Dingledine, R. (1983). N-methyl-aspartate activates voltage-dependent calcium conductance in rat hippocampal pyramidal cells. *J. Physiol.*, **343**, 385–405.
Doyle, W. (1962). Operations useful for similarity-invariant pattern recognition. *J. ACM*, **9**, 256–267.
Durbin, R. and Rumelhart, D. E. (1989). Product units: a computationally powerful and biologically plausible extension to back-propogation networks. *Neural Computation*, **1**, 133–142.
Feldman, J. A. (1982). Dynamic connections in neural networks. *Biol. Cybern.*, **46**, 27–39.
Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. NY: Academic Press.
Giles, C. L. and Maxwell, T. (1987). Learning, invariances, and generalization in high-order neural networks. *Appl. Opt.*, **26**, 4972–4978.
Glunder, H. (1986). Neural computation of inner geometric pattern relations. *Biol. Cybern.*, **55**, 239–251.
Glunder, H. (1987). Invariant description of pictorial patterns via generalized autocorrelation functions. In *ASST '87*. ed. Meyer-Ebrecht, D. pp. 84–87. Berlin: Springer Verlag.
Hadeler, K. P. (1974). On the theory of lateral inhibition. *Kybernetik*, **14**, 161-165.
Kelso, S. R., Ganong, A. H. and Brown, T. H. (1986). Hebbian synapses in hippocampus. *Proc. Natl. Acad. Sci. USA*, **83**, 5326–5330.
Kröse, B. J. A. (1985). A structure description of visual information. *Patt. Recogn. Lett.*, **3**, 41–50.
Lippmann, R. P. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, **4**, 4–22.
Lohmann, A. W. and Wirnitzer, B. (1984). Triple correlations. *Proc. IEEE*, **72**, 889–901.
MacDermott, A. B. and Dale, N. (1987). Receptors, ion channels and synaptic potentials underlying the integrative actions of excitatory amino acids. *TINS*, **10**, 280–284.
McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of ideas immanent in nervous activity. *Bull. Math. Biophys.*, **5**, 115–133.
McLaughlin, J. A. and Raviv, J. (1968). Nth-order autocorrelations in pattern recognition. *Information and Control*, **12**, 121–142.
Marr, D. (1982). *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: Freeman.
Minsky, M. L. and Papert, S. A. ([1]1969, [2]1988). *Perceptrons. An Introduction to Computational Geometry*. Cambridge, MA: MIT Press.
Moore, D. J. H. and Parker, D. J. (1974). Analysis of global pattern features. *Pattern Recognition*, **6**, 149–164.
Nicoll, R. A., Kauer, J. A. and Malenka, R. C. (1988). The current excitement in long-term potentiation. *Neuron*, **1**, 97–103.
Pao, Y.-H. (1989). *Adaptive Pattern Recognition and Neural Networks*. Reading, MA: Addison-Wesley.
Phillips, C. G., Zeki, S. and Barlow, H. B. (1984). Localization of function in the cerebral cortex. *Brain*, **107**, 328–361.
Pitts, W. and McCulloch, W. S. (1947). How do we know universals. The perception of auditory and visual forms. *Bull. Math. Biophys.*, **9**, 127–147.
Rosenblatt, F. (1962). *Principles of Neurodynamics – Perceptrons and the Theory of Brain Mechanisms*. Washington, DC: Spartan Books.
Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel Distributed Processing 1*. Cambridge, MA: The MIT Press.
Rumelhart, D. E., Hinton, G. E. and McClelland, J. L. (1986). A general framework for parallel distributed processing. In *Parallel Distributed Processing 1*. eds. Rumelhart, D. E. and McClelland, J. L. pp. 45–76. Cambridge, MA: MIT Press.
Schürmann, J. (1977). *Polynomklassifikatoren für die Zeichenerkennung*. Munich: Oldenbourg Verlag.
Sebestyen, G. S. (1962). *Decision-making Processes in Pattern Recognition*. NY: Macmillan.
Uesaka, Y. (1971). Analog perceptrons: on additive representations of functions. *Information and Control*, **19**, 41–65.
Uesaka, Y. (1975). Analog perceptron: its decomposition and order. *Information and Control*, **27**, 199–217.
Watt, R. (1988). *Visual Processing: Computational, Psychophysical, and Cognitive Research*. Hove, UK: Erlbaum.

Knill, D., & Kersten, D. (1991). Ideal perceptual observers for computation, psychophysics and neural networks. In R. J. Watt (Ed.), Pattern recognition by man and machine (vision and visual dysfunction). McMillan.

7 Ideal Perceptual Observers for Computation, Psychophysics and Neural Networks

David C. Knill and Daniel Kersten

We do not see optical images in an optical space, but we perceive the bodies round about us in their many and sensuous qualities. Ernst Mach, 1897.

Introduction

That our phenomenal perceptual world, or a part of it, is composed of the properties of objects and surfaces arrayed in the richness of three dimensions provides one of the great puzzles in psychology: How are such properties perceived when the information available to an organism takes the form of a two-dimensional layout of light energy impinging on the surface of the retina? A more specific problem is that the variables describing surface properties such as reflectance, transmittance, shape and illumination are locally confounded in the perspective map to an image. How, in the face of apparent ambiguity, does the human visual system arrive at a stable and accurate percept of scenes?

The accuracy of percepts in our normal behavioural environment implies the existence of ecological constraints on the structure of scene attributes which serve to remove some, if not all, of the ambiguity. Furthermore, the visual system apparently takes advantage, either implicitly or explicitly, of these constraints. The very fact of perception implies the existence of some rudimentary structure in the environment. The organization of the enviroment into cohesive objects with surfaces imposes a tremendous amount of structure, reducing by one the dimensionality of the space needed to describe a scene (taking us from a description of points in a volume to a description of points on disjoint surfaces). There are other strong constraints on the environment; object motions are continuous and often rigid, scenes are illuminated by a small number of light-emitting surfaces (often not visible), and many more.

What are the constraints on environmental structure? To what extent does the human visual system take advantage of these constraints? Are these constraints enough to make the apparently 'ill posed' problem of visual perception, as it is often presented in computer vision (Poggio *et al.*, 1985) work, well-posed? In this chapter, we will introduce a probabilistic framework for understanding visual perception which serves to organize inquiry into these issues. The development is an extension of an earlier proposal of a Bayesian model for ideal observers applied to the perception of scene attributes (Kersten, 1990). It is also similar in spirit to the general model for perception proposed in *Observer Mechanics* (Bennett *et al.*, 1989), though our development differs in its emphasis on the components of a Bayesian formulation of ideal observers.

A Probabilistic Approach to Perception

Percepts are statistical inferences about the scene which an observer is viewing. They are an observer's best guess about whatever characteristic of a scene are of interest to that observer. Somewhat more formally, we say that the percept of a scene results from the selection of the most likely scene to have given rise to an image based on a conditional distribution, $p(scene|image)$ (the probability of a scene conditional on an image). The form of the distribution is implicitly defined in the visual system of the observer.

At first reading, this may appear to be a strong claim about the nature of perceptual processing. Note, however, that the form of the distribution governing the statistical inferences is defined by the processing characteristics of the system. Such a probabilistic framework can therefore accommodate a broad range of possible systems. In order

to create a predictive model within the framework, one must specify some of the characteristics of the conditional distribution $p(scene|image)$ embodied in an observer. As an example, consider Gibson's hypothesis of direct perception. Within a probabilistic framework, this would be rephrased as an hypothesis that the conditional distribution is a Dirac delta function supported only at one point in an appropriate representational space.

What are the advantages of studying visual perception within a probabilistic framework? First, it is a natural framework for developing theories of competence for different perceptual tasks, and thus for measuring human performance on these tasks. Secondly, an expansion of the conditional distribution using Bayes' rule isolates the qualitatively different components of the problem of perception and leads to an empirical programme of research which deals with issues fundamental to different perspectives on perception, including both ecological and information processing perceptives. Both of these issues will be elucidated in the following sections, in which the concept of an ideal observer is defined and discussed.

Ideal Observers

Before defining the concept of an ideal observer, we will need some notation to characterize the environment and the information available to an observer for the perception of the environment. Let S be a complete characterization of a scene viewed by an observer. We consider S to be an element of a continuous stochastic ensemble Λ_S, with an associated probability density function $p_S(S)$. Λ_S specifies the environment of an observer. The ecological constraints on this environment are embodied in the density function, $p_S(S)$. As an example, scenes with rigidly moving objects are much more likely than scenes with elastically deforming objects. This would be reflected in the relative values of $p_S(S)$ for the two different types of scenes.

