The Theory and Practice of Bayesian Image Labeling

by

Paul Bao-Luo Chou

Submitted in Partial Fulfillment
of the
Requirements for the Degree

Doctor of Philosophy

Supervised by Christopher M. Brown

Department of Computer Science
College of Arts and Sciences

University of Rochester

Rochester, New York

1988

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR 258	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* The Theory and Practice of Bayesian Image Labeling		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Paul Bao-Luo Chou		8. CONTRACT OR GRANT NUMBER(s) DACA76-85-C-0001 N00014-82-K-0193
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Computer Science 734 Computer Studies Bldg. Univ. of Rochester, Rochester, NY 14627		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS DARPA / 1400 Wilson Blvd. Arlington, VA 22217		12. REPORT DATE August 1988 13. NUMBER OF PAGES 143
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. *(of this report)* unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Distribution of this document is unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Image Segmentation
Markov Random Fields
Relaxation Labeling
Image Analysis
Sensor Fusion

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Integrating disparate sources of imformation has been recognized as one of the keys to the success of general purpose vision systems. Image clues such as shading, texture, stereo disparities and image flows provide uncertain, local and incomplete information about the three-dimensional scene. Spatial a priori knowledge plays the role of filling in missing information and smoothing out noise. This thesis proposes a solution to the longstanding open problem of visual integration. It reports a framework, based on Bayesian probability theory, for computing an intermediate representation of the scene

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

20. ABSTRACT (Continued)

from disparate sources of information.

The computatuion is formulated as a labeling problem. Local visual observations for each image entity are reported as label likelihoods. They are combined consistently and coherently an hierarchically structured label trees with a new, computationally simple procedure. The pooled label likelihoods are fused with the a priori spatial knowledge encoded as Markov Random Fields (MRF The a posteriori distribution of the labelings are thus derived in a Bayesian formalism. A new inference method, Highest Confidence First (HCF) estimation, is used to infer a unique labeling from the a posteriori distribution. Unlike previous inference methods based on the MRF formalism, HCF is computationally efficient and predictable while meeting the principles of graceful degradation and least commitment. The results of the inferences process are consistent with both observable evidence and a priori knowledge.

The effectiveness of the approach is demonstrated with experiments on two image analysis problems: intensity edge detection and surface reconstruction. For edge detection, likelihood outputs from a set of local edge operators are integrated with a priori knoledge represented as an MRF probability distribution. for surface reconstruction, intensity information is integrated with sparse depth measurements and a priori knowledge. Coupled MRF's provide a unified treatment of surface reconstruction and segmentation, and an extension of HCF implements a solution method. Experiments using real image and depth data yield robust results. The framework can also be generalized to higher-level vision problems, as well as to other domains.

Acknowledgements

I would like to express my deep gratitude to Chris Brown, my adviser. Without his encouragement, support, and guidance, this work would have been impossible.

I would also like to thank my committee. Dana Ballard and Jerry Feldman have helped me in many ways. Their work has always been a source of inspiration for me. Henry Kyburg has provided many valuable suggestions. I am grateful for his advice and attention. Edwin Kinnen suggested many improvements that enhance the readability of this thesis.

I am indebted to Dave Sher for suggesting the edge models used in this work, and for the use of edge-detection software developed by him for our experiments. Rajeev Raman, Lata Narayanan, and Myra VanInwegen have also contributed to the experiments. Much credit should go to Rajeev for his help in designing and implementing the MRF simulator.

Paul Cooper has helped me at various stages in the development of this work. He provided his stereo system that made some of the experiments possible, proofread this document, and corrected many twisted sentences. Stuart Friedberg and Mark Fanty taught me many things during the past 5 years. I wish to thank them for their help and friendship.

I have been fortunate to be able to interact with Mike Swain, Yiannis Aloimonos, and Amit Bandopadhay. The Rochester vision group has stimulated my intellectual growth. Others at Rochester have also helped to make my stay here more enjoyable. Among them, I would like to mention Doug Ierardi and Tom Blenko. I also wish to thank Professor Tomaso Poggio and Professor Max Mintz for their encouragement and suggestions.

Finally, I wish to express my special gratitude to my wife Jihan for all the support and understanding she provided during the past years.

This work was supported in part by U.S. Army Engineering Topographic Laboratories research contract no. DACA76-85-C-0001, in part by NSF Coordinated Experimental Research grant no. DCR-8320136, and in part by ONR/DARPA

research contract no. N00014-82-K-0193. We thank the Xerox Corporation University Grants Program for providing equipment used in the preparation of this thesis.

Abstract

Integrating disparate sources of information has been recognized as one of the keys to the success of general purpose vision systems. Image clues such as shading, texture, stereo disparities and image flows provide uncertain, local and incomplete information about the three-dimensional scene. Spatial a priori knowledge plays the role of filling in missing information and smoothing out noise. This thesis proposes a solution to the longstanding open problem of visual integration. It reports a framework, based on Bayesian probability theory, for computing an intermediate representation of the scene from disparate sources of information.

The computation is formulated as a labeling problem. Local visual observations for each image entity are reported as label likelihoods. They are combined consistently and coherently on hierarchically structured label trees with a new, computationally simple procedure. The pooled label likelihoods are fused with the *a priori* spatial knowledge encoded as Markov Random Fields (MRF's). The *a posteriori* distribution of the labelings are thus derived in a Bayesian formalism. A new inference method, Highest Confidence First (HCF) estimation, is used to infer a unique labeling from the *a posteriori* distribution. Unlike previous inference methods based on the MRF formalism, HCF is computationally efficient and predictable while meeting the principles of graceful degradation and least commitment. The results of the inference process are consistent with both observable evidence and *a priori* knowledge.

