Stereo Integration, Mean Field Theory and Psychophysics

Alan L. Yuille
Division of Applied Science,
Harvard University, Cambridge, MA 02138

Davi Geiger
Siemens Research, Princeton, NJ 08540

and

Heinrich H. Bülthoff
Department of Cognitive Science and Linguistic Sciences,
Brown University, Providence, RI 02912

Abstract

We describe a theoretical formulation for stereo in terms of the Bayesian approach to vision. This formulation enables us to integrate the depth information from different types of matching primitives, or from different vision modules. We solve the correspondence problem using compatibility constraints between features and prior assumptions on the interpolated surfaces that result from the matching. We use techniques from statistical physics to show how our theory relates to previous work and to develop novel algorithms. Finally we show that, by a suitable choice of prior assumptions about surfaces, the theory is consistent with some psychophysical experiments which investigate the relative importance of different matching primitives.

1 Introduction

Computational vision (Marr, 1982) attempts to construct theoretical models of visual processes and relate them to psychophysical and physiological experiments. With the vast number of research papers on these topics it is desirable to develop an overall framework for such theories.

In this paper we introduce a theoretical formulation for stereo in terms of the Bayesian approach to vision, in particular in terms of coupled Markov Random Fields. We show that this formalism is rich enough to contain most of the elements used in standard stereo theories. Techniques from statistical physics are used to show its relations to previous theories and to suggest novel algorithms. Finally we show that the theory is consistent with some psychophysical experiments.

The fundamental issues of stereo are: (i) what primitives are matched between the two images, (ii) what *a priori* assumptions are made about the scene to determine the matching and thereby compute the depth, and (iii) how is the geometry and calibration of the stereo system determined. For this paper we assume that (iii) is solved, and so the corresponding epipolar lines (see, e.g., Horn, 1986) between the two images are known. Thus we use the epipolar line constraint to reduce the matching problem to a one-dimensional search. Some support for this is given by the work of Bülthoff and Fahle (1989), described in more detail in Section 5.

Our framework combines cues from different matching primitives to obtain depth perception. These primitives can be weighted according to their robustness. For example, depth estimates obtained by matching intensity are sometimes unreliable since small fluctuations in intensity (due to illumination or detector noise) might lead to large fluctuations in depth. Therefore intensity matches are less reliable than estimates from matching edges. The formalism also can be extended to incorporate information from other depth modules (see psychophysical experiments in Bülthoff and Mallot, 1987, 1988). It provides also a model for sensor fusion that can include domain dependent knowledge (Clark and Yuille, 1990).

Unlike many previous theories of stereo that first solved the correspondence problem and then constructed a surface by interpolation Grimson (1981), we propose combining

the two stages. The correspondence problem is solved to give the disparity field that best satisfies the *a priori* constraints. Our model involves the interaction of several processes. We will introduce it in three stages at different levels of complexity in sections (2.2), (2.3) and (2.4). It involves fields for matching, disparity and discontinuities.

Using standard techniques from statistical physics we can eliminate certain fields and obtain effective energies for the remaining fields (see Geiger and Girosi, 1989). By these methods we can relate our framework to the apparently rather different cooperative stereo algorithms (Dev, 1975; Marr and Poggio, 1976) and the disparity gradient limit theories (Prazdny, 1985; Pollard et al., 1985).

These techniques also suggest novel algorithms for stereo computation that incorporate constraints about the set of possible matches differently from the cooperative stereo algorithms. They can be directly related (Yuille, 1990) to analog methods for solving the traveling salesman problem (T.S.P.). The greater empirical success of the elastic net algorithm (Durbin and Willshaw, 1989) compared with the Hopfield and Tank method (1985) strongly suggests that our novel stereo algorithms will be more successful than the cooperative algorithms.

We relate this model to psychophysical experiments (Bülthoff and Mallot, 1988; Bülthoff and Fahle, 1989; Bülthoff et al., 1991) in which perceived depth for different matching primitives and disparity gradients are precisely measured. These experiments suggest that several types of primitive are used for correspondence, but that some primitives are better than others. Our model is consistent with the data from these experiments assuming standard prior assumptions about the surfaces.

Some of the energy functions formulated here, but without the statistical methods or the Bayesian framework, were developed in collaboration with Mike Gennert (Gennert et al., 1990) based on a model for long range motion correspondence by Yuille and Grzywacz (1989).

This paper uses differential operators to impose prior constraints on surfaces. For the discretized case this corresponds to a Markov Random Field. We would like to emphasize, however, that this choice of priors is motivated by consistency with a specific class of psychophysical experiments. Our framework allows for a far more general class of prior.

An attractive possibility (Clark and Yuille, 1990) is to use domain specific priors, if domain knowledge is available, or competitive and adaptive priors.

The plan of this paper is as follows: in Section 2 we review the Bayesian approach to vision and describe our framework. Section 3 introduces techniques from statistical physics and uses them to analyze the theory. This analysis is used in Section 4 to compare the model to existing stereo theories and to suggest novel algorithms. Finally, in Section 5, we relate the theory to psychophysical data.

2 The Bayesian Approach to Stereo

There has been a vast amount of work on stereo. For overviews, see Grimson (1981) and Barnard and Fischler (1982). We first briefly review the problem of stereo and give an overview of our theory. We describe this theory in terms of an energy function and finally put into a probabilistic framework.

2.1 The Matching Problem

The input to any binocular stereo system is a pair of images. The task is to match primitives of the two images, thereby solving the correspondence problem. The depth of objects in the scene can then be determined by triangulation, assuming the orientations of the cameras (and other camera parameters) are known. In many stereo theories the disparity, the relative distance between matched features, is first computed. The depth of the feature from the fixation point is then, to first order approximation, linearly dependent on its disparity.

There are several choices of matching primitives such as edges and peaks in image intensity, intensity itself, or varieties of measures of texture. It is unclear which primitives the human visual system uses. Psychophysics described in Section 5 suggests that at least edges and image intensity are used as primitives.

It is desirable to build a stereo theory that is capable of using all these different types of primitives. This will allow to reduce the complexity of the correspondence problem and will enhance the robustness of the theory and its applicability to natural images. However, not all primitives are equally reliable. A small fluctuation in the image intensity might lead to

a large change in the measured disparity for a system that matches intensity. Thus image intensity is usually less reliable than features such as edges.

Some assumptions about the scene are usually necessary to solve the correspondence problem. These can be thought of as natural constraints (Marr, 1982) and are needed because of the ill-posed nature of vision (Poggio et al., 1985). There are two types of assumption: (i) assumptions about the matching primitives, for example the *compatibility constraint* that similar features match, and (ii) assumptions about surfaces. For (ii) one typically assumes that either the surface is close to the fixation point (the disparity is small) or that the surface's orientation is smoothly varying (the disparity gradient is small) with a few possible discontinuities. We discuss specific smoothness measures in Section 2.2.