In general, humans perceive some subset and/or some high level function of the complete scene attributes in S. Let us therefore define S^* as characterizing those aspects of a scene which the visual system attempts to perceive. We call S^* the diorama – the term used for a three-dimensional museum display such as one built out of small figurines, papier mâché landscape, miniature trees and bushes. S^* might contain descriptions of such scene attributes as relative surface reflectances, surface shape and illumination direction. S^* is given by a one-to-one map η, $S^* = \eta(S)$, and is a member of a stochastic ensemble Λ_{S^*} (see Fig. 7.1). Λ_{S^*} may be continuous, discrete or of mixed type. For the sake of discussion, we will assume it to be continuous. (One can accommodate discrete ensembles, Λ_{S^*} and Λ_{I^*}, by replacing the probability density functions by probability mass functions, $P_{S^*}(S^*)$ and $P_{I^*}(I^*)$,

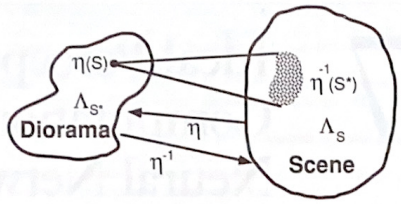

Fig. 7.1 *Schematic of the scene-to-diorama map, $\eta : \Lambda_S \rightarrow \Lambda_{S^*}$. Often, η will be many-to-one, and its inverse, one-to-many, as shown here.*

and the appropriate integrals by summations.) The probability density function associated with Λ_{S^*}, $p_{S^*}(S^*)$, is simply related to $p_S(S)$ by

$$p_{S^*}(S^*) = \int_{\eta^{-1}(S^*)} p_S(S)\,dS \tag{7.1}$$

where $\eta^{-1}(S^*)$ is the subset of Λ_S whose elements map under η to S^*.

A number of different formulations are possible for characterizing the information available to an observer. One can describe the information potentially available at each point in space prior to imaging. Leonardo da Vinci recognized this possibility. After describing the principle of pinhole projection, he said '... any object, ..., diffuses itself in circles, and fills the surrounding air with infinite images of itself. And is repeated, the whole everywhere, and the whole in every smallest part' (Da Vinci, 1970). Gibson (1979) calls this an optic array, which he defines as the bundle of light rays coming to each point in space. It can be more formally defined using the 'holoscopic' function (Adelson and Bergen, 1990), $H(\lambda, V_x, V_y, V_z, \phi, \theta, t)$, specifying the light level at wavelength λ projected to the point (V_x, V_y, V_z) from the direction in spherical coordinates ϕ and θ at time t. A more common characterization of visual information is as a pair of idealized retinal images (i.e. imaged through an ideal optical system, assuming no diffraction or optical aberrations). We define each image as a function, $I(\lambda, \phi, \theta, t)$, specifying the light level at wavelength λ impinging on the idealized retina at a point given by the spherical coordinates ϕ and θ at time t. The retina is often approximated as being planar, in which case the spherical coordinates ϕ and θ are replaced by rectangular coordinates x and y.

Gibson's formulation represents the information *potentially* available to an observer, while the retinal image represents the information actually sampled by the observer. The optic array is perhaps the appropriate representation to use when considering aspects of perception related to the dynamic interaction of observers with the environment; however, for consideration of other aspects of perception (e.g. the case of a static observer), a representation

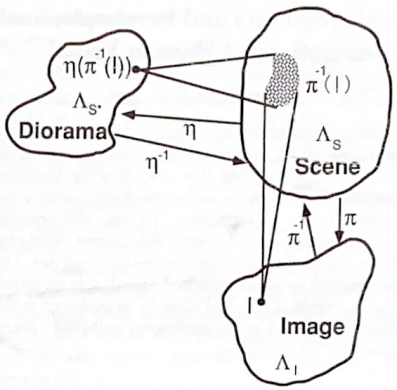

Fig. 7.2 *Schematic of the scene-to-image map, $\pi : \Lambda_S \rightarrow \Lambda_I$. π, being a projective map, is many-to-one, and its inverse, one-to-many, as shown here.*

in terms of a dynamic two-dimensional image is often more convenient. For the development of the ideal observer, we will find it convenient to represent visual information as an image, where we use the term image to refer generically to the input to the visual system, be it a static monocular image, a binocular pair of static images or a sequence of images over time.

Each image results from the application of a projective map, π, to a scene, $I = \pi(S)$. It is, therefore, an element of a stochastic ensemble, Λ_I, with an associated probability density function, $p_I(I)$ given by

$$p_I(I) = \int_{\pi^{-1}(I)} p_S(S)\,dS \tag{7.2}$$

where $\pi^{-1}(S)$ is the subset of Λ_S whose elements map under π to I; that is, the set of scenes which could have given rise to an image under the assumed projective map (see Fig. 7.2).

We define an ideal observer for a given environment to an observer which selects as its estimate of the diorama, the element S^*, $S^* \varepsilon \Lambda_{S^*}$, which maximizes the conditional probability $p(S^*/I)$ (from this point on in the text, we will not specify the subscript for probability density functions where they are implied by the context). An expansion of the conditional probability density function using Bayes' rule will help to elucidate the different components of the ideal observer formulation. The conditional probability is given by

$$p(S^*/I) = \frac{p(I/S^*)p(S^*)}{p(I)} \tag{7.3}$$

or, equivalently,

$$p(S^*/I) = \frac{\int \eta^{-1}(S^*)\, p(I/S)p(S)\,dS}{p(I)} \tag{7.4}$$

This definition of the ideal observer corresponds to what is called in statistical estimation theory the maximum a-posteriori (MAP) estimator, or, in the case that the diorama S^* is categorical, a Bayesian classifier (Duda and Hart, 1973). The density function $p(S^*/I)$ is referred to as the posterior conditional probability density function.

We will find it convenient in the discussion that follows to refer to Equation 7.4. The first term in the numerator, $p(I/S)$, is the conditional probability of obtaining an image I from a scene, S. For the case so far described, in which the image information available to the observer is completely reliable (i.e. it has not been corrupted by noise), $p(I/S)$ serves the function of selecting those scenes which could have given rise to an image. More formally,

$$p(I/S) = \begin{cases} 1 \text{ if } S \varepsilon \pi^{-1}(I) \\ 0 \text{ otherwise} \end{cases} \tag{7.5}$$

The second term in the numerator, $p(S)$, is the prior probability of a scene occurring in the observer's environment. As discussed above, $p(S)$ embodies the ecological constraints on the observer's environment. The denominator, $p(I)$, is the prior probability of obtaining the image I, and is a constant which acts to normalize $p(S^*/I)$.

For specific problems in visual perception, one may find that a higher-level representation of visual information is convenient. We will refer to such a representation as the sketch I^*. Formally, the sketch would be given by a map χ, $I^* = \chi(I) = \chi(\pi(S))$ (see Fig. 7.3). The sketch I^* may or may not represent all the information available in the image I. An example would be the optic flow of retinal intensities as a consequence of motion. Another example of a sketch is a discretely sampled, blurred image, such as would be input to any robotic or human visual system. I^* is an element of a stochastic ensemble Λ_{I^*} with an associated probability density function $P(I^*)$ given by

$$p(I^*) = \int_{\chi^{-1}(I^*)} \int_{\pi^{-1}(I)} p_S(S)\,dS\,dI \tag{7.6}$$

An ideal observer having access to a sketch I^* would select an estimate of the diorama S^* so as to maximize the posterior density, $p(S^*/I^*)$. $p(S^*/I^*)$ can be expanded using Equations 7.3 and 7.4, replacing I with I^*.

