mature and theoretically the most advanced framework for studying perception (e.g. Luce, Bush, & Galanter, 1963; Falmagne, 1985).

The main goal of this paper is to describe and discuss a framework for studying perception that avoids the methodological limitations of Fechnerian psychophysics. *This framework emphasizes the relationship between the distal and proximal stimulus and treats perception as a solution of an inverse problem.* The next sections describe this approach and its implications for the study of perception, and review empirical and theoretical results that demonstrate the benefit of this framework.

2. Inverse problems and methods of solving them

Consider a mapping A from the distal stimulus X (e.g. a 3D object) to the proximal stimulus Y (e.g. its retinal image). If the object and its image are represented by homogeneous coordinates, the perspective mapping A is a linear transformation. Thus, one can write the following equation:

$$Y = AX. \tag{1}$$

Finding the proximal stimulus for a given distal stimulus is a direct (forward) problem and is expressed in the rules of physics. Direct problems in natural sciences are usually easy. Specifically, they are well posed and well conditioned. A problem is well posed when: there is a solution, the solution is unique, and the solution depends continuously on the data. If one or more of these criteria are not satisfied, the problem is ill posed. A problem is well conditioned when the solution is computationally stable (robust against noise).

In contrast to the problem represented by Eq. (1), an observer is faced with an inverse problem. Namely, *perception is about inferring the properties of the distal stimulus X given the proximal stimulus Y*:

$$X = A^{-1}Y. \tag{2}$$

This inverse problem is ill posed and/or ill conditioned. This is related to the fact that finding a unique and stable A^{-1}, which is needed to determine X, is difficult.

Consider an example of the image of a cube. This retinal image determines an infinite number of objects whose faces do not have to be planar, edges do not have to be straight-line segments, and the object does not have to be symmetric. Clearly, the problem of visual interpretation of the retinal image is ill posed, yet whenever the retinal image could have been produced by a cube, the observer perceives a cube (Attneave & Frost, 1969; Perkins, 1972), except for degenerate cases (Hochberg & McAlister, 1953). It is obvious that the visual system is imposing constraints on the family of possible perceptual interpretations. For example, when the retinal image is produced by a polyhedron, the observer's percept corresponds to a polyhedron in which the variance of all angles is minimal (Attneave & Frost, 1969; Marill, 1991; Leclerc & Fischler, 1992; Sinha & Adelson, 1992).

It is clear that imposing *a priori* constraints on the family of possible solutions is likely to remove the ill posedness of an inverse problem and thus leads to a unique solution. It is not clear, however, whether this approach will lead to veridical percepts. For this to happen, the constraints should reflect the properties of objects in the natural environment. Objects 'out there' are not completely arbitrary and random. Objects tend to be continuous, piece-wise smooth, symmetric, and are usually familiar to the observer. It seems reasonable to assume that an effective perceptual system should obtain and use knowledge of the constraints that characterize the distal stimuli to 'make up' for the information that has been lost in the transformation from the distal to proximal stimulus.

Let the constraints be denoted by P_X. Let Y_X be the partial information that the observer obtains about the distal stimulus X from the proximal stimulus Y. The percept of the distal stimulus X will be veridical if the information obtained from the constraints P_X and the proximal stimulus Y_X is sufficient to compute (reconstruct) X, or a good estimate $\hat{X}$:

$$\hat{X} = m_X(Y_X, P_X) \tag{3}$$

In this equation, m_X is a many-to-one mapping that represents what and how is being computed. The subscript in m_X refers to the fact that the mapping from (Y_X, P_X) to $\hat{X}$ depends, in the general case on X. This means that *a priori* constraints can be incorporated not only explicitly in the form of P_X, but also implicitly in the form of assumptions involved in computational methods m_X. For example, Ullman's (1979) structure from motion algorithm assumes that a 3D rotating object is rigid. When the rigidity assumption (constraint) is satisfied, the algorithm produces a unique solution (up to depth reversal). Otherwise, the solution may not exist.

Determining Y_X given Y is not trivial. The perceptual ability to solve this problem is related to the phenomenon of figure–ground segregation, which in most cases is itself an inverse problem. Since Y_X is a subset of Y, the former can be thought of as resulting from a mapping $f_X(Y) \rightarrow Y_X$ that selects the right features from the proximal stimulus Y for the problem of reconstructing the distal stimulus X. Thus, Eq. (3) can be written more fully as:

$$\hat{X} = m_X[f_X(Y), P_X]. \tag{4}$$

In some cases, the reconstruction problem can be simplified. The constraints P usually operate on the distal stimulus X because properties of objects are more stable than properties of images. But some properties of objects behave well under the transformation from the distal to proximal: a straight line in 3D always projects to a straight line under perspective transformation, a symmetrical planar figure gives rise to a nearly symmetrical image, etc. In such cases, one can apply constraints P to the proximal stimulus Y. For example, Gestalt rules of perceptual organization, which involve constraints of good continuation, proximity, similarity, etc., operate on the proximal stimulus and do not depend strongly on the distal stimulus. Furthermore, in some cases, the mappings m and f may not depend on the distal stimulus X. For example, the problem of figure–ground segregation can often be solved without top-down knowledge about the distal stimulus, although such knowledge can be beneficial (e.g. Pizlo, Salach-Golyska, & Rosenfeld, 1997). Under such conditions, Eq. (4) leads to a simpler form:

$$\hat{X} = m[f(Y), P_Y] \tag{5}$$

Examples of perceptual problems that can be adequately represented by Eq. (5) include contour detection (e.g. Pizlo et al., 1997), and binocular correspondence (e.g. Pollard, Mayhew, & Frisby, 1985).

If the proximal stimulus Y contains noise, there may not exist a distal stimulus $\hat{X}$ that is consistent with the proximal stimulus Y_X and at the same time satisfies the constraints P_X or P_Y exactly. In such cases, one has to deal with the compromise between the two requirements. This compromise can be addressed in a natural way by regularization and Bayesian methods. These methods seem to be especially interesting because they have strong theoretical foundations on the one hand, and they lead to plausible models of biological vision, on the other hand. However, other strategies for applying constraints in solutions of inverse problems have been explored.

The regularization method of solving ill-posed inverse problems was formulated by Tikhonov in the early 1960s (Tikhonov & Arsenin, 1977). In this method, the solution is obtained by finding $\hat{X}$, which minimizes a functional:

$$E_1 = \|AX - Y_X\|^2 + \lambda \|P_X\|^2 \tag{6}$$

where λ is a regularization parameter. The first norm evaluates how close the distal stimulus is to the proximal stimulus, and the second norm evaluates how well the *a priori* constraints are satisfied. If the proximal stimulus is reliable, λ should be small, otherwise λ should be large. In Tikhonov's theory, A is assumed to be a linear operator, P_X, a linear combination of the first p derivatives of the distal stimulus X, and the norms are quadratic. Poggio and colleagues (Poggio, Torre & Koch, 1985) were the first to draw attention to Tikhonov's work in the vision community, even though these methods were used earlier. Regularization methods have been applied to a wide range of visual tasks such as contour detection (Shashua & Ullman, 1988), motion reconstruction (Horn & Schunck, 1981; Hildreth, 1984; Shulman & Aloimonos, 1988), shape reconstruction (Ikeuchi & Horn, 1981; Grimson, 1982), and color perception (Horn, 1974).

There is a stochastic version of the regularization theory involving Bayes' rule (Poggio et al., 1985; Bouman & Sauer, 1993; Kersten, 1999). Bayes' rule allows the computation of the posterior probability $p(X|Y_X)$ as follows:

$$p(X|Y_X) = \frac{p(Y_X|X)p(X)}{p(Y_X)} \tag{7}$$

$p(X)$ is the prior probability distribution for the distal stimulus and represents *a priori* knowledge about objects in the environment. $p(Y_X|X)$ is the likelihood function for X. This function represents the transformation from the distal to proximal stimulus and includes information about noise in the proximal stimulus. Finally, $p(Y_X)$ is the probability of obtaining the proximal stimulus. The inverse problem of determining the distal stimulus based on the proximal stimulus can be solved by finding $\hat{X}$ that maximizes the posterior probability, $p(X|Y_X)$. Such an $\hat{X}$ is called a Maximum A Posteriori (MAP) estimator. The maximum of the probability distribution is not the only choice. One can use other properties like the mean of the distribution. The term $p(Y_X)$ is a normalizing constant in Eq. (7) that does not affect $\hat{X}$ and can be omitted in solving the optimization problem.

