

NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:
The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

Cooperative algorithms for solving random-dot stereograms

Richard Szeliski

Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213

June 1986

Abstract

This report examines a number of parallel algorithms for solving random-dot stereograms. A new class of algorithms based on the Boltzmann Machine is introduced and compared to previously developed algorithms. The report includes a review of the stereo correspondence problem and of cooperative techniques for solving this problem. The use of energy functions for characterizing the computational problem, and the use of stochastic optimization techniques for solving the problem are explained.

Copyright © 1986 Richard Szeliski

This research was sponsored in part by grant number IST8520359/8520359 from the National Science Foundation.

Table of Contents

1. Introduction
 - 1.1. Motivation
 - 1.2. Previous work
 - 1.3. Organization
2. The stereo correspondence problem
 - 2.1. Computational theory
 - 2.1.1. Image formation
 - 2.1.2. Compatibility
 - 2.1.3. Uniqueness
 - 2.1.4. Continuity
 - 2.2. Representation
 - 2.3. Algorithms
3. The Boltzmann Machine
 - 3.1. Constraint satisfaction networks
 - 3.2. Energy minimization
 - 3.3. Search
 - 3.4. The Boltzmann Machine
 - 3.5. Learning
4. Simple networks
 - 4.1. Hand-wired networks
 - 4.2. Derived networks
 - 4.3. Multi-level units
5. More complex models
 - 5.1. Line process
 - 5.2. Separate correspondence and interpolation
 - 5.3. Coarse coding
 - 5.4. Real-valued units
6. Comparison of algorithms
 - 6.1. Simple networks with annealing (Boltzmann Machine)
 - 6.2. A Cooperative Stereo Algorithm (Marr-Poggio)
 - 6.3. Winner-take-all (Marroquin)
 - 6.4. Coherence (Prazdny)
7. Conclusions
- References

List of Figures

Figure 2-1:	Projective geometry	4
Figure 2-2:	Geometry of random-dot stereogram formation	4
Figure 2-3:	A random-dot stereogram pair	4
Figure 2-4:	Strip depth representation and display	5
Figure 2-5:	A noisy random-dot stereogram pair	8
Figure 3-1:	Binary threshold unit	12
Figure 3-2:	Sigmoid function	14
Figure 4-1:	Network derived from image formation model	19
Figure 4-2:	Plots of excitation and inhibition from dense compatibilities	20
Figure 4-3:	Plots of excitation and inhibition from sparse compatibilities	21
Figure 6-1:	Sample stereograms	26
Figure 6-2:	Associated compatibilities	27
Figure 6-3:	Simulated annealing results	29
Figure 6-4:	Marr-Poggio algorithm results	30
Figure 6-5:	Winner-take-all results	31
Figure 6-6:	Coherence principle results	31

1. Introduction

This report examines a number of parallel algorithms for solving random-dot stereograms. A new class of algorithms based on the Boltzmann Machine is introduced and compared to previously developed algorithms. The report includes a review of the stereo correspondence problem and of cooperative techniques for solving this problem. The use of energy functions for characterizing the computational problem, and the use of stochastic optimization techniques for solving the problem are explained.

1.1. Motivation

Stereo vision usually operates on real images acquired either through the retina or through photographic/electronic means. It is also possible to obtain depth perception from pairs of images containing seemingly random and uncorrelated dots. If the dots are replicated in both images with some areas being shifted horizontally, they can be matched to obtain a *correspondence*, from which depth values can be derived.

The restricted domain of random-dot stereograms is worth studying for several reasons. For researchers trying to understand human visual perception, it is interesting because the presence of a depth effect in the absence of any obvious monocular cues indicates that stereo fusion is possible without the necessity of object recognition or segmentation¹. For researchers studying the computational side of stereo, it is interesting because there exist a large number of possible matches without any external source of information to disambiguate them. This makes the correspondence problem more difficult, but allows the process to be studied using *tokens*, without making a commitment to any particular primitive feature to be used for matching.

The second major component of the work described in this report deals with cooperative² algorithms. These algorithms are also known as "relaxation", "connectionist" or "local" algorithms or architectures. One attraction of these systems is that very high performance can be obtained through massive parallelism, i.e. the replication of relatively slow/simple processors with only local interconnections. Another attraction is that this architecture maps more closely to the neural machinery found in the brain than does a serial machine.

Within the class of cooperative algorithms, we are most interested in those whose behaviour can be

¹However, band-pass filtered versions of the stereograms will exhibit low-frequency "blobs".

²Marr [Marr 82] defines cooperative methods as "a type of nonlinear, iterative algorithm" and later as "local operations [that] appear to cooperate in forming global order".

characterized by the use of energy functions. This allows the computation (a functional minimization) to be specified and analyzed cleanly. The use of energies can also be related to probabilistic models through the use of "exponential families", in which the log probability of a global configuration is proportional to the negative of its energy. The stochastic algorithms (based on the Boltzmann Machine) that we will examine can find globally optimal solutions to problems that cannot be solved with deterministic gradient descent techniques.

1.2. Previous work

The first investigations into random-dot stereograms began with their invention by Julesz [Julesz 60]. In addition to generating the stereograms, Julesz also proposed a simple "spring" model that could solve the correspondence problem [Julesz 71]. More recent algorithms for solving these stereograms include [Sperling 70, Dev 74, Marr 76, Marroquin 83, Prazdny 85, Mayhew 80, Mayhew 85, Szeliski 85]. Research into recovering depth information from real stereo image pairs also has a long history. More recent examples of such stereo algorithms include [Marr 79, Grimson 80, Mayhew 81, Baker 82, Arnold 83, Kass 84, Ohta 85, Yuille 84]. These stereo algorithms are discussed in more detail in Section 2.

The use of cooperative (relaxation) algorithms to vision problems began with Waltz filtering [Waltz 75] and the development of relaxation labeling techniques [Rosenfeld 76, Hinton 77, Hummel 83]. The development of the Boltzmann Machine algorithm [Hinton 84] was a combination of ideas from neural networks [Hopfield 82] and stochastic optimization techniques [Metropolis 53, Kirkpatrick 83]. Other examples of cooperative algorithms include surface interpolation [Terzopoulos 84a] and the use of regularization techniques [Poggio 84]. These cooperative algorithms are discussed in more detail in Section 3.

1.3. Organization

This report is organized as follows. Section 2 formulates the stereo correspondence problem and analyses it in terms of three aspects: computational theory, representation and algorithms. Section 3 introduces the Boltzmann Machine as well as related cooperative algorithms, computational networks and optimization techniques. Section 4 gives some examples of the new algorithms based on the Boltzmann Machine that can be used to solve the correspondence problem. Section 5 discusses more complex models that could be used, but that have not yet been implemented. Section 6 compares the new stochastic cooperative algorithm with three previously developed algorithms. Section 7 concludes by discussing how the new techniques offer better performance, and what problems remain to be investigated.

2. The stereo correspondence problem

This section discusses in more detail the stereo correspondence problem, using the paradigm suggested by Marr [Marr 82]. First, the *computational theory* is described, i.e. how stereo images are generated, and what is involved in solving the correspondence problem. Second, various alternative *representations* are discussed. Third, possible *algorithms* for the solution are introduced. While this section surveys the range of possible representations and algorithms, the detailed description of particular algorithms is left to subsequent sections.

2.1. Computational theory

An examination of the computational theory of stereoscopic depth reconstruction begins with the image formation process. From the resulting "stereo" pair of images, the set of candidate correspondences can then be calculated. The selection of the "correct" correspondences is then formulated as the computational problem to be solved. This task involves finding solutions that satisfy three constraints: compatibility, uniqueness and continuity. Each of these is discussed in turn.

2.1.1. Image formation

In the general case of stereo depth perception (e.g. normal human vision, stereoscopic slide viewers, or aerial photograph interpretation), the left and right images are formed by the usual process of perspective projection and photometric optics. The geometric aspect (projective geometry) determines how three dimensional points in the world map into the two dimensional left and right image planes (Figure 2-1). The photometric optics dictate how the effects of lighting, surface properties and viewing angle interact to produce a particular intensity at a point on the image plane. In the general problem of vision, all of these factors (geometry, lighting, surface properties) must be extracted from the perceived images.

Random-dot stereograms provide a much more limited domain than general vision. A very simple geometric arrangement, with orthographic projection and epipolar scanline geometry is used (Figure 2-2). As well, surfaces in the world are assumed to be painted a uniform white, with black dots distributed at random over the surfaces. The surfaces are fronto-parallel, and the dot positions and size are quantized to a uniform grid. The resulting image pair is devoid of any interesting characteristics when viewed monocularly (Figure 2-3).

The actual method of stereogram formation used in this report is as follows. The original depth data is assumed to be quantized to 7 depth levels (-3 to +3). Figure 2-4(a) shows the original depth "stack" that has been unfolded by laying the depth layers side by side (the leftmost plane is the nearest). To create the random-dot stereogram, we start with the deepest (rightmost) plane, and