Noise

How does the formulation of the ideal observer change when we consider the information available to it to be corrupted by noise. For simplicity, let us assume that the

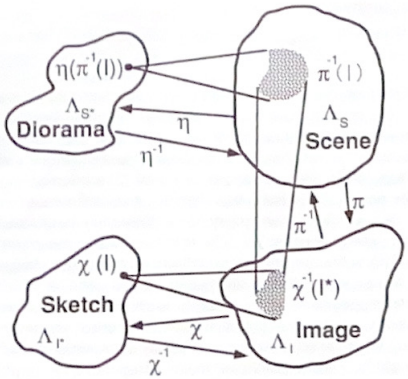

Fig. 7.3 *Schematic including the image-to-sketch map, χ: $\Lambda_I \rightarrow \Lambda_{I^*}$. χ is often many-to-one, and its inverse, one-to-many, as shown here.*

noise, $N(\lambda, \phi, \theta, t)$ is combined with $\pi(S)$ through an invertible operator $\oplus$, with inverse $\ominus$ (note that the noise is defined in the same representational space as the image). The image is given by

$$I = \pi(S) \oplus N \tag{7.7}$$

The density function $p(I/S)$ can be simply related to the noise process by

$$p(I/S) = p(N = I \ominus \pi(S)) \tag{7.8}$$

The posterior conditional density $p(S^*/I)$ becomes

$$P(S^*/I) = \frac{\int_{S \in \eta^{-1}(S^*)} p(N = I \ominus \pi(S)) p(S) dS}{p(I)} \tag{7.9}$$

$p(I)$ may be expanded to

$$p(I) = \int_{\Lambda S} p(N = I \ominus \pi(S)) p(S) dS \tag{7.10}$$

As before, $p(S^*/I)$ can be easily modified to accommodate the use of a sketch I^*.

The essential change in the formulation is that $p(I/S)$ is now spread out over the scene space Λ_S, and is no longer supported only on the subset $\pi^{-1}(I)$. It still maintains some of its characteristics as a selection function, as $p(I/S)$ will generally be concentrated around $\pi^{-1}(I)$. As the level of noise increases, however, the selectivity of $p(I/S)$ decreases, and the influence of the prior density function $p(S)$ on $p(S/I)$ increases.

Ideal Observers and Psychophysical Investigations of Human Visual Perception

A number of researchers have suggested the use of a Bayesian framework for problems in computer vision (Marroquin, 1985; Szeliski, 1990). The analysis of regularization tehniques presented in the last section of the chapter is one example of such an application. We may also organize empirical psychophysical research into human visual perception around the central concept of ideal observers. The main research issues are organized according to the different components of the ideal observer as expressed in the Bayesian expansion of the conditional probability distribution $p(S^*/I)$. These are:

1. The environment, specified as a stochastic scene ensemble, Λ_S, with an associated probability density function, $p(S)$. $p(S)$ embodies the ecological constraints on the structure of the environment.
2. The functional goals of the observer. We have represented this as a scene-to-diorama map, η, applied to a complete scene description to obtain the scene attributes of interest to the observer, the diorama.
3. The image formation process, in the form of a projective map π from scenes in the environment to images. Noise may also be incorporated into the image formation process.

The image formation process determines the conditional distribution $p(I/S)$, which serves as a selection function to select that subset of Λ_S which could have given rise to an image. We may also include the specification of a sketch, given by a map, χ, applied to the raw image. Use of a sketch is usually a matter of analytical convenience, but may correspond to an early neural processing of the visual image.

Each of these components isolates for study some broad aspect of the problem of visual perception. The following sections outline the research questions related to each of the three components. The notion of ideal observers suggests two perspectives from which questions may be asked. One perspective is to first attempt to understand ideal observers for humans' normal living environment, and then to compare human perception to that of the ideal observer. The second perspective is to consider humans as ideal observers in some environment, and to ask questions about that environment. In either case, the Bayesian framework leads to an empirical focus on the constraints on scene interpretation, imposed either by environmental structure or by the image.

The Environment

The two obvious questions to ask about the environment and its relation with human visual perception are: 'what is $p(S)$ for the environment in which humans live' and 'what is $p(S)$ for the environment for which humans are ideal observers?' Of course, one cannot hope to completely specify the appropriate density functions, nor should that necessarily be our goal. Even if one were able to write out equations for the density functions, they might provide no more insight into the structure of the environment than the defining differential equations provide into the behaviour of a non-linear dynamical system. What can be done is to characterize different aspects of the structure in the environment, much as stochastic processes are characterized by their means, their correlation structure and so forth. We will, therefore, rephrase the two questions given above as follows.

What are the ecological constraints on the structure of scenes in the normal living environment of humans?
Ecological constraints are of two types; nomothetic and statistical. Nomothetic constraints are those which strictly apply to each individual scene in an environment. Within a probabilistic framework, statistical constraints reflect tendencies in the environment whereas nomothetic constraints are laws that hold with probability one. As an example of a nomothetic constraint, consider a representation of differentiable surfaces in terms of local surface orientation. The local orientations must satisfy the condition that they be integrable; that is, integration of orientations along a closed path on the surface must result in no change in depth. Another example of a nomothetic constraint is the fact that the sum of a surface's reflectance and transmittance coefficients is always less than or equal to 1. Examples of possible statistical constraints are that changes in surface reflectance tend to be discontinuous (surface reflectance is piece-wise constant) and that surfaces are smooth, or piece-wise smooth. The last two constraints are commonly used in computational models of reflectance and shape estimation (Land and McCann, 1971; Horn, 1974), though their ecological validity has not been rigorously tested.

What are the constraints on scene attributes which are incorporated into human visual system processing?
These constraints may also be one of two types, nomothetic or statistical. The constraints may or may not match the ecological constraints of the physical environment. Consider the image of three cylinders viewed through an aperture shown in Fig. 7.4. The cylinders appears to be painted with stripes of different colour, and appear to be oriented in different direction relative to the viewer. The orientations of the cylinders are formally ambiguous. The strength of the percept, therefore, indicates that the visual system assumes some constraint on the relationship between the contours formed by the reflectance edges and the shape of the cylinders which serves to disambiguate the orientations of the cylinders. A specific proposal for this constraint has been discussed, suggesting that it reflects a perceptual bias toward an interpretation of reflectance edges as geodesics of a surface (Knill, 1991).

Consideration of the above two questions leads to another question.

Is any given ecological constraint on scene attributes incorporated into visual system processing, or is any given constraint which appears to be enforced by the visual system ecologically valid?
There are, of course, just two ways of asking the same question, and how it is asked in any particular case is a matter of methodological convenience.

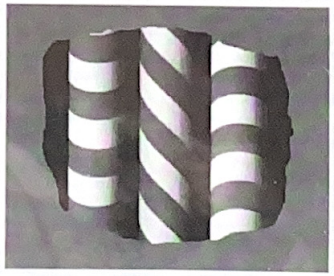

Fig. 7.4 *Three cylinders painted with barber pole patterns, each of which appears at different orientations from the line of sight.*

Psychological research into the environmental aspect of perception should be focussed on the second and third questions, as the first question is properly the domain of the physical sciences. A teleological argument, however, suggests that this research should always be tempered by consideration of the ecological constraints. One might expect, after all, that the human visual system had adapted to bring the phenomenal perceptual world into as close a correspondence with the real world as possible given physiological limitations. Moreover, the apparent accuracy of our percepts argues for the ecological validity of the constraints imposed by the visual system.

The Functional Goals of Human Observers

Ultimately, the functional goal of visual perception is to guide behaviour in the naturel environment; either immediate or future. Consideration of the perception-action loop as a whole is sometimes possible, and appropriate. In general, however, the complexity of the intervening processes makes it necessary to consider perception as a separate process and, furthermore, to isolate different aspects of perception for study. The different aspects may also sometimes be considered to reflect different levels of processing in the visual system (e.g. image coding and transduction *vs* object recognition). The functional goals which may be associated with the different aspects of perception are often assumed in studies of perception; however, what they are is itself a question open to empirical investigation. The ideal observer framework is well-suited to dealing with this question.

Within the ideal observer framework, functional goals are identified with those characteristics of the environment which the visual system extracts from images. These characteristics may be implicit in a perception-action loop; that is, they are those characteristics of the environment required for the control of a certain action. They may also be those characteristics which are made explicit in the system for further cognitive processing or for the formation of memories. The appropriate question to ask concerning the functional goals of visual perception is:

What scene attributes does the human visual system extract from images?

In the formal specification of the ideal observer, this corresponds to inquiring into the form of the scene-to-diorama map, η, which maps S to S^*. Intuitively, the way to approach the problem empirically is to design experimental tasks for which a subject requires information about different scene attributes, and to search for those tasks for which subjects' performance most closely approaches that of the ideal observer for the task. One can argue that the scene attributes required for performance of these tasks are the ones which the visual system is designed to extract from images. Some studies of image coding and transduction properties of the visual system and of simple image domain tasks have successfully made use of this approach (Barlow, 1978; Kersten, 1984; Geisler, 1989); however, it has yet to be applied to the so-called 'higher level' problems of scene production.