Our theory requires both assumptions, though their relative importance depends on the scene. If the features in the scene are sufficiently different then assumption (i) is often sufficient to obtain a good match. If all features are very similar, assumption (ii) is necessary. For really jagged surfaces, such as Bryce Canyon, stereo data cannot be fit to smooth surfaces, even with discontinuities. Therefore compatibility constraints are far more important than prior surface assumptions. Fortunately the more jagged the surface the more structure there will be in the image and the more powerful the compatibility constraint. Note that although there may be severe correspondence problems for simple edge features, there will usually be unique matches for clusters of edges or other primitives including intensity. In general we require that the matching is chosen to obtain the smoothest possible surface, so interpolation and matching are performed simultaneously. The next section formalizes these ideas.

2.2 The First Level: Matching Field and Disparity Field

The basic idea is that there are several possible primitives that could be used for matching and that these all contribute to a disparity field $d(x)$. This disparity field exists even where there is no source of data. The primitives we will consider here are features, such as edges in image brightness. Edges typically correspond to object boundaries, and other significant events in the image. Other primitives, such as peaks in the image brightness or texture features, can also be added. In the following we describe the theory for the one-dimensional

case.

We assume that the edges and other features have already been extracted from the image in a preprocessing stage. The matching elements in the left eye consist of the features $\{x_{i_L}\}$, for $i_L = 1, ..., N_l$. The right eye contains features $\{x_{a_R}\}$, for $a_R = 1, ..., N_r$. We define a set of binary matching elements $\{V_{i_L a_R}\}$, the matching field, such that $V_{i_L a_R} = 1$ if point i_L in the left eye corresponds to point a_R in the right eye, and $V_{i_L a_R} = 0$ otherwise. A *compatibility field* $\{A_{i_L a_R}\}$ is defined to be small if i_L and a_R are compatible (i.e., features of the same type) and large if they are incompatible (an edge cannot match a peak). For example, for random dot stereograms the features are all equally compatible and we can set $A_{i_L a_R} = 1$ for all i_L, a_R.

We now define a cost function $E(d, V)$ of the disparity field and the matching elements. We will interpret this in terms of Bayesian probability theory in the next section. This will suggest several methods to estimate the fields d, V given the data. A standard estimator is to minimize $E(d, V)$ with respect to d, V.

$$E(d,V) = \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2$$
$$+\lambda \{ \sum_{i_L} (\sum_{a_R} V_{i_L a_R} - 1)^2 + \sum_{a_R} (\sum_{i_L} V_{i_L a_R} - 1)^2 \}$$
$$+\gamma \int_M (Sd)^2 dx. \qquad (1)$$

The first term gives a contribution to the disparity obtained from matching i_L to a_R. The 4th term imposes a prior constraint on the disparity field imposed by a suitable differential operator S.

The second and third terms encourage features to have a single match, they also can be imposed using techniques from statistical physics which ensure that each column and row of the matrix $V_{i_L a_R}$ contains at most one 1. In Section 3 we will argue that it is better to impose constraints in this way, therefore the second term will only be used in our final theory to give a penalty λ for unmatched points. We will keep it in our energy function, however, since it will help us to relate our approach to alternative theories.

Minimizing the energy function with respect to $d(\vec{x})$ and $V_{i_L a_R}$ will cause the matching which results in the interpolated disparity field best satisfying the prior constraints. We

discuss ways of doing this minimization in Section 3.

The coefficient γ determines the amount of *a priori* knowledge required. If all the features in the left eye have only one compatible feature in the right eye then little *a priori* knowledge is needed and γ may be small. If all the features are compatible then there is matching ambiguity which the *a priori* knowledge is needed to resolve, requiring a larger value of γ and therefore more smoothing. In Section 5 we show that this gives a possible explanation for some psychophysical experiments.

Compatibility is enforced multiplicatively in the first term of (1). An alternative would be to choose $\sum_{i_L,a_R} V_{i_L a_R}\{(d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 + A_{i_L a_R}\}$.

The theory can be extended to two-dimensions in a straightforward way. The matching elements $V_{i_L a_R}$ must be constrained to allow only for matches that use the epipolar line constraint. The disparity field will have a smoothness constraint perpendicular to the epipolar line which will enforce figural continuity.

Finally, and perhaps most importantly, we must choose a form for the differential operator S which imposes the prior constraints. Marr (1982) proposed that, to make stereo correspondence unambiguous, the human visual system assumes that the world consists of smooth surfaces. This suggests that we should choose a smoothness operator that encourages the disparity to vary smoothly spatially. In practice the assumptions used in Marr's two theories of stereo are somewhat stronger. Marr and Poggio I (1976) encourages matches with constant disparity, thereby enforcing a bias to the fronto-parallel plane. Marr and Poggio II (1979) uses a coarse to fine strategy to match nearby points, therefore encouraging matches with least disparity and by that giving a bias toward the fixation plane.

The simplest smoothness constraint corresponds to approximating a membrane surface. This gives an operator $S = \partial/\partial x$ and leads to a discretized term $\sum_k (d_k - d_{k+1})^2$. We will use this operator as a default choice for our theory. We will use it in conjunction with discontinuity fields, introduced in the next section, which break the smoothness constraint. This specific choice of operator is consistent with the experiments described in Section 5.

We would like to reiterate that the choices of prior are motivated by consistency with the psychophysical experiments described in Section 5. Alternate choices, which need not necessarily be implemented by differential operators, may be preferable for computer vision

systems.

2.3 The Second Level: Adding Discontinuity Fields

The first level theory is easy to analyze but makes the *a priori* assumption that the disparity field is smooth everywhere, which is false at object boundaries. There are several standard ways to allow smoothness constraints to break (Blake, 1983; Geman and Geman, 1984; Mumford and Shah, 1989). We introduce a discontinuity field $l(x)$ represented by a set of curves C.

Introducing the discontinuity fields C gives an energy function

$$\begin{aligned} E(d(x), V_{i_L a_R}, C) = & \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 \\ & + \lambda \{ \sum_{i_L} (\sum_{a_R} V_{i_L a_R} - 1)^2 + \sum_{a_R} (\sum_{i_L} V_{i_L a_R} - 1)^2 \} \\ & + \gamma \int_{M-C} (Sd)^2 dx + M(C). \end{aligned} \tag{2}$$

where smoothness is not enforced across the curves C and $M(C)$ is the cost for enforcing breaks. Again the constraints in the second term on the right hand side of (2) will be imposed by statistical physics techniques.