To see more directly the relation between the deterministic and stochastic methods of solving inverse problems, we take the logarithm of both sides of Eq. (7) (ignoring the term $p(Y_X)$):

$$-\log\ p(X|Y_X) = -\log\ p(Y_X|X) - \log\ p(X). \tag{8}$$

Eqs. (6) and (8) are analogous (the regularizing parameter λ in Eq. (6) is implicitly represented in Eq. (8) by the ratio of the variances of $p(Y_X|X)$ vs. $p(X)$). In fact, under some assumptions about the probability distributions, maximizing the posterior probability $p(X|Y_X)$ (Eq. (7)) is equivalent to minimizing the functional E_1 (Eq. (6)). The Bayesian framework, however, is more general than the standard regularization. For example, in Bayesian formulation, the uniqueness of the solution is represented by the uniqueness of the posterior probability distribution, not by the fact that this distribution is unimodal, as in standard regularization (Tarantola, 1987).

It is obvious why some tasks, like perception of a 3D object from a single image, are ill-posed, but it may be less clear why binocular reconstruction of a 3D scene, or perception of shape from motion should be ill posed or ill conditioned. In fact, Gibson (1950, 1979) and Rock (1983) claimed that in full-cue ('ecologically valid') conditions, there is no ambiguity, and all the important properties of the 3D scene can be simply 'picked up' or computed from the retinal images. Consider an example in which an observer is viewing a 3D object binocularly. Assume that the observer knows exactly the position of one eye relative to the other (i.e. interocular distance, vergence, etc.), that there is no noise in the perceptual system and that the correspondence of the points (or features) on the retinas is known. Under such assumptions, the problem of reconstructing any point in 3D space from the two images is well posed and well conditioned. The observer can simply find the intersection of the visual rays emanating from the corresponding retinal image points and going through the centers of projection (von Helmholtz, 1910/2000). This intersection must exist and is unique.

Now consider the more realistic case, which acknowledges the presence of perceptual and motor noise. In this case, the two corresponding visual rays almost never actually intersect (i.e. there is no solution). Clearly, binocular reconstruction is an ill-posed problem. The ill posedness can be removed by any approximation method (e.g. the mid-point method, Hartley and Sturm, 1997), but such approximations are quite unstable.[1] Even a small amount of image noise (a standard deviation of 1% of image diameter) leads to very large errors (a standard deviation of 80%) in the reconstructed ratios of distances (Chan, Pizlo, & Chelberg, 1999). This instability demonstrates that the inverse problem of binocular reconstruction is ill conditioned. As a result, a binocular vision system must use constraints in order to stabilize the solution.

The fact that the percept involves *a priori* constraints has been commonly recognized since the beginning of the previous century when the Gestalt psychologists described the operation of the rules of perceptual organization (Koffka, 1935). Surprisingly enough, however, the current research on human perception is still dominated by the Fechnerian framework: research questions concentrate on perceptual and neurophysiological coding of sensory information. But *constraints do not merely contribute to perceptual coding. They represent a critical tool in solving (veridically and reliably) an inverse problem of perceptual interpretation.*

The next section presents details of the framework of inverse problems as applied to human perception. It begins with theoretical aspects. This is followed by a description of experimental methodology. Then, prior research on human visual perception is reviewed and compared to the main assumptions of the framework of inverse problems.

3. Framework of inverse problems in human perception

3.1. Theoretical aspects

The discussion of the framework of inverse problems presented in the previous section concentrated on computational methods and did not incorporate aspects that are characteristic of solving inverse problems by human observers. We identify three such aspects of human perception that have to be explicitly incorporated: (i) the locality of constraints, (ii) the phenomenon of perceptual constancy and familiarity with the environment, and (iii) the dynamic nature of the proximal and distal stimuli.

[1] See our demo illustrating this statement: http://bigbird.psych.purdue.edu/binshape

3.1.1. Locality of constraints

The first question is how the observer forms good priors (constraints) for objects and events that can be encountered. In principle, in order to have an estimate of the probability of every event (distal stimulus), the observer would have to experience every event more than once. This is, of course, unrealistic (Maloney, in press). To overcome this problem, the observer could use the universal priors, which involve the concept of simplicity and do not require any experience with the stimuli (Li & Vitanyi, 1997). Although there is good evidence that human observers use simplicity constraints in perception, these constraints are 'local' in that they are formulated only in some subspaces of the space of possible distal stimuli. This is reasonable from a computational point of view. Information about the distal stimulus that can be reliably obtained from the proximal stimulus does not need to be replicated in *a priori* constraints. Consider the visual perception of an object in the 3D space. If there are no occlusions, the geometrical properties such as shape specified in the 2D subspace corresponding to the frontal plane of the observer are reliably represented on the retina and do not require any priors to interpret. The orthogonal space corresponding to depth is not reliably represented on the retina. Thus, only perceptual interpretations involving the depth dimension require priors.

The use of priors that are defined only in a subspace of the space of properties of distal stimuli allows the perceptual system to solve the inverse problem for a wide range of distal stimuli, including those that the observer has never perceived before (as in the case of the application of a smoothness constraint to an unfamiliar contour in a noisy image). In the case of objects from our natural environment, such local priors tend to apply to all properties (subspaces) that are not reliably represented in the proximal stimulus, and as a result, the percept is often veridical. The combination of the information present in the proximal stimulus and information provided by the constraints may sometimes be insufficient for a unique and veridical percept. This happens only when a distal stimulus does not have any features in the subspace in which the constraints operate, or when the proximal information is ambiguous in the subspace in which the constraints do not operate.

3.1.2. Perceptual constancy and familiarity with the environment

Perceptual constancy refers to the fact that the percept of a distal stimulus is constant despite changes in its proximal stimulus. The proximal stimulus may change because of changes in the viewing conditions. For example, when the orientation of an object relative to an observer changes, the object's retinal image changes, but the percept of the object's shape does not. Let the observer's percepts of X produced by two different proximal stimuli Y_i and Y_j be $\hat{X}_i$ and $\hat{X}_j$, respectively. The constancy of the percept is represented by:

$$\|\hat{X}_i - \hat{X}_j\| = 0 \tag{9}$$

Perceptual constancy is closely related to veridicality of the percept, which is represented by:

$$\|\hat{X}_i - X\| = 0 \tag{10}$$

Perceptual constancy is a necessary condition for veridicality. If the percept is not constant, then it cannot be veridical. Perceptual constancy is commonly accepted as an adequate measure of the veridicality of the percept. This is justified because the veridicality of a percept (defined as in Eq. (10)) cannot always be verified. In the case of shape, the veridicality of the percept can be verified because: (i) shape does not involve measurement units (it is defined by ratios of distances), and (ii) the percept and the distal stimulus are described in spaces with the same number of dimensions (here three). In other cases, however, establishing whether the percept is identical to the distal stimulus can be problematical. Consider the case of surface color. Surface reflectance is represented by a point in an infinitely dimensional space, whereas the percept of the reflectance is represented by a point in a 3D space. A 3D property is almost never equal to an infinitely dimensional one. In such cases, it is more convenient to define perceptual veridicality in terms of the constancy of the percept. After all, the constancy of the percept is equivalent to the veridicality of the percept of differences.

Another reason for the centrality of the concept of perceptual constancy is that human observers usually operate in a familiar environment. Familiarity with a distal stimulus X implies that the observer already has in her memory a perceptual representation X_M of this stimulus. When presented with the stimulus X, the task is simply to verify whether it is the same stimulus that produced X_M (perceptual constancy task). This can be done by performing recognition, where the observer matches the memory representation X_M to the proximal stimulus:

$$E_2 = \|AX_M - Y_X\| \tag{11}$$

If E_2 is close to zero, the observer can conclude that the distal stimulus X is identical with X_M (Pizlo, 1994; Pizlo & Loubier, 2000). If the memory representation were produced by a stimulus similar, but not identical to the distal stimulus, the memory representation X_M may serve as a source of additional constraint. This constraint can be incorporated by extending E_1 (Eq. (6)) as follows:

$$E_3 = \|A\hat{X} - Y_X\| + \lambda\|P_X\| + \alpha\|\hat{X} - X_M\| \tag{12}$$

E_3 is a combination of conventional regularization (Eq. (6)) with perceptual constancy (Eq. (9)). In this case, the memory representation can be used as a starting point in the task of finding the minimum of E_3. One can, in a similar way, incorporate recognition to the Bayesian method of perceptual interpretation.

Objects from our environment often form 'families' consisting of objects that share some geometrical, physical or functional properties. For example, all human faces contain eyes, ears, a nose and a mouth. The family of chairs contains objects that vary with respect to physical and geometrical properties, but share a functional property. A family of similar objects can be represented by a prototype, or by some characteristic features of the prototype. This kind of memory representation is a useful source of constraint that can be incorporated in perception by modifying the third term on the right-hand side of Eq. (12):

$$E_4 = \|A\hat{X} - Y_X\| + \lambda\|P_X\| + \alpha\|B\hat{X} - M_X\| \tag{13}$$

B is a projection of $\hat{X}$ onto the memory, and M_X is the memory representation of a prototype, which involves a subspace of the space characterizing the distal stimulus. This type of memory is quite efficient: storing 'copies' of all distal stimuli requires much more memory space than storing only some characteristic aspects of the stimuli. Furthermore, recognition based on prototypes leads to a hierarchical organization of memory that speeds up the memory search (Pizlo & Loubier, 2000).