The Image Formation Process

As with questions about the relationship between environmental structure and perception, questions about the image formation process may be asked in one of two ways; from the point of view of the ideal observer and from the point of view of the human observer. Thus we can either ask questions about the projective map, π, and the image noise, which together make up the image formation process, or we can ask questions directly about the selection function $p(I/S)$. The general questions are as follows.

What is the nature of the projective map and the noise, if any, which corrupts the image?

A complete characterization of the projective map involves both the mathematics of perspective projection and the physics of light reflection and refraction by objects, liquids and gases. This is well enough understood for the generation of realistic images through computer graphics; in which the limitations to realism are primarily found in algorithmic complexity and the accurate modelling of scenes. Noise must be considered only if one takes the image to be represented at some stage of processing in a real visual sensing system, such as the light captured by retinal photoreceptors, or the outputs of the photoreceptors.

What is the nature of the sketch I^*?

One research goal is to seek out correspondences between known neural mechanisms and sketches, I^*. As an example, a sketch of band-pass filtered images bears a strong resemblance to output of retinal ganglion cells. At the functional level, sketches must be found which make explicit the information required to estimate specific scene characteristics. An example would be measurements of dilatation and rotation components of the optic flow field as information for estimating direction of heading.

How does the ideal observer's selection function, $p(I/S)$, constrain the scene which could have given rise to an image?

These constraints are derived from the characteristics of the projective map and the noise. If we take as the starting point for our investigation an idealized image, uncorrupted by noise, we can rephrase the question as, 'what are the attributes of the scenes contained in $\pi^{-1}(I)$?' A good example of research into this question for natural scenes is the work of Koenderink and van Doorn on invariant relations between surface shape and the projected image (Koenderink and van Doorn, 1976, 1984; Koenderink, 1984).

How does the human observer's selection function, p(I/S), constrain the scene which could have given rise to an image?

Does the human visual system enforce the constraints imposed by the ideal observer's selection function? How similar are the attributes of scenes perceived by humans to the same attributes of the scenes in $\pi^{-1}(I)$. Can the differences be reasonably explained by postulating the injection of noise somewhere in the processing stream, or are they due to improper adaptation to the environment? These are some of the detailed questions which one may pose about the observer's selection function.

Further Questions

The questions listed above lead to further questions which bridge some of the different aspects of perception we have outlined.

How are the different constraints weighted?

An important aspect of this problem is the question of how the constraints imposed by the image are weighted relative to ecological constraints built into visual system processing. As illustrated by the example of regularization techniques given later in the chapter, the relative weighting for the ideal observer is determined by the form of the image noise. In the case of a noise-free image, constraints on perceived scene attributes imposed by the image are 'hard constraints' which should never be violated. Do the relative weights apparent in the human visual system reflect simply the effects of noise or are they examples of misadaptation to the environment?

What constraints are learned and how are they learned?

Part of this problems is also the question of what weights are learned and how they are learned.

What algorithms and mechanisms are used in the visual system for the implementation of the different constraints?

Comparison with Other Approaches

The statement that perception is a process of statistical inference has the familiar ring of Helmholtz's theory of unconscious inference. A rough summary of Helmholtz's idea is that an observer perceives in an image those scene attributes which would have normally given rise to the image. The observer unconsciously applies knowledge of the structure of the environment, gained through associational learning, to disambiguate an image (Helmholtz, 1925). The theory was developed more thoroughly by the transactionalists (Ittleson, 1960) and is present in contemporary information processing approaches to perception (Gregory, 1973; Rock, 1977).

A number of the concepts of the transactionalist approach map onto the probabilistic framework presented here. The notion of equivalent scene configurations for an image corresponds to the inverse projective map, π^{-1}, and the assumptions about environmental structure proposed to be used by the visual system to disambiguate images correspond to characteristics of the visual system's prior distribution, $p(S)$. The focus of the transactionalist, and later information processing approaches, however, has been on the study of depth and shape cues used by the visual system to 'fill in' the depth dimension. The traditional cues studied include retinal accommodation, size, interposition, linear perspective, aerial perspective, binocular disparity, convergence and motion parallax. Cues which have begun to receive significant attention are texture gradients, shading and contour form. One problem with the cue concept is the equivocality of its definition. As Ittleson says about the definition of cues,

> *The most obvious drawback to the descriptions is their lack of consistency. Some cues are described primarily in terms of the attributes of the physical object, some in terms of the light energy, some with reference to physiological excitation and some entirely in terms of psychological factors. This heterogeneity is not accidental. It reflects a basic property of the cue concept. A cue is not something that can be pointed to; rather it represents a complex interrelationship between a number of aspects that must be taken into account in the definition of the cue (Ittleson, 1960.)*

Many of the image cues (as opposed to physiological cues like accommodation and convergence) confound constraints on the image mapping of scene attributes with ecological constraints on the structure of these characteristics. As an example of this, consider the images shown in Fig. 7.5. The converging contours and the brightness differences in the perspective painting of a cubic block shown in Fig. 7.5(a) provide a cue to the 3-D shape of the block only in so much as the contours have been labelled as corners of a polyhedral object and the shading has been attributed to differences in orientation of the sides of the object. The pattern of convergence of the contours, however, also contributes to the labelling itself of the contours as polyhedral corners. A similar image shown in Fig. 7.5(b) appears as a flat surface and the contours as discontinuities in surface reflectance. How then can the contours be said to be a cue for shape? One might argue that the contour pattern provides a cue to the shape of a surface in the sense that it constrains the percept to be one of a cube. This is

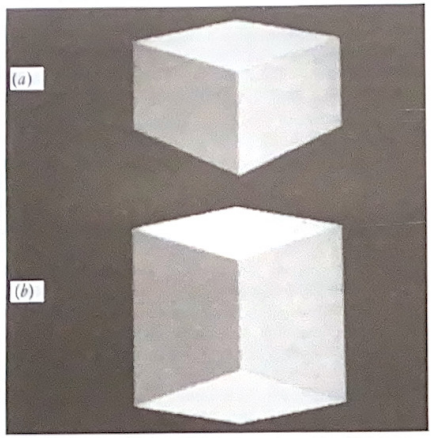

Fig. 7.5 *The image in (a) appears as a perspective view of a shaded block, while the image in (b) appears as a multi-coloured flat surface.*

undoubtedly true, but is no more than a restatement of the perceptual phenomenon. A deeper understanding of how and why the contours constrains percepts in this way requires an investigation into the different constraints which go into making up the cue.

Perhaps the major problem with the notion of depth or shape cues, is that it has led to the implicit modularization of visual perception in empirical research. Modularization appears in two forms; first, in the attempt to isolate individual cues for study, and secondly, in the inherent treatment of depth and shape perception as the result of separate processes from those involved in the perception of other scene attributes such as surface reflectance and illumination. The modularization of visual information processing, used by the transactionalists as a necessary simplying assumption for empirical research, was raised to the level of a principle in the computational approach proposed by Marr (1980):

> *. . . the idea that a large computation can be split up and implemented as a collection of parts that are as nearly independent of one another as the overall task allows, is so important that I was moved to elevate it to a principle, the principle of modular design . . . Information about the geometry and reflectance of visible surfaces is encoded in the image in various ways and can be decoded by processes that are almost independent.*

Interestingly, studies (Gilchrist, 1980; Kersten, 1990) have shown that the visual system does compute some information about surface geometry such as orientation and curvature, cooperatively with surface reflectance. For the purpose of studying human visual perception, untested assumptions of modularity unnecessarily limit the scope of inquiry. A study of constraints does not rely on assumptions of modularity or non-modularity. Modularity is left as a hypothesis, open to empirical investigation, about the nature of the processing mechanism which incorporates the constraints in the visual system.

While the notion of probabilistic inference is inherent to an information processing view of perception, the focus on constraints and the view of perception as constraint satisfaction is reminiscent of Gestalt theories. The central concept in Gestalt theories of perception is that of Praegnanz; that 'psychological organization will be as good as the prevailing conditions allow' (Koffka, 1935). The central problem for Gestaltists was the elucidation of the organizing principles which govern the formation of percepts. Toward this end, Koffka identified 'goodness' with minimum-maximum properties; such as maximum symmetry, minimum variation of form, and so on. This fundamental concept has led to a number of general Praegnanz principles, such as energy minimization in 'soap bubble' systems (Attneave, 1982), minimal coding of scenes (Leeuwenburg, 1971), and minimization of 'changes' in scene attributes (Hochberg and McAlister, 1953). The last two are examples of so-called principles of efficiency. As early as the late nineteenth century, Mach pointed out the relationship between an efficiency principle and a probability principle (Mach, 1980). A similar relationship holds for minimum energy principles. In either case, the organizing principles can be re-posed as characteristics of the probability distribution assumed by observers for scene attributes in the environment.

