STEREO AND RECONSTRUCTION

Stereo Integration, Mean Field Theory and Psychophysics

A. L. Yuille (D.A.S., Harvard University), D. Geiger (A.I. Lab, M.I.T.)
H. Bülthoff (Dept. Cog. Sci, Brown)

Abstract

We describe a theoretical formulation for stereo in terms of the Markov Random Field and 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 treat the correspondence problem and surface interpolation as different aspects of the same problem and solve them simultaneously, unlike most previous theories. We use techniques from statistical physics to compute properties of our theory and show how it relates to previous work. These techniques also suggest novel algorithms for stereo which are argued to be preferable to standard algorithms on theoretical and experimental grounds. It can be shown (Yuille, Geiger and Bülthoff 1989) that the theory is consistent with some psychophysical experiments which investigate the relative importance of different matching primitives.

1 Introduction

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.

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 between the two images are known. Thus we use the epipolar line constraint for matching, some support for this is given by the work of Bülthoff and Fahle (1989).

Our framework combines cues from different matching primitives to obtain an overall perception of depth. 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, hence they are less reliable than estimates from matching edges. The formalism can also be extended to incorporate information from other depth modules (Bülthoff and Mallot, 1987, 1988) and provide a model for sensor fusion (Clark and Yuille, 1990). This framework was initially described in Yuille and Gennert (1988).

Unlike previous theories of stereo which first solved the correspondence problem and then constructed a surface by interpolation Grimson (1981), our theory proposes combining the two stages. The correspondence problem is solved to give the disparity field which best satisfies the *a priori* constraints.

Our model involves the interaction of several processes and is fairly complex. We will introduce it in three stages at different levels of complexity.

At the first level features (such as edges) are matched, using a binary matching field V_{ia} determining which features correspond. In addition smoothness is imposed on the disparity field $d(\vec{x})$ which represents the depth of the surface from the fixation plane. In this case the correspondence problem, determining the V_{ia}, is solved to give the smoothest possible disparity field. The theory is related to work by Yuille and Grzywacz (1988a, 1988b) on motion measurement and correspondence, and, in particular, to work on long-range motion.

At the second level we add line process fields $l(\vec{x})$ (which represents depth discontinuities) (Geman and Geman, 1984) to break the surfaces where the disparity gradient becomes too high.

The third level introduces additional terms corresponding to matching image intensities. Such terms are used in the theories of Gennert (1987) and Barnard (1986) which, however, do not have line process fields or matching fields. A psychophysical justification for intensity matching is given by the work of Bülthoff and Mallot (1987, 1988). Thus our full theory is expressed in terms of energy functions relating the disparity field $d(\vec{x})$, the matching field V_{ia}, and the line process field $l(\vec{x})$.

By the use of standard techniques from statistical physics, particularly the mean field approximation, we can eliminate certain fields and obtain effective energies for the remaining fields (see Geiger and Girosi, 1989; Geiger and Yuille, 1989). As discussed in Yuille (1989b) (following Wyatt - private communication) this can be interpreted as computing marginal probability distributions. We use this to show (Yuille, Geiger and Bülthoff 1989) that several existing stereo theories, such as the cooperative stereo algorithms (Dev, 1975; Marr and Poggio, 1976) and disparity gradient limit theories (Prazdny 1985; Pollard, Mayhew and Frisby, 1987), are closely related to versions of our model.

These techniques, however, also suggest novel algorithms for stereo computation. We argue that these algorithms incorporate constraints about the set of possible matches better than previous algorithms. They can also be directly related (Yuille 1989b) to analog methods for solving the travelling salesman problem. Moreover the greater empirical success of the elastic net algorithm (Durbin and Willshaw 1987) compared with the Hopfield and Tank method (1985) strongly suggests that our novel stereo algorithms will be more sucessful than some existing algorithms.

Our model can be related (Yuille, Geiger and Bülthoff 1989) to some psychophysical experiments (Bülthoff and Mallot, 1987, 1988; Bülthoff and Fahle, 1989) in which perceived depth for different matching primitives and disparity gradients are precisely measured. Their results suggest that several types of primitive are used for correspondence, but that some primitives are better than others. Our model is in good general agreement with the data from these experiments.