3.1.3. Dynamic nature of the proximal and distal stimuli

If an approach based on Eq. (4) or Eq. (5) is used to solve the inverse problem of perceptual interpretation, there is always a possibility that the solution will be non-unique or unstable despite the use of constraints. This may come about when the proximal stimulus is too impoverished, as in the case of severe occlusion of the distal stimulus, and does not allow the application of all of the necessary constraints. Since the proximal stimulus is a function of time (t), because of random noise in the perceptual and motor systems and because the distal stimulus is rarely static, the reconstruction method based on Eq. (4) or Eq. (5) cannot assure that $\hat{X}(t+1)$ will stay close to $\hat{X}(t)$. This means that perceptual constancy is likely to fail. Such failures can be avoided when the regularization or Bayesian method is used, by adding to E_1 (Eq. (6)) a term stabilizing the percept in time (adding this term resembles, although is not equivalent to, Bayesian updating—Maloney, in press):

$$E_5 = \|A\hat{X}(t+1) - Y_X\| + \lambda\|P_X\| + \beta\|\hat{X}(t+1) - \hat{X}(t)\| \tag{14}$$

Regularization based on Eq. (14) has some resemblance to Ullman's (1984) maximizing rigidity approach to the shape from motion problem. The temporal stability term in Eq. (14) does not necessarily represent rigidity—it refers to small changes in the reconstructed distal stimulus, in general. An object could be nonrigid, and this term could represent stability of the sizes of individual parts or the color of its surface.

Consider an example. It is known that binocular judgments of distances in 3D are quite imprecise. When the observer is asked to judge a distance along the line of sight between two points, the threshold is quite large (McKee, Levi, & Bowne, 1990). This illustrates the fact that the proximal stimulus does not provide reliable information and that this kind of impoverished distal stimulus (two points) does not allow the visual system to apply strong constraints. Nevertheless, in a single trial, the two dots are perceptually stable and do not appear to move relative to one other. Such temporal stability of the perceived distance in conjunction with the large threshold in judging distances suggests that the percept involves some version of the cost function E_5 (Eq. (14)).

To summarize, the human perceptual system solves the inverse problem of perceptual interpretation by using a combination of constraints. The constraints are applied to a subspace of the feature space of the distal stimulus, and include not only simplicity (or likelihood), but also familiarity (memory) and temporal stability.

3.2. Experimental methodology

Early research on human perception emphasized bottom-up processes, because they are conceptually compatible with the causal chain of events in Fechnerian psychophysics. These processes include the coding of the sensory information, as well as integrating this information across sensory channels. There has been a gradual shift towards including complex stimuli in psychophysical research during the past half century. Complex stimuli allowed the researcher better insight into the mechanisms underlying the integration of a number of cues. The writings of Brunswik, Gibson and Marr represent this shift, which was greatly facilitated by developments in computer technology. *Complex stimuli did not fit into the Fechnerian measurement framework, and this fact caused theoretical psychophysics to lag behind developments in experimental psychophysics.*

In the framework of inverse problems, the classification of experimental stimuli, and the corresponding experimental methodology for studying psychophysics, is different. There are two criteria for this classification: (i) whether the perceptual system attempts to solve an inverse problem, and (ii) whether effective *a priori* constraints can be applied to a given stimulus.[2] When both these criteria are satisfied, the percept is quite different from the proximal stimulus, and the percept is stable (reliable across trials and subjects) as well as veridical. This experimental methodology is called here 'inverse psychophysics'. It often involves relatively complex and structured stimuli, like those encountered in everyday life. In such cases, the distal stimulus is different (and richer) from the proximal stimulus, and the percept resembles the distal rather than proximal stimulus (perceptual constancy).

If effective constraints cannot be applied, but the proximal stimulus and contextual cues provide information about the differences between the distal and proximal stimulus, then the perceptual system will attempt to solve the inverse problem of perceptual interpretation. The percept will be unstable, non-unique and non-veridical because the inverse problem remains ill posed and ill conditioned. This type of methodology is called here 'underconstrained inverse psychophysics'. It always involves unstructured, impoverished stimuli. Binocular judgments of spatial relations among light points in total darkness (Luneburg, 1947) are an example of underconstrained inverse psychophysics. Effective constraints do not exist in this case because the stimulus is very simple, but binocular disparity (a depth cue) provides information about the fact that the distal stimulus is three-dimensional, and thus, the visual system should provide such an interpretation. Underconstrained inverse psychophysics is often used to provide a baseline for perceptual performance. It reveals the worst possible performance, i.e. poor reliability and/or non-veridicality.

Finally, if effective constraints cannot be applied, and the proximal stimulus and contextual cues do not provide any information about the differences between the distal and proximal stimulus, the perceptual system will not attempt to solve the inverse problem of perceptual interpretation. In this case, the percept will be a mere coding of the proximal stimulus. The percept is likely to be stable and unique, but it may be veridical only if the distal and proximal stimuli are identical. This type of methodology is called here 'forward psychophysics'. It usually involves very simple, impoverished stimuli. Brightness detection and color discrimination of lights are examples of forward psychophysics. Forward psychophysics is usually used to study the nature of coding stimulus properties and to measure perceptual noise, which is represented by the likelihood function in Eq. (7) and by the regularizing parameter λ, in Eq. (6).

To learn about perceptual mechanisms that are involved in solving an inverse problem, an experimental study should establish first for which conditions and stimuli the perceptual system actually solves the problem and thus achieves veridical and reliable percept. It seems, therefore, that the study should begin with the methodology of inverse psychophysics. Next, one should include versions of the stimulus with 'degraded' structure. A stimulus with degraded structure is one whose structure has been modified in such a way that some constraints can no longer be applied, or the constraints lead to different interpretations. In the extreme case, a stimulus with highly degraded structure may produce an experimental condition described as underconstrained inverse psychophysics. If the observer's performance is strongly affected by this manipulation, called here the 'structure degradation procedure', one can conclude that the constraints under consideration are incorporated by the perceptual system in solving the inverse problem of perceptual interpretation.[3] Ideally, an experimental study using the structure degradation procedure should be accompanied by formulating a computational model of the perceptual mechanisms and testing the importance and effectiveness of the hypothesized constrains in simulations.

One can discover constraints without using the structure degradation procedure. A commonly used alternative concentrates on studying perceptual illusions. However, illusions themselves have weaker implications for our understanding of how the perceptual system solves an inverse problem: An application of *a priori* constraints to a distal stimulus that violates the underlying assumptions does imply a perceptual illusion. However, the converse is not true. That is, occurrence of an illusion may or may not indicate the operation of *a priori* constraints in solving an inverse problem. Even if it does, it is still not clear which perceptions benefit from the constraints. For example, the perceptual significance of the vertical–horizontal illusion is at present unclear. One can easily incorporate this illusion in a computational model by stretching the vertical dimension by several per cent. But this manipulation would not represent the application of the framework of inverse problems to perception. In the framework of inverse problems, one would have to demonstrate which percepts benefit from constraints that produce a given illusion. Such a demonstration requires the methodology of inverse psychophysics. Note that the conditions producing illusion in conjunction with the corresponding conditions of inverse psychophysics represent the structure degradation procedure.

[2] It is quite commonly accepted that most (if not all) constraints are built into perceptual processing mechanisms that automatically operate on the proximal stimulus. But this does not necessarily imply that all constraints are applied to every proximal stimulus. An automatic perceptual mechanism could be 'smart' in a sense that it decides which constraints are applied to a given proximal stimulus depending on the content of the stimulus. In other words, the perceptual processing does not have to be purely bottom-up as envisioned by Fechner, Helmholtz, Koffka, Gibson and even Marr. A perceptual mechanism may even incorporate such top-down effects as attention to locations and features (Palmer, 1994; Tsotsos et al., 1995; Pizlo et al., 1997).

[3] As pointed out above, constraints may be incorporated either explicitly in the form of P_X and P_Y in Eqs. (3)–(5), or implicitly, as assumptions underlying the computational methods.