Gibson criticizes Gestalt theory for being focussed on the constructive role played by the observer in perception at the cost of disregarding the relationship between the observer and the environment (Gibson, 1982). Attneave and Frost (1969) make this perspective clear, saying, 'A Praegnanz principle assume a teleological system (as Koffka, 1935, explicitly recognized) in which simplicity has the status of a final cause, or goal-state.'

No reference is made to functional goals of a behaving organism. Consideration of Praegnanz principles as characterizing an observer's model of environmental structure is the necessary step in bridging the gap between the observer and the environment. In this light, the Gestalt programme of research deals with only half the problem of perception, as it is also necessary to characterize the properties of ideal observers. The two aspects of the problem should be studied jointly, as understanding of one can guide research into the other.

The claim that visual perception should be understood as a relationship between human observers and ideal observers is equivalent to the point by Gibson and Brunswik that perception should be considered in its relationship to the normal environment of an organism. Gibson further claims that the information for the perception of functionally important attributes of scenes is unambiguously represented in the image (or what he referred to as the 'ambient optic array'), and moreover, that this information is picked up 'directly' by an observer (Gibson, 1979). As mentioned previously, the first claim amounts to a claim that for the appropriate S^*, the conditional distribution, $p(S^*/I)$, is a Dirac delta function supported at one point in the ensemble, Λ_{S^*}. The second claim, developed further, states that one can find image properties which are invariantly related to each individual scene characteristic in S^*. This claim brings to mind the modularity of processing referred to earlier as a characteristic assumption of cue processing approaches. Implicit in the claim is the assumption that each of the attributes in S^* can be independently detected in an image.

Within the probabilistic framework, the two aspects of Gibson's theory can be formulated as constraints on the ideal observer's posterior conditional distribution, $p(S^*/I)$. The first is that with the appropriate selection of scene attributes, the posterior distribution can be expanded into the product of independent distributions of the individual scene attributes conditional on the image; thus, letting $S^* = (s_1, s_2, s_3, \ldots, s_n)$, we have

$$p(S^*/I) = p(s_1/I)p(s_2/I)p(s_3/I)\ldots p(s_n/I) \qquad (7.11)$$

A search for scene attributes which depend independently on the image is important; however, the independence criterion is not enough to determine the functional relevance of a given scene characteristic for human perception. Furthermore, many cases may be found in which the human visual system apparently does not generate percepts of different scene attributes independently (Hochberg, 1974; Epstein, 1977), including the previously mentioned example of cooperativity in the perception of lightness and spatial layout.

(The counter-argument to these examples is generally that researchers have simply considered inappropriate characterizations of scenes, and an appropriate redefinition of the scene attributes of interest will lead to ones which are independently specified in the image. An example of this is the postulate of 'shape-at-a-slant' as the appropriate psychological variable for the perception of skewed figures (Beck and Gibson, 1955). Whether this is, in fact, the functionally appropriate variable should be opened to empirical investigation, and not simply assumed based on the argument that it matches the independent criterion for the types of stimuli considered. As we have pointed out, the ideal observer construct provides an ideal mechanism for testing these sorts of assumptions.)

The second aspect of Gibson's theory, that appropriately defined scene attributes are unambiguously represented in the image implies that each $p(s_j/I)$ in Equation 7.11 is a Dirac delta function. This condition clearly does not hold for all perceived scene attributes in all viewing conditions. All other considerations aside, physiological limitations on the pick-up of information from images, such as image blurring outside the foveal region and the addition of noise in the nervous system, preclude this possibility for human observers (Hochberg, 1982). These limitations increase the importance of statistical constraints on environmental structure in perception. Consideration of statistical ecological constraints was central to Brunswik's version of 'ecological' psychology. Realizing the importance of these constraints led Brunswik to develop an empirical framework based on correlation studies of scenes, image cues and percepts, what he called representative functionalism (Brunswik, 1956). He considered the environment, however, to be too complex to allow the more in-depth analysis of its relation to perception proposed by Gibson. The probabilistic framework presented here is really a generalization of Gibson's analytical approach to consideration of a stochastically defined environment as envisaged by Brunswik.

The probabilistic framework is most closely related to the natural computation approach of Whitman Richards. In fact, Richards explicitly recognizes the link between his approach and a probabilistic view of perception (Richards, 1988). The one difference in the approach presented here is in the emphasis on the ideal observer formulation as a tool for empirical research, particularly, in its potential use for testing hypotheses about the functional goals of human observers. Otherwise, the research questions presented here as derivative of a probabilistic view of perception are the same as those considered by Richards to be of basic importance.

Some may argue that a probabilistic framework is too general to be of practical use for the study of perception. Not only does it have little predictive power as concerns the mechanism of perception, but is it general enough to accommodate as special cases theoretical approaches as divergent as information processing psychology, Gestalt psychology and ecological psychology. It does, however, make explicit as the fundamental objects of research those elements of the problem of perception which are common to each of these approaches; namely the constraints on the interpretation of the physical causes of images in the environment. These constraints are at the root of the structure of traditional image cues, Gestalt rules of organization and Gibson's invariants (Flock, 1964). The probabilistic framework, therefore, provides an organizational structure in which one can consistently consider concepts often supposed to be diametrically opposed to one another.

Application of Ideal Observer Analysis to Neural Networks and Natural Computation

A few concrete applications will elucidate the usefulness of ideal observer analysis. In the first application we will show how the ideal observer relates to specific mechanisms of neural networks. In the second application we will develop the ideal observer for a world consisting only of thin wires. Although only a 'toy' world, this analysis quantifies and makes explicit the nature of the line projection assumptions used in some models of visual recognition (Biederman, 1987).

Probability to Neural Networks

One early connection between ideal observers and processing by single neurones was suggested by the high performance of human observers relative to ideal for the contrast detection and discrimination of Gaussian windowed sinusoidal luminance patches in visual noise (Burgess *et al.*, 1981; Kersten, 1984). High performance and the way in which human detection efficiency varied with width suggested that cortical simple cell processing might account for the data. Ideal performance is calculated by maximizing the likelihood of a known signal against noise and can be achieved by a linear cross-correlation of the input signal with the known signal. Highest detection performance is expected when the visual image most closely matches the receptive field of the neurone.

Image understanding problems require going beyond simple detection to the estimation of a typically large set of parameters. As we have seen earlier, this requires a more sophisticated modelling of prior constraints than for detection. We would expect that estimation mechanisms involve computation by a large collection of neurones. Most current work is only suggestive of what real neural networks might achieve, but serves to illustrate how one can establish a bridge between the abstract formalism of Bayesian ideal observers and neural mechanism. One computational approach which has been mapped to neural networks is regularization theory (Poggio *et al.*, 1985). By reformulating standard regularization theory within a probabilistic framework, one can show the relationship between linear neural network models and ideal observers.

A number of computational models for problems as diverse as structure from motion and shape from shading may be reformulated under the umbrella framework of regularization theory. The basic method in this approach is to select as the interpretation of the scene which projected to a given image that scene which minimizes some error functional. The error functional consists of an image error term, which reflects how well a scene matches a given image, and a penalty term which incorporates prior constraints (typically in the form of smoothness) on scenes. The models typically use representations of images and scenes which may be expressed as functions of the x and y coordinates of a planar image. For the case of a static, monochromatic image, I would be a function $I(x,y)$ specifying the light intensity at each point in the image. S is generally some vector function specifying local attributes of surfaces in a retino-centric coordinate system, $\mathrm{S}(x,y)$. The total error over a bounded region R in an image may be expressed as

$$E(I,\mathrm{S}) = \iint_R \lambda(I(x,y) - \pi(\mathrm{S}(x,y)))^2 + P(\mathrm{S}(x,y))^2 dxdy \tag{7.12}$$

where $(I(x,y) - \pi(\mathrm{S}(x,y)))^2$ is the image error and $P(\mathrm{S}(x,y))^2$ is the penalty function (often referred to as the regularizing function). The constant λ is a Lagrange multiplier which weights the relative contributions of the image error and penalty terms.