2.4 The Third Level: Adding Intensity Terms

The final version of the theory couples intensity based and feature based stereo. Psychophysical results (see Section 5.2) suggest that this is necessary. Our energy function becomes

$$\begin{aligned} E(d(x), V_{i_L a_R}, C) = & \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 \\ & + \mu \int \{ L(x) - R(x + d(x)) \}^2 dx \\ & + \lambda \{ \sum_{i_L} (\sum_{a_R} V_{i_L a_R} - 1)^2 + \sum_{a_R} (\sum_{i_L} V_{i_L a_R} - 1)^2 \} \\ & + \gamma \int_{M-C} (Sd)^2 dx + M(C). \end{aligned} \tag{3}$$

If certain terms are set to zero in (3) it reduces to previous energy function formulations of stereo. If the second and 4th terms are kept, without allowing discontinuities, it is similar to work by Gennert (1987) and Barnard (1989). If we add the fifth term, and allow discontinuities, we get connections to some work described in Yuille (1989) (done in collaboration with T. Poggio). The third term will again be removed in the final version of the theory.

Thus the cost function (3) reduces to well-known stereo theories in certain limits. It also shows how it is possible to combine feature and brightness data in a natural manner.

2.5 The Bayesian Formulation

Given an energy function model one can define a corresponding statistical theory. If the energy $E(d, V, C)$ depends on three fields: d (the disparity field), V the matching field and C (the discontinuities), then (using the Gibbs distribution – see Parisi 1988) the probability of a particular state of the system is defined by

$$P(d, V, C|g) = \frac{e^{-\beta E(d,V,C)}}{Z} \tag{4}$$

where g is the data, β is the inverse of the temperature parameter and Z is the partition function (a normalization constant).

Using the Gibbs Distribution we can interpret the results in terms of Bayes' formula (Bayes, 1783),

$$P(d, V, C|g) = \frac{P(g|d, V, C)P(d, V, C)}{P(g)} \tag{5}$$

where $P(g|d, V, C)$ is the probability of the data g given a scene d, V, C. $P(d, V, C)$ is the *a priori* probability of the scene and $P(g)$ is the *a priori* probability of the data. Note that $P(g)$ appears in the above formula as a normalization constant, so its value can be determined if $P(g|d, V, C)$ and $P(d, V, C)$ are assumed known.

This implies that every state of the system has a finite probability of occurring. The more likely ones are those with low energy. This statistical approach is attractive because the β parameter gives us a measure of the uncertainty of the model temperature parameter $T = \frac{1}{\beta}$. At zero temperature ($\beta \to \infty$) there is no uncertainty. In this case the only state of

the system that have nonzero probability. Therefore probability 1, is the state that globally minimizes $E(d,V,C)$. Although in some nongeneric situations there could be more then one global minimum of $E(d,V,C)$.

Minimizing the energy function will correspond to finding the most probable state, independent of the value of β. The mean field solution,

$$\bar{d} = \sum_{d,V,C} dP(d,V,C|g), \tag{6}$$

is more general and reduces to the most probable solution as $T \rightarrow 0$. It corresponds to defining the solution to be the mean fields, the averages of the f and l fields over the probability distribution. This enables us to obtain different solutions depending on the uncertainty.

In this paper we concentrate on the mean quantities of the field. An additional justification for using the mean field is that it represents the minimum variance Bayes estimator (Gelb 1974). More precisely, the variance of the field d is given by

$$Var(d:\bar{d}) = \sum_{d,V,C} (d-\bar{d})^2 P(d,V,C|g) \tag{7}$$

where $\bar{d}$ is the center of the variance and the $\sum_{d,V,C}$ represents the sum over all the possible configurations of d,V,C. Minimizing $Var(d:\bar{d})$ with respect to all possible values of $\bar{d}$ we obtain

$$\frac{\partial}{\partial \bar{d}} Var(d:\bar{d}) = 0 \rightarrow \bar{d} = \sum_{d,V,C} dP(d,V,C). \tag{8}$$

This implies that the minimum variance estimator is given by the mean field value.

3 Statistical mechanics and mean field theory

In this section we discuss methods for calculating the quantities we are interested in from the energy function and propose novel algorithms. One can estimate the most probable states of the probability distribution (5) by, for example, using Monte Carlo techniques (Metropolis et al., 1953) and the simulated annealing approach (Kirkpatrick et al., 1983). The drawback of these methods is the amount of computer time needed for the implementation.

There are yet many other techniques from statistical physics that can be applied. They have recently been used to show (Geiger and Girosi, 1989; Geiger and Yuille, 1989) that several seemingly different approaches to image segmentation are closely related.

There are two main uses of these techniques: (i) we can eliminate (or average out) different fields from the energy function to obtain effective energies depending on only some fields (therefore relating our model to previous theories) and (ii) we can obtain methods for finding deterministic solutions.

There is an additional important advantage in eliminating fields — we can impose global constraints on the possible fields by only averaging over fields that satisfy these constraints. For example, Geiger and Yuille (1989) describe two possible energy function formulations of a winner-take-all network in which binary decision units determine the "winner" from a set of inputs. The global constraint that there is only one winner can be expressed by: (i) introducing a term in the energy function to penalize configurations with more than one winner, or (ii) computing the mean fields by averaging only over configurations with a unique winner. The second method, known as *strong constraints*, is preferable because it gives the correct solution directly with least computation while the first method leads to a dynamic system that may not converge to the correct solution.

For the first level theory, see Section 3.1, it is possible to eliminate the disparity field to obtain an effective energy $E_{eff}(V_{ij})$ depending only on the binary matching field V_{ij}, which is related to cooperative stereo theories (Dev, 1975; Marr and Poggio, 1976). Alternatively, see Section 3.2, we can eliminate the matching fields to obtain an effective energy $E_{eff}(d)$ depending only on the disparity. Since the first method involves energy biases to impose global constraints we believe it will be less effective than the second approach, where the constraints are built in.[1]

We also can average out the line process fields and the matching fields or both for the second and third level theories. This leaves us again with a theory depending only on the

[1] It can be shown (Yuille, 1990) that there is a direct correspondence between these two theories (with $E_{eff}(V_{ij})$ and $E_{eff}(d)$) and analog models for solving the traveling salesman problem by Hopfield and Tank (1985) and Durbin and Willshaw (1987) (see Simic 1990 for the relation between these models). The far greater empirical success of the Durbin and Willshaw algorithm suggests that the first level stereo theory based on $E_{eff}(d)$ will be more effective than the cooperative stereo algorithms.

disparity field, for details see Yuille, Geiger and Bülthoff (1989).