The effectiveness of the approach is demonstrated with experiments on two image analysis problems: intensity edge detection and surface reconstruction. For edge detection, likelihood outputs from a set of local edge operators are integrated with *a priori* knowledge represented as an MRF probability distribution. For surface reconstruction, intensity information is integrated with sparse depth measurements and *a priori* knowledge. Coupled MRF's provide a unified treatment of surface reconstruction and segmentation, and an extension of HCF implements a solution method. Experiments using real image and depth data yield robust results. The framework can also be generalized to higher-level vision problems, as well as to

other domains.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1. Introduction

Current computer vision systems are neither as robust nor as flexible as biological vision systems. However, there do exist computer programs that solve particular visual problems in restricted environments [Ballard and Brown 1982]. One common characteristic of these efforts is that they use only a small fraction of the information conveyed in the input images. Examples include the computation of intrinsic images [Barrow and Tenenbaum 1978], such as structure from motion, shape from shading, and shape from texture. Figure 1.1 shows an image of stereo disparity data found by a typical feature-based stereo system. Such data provide partial, uncertain, and incomplete information about the scene. Other image features such as texture, shading, contours, and optical flow can all provide information valuable for interpretation of an image.

Much of computer vision research in the last decade has had as a goal the recovery of physical properties (such as depth, (spectral) reflectance, or surface orientation) from images. Until recently, this research in deriving physical characteristic X from image characteristic Y was pursued under the assumption that

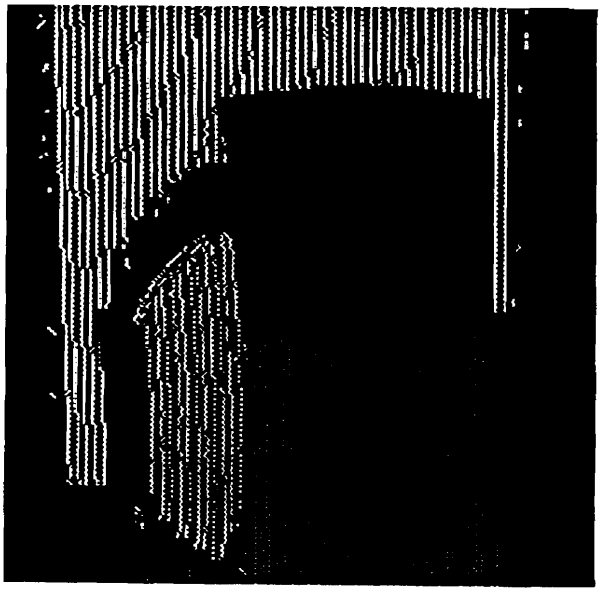

Figure 1.1. Stereo Disparity Image

only variations in X causes variations in Y. For example, all shading variations in an image would be assumed to arise from variations in surface orientation. (Shadows and paint would not be accounted for.) This work leans to the idea of "visual modules" that work separately under their own assumptions. Usually the physical characteristic is underdetermined by the image cue, so *a priori* knowledge and physical constraints are used to render the problem solvable.

This work on reconstruction of intrinsic characteristics from single cues raised the obvious question: How can multiple image cues work together? Together they provide a much richer description of the scene than any image cue does alone. Human observers seem to exploit multiple cues in a coherent manner, but existing vision systems lack the ability to integrate multiple image cues. Furthermore, much information is lost in the three-dimension to two-dimension imaging process. To infer a robust representation of the scene structure, it is necessary to incorporate appropriate *a priori* knowledge to fill in missing information and to smooth out noise. The employment of natural constraints in visual computation has been of interest for some time [Brown 1984], but there still lacks a mechanism to combine such constraints consistently with disparate input data. One approach was followed by Aloimonos [1986], who explored the increased mathematical constraints imposed by several sets of image characteristics. The work reported here is different in character: individual modules form their opinions, which are combined by a method based on Bayesian probability theory. One advantage of this approach is that it can proceed over a hierarchy of abstractions. It has a parallel-iterative computational basis. The idea of parallel-cooperative computation of intrinsic parameters has been articulated [Barrow and Tenenbaum 1978] [Ballard 1984] [Poggio 1985], but, due to the complex physical couplings of intrinsic parameters, a satisfactory implementation of this idea is not yet seen.

This thesis proposes a solution to the longstanding open problem of visual integration. It reports a framework, based on Bayesian-probability theory, for computing an iconic (image-like, or raster) representation of the scene from disparate

sources of information. The computation is formulated as a labeling problem: for each image location find the appropriate attribute or label (such as depth, or whether it is an object boundary) describing the corresponding portion of the scene. The thesis confronts and solves problems in the representation of knowledge, reasoning procedures for combining distinct bodies of knowledge, and inference methods for using available knowledge to infer scene properties. We have successfully applied the framework to two instances of the labeling problem - boundary detection and surface reconstruction, The results using range data and irradiance inputs are also reported here. The thesis deals with the following open problems.

(1) How to integrate multi-modal visual data in a hierarchically structured hypothesis space. Pearl [1986] has pointed out the utility of employing hypothesis hierarchies, and has provided probabilistic interpretations of the father-son relation. However, his belief updating procedure is not suitable for the labeling problem, in which prior knowledge is about spatial interactions between image locations.