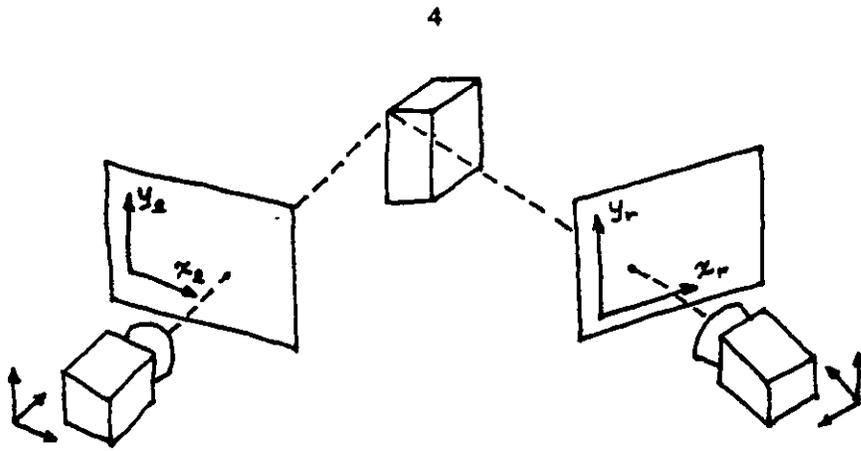


Figure 2-1: Projective geometry

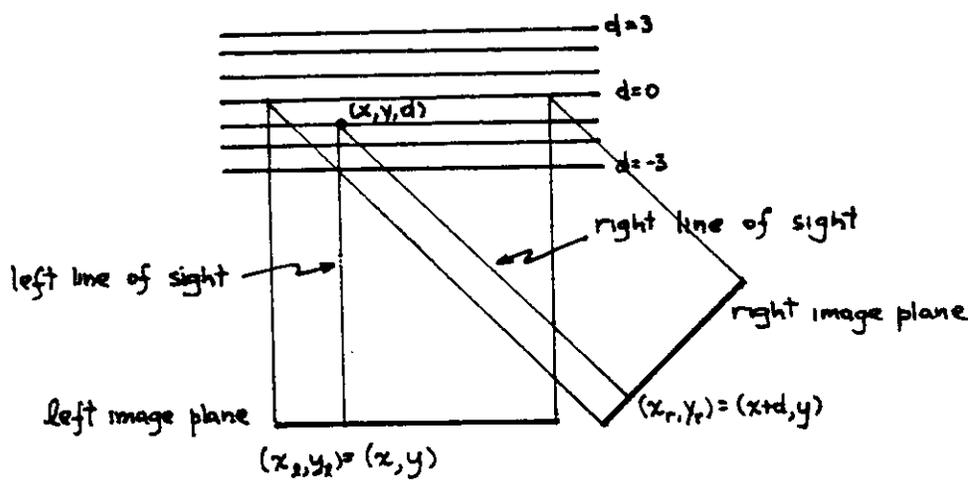


Figure 2-2: Geometry of random-dot stereogram formation

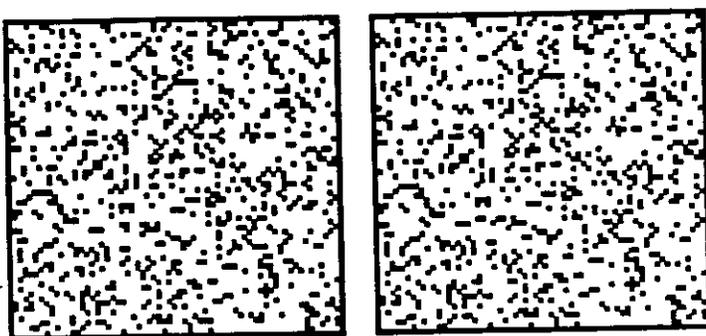


Figure 2-3: A random-dot stereogram pair

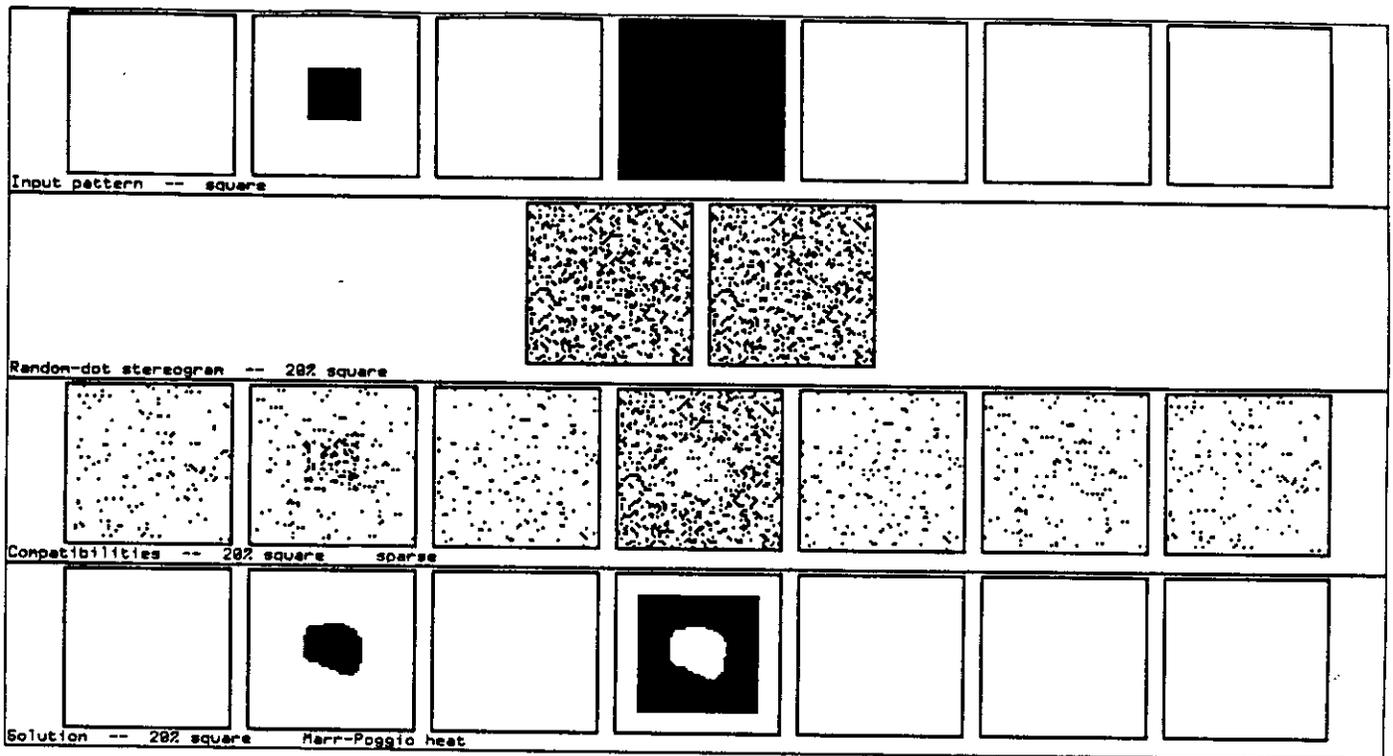


Figure 2-4: Strip depth representation and display
 a) original surface b) stereogram c) compatibilities d) solution

paint each surface point (i.e. where there exists a surface patch) either white or black at random. The dot density (probability of generating a black dot) is a user-controlled parameter. Each (x,y,d) point maps to (x,y) in the left image and $(x+d,y)$ in the right. If opaque surfaces are being modeled, then both white and black points overpaint the left and right images. If transparent surfaces are used, then only black points get painted (the background is initially white). The stereogram corresponding to Figure 2-4(a) using an opaque surface model is shown in Figure 2-4(b).

The recovery of depth information from a pair of stereo images can be thought of as the inversion of this image formation process [Horn 77, Marr 82]. Three constraints which are useful for solving this problem (compatibility, uniqueness and continuity) are examined next.

2.1.2. Compatibility

The compatibility constraint can be expressed as "points (features) in one image should match like points in the other image". This constraint can be expressed as an array of compatibilities³, where each compatibility can be thought of as the answer to the question "can point l from the left image match point r from the right image?", i.e. could they arise from the same surface point? Any such matching (correspondence) implies that a piece of surface exists at some known spatial location (which can be determined by triangulation if the camera model is known). The determination of which matches are true and which correspond to *false targets* is the stereo correspondence problem.

For stereo matching on real images, there are many possible choices for compatibility measures. Algorithms that try to match each pixel directly are called *correlation based*, while those which first extract features and then match them are called *feature based*. Within this latter class, commonly used compatibility measures include intensity data [Ohta 85], zero-crossing in the band-pass images [Marr 79, Grimson 80], intensity derivatives [Kass 84], edge orientations [Arnold 83] or edge intensity profiles [Baker 82]. These compatibilities can have associated confidence (or cost) values derived from the intensity data.

In our restricted random-dot stereogram domain, points can match only if they are on the same scanline ($y_l = y_r$), are within 3 pixels of each other ($|x_l - x_r| \leq 3$ since there are only 7 disparity levels), and have the same intensity. Some researchers working in solving random-dot stereograms consider matches only on black-to-white or white-to-black transitions (edges) [Prazdny 85]. For the purpose of this research however, only two possible simple matching rules are considered. Compatibilities can be defined either as matches between two points of the same intensity (black-black or white-white) or

³The term compatibility used here (first introduced in [Marr 76]) should not be confused with the compatibilities used in relaxation labeling [Hummel 83].

only between two black points. The resulting data in the former case will be called the *dense* compatibility map, while the latter will be called the *sparse* compatibility map. Note that for low-density stereograms with sparse compatibilities, and for high-density stereograms with dense compatibilities, the maps using intensity based matches are similar to the edge based compatibilities.

The compatibility maps (sparse and dense) arising from the random-dot stereogram previously introduced are shown in Figure 2-4(c) and (d). The compatibilities are calculated as:

$$c[x,y,d] = \text{match}(\text{left}[x,y], \text{right}[x+d,y]) \quad (1)$$

These compatibilities (and not the stereo pair) are what form the input to the correspondence algorithm.

2.1.3. Uniqueness

The uniqueness constraint can be paraphrased as "almost always, a feature in one image matches exactly one feature in the other image". In early work on the computer solution of random-dot stereograms [Julesz 71, Sperling 70, Dev 74], the uniqueness constraint was applied along the depth axis, in effect requiring that each (x,y) location have a unique depth value. Marr and Poggio [Marr 76] pointed out that the uniqueness constraint should operate along both lines of sight (see [Marr 82] for a more detailed discussion).

For the case of random-dot stereograms, the uniqueness constraint implies that there should only be one point in a given (x,y,d) column and only one in a given $(x+d,y,d)$ column, where d is the free variable. Because of occlusion, this constraint cannot always be met. There are two possible types of violations: either some points can have no matches, or some points can have two matches. Thus, depending on which types of violations are permitted, we have two approaches to implementing the uniqueness constraint.

The first approach, allowing some points to have no matches, is consistent with what happens during occlusion, i.e. when some surface points are visible from one eye only (assuming opaque surfaces). Note that if the image data is noisy (i.e. some features may be detected only in one image), then this ability to skip matches is necessary. Stereo algorithms that work with real images [Ohta 85] commonly include this ability in the matching process.

The second approach, allowing some points to have more than one match, is more in keeping with the results observed when humans obtain fusion in a random-dot stereogram. Figure 2-5 shows a stereogram to which noise has been added. Observe that when fused, the noise points seem to "float", i.e. they match the nearest possible points, rather than becoming invisible. This may have to

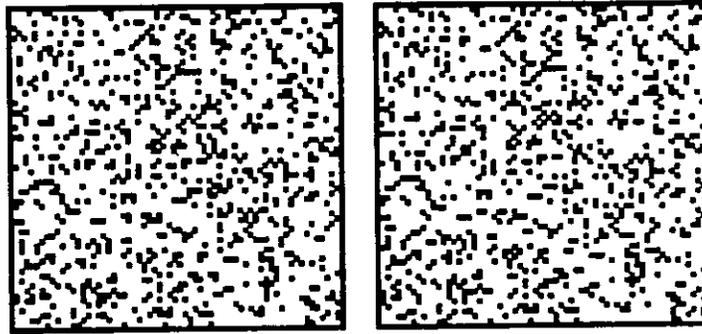


Figure 2-5: A noisy random-dot stereogram pair

do with the sharpness of the point features, and indicates a tendency to interpret these stereograms as having "transparent" qualities when necessary.

The uniqueness constraint causes the stereo correspondence problem to have many locally optimal solutions (since the inclusion or exclusion of a match in the solution will conflict with other matches). Thus, some kind of search is required to find the correct correspondence. The implications of this search requirement are discussed in Section 3.3.

2.1.4. Continuity

The third constraint that applies to the correspondence problem is continuity, i.e. "neighboring points have similar depth values". The most straightforward implementation of this constraint occurs in relaxation schemes (e.g. [Marr 76]) where neighboring match hypotheses at similar depths reinforce each other. This is the approach taken in the algorithms examined in this report. Other interpretations of the continuity constraint include "figural continuity" [Mayhew 81] and "inter-scanline matching" [Ohta 85]. If the features in the image are sufficiently distinct, matching can be performed independently on each epipolar line using no continuity assumptions (e.g. the 2-D version of [Ohta 85]). The continuity constraint for stereo matching is related to the smoothness constraint used for surface interpolation [Poggio 84].

As has been pointed out by Prazdny [Prazdny 85], the continuity constraint is valid only for opaque surfaces, and away from depth discontinuities. The alternative that he has suggested is the *coherence principle*: points must lie on coherent (smooth) surfaces, but neighboring points are not required to lie on the same surface. This principle has been the basis for several stereo matching algorithms, using both global support functions [Prazdny 85, Mayhew 85] and local relaxation [Szeliski 85].

Another constraint related to continuity is the *ordering constraint*, which deals with the presence of

"reversals" between the two images. The image formation geometry is used to derive consistency rules for matches (assuming opaque surfaces). Stereo matching algorithms that use the dynamic programming approach [Baker 82, Arnold 83, Ohta 85] generally have some simple ordering constraint implicitly built in. A more comprehensive treatment of ordering constraints can be found in [Yuille 84].

When implementing the continuity constraint, a distinction must be made between those algorithms that explicitly interpolate depth values so that a *dense* depth map is obtained, and those where depth values are only represented at the matches (*sparse* depth map). This decision, which should properly be discussed in the representation section, influences the computational definition of continuity. Most methods that use relaxation schemes (e.g. [Marr 76] and the new algorithm presented in this report) use the dense map, and thus define continuity to be a property of the depth surface (points must have similar depth to their immediate spatial neighbors). Schemes that are based on dynamic programming (e.g. [Ohta 85]) or on support functions (e.g. [Prazdny 85]) use the sparse depth map, and define continuity to be agreement with some other matches, which may be spatially far removed.

2.2. Representation

The choice of representations for solving the correspondence problem can be important. Specifying the representation includes both choosing the frame of reference and the method of encoding the data within that frame.