The usefulness of a constraint in solving the inverse problem of perceptual interpretation can be verified in computational studies. However, such studies must be accompanied by psychophysical experiments in order to determine whether this constraint is actually being used by perceptual mechanisms. To illustrate this point, we analyze an example involving 3D unstructured (nonsense) wire objects. Wire objects received widespread attention in the vision community after Rock and DiVita (1987) demonstrated a complete failure of shape constancy with these stimuli. Consider the case of viewing a rigid wire object rotating around a vertical axis. Several potentially useful constraints are present in this case: rigidity of the object, smoothness of rotation, fixed axis of rotation and physical realizability. But the fact that these constraints are present does not mean that the human visual system uses them to infer the 3D shape of an object. The system may not use them either because they are not effective or because they are difficult to implement. Pizlo and Stevenson (1999) showed that shape recognition from novel views of a 3D rotating wire object is indeed close to chance level. At the same time, recognition performance in the case of structured objects like polyhedra is very reliable. The poor performance in the case of wire objects is not an artifact related to the use of highly similar stimuli. Phenomenologically, a wire object looks quite different when viewed from different viewing directions (Rock, Wheeler, & Tudor, 1989). Clearly, a subject's shape recognition performance indicates that the constraints that are potentially useful in the case of wire objects are not used by the human visual system. It follows that a result of the experiment with wire objects says little about how the visual system solves the inverse problem of shape perception. In other words, wire objects seem to 'force' the visual system to operate as if perception were a forward problem, and thus, consistent with Fechnerian psychophysics, rather than an inverse problem.

4. Psychological relevance of the framework of inverse problems

4.1. Development of the main concepts

Gestalt psychologists were the first to recognize that the percept does not reflect the nature of the proximal stimulus. The percept follows the application of *a priori* rules of organization (good continuation, proximity, similarity, symmetry, etc.) to the proximal stimulus. The rules of organization represent the operation of a minimum principle (Law of Prägnanz; see Koffka, 1935). According to the Gestalt psychologists, the perceptual system chooses an interpretation that is consistent with the proximal stimulus and is as simple as possible (see Mach 1906/1959, who anticipated this idea). This view fits the main idea underlying the framework of inverse problems with simplicity corresponding to the constraints P_X (or P_Y) in Eqs. (3)–(6). Thus, the Gestalt psychologists were close to changing the theoretical framework of psychophysics from Fechnerian to one like that described in this paper. They did not, however, succeed in changing it.

It seems likely that the main obstacle was that the Gestalt psychologists did not think of an observer as an information-processing system, which is trying to accomplish some goals, such as obtaining a veridical perceptual representation of the environment. Goal-directed behavior, obtaining, analyzing and communicating information, are elements that were first emphasized by researchers who launched the cognitive revolution (Craik, 1943; Rosenblueth, Wiener, & Bigelow, 1943; Bruner, Goodnow, & Austin, 1956; Attneave, 1959; Miller, Galanter, & Pribram, 1960; Neisser, 1967). The Gestaltists instead adopted the Fechnerian framework and tried to explain percepts in terms of a causal chain of events, with percepts being a reaction to stimulation. This way of thinking neither generated plausible theories nor led to experimental methodology adequate for studying the operation of the minimum principle. Approaches to perception that came after Gestalt psychology, namely, Gibson's (1950, 1979) 'ecological (direct) perception' and Rock's (1983) elaboration of Helmholtzian unconscious inference, also adopted the Fechnerian framework. Gibson's 'direct perception' was the farthest removed from the idea of perception as an inverse problem. Helmholtzian theory of the kind proposed by Rock was a bit closer because it was based on a likelihood principle that included elements resembling the use of Bayes' rule to maximize a posterior probability (Rock, 1983, chapter 6).

After Shannon's (1948) formulation of information theory, there were several attempts to revive the Gestalt approach by defining the simplicity principle in a more precise way. In the case of perception of objects from a single image, the simplicity principle was defined by the minimum variance of angles, of lengths, as well as symmetry (Hochberg & McAlister, 1953; Attneave & Frost, 1969; Perkins, 1972, 1976). These studies showed that the subject's percept was systematically affected by the degree of simplicity of the 3D interpretations (this manipulation is equivalent to the structure degradation procedure described in the previous section), and that the judgments were quite reliable (consistent) across subjects. All these studies demonstrated that the visual system uses constraints, and these constraints lead to a unique solution of the inverse problem of perceptual interpretation. Furthermore, the results appeared consistent with the proposed definitions of simplicity.

Hatfield and Epstein (1985) raised some criticism. They pointed out that these studies confounded two theoretical statements: (i) the claim that the perceptual system uses the simplicity (minimum) principle and (ii) the definition of simplicity itself. By changing (ii), one can either support or reject (i). It is obvious that experimental studies cannot test the two statements separately. Note, however, that whereas this criticism is valid in the context of Fechnerian framework, it is not valid in the context of the framework of inverse problems. In the framework of inverse problems, one does not need to perform an experiment to verify whether constraints such as simplicity are involved in obtaining accurate percepts. The fact that constraints are involved is a mathematical consequence of the nature of the problem: constraints are needed because a solution of an inverse problem tends to be non-unique and unstable.

The simplicity of the interpretation is not the only constraint that could be used in perception. Mach (1906/1959) and von Helmholtz (1910/2000) suggested a potential role for a likelihood principle that states that the percept is the most likely interpretation of the proximal stimulus. (Mach pointed out that simplicity and the likelihood of the percept may represent two aspects of the same perceptual mechanism.) Even though the likelihood principle was introduced to psychology at the end of the 19th century, it did not play an important role in theories of perception for several dozen years. This principle received attention only after Gestalt psychologists pointed out the fundamental significance of *a priori* constraints in perception (Wertheimer, 1923/1958). Little was added to this until Brunswik (1956) illustrated empirically what Mach had conjectured half a century earlier, namely, that the simplicity principle may often lead to the most likely interpretation. The likelihood principle was not taken seriously before Gestaltists came to scene probably because most interesting applications of simplicity and likelihood are those that involve context (Wertheimer, 1925/1938). The role of context (configuration, Gestalt) was greatly underestimated by Helmholtzian and Wundtian theorists, who instead emphasized the role of local processing, expressed in the so-called ‘constancy hypothesis’ (Koffka, 1935). More recently, there has been a vigorous discussion (and antagonism) between the proponents of each of these principles (Hatfield & Epstein, 1985; Pomerantz & Kubovy, 1986; Leeuwenberg & Boselie, 1989; Sutherland, 1989). Such antagonism is not surprising because these two principles have usually been associated with the antagonistic philosophical traditions of nativism and empiricism. Despite many efforts (both theoretical and empirical), little progress has been made to establish which of these two principles is more important or more fundamental in human perception.

Recent work in the area of information theory has shed new light on this discussion, showing that the two principles are, in fact, closely related. Specifically, if one knows the probabilities in Eq. (7), one can find an optimal description language that leads to the shortest expected length of the description (Rissanen, 1983; Leclerc, 1989; Chater, 1996; Mumford, 1996; Li & Vitanyi, 1997). For such an optimal language, the length (in bits) of a description is equal to $-\log_2 p$. Thus, if the logarithms in Eq. (8) have a base of ‘2’, the right-hand side becomes an expression that represents the length of the code describing (i) the data (image) produced by a given object ($Y_X|X$) and (ii) the object (X). Clearly, maximizing the posterior probability (Eq. (7)) is mathematically equivalent to minimizing the description length, which represents simplicity (or economy) of the description.

4.2. Experimental results

Studying how the perceptual system solves the inverse problem of perceptual interpretation pertains to two different, although related, questions. The first asks which constraints are used by the human perceptual system, and how they are used, to guarantee a unique percept. Studies addressing this question often involve a restricted range of impoverished stimuli and viewing conditions. The second (more difficult and more interesting) question asks which constraints are used by the human perceptual system, and how they are used, to guarantee a veridical percept. Studies addressing this question should involve a wide range of complex stimuli and viewing conditions. In principle, one could study the veridicality of the percept for a single stimulus and a single viewing condition, e.g. a cube viewed monocularly from one particular direction. But a veridical percept in such a case could merely represent a bias towards the cube interpretation rather than the ability to solve an inverse problem of 3D shape perception. Prior research on the role of constraints concentrated on the uniqueness and stability of the percept. The role of constraints in achieving veridical percepts received much less attention, but it is the emphasis on veridical percepts that marks the adoption of the framework of inverse problems.

The review presented below shows that prior psychophysical experiments testing the operation of constraints almost invariably used the structure degradation procedure described earlier in this paper. It is important to point out, however, that not all aspects of the mechanisms involved in solving the inverse problem of perceptual interpretation can be answered by using the structure degradation (or any other psychophysical) procedure. One of the critical questions is whether a constraint guarantees a veridical and reliable percept for a given set of distal stimuli. To answer this question, one has to formulate mathematical or computational models of perceptual functions and test the models with distal stimuli.