(2) How to incorporate *a priori* spatial knowledge with visual observations. Poggio [1985] has proposed to use coupled Markov Random Fields (MRF) to integrate disparate visual cues. His proposal is hard to realize because it requires the knowledge of underlying interactions among various physical processes. It is not clear whether such knowledge is available and how to incorporate statistical data.

(3) How to compute the optimal solution given a complex *a posteriori* probability distribution of the possible solutions. The existing stochastic simulation [Geman and Geman 1984] [Marroquin 1985] and iterative relaxation [Besag 1986] methods have proven inadequate because of their expensive computations and unpredictable results.

(4) How simultaneously to reconstruct and segment the surfaces in a three-dimensional scene using sparse depth measurements and intensity observations. Gamble and Poggio [1987] have demonstrated some good

results on this subject. However, statistical knowledge about the relation between intensity discontinuities and depth discontinuities cannot be incorporated in their scheme. Moreover, their solutions for surface depth and discontinuities are computed with an expensive, interlacing procedure.

This thesis provides an answer for each of these four open questions. There are thus four main contributions.

(1) Consistent and coherent integration of early visual observations on hierarchically structured label trees. A computationally simple procedure and its probabilistic justification, based on well-specified assumptions, are provided [Chou and Brown 1987b] [Chou and Brown 1987a].

(2) Successfully incorporating *a priori* spatial knowledge encoded as potential energy functions that determine the distribution of the corresponding MRF. The *a posteriori* probability distribution is thus derived by combining the pooled external observations and the *a priori* distribution [Chou and Brown 1987b] [Chou and Brown 1987a].

(3) A robust and efficient estimation method for solving the labeling problem based on the *a posteriori* probability distribution. The Highest Confidence First (HCF) method appears to meet the principles of graceful degradation and least commitment. It is a sequential deterministic calculation whose running time is predictable and small [Chou and Raman 1987] [Chou, Brown, and Raman 1987].

(4) Development of a unified treatment of reconstruction and segmentation of three-dimensional surfaces with sparse depth observations and intensity discontinuity information. HCF is extended to handle both symbolic and numerical labels simultaneously using coupled MRF's [Chou and Brown 1988].

1.1. The Labeling Problem

Many visual computations can be formulated as instances of the image labeling problem. In those computations, a finite two-dimensional region R, corresponding to the (retinotopic) projection of the three dimensional scene, is partitioned into a set of *sites*. The sites are considered non-overlapping in general. Usually, the sites are predetermined, independent of the scene, and regularly structured. They are of the same size and shape, with the identical spatial neighboring adjacency except for those near the boundary of R. A typical example is the rectangular array of picture elements (pixels) in an digital image. (There are situations, usually in higher-level computations, such as semantically labeling regions, in which the image partition is *a posteriori*, dependent on some scene properties such as texture and color -- such sites are irregularly structured.)

There is a set of labels associated with each site. The labels can be numerical, representing measurements of some underlying continuous-valued variables such as intensity, or symbolic, representing some underlying qualitative properties of the three-dimensional scene such as the types of the surfaces in view. A labeling is an assignment of labels to the sites: one label, of the corresponding label set, per site. A premise of the labeling problem is that there exists a correct labeling that truly reflects the underlying scene properties. The goal is to recover the correct but unknown labeling.

Intensity image restoration is a well-known instance of the labeling problem. The idea is to recover the "true" intensity measures of the pixels of a discrete image. The image measures are usually corrupted by (perhaps known) sampling, quantization, and random noise processes. The recovery of the true intensities would enable the inference of the amount of light emitted from the underlying surfaces in the scene - an important cue for understanding the scene structure. Usually, the labels are the available gray scales of some imaging device and the sites correspond to the image pixels. It might be desirable to use finer or coarser scales in measurement or spatial resolution, but the underlying problem remains the same.

Besides the corrupted image measures, inexact *a priori* knowledge such as "nearby pixels tend to have similar intensities" can be used as another source of knowledge that contributes to the solutions of the problem.

The labeling problem can be defined with respect to a general graph for irregularly structured sites. Each node of a graph corresponds to a site, and usually has unordered, symbolic labels. The links represent the direct interactions, spatially or causally, between the sites. For example, in polyhedral line labeling, the sites correspond to the vertices (junctions) with multi-variate labels. The number of dimensions of a label is identical to the degree of the vertex. Each dimension represents the possible interpretations of a line (e.g. concave, convex, shadow) attached to the vertex. There has been much effort in analyzing two types of (exact) knowledge that constrain the possible solutions to a manageable number [Huffman 1971] [Clowes 1971] [Waltz 1975] [Kanade 1980]. First, local external evidence provided by the shape of the junctions (arrow, fork, etc.) initializes the sites to a relatively few possible labels. Second, another form of *a priori* knowledge, the line-coherence rule, which states that no line may change its interpretation between vertices, further constrains the corresponding label values to be matched at the two ends of a line.