For the stereo correspondence task, the frame of reference usually chosen is *retinotopic*, i.e. aligned with the vertical, horizontal and depth coordinates of the imaging system (or eye). The representation of depth and other surface information in a retinotopic⁴ coordinate system is the basis of Marr's $2\frac{1}{2}$ -D sketch [Marr 82]. The mapping from this representation to *egocentric* or *object centered* reference frames is left to higher-order processes.

Within the retinocentric framework, a number of methods can be used for encoding the disparity (depth) information. Those matching techniques that produce only a sparse set of matches (e.g. dynamic programming) have an implicit representation that defines a sparse set of disparity values. For cooperative algorithms, however, the method of encoding is of greater significance. The three encoding methods that we will examine here are: rate encoding, non-overlapping interval encoding, and overlapping ("coarse") encoding.

⁴Marr also uses the term "viewer-centered".

The first type of depth representation, rate encoding, is the simplest. In this representation, there is a single real value for representing the depth at each (x,y) point. A neural analog of this would be if the firing rate of a neuron were related to the depth at that point. This representation is the one commonly used for surface interpolation algorithms (e.g. [Terzopoulos 84a]). Its drawbacks are: (a) it only works for opaque surfaces (because it cannot encode several depths at one point), (b) it makes implementation of the true (double column) uniqueness difficult, and (c) it can't represent multiple (competing) hypotheses about the depth value during the search. Its advantage is that it is easy to manipulate, and convergence results can be proved in simple cases (e.g. for smooth surface interpolation).

The second representation, interval encoding, uses a different unit to represent each of several possible discrete depth hypotheses at a given location. The value (level) of each unit is then free to represent the confidence in the given depth hypothesis. The neural analog of this is when each neuron represents a different (x,y,d) location. This is the standard representation used in cooperative stereo matching algorithms. The ability to represent multiple hypotheses also means that it is more difficult to ensure that the system converges to a correct answer (or that it converges in a reasonable time).

The third representation is distributed (or "coarse") encoding. Here, several units with overlapping responses are used to represent a depth hypothesis. The pattern of activity of the units encodes the actual depth value. The neural analog of this method is similar to the overlapping response characteristics of cortical neurons. The advantage of this method is that the convergence properties of the collection of units can be improved (in terms of the energy formalism introduced in 3.2, this helps the system get over "energy barriers"). Distributed encoding includes the possibility of using multiple-resolution representations to improve the performance of the system.

Having chosen the method of disparity encoding, there are still several choices for the representation of individual hypotheses. Three such possibilities are: deterministic binary, analog, or stochastic. In a binary representation, each unit is either on or off (0/1). One example of such units is the binary threshold units used in the Marr-Poggio cooperative algorithm [Marr 76]. In an analog representation, each unit takes on a real value, which can be related to the value being encoded (activation level), or the strength of the hypothesis. It is possible for such analog networks to have behaviors that are in the limit binary [Hopfield 84]. Finally, the value of a unit can be encoded by a probabilistic distribution over a range of discrete values. This method is used in the Boltzmann Machine (see Section 3.4).

2.3. Algorithms

A large number of algorithms for stereo matching have been proposed in the past. Many of these algorithms have already been mentioned in this chapter in the context of the computational theory (i.e. *what* is to be solved). What further distinguishes various algorithms is in *how* a solution is found that [optimally] satisfies the constraints expressed in the computational theory.

For *feature based* stereo, the most popular approaches for finding a good set of correspondences are based on the ideas of local relaxation, dynamic programming and/or multiple-resolution processing. Local relaxation (cooperative) algorithms use a high degree of parallelism combined with low connectivity and iteration to arrive at solutions that are locally consistent (or locally optimal). Dynamic programming techniques use a serial search to find the globally optimal solution, but generally do not have parallel implementations (but see [Guerra 83] for an example of limited parallelism). Multiple-resolution techniques first perform the matching on small low-resolution images, and then use these results to guide the matching on larger high-resolution images.

While cooperative algorithms usually arrive at solutions that are only locally optimal, it is possible to obtain globally optimal solutions by using stochastic (probabilistic) algorithms. These algorithms are discussed in the next section.

3. The Boltzmann Machine

This section describes the Boltzmann Machine, the parallel computation network that is the basis for the new stereo matching algorithm presented in this report. The Boltzmann Machine evolved from two ideas: constraint satisfaction networks (in particular Hopfield's "neural networks" with *asynchronous* updating), and the use of energy minimization as a characterization of the computational task (including Kirkpatrick's *simulated annealing* optimization techniques). These are presented first, followed by a description of the Boltzmann Machine. The learning algorithm used with the Boltzmann Machine is also mentioned.

3.1. Constraint satisfaction networks

Constraint satisfaction networks have long been used in vision to arrive at interpretations that not only agree with the data, but that are also self-consistent. Early work concentrated on finding all possible labelings that were consistent under some fixed rules (*strong* constraints). It was found that efficient cooperative algorithms were possible with *relaxation* (iterative methods) where constraints are propagated in a local fashion [Waltz 75]. Relaxation labeling techniques were developed where confidence (plausibility) values were associated with each possible label at a node. Updating rules

based on the compatibilities between labels force the system into locally consistent interpretations [Hummel 83].

A separate class of cooperative algorithms has been studied by researchers interested in neural modeling. The Boltzmann Machine is based on the work done by Hopfield in neural networks [Hopfield 82]. In such a network, each unit is a "binary threshold unit" (Figure 3-1) that takes of a value of 1 or 0, depending on whether the total input from the other units exceeds a threshold⁵ (Figure 3-1).

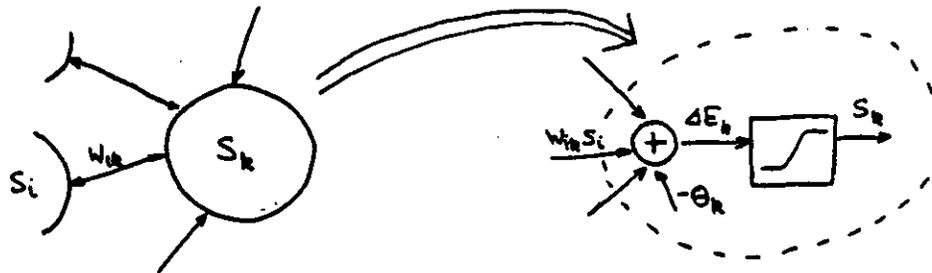


Figure 3-1: Binary threshold unit

The total input is simply a linear sum of the states of the neighboring nodes multiplied by the weights on the arcs

$$\Delta E_k = \sum_i w_{ki} s_i - \theta_k \quad (2)$$

This formulation is identical to that used in the Marr-Poggio cooperative stereo algorithm [Marr 76]. Where the two algorithms differ is in the sequencing of the updates. The Marr-Poggio algorithm uses *synchronous* updating, i.e. all units change state simultaneously. In contrast, Hopfield's nets use *asynchronous* updating, i.e. only one unit changes its value at a time. When asynchronous updating is used in conjunction with *symmetric* weights ($w_{ij} = w_{ji}$), the system settles into a configuration that is a local minimum of the energy function

$$E_\alpha = - \sum_{k,j} w_{kj} s_k^\alpha s_j^\alpha + \sum_i \theta_i s_i^\alpha \quad (3)$$

While Hopfield nets were originally designed as associative memories (where each local minimum encodes a datum), they can obviously be used for cooperative stereo algorithms (being a variant of the Marr-Poggio algorithm).

⁵Hopfield uses values of +1 and -1, but his systems are equivalent to 1,0 systems with suitably adjusted thresholds.

3.2. Energy minimization

The ability to characterize the behaviour of a cooperative network in terms of an energy function may not seem at first to have much advantage. There are, however, several good reasons why this energy formulation is very useful:

- It allows the definition of the computational problem to be separated from any particular implementation (e.g. whether dynamic programming search or local relaxation is used to find the minimum).
- Various tradeoffs between desired constraints can be expressed as sums of potential (cost) functions. This approach is known as *weak constraint satisfaction*.
- The energy of a given state (or configuration) can be related to its probability through the use of exponential families [Geman 84] (the technical details are discussed below). This allows the overall probability distribution to be built up by specifying the global energy as a sum of local energy contributions, without requiring that the local probabilities be independent.
- Cooperative algorithms can be found that compute the globally optimal solution. When the problem is convex, it has only one minimum, and gradient descent techniques can be used. Examples of such systems include the finite-element solutions of some variational problems [Terzopoulos 84a]. When many local minima exist, the stochastic gradient techniques discussed below can be used.

3.3. Search

Systems whose energy functions have many local minima are said to have *frustration*. Systems which have multiple optimal solutions are typically underconstrained and are said to be *ill-posed*. One approach to solving such systems is to use *regularization* (additional smoothness constraints) to make the system convex [Poggio 84]. The alternative is to actually search for the globally minimum solution.

When a cooperative algorithm is desired, adding noise to the system (using a stochastic updating rule) will allow it to escape local minima. If the updating rule is based on the Boltzmann distribution, then the system can be analysed by the use of statistical mechanics [Metropolis 53]. The *steady-state* probability of a given system state is then related to its energy by the Boltzmann distribution

$$p_{\alpha} = \frac{1}{Z} e^{-E_{\alpha}/T}, \text{ where } Z = \sum_{\beta} e^{-E_{\beta}/T} \quad (4)$$

The probability of being in the lowest energy state at equilibrium can thus be made arbitrarily high by decreasing the value of the parameter T , which is usually called "temperature".

While the steady-state distribution is thus readily characterized, these distributions tell us nothing about how long it will take to achieve equilibrium. For low temperatures, the time to reach equilibrium

becomes progressively longer. One way to achieve equilibrium faster is to use *simulated annealing*, i.e. to start with a high temperature and to slowly cool the system. This approach has been successfully used for the (approximate) solution of combinatorial optimization problems (such as circuit placement) in systems with a high degree of frustration [Kirkpatrick 83]. For the stereo correspondence problem, we expect the degree of frustration to be lower, so that faster annealing schedules and fewer iterations should be possible.

3.4. The Boltzmann Machine

The Boltzmann Machine is an algorithm/network⁶ that combines the ideas presented above. The binary threshold unit in Hopfield's nets is replaced by a stochastic binary threshold unit, whose probability of taking on an output of 1 when sampled is determined by the sigmoid distribution

$$p_k = \sigma(\Delta E_k) = \frac{1}{1 + e^{-\Delta E_k/T}} \quad (5)$$

The shape of this curve and its dependence on T are shown in Figure 3-2.

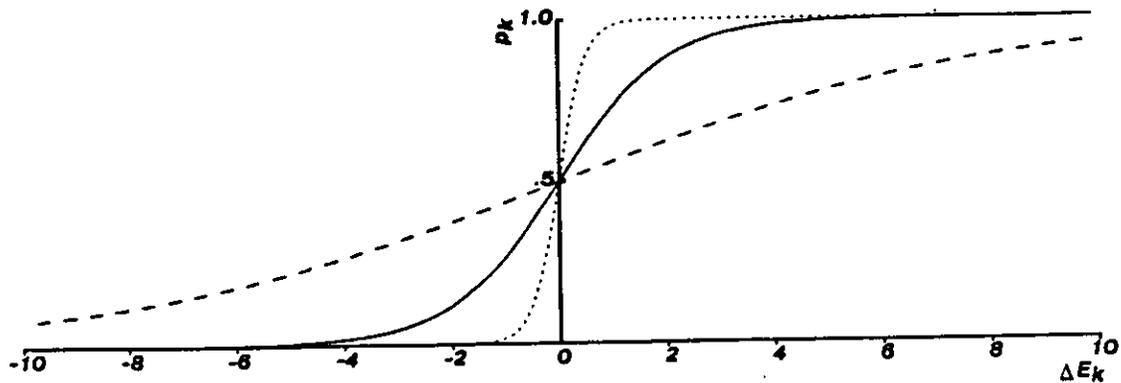


Figure 3-2: Sigmoid function

Units are sampled asynchronously, either in random, permuted or sequential order. While the sampling strategy does not affect the equilibrium distribution, permuted sampling has empirically been found to have the fastest convergence. The temperature of the system is started at some high value, and is then cooled according to some *annealing schedule* until the network has "frozen" into its final configuration.