The plan of this paper is as follows: in Section 2 we review the Bayesian approach to vision and describe our theory. Section 3 introduces techniques from statistical physics and uses them to analyse the theory. The paper Yuille, Geiger and Bülthoff (1989) describes this work in more detail, in particular the comparison between theories and the relations to psychophysics.

2 The Bayesian Approach to Stereo

There has been a vast amount of work on stereo. Barnard and Fischler (1982) gives a good survey of the literature. We first review the problem of stereo and give an overview of our theory. Our theory is then described in terms of an energy function and finally put into a probabilistic framework.

2.1 The Matching Problem

There are several choices of matching primitives for stereo. Some theories use features such as edges or peaks in the image intensity (e.g., Marr and Poggio, 1976; Pollard, Mayhew and Frisby, 1987; Prazdny, 1985) while others match the image intensity directly (e.g., Barnard, 1986; Gennert, 1987). Yet another class of theory acts on the Fourier components of the images (e.g., Sanger 1988; Jepson and Jenkin, 1989) and hence is particularly sensitive to texture. It is unclear which primitives the human visual system uses. Current research (Bülthoff and Mallot, 1987, 1988; Bülthoff and Fahle, 1989) 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. But not all primitives are equally reliable, however. A small fluctuation in the image intensity might lead to a large change in the measured disparity for a system which matches intensity. Thus image intensity tends to be less reliable than features such as edges.

Some assumptions about the scene being viewed are usually necessary to solve the correspondence problem. These can be thought of as natural constraints (Marr and Poggio 1976) and are needed because of the ill-posed nature of vision (Poggio and Torre, 1984). There are two types of assumption: (i) assumptions about the matching primitives, i.e., that similar features match (*compatibility constraint*), and (ii) assumptions about the surface being viewed (*continuity constraint*). For (ii) one typically assumes that either the surface is close to the fixation point (disparity is small) or that the surface's orientation is smoothly varying (disparity gradient is small) with possible discontinuities (we discuss possible smoothness measures in Section 2.2).

Our theory requires both assumptions but 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. 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 a number of 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. We will 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 over the range $[0, 1]$. For example, it is 1 if i_L and a_R are compatible (i.e. features of the same type), 0 if they are incompatible (an edge cannot match a peak),

We now define a cost function $E(d(x), V_{i_L a_R})$ 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 $d(x), V_{i_L a_R}$ given the data. A standard estimator is to minimize $E(d(x), V_{i_L a_R})$ with respect to $d(x), V_{i_L a_R}$.

$$E(d(x), V_{i_L a_R}) = \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2$$

$$+\lambda\{\sum_{i_L}(\sum_{a_R} V_{i_L a_R} - 1)^2 + \sum_{a_R}(\sum_{i_L} V_{i_L a_R} - 1)^2\} + \gamma \int_M (Sd)^2 dx. \quad (1)$$

The first term gives a contribution to the disparity obtained from matching i_L to a_R. The third term imposes a smoothness constraint on the disparity field imposed by a smoothness operator S.

The second term encourages features to have a single match, it can be avoided by requiring that each column and row of the matrix $V_{i_L a_R}$ contains only one 1. In Section 3 we will argue that it is better to impose constraints in this way, hence the second term will not be used in our final theory. However we will keep it in our energy function for the present since it will help us 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 smoothest disparity field. 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 hence more smoothing. In Yuille, Geiger and Bülthoff (1989) we show that this gives a possible explanation for some psychophysical experiments.

The theory can be extended to two-dimensions in a straightforward way. The matching elements $V_{i_L a_R}$ must be constrained to only allow 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 smoothness operator S. Marr and Poggio (1976) 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 which 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, hence encouring matches with minimal disparity and thereby giving a bias towards the fixation plane.

Considerations for the choice of smoothness operator (Yuille and Grzywacz 1988a, 1988b) are discussed in Yuille, Geiger and Bülthoff (1989). They emphasize the importance of choosing an operator so that the smoothness interaction decreases with distance. An alternative approach is to introduce discontinuity fields which break the smoothness constraint, see next section. For these theories the experiments described in Yuille, Geiger and Bülthoff (1989) are consistent with S being a first order derivative operator. This is also roughly consistent with Marr and Poggio I (1976). We will therefore use $S = \partial/\partial x$ as a default choice for our theory.