Alternatively we can use (Yuille et al., 1989) mean field theory methods to obtain deterministic algorithms for minimizing the first level theory $E_{eff}(V_{ij})$. These differ from the standard cooperative stereo algorithms and should be more be more effective, though not as effective as using $E_{eff}(d)$, since they can be interpreted as performing the cooperative algorithm at finite temperature thereby smoothing local minima in the energy function.

Our proposed stereo algorithms, therefore, consist of eliminating the matching field and the line process field by these statistical techniques leaving an effective energy depending only on the disparity field. This formulation will depend on the parameter β (the inverse of the temperature of the system). We then intend to minimize the effective energy by steepest descent while lowering the temperature (increasing β). This can be thought of as a deterministic form of simulated annealing and has been used by many algorithms, for example (Hopfield and Tank, 1985; Durbin and Willshaw, 1989; Geiger and Girosi, 1989). It is also related to continuation methods (Wasserstrom, 1973).

3.1 Averaging out the Disparity Field

We now show that, if we consider the first level theory, we can eliminate the disparity field and obtain an energy function depending on the matching elements V only. In Section 4 we will relate this to cooperative stereo algorithms.

The disparity field is eliminated by minimizing and solving for it as a function of the V. Since the disparity field occurs quadratically this is equivalent to doing mean field over the disparity (Parisi, 1988).

For the first level theory, assuming all features are compatible, our energy function becomes

$$E(d(x), V_{i_L a_R}) = \sum_{i_L, a_R} V_{i_L a_R}(d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 + \mu \sum_{i_L}(\sum_{a_R} V_{i_L a_R} - 1)^2 + \mu \sum_{a_R}(\sum_{i_L} V_{i_L a_R} - 1)^2 + \lambda \int_M (Sd)^2 dx, \qquad (9)$$

with Euler-Lagrange equations

$$\lambda S^2 d(x) = \sum_{i_L, a_R} V_{i_L a_R}(d(x_{i_L}) - (x_{a_R} - x_{i_L}))\delta(x - x_{i_L}). \quad (10)$$

The solutions of this equation are given by

$$d(x) = \sum_{i_L} \alpha_{i_L} G(x, x_{i_L}), \quad (11)$$

where the $G(x, x_{i_L})$ are the Green function of the operator S^2, and the α_{i_L} obey

$$\sum_{i_L} \alpha_{i_L}(\lambda \delta_{i_L a_R} + G(x_{i_L}, x_{a_R})) = \sum_{i_L} V_{i_L a_R}(x_{i_L} - x_{a_R}). \quad (12)$$

Substituting this back into the energy function (assuming that each feature is matched precisely once, *uniqueness constraint*) eliminates the disparity field and yields

$$E(V_{i_L a_R}) = \lambda \sum_{i_L, j_L} (\sum_{a_R} V_{i_L a_R}(x_{i_L} - x_{a_R}))(\lambda \delta_{i_L j_L} + G(x_{i_L}, x_{j_L}))^{-1} (\sum_{b} V_{j_L b_R}(x_{j_L} - x_{b_R}))$$
$$+\lambda \sum_{i_L} (\sum_{a_R} V_{i_L a_R} - 1)^2 + \lambda \sum_{a_R} (\sum_{i_L} V_{i_L a_R} - 1)^2 \quad (13)$$

This calculation shows that the disparity field is, strictly speaking, unnecessary since the theory can be formulated as in (13). We discuss the connection to cooperative stereo algorithms in Section 4. Note that the global constraints are imposed by energy biases[2].

3.2 Averaging out the matching fields

We prefer an alternative way of writing the first level theory. This can be found by using techniques from statistical physics to average out the matching field while imposing global constraints, leaving a theory that depends only on the disparity field.

The partition function for the first level system, again assuming compatibility between all features, is defined to be

[2] A similar calculation (Yuille and Grzywacz, 1989) showed that minimal mapping theory (Ullman, 1979) was a special case of the motion coherence theory.

$$Z = \sum_{V,d} e^{-\beta E(V,d)}, \tag{14}$$

where the sum is taken over all possible states of the system determined by the fields V and d.

It is possible to perform explicitly the sum over the matching field V yielding an effective energy for the system depending only on the disparity field d.

To compute the partition function we must first decide what class of V we wish to sum over. We might try to impose the constraint that each point has a unique match by only summing over $\{V_{i_L a_R}\}$ which contain a single 1 in each row and each column. We might further restrict the class of possible matches by requiring that they satisfied the ordering constraint. However, we cannot compute the partition function analytically for these cases. To get analytic results we must be more pragmatic.

For this section we will initially restrict that each feature in the left image has at most one match in the right image, but not vice versa. This simplifies the computation of the partition function, but we will relax it at the end of the section. The requirement of smoothness on the disparity field should ensure that unique matches occur. This follows from a mathematical analysis of a similar algorithm used for an elastic network approach to the T.S.P. (Durbin et al., 1989). It also should imply that the ordering constraint is satisfied.

Since we are attempting to impose the matching constraint by restricting the class of V's the only effect of the $\lambda \sum_{i_L} (\sum_{a_R} V_{i_L a_R} - 1)^2 + \lambda \sum_{a_R} (\sum_{i_L} V_{i_L a_R} - 1)^2$ terms is to impose a penalty λ for unmatched points. We can now write the partition function as

$$Z = \sum_{V,d} \prod_{i_L} e^{-\beta \{ \sum_{a_R} V_{i_L a_R} \left(d(x_{i_L}) - (x_{a_R} - x_{i_L}) \right)^2 + \int_M \left(Sd \right)^2 dx \}}. \tag{15}$$

For fixed i_L we sum over all possible $\{V_{i_L a_R}\}$, such that $V_{i_L a_R} = 1$ for at most one a_R (this ensures that points in the left image have at most one match to points in the right image). This gives

$$Z = \sum_{d} \prod_{i_L} \{ \sum_{a_R} e^{-\beta \left(d(x_{i_L}) - (x_{a_R} - x_{i_L}) \right)^2} \} e^{-\beta \int_M \left(Sd \right)^2 dx}. \tag{16}$$

This can be written using an effective energy $E_{eff}(d)$ as

$$Z = \sum_{d} e^{-\beta E_{eff}(d)}. \tag{17}$$

where

$$E_{eff}(d) = \frac{-1}{\beta} \sum_{i_L} log\{e^{-\beta\lambda} + \sum_{a_R} e^{-\beta\left(d(x_{i_L}) - (x_{a_R} - x_{i_L})\right)^2}\} + \int_M (Sd)^2 dx. \tag{18}$$

It is probably preferable, however, to impose symmetry between the two eyes. We do this by summing over states where the points in the right image have a unique match to points in the left image. This modifies the effective energy by the addition of a term $E_{aux}(d)$, where

$$E_{aux}(d) = \frac{-1}{\beta} \sum_{a_R} log\{e^{-\beta\lambda} + \sum_{i_L} e^{-\beta\left(d(x_{a_R}) - (x_{a_R} - x_{i_L})\right)^2}\}. \tag{19}$$

Preliminary experiments suggest that the symmetric energy function $E_{eff}(d) + E_{aux}(d)$ has fewer local minima than $E_{eff}(d)$ and is easier to compute.