The Boltzmann Machine is thus an example of a stochastic optimization technique applied to a very simple energy function (a sum of quadratic terms). It is also a simple case of the more general Markov Random Field model, where units are not restricted to being binary, and energy functions can be

⁶The network (setting of weights) determines the computation to be performed, while the updating rules (sampling and temperature control) complete the description of the algorithm.

more complex (see [Geman 84] for an overview). The use of MRF models, while not being part of the "pure" Boltzmann Machine model, has been studied in the context of stereo correspondence algorithms and is discussed in this report. When the temperature is reduced to 0, the Boltzmann Machine performs gradient descent (it becomes equivalent to a Hopfield net), and can thus be used to solve convex problems as well. Finally, it should be noted that under some restrictive conditions, the stochastic gradient-descent algorithm can be replaced by a "winner-take-all" network [Marroquin 83] (this algorithm is discussed in a later section).

3.5. Learning

In addition to implementing weak constraint satisfaction and finding globally optimal solutions to non-convex problems, the Boltzmann Machine can be made to actually *learn* the weights that determine a particular algorithm. The learning algorithm is based on the elegant relationship between the performance of the network (characterized as an entropy measure G) and the joint probabilities of node pairs p_{ij} (see [Hinton 84] for a detailed explanation).

The stereo correspondence problem could in theory be "learned" by studying examples of random-dot pairs (or more specifically the input compatibilities) and the final "correct" surface interpretations. Only a few weights would have to be learned because of the spatial isotropy inherent in the system. While this does not appear to be a difficult task to implement, it has not yet been tried, and remains one of the areas for future research.

4. Simple networks

This section introduces some new stereo correspondence algorithms based on the Boltzmann Machine. Three different kinds of networks are presented. The first subsection shows how hand-built networks can be used. The second subsection shows how a network can be derived from the *a priori* and image formation statistics, in effect solving the *inverse problem* of optimally estimating the original surface from the stereo pair. The third subsection extends the binary Boltzmann Machine model to use n-ary (rate) encodings.

4.1. Hand-wired networks

Hand-wired networks are the simplest to implement. The networks used in previous cooperative stereo algorithms have all been of this kind (e.g. [Dev 74, Marr 76]). Among the parameters that can be varied when specifying the network are the number of excitatory neighbors, the shape of the inhibition (single or double column) and the relative weights assigned to the excitatory, inhibitory, threshold and input links.

One such simple network is a four neighbor excitatory region with single column inhibition. The equation for the input to the decision unit is

$$-\Delta E_{x,y,d} = \alpha \sum_{x',y',d' \in N_4(x,y,d)} s_{x',y',d'} - \beta \sum_{x',y',d' \in M_6(x,y,d)} s_{x',y',d'} - \gamma + \delta \cdot c_{x,y,d} \quad (6)$$

with the neighborhoods

$$N_4(x,y,d) = \{(x',y',d') | (x-x')^2 + (y-y')^2 \leq 1\} \quad (7)$$

and

$$M_6(x,y,d) = \{(x,y,d') | d' \in 0 \dots 6 \text{ and } d' \neq d\} \quad (8)$$

for the model with 7 disparity levels. The resulting energy equation (c.f. equation (2)) has the following weights:

$$w_{x,y,d,x',y',d'} = \begin{cases} \alpha & \text{if } (x',y',d') \in N_4(x,y,d) \\ -\beta & \text{if } (x',y',d') \in M_6(x,y,d) \end{cases}$$

and

$$\theta_{x,y,d} = \gamma - \delta \cdot c_{x,y,d}$$

The choice of the α, β, γ and δ parameters has to be determined empirically. Their values should be chosen so that in the final solution, there is exactly one point on at each depth, neighboring points have similar depth values, and the surface follows areas of high "compatibility" density.

Other simple networks can be obtained by changing the size and shape of the excitatory and inhibitory regions. For example, the excitatory N_4 region can be replaced by the

$$N_8(x,y,d) = \{(x',y',d') | (x-x')^2 + (y-y')^2 \leq 2\} \quad (9)$$

or

$$N_{12}(x,y,d) = \{(x',y',d') | (x-x')^2 + (y-y')^2 \leq 4\} \quad (10)$$

neighborhoods. The single column M_6 inhibitory region can be replaced by the double column

$$M_{12}(x,y,d) = M_6 \cup \{(x'+d'-d, y, d') | d' \in 0 \dots 6 \text{ and } d' \neq d\} \quad (11)$$

inhibitory region.

Various combinations of these neighborhoods and parameters have been proposed in previous work. For example, Marr and Poggio [Marr 76] use N_{12} and M_{12} neighborhoods, with parameter values $(\alpha, \beta, \gamma, \delta) = (1, 2, 4, 0)$. These values were chosen to work well with their cooperative algorithm, and are justified in a subsequent paper [Marr 78]. Marroquin's "winner-take-all" network [Marroquin 83] uses a N_8 neighborhood with no inhibition built into the local cost function. These two algorithms are

discussed in more detail in Section 6.

4.2. Derived networks

The previous section has presented a number of networks that can be used to solve the correspondence problems. The parameters in these networks are adjusted by hand to give good performance on the sample images that are used. An alternate approach is to start with an *a priori* model of the surface and of the random coloring process that generates the stereo pair, and to derive the network to be used to solve the correspondence. This is similar to the work in "optimal restoration" that has been done by Geman and Geman [Geman 84], and is a direct solution of the *inverse problem* of estimating scene parameters (in this case depth) from the image (stereo pair).

The *a priori* surface model that we will start with will be the Markov Random Field (MRF) model. Such models are characterized by an energy function defined over the possible states of the system (in this case, over the possible surfaces that can be generated). The probability of any particular surface being generated is then related to its energy by the Boltzmann distribution. These models are related to the "regularization" methods for solving surface interpolation [Poggio 84], in that the cost function that is being minimized can also be used as an energy function in the generative surface model.

The particular MRF model that we will consider is a modification of the one used in [Geman 84]. The energy of a given surface is of a similar form to those proposed for the solution of the correspondence problem in the previous subsection. It has an excitatory neighborhood that implements the smoothness constraint and an inhibitory neighborhood (single or double column) that implements uniqueness. There is no contribution from the compatibilities (i.e. $\delta = 0$), since this is a surface model.

Given an *a priori* model of the surfaces, we can then derive the (stochastic) equations of left and right image formation, and then the equations for the compatibility matrix. We will characterize the dot density of the random painting process by the parameter d . As well, when the dots are being projected to the left and right image planes, there is a probability n that the dot intensity will be reversed (n stands for noise).

Assume that there is actually a surface patch passing through the point (x,y,d) , i.e. $S_{x,y,d} = 1$, where S stands for the surface. This point (x,y,d) projects to the point (x,y) in the left image and $(x+d,y)$ in the right. Ignoring the problems of occlusion that occur near depth edge boundaries, we can calculate the probability of the four possible combinations of intensities for the left and right image points

Prob($L,R|S=1$) as

$$\begin{aligned} P_{LR}(0,0|1) &= (1-d)(1-n)^2 + d n^2 \\ P_{LR}(0,1|1) &= P_{LR}(1,0|1) = n(1-n) \\ P_{LR}(1,1|1) &= d(1-n)^2 + (1-d)n^2 \end{aligned}$$

The [marginal] probability of any point in the left or right image being on is

$$p_{on} = d(1-n) + (1-d)n.$$

Thus, if the two points $L(x,y)$ and $R(x+d,y)$ are *not* in correspondence, i.e. $S_{x,y,d}=0$, then we have the probabilities

$$\begin{aligned} P_{LR}(0,0|0) &= (1-p_{on})^2 \\ P_{LR}(0,1|0) &= P_{LR}(1,0|0) = p_{on}(1-p_{on}) \\ P_{LR}(1,1|0) &= p_{on}^2 \end{aligned}$$

Knowing both the *a priori* surface model and the image formation equations, we can in theory find the *Maximum a Posteriori* (MAP) estimate of the original surface given the stereo pair. Instead of using this image pair directly, however, we will first calculate the compatibilities (as discussed in Section 2.1.2), and then use these as the input to the correspondence algorithm. Note that in this process, some of the information inherent in the stereo pair is lost. However, it makes the subsequent processing simpler since (as we will see) the compatibilities end up adding a linear term to the surface model equation.

The formulation for the compatibilities depends on whether we use the *dense* or *sparse* compatibilities.

$$\begin{aligned} P_{dense}(0|S) &= P_{LR}(0,1|S) + P_{LR}(1,0|S) \quad \text{and} \quad P_{dense}(1|S) = P_{LR}(0,0|S) + P_{LR}(1,1|S) \\ P_{sparse}(0|S) &= P_{LR}(0,0|S) + P_{LR}(0,1|S) + P_{LR}(1,0|S) \quad \text{and} \quad P_{sparse}(1|S) = P_{LR}(1,1|S) \end{aligned}$$

The MAP estimate is that configurations of surface points $\{S_{x,y,d}\}$ which maximizes the conditional probability

$$P(S|C) = \frac{P(S,C)}{P(C)} = \frac{P(S)P(C|S)}{P(C)}$$

which is equivalent to minimizing

$$E = -\log P(S) - \log P(C|S) \approx E' - \sum_{xyd} \log P(C_{xyd}|S_{xyd})$$

The first term is the energy associated with the surface model, while the second relates to the log likelihood of the compatibilities given the surface. Note that expansion of the $P(C|S)$ term into a

product of local probabilities assumes independence between the $P(C_{xyd}|S_{xyd})$, which is only approximately true.⁷ We can calculate the difference in energies between a unit $S_{x,y,d}$ being on and being off

$$-\Delta E_{x,y,d} = -\Delta E'_{x,y,d} + \log \frac{P(C_{xyd}|1)}{P(C_{xyd}|0)}$$

When $C_{xyd}=1$, then the log term is positive (excitatory), and when $C_{xyd}=0$, then the log term is negative (inhibitory). The overall contribution of the compatibility information to the solution can be modeled by a link between the compatibility and surface element units, and an added inhibition term (Figure 4-1).

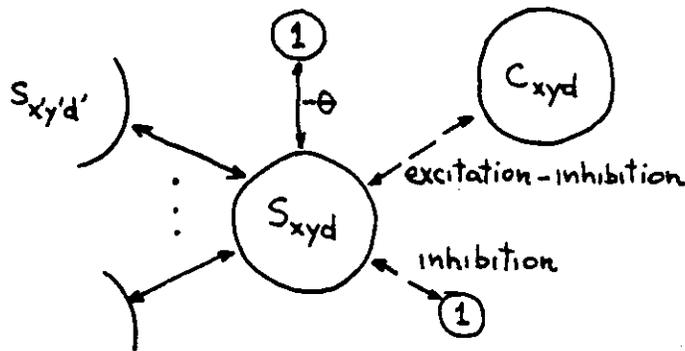


Figure 4-1: Network derived from image formation model

The plots of the excitation, inhibition and (excitation - inhibition) terms as a function of the dot density d and the noise n are shown in Figures 4-2 and 4-3. From these graphs, we see that using dense compatibilities works best for densities close to 50%, and that most of the contribution comes from *negative information* (i.e. inhibition) when the compatibility is 0. Sparse compatibilities work best at low dot densities, using *positive information* when the compatibility is 1. The effect of noise is in general to diminish the relative weight of the compatibilities as compared to the *a priori* surface model (smoothness and uniqueness constraints).