For what environment would such a model be an ideal observer? The relationship between optimal Bayesian estimators and regularization methods has been developed by others (see, e.g. Marroquin, 1985; Kersten *et al.*, 1987; Szeliski, 1987) and is well understood in the computer vision community. Minimizing Equation 7.12 is equivalent to maximizing the probability density function

$$p(\mathrm{S}/I) = k \cdot exp[-\iint_R \lambda(I(x,y) - \pi(\mathrm{S}(x,y)))^2 + P(\mathrm{S}(x,y))^2 + dxdy] \tag{7.13}$$

where k is selected to normalize the distribution. We can rewrite Equation 7.13 as

$$p(\mathrm{S}/I) = k \cdot exp[-\frac{1}{2\sigma^2}\iint_R (I(x,y) - \pi(\mathrm{S}(x,y)))^2 dxdy] \cdot exp[-\iint_R P(\mathrm{S}(x,y))^2 dxdy] \tag{7.14}$$

where we have replaced λ with $1/2\sigma^2$. If we let

$$p(I/\mathrm{S}) = k_1 \cdot exp[-\frac{1}{2\sigma^2}\iint_R (I(x,y) - \pi(\mathrm{S}(x,y)))^2 dxdy] \tag{7.15}$$

$$p(\mathrm{S}) = k_2 \cdot exp[-\iint_R P(\mathrm{S}(x,y))^2 dxdy] \tag{7.16}$$

and

$$p(\mathrm{I}) = \frac{k_1 k_2}{k} \tag{7.17}$$

where k_1 and k_2 are normalizing constants, we can relate Equation 7.14 to the posterior density function for the ideal observer for noisy images given in Equation 7.9, in which $p(I/\mathrm{S})$ is the conditional density function for an image corrupted by additive white Gaussian noise with variance σ. Regularization models which make use of Equation 7.12 are ideal observers in a world in which the statistical structure of the environment can be characterized by a Gibbs distribution as in Equation 7.16, and in which the image available to the observer is corrupted by additive Gaussian noise. In so-called standard regularization where π and P are linear operators, S can be estimated from I by a linear neural network (Poggio *et al.*, 1985). Whether this sort of linear approximation is actually a good model of neural processing is an open question. One possible candidate is the solution of the aperture problem by projections from area V1 to MT of cortex (Sereno *et al.*, 1988).

If the precise form of the prior constraint operator P, or the image formation function π are unknown, it is possible to learn the best linear approximator to estimate scene parameters from image data (Kersten *et al.*, 1987; Knill and Kersten, 1990). This estimator corresponds to the ideal MAP estimator when the image formation function is linear and the statistical distribution of the image noise and scene prior are Gaussian.

Ideal observer analysis is not restricted to linear models of neural networks. One of the major contributions to the theory of neural networks was to show the existence of an 'energy' function for both linear threshold and linear summation followed by sigmoid non-linearity models of neurones (Hopfield, 1984). The energy function depends on the state of the network and its value decreases as the neural elements are updated. It is straightforward to interpret the energy function as the negative exponent of a Gibbs distribution that the network is trying to maximize, and thus as a mechanism for seeking the MAP ideal observer solution for the world the connection strengths represent. Golden has shown that several major theoretical models of neural networks, including back-propagation, can be interpreted as MAP estimation (Golden, 1988).

Wire World: Ideal Observer Analysis and Natural Computation

Contours formed by discontinuities in luminance in a static image are caused by discontinuities in one or another characteristic of a scene. They may be caused by discontinuities in surface reflectance, surface orientation or illumination, or by the self-occlusion of a smooth object relative to the viewpoint of the observer. Such contours provide highly reliable information for the perception of such scene attributes as surface shape and illumination conditions, or for the categorization of objects. A number of properties of the contours are taken to have an invariant relationship to the generating discontinuities in the scene. The invariants are either preserved exactly in the projective map, or are assumed to hold in almost all cases, except for accidents (Waltz, 1974; Biederman, 1987). Examples of the latter are collinearity, parallelism (under orthographic projection) and straightness. These invariants are assumed by many people to depend only on the assumption of a general viewpoint on a scene and not on any statistical structure in the environment. In the example to follow, we will show that one of the assumed invariants, that of straightness, does in fact depend on the environment having a certain statistical structure. In an environment in which the shapes of physical discontinuities are arbitrary, the straightness invariant does not hold. It does not hold even in some environments in which straight discontinuities are the most likely discontinuities to occur. For the invariant to be nomothetic, that is hold with probability one, the discontinuities, in the environment must have some categorical structure, with a non-zero proportion of discontinuities being straight, and the rest being curved, though the actual proportion of discontinuities which are straight does not affect the result, and it could be quite small.

We will analyse the validity of the straightness invariant by developing an ideal observer in a simple toy world consisting only of open thin wires. We will assume for simplicity that the observer only sees one wire at a time and that both the viewpoint of the observer and the wires are static. We will approximate the wires as open curves in a three-dimensional Euclidean space and their projected images as open curves in a two-dimensional Euclidean space. To further simplify the discussion, we will assume

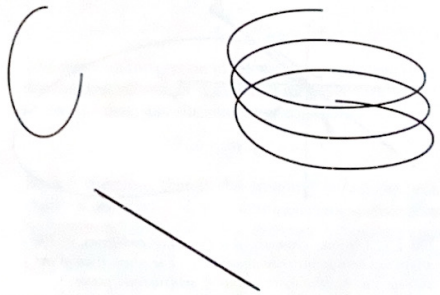

Fig. 7.6 *Examples of the types of wire in the wire world – circular arcs, helical arcs and straight lines. These are the three types of shapes which could be created by physically bending or twisting a thin straight wire at its end-points.*

that the wires have constant curvature and torsion, constraining the wires in the toy world to being either straight segments, circular arcs or helical arcs (see Fig. 7.6). To visualize the curvature and torsion of a wire, imagine laying a straight wire on a flat table and bending it. If the wire is bent without twisting, it will stay flat on the table. If it is twisted when it is bent, however, the wire will pull away from the table. The curvature of the wire specifies the degree to which it is bent at each point, while the torsion specifies the degree to which it is twisted away from being flat. We will assume orthographic projection, though the result generalizes easily to the case of perspective projection.

A scene consists of one wire floating in space, so our model of the scene, S, must specify the spatial configuration of a wire, which, in our toy world, is completely defined by the position of one end-point of the wire, the orientation of the wire at that end-point, its curvature, its torsion and its length. Thus, we have for S

$$S=\{X_w,\ Y_w,\ Z_w,\ \Phi_w,\ \Theta_w,\ \Sigma_w,\ K_w,\ T_w,\ L_w\} \tag{7.18}$$

$(0<X_w<1,\ 0<Y_w<1,\ 0<Z_w<1,\ 0\leq\Phi_w<2\pi,\ 0\leq\Sigma_w\leq\pi,\ \leq\Theta_w<2\pi,\ K_w\geq 0,\ L_w>0)$. The triple, $\{X_w,\ Y_w,\ Z_w\}$ specifies the position of an end-point of the wire. The end-point is arbitrarily chosen. We have constrained the position of the end-point of the wire to a unit cube. Φ_w, Σ_w and Θ_w specify the orientation of the wire at the end-point, with Φ_w being the tilt of the wire, its orientation away from the horizontal in a fronto-parallel plane, Σ_w the slant of the wire, its orientation out of the fronto-parallel plane, and Θ_w the orientation of the normal vector of the wire, relative to the line of sight (Θ_w specifies the direction in which the wire curves at its end-point). Since a straight line has no curvature, its normal vector is indeterminate; therefore, we will assume that $\Theta_w=0$ for straight lines, K_w and T_w specify the curvature and torsion of the wire, and L_w specifies the wire's length (in the case of a planar wire, for which some of the wire might be occluded from the observer, L_w represents the length of the portion of the wire which is visible). Fig. 7.7 summarizes the geometry of the scene. From Equation 7.18, we see that $\Lambda_S\subset\mathscr{R}^9$, where Λ_S is the stochastic ensemble from which particular instances of S are drawn.