Other interesting instances of the labeling problem include low-level line finding, intermediate-level region growing (e.g. [Feldman and Yakimovsky 1974]), and high-level object recognition (e.g. [Cooper and Hollbach 1987]). One major difficulty involved in solving labeling problems is the combinatorial complexity of solution spaces. The central issue is how to search for solutions intelligently without enumerating all possible solutions. Iterative constraint propagation and relaxation labeling procedures have been developed and applied to these problems [Waltz 1975] [Mackworth 1977] [Mackworth and Freuder 1985] [Hinton 1977] [Freuder 1978] [Cooper 1988] [Swain and Cooper 1988]. These procedures do not always result in unique labelings but eliminate the ones that conflict with the constraints. It may require further processing, often equivalent to tree search, to

identify a labeling that best interprets the pictures. It is interesting, as noticed by Waltz and many others, that pairwise consistency of vertices usually results in global consistency. Many labeling problems are concerned with discrete choice of exactly one label per site. However, continuous (probabilistic) relaxation methods have also been developed that can assign a degree of confidence to each label for a site [Davis and Rosenfeld 1981] [Hummel and Zucker 1983]. We will discuss them further in Chapter 3.

We have seen that labeling is a general image understanding paradigm and that there is a large diversity of domains and their corresponding knowledge characteristics. However, all instances of the labeling problem are amenable to the treatment developed in this thesis. All require representations of knowledge, reasoning procedures that combine bodies of knowledge, and inference rules that choose solutions among the possibilities. In general, two sources of knowledge are available for solving the labeling problem: the *external evidence*, which relates some image measurements to the labels, and the *a priori knowledge* about the interactions between the labels of adjacent sites. This thesis represents knowledge by statistical measurements. Bodies of knowledge are combined in a Bayesian formulation, and solutions are estimated based on the resulting *a posteriori* distributions using a new, deterministic procedure. The two sets of applications described in Chapter 5 and Chapter 6 use a retinotopic representation of images. However, the proposed framework can be extended to problems using other graph representations as well.

1.2. Early Visual Modules and Image Observations

In this thesis, early visual computation begins with a set of independent modules each extracting a particular scene property. There are variety of motivations for this design including practical considerations and biological discoveries. Modularity has been recognized as one of the most important concepts in problem solving: complex problems are broken into manageable subproblems. Each subproblem is independently studied and solved. A module logically corresponds to a piece of knowledge and machinery for solving a particular subproblem. Together, the

modules cooperatively solve the global problem. A basic open question, and the one addressed here, is how the modules cooperate to produce a single answer.

Interestingly, biological vision systems appear consistent with the modularity concept. Visual cues such as form, color, and spatial information are processed along separate pathways in the brain [Cavanagh 1987] [Livingstone 1988]. Removing one of the cues does not dramatically reduce the visual ability.

It is important to understand how the visual modules interact from the computational point of view. More precisely, the following questions concerning visual integration need to be addressed.

(1) At what stage of the visual computation should the integration take place?

(2) What representations of visual information are suitable for integration?

(3) What is the integration mechanism?

These questions cannot be separately answered, and have become the focus of much research on robot and human vision systems. Some relevant work is reviewed in Chapter 2.

This thesis assumes the existence of a set of independent early visual modules. The outputs, or opinions, of the modules comprise the external evidence for solving the labeling problem. Each module represents a particular piece of knowledge that relates some types of image features to some labels of interest. For example, an intensity edge module may compute a measure of edge strength for each site according to the intensity variation of the image area corresponding to the neighborhood of the site. The edge strength measures are then used to support or refute the hypotheses (labels) concerning the presence of intensity discontinuities. The common characteristics of the early modules are:

Local computation: Modules' opinions as to the label of an individual site are based on local image features. The definition of "local" is in terms of site connectivity, which can encode either geometric or logical relationships.

Probabilistic representation: Each opinion is encoded as a likelihood ratio representing the degree of confirmation or refutation of a particular label.

A procedure that combines early modules' opinions using a label tree is presented in Chapter 3. Organizing labels as hierarchically structured trees allows a particular piece of knowledge to be presented at an appropriate level of abstraction. In this way, early modules corresponding to expert knowledge about certain subsets of labels can be built independently of each other. The evidence combination procedure consistently and coherently pools the early modules' opinions about the labels, and the pooled opinions can be combined with the other source of imperfect knowledge in an abstraction hierarchy whose higher levels may correspond to physically or semantically meaningful objects as described next.

1.3. Prior Knowledge for Inverse Problems

The labeling problem is an inverse problem in that the goal is to recover the properties of a three-dimensional scene from its two dimensional projections. Since much information is lost in the 3D to 2D projection process, the problem cannot be solved, in the sense of well-posedness, without the use of *a priori* knowledge about the scene [Poggio, Torre, and Koch 1985]. A problem is well-posed if a unique solution exists and depends continuously on the initial data [Hadamard 1923]. Incorporating *a priori* knowledge "restores" the well-posedness by, for example, imposing restrictions on the possible solutions or constraining the solutions to have particular statistical properties.

Bodies of *a priori* knowledge are available at various levels of specificity and certainty. The success of an intelligent system largely depends on its ability to retrieve the relevant knowledge in particular circumstances. Therefore, issues concerning organization and representation of *a priori* knowledge are of vital interest to computer vision. As mentioned previously, the labeling problem is the abstraction of many early visual computations. A commonly accepted characteristic of the early computations is that they do not use domain dependent knowledge, but only natural constraints about the world such as surface smoothness, continuity, and rigidity.

[Brown 1984] [Poggio, Torre, and Koch 1985]. Such constraints are valid for most environments, thus the results can be passed to many higher-level domain specific tasks. There is an argument that such constraints may be innate in organisms that have evolved under a basically unvarying set of physical laws and a slowly-varying environment.