The above model is a simplification of the image formation process for the stereo pair. For one thing, the original surface model is not necessarily a good match to the surface, which is usually generated by hand (e.g. the "floating square" or "wedding cake") and not by taking a sample from the surface model MRF. As well, the simplified imaging model does not take into account the problems of occlusion. However, the noise parameter can in some way compensate for this, since

⁷The $P(C_{xyd}|S_{xyd})$ have 0 correlation for $d=50\%$ and dense compatibilities, and correlations of less than 0.1 for commonly used values of d and n .

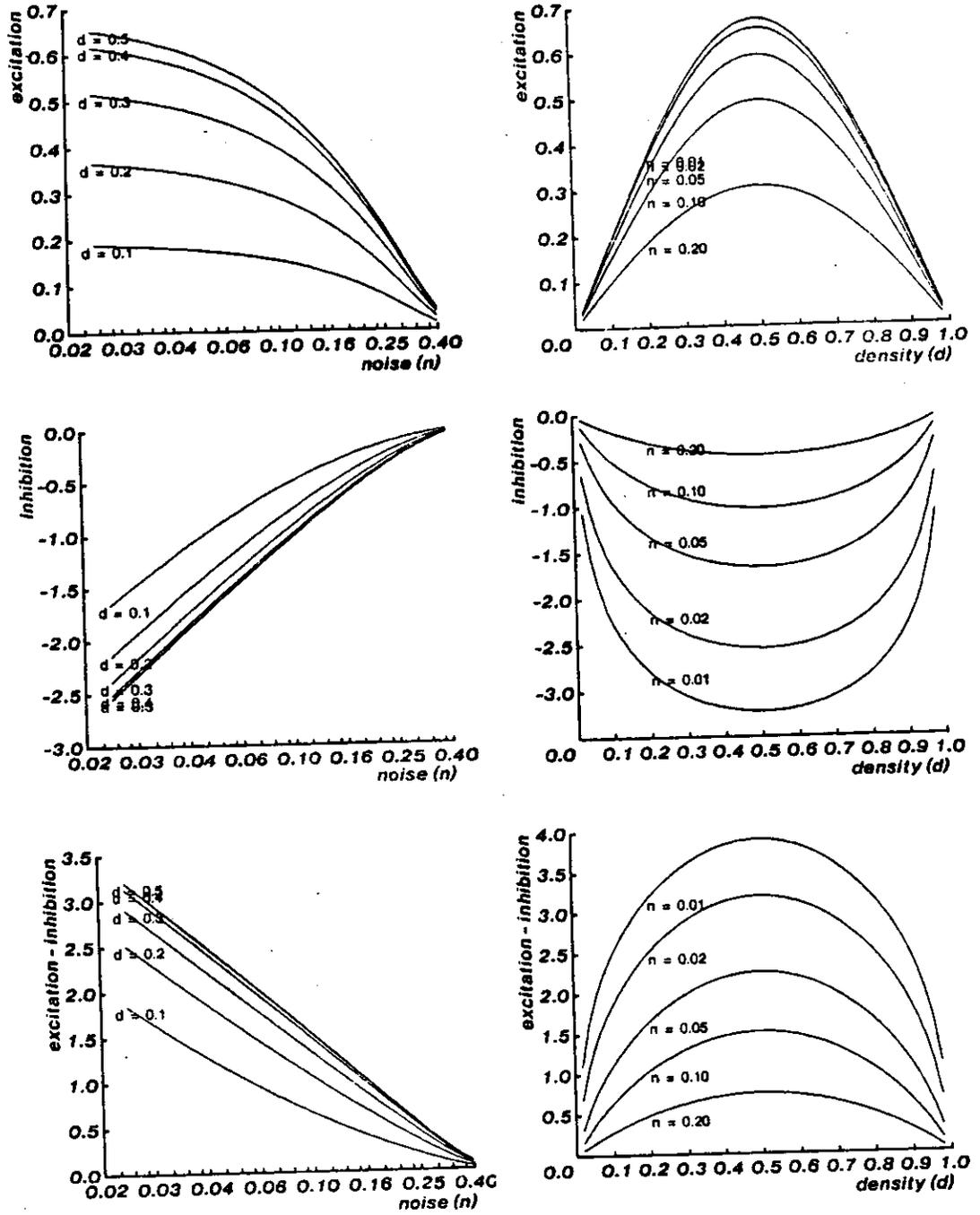


Figure 4-2: Plots of excitation and inhibition from dense compatibilities

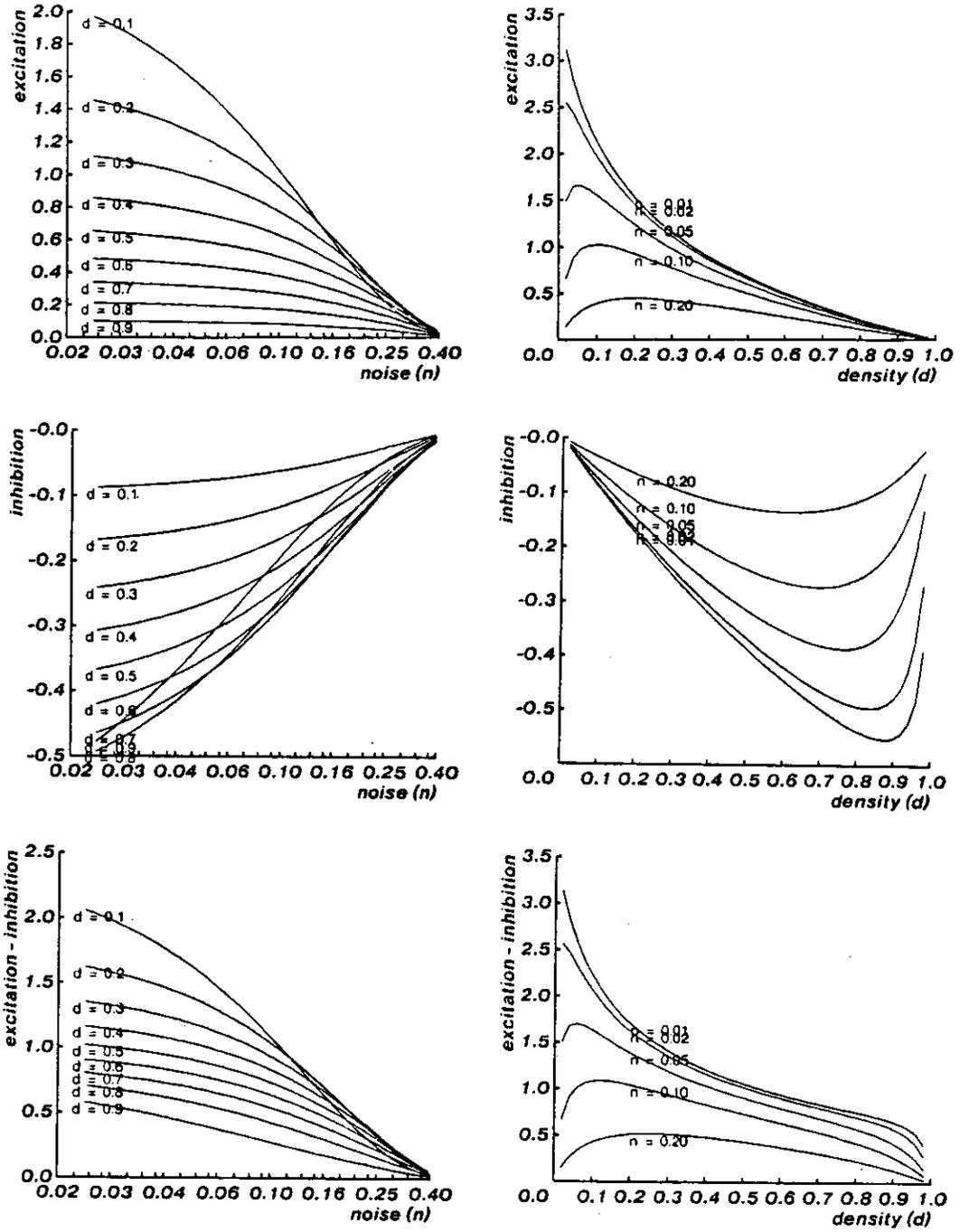


Figure 4-3: Plots of excitation and inhibition from sparse compatibilities

occlusions are sufficiently rare that their occurrence can be treated (in the first order) as random noise.

4.3. Multi-level units

The networks introduced thus far have all been composed of binary units, where each unit represents a particular correspondence, or equivalently, a depth point $S_{x,y,d}$. It is also possible to use a *rate encoding*, where only one unit is used for each (x,y) position, and its value represents the current depth, i.e. $d(x,y)$. To use these multi-level (rate encoded) units, we need only change the updating rule of the correspondence algorithm so as to leave only one (binary) unit on in each (x,y) column after each step. There are two different updating rules possible.

The first rule picks a new candidate level $d_{new}(x,y)$ and calculates the energy gap between this state and the old state. A transition is then taken under the same conditions as that used with the Boltzmann Machine, i.e. if the sigmoid of the energy gap exceeds a uniform random number.⁸ The second rule calculates the energy gaps for *all* of the possible $d(x,y)$ values. The probability that level d with energy gap ΔE_d is chosen is then

$$p(d) = \frac{1}{Z} e^{-\Delta E_d/T} \quad \text{with} \quad Z = \sum_d e^{-\Delta E_d/T}$$

This second rule is called the *heat-bath* equation, since it directly places the unit (x,y) in thermal equilibrium with respect to the other [fixed] units. Using the heat-bath equation leads the system to a faster thermal equilibrium, but is computationally more expensive for each step.

When using this multi-level representation, the single column uniqueness constraint is directly satisfied, since only one possible correspondence exists in each (x,y) column. Note that it is also now impossible to have 0 or 2 matches in a column, i.e. the weak uniqueness constraint (characterized by a penalty term added to the energy function) has turned into a strong constraint. To implement the true double column uniqueness, a penalty can be added to the energy term (as in the binary case) whenever $d(x+\delta,y) = d(x,y) + \delta$.

The multi-level representation is the usual choice for many image processing tasks such as surface interpolation [Terzopoulos 84a] or gray-level restoration [Geman 84]. For the stereo correspondence problem, it directly implements one kind of uniqueness constraint, and may thus lead to a faster solution. However, this representation is unable to represent competing depth hypotheses simultaneously, and thus cannot be used to solve certain kinds of stereograms, such as those with

⁸ If $|d_{new}(x,y) - d_{old}(x,y)| = 0$ or 1, we have a random Markov walk. This can be approximated by an analog network [Koch 85] with noise.

transparency.

5. More complex models

This section discusses some possible extensions to the models introduced in the previous section. The first two extensions, line processes and separate representations for correspondence and interpolation, could be achieved by simply increasing the size of the networks and using more complicated (e.g. ternary) energy terms. The third extension, coarse coding, can also be accommodated within the Boltzmann Machine model, and deals with the question of structuring the energy landscape for easier search. The last set of extensions are in reality alternative cooperative networks that could be used instead of the Boltzmann Machine. None of the extensions described in this section have yet been implemented or evaluated.

5.1. Line process

In addition to the usual (x,y,d) lattice for solving the correspondence, it is possible to introduce a "dual" lattice between adjacent (x,y) points [Geman 84]. This dual lattice encodes the presence of "edge elements" which indicate a discontinuity in depth. Such edge elements receive support from nearby aligned edge elements, and inhibit the continuity constraint across the edge. In [Geman 84] these edge elements are called a "line process" because the edge element interactions are along the discontinuity contour. Terzopoulos [Terzopoulos 84a, Terzopoulos 84b] deals with how depth (and orientation) discontinuities can be used in a finite-element (relaxation) implementation of surface interpolation. Marroquin [Marroquin 84] shows how stochastic minimization can be used to locate these discontinuities.

5.2. Separate correspondence and interpolation

In the models that we have examined so far, the correspondence problem (determining which points match) and the depth interpolation problem (determining the depth at each pixel) are solved simultaneously using the same representation. This is necessary since all of the interactions among candidate correspondences are mediated by strictly local (4, 8 or 12 neighbor) interactions.