4.2.1. Binocular vision

Quite possibly the first study on the operation of constraints in perception was that of Panum in 1858 (see Howard & Rogers, 1995). He showed that a point P on one retina can be fused with a point Q on the other retina only when P falls inside some area S centered around a point Q' corresponding to Q. The area S representing the spatial limit for binocular fusion is now called 'Panum's area'. Restricting the search for corresponding points to Panum's area reduces the number of possible matches and thus makes it easier to obtain a unique and accurate solution to the inverse problem of binocular correspondence. Tyler (1973) and then Burt and Julesz (1980) generalized the concept of a disparity limit to a disparity gradient limit. Ecological justification for the operation of a disparity gradient limit was demonstrated in computational studies: corresponding points produced by an object's surface often give rise to a disparity gradient below *1.0*; conversely, disparity gradients exceeding this limit are usually produced by incorrect matches (Pollard et al., 1985). Note that disparity limits represent the continuity constraint in Marr and Poggio's theory (Marr, 1982; Marr & Poggio, 1976). Mitchison (1988) tested the psychological plausibility of computational models that incorporate disparity constraints. He showed that the human visual system solves the correspondence problem by imposing planarity constraint. Specifically, it first establishes the correspondence for the boundaries and then for regions, by minimizing departure of the 3D points from planarity. Marr and Poggio also described other constraints that might be involved in solving binocular correspondence problem, namely, uniqueness and similarity (compatibility). The role of a similarity constraint was demonstrated by Prazdny (1985), who showed that the human visual system can fuse points even when their disparity gradient is greater than *1.0*, provided the non-corresponding points differ in contrast or size.

The correspondence problem could be simplified by imposing the epipolar constraint. A given physical point F and the centers of projections of the two eyes determine a plane, which intersects the two retinas. These intersections are called epipolar lines. It follows that the two images of F must be on the corresponding epipolar lines. As a result, the search for a corresponding point on the second retina can be restricted to the epipolar line. However, it does not seem that the visual system uses the epipolar constraint probably because it does not have accurate information about the torsion of the eyes (Howard & Rogers, 1995).

Next, consider Nakayama and Shimojo's (1992) experiment. Their stimuli provided unambiguous disparity for the ends of a horizontal line and for the intersection of this line with a vertical line. The regions between these points did not provide information about disparity. As a result, the orientation of the horizontal line in depth was ambiguous: the line could be perceived either as being in the frontal plane or as consisting of two parts slanted (or curved) in depth. In other words, the interpolation problem had more than one solution. Despite this mathematical ambiguity, almost all observers systematically reported a single percept corresponding to a line in the frontal plane. Evidently, the visual system is able to solve this problem, and it does so by using constraining assumptions. The explanation provided by Nakayama and Shimojo involves a likelihood principle, similar to that described by Mach (1906/1959)(pp. 213–214). Specifically, the given retinal image could have been produced by a number of viewing directions in the case of a line in the frontal plane, and by only one viewing direction in the case of a line slanted in depth. Thus, the former case is more likely. Next, they demonstrated by a simple modification of the proximal stimulus that the perceptual interpretation changes from a line in the frontal plane to a transparent patch in front of a cross. This new interpretation is again consistent with the likelihood principle.

Grimson (1982) addressed the problem of reconstructing a 3D surface from 3D points or features, assuming that the points have already been computed from binocular disparity or motion. Because the points provide sparse information but the percept corresponds to continues surfaces and contours, the stage of solving correspondence and computing the 3D coordinates of points must be followed by interpolation. Grimson's interpolation involved regularization algorithm that maximized surface smoothness.

4.2.2. Motion perception

As was the case with binocular vision, the perception of motion requires the visual system to solve an ill-posed problem of establishing a correspondence between features in two successive images. Ullman (1979) drew attention of the vision community to the importance of this problem. He conducted a series of psychophysical experiments to determine the operation of a constraint called 'affinity' between two competing choices. Affinity was defined by such geometrical properties of the image as distance ratio, size ratio and orientation difference. By changing these properties, he produced systematic and reliable changes in the direction of the perceived apparent motion. Thus, he showed that similarity and proximity are important in achieving a unique and stable perceptual interpretation. Ullman's work was subsequently extended by Burt and Sperling (1981), who showed that spatial and temporal proximity are much stronger predictors of motion correspondence than geometrical similarity among the elements inducing apparent motion. All these results demonstrated that motion correspondence involves local constraints operating on the retinal image. Later, He and Nakayama (1994) showed that motion correspondence also involves 3D global constraints.

The studies of motion correspondence described so far concentrated on determining constraints that are involved in achieving a unique and reliable percept. The veridicality of motion correspondence was studied by Hildreth (1984). She analyzed the problem of determining the velocity field from the changing retinal image produced by a motion in 3D space. The retinal image was assumed to contain one or more contours, and the computation was based on the motion of the elements of the contours. This problem is ill posed, in part due to an aperture problem, which refers to the fact that a local mechanism analyzing motion of a line segment can measure only one component of this motion, the one that is normal to the line segment. The component tangent to the line segment is not visible to any mechanism. Hildreth showed that if image motion is produced by a rigid 3D object whose edges are straight-line segments, the smoothness constraint guarantees a veridical interpretation. In other cases, the interpretation may not be veridical; in such cases, however, the interpretation tends to agree with illusory perceptions of human observers (see Horn & Schunck, 1981 for an analogous study in the case of a flow field without contours).

Next, consider the problem of inferring the depth and shape of a solid object from a moving image. Wallach and O'Connell (1953) demonstrated that a shadow of a rotating solid object often gives rise to the percept of the object (kinetic depth effect). By manipulating the geometry of the stimulus, they determined that the 3D percept is produced only when the object's shadows 'display contour lines that change their direction and their length'. Wallach and O'Connell did not formulate a theory explaining the kinetic depth effect. Such a theory (called structure from motion) was proposed by Ullman (1979). The theory says that three orthographic views of four (or more) 3D non-coplanar points rotating in a rigid fashion allow for a veridical reconstruction of the 3D structure of the points (up to depth reflection). This theory accounts for a number of results described by Wallach and O'Connell. In particular, the experimental conditions that did not produce a 3D percept in Wallach and O'Connell's study do not satisfy the necessary condition for a unique interpretation in Ullman's theory. Note that the rigidity assumption in Ullman's theory plays a role of an implicit constraint. The reconstruction is veridical only if the object is rigid. Subsequently, Ullman (1984) generalized his theory by using rigidity as an explicit constraint. This new algorithm handles both rigid and non-rigid objects and can account for some results where the perceptual interpretation of a rigid rotating object is not veridical (Loomis & Eby, 1988). More recent studies extended Ullman's work to other cases, in which constraints of the fixed axis or planar motion were included (e.g. Hoffman & Bennett, 1986). To summarize, prior studies demonstrated that human observers are able to infer 3D relations from a moving image by applying one or more constraints. Psychophysical experiments concentrated, however, on a very restricted set of 3D shapes and viewing conditions. Therefore, these results did not shed much light on which constraints are used, and how they are used in achieving a veridical 3D percept. Computational studies showed that such constraints as rigidity may lead to a veridical interpretation, but only when several revolutions are used. Thus, the question arises as to whether veridical shape perception from motion involves more visual information (several revolutions) or more effective constraints. Existing results support the latter; this point will be elaborated in Section 4.2.5.

4.2.3. Color and lightness perception

Systematic studies of lightness and color constancy began with experiments by Katz in 1911 (see Gilchrist, 1994) and Gelb (1929/1938). When there is only one surface in the field of view (and this surface is homogeneous with respect to albedo, i.e. there is no visible texture), then the percept of the color of the surface along the black–gray–white dimension (its 'lightness') is neither veridical nor stable. Instead, the percept is correlated with the intensity of the illuminating light. This result is related to the fact that the intensity of the reflected light confounds the albedo with the intensity of the incident light. In order for the percept to be stable, there must be at least two surfaces of different albedos in the field of view. The percept of albedo is determined by the ratio of the intensities of light reflected by the two surfaces, at least for simple configurations involving Lambertian surfaces in the frontal plane (Hess & Pretori, 1894/1970; Wallach, 1948). For a given pair of surfaces, changes in the intensity of the illuminating light do not change the ratio of the intensities of the reflected light. Thus, the constancy of the ratio is a necessary condition for the constancy of the albedos, and it can be used in recognition of the albedo of a surface. The ratio rule is useful only under the assumption (implicit constraint) that spatial changes of the intensity of the illuminant are smooth, and changes of the albedo are abrupt.

Gilchrist and Jacobsen (1983) tested the generality of the ratio rule. The subject viewed a 3D scene containing familiar or unfamiliar objects and was asked to judge albedos of the 3D objects. The viewing was either direct or through a veiling luminance. The veil component increased all luminances by the same additive factor, reducing ratios by a multiplicative factor of up to 15. The judgments of the albedos were veridical in both veil and no-veil viewing. It follows that the judgments agreed with the ratio rule only in the no-veil condition, and they violated the rule in the veil condition. When the veil and no-veil viewing conditions were applied to a set of planar patches in the frontal plane (Mondrian stimulus), the judgments in both conditions agreed with the ratio rule and were veridical only in the no-veil condition. These results indicate that color perception is aided (constrained) by the information about the geometry of the 3D scene and that color constancy operates over the range of viewing conditions that are more general than those to which the ratio rule applies. It is known that 3D shape and space perception involves *a priori* constraints. Thus, Gilchrist and Jacobsen's results do not represent a mere integration of visual cues across visual channels (color and depth). Instead, they represent an integration of constraints across visual channels, which seems to be characteristic for solutions of difficult inverse problems.