To minimize $E_{eff}(d) + E_{aux}(d)$ we use a deterministic annealing strategy, minimizing for large temperature and tracking the solution as $T \to 0$.

Thus our first level theory of stereo can be formulated in terms of disparity without explicitly using a matching field. The theory imposes the global matching constraints strongly.

A second modification (which arose during conversations with S. Mallat) is to allow features to be unmatched, provided they pay a penalty μ. This gives an additional configuration (zero match) that is must be included when evaluating the partition function. It results in an effective energy (with the symmetric formulation) of

$$\begin{aligned} E_{eff}(d) = \frac{-1}{\beta} \sum_{i_L} log\{e^{-\beta\mu} + \sum_{a_R} e^{-\beta\left(d(x_{i_L}) - (x_{a_R} - x_{i_L})\right)^2}\} + \int_M (Sd)^2 dx \\ + \frac{-1}{\beta} \sum_{a_R} log\{e^{-\beta\mu} + \sum_{i_L} e^{-\beta\left(d(x_{a_R}) - (x_{a_R} - x_{i_L})\right)^2}\}. \end{aligned} \tag{20}$$

3.3 Averaging out fields for the second and third level theories

The second and third level theory includes a binary discontinuity field in addition to the matching field and the disparity field. This discontinuity field can be averaged out using the techniques developed in Geiger and Girosi (1989). This leads to an effective energy that depends only on the disparity field (Yuille et al., 1989).

4 Comparisons with other theories

In the previous section we have used methods from statistical physics to analyze our theory. In particular, we have shown how we can eliminate some fields to obtain equivalent, but apparently different, theories. In this section we show that this helps compare our work to previous theories.

4.1 The Cooperative Stereo Algorithm

The cooperative stereo algorithm (Dev, 1975; Marr and Poggio, 1976) contains binary units C_{ia} for each lattice point i in the left image and a in the right image. Note that, unlike our V_{ia}, these units do not occur only where there are features. The C_{ia} are initialized to be 1 if there are features at i and a in the left and right images, therfore a potential match, otherwise they are 0.

The algorithm corresponds to doing iterative gradient descent with an energy function

$$E[C] = \frac{1}{2} \sum_{i,j,a,b} T_{ijab} C_{ia} C_{jb} + \sum_{i,a} C_{ia}(0) C_{ia}, \tag{21}$$

where T_{ijab} imposes excitation for matches with similar disparity and inhibition for more than one match per feature. A specific choice of T_{ijab} might be

$$\begin{aligned} T_{ijab} = k_1\{\delta_{i,j-1}\delta_{a,b-1} + \delta_{i,j+1}\delta_{a,b+1}\} \\ -k_2\{\delta_{a,b}\delta_{i,j+1} + \delta_{a,b}\delta_{i,j-1} + \delta_{ij}\delta_{a,b+1} + \delta_{i,j}\delta_{a,b-1}\}. \end{aligned} \tag{22}$$

We now compare cooperative stereo algorithm to our theory without line processors. This has energy function

$$E(V) = \alpha \sum_{i,j} (\sum_{a} V_{ia}(x_i - x_a))(\alpha\delta_{ij} + G(x_i, x_j))^{-1} (\sum_{b} V_{jb}(x_j - x_b)) + \alpha \sum_{i_L} (\sum_{a_R} V_{i_L a_R} - 1)^2 + \alpha \sum_{a_R} (\sum_{i_L} V_{i_L a_R} - 1)^2 \quad (23)$$

Comparing the two energy functions we see several similarities. There is a general excitation for a match in the direction of constant disparity, due to the first term, and inhibition of matching in the viewing directions, due to the second term. Both theories suffer from imposing global constraints as energy biases. The principle difference is that our method only has matching elements at feature points, rather than everywhere in the image.

The cooperative stereo algorithm can be seen (Yuille et al., 1989) to be the zero temperature limit of a statistical physics algorithm. This suggests that convergence would improve for a deterministic annealing algorithm.

The disparity gradient limit theories (Pollard et al., 1985; Prazdny, 1985) also can be thought of as extremizing a cost function similar to (21). The difference being that the support for a specific match is over all matches in the neighborhood whose disparity gradient is below a certain cut-off value. Thus the directions of excitation are more general than for the cooperative stereo algorithm, therefore the theory makes fewer prior assumptions about the surface.

Moreover these theories also use a version of strong constraints for matching (instead of using energy biases). The strategy is to fix the indices i, a and then for each j in a suitable neighborhood of i to do winner-take-all on b (imposing a unique winner). This gives the "support" for i to be matched to a. Winner-take-all is again employed to find the best match for i. The process is repeated for matching from the other eye and the results are checked for consistency.

The disparity gradient limit theories seem preferable to the original cooperative stereo algorithms since they have a more general prior and use a version of strong constraints. Moreover the forty-second computer implementation of the theory gives very impressive results on real data (Mayhew, Frisby - many personal communications). Despite the many advantages of the disparity gradient theories we argue that it is preferable to represent

depth explicitly by surfaces, using more than one surface (Yuille et al., 1990) if transparency occurs. Such a representation in terms of surfaces is easier to integrate with other depth information and can be used to incorporate domain specific knowledge (Clark and Yuille, 1990). Moreover, see next section, it seems that the disparity gradient theories cannot account for the experimental data reported here.

5 Comparisons to Psychophysics

In this section we describe the relationship between our theory and psychophysics. We will chiefly be concerned with two experiments (Bülthoff and Fahle, 1989; Bülthoff and Mallot, 1987) which provide measurements of the perceived depth as a function of the real depth and the matching primitives by use of reference systems (or depth probes). These experiments are particularly useful for our purposes because of: (i) the quantitative depth information they supply and (ii) their investigation of which features are used for matching and how the perceived depth depends on these features.

One general conclusion from these experiments is that objects perceived stereoscopically tend to be biased toward the fronto-parallel plane and the degree of this bias depends on the features being matched. This is in general agreement with our theory in which the disparity smoothness term causes such a bias with a magnitude depending on the robustness and discriminability of the features (as imposed by the compatibility terms).