2.3 The Second and Third Level Theories: Adding Discontinuity and Intensity Fields

The first level theory is easy to analyse but makes the *a priori* assumption that the disparity field is smooth everywhere, which is false at object boundaries. The second level theory introduces discontinuity fields $l(x)$ to break the smoothness constraint (Blake 1983, Geman and Geman 1984, Mumford and Shah 1985). The third level theory adds intensity fields. Our energy function becomes

$$E(d(x), V_{i_L a_R}, C) = \sum_{i_L, a_R} A_{i_L a_R} V_{i_L a_R} (d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2$$
$$+\mu \int \{L(x) - R(x + d(x))\}^2 dx$$

$$+\lambda\{\sum_{i_L}(\sum_{a_R} V_{i_L a_R} - 1)^2 + \sum_{a_R}(\sum_{i_L} V_{i_L a_R} - 1)^2\} + \gamma \int_{M-C} (Sd)^2 dx + M(C). \quad (2)$$

If certain terms are set to zero in (3) it reduces to previous theories of stereo. If the second and fourth terms are kept, without allowing discontinuities, it is similar to work by Gennert (1987) and Barnard (1986). If we add the fifth term, and allow discontinuities, we get connections to some work described in Yuille (1989a) (done in collaboration with T. Poggio). The third term will again be removed in the final version of the theory.

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

$$P(d, V, C|g) = \frac{P(g|d, V, C)P(d, V, C)}{P(g)} \quad (3)$$

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, hence 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,

$$\bar{d} = \sum_{d,V,C} dP(d, V, C|g), \quad (4)$$

is more general and reduces to the most probable solution as $T \to 0$. 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 (these can be related to the minimum of the energy function in the zero temperature limit). A justification to use the mean field as a measure of the fields resides in the fact that it represents the minimum variance Bayes estimator (Gelb 1974).

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 (Kirpatrick et al 1983) approach. The drawback of these methods are the amount of computer time needed for the implementation.

There are, however, a number of 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 a number of 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 of the fields (hence 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 constraints on the possible fields by only averaging over fields that satisfy these constraints.

For the first level theory, see Section 3.1.1, 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, Section 3.1.2, we can eliminate the matching fields to obtain an effective energy $E_{eff}(d)$ depending only on the disparity. We believe that the second approach is better since it incorporates the constraints on the set of possible matches implicitly rather than imposing them explicitly in the energy function (as the first method does).

Moreover it can be shown Yuille (1989b) 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 travelling salesman problem by Hopfield and Tank (1985) and Durbin and Willshaw (1987). 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.

We can also average out the line process fields or the matching fields or both for the second and third level theories. This leaves us again with a theory depending only on the disparity field, see Sections 3.1.2, and 3.1.3.

Alternatively we can use (Yuille, Geiger and Bülthoff 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 efective energy depending only on the disparity field. This formulation will depend on a parameter β (which can be interpreted as 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 (Kirkpatrick et al 1983) and has been used by many algorithms, for example (Hopfield and Tank 1985), (Durbin and Willshaw 1987) (Geiger and Girosi 1989). It is also related to continuation methods (Wasserstrom 1973).

3.1 Averaging out Fields

In the next few sections we show that, for the first and second level theories, we can average out fields to obtain equivalent, though apparently different, formulations. As discussed in Yuille (1989b) (following Wyatt - private communication) this can be interpreted as computing marginal probability distributions.

3.1.1 Averaging out the Disparity Field for the first level theory

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. This can be related (Yuille, Geiger and Bülthoff 1989) to cooperative stereo algorithms and it does impose the matching constraints optimally.