The ratio rule can be applied to all three types of cones independently. This mechanism of color constancy is consistent with von Kries's (1902/1970) adaptation hypothesis, and it has been used in several theories of color constancy (Horn, 1974; Land, 1986; Brainard & Wandell, 1992; Foster & Nascimento, 1994). It is known that the spectrum of daylights varies with only two degrees of freedom: intensity and the correlated color temperature. It follows that it should be possible to reformulate von Kries's rule by removing one degree of freedom. This was done by Lee (1990) and then by Wei, Pizlo, Wu, and Allebach (1999). Spectral properties of stimuli are coded in the visual system by cones and then are transformed to opponent channels, which code colors: red–green, yellow–blue, bright–dark (Hurvich & Jameson, 1951). When a Munsell surface is illuminated by one daylight first and then by another daylight, the response of the red–green channel is the same in these two cases (the red–green channel is invariant). The chromatic change across the two illuminants is represented in the yellow–blue channel, and the intensity change is represented in the brightness channel. Clearly, the organization of the human color system, i.e. the use of blue–yellow and red–green channels for coding hues, reflects the highly constrained nature of natural illuminants. This does not imply that this constraint is used by the visual system, but it indicates that the visual system 'knows' the properties of this constraint. Preliminary experiments showed that human color constancy is more reliable for daylights than for non-daylights (Wei et al., 1999). These results suggest that the constrained nature of daylight is actually used in the perceptual mechanisms responsible for color constancy.

4.2.4. Figure–ground segregation

Figure–ground segregation refers to the perceptual ability to perceive which regions and contours in the retinal image represent the object (figure) as opposed to the background. Consider the detection of object contours. Uttal (1975) was the first to test the role of constraints in contour detection. His stimulus consisted of randomly positioned dots and a target, which was a set of dots representing a line. The line was either a straight-line segment or a curve. The subject's task was to decide whether the target was present in the stimulus. Clearly, the task of finding a set of dots representing a line in a stimulus consisting of N dots is ill posed (each subset of dots represents a number of possible contours). To solve the task, one has to impose constraints on the properties of contours. Uttal examined three constraints, namely, proximity, smoothness and regularity. All of these constraints are satisfied when the target is represented by a set of equally spaced dots that form a straight-line segment with small inter-dot distances. Indeed, such a target was the easiest to detect. When the inter-dot distances were large or irregular, the line was not straight, or it was not smooth, detectability dropped. These results suggest that all three constraints are necessary in contour detection (it is possible that some of these constraints are implicit, or that one constraint could result from the implementation of the others).

Uttal's results were generalized by Pizlo et al. (1997), who analyzed the role of global constraints. They showed that the knowledge (memory) of the target's shape and orientation improved its detectability. However, the memory of the target shape alone, when the orientation of the target was random, did not improve detectability. It seems that the visual system can use only such global information as can be translated into local constraints (although the local constraints can be applied on several levels of scale and resolution—Pizlo, Rosenfeld, & Epelboim, 1995). Other studies of constraints in solving the inverse problem of contour detection and figure–ground segregation include those of Bouman and Liu (1991), Vos and Helsper (1991), Field, Hayes and Hess (1993), Lee (1995), Alter and Basri (1998), Nakayama, Shimojo, and Silverman (1989), Shashua and Ullman (1988), and Geisler, Perry, Super, and Gallogly (2001). These studies demonstrated the operation of local constraints such as smoothness of contours and surfaces, and global constraints such as homogeneity of regions and saliency.[4]

4.2.5. Shape perception

The approach of Hochberg and McAlister (1953), Attneave and Frost (1969) and Perkins (1976) to the role of constraints in shape perception was adopted and subsequently elaborated by psychophysicists in the 1990s. Knill (1992) tested the role of a geodesic constraint. Geodesics can be defined as the shortest or straightest lines on a surface (Hilbert & Cohn-Vossen,

[4] As Ullman (1979) and Marr (1982) pointed out, at least some of the global constraints can be implemented in a purely bottom-up fashion by performing local operations.

1990) and fully characterize the intrinsic properties of surfaces. Knill's study was motivated by Stevens' (1981, 1986) theoretical and empirical studies suggesting that human perception of surfaces involves the assumption that the surface contours are planar geodesics, which are also lines of curvature. Knill presented subjects with a picture of a shaded solid object on which he superimposed a patch whose contours as projected on the image were either straight-line segments (geodesics of a planar surface) or geodesics of the surface of the object. If surface geodesics are involved in the perception of shape from shading, the observer should be able to classify the patches into two categories: a planar, transparent patch 'floating' in front of the object or a part of the surface of the 3D object. Knill found that the subjects could reliably perform this classification, supporting the claim that the visual system uses a geodesic constraint. Observers are able to interpret shapes of surfaces from lines that are not geodesics. Therefore, Knill also tested the usefulness of two soft measures of the geodesic constraint that evaluated the variability of geodesic curvature and showed that subjects classified curves as surface contours more often when the variability of geodesic curvature was small.

Mamassian and Landy (1998) tested the perception of the 3D shape represented by surface contours. The subject's task was to classify the surface as elliptic or hyperbolic. Subjects' classifications were systematically affected by the stimulus geometry suggesting that the percept involved some constraints. The authors formulated a Bayesian model, which incorporated preference for convex surfaces, for surface orientation corresponding to viewing the object from above, and for surface contours representing lines of curvature that are geodesics.

To summarize, prior studies showed that perception of color, depth, motion, shape and figure–ground segregation require *a priori* constraints because the information available in the proximal stimulus is inherently ambiguous and thus determines an infinite number of possible interpretations. A question arises, however, about the generality of these results. Specifically, one can argue that constraints were needed because the stimuli in the studies described were impoverished. Perhaps constraints would no longer be necessary, or their role would be minimized, if the proximal stimulus contained more information. By obtaining more cues about the distal stimulus and its context, the observer might be able to disambiguate the proximal stimulus and obtain a veridical interpretation of the distal stimulus in a bottom-up fashion, without using *a priori* constraints. Thus, the question asked here is whether constraints are used by the perceptual system *only* to disambiguate the sensory information. Perhaps they are essential for perceptual processing and are always involved in solving inverse problems of perceptual interpretation. According to the framework of inverse problems, constraints are always involved. But does this framework apply to all cases of perceptual interpretation, including those that provide the observer with strong contextual cues?

Recently, Pizlo and Stevenson (1999) and Pizlo, Chan, and Stevenson (1999) addressed this question in studies of shape perception involving motion and binocular disparity cues. In one study, subjects were shown two sequences of motion (20 images) of a rigid 3D structure and were asked to judge whether the shapes of the structures were the same (Pizlo & Stevenson, 1999).[5] The two sequences of motion were 90 deg apart. As a result, the sequences contained different images, regardless of whether the shapes presented in the two sequences were the same or different. In the first condition, the stimuli were symmetric polyhedra represented by planar contours. In this condition, performance was very reliable. In the second condition, the contours were no longer planar. This caused performance to drop by a factor of two, which indicates that perceptual mechanisms use the planarity constraint, either explicitly by incorporating a measure of non-planarity in the cost function, or implicitly by using algorithms that assume that the contours are planar (Hildreth, 1984). In the latter case, an object with non-planar contours would be likely to lead to a percept of a non-rigid object with planar contours. The subjects reported seeing rigid objects with non-planar contours, which suggests that the planarity constraint is included explicitly in the cost function. In the third condition, the subject was presented with only three planar faces of the polyhedron from the first condition, each of the two pairs of the faces sharing one vertex. In this condition, the mutual constraining among the faces of the object produced by topological relations among them was minimized. Again, performance dropped, in comparison to the first condition, by a factor of two. This performance drop suggests that perceptual mechanisms use constraints that rely on the 3D topology of the object. In the last condition, where several constraints were removed (the stimulus was a polygonal line connecting the vertices of the polyhedron in a random order), performance dropped to chance level. All of these results were replicated with binocular viewing of stationary shapes from a close viewing distance (Pizlo et al., 1999).

Results of Pizlo and Stevenson's (1999) and Pizlo et al.'s (1999) studies show that *shape perception involves a priori constraints even when depth cues are strong, and that these constraints are used not only to disambiguate the proximal stimulus. These constraints are used because the perceptual interpretation is not only an ill-posed problem but also an ill-conditioned problem.*

[5] See our Web page for examples of the stimuli: http://bigbird.psych.purdue.edu/shapedemo.