5.1 The Disparity Gradient Experiments

For these experiments (Bülthoff and Fahle, 1989) the observer was asked to estimate the disparity of stereo stimuli relative to a set of reference lines. The stimuli were either lines at various angles or pairs of dots or simple geometric features. The perceived depth was plotted as a function of the disparity and the disparity gradient. These were calculated, assuming features at x_1, x_2 $(x_1 > x_2)$ and y_1, y_2 $(y_1 > y_2)$ in the left and right eyes, using the formulae (Burt and Julesz, 1980) for: (i) disparities $d_1 = (x_1 - y_1)$, $d_2 = (x_2 - y_2)$, (ii) the binocular disparity $d_b = d_1 + d_2$, and (iii) the disparity gradient $d_{grad} = d_b / \{(x_1 + y_1) - (x_2 + y_2)\}$.

The experiments showed that the perceived disparity decreased as a function of the

disparity gradient. This effect was: (i) strongest for horizontal lines, (ii) strong for pairs of dots or similar features, (iii) weak for dissimilar features and (iv) weak for non-horizontal lines (Fig. 1).

Figure 1 about here

Our explanation assumes these effects are due to the matching strategy and is based on the second level theory, with energy function given by (1). The idea is that the smoothness term (the third term) is required to give unique matching but that its importance, measured by γ, increases as the features become more similar. If the features are sufficiently different (perhaps pre-attentively discriminable) then there is no matching ambiguity, so the correct disparities are obtained. If the features are similar then smoothness (or some other a priori assumption) must be used to obtain a unique match, leading to biases toward the fronto-parallel plane. The greater the similarity between features the more the need for smoothness and therefore the stronger the bias toward the fronto-parallel plane. The discontinuity field is switched on at both the points ensuring that smoothness is only imposed between the two points. Thus the two points are considered the boundaries of an object and only the object itself is smoothed.

We now analyze the second level theory and show that it predicts the falloff of perceived disparity with disparity gradient, provided we choose the smoothness operator to be the first derivative of the disparity. The change of rate of falloff for different types of features is due to varying γ as described above.

Suppose we have features x_1, x_2 in the left eye and y_1, y_2 in the right eye. We define matching elements V_{ai} so that $V_{ai} = 1$ if point a in the first image matches point i in the second image, $V_{ai} = 0$ otherwise. The matching strategy assumes we minimize

$$E(V_{ai}, d(x)) = \sum_{a,i} V_{ai}\{d(x_a/2 + y_i/2) - (x_a - y_i)\}^2 + \gamma \int \{Ld(x)\}^2 dx \qquad (24)$$

with respect to the V_{ai} and the disparity field $d(x)$. We assume $d(x)$ is only defined between the dots, due to the discontinuities at the boundaries and L is a differential operator. Apparently only some operators L are consistent with the experiments.

Assume that $L = \partial/\partial x$, which corresponds to a smoothness term

$$\gamma \int_{z_1}^{z_2} \{\frac{\partial d(x)}{\partial x}\}^2 dx, \tag{25}$$

i.e., just minimizing the gradient of the disparity field.

Suppose the minimum energy state corresponds to matching x_1 with y_1 and x_2 with y_2. Then we have points $z_1 = x_1/2 + y_1/2$ and $z_2 = x_2/2 + y_2/2$ with measured disparities $\alpha_1 = x_1 - y_1$ and $\alpha_2 = x_2 - y_2$ (assume $z_2 > z_1$).

The perceived disparity will correspond to the field $d(x)$. It is obtained by minimizing

$$E(d(x)) = \{d(z_1) - \alpha_1\}^2 + \{d(z_2) - \alpha_2\}^2 + \gamma \int_{z_1}^{z_2} \{\frac{\partial d(x)}{\partial x}\}^2 dx. \tag{26}$$

The minimum will occur when $d(x)$ is a straight line passing through $d(z_1)$ at z_1 and $d(z_2)$ at z_2 ($d(z_1)$ and $d(z_2)$ are to be determined). This gives

$$E(d_1, d_2) = \{d(z_1) - \alpha_1\}^2 + \{d(z_2) - \alpha_2\}^2 + \gamma \frac{\{d(z_1) - d(z_2)\}^2}{z_2 - z_1}. \tag{27}$$

Minimizing with respect to $d(z_1)$ and $d(z_2)$ gives, assuming $\gamma << (z_2 - z_1)$,

$$\{d(z_1) - d(z_2)\} = \{\alpha_1 - \alpha_2\} - \frac{2\gamma\{\alpha_1 - \alpha_2\}}{(z_2 - z_1)} + O(\gamma/(z_2 - z_1))^2. \tag{28}$$

Thus the perceived disparity equals the true disparity minus 2γ times the disparity gradient. This seems in general agreement with the experiments. As the features become less distinguishable more smoothness is required to give an unambiguous match causing γ to be increased and increasing the gradient of the falloff.

The results are not consistent with several possible choices of the smoothness operator, such as the second derivative of the disparity $\partial^2 d/\partial x^2$. It is straightforward to calculate that this choice will give $d(z_1) = \alpha_1$ and $d(z_2) = \alpha_2$, and therefore does not bias toward the fronto-parallel plane. It is likely that the smoothness operator S must contain a $\partial/\partial x$ term to ensure the observed fronto-parallel bias.

5.2 The Matching Primitive Experiments

These experiments (Bülthoff and Mallot, 1987; Bülthoff and Mallot, 1988) compared the relative effectiveness of image intensity and edges as matching primitives. The stimuli were chosen to give a three dimensional perception of an ellipsoid. The observer used a stereo depth probe to make a pointwise estimate of the perceived shape (Fig. 2a). The data show that depth can be derived from images with disparate shading even in the absence of disparate edges. The perceived depth, however, was weaker for shading disparities (about seventy percent of the true depth).

Figure 2a, b about here

Putting in edges or features helped improve the accuracy of the depth perception. But in some cases these additional features appeared to decouple from the intensity and were perceived to lie above the depth surface generated from the intensity disparities. These results are again in general agreement with our model. The edges give good estimates of disparity and so little *a priori* smoothness is required. The disparity estimates from the intensity, however, are far less reliable because small fluctuations of intensity might yield large fluctuations in the disparity. Therefore more *a priori* smoothness is required to obtain a stable result. This causes a weaker perception of depth or a bias to fronto-parallel.

The use of the intensity peak as a matching feature is vital (at least for the edgeless case) since it ensures that the image intensity is accurately matched. Some stereo theories based purely on intensity give an incorrect match for these stimuli (Bülthoff and Yuille, 1990). For the ellipsoids, however, the peak is difficult to localize and depth estimates based on it are not very reliable. Thus the peak cannot pull the rest of the surface to the true depth.