Natural constraints may conveniently be encoded, generally speaking, in the form of a global goodness measure. A better (more consistent with the constraints) solution is the one with a higher goodness measurement. This measurement has been expressed as "regularizing functional" [Poggio, Torre, and Koch 1985]. On the other hand, constraints such as smoothness and continuity are actually local geometrical properties; they are about the local interactions between the labels of nearby sites. Thus measurements of such properties may contribute to the global goodness measure. Some previous work on exploiting natural constraints is reviewed in Chapter 2.

The thesis uses a probabilistic - Markov Random Field - encoding of the *a priori* knowledge for the labeling problem. The noncausal Markovian interactions nicely capture the essence of natural constraints in that the local conditional probabilities measure the local smoothness or continuity while the global joint probabilities describe the plausibility of the solutions. In Chapters 5 and 6, bodies of qualitative knowledge such as line coherence and surface smoothness are successfully integrated in experimental applications, via the specification of several quantitative parameters. The success is partially due to the Hammersley-Clifford theorem that relates measures of local potential energy to global joint probability distributions of MRF's.

The probabilistic encoding also allows a Bayesian approach to the integration of the *a priori* knowledge with the external evidence. The resulting *a posteriori* distribution, also a Gibbs distribution characterizing an MRF, forms the basis for estimating the true labeling of the problem.

1.4. Cooperative Networks and Estimation

The computation of the label estimates can naturally be mapped onto a network consisting of a set of cooperative computing units. Each unit corresponds to a site, and makes decisions about the labels in accordance with the external input - the pooled external opinion concerning the site, and the outputs (current decisions) of its neighboring units. The hope is that the network will eventually stabilize to a configuration corresponding to the correct solution. Such networks, reminiscent of interconnected neurons and sometimes referred to as connectionist architectures, have brought many new insights to the field of computer science, especially artificial intelligence [Marr and Poggio 1976] [Feldman 1982] [Feldman and Ballard 1982] [Ballard 1984] [Feldman 1985] [Ballard 1987b] [Feldman et al. 1988]. In fact, the choices of units' functions and interactions for connectionist networks are very flexible [Feldman and Ballard 1982]. The computation of the labeling problem described here can be considered as a specialization of connectionist nets. The utility of using a large number of simple computing units is not only the exploitation of massive parallelism but also a conceptually different way of modeling computations involving large number of variables (as in the case of the labeling problem).

The global behavior of a network depends on, in addition to the units and connections, the inputs, the initial configuration, and the "firing" pattern. To study it, it is convenient to define a global functional of the units' states in the way that the correct solution corresponds to an extremum, say the minimum, of the functional. Then the problem becomes whether the network will evolve to the configuration corresponding to the correct solution from the initialization. Unfortunately, for most interesting problems, "good" global functionals that discern possible solutions by the functional values are difficult to find. Ideally, one would like the functional to be a cost measure, depending continuously and smoothly on the variables and reflecting how far a configuration is from the correct solution. In practice, the absolute and relative goodness of solutions is difficult to decide *a priori*. Moreover, even if such a functional can be found, it is usually very complex, with many local extrema. The

computing units are thus required to perform complicated multidimensional optimization computations, such as stochastic simulated annealing, to achieve global optimality.

For our problem, there are two related questions. First, what is an appropriate goodness measure for the labelings, based on their *a posteriori* probabilities? Second, how do we effectly compute this goodness measure? The thesis answers these two questions simultaneously with a new method - Highest Confidence First estimation. Several basic ideas make this estimation distinct from previous approaches:

Optimality and Computation: The optimality of the solutions defines the computation procedure and *vice versa*. The algorithm attempts to optimize the functional globally by using a local goodness measure to guide the computation. Still, the network behavior is clearly defined, dependent only on input data.

Parallel vs. Sequential: The thesis shows that the units can make better decisions, at least for the labeling problem, by following a global and dynamic ordering that decides which unit to fire next. As a consequence, the computation is intrinsically sequential rather then parallel, as in most connectionist computations.

Commit or not Commit: The estimation procedure introduces a new dimension, "uncommitted", to the unit labels. That is, a unit can decide to commit to a particular label, or not to commit to any one. This additional dimension allows the units to make better judgements based on local information.

HCF is a general technique that it can be used as a heuristic search strategy for any large state-space optimization problems. It has been demonstrated to find good results, both qualitatively and in terms of quantitative energy minimization, in our image-understanding experiments. It would be interesting to see it applied in other contexts.

1.5. Thesis Outline

The thesis is organized so as to be best read in order, with the exception of Chapter 2. Chapter 2 is a review of previous research on several related topics. Its goal is to provide the reader with the background and motivation of the thesis research. However, relevant previous work is also discussed in the technical chapters whenever necessary. Skipping Chapter 2 should not affect the understanding of the theme or results of the thesis.

Chapter 3 presents a novel approach to knowledge representation and reasoning for the labeling problem addressed by the thesis. A hierarchically structured label tree is used to accrue external evidence concerning the labels for each site. A probabilistically justifiable procedure combines distinct bodies of evidence, represented as label likelihood ratios in the tree. The combination is commutative and associative, and has a simple message-passing implementation. More importantly, this procedure accumulates external evidence in terms of likelihood ratios rather then *a posteriori* probability distributions. This feature enables the integration of the *a priori* knowledge, encoded in terms of a joint probability distribution of all sites, with the pooled external evidence in a Bayesian formalism. The utility of Markov Random Field modeling is also discussed. The chapter is concluded with the derivation of the *a posteriori* distributions of the labelings on which the solutions are based.