It is also possible to solve the correspondence problem by using two separate but interacting representations. The first representation implements the uniqueness requirement, i.e. the inhibition among competing matches within a line of sight. The second representation implements the smoothness constraint, i.e. the depth interpolation using the candidate matches provided by the first representation. The first representation can use the usual binary (x,y,d) hypotheses, while the second is an implementation of a standard depth interpolator (e.g. [Terzopoulos 84a]) using multi-level (rate

encoded) units. The overall system still has a well defined energy function that can be minimized by using a combination of stochastic and deterministic gradient descent algorithms.

5.3. Coarse coding

The two types of depth representation that have been studied thus far are *rate encoding* and *interval encoding*. A third possible representation is distributed (or coarse) coding, where the units have overlapping response fields. One advantage of such units is that it is possible to have more "degeneracy" in the system. At high temperatures, the "virtual" energy barriers between various states (e.g. rivalrous correspondences) are reduced because there are many different ways to move from one representation to another [Hinton 86a]. An energy barrier with many paths over it acts like a lower "virtual" barrier with few paths over it. Another advantage is that the coarse coding of higher dimensional spaces (such as the (x,y,d) space) can result in greater accuracy with fewer units [Hinton 86b].

An example of using coarse coding (in one dimension) would be if each unit represented three adjacent depths, and each depth were represented by the conjunction of three overlapping units. In this case, it is easier to move from one depth hypothesis to another, since there are many ways to partially turn on the new representation while turning off the old (more degeneracy).

5.4. Real-valued units

Another variation on the binary stochastic (Boltzmann Machine) model previously presented is to replace the units with *real-valued* deterministic units [Hopfield 84]. These units are linear summation units followed by a saturating non-linearity, and can thus be regarded as the mean-field approximations to the Boltzmann Machine units [Rumelhart 86] [Hopfield 85]. The amount of non-linearity is initially low, with most of the units operating in their linear range. As the non-linearity is increased, the units tend to 0 or 1. Thus, the solution moves from the center of the hypercube that forms the state space to one of its corners.

The advantage of such units is that local confidence in hypotheses can be directly transmitted using a real number, rather than waiting for the *probabilistic* effect of one unit on another. These networks, if properly formulated, can also be shown to be monotonically decreasing in energy, but like deterministic binary networks, they can get trapped in local minima. However, these local minima are not as harmful, since initially the search proceeds by first moving into areas of broad global minima (principle of least commitment).

The converse of using real-valued units to represent hypotheses is to use probabilistic units to

represent real values. In this model, the "flicker" of a stochastic unit is used to encode the real value [Hopfield 84]. The types of interaction that such units can have, however, is limited to simple summation, possibly followed by a non-linearity.

One place where such units could be used is in those problems that could be solved by an "analog" network [Koch 85]. These networks are used to solve simple finite difference equations such as those used for surface interpolation [Terzopoulos 84a] or shape from shading [Horn 86]. At each step, the value of a node is replaced by some weighted average of its neighbors. A binary stochastic unit whose output probability is a linear sum of the probabilities of its inputs could be used to simulate such analog units. This system would have much slower convergence to the desired solution (if it converged at all), but would have a lower communication bandwidth between nodes since only binary values are transmitted.

6. Comparison of algorithms

This section compares the performance of the new correspondence algorithm introduced in Section 4.1 against three other previously developed methods: Marr and Poggio's cooperative stereo algorithm [Marr 76], Marroquin's winner-take-all networks [Marroquin 84], and Prazdny's coherence principle [Prazdny 85].

Three sample stereograms are used to illustrate the performance of the algorithms. These stereograms are shown in Figure 6-1, along with their associated dense and sparse compatibilities (Figure 6-2). The first stereogram is a floating central square, with dot density 20%. The second is a pyramid (or "wedding cake") with dot density 20%. The third is a set of transparent steps going through a central transparent plane, with a dot density of 20%.

6.1. Simple networks with annealing (Boltzmann Machine)

As discussed in Section 4.1, simple networks can be used to solve random-dot stereograms if simulated annealing is used. The simplest possible has a four neighbor excitatory region, and single column inhibition. It will solve stereograms if it is annealed sufficiently slowly (Figure 6-3(a)), but does not work well if simple gradient descent is used (Figure 6-3(b)).

A slightly more complex network with an eight neighbor excitatory network can be used to solve more difficult stereograms such as the pyramid (Figure 6-3(c)). However, even careful annealing does not guarantee that the lowest energy solution is found. In the previous example, if the broken ring on the pyramid is filled in by hand (Figure 6-3(d)), a lower energy solution results. This indicates that even simulated annealing cannot always find (in a reasonable amount of time) the best solution in

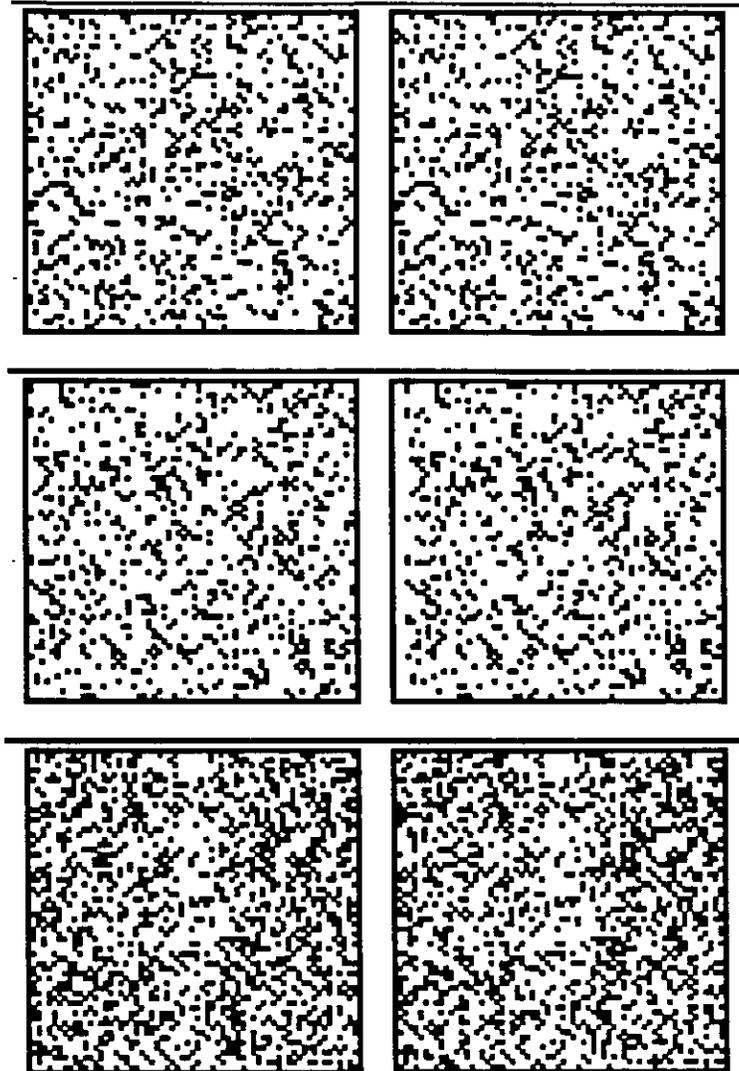


Figure 6-1: Sample stereograms
(a) 20% square (b) 20% pyramid (c) 20% steps

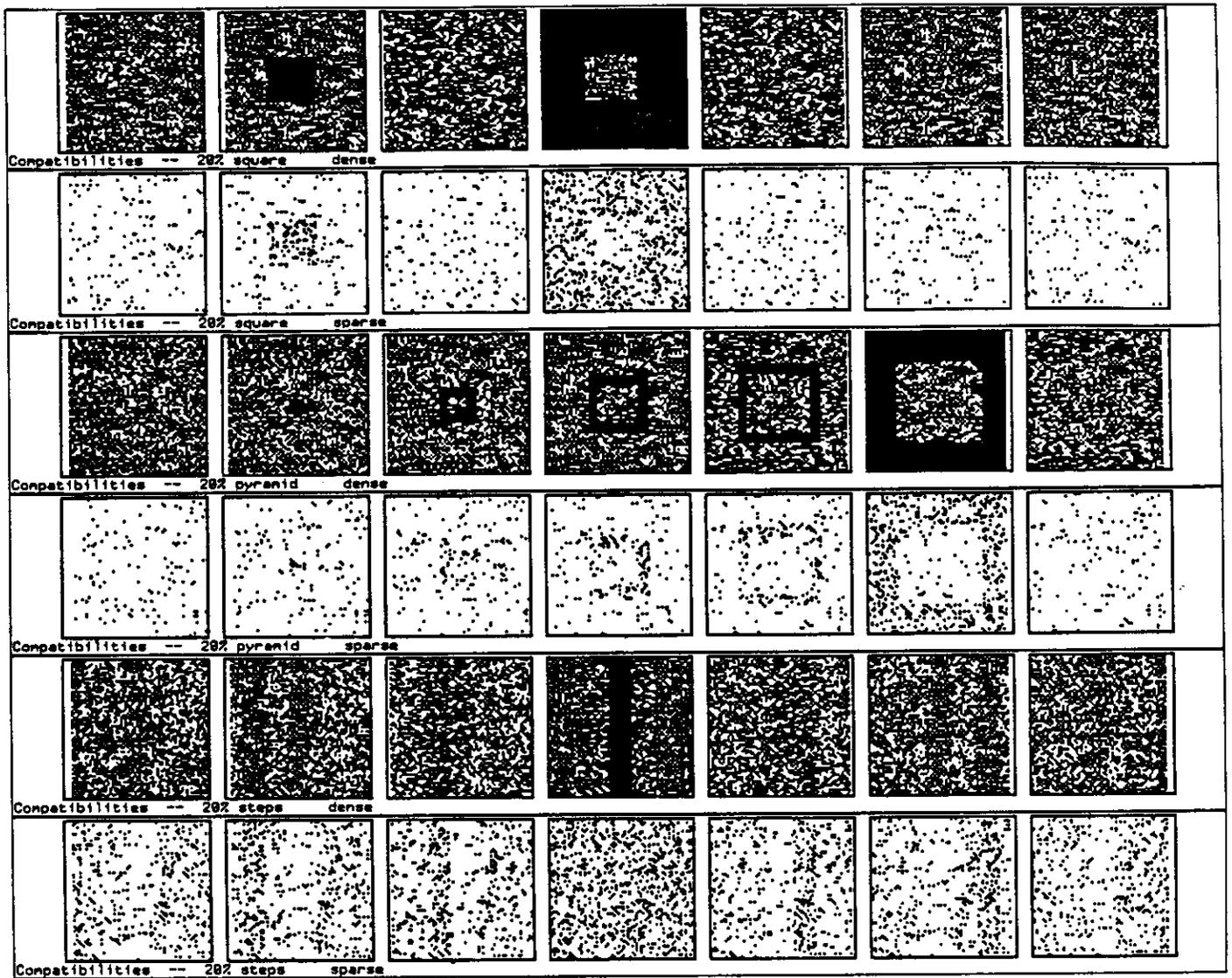


Figure 6-2: Associated compatibilities

- (a) 20% square dense compatibilities (b) sparse compatibilities
- (c) 20% pyramid dense compatibilities (d) sparse compatibilities
- (e) 20% steps dense compatibilities (f) sparse compatibilities

the presence of large numbers of almost equally good minima and high energy barriers.

6.2. A Cooperative Stereo Algorithm (Marr-Poggio)

This algorithm, first described in [Marr 76] and later analysed in [Marr 78], is a simple hand-wired network (as described in Section 4.1) which uses a *deterministic* cooperative algorithm with synchronous updating. The network is initialized to the compatibility array, and the inhibitory and excitatory neighborhoods designed so that the network settles into a reasonable interpretation of the stereogram.