The fact that the perception of shapes relies critically on *a priori* constraints, regardless of the amount of information present in the proximal stimulus (single view vs. two views vs. 20 views), *suggests* the following hypothetical order of perceptual processing. Perceptual processing begins with the reconstruction of a 3D shape from one of the retinal images by applying constraints. If more retinal images are provided from binocular vision or from motion, the 3D percept is corrected (see Chan, Stevenson, & Pizlo, submitted for publication for details of the model). As pointed out earlier, a 3D reconstruction based on binocular disparity is extremely unstable in the presence of visual or motor noise, while monocular shape reconstruction from a noisy image with the use of constraints such as minimum variance of angles and planarity is quite stable. Furthermore, monocular shape reconstruction almost never produces topological errors (Leclerc & Fischler, 1992). These observations suggest that shape reconstruction from binocular disparity and motion should begin with monocular shape reconstruction.

The experiments of Pizlo and colleagues resolved an apparent controversy in the perceptual literature. Specifically, Rock and DiVita (1987) demonstrated a complete failure of shape constancy in the case of wire objects viewed binocularly. Biederman and Gerhardstein (1993), however, demonstrated reliable shape constancy from a single image of a structured object. These two results appear contradictory within the framework of Fechnerian psychophysics. How is it possible that when more depth cues are provided, the percept is much less veridical? The answer is straightforward within the framework of inverse problems: *The percept of the shape of an object is not a mere result of the stimulation of the retina; the percept depends critically on the operation of a priori constraints. For unnatural objects, such as nonsense wire stimuli, adding depth cues will never compensate for the absence of effective constraints*. This example illustrates the kind of explanatory power that one might require before the new framework can replace the old one.

5. Concluding remarks

Following the publication of Marr's (1982) seminal book, many studies began to treat the perceptual system as a set of individual modules: binocular, motion, texture, etc. An inverse problem was formulated for each module. It is obvious that these modules are not independent. They interact (e.g. shape with lightness). It follows that one should study more than a single type of perceptual information or constraint. The analysis of the nature of inverse problems in perception, presented in this paper, suggests that perception should be modeled as a *hierarchy of inverse problems* (e.g. figure–ground segregation, binocular and motion correspondence, depth reconstruction, surface interpolation). Once the need for such a hierarchy is acknowledged, a number of questions arise. Is such a hierachical approach realizable computationally? Furthermore, would it be plausible psychologically? How should the hierarchy be structured to produce best results (i.e. stable interpretations, short reaction times, etc.)? For example, does the visual system always solve binocular and motion correspondence before proceeding to reconstruct depth and objects? Is the order of these stages fixed or dependent on the proximal stimulus? This kind of question arises as soon as one adopts the framework of inverse problems. Note that they have only been touched on earlier in this paper and have received little attention by those who have adopted the inverse problem framework so far.

Prior research emphasized the use of knowledge about the distal stimuli. It is reasonable to assume that the human perceptual system is familiar not only with the distal stimuli but also with some properties of the cost functions that are used to solve the inverse problem of perceptual interpretation (Eqs. (6)–(14)). This knowledge can be used to improve the *efficiency* of the search for the global minimum of the function by speeding up this search or avoiding local minima. The continuation method in reconstructing shapes of polyhedra (Leclerc & Fischler, 1992; Chan et al., submitted for publication) is one example of a realization of this goal. It avoids local minima when the cost function is sufficiently regular. A second example is the choice of a good starting point to hasten the search for the global minimum. Consider the well-known perceptual bias for 3D interpretations with right angles in reconstructing the shapes of 3D objects (Perkins, 1976). In the past, the bias towards rectangular interpretations was assumed to reflect a constraint of the environment, but it is possible that the bias for 3D interpretations with right angles reflects a constraint of the process that enhances efficiency of computations.

Now, consider the ideal-observer analysis as an approach to modeling the percept within the framework of inverse problems (Kersten & Schrater, in press). Conventionally, the ideal observer has been used as a method of evaluating of how much information is *lost* in the process of coding a proximal stimulus in the perceptual system (Geisler, 1989). In the framework of inverse problems, however, we are concerned with how much and what kind of information is *added* to the proximal stimulus in the process of inferring properties of a distal stimulus. Thus, the conventional use of the ideal-observer analysis is not applicable to inverse problems. Kersten and colleagues generalized the ideal-observer analysis by using Bayesian models of perceptual decisions (Knill & Kersten, 1991). The priors in the Bayesian models represent the properties of the natural environment of the observer. Note, however, that these properties are not known, which means that the ideal observer may be less than *ideal* for a given set of stimuli or a given task. To avoid this problem, the researcher should formulate several ideal-observer models, with different models representing different amounts of information about the physical environment (see Liu et al.'s 1995 ideal vs. 'true' ideal vs. super-ideal). It follows that the general ideal observer comes with a cost. It requires a number of free parameters, which makes it less attractive than the parameter-free, conventional ideal observer (Liu, Knill, & Kersten, 1995; Liu & Kersten, 1998).

In conclusion, now that we are approaching the 150th anniversary of the publication of Fechner's influential book, which launched the psychophysical study of perception, it is clear that a great deal remains to be done. We are still far from understanding the nature of the processes that allow our percepts to be veridical. This paper set out to make it clear that an alternative approach to Fechner's is required; Fechner's approach cannot provide the answers we desire. The nature of one such alternative approach, in which perception is considered as an inverse problem, has been described. Its current status, methods, and virtues have been sketched out. One hopes that perception researchers will explore the possibilities it opens up.

Acknowledgements

I thank Dr. Ellen Hildreth for her suggestions, comments and criticism that led to a substantial improvement of this paper. I would also like to acknowledge comments and suggestions of Drs. Azriel Rosenfeld and Robert Steinman, as well as those of the anonymous reviewers.

References

Alter, T. D., & Basri, R. (1998). Extracting salient curves from images: an analysis of the saliency network. *International Journal of Computer Vision*, *27*, 51–69.

Attneave, F. (1959). *Applications of information theory to psychology*. New York: Holt.

Attneave, F., & Frost, R. (1969). The determination of perceived tridimensional orientation by minimum criteria. *Perception & Psychophysics*, *6*, 391–396.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions from three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, *19*, 1162–1182.

Bouman, C., & Liu, B. (1991). Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *13*, 99–113.

Bouman, C., & Sauer, K. (1993). A generalized gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing*, *2*, 296–310.

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color matching: how color appearance depends on the illuminant. *Journal of the Optical Society of America. Part A, Optics, Image Science and Vision*, *A9*, 1433–1447.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. *Perception*, *9*, 671–682.

Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Pychological Review*, *88*, 171–195.

Chan, M. W., Pizlo, Z., & Chelberg, D. (1999). Binocular shape reconstruction: psychological plausibility of the 8 point algorithm. *Computer Vision & Image Understanding*, *74*, 121–137.

Chan, M. W., Stevenson, A. K. & Pizlo, Z. (submitted for publication). A model of human binocular shape reconstruction based on a regularization approach.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, *103*, 566–581.

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge: University Press.

Falmagne, J.-C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.

Fechner, G. T. (1860/1966). *Elements of psychophysics*. New York: Holt, Rinehart & Winston.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local association field. *Vision Research*, *33*, 173–193.

Foster, D. H., & Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London. Series B: Biological Sciences (London)*, *257*, 115–121.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discrimination. *Psychological Review*, *96*, 267–314.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, *41*, 711–724.

Gelb, A. (1929/1938). Colour constancy. In W. D. Ellis, *A source book of Gestalt psychology* (pp. 196–209). London: Routledge & Kegan Paul.

Gibson, J. J. (1950). *The perception of the visual world*. Boston, MA: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Gilchrist, A. L. (1994). *Lightness, brightness, and transparency*. Hillsdale, NJ: Lawrence Erlbaum.

Gilchrist, A. L., & Jacobsen, A. (1983). Lightness constancy through a veiling luminance. *Journal of Experimental Psychology: Human Perception & Performance*, *9*, 936–944.

Grimson, W. E. L. (1982). A computational theory of visual surface interpolation. *Philosophical Transactions of the Royal Society (London) B*, *298*, 395–427.

Hartley, R. I., & Sturm, P. (1997). Triangulation. *Computer Vision and Image Understanding*, *68*, 146–157.

Hatfield, G., & Epstein, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin*, *97*, 155–186.

He, Z. J., & Nakayama, K. (1994). Perceived surface shape not features determines correspondence strength in apparent motion. *Vision Research*, *34*, 2125–2135.

Hess, C., & Pretori, H. (1894/1970). Qantitative investigation of the lawfulness of simultaneous brightness constrast (H. Flock & J. H. Tenney, Trans.). *Perceptual and Motor Skills*, *31*, 947–969.

Hilbert, D., & Cohn-Vossen, S. (1990). *Geometry and imagination*. New York: Chelsea.

Hildreth, E. C. (1984). The computation of the velocity field. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *221*, 189–220.

Hochberg, J., & McAlister, E. (1953). A quantitative approach to figural 'goodness'. *Journal of Experimental Psychology*, *46*, 361–364.

Hoffman, D., & Bennett, B. (1986). The computation of structure from fixed axis motion: rigid structures. *Biological Cybernetics*, *54*, 1–13.