Chapter 4 presents the Highest Confidence First estimation algorithm. Practical concerns are the primary motivation of this novel inference method. Previous approaches based on MRF's using stochastic simulation or deterministic iterative techniques have been observed to be computationally expensive and unpredictable. Also their results tend to be affected significantly by the large-scale characteristics of MRF's and the possible error involved in *a priori* models. This chapter discusses the reasons why HCF is a better alternative. A priority-heap implementation of HCF is presented, and its convergence properties are discussed.

Chapters 5 and 6 present two case studies of the labeling problem. The boundary detection problem, still interesting and unsolved after twenty years of vision research, is dealt with in Chapter 5, using the formulation presented in Chapter 3 and 4. A novel aspect of this work is the use of an MRF to model an explicit line process, and the use of outputs from a set of local edge operators for the external evidence. Experimental results using several estimation methods are compared, and HCF clearly outperforms the others in both robustness and efficiency.

Chapter 6 deals with the problem of surface reconstruction by integrating intensity information with sparse depth measurements. Coupled MRF's provide a unified treatment of reconstruction and segmentation, and an extension of HCF successfully implements a solution method. Experiments are shown to demonstrate the effectiveness of the approach. This work uses not only the results of Chapters 3 and 4, but also the boundary detection package described in Chapter 5.

Finally, Chapter 7 summarizes the thesis and sketches some future research directions.

Chapter 2 Related Research

2. Related Research

This chapter provides an intellectual background for the thesis research with selected references. More detailed references are given in technical sections. We focus on previous research on the three topics dealt with in this thesis. First, we provide an overview of visual integration. We are particularly interested in the computations of consistent intrinsic images. The large literature on multi-sensor data fusion applications (e.g. [Brooks, Flynn, and Marill 1988]) is not discussed, although the reader may find its subject matter somewhat related to that of this thesis. Second, we review previous work on regularization. We review the uses of variational principles and Bayesian estimation, and compare mechanical surface models and Markov random field models. Last, we discuss computation algorithms and architectures, especially for energy minimization problems.

2.1. Visual Integration

Visual integration is, generally speaking, the process of building an intermediate representation of visual information for higher-level tasks, such as recognition and obstacle avoidance, using results from early visual processing. Typical intermediate representations are Barrow and Tennenbaum's intrinsic images [1978], Marr's 2-½D sketch [1982], and Feldman's stable feature frames [1985]. The contents of the intermediate representations are the "intrinsic" scene parameters. Surface discontinuities, range, orientation, reflectance, texture, and motion velocity are all spatially indexed properties of the scene. Since some low-level vision modules provide intrinsic images themselves, the task of integration is to combine outputs from several low-level modules to produce improved results.

Much vision work for the past decade has been focused on early processing [Ballard and Brown 1982] [Aloimonos 1986]. An early processing module computes intrinsic parameters from one characteristic of the input intensity arrays. Examples are the computation of optical flow, shape from shading, shape from texture, structure from motion, binocular stereo fusion, and boundary detection. Since individual problems are essentially ill-posed, "natural" constraints such as

smoothness are often imposed to ensure unique and well-behaved solutions can be obtained. Sometimes, intrinsic parameters are interrelated according to physical laws. Often, however, choices of the constraints are decided by computational considerations rather than the physical plausibility of the constraints themselves. One important problem that arises in visual integration, and a problem addressed in this thesis, is how to ensure consistency among the intrinsic parameters.

Loose Coupling Vs. Tight Coupling:

Barrow and Tennenbaum [1978] stated that the intra- and inter-image constraints can be satisfied simultaneously by adjusting the parameter values with local operations in parallel. Ballard [1984] proposed to use a multi-level connectionist network that computes the intrinsic images simultaneously along with non-spatially indexed features such as light source directions, in a parallel-iterative manner. The interactions between individual modules and their outputs were loosely coupled by the connections.

Similar to this line of thought, Poggio [1985] sketched a computational model for computing intrinsic images by using coupled MRF's and stochastic estimation techniques. The early visual modules in his design are loosely coupled, in the sense that each module operates on a particular cue independent of the others, and the intra-image consistency is maintained through the coupling of MRF's. Practical concerns are the motivations for loosely coupled designs. In the absence of certain visual cues, a visual system should still work using the available cues. This sort of robustness is a characteristic of loosely-coupled systems.

Based on a different philosophy, Aloimonos [1986] studied the physical and mathematical relations between intrinsic parameters and multiple image features such as shape from shading and motion. The general idea is the use of multiple image cues to setup intrinsic parameter computations that robustly yield unique answers. As a consequence, less *a priori* knowledge needs to be assumed in the computation, and the resultant intrinsic images are guaranteed to be consistent with multiple image cues. Similar ideas have appeared in the work on computing surface structure by

coupling motion and stereo modules [Waxman and Duncan 1985] and stereo and shading modules [Grimson 1984]. In this work, the individual modules are tightly-coupled by being part of a single mathematical formulation. When all the data is available, the computation is robust. When data is missing, the computation cannot be done.

The work described in this thesis is in a loosely coupled style. This thesis decouples the notion of *a priori* knowledge from observable visual evidence. Early modules perform only local operations: relating local visual observations to labels of sites. No inter-image consistency is assumed. Thus individual modules do not impose spatial constraints (such as smoothness) in their computations. The outputs of early modules are combined by a method based on Bayesian probability theory. Spatial *a priori* knowledge is incorporated and inter-image consistency is maintained with respect to a particular labeling problem by using MRF modeling and HCF estimation. Thus the *a priori* knowledge is applied as part of global decision making process, rather than at a local level by low-level modules.