The performance of this algorithm on some random-dot stereograms is quite reasonable (Figure 6-4(a)). Using a small amount of heat after the initial solution has been found can slightly improve the results (Figure 6-4(b)). However, doing a full annealing, which finds the global minimum, yields a poor result (Figure 6-4(c)). This is because the global minimum of the energy function used does not correspond to the desired solution. In this case, it is not simulated annealing which results in poor performance. Rather, the attempt to interpret the updating rules as doing a functional minimization fails since the energy function contains no term for the compatibilities (these are in the initial conditions). On some more difficult stereograms (Figure 6-4(d)), the Marr-Poggio algorithm does not produce very good results.

6.3. Winner-take-all (Marroquin)

A winner-take-all [Marroquin 84] network is a recently developed variation on the Marr-Poggio algorithm. Like the latter algorithm, the solution matrix is initialized to the compatibilities, updating is synchronous, and each unit receives support from neighbors at the same disparity. However, there is no inhibitory input from units in the same line of sight. Instead, at each step units within one depth column that do not have a maximal input are turned off. Thus, the solution must be initialized to the *dense* compatibility matrix. Eventually, a single winner remains in each column.

This algorithm has been proven to work correctly unless sufficiently large "incorrect" clumps exist in the compatibility matrix. In practice, it works well on random-dot stereograms (Figure 6-5(a)), but has problems dealing with occlusions, since the initial compatibilities in such areas may be 0. On more difficult stereograms (Figure 6-5(b)) it performs similarly to simulated annealing.

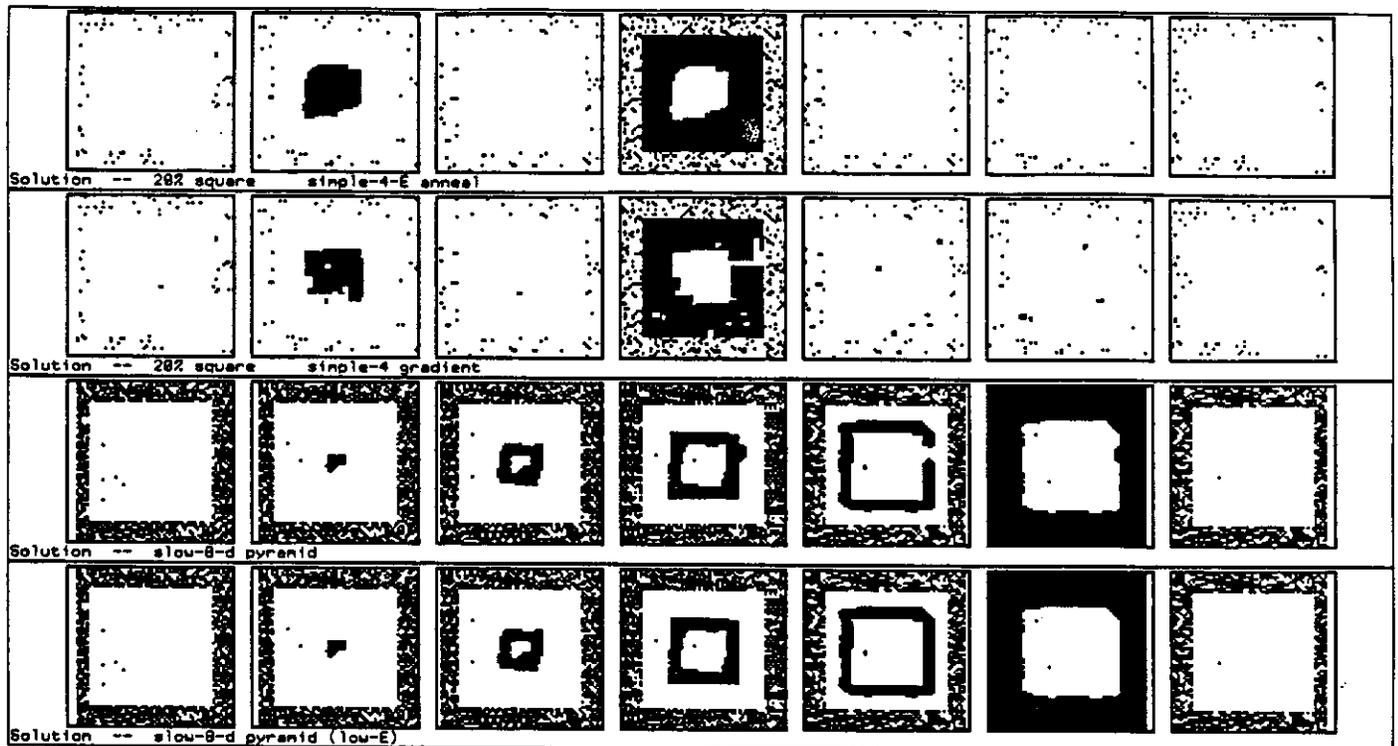


Figure 6-3: Simulated annealing results

- (a) 20% square with slow annealing (b) with gradient descent
(c) 20% pyramid with full annealing (d) lower energy configuration

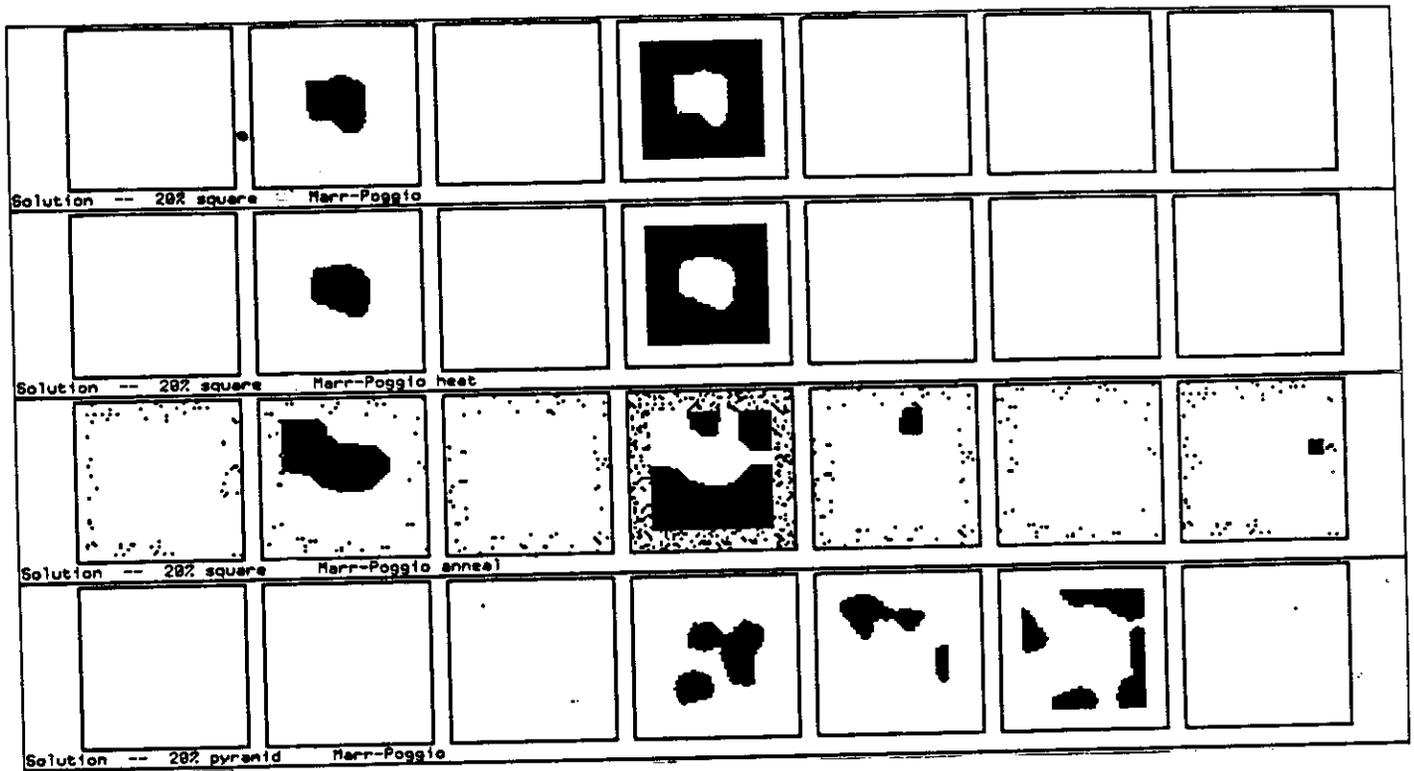


Figure 6-4: Marr-Poggio algorithm results

(a) 20% square (b) with some heat
(c) with full annealing (d) 20% pyramid

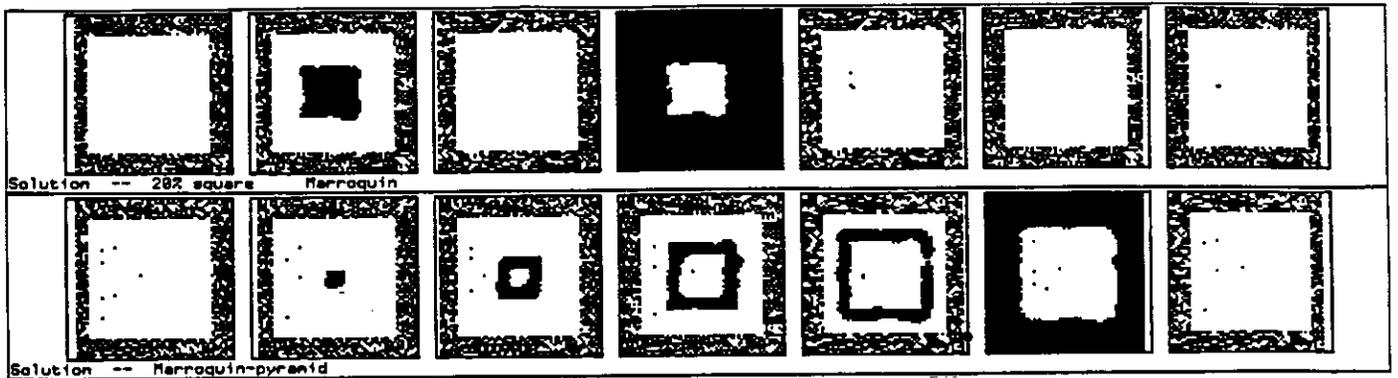


Figure 6-5: Winner-take-all results
 (a) 20% square (b) 20% pyramid

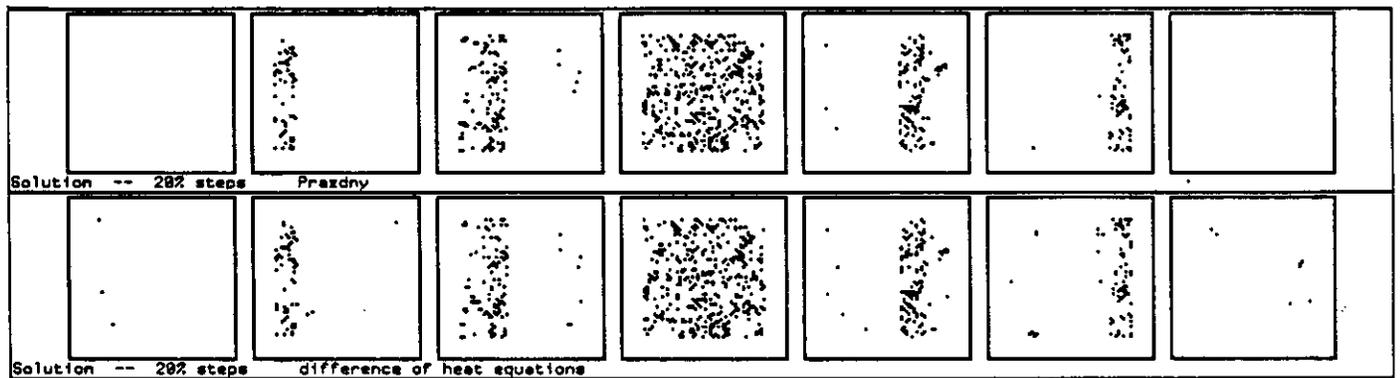


Figure 6-6: Coherence principle results
 (a) 20% transparent steps (b) iterative *diffusion* version

6.4. Coherence (Prazdny)

The *coherence principle* is an alternative to the continuity constraint that can be used as part of a parallel non-iterative correspondence algorithm [Prazdny 85]. The principle requires that points lie on some coherent (relatively smooth) surface, but does *not* require neighboring matches to be on the same surface. This allows the solution of transparent random-dot stereograms (Figure 6-6(a)).