Bülthoff and Mallot (1988) found that pulling up did occur for the edgeless case if a dot was added at the peaks of the images (Fig. 3). This is consistent with our theory since, unlike the peaks, the dots are easily localized and matching them would give a good depth estimate. Our present theory, however, is not consistent with a perception that sometimes occurred for this stimulus. Some observers saw the dots as lying above the surface rather than being part of it. This may be explained by the extension of our theory to transparent surfaces (Yuille et al., 1990).

Figure 3 about here

5.3 Summary of Psychophysics

We have shown that the experiments described above can be explained in terms of a Bayesian model (of our type) with suitably chosen priors. It is unclear whether other existing stereo theories could explain them. Most theories, such as the coarse-to-fine algorithm (Marr and Poggio, 1979) and the disparity gradient limit theories (Prazdny, 1985; Pollard et al., 1985), are designed to get the correct depth if the correspondence problem is correctly solved. Therefore these theories will not give the types of fronto-parallel biases reported here.

For another class of experiments (Mitchison - personal communication) there is a very strong bias toward the fronto-parallel plane which seems inconsistent with our theory. Mitchison's explanation, based on invariance under head-eye movements is, however, unable to account for our effects. There seems no way to get the differential fronto-parallel bias depending on the form of the features discussed in section 5.1 nor to obtain the squashing of the ellipsoid described in section 5.2. Since our theory and Mitchison's are complementary it seems likely that they could be combined in an extended version that includes head and eye movements.

6 Conclusion

We have derived a theory of stereo on theoretical grounds using the Bayesian approach to vision. This theory can incorporate most of the desirable elements of stereo and it is closely related to several existing theories. The theory can combine information from matching different primitives, which is desirable on computational and psychophysical grounds. The formulation can be extended to include monocular depth cues, domain dependent priors and competitive priors (Clark and Yuille, 1990).

A basic assumption of our work is that correspondence and interpolation should be performed simultaneously. This is related to the experimental and theoretical work of Mitchison (1988) and Mitchison and McKee (1987) which suggests that the initial matching

uses a planarity assumption.

The use of mean field theory allows us to average out fields and to make mathematical connections between different formulations of stereo. It also suggests novel algorithms for computing the estimators due to enforcing the global matching constraints while performing the averaging, see Section 3.2. We argue that these algorithms are likely to be more effective than existing cooperative algorithms which impose global constraints with energy biases.

Finally, for a suitable choice of priors, the theory is consistent with some psychophysical experiments (Bülthoff and Mallot, 1987; Bülthoff and Mallot, 1988; Bülthoff and Fahle, 1989). More complex scenes probably require more sophisticated priors than those which we have used to explain our experiments on simple scenes. Our ongoing research in computer graphics psychophysics linked to bayesian computer vision attempts to generalize these models to more natural scenes.

7 Acknowledgements

A.L.Y. would like to acknowledge support from the Brown/Harvard/MIT Center for Intelligent Control Systems with U.S. Army Research Office grant number DAAL03-86-K-0171 and from DARPA with contract AFOSR-89-0506. H.H.B. work at MIT was supported by the Office of Naval Research, Cognitive and Behavioral Sciences Division. Some of these ideas were initially developed with Mike Gennert. We would like to thank Jim Clark, Manfred Fahle, Norberto Grzywacz, Stephan Mallat, David Mumford and Tommy Poggio for many helpful discussions.

References

Barnard, S. (1989). Stochastic stereo matching over scale. *International Journal of Computer Vision*, 3:17–33.

Barnard, S. T. and Fischler, M. A. (1982). Computational stereo. *ACM Comput. Surveys*, 143:553–572.

Bayes, T. (1783). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, 53:370–418.

Blake, A. (1983). The least disturbance principle and weak constraints. *Pattern Recognition Letters*, 1:393–399.

Bülthoff, H. H. and Fahle, M. (1989). Disparity gradients and depth scaling. A.I. Memo No. 1175, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Bülthoff, H. H., Fahle, M., and Wegmann, M. (1991). Disparity gradients and depth scaling. *Perception.* in press.

Bülthoff, H. H. and Mallot, H. A. (1987). Interaction of different modules in depth perception. In *Proceedings of the 1st International Conference on Computer Vision*, pages 295–305.

Bülthoff, H. H. and Mallot, H. A. (1988). Interaction of depth modules: stereo and shading. *Journal of the Optical Society of America*, 5:1749–1758.

Bülthoff, H. H. and Yuille, A. (1990). Bayesian models for seeing shapes and depth. Harvard Robotics Laboratory Technical Report 90-11, Harvard University.

Burt, P. and Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, 208:615–617.

Clark, J. and Yuille, A. (1990). *Data Fusion for Sensory Information Processing Systems.* Kluwer Academic Publishers, Boston, MA.

Dev, P. (1975). Perception of depth surfaces in random-dot stereograms: A neural model. *Int. J. Man-Machine Stud.*, 7:511–528.

Durbin, R., Szeliski, R., and Yuille, A. L. (1989). An analysis of the elastic net approach to the travelling salesman problem. *Neural Computation*, 1:348–358.

Durbin, R. and Willshaw, D. (1989). An analog approach to the travelling salesman problem using an elastic net method. *Nature*, 326:689–691.

Geiger, D. and Girosi, F. (1989). Parallel and deterministic algorithms from mrfs :integration and surface reconstruction. A.I. Memo No. 1114, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Geiger, D. and Yuille, A. (1989). A common framework for image segmentation. Harvard Robotics Laboratory Technical Report 89-7, Harvard University.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Gennert, M. (1987). *A Computational Framework for Understanding Problems in Stereo Vision.* PhD thesis, MIT.

Gennert, M., Ren, B., and Yuille, A. L. (1990). Stereo matching by energy function minimization. In *Proc. SPIE: Visual Communications and Image Processing.*

Grimson, W. E. L. (1981). *From Images to Surfaces.* MIT Press, Cambridge, MA.

Hopfield, J. J. and Tank, D. W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52:141–152.

Horn, B. K. P. (1986). *Robot vision.* MIT Press, Cambridge, Mass.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220:671–680.

Marr, D. (1982). *Vision.* W. H. Freeman, San Francisco, CA.

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

Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B*, 204:301–328.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Phys. Chem.*, 21:1087–1091.

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

Mitchison, G. J. and McKee, S. (1987). The resolution of ambiguous stereoscopic matches by interpolation. *Vision Research*, 27:285–294.