Coordinate Systems

The choice of coordinate systems for representing intrinsic parameters is also an important issue for visual integration. Since the intermediate representation bridges early vision and late vision, the choice of the representation depends not only on how the early processes compute, but also on how the representation is used.

Marr's 2-½D sketch [1982] is viewer-centered. He argued that the results of early processes are combined in some kind of retinocentric frame because they are delivered in this form and that it is consistent with the capability of a fovea. Feldman [1985] took eye-movements into account. His stable feature frame is also viewer-centered, but is aligned with respect to the observer's head position. Retinotopic information can be integrated over saccades using the knowledge of gaze. Thus the descriptions of a scene would appear stable to the observer.

There exist some efforts in computer vision that integrate information over camera movements [Ayache and Faugeras 1986] [Matthies, Szeliski, and Kanade 1987]. The underlying fusion mechanism used by them is the Kalman filtering. Ayache *et al* used a feature-based stereo system mounted on a mobile robot to compute a sparse map (position and orientation) of features with respect to each robot's position. The transformation between two frames is estimated by matching overlapping features. Such information is used to refine the feature estimates in each map. Matthies *et al* also used Kalman filtering to integrate new depth measurements and reduce uncertainty over time, in a pixel-wise retinotopic representation. In their setup, the camera is controlled to move laterally at fixed speed. Depth measurements at each pixel at a certain time can thus be predicted and updated in accordance with the camera motion, the previous estimate, and the new depth measurement.

Recently, Ballard [Ballard 1987a] has advocated that the representation of the products of early vision is best expressed in a fixation frame of an eye-movement system. Fixation frames are object-centered but viewer-oriented. He argued that an active visual system can represent the calculations more correctly in object-centered coordinates, and thus spatial relations of objects can be encoded as transformations between frames.

Although we use a retinotopic frame in our case studies, the thesis framework is by no means limited to this choice. We believe that different tasks may require different coordinate systems. For example, object-centered representations ease the task of object recognition, but are awkward for the obstacle avoidance problem, for which view-centered representations may be more suitable.

2.2. Regularization: Mechanical and Probabilistic Models

Regularization is used as a general term for any method to make an ill-posed problem well-posed [Poggio, Torre, and Koch 1985] Early vision, as mentioned previously, is an inverse problem: given the projection process P and the image observation O, recover the state of the scene S such that $O = P\ S$. In general, the observation O does not determine S in a unique and stable way, due to the loss of

information in the three- to two-dimension projection and noise corruption of the observation during the imaging process. To solve such problems, it is customary to impose constraints on the solutions to ensure uniqueness and stability. That is, to regularize the otherwise ill-posed problems.

One question arises, that is: what constraints are physically plausible and yet powerful enough to regularize a problem? Imposing strong (domain specific) constraints would certainly ease the solutions to inverse problems, but the generality of the solutions would be lost. Much research has exploited natural (generic) constraints for early vision. There are basically two approaches to regularization that employ natural constraints. One approach uses variational principles to restrain the possible solutions. Typically, a norm and a stabilizing functional for the solutions are chosen so that the most "regular" (stable and consist with input data) solution can be computed. The stabilizing functionals used so far are mainly the linear combinations of the first few derivatives of solutions, encoding various degrees of continuity and smoothness [Horn and Schunk 1981] [Ikeuchi and Horn 1981] [Grimson 1981] [Hildreth 1984] [Terzopoulos 1986a] [Terzopoulos 1986b] [Blake and Zisserman 1987]. Using such stabilizers corresponds to fitting splines to the data.

For example, in the context of surface reconstruction, Grimson [1981] and Terzopoulos [1986a] model smooth surfaces with membranes and thin plates. A membrane, being characterized by first derivatives, is sensitive to depth variations but insensitive to changes in surface orientation. On the other hand, a thin plate bends but cannot crease, because it has a second order energy (function of second derivatives). Observing this, Terzopoulos proposed to fit membranes at locations of orientation discontinuities and plates in homogeneous regions. Although depth and orientation discontinuities are assumed to be known in this work, Terzopoulos [1986b] pointed out that discontinuities can be detected as locations of abnormally high strain in the modeled surface. Blake *et al* [1987] argued that the primary purpose of surface reconstruction is to detect discontinuities. They proposed "weak" surface models to incorporate discontinuities in surface energy formulations. Using weak models, a surface breaks if that reduces the total energy. Discontinuities can

thus be computed in the process of surface reconstruction.

The other approach to regularization, the one used in this thesis, is based on Bayesian estimation and Markov Random Field models. The world S is modeled as a random field, and the observation O is a set of noisy measurements. The *a priori* knowledge about the world is expressed in the form of a probability distribution. Using this distribution together with a probabilistic description of the noise that corrupts O, the *a posteriori* distribution of the world can be derived. The inverse problem is solved by finding an optimal estimate based on some statistical criterion, such as maximizing the *a posteriori* probability (MAP) or minimizing the expected value of some cost measure (see Chapter 4).

There are two reasons for using Markov Random Field models: 1. Spatial *a priori* knowledge can be naturally expressed in terms of local conditional probabilities or local potential energy measures (see Chapter 3). 2. It is powerful. All nondegenerate random fields are MRF's with respect to different neighborhood systems. Natural constraints about smoothness and continuity can usually be encoded adequately with small neighborhood systems.