Prazdny's algorithm works by determining for each possible match a single "disparity cell" value that is the sum of the similarity support received from its neighbors. The support function

$$w(i,j) = \frac{1}{c|i-j|\sqrt{2\pi}} e^{-|d_j-d_i|^2/2c^2|i-j|^2}$$

is based on the *disparity gradient* $\frac{|d_j-d_i|}{|i-j|}$ which is the slant of the line through the two points i and j . The strength of the support function is inversely proportional to the image plane distance between the points. Coherence is implemented by choosing only one value at point j (the one nearest to d_j) to add to the disparity cell for d_j . After all of the disparity cell values have been calculated, the match with the highest support within a disparity column is selected. This process is done separately for each eye, and the results are combined.

An alternative support function that uses some iteration and in addition has a fixed disparity gradient limit has also been proposed by Mayhew [Mayhew 85]. It is also possible to approximate the effect of the support function by using *heat diffusion*. This allows the high connectivity inherent in Prazdny's algorithm to be simulated by iteration on a network with strictly local connectivity [Szeliski 85]. The resulting performance on random-dot stereograms is quite similar (Figure 6-6(b)).

7. Conclusions

This report has examined a number of cooperative (parallel) algorithms for the solution of random-dot stereograms. A summary of its contents is as follows.

The stereo correspondence problem can be analysed in terms of its computational theory, representation and algorithms. A computational theory based on Marr's three rules of compatibility, uniqueness and continuity is implemented by most algorithms. Recently, the coherence principle has been proposed as an alternative to the continuity constraint. The representation of depth for correspondence algorithms can be either rate encoded (unique depth), interval encoded (multiple depth hypotheses) or coarse coded (overlapping fields). The most common stereo matching programs rely either on dynamic programming, multiple-resolution techniques, or parallel (cooperative) algorithms.

The problem of stereo correspondence can be formulated as the global minimization of an energy function. This formulation defines the computation separately from the implementation, and can also be related to Bayesian models of image formation. The calculation of the global minimum usually requires search, and stochastic optimization techniques can be used to find this minimum. In certain cases, this stochastic search can be implemented by a Boltzmann Machine, a network of simple stochastic binary threshold units.

The required network for solving the random-dot stereograms can be built by hand. An alternative is to derive the correct network from the *a priori* surface and image formation models. It is also possible to extend the Boltzmann Machine model to use n-ary (instead of binary) units. Other extensions to the basic model are possible, but have not been implemented. The extensions discussed in this report are: line processes, separating the correspondence from the interpolation, coarse coding, and real-valued units.

An informal comparison has been made between the new correspondence algorithm based on the Boltzmann Machine and three previously developed algorithms. Very simple (four neighbor) networks used in conjunction with annealing (stochastic optimization) can do a reasonable job on easy stereograms. However, even slow annealing combined with larger networks will not always find the minimum energy configuration. To find the optimal solution, it may be necessary to structure the network to have a better energy landscape. Algorithms that look for local minima (locally consistent interpretations) can do a reasonable job, but cannot handle more difficult cases (like very sparse stereograms). Algorithms (such as winner-take-all) tailored specifically to random-dot stereograms can do well. Using support functions in conjunction with the coherence principle makes it possible to solve a larger class of stereograms (such as transparent stereograms).

While the new algorithm presented in this report and the three comparison algorithms all do a reasonable job on most stereograms, there are still many interesting questions associated with this work. The more complex models that were introduced have yet to be implemented, and may prove to have significant advantages over existing techniques. Finally, it remains to be seen if these techniques can be extended to recover depth from stereo using real-world image pairs.

References

- [Arnold 83] Arnold, R. D.
Automated Stereo Perception.
 Stanford Artificial Intelligence Laboratory AIM-351, Stanford University, March, 1983.
- [Baker 82] Baker, H. H.
Depth from Edge and Intensity Based Stereo.
 Stanford Artificial Intelligence Laboratory AIM-347, Stanford University, September, 1982.
- [Dev 74] Dev, P.
Segmentation Processes in Visual Perception: A Cooperative Neural Model.
 COINS Technical Report 74C-5, University of Massachusetts at Amherst, June, 1974.
- [Geman 84] Geman, S. and Geman, D.
 Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images.
IEEE Trans. Pattern Anal. Machine Intell. PAMI-6(6):721-741, November, 1984.
- [Grimson 80] Grimson, W. E. L.
A Computer Implementation of a Theory of Human Stereo Vision.
 Artificial Intelligence Laboratory A.I. Memo No. 565, Massachusetts Institute of Technology, January, 1980.
- [Guerra 83] Guerra, C. and Kanade, T.
 A Systolic Algorithm for Stereo Matching.
 (Carnegie-Mellon University), 1983.
- [Hinton 77] Hinton, G. E.
Relaxation and its Role in Vision.
 PhD thesis, University of Edinburgh, 1977.
- [Hinton 84] Hinton, G. E., Sejnowski, T. J., and Ackley, D. H.
Boltzmann Machines: Constraint Satisfaction Networks that Learn.
 Technical Report CMU-CS-84-119, Carnegie-Mellon University, May, 1984.
- [Hinton 86a] Hinton, G. E. and Sejnowski, T. J.
 Learning and relearning in boltzman machines.
 In D. E. Rumelhart, J. L. McClelland, and the PDP research group (editors), *Parallel distributed processing: Explorations in the microstructure of cognition.*
 Bradford Books, Cambridge, MA, 1986.
- [Hinton 86b] Hinton, G. E., McClelland, J. L., and Rumelhart, D. E.
 Distributed representations.
 In D. E. Rumelhart, J. L. McClelland, and the PDP research group (editors), *Parallel distributed processing: Explorations in the microstructure of cognition.*
 Bradford Books, Cambridge, MA, 1986.
- [Hopfield 82] Hopfield, J. J.
 Neural networks and physical systems with emergent collective computational abilities.
Proceedings of the National Academy of Sciences U.S.A. 79:2554-2558, April, 1982.

- [Hopfield 84] Hopfield, J. J.
Neurons with graded response have collective computational properties like those of two-state neurons.
Proceedings of the National Academy of Sciences U.S.A. 81:3088-3092, May, 1984.
- [Hopfield 85] Hopfield, J. J. and Tank, D. W.
"Neural" Computation of Decisions in Optimization Problems.
Biol. Cybern. 52:141-152, 1985.
- [Horn 77] Horn, B. K. P.
Understanding Image Intensities.
Artificial Intelligence 8:201-231, April, 1977.
- [Horn 86] Horn, B. K. P. and Brooks, M. J.
The Variational Approach to Shape from Shading.
Computer Vision, Graphics, and Image Processing 33:174-208, 1986.
- [Hummel 83] Hummel, R. A. and Zucker, S. W.
On the foundations of relaxation labeling processes.
IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-5:267-287, 1983.
- [Julesz 60] Julesz, B.
Binocular depth perception of computer generated patterns.
Bell System Technical Journal 39(5):1125-1162, September, 1960.
- [Julesz 71] Julesz, B.
Foundations of Cyclopean Perception.
Chicago University Press, Chicago, IL, 1971.
- [Kass 84] Kass, M. H.
Computing Stereo Correspondence.
Master's thesis, Massachusetts Institute of Technology, May, 1984.
- [Kirkpatrick 83] Kirkpatrick, S., Gelatt, C. D. Jr., and Vecchi, M. P.
Optimization by simulated annealing.
Science 220:671-680, 1983.
- [Koch 85] Koch, C., Marroquin, J., and Yuille, A.
Analog 'neuronal' networks in early vision.
Artificial Intelligence Laboratory A.I. Memo No. 751, Massachusetts Institute of Technology, 1985.
- [Marr 76] Marr, D. and Poggio, T.
Cooperative computation of stereo disparity.
Science 194:283-287, October, 1976.
- [Marr 78] Marr, D., Palm, G., and Poggio, T.
Analysis of Cooperative Stereo Algorithm.
Biological Cybernetics 28:223-239, 1978.
- [Marr 79] Marr, D. C. and Poggio, T.
A Computational Theory of Human Stereo Vision.
Proc. Royal Soc. London B 204:301-328, 1979.

- [Marr 82] Marr, D.
Vision, A Computational Investigation into the Human Representation and Processing of Visual Information.
W. H. Freeman and Company, San Francisco, 1982.
- [Marroquin 83] Marroquin, J. L.
Design of Cooperative Networks.
Artificial Intelligence Laboratory Working Paper 253, Massachusetts Institute of Technology, July, 1983.
- [Marroquin 84] Marroquin, J. L.
Surface Reconstruction Preserving Discontinuities.
A. I. Memo 792, Massachusetts Institute of Technology, August, 1984.
- [Mayhew 80] Mayhew, J. E. W., and Frisby, J. P.
The Computation of Binocular Edges.
Perception 9:69-87, 1980.
- [Mayhew 81] Mayhew, J. E. W. and Frisby, J. P.
Psychophysical and Computational Studies towards a Theory of Human Stereopsis.
Artificial Intelligence 17(1-3):349-408, August, 1981.
- [Mayhew 85] Mayhew, J. E. W.
A stereo correspondence algorithm using a disparity gradient limit.
1985.
Personal communication.
- [Metropolis 53] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.
Equations of state calculations by fast computing machines.
Journal of Chemical Physics 21:1087-1091, 1953.
- [Ohta 85] Ohta, Y. and Kanade, T.
Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming.
IEEE Trans. Pattern Analysis and Machine Intelligence PAMI-7(2):139-154, March, 1985.
- [Poggio 84] Poggio, T. and Torre, V.
Ill-Posed Problems and Regularization Analysis in Early Vision.
In *IUS Workshop*, pages 257-263. ARPA, October, 1984.
- [Prazdny 85] Prazdny, K.
Detection of Binocular Disparities.
Biological Cybernetics 52:93-99, 1985.
- [Rosenfeld 76] Rosenfeld, A., Hummel, R. A., and Zucker, S. W.
Scene labeling by relaxation operations.
IEEE Transactions on Systems, Man, and Cybernetics SMC-6:420-433, 1976.
- [Rumelhart 86] Rumelhart, D. E., Hinton, G. E., and Williams, R. J.
Learning internal representations by error propagation.
In D. E. Rumelhart, J. L. McClelland, and the PDP research group (editors), *Parallel distributed processing: Explorations in the microstructure of cognition.*
Bradford Books, Cambridge, MA, 1986.

- [Sperling 70] Sperling, G.
Binocular Vision: A Physical and Neural Theory.
J. Am. Psychol. 83:461-534, 1970.
- [Szeliski 85] Szeliski, R. and Hinton, G.
Solving Random-Dot Stereograms Using the Heat Equation.
In *Proceedings of the Conference on Computer Vision and Pattern Recognition*.
IEEE Computer Society, San Francisco, CA, June, 1985.
- [Terzopoulos 84a] Terzopoulos, D.
Multiresolution Computation of Visible-Surface Representations.
PhD thesis, Massachusetts Institute of Technology, January, 1984.
- [Terzopoulos 84b] Terzopoulos, D.
Controlled-Smoothness Stabilizers for the Regularization of Ill-Posed Visual
Problems Involving Discontinuities.
In *IUS Workshop*, pages 225-229. ARPA, October, 1984.
- [Waltz 75] Waltz, D. L.
Understanding line drawings of scenes with shadows.
In P. Winston (editor), *The Psychology of Computer Vision*. McGraw-Hill, New York,
1975.
- [Yuille 84] Yuille, A. L. and Poggio, T.
A Generalized Ordering Constraint for Stereo Correspondance.
A. I. Memo 777, Massachusetts Institute of Technology, May, 1984.