Mumford, D. and Shah, J. (1989). Optimal approximation of piecewise smooth functions and associated variational problems. *Comm. in Pure and Appl. Math.*, 42:577–685.

Parisi, G. (1988). *Statistical Field Theory.* Addison-Wesley, Reading, MA.

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

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

Prazdny, K. (1985). Detection of binocular disparities. *Biological Cybernetics*, 52:93–99.

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

Wasserstrom, E. (1973). Numerical solutions by the continuation method. *SIAM Review*, 15:89–119.

Yuille, A., Geiger, D., and Bülthoff, H. H. (1989). Stereo integration, mean field theory and psychophysics. Harvard Robotics Laboratory Technical Report 89-11, Harvard University.

Yuille, A., Yang, T., and Geiger, D. (1990). Robust statistics, transparency and correspondence. Harvard Robotics Laboratory Technical Report 90-7, Harvard University.

Yuille, A. L. (1989). Energy functions for early vision and analog networks. *Biological Cybernetics*, 61:115–123.

Yuille, A. L. (1990). Generalized deformable models, statistical physics, and matching problems. *Neural Computation*, 2:1–24.

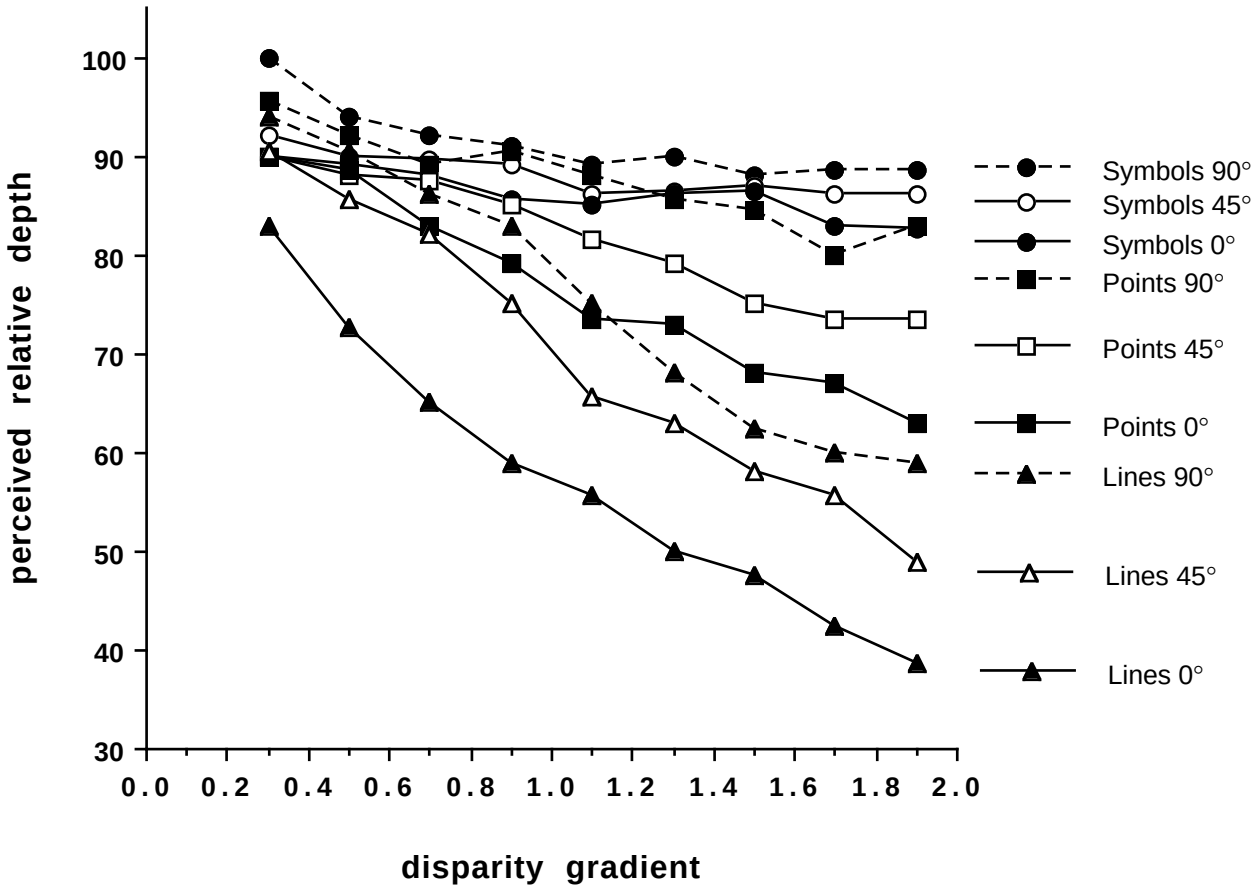
Yuille, A. L. and Grzywacz, N. M. (1989). A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Computation*, 1:334–347.

Figure captions

Figure 1: Disparity Gradient and Matching Primitives: Perceived depth decreases with increasing depth gradient and depends largely on the matching primitive (points, lines or symbols) and orientation (0 deg, 45 deg, 90 deg) relative to the epipolar line. Each data item represents the mean of nine different disparities (3 – 27 arc min) tested with 10 subjects. The standard errors of the means are in the order of the symbol size. Redrawn from Bülthoff and Fahle (1990).

Figure 2: Edge versus Intensity-based Stereo: **(a)** Ellipsoidal surfaces with or without edge information were stereoscopically displayed on a CRT-monitor. The perceived depth map of flat-shaded (edges) and smooth shaded (no edges) ellipsoids of rotation was measured with a local stereo depth probe for different elongations c of the main axis. The main axis was perpendicular to the display screen, i.e., the objects were viewed end-on. **(b)** The coefficient of the first principal component is used as a global measure of perceived elongation for different displayed elongations. Less reliable information (smooth surfaces without edges) puts more weight on surface priors and leads to a fronto-parallel bias of surface perception. Redrawn from Bülthoff and Mallot (1988).

Figure 3: The top 4 figures shows surface interpolation for: (i) dense edges and intensity based stereo, (ii) sparse edges and intensity based stereo, (iii) intensity based stereo, and (iv) sparse edges in stereo and shape-from-shading. The bottom 2 figures illustrate pulling: an edge token in front of intensity patterns with no relative disparity (no intensity based stereo) pulls the surface up (left figure) but can sometimes cause a transparency percept (right figure) of the token lying in front of the intensity surface.


perceived relative depth
100
90
80
70
60
50
40
30
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
disparity gradient
Symbols 90°
Symbols 45°
Symbols 0°
Points 90°
Points 45°
Points 0°
Lines 90°
Lines 45°
Lines 0°