Geman and Geman [1984] presented an excellent treatment of MRF models. The two important contributions of their work are the MAP estimation procedure using simulated annealing optimization and the use of coupled MRF's (one for the intensity process and one for the implicit discontinuity process) for intensity image restoration. Marroquin [1985] pointed out that for some problems MAP estimation may not be desirable, based on Bayesian decision theory. For example, for the labeling problem, the Maximizer of Posterior Marginals (MPM) is a better estimate than MAP because it minimizes the expected value of the number of mislabeled sites. He also described a Monte Carlo procedure for computing MPM.

Interestingly, the mechanical (e.g. membrane and plate) models are not irrelevant to the MRF models. It can be proven that fitting splines over discrete lattice sites is equivalent to MAP estimation using MRF models with appropriate neighborhood systems and potential functions [Szeliski 1987]. But, the probabilistic

approach is more flexible. One can easily adjust one's priors by modifying the local potential functions. Furthermore, the probabilistic approach allows the development of other, perhaps better, estimates such as MPM and HCF.

Other advantages of the probabilistic approach include the well-known semantics of model parameters and estimates, the easy incorporation of statistical data, and the ability to integrate disparate sources of information.

2.3. Computation: Energy Minimization and Relaxation

Many problems can be formulated as either optimizing an objective function or satisfying a set of constraints or both [Feldman and Yakimovsky 1974] [Hummel and Zucker 1983] [Kirkpatrick, Gelatt, and Vecchi 1983] [Feldman 1985] [Poggio, Torre, and Koch 1985]. Usually, brute-force exhaustive search for solution is computationally prohibited because of the large number of variables involved in a problem. Furthermore, some problems are intrinsically difficult. For example, the well-known polyhedral line labeling is proven to be NP-complete [Kirousis and Papadimitriou 1985]. To "solve" such problems, one must exploit possible parallelism in the computation or settle for approximation methods that provide "good" solutions most of the time.

One class of problems is relatively easy to solve: the minimization of convex functions. Convex functions are well-behaved in the sense that they have unique minima when bounded from below. Any procedure that keeps searching for solutions with lower function values is guaranteed to find the optimal answer. Therefore, iterative relaxation techniques can be developed (e.g. [Terzopoulos 1983] [Terzopoulos 1986c]). At any instant, each variable is considered to change its value depending on whether the change would result in a smaller function value. Such techniques can be mapped onto parallel network architectures; each computing unit corresponds to a variable, and the connections depend on the interdependencies among the variables. The problem of surface interpolation is a typical example. In a finite element formulation, Terzopoulos [1986a] developed a set of computational molecules for fitting membrane and thin plate surfaces, all having quadratic energy,

over an image grid. The computational molecules, representing the connections among the (depth) variables, are small in size, and the computations performed are linear (weighted sum of the variables).

However, most interesting problems are not convex, and computations are not linear. In fact, any network that can make decisions cannot be linear [Blake and Zisserman 1987]. For example, if discontinuities are to be computed simultaneously with surface reconstruction in the above energy formulation, decisions on the presence or absence of a discontinuity must be made at every candidate location. For each discontinuity configuration, there is a corresponding optimal surface configuration. Therefore, there are as many "local" optimal solutions as the number of possible discontinuity configurations. Blake *et al* devised an algorithm that uses a series of functions (decreasing in "convexity") to approximate the original non-convex energy function. Their method was proven to converge to the global minimal-energy solution for a restricted class of energy functions. For example, it requires depth measurements everywhere for surface reconstruction.

Simulated annealing is a general paradigm for solving combinatorial optimization problems [Kirkpatrick, Gelatt, and Vecchi 1983]. The idea is to allow, stochastically, energy to increase during the energy minimization process in order to escape from local minima. It was proven that if the temperature of the annealing system decreases slowly enough, a globally optimal solution is guaranteed [Geman and Geman 1984] [Gidas 1985]. Annealing has been applied to image restoration [Geman and Geman 1984], texture segmentation [Geman and Graffigne 1986] [Simchony and Chellappa 1988], the computation of optical flow [Murray and Buxton 1987], and binocular stereo matching [Barnard 1987].

On a related topic, Marroquin [1985] proposed a Monte Carlo stochastic sampling procedure. The idea is to generate system configurations stochastically in accordance with the likelihoods of the configurations (see Chapter 4). In this way, estimates such as MPM can be approximated by collecting sample statistics. For surface reconstruction, he proposed to use a hybrid network in which discontinuity process (whose state is updated digitally by a stochastic procedure) acts as a set of

switches between the nodes (of an analog network) whose voltages represent the depth process. Hutchinson *et al* [Hutchinson et al. 1988] used a similar hybrid network to compute optical flow. The digital subnetwork for the discontinuity process deterministically updates the line configuration after every run of the analog subnetwork; the analog subnetwork computes the optimal flow configuration according to the current line configuration. This interlacing relaxation procedure is guaranteed to converge. However, the results might not be globally optimal.

Building an analog network for optimizing a quadratic function, as in surface reconstruction, is relatively easy. The problem of constructing an analog network that makes decisions is somewhat more difficult [Hopfield and Tank 1985] [Koch, Marroquin, and Yuille 1986]. Hopfield *et al* [1985] studied the classic traveling salesman problem. They formulated this discrete decision problem with a continuous energy function in which valid and short paths are favored. They designed an analog network of which the stable states are locally optimal with