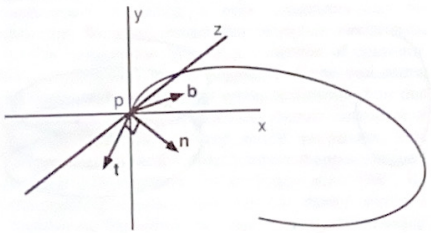

Fig. 7.7 *The local geometry of a wire at its end-point,* p. *The origin has been arbitrarily placed at* p. *The orientation of the wire at* p *is specified by the triple of orthonormal vectors,* $\{t,n,b\}$. t *is the tangent vector of the wire,* n *is its normal vector, pointing in the direction that the wire bends, and* b *is the binormal vector of the wire, pointing parallel to the direction that the wire twists out of the plane.*

The image of a wire can be characterized as the shape of the contour to which the wire projects. We will represent this using the position and orientation of one end of the contour and the curvature of the contour as a function of arc-length. Thus, we have for I

$$I=\{X_i,\ Y_i,\ \Phi_i,\ K_i(s),\ L_i\} \tag{7.19}$$

$(0<X_i<1,0<Y_i<1,0\leq\Phi_i<2\pi,K_i(s)\geq 0,L_i>0)$ where the pair, $\{X_i,\ Y_i\}$ specifies the position of the end-point of the contour, Φ_i specifies the orientation of the contour at its end-point, $K_i(s)$ specifies the curvature of the contour as a function of arc-length, and L_i is the length of the contour. Note that $K_i(s)$ is a function, whereas K_w is a scalar variable.

We are now ready to formulate a statement of the straightness invariant in a probabilistic framework. Let us define a diorama to characterize the straightness of a wire,

$$S^*=\begin{cases}\text{straight; if } K_w=0 \text{ and } T_w=0\\ \text{curved; otherwise}\end{cases} \tag{7.20}$$

An image with a straight contour is given by

$$\hat{I}=\{X_i=x_i,\ Y_i=y_i,\ \Phi_i=\phi_i,\ K_i(s)=0;\ \forall s,\ L_i=l_i\} \tag{7.21}$$

where x_i, y_i is the endpoint of the contour, ϕ_i is its orientation and l_i is its length. With these definitions, the straightness invariant may be stated formally as

$$p(S^*=\text{straight}/I=\hat{I})=1 \tag{7.22}$$

that is, given an image of a straight line, the probability that the wire in the scene is straight is one.

We will expand the invariant Equation 7.22 using Baye's rule to analyse the implications it has for the structure of the wire world environment. We have

$$p(S^*=\text{straight}/I=\hat{I})=\frac{\int_{S\in\eta^{-1}(\text{straight})}p(I=\hat{I}/S)p(S)dS}{p(I=\hat{I})}=\frac{\int_{S\in\eta^{-1}(\text{straight})}p(I=\hat{I}/S)p(S)dS}{\int_{S\in\Lambda_S}p(I=\hat{I}/S)p(S)dS} \tag{7.23}$$

where η, in this case, is a projective map which takes S to S^*. The set, $\eta^{-1}(\text{straight})\subset\Lambda_S$, is a seven-dimensional slice through Λ_S along the hyperplane defined by $K_w=0$, $T_w=0$. The density function $p(I=\hat{I}/S)$ has the following characteristics. The position of the end-point of the contour partially determines the position of the endpoint of the wire ($X_w=x_i,\ Y_w=y_i$, under orthographic projection), the orientation of the contour fully determines the tilt of the wire at its endpoint ($\Phi_w=\phi_i$), the straightness of the contour determines the orientation of the wire's normal vector ($\Theta_w=0$) and the torsion of the wire ($T_w=0$), and the length of the contour determines the length of the wire as a function of the slant and curvature of the wire ($L_w=f(\Sigma_w,K_w,l_i)$), which, following a simple derivation, is given by

$$L_w=f(\Sigma_w,K_w,l_i)=\begin{cases}\dfrac{1}{K_w}[cos^{-1}(cos\Sigma_w-K_wl_i)-\Sigma_w] & ;\text{ if } K_w>0\\ l_i/sin\Sigma_w & ;\text{ if } K_w=0\end{cases} \tag{7.24}$$

From these facts, we can derive an explicit expression for $p(I=\hat{I}/S)$, given by

$$p(I=\hat{I}/S)=\begin{cases}1\text{; if } X_w=x_i,\ Y_w=y_i,\ \Phi_w=\phi_i,\\ T_w=0,\ \Theta_w=0,\ K_wl_i-cos\Sigma_w<1 \text{ and}\\ L_w=f(\Sigma_w,K_w,l_i)\\ 0\text{; otherwise}\end{cases} \tag{7.25}$$

The constraint that $K_wl_i-cos\,\Sigma_w<1$, *for* $p(I=\hat{I}/S)=1$, is derived from Equation 7.24 for the length of the wire. An intuitive explanation of this constraint is that a circular arc must have a radius of curvature $(1/K_w)$ greater than a certain limiting size in order to project to a contour of a given length.

We will use these facts to show that if $p(S)$ is a smooth density function (e.g. a uniform distribution over the variables in S), then $p(S^*=\text{straight}/I=\hat{I})=0$. The result will follow from the fact that $\hat{I}$ constrains the possible values of S to a three-dimensional volume in Λ_S, whereas the subspace of Λ_S for which $S^*=$ straight is a two-dimensional surface in this volume. If $p(S)$ is smooth, then the probability of any given surface in Λ_S occurring in a given volume is zero. To understand this result, consider by analogy the case of a real scalar random variable, X, with smooth probability density function, $p(X)$. The probability that X takes on any given value, $X=x$, is zero. We can view this as the probability of a zero-dimensional subspace of a one-dimensional stochastic ensemble, $\Lambda_X\subset\mathscr{R}^1$. In order for the probability of a given value of X, $X=x$, to be greater than zero, the probability density function must include a Dirac delta function defined at the point, x, in Λ_X. The Dirac delta function serves the purpose of concentrating a non-zero proportion of the probability density at the point, x. Similarly, any m-dimensional subspace of an n-dimensional stochastic ensemble, $\Lambda_X\subset\mathscr{R}^n$, where $m<n$, has probability zero, if the probability density function, $p(X)$, is smooth. Again, a Dirac delta function must be included in the definition of $p(X)$ to concentrate a non-zero proportion of the probability density in the given m-dimensional subspace of Λ_X.

Returning to the problem at hand, we can rewrite the numerator of Equation 7.24 using Equation 7.25 as

$$\int_{S\in\eta^{-1}(\text{straight})}p(I=\hat{I}/S)dS=\iint_{R_n}p(X_w=x_i,\ Y_w=y_i,\ Z_w=z_w,\ \Phi_w=\phi_i,\ \Sigma_w=\sigma_w,\ \Theta_w=0,\ K_w=0,\ T_w=0,\ L_w=f(\Sigma_w,K_w,l_i))dz_wd\sigma_w \tag{7.26}$$

where R_n is the region $\{0<z_w<1,0\leq\sigma_w\leq\pi\}$. We can rewrite the deonimator using a similar expansion, to obtain

$$\int_{S\in\Lambda_S}p(I=\hat{I}/S)dS=\iiint_{R_d}p(X_w=x_i,\ Y_w=y_i,\ Z_w=z_w,\ \Phi_w=\phi_i,\ \Sigma_w=\sigma_w,\ \Theta_w=0,\ K_w=k_w,\ T_w=0,\ L_w=f(\Sigma_w,K_w,l_i))dz_wd\sigma_wdK_w \tag{7.27}$$

where R_d is the region, $\{0<z_w<1,0\leq\sigma_w\leq\pi,\ k_w\geq 0,\ \kappa_wl_i-cos\ \sigma_w<1\}$. The region of integration for the numerator is the two-dimensional surface in Λ_S defined by seven equations in nine unknowns,

$$\begin{aligned}K_w&=0\\ T_w&=0\\ X_w&=x_i\\ Y_w&=y_i\\ \Phi_w&=\phi_i\\ \Theta_w&=0\\ L_w&=f(\Sigma_w,\kappa_w,T_w)\end{aligned} \tag{7.28}$$

The region of integration in the denominator is the three-dimensional volume in Λ_S defined by the last six of these seven equations and the constraint equation

$$K_wl_i-cos\ \Sigma_w<1 \tag{7.29}$$

We see, therefore, that $p(S^*=\text{straight}/I=\hat{I})$ is the probability of occurrence of the two-dimensional surface given by Equation 7.28 in the three-dimensional volume given by Equation 7.29; therefore, If $p(S)$ is smooth over Λ_S, we have $p(S^*=\text{straight}/I=\hat{I})=0$. Even if $p(S)$ were defined so that the $K_w=0$, $T_w=0$ was a mode of the distribution, the result would hold, as it only depends on the condition that $p(S)$ be smooth. In order for the straightness invariant to hold, the definition of $p(S)$ must include a Dirac delta

function at $K_w = 0$, $T_w = 0$. An example of an appropriate density function for S is

$$p(S) = (1-q)p_s(S) + q\,k\delta(K_w^2 + T_w^2) \qquad (7.30)$$

where $p_s(S)$ is a smooth density function over S, q is the proportion of straight wires in the environment, and k is a uniform probability density spread over the six-dimensional subspace of Λ_S defined by $K_w = 0$, $T_w = 0$. $\delta()$ is the Dirac delta function. The straightness invariant will hold in an environment with a structure characterized by the modified $p(S)$, regardless of the value of q.