Horn, B. K. P. (1974). Determining lightness from an image. *Computer Graphics and Image Processing*, *3*, 277–299.

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, *17*, 185–203.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. New York: Oxford University Press.

Hurvich, L. M., & Jameson, D. (1951). The binocular fusion of yellow in relation to color theories. *Science*, *114*, 199–202.

Ikeuchi, K., & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, *17*, 141–184.

Kersten, D. (1999). High-level vision as statistical inference. In M. S. Gazzaniga, *The new cognitive neurosciences* (pp. 353–363). Cambridge, MA: MIT Press.

Kersten, D. & Schrater, P. (in press). Pattern inference theory: a probabilistic approach to vision. In R. Mausfeld & D. Heyer (Eds.), *Perception and the physical world*. Chichester, UK: Wiley.

Knill, D. C. (1992). Perception of surface contours and surface shape: from computation to psychophysics. *Journal of the Optical Society of America. Part A, Optics, Image Science and Vision*, *A9*, 1449–1464.

Knill, D. C., & Kersten, D. (1991). Ideal perceptual observers for computation, psychophysics and neural networks. In R. J. Watt, *Pattern recognition by man and machine* (pp. 83–97). London: MacMillan.

Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement*, vol. 1. San Diego, CA: Academic Press.

Land, E. H. (1986). Recent advances in retinex theory. *Vision Research*, *26*, 7–22.

Leclerc, Y. G. (1989). Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, *3*, 73–102.

Leclerc, Y. G., & Fischler, M. A. (1992). An optimization-based approach to the interpretation of single line drawings as 3D wire frames. *International Journal of Computer Vision*, *9*, 113–136.

Lee, H.-C. (1990). A computational model for opponent color encoding. In *Proceedings of the annual conference of the Society for Imaging Science & Technology*. *Springfield*, *VA* (pp. 178–181).

Lee, T. S. (1995). A Bayesian framework for understanding texture segmentation in the primary visual cortex. *Vision Research*, *35*, 2643–2657.

Leeuwenberg, E., & Boselie, F. (1989). How good a bet is the likelihood principle? In B. A. G. Elsendoorn, & H. Bouma, *Working models of human perception*. New York: Academic Press.

Li, M., & Vitanyi, P. (1997). *An introduction to Kolmogorov complexity and its applications*. New York: Springer.

Liu, Z., & Kersten, D. (1998). 2D observers for human 3D object recognition? *Vision Research*, *38*, 2507–2519.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, *35*, 549–568.

Loomis, J. M., & Eby, D. W. (1988). Perceiving structure from motion: failure of shape constancy. In *Proceedings of the International Conference on Computer Vision* (pp. 383–391). Washington, DC: Computer Society Press.

Luce, R. D., Bush, R. R., & Galanter, E. (1963). *Handbook of mathematical psychology*, vol. 1. New York: Wiley.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement*, vol. 3. San Diego, CA: Academic Press.

Luneburg, R. K. (1947). *Mathematical analysis of binocular vision*. Princeton, NJ: Princeton University Press.

Mach, E. (1906/1959). *The analysis of sensations*. New York: Dover.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, *30*, 1763–1779.

Maloney, L. T. (in press). Statistical decision theory and biological vision. In R. Mausfeld & D. Heyer (Eds.), *Perception and the physical world*. Chichester, UK: Wiley.

Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research*, *38*, 2817–2832.

Marill, T. (1991). Emulating the human interpretation of line drawings as three-dimensional objects. *International Journal of Computer Vision*, *6*, 147–161.

Marr, D. (1982). *Vision*. New York: W.H. Freeman.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, *194*, 283–287.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt.

Mitchison, G. (1988). Planarity and segmentation in stereoscopic matching. *Perception*, *17*, 753–782.

Mumford, D. (1996). Pattern theory: a unifying perspective. In D. C. Knill, & W. Richards, *Perception as Bayesian inference* (pp. 25–62). New York: Cambridge.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357–1363.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, *18*, 55–68.

Narens, L. (1985). *Abstract measurement theory*. Cambridge, MA: MIT Press.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton.

Palmer, J. (1994). Set-size effects in visual search: the effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*, 1703–1721.

Perkins, D. N. (1972). Visual discrimination between rectangular and nonrectangular parallelopipeds. *Perception & Psychophysics*, *12*, 396–400.

Perkins, D. N. (1976). How good a bet is good form? *Perception*, *5*, 393–406.

Pizlo, Z. (1994). A theory of shape constancy based on perspective invariants. *Vision Research*, *34*, 1637–1658.

Pizlo, Z., Chan, M. W., & Stevenson, A. K. (1999). Binocular shape perception: top-down approach. *Investigative Ophthalmology & Visual Science*, *40*(Supplement), 374.

Pizlo, Z., Rosenfeld, A., & Epelboim, J. (1995). An exponential pyramid model of the time-course of size processing. *Vision Research*, *35*, 1089–1107.

Pizlo, Z., & Loubier, K. (2000). Recognition of a solid shape from its single perspective image obtained by a calibrated camera. *Pattern Recognition*, *33*, 1675–1681.

Pizlo, Z., Salach-Golyska, M., & Rosenfeld, A. (1997). Curve detection in a noisy image. *Vision Research*, *37*, 1217–1241.

Pizlo, Z., & Stevenson, A. K. (1999). Shape constancy from novel views. *Perception & Psychophysics*, *61*, 1299–1307.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*, 314–319.

Pollard, S. B., Mayhew, J. E. W., & Frisby, J. P. (1985). PMF: a stereo correspondence algorithm using a disparity gradient limit. *Perception*, *14*, 449–470.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization. In K. R. Boff, L. Kaufman, & J. P. Thomas, *Handbook of perception and human performance*. New York: Wiley.

Prazdny, K. (1985). On the disparity gradient limit for binocular fusion. *Perception & Psychophysics*, *37*, 81–83.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Annals of Statistics*, *11*, 416–431.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, *19*, 280–293.

Rock, I., Wheeler, D., & Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology*, *21*, 185–210.

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, purpose and teleology. *Philosophy of Science*, *10*, 18–24.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 623–656, 379–423.

Shashua, A., & Ullman, S. (1988). Structural saliency: the detection of global salient structures using a locally connected network. In *Proceedings of the international conference on computer vision* (pp. 321–327). Washington, DC: IEEE Computer Society Press.

Shulman, D., & Aloimonos, J. Y. (1988). Non-rigid motion interpretation: a regularized approach. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *233*, 217–234.

Sinha, P., & Adelson, E. H. (1992). Recovery of 3-D shape from 2-D wireframe drawings. *Investigative Ophthalmology & Visual Science*, *33*(Supplement), 825.

Stevens, K. A. (1981). The visual interpretation of surface contours. *Artificial Intelligence*, *17*, 47–73.

Stevens, K. A. (1986). Inferring shape from contours across surfaces. In A. P. Pentland, *From pixels to predicates*: *recent advances in computational & robotic vision* (pp. 93–110). Norwood, NJ: Ablex.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement*, vol. 2. San Diego, CA: Academic Press.

Sutherland, S. (1989). Simplicity is not enough. In B. A. G. Elsendoorn, & H. Bouma, *Working models of human perception*. New York: Academic Press.

Tarantola, A. (1987). *Inverse problem theory*. New York: Elsevier.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of ill-posed problems*. Washington, DC: Winston.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*, 507–545.

Tyler, C. W. (1973). Stereoscopic vision: cortical limitations and a disparity scaling effect. *Science*, *181*, 276–278.

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

Ullman, S. (1984). Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, *13*, 255–274.

Uttal, W. R. (1975). *An autocorrelation theory of form detection*. Hillsdale, NJ: Lawrence Erlbaum.

von Helmholtz, H. (1910/2000). *Treatise on physiological optics*. Bristol: Thoemmes (translated from German, J. P. C. Southall).

von Kries, J. (1902/1970). Chromatic adaptation. In D. L. MacAdam, *Sources of color science* (pp. 109–119). Cambridge, MA: MIT Press.

Vos, P. G., & Helsper, E. L. (1991). Modelling detection of linear structure in dot patterns. In I. Dvorak, & A. V. Holden, *Mathematical approaches to brain functional diagnostics* (pp. 429–440). New York: Manchester University Press.

Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, *38*, 310–324.

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

Wei, G., Pizlo, Z., Wu, W., & Allebach, J. P. (1999). Opponent channels and color constancy. *Investigative Ophthalmology & Visual Science*, *40*(Supplement), 749.

Wertheimer, M. (1923/1958). Principles of perceptual organization. In D. C. Beardslee, & M. Wertheimer, *Readings in perception* (pp. 115–135). New York: D. van Nostrand.

Wertheimer, M. (1925/1938). Gestalt theory. In W. D. Ellis, *A source book of Gestalt psychology* (pp. 1–11). London: Routledge & Kegan Paul.