The disparity field is eliminated by minimizing and solving for it as a function of the V (Yuille and Grzywacz 1988a,b). 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

$$E(d(x), V_{i_L a_R}) = \sum_{i_L, a_R} V_{i_L a_R}(d(x_{i_L}) - (x_{a_R} - x_{i_L}))^2 + \mu \sum_{i_L}(\sum_{a_R} V_{i_L a_R} - 1)^2 + \mu \sum_{a_R}(\sum_{i_L} V_{i_L a_R} - 1)^2 + \lambda \int_M (Sd)^2 dx. \quad (5)$$

Since the energy function is quadratic in the d's the Euler-Lagrange equations are linear in d. We can (Yuille, Geiger and Bülthoff 1989) solve these equations for the d's as functions of the matching fields and substitute back into the energy function obtaining:

$$E(V_{i_L a_R}) = \lambda \sum_{i_L, j_L}(\sum_{a_R} V_{i_L a_R}(x_{i_L} - x_{a_R}))(\lambda \delta_{i_L j_L} + G(x_{i_L}, x_{j_L}))^{-1}(\sum_{b} V_{j_L b_R}(x_{j_L} - x_{b_R})) + \lambda \sum_{i_L}(\sum_{a_R} V_{i_L a_R} - 1)^2 + \lambda \sum_{a_R}(\sum_{i_L} V_{i_L a_R} - 1)^2 \quad (6)$$

where G is the Green function of the operator S^2. This calculation shows that the disparity field is strictly speaking unnecessary since the theory can be formulated as in (13), the connection to cooperative stereo algorithms is discussed in Yuille, Geiger and Bülthoff (1989). A similar calculation (Yuille and Grzywacz 1988a, 1988b) showed that minimal mapping theory (Ullman 1979) was a special case of the motion coherence theory.

A weakness of this formulation of the theory in (13), and the cooperative stereo algorithms based on it, is that the uniqueness constraints are imposed as penalties in the energy function, by the second and third terms on the right hand side of (13). As mentioned earlier we believe it is preferable to use mean field theory techniques which enforce the constraints strictly, see Section 3.1.2.

3.1.2 Averaging out the matching fields for the first level theory

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, leaving a theory which depends only on the disparity field.

The partition function for the first level system, again assuming compatibility between all features, is defined to be $Z = \sum_{V_{i_L a_R}, d(x)} e^{-\beta E(V_{i_L a_R}, d(x))}$ where the sum is taken over all possible states of the system determined by the fields V and d.

It is possible to explicitly perform the sum over the matching field V yielding an effective energy for the system depending only on the disparity field d. Equivalently we could obtain the marginal probablity distribution $p(d|g)$ from $p(d, V|g)$ by integrating out the V field (Wyatt -personal communication).

To compute the partition function we must first decide what class of $V_{i_L a_R}$ we wish to sum over. We could sum over all possible $V_{i_L a_R}$ and rely on the $\lambda \sum_{i_L}(\sum_{a_R} V_{i_L a_R} - 1)^2 + \lambda \sum_{a_R}(\sum_{i_L} V_{i_L a_R} - 1)^2$ term to bias against multiple matches. Alternatively we could 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 could further restrict the class of possible matches by requiring that they satisfied the ordering constraint.

For this section we will initially restrict that each feature in the left image has a unique match in the right image, but not vice versa. This simplifies the computation of the partition function, but it can be relaxed (Yuille, Geiger and Bülthoff 1989). The requirement of smoothness on the disparity field should ensure that unique matches occur, this is suggested by mathematical analysis of a similar algorithm used for an elastic network approach to the T.S.P. (Durbin, Szeliski and Yuille 1989).

Since we are attempting to impose the unique matching constraint by restricting the class of V's 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 do not need to be included in the energy function. We can now write the partition function as

$$Z = \sum_{V,d} \prod_{i_L} e^{-\beta \{\sum_{a_R} V_{i_L a_R} \left(d(x_{i_L}) - (x_{a_R} - x_{i_L})\right)^2 + \int_M (Sd)^2 dx\}}. \tag{7}$$

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

$$Z = \sum_{d} \prod_{i_L} \{\sum_{a_R} e^{-\beta \left(d(x_{i_L}) - (x_{a_R} - x_{i_L})\right)^2}\} e^{-\beta \int_M (Sd)^2 dx}. \tag{8}$$

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

$$Z = \sum_{d} e^{-\beta E_{eff}(d)}, \text{ where } E_{eff}(d) = \frac{-1}{\beta} \sum_{i_L} log\{\sum_{a_R} e^{-\beta \left(d(x_{i_L}) - (x_{a_R} - x_{i_L})\right)^2}\} + \int_M (Sd)^2 dx. \tag{9}$$

Thus our first level theory of stereo can be formulated in this way without explicitly using a matching field. We are not aware, however, of any existing stereo theory of this form. Since it has formulated the matching constraints in computing the partition function we believe it is preferable to standard cooperative stereo algorithms.

3.1.3 Averaging out the matching and discontinuity fields for the third level theory

We can apply the same techniques to the second and third theories eliminationg both the matching field and the discontinuity fields (Yuille, Geiger and Bülthoff 1989). This gives an effective energy for the third level theory:

$$E_{eff}(d) = \frac{-1}{\beta} \sum_{i_L} log\{\sum_{a_R} e^{-\beta \left(d(x_{i_L}) - (x_{a_R} - x_{i_L})\right)^2}\} \frac{-1}{\beta} \sum_{a_R} log\{\sum_{i_L} e^{-\beta \left(d(x_{a_R}) - (x_{a_R} - x_{i_L})\right)^2}\}$$
$$-\frac{1}{\beta} ln(1 + e^{-\beta[\alpha(d_k - d_{k-1})^2 - \gamma]}) + \mu \int \{L(x) - R(x + d(x))\}^2 dx. \tag{10}$$

Again a deterministic annealing approach should yield good solutions to this problem.

Averaging out the discontinuity field from the second level theory will give (Yuille, Geiger and Bülthoff 1989) a theory, depending only on the disparity field and the matching field, which is reminiscent of disparity gradient limit theories (Pollard, Mayhew and Frisby, 1987; Prazdny, 1985).

The mean field theory approach can also yield deterministic algorithms for theories including the binary matching elements (although we believe these algorithms will be inferior to methods which eliminate the matching fields for the reasons discussed in Section 3.1.2) which should be superior to the cooperative stereo algorithms (the cooperative algorithms are the zero temperature limit of these equations, and hence are less able to escape local minima).

4 Conclusion

We have derived a theory of stereo on theoretical grounds using the Bayesian approach to vision. This theory is able to incorporate most of the desirable elements of stereo and it is closely related to a number of 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 for stereo correspondence (Clark and Yuille. 1990).

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

The use of mean field theory enables us to average out fields, enabling us to make mathematical connections between different formulations of stereo. It also suggests novel algorithms for computing the estimators (due to enforcing the matching constraints while performing the averaging, see Section 3.1.2) and we argue that these algorithms are likely to be more effective than a number of existing algorithms.

Finally the theory agrees well with some psychophysical experiments (Bülthoff and Mallot, 1987, 1988; Bülthoff and Fahle, 1989). Though further experiments to investigate the importance of different stereo cues are needed.

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. Some of these ideas were initially developed with Mike Gennert. We would like to thank Mike Gennert, Manfred Fahle, Jim Clark, and Norberto Grzywacz for helpful conversations.

References

Barnard, S. *Proc. Image Understanding Workshop*, Los Angeles, 1986.

Barnard, S. and Fischler, M.A, "Computational Stereo". *Computing Surveys*, **14**, No. 4, 1982.

Blake, A. "The least disturbance principle and weak constraints," *Pattern Recognition Letters*, **1**, 393-399, 1983.

Blake, A. "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction," *PAMI*, Vol. 11, No. 1, 2-12, 1989.

Bülthoff, H. and Mallot, H-P. "Interactions of different modules in depth perception". In *Proceedings of the First International Conference on Computer Vision*, London, 1987.

Bülthoff, H. and Mallot, H-P. "Integration of depth modules: stereo and shading". *J. Opt. Soc. Am.*, **5**, 1749-1758, 1988.

Bülthoff, H. and Fahle, M. "Disparity Gradients and Depth Scaling". *Artificial Intelligence Memo 1175*, Cambridge, M.I.T., 1989.

Burt, P. and Julesz, B. "A disparity gradient limit for binocular fusion". *Science* **208**, 615-617, 1980.

Clark, J.J. and Yuille, A.L. **Data Fusion for Sensory Information Processing Systems.**, Kluwer Academic Press, 1990.

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

Durbin, R., Szeliski, R. and Yuille, A.L. "The elastic net and the travelling salesman problem". Harvard Robotics Laboratory Technical Report. No. 89-3, 1989.

Durbin, R. and Willshaw, D. "An analog approach to the travelling salesman problem using an elastic net method". *Nature*, **326**, 689-691, 1987.

Duchon, J. *Lecture Notes in Mathematics.* 571. (Eds Schempp, W. and Zeller, K.), 85-100 (Berlin, Springer-Verlag, 1979).

Geiger, D. and Girosi, F., "Parallel and deterministic algorithms from MRFs :integration and surface reconstruction". *Artificial Intelligence Laboratory Memo 1114*. Cambridge, M.I.T., June 1989.

Geiger, D. and Yuille, A., "A common framework for image segmentation". Harvard Robotics Laboratory Technical Report. No. 89-7, 1989.

Geiger, D. and Yuille, A., "Stereopsis and eye movement", *Proceedings of the First International Conference on Computer Vision.* London, pp 306-314, 1987.

Gelb, A. **Applied Optimal Estimation.** M.I.T. Press. Cambridge, Ma., 1974.

Geman, S. and Geman, D. "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images". *IEEE Trans. PAMI*, **6**, 721-741, 1984.

Gennert, M. "A Computational Framework for Understanding Problems in Stereo Vision". M.I.T. AI Lab PhD. Thesis, 1987.

Grimson, W.E.L. **From Images to Surfaces: A computational study of the human early visual system.** M.I.T. Press. Cambridge, Ma., 1981.

Hopfield, J.J. and Tank, D.W. "Neural computation of decisions in optimization problems". *Biological Cybernetics*, **52**, 141-152, 1985.

Jepson, A.D. and Jenkin, M.R.M. "The fast computation of disparity from phase differences". *Proceedings Computer Vision and Pattern Recognition '89.* pp 398-403, San Diego, 1989.

Kirkpatrick, S., Gelatt, C.D. Jr. and Vecchi, M.P. "Optimization by simulated annealing". *Science*, **220**, 671-680, 1983.

Marr, D. and Poggio, T. "Cooperative computation of stereo disparity". *Science*, **194**, 283-287, 1976.

Marr, D. and Poggio, T. "A computational theory of human stereo vision". *Proc. R. Soc. Lond. B.* Vol 204, pp 301-328, 1979.

Marroquin, J. In *Proceedings of the First International Conference on Computer Vision.* London. 1987.

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

Mitchison, G.M. "Planarity and segmentation in stereoscopic matching. *Perception*, **17**, 753-782, 1988.

Mitchison, G.M. and McKee, S. "The resolution of ambiguous stereoscopic matches by interpolation". *Vision Research*, Vol 27, no 2. pp 285-294, 1987.

Mumford, D. and Shah, J. "Boundary detection by minimizing functionals, I", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, 1985.

Parisi, G. **Statistical Field Theory**. Addison-Wesley, Reading, Mass. 1988.

Pollard, S.B., Mayhew, J.E.W. and Frisby, J.P. "Disparity Gradients and Stereo Correspondences". *Perception*, 1987.

Poggio, T. and Torre, V. "Ill-posed problems and regularization analysis in early vision". M.I.T. A.I. Memo No. 773, 1984.

Prazdny, K. "Detection of Binocular Disparities". *Biological Cybernetics*, **52**, 93-99, 1985.

Sanger, T. "Stereo disparity computation using Gabor filters," *Biological Cybernetics*, **59**, 405-418, 1988.

Ullman, S. **The Interpretation of Visual Motion**. Cambridge, Ma. M.I.T. Press, 1979.

Wasserstrom, E. "Numerical solutions by the continuation method". *SIAM Review*,,**15**, 89-119, 1973.

Yuille, A.L. "Energy Functions for Early Vision and Analog Networks". *Biological Cybernetics*, **61**, 115-123, 1989a.

Yuille, A.L. Harvard Robotics Laboratory Technical Report 89-12. 1989b.

Yuille, A.L., Geiger, D. and Bülthoff, H. "Stereo Integration, Mean Field Theory and Psychophysics". Harvard Robotics Laboratory Technical Report 89-11.

Yuille, A.L. and Gennert, M. Preprint. 1988.

Yuille, A.L. and Grzywacz, N.M. "A Computational Theory for the Perception of Coherent Visual Motion". Nature, 1988a.

Yuille, A.L. and Grzywacz, N.M. "The Motion Coherence Theory". *Proceedings of the Second International Conference on Computer Vision.* pp 344-353. Tampa, Florida. 1988b.