If we were to generalize the wire world model to allow wires with non-constant curvature and torsion, we would obtain the same result. The inclusion of other wires would lead to an increase in the dimensionality of Λ_S, but this increase would be entirely contained in an increase of the dimensionality of the subspace of curved wires, leaving the dimensionality of the subspace of straight wires constant (the constant curvature and torsion world already contains all the possible straight wires). In the formulation given above, this would lead to an increase in the dimensionality of the subspace over which the denominator in Equation 7.24 is integrated, while the dimensionality of the subspace over which the numerator is integrated would remain two. Thus $p(S^* = \text{straight}/I = \hat{I})$ would be the probability of occurrence of a two-dimensional surface in an n-dimensional volume, where $n > 3$. As before, a Dirac delta function which concentrated a proportion of $p(S)$ in the subspace of straight wires would be necessary to obtain the straightness invariant.

Acknowledgements

This work was supported by NSF grant BNS-8708532 to Daniel Kersten, and by NSF grant BNS-8518675 to James A. Anderson. We would like to thank William Warren and James Anderson for useful comments on this work.

References

Adelson, E. H. and Bergen, J. R. (1990). The holoscopic function and the elements of early vision. In *Computational Models of Early Vision*. eds. Landy, M. and Movshon, A. Cambridge, MA: MIT Press.

Attneave, F. (1982). Pragnanz and soap bubbles systems: A theoretical explanation. In *Organization and Representation in Perception*. ed. Beck, J. Hillsdale, NJ: Lawrence Erlbaum Ass.

Attneave, F. and Frost, R. (1969). The determination of perceived tridimensional orientation by minimum criteria. *Percept. Psychophys.* 6, 391–396.

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Res.*, 18, 637–650.

Beck, J. and Gibson, J. J. (1955). The relation of apparent shape to apparent slant in the perception of objects. *J. Exp. Psychol.* 50, 125–133.

Bennett, B. M., Hoffman, D. D. and Prakash, C. (1989). *Observer Mechanics: A Formal Theory of Perception*. NY: Academic Press, Inc.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychol. Rev.*, 94 (2), 115–147.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley, CA: Univ. of California Press.

Burgess, A. E., Wagner, R. F., Jennings, R. J. and Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, 214, 93–94.

Da Vinci, L. (1970). *The Notebooks of Leonardo da Vinci. Vol. I*. NY: Dover.

Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. NY: John Wiley and Sons.

Epstein, W. (1977). What are the prospects for a higher-order stimulus theory of perception? *Scand. J. Psychol.* 18, 164–171.

Flock, H. R. (1964). Three theoretical views of slant perception. *Psychol.* 62, 110–121.

Geisler, W. (1989). Sequential ideal- observer analysis of visual discriminations. *Psychol. Rev.*, 96, 267–314.

Gibson, E. (1982). Contrasting emphasis in Gestalt theory, information processing and the ecological approach to perception. In *Organization and Representation in Perception*. ed. Beck, J. Hillsdale, NJ: Lawrence Erlbaum Ass.

Gibson, J. J. (1979). *The Ecological Approach to Vision*. Boston, MA: Houghton-Mifflin.

Gilchrist, A. (1980). When does perceived lightness depend on perceived spatial arrangement?, *Percept. Psychophys.* 28 (6), 527–538.

Golden, R. (1988). A unified framework for connectionist systems. *Biol. Cybern.* 59, 109–120.

Gregory, R. L. (1973). *Eye and Brain: The Psychology of Seeing*. NY: McGraw-Hill.

Helmholtz, H. (1925). *Physiological Optics, Vol. III: The Perceptions of Vision*. (J. P. Southall, Trans.). Rochester, NY: Optical Society of America. (Original publication in 1910.)

Hochberg, J. (1974). Higher-order stimuli and inter-response coupling in the perception of the visual world. In *Perception: Essays in Honor of James J. Gibson*. eds. MacLeod, R. B. and Pick, H. L. NY: Cornell Univ. Press.

Hochberg, J. (1982). How big is a stimulus. In *Organization and Representation in Perception*. ed. Beck, J. Hillsdale, NJ: Lawrence Erlbaum Ass.

Hochberg, J. and McAlister, E. (1953). A quantitative approach to figural 'goodness'. *J. Exp. Psychol.*, 46, 361–364.

Hopfield, J. J. (1984). Neurons with graded response have collection computational properties like those of two-state neurones. *Proc. Nat. Acad. Sci. USA*, 81, 3088–3092.

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

Ittleson, W. H. (1960). *Visual Space Perception*. NY: Springer Publishing Co.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Res.*, 24, 1977–1990.

Kersten, D. (1990). Statistical limits to image understanding. In Blakemore, C. *Vision: Coding and Efficiency*. Cambridge, UK: Cambridge Univ. Press.

Kersten, D. and Knill, D. C. (1990). Human edge labelling: Examples of cooperative computation. *Invest. Ophthalmol. Vis. Sci. (Suppl.)*, 31 (4), 325.

Kersten, D., O'Toole, A. J., Sereno, M. E., Knill, D. C. and Anderson, J. A. (1987). Associative learning of scene parameters from images. *Appl. Opt.*, 26 (23), 4999–5006.

Knill, D. C. (1991). *The Role of Cooperative Processing in the Perception of Surface Shape and Reflectance*. PhD, Brown University, Providence, RI.

Knill, D. C. and Kersten, D. (1990). Learning a near-optimal estimator for surface shape from shading. *Computer Vision Graphics Image Processing*, 50 (1), 75–100.

Koenderink, J. J. (1984). What does occluding contour tell us about solid shape. *Perception*, 13, 321–330.

Koenderink, J. J. and van Doorn, A. J. (1976). The singularities of the visual mapping. *Biol. Cybern*, 24, 51–59.

Koenderink, J. J. and van Doorn, A. J. (1984). The shape of smooth objects and the way contours end. *Perception*, 11, 129–137.

Koffka, K. (1935). *Principles of Gestalt Psychology*. NY: Harcourt, Brace and Co.

Land, E. H. and McCann, J. J. (1971). Lightness and retinex theory, Opt. Soc. Am., 61, 1–11.

Leeuwenburg, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. *Am. J. Psychol.*, 84, 307–349.

Mach, E. (1980). *Contributions to the Analysis of the Sensations*. (C. M. Williams, Trans.). Chicago, IL: Open Court Publishing Co., (Original work published in 1890).

Marr, D. (1980). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. NY: W. H. Freeman and Company.

Marroquin, J. L. (1985). *Probabilistic Solution of Inverse Problems*. MIT-AI technical Report 860.

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

Richards, W. (1988). Introduction to natural computation. In *Natural Computation*. Cambridge, MA: MIT Press.

Rock, I. (1977). In defense of unconscious inference. In *Stability and Constancy in Visual Perception: Mechanism and Processes*. ed. Epstein, W. NY: John Wiley and Sons.

Sereno, M. E., Kersten, D. J. and Anderson, J. A. (1988). A neural network model of an aspect of motion perception. In *Annual Research Report of the Consortium for Scientific Computing*, Science at the John von Neumann National Supercomputer Center, pp. 173–178.

Szeliski, R. (1987). Regularization uses fractal priors. In *Proc. AAAI-87* pp. 749–745, Seattle, WA.

Szeliski, R. (1990). *Bayesian Modelling of Uncertainty in Low-Level Vision*. Norwell, MA: Kluwer Academic Press.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In *The Psychology of Computer Vision*. ed. Winston, P. H. pp. 19–91. NY: McGraw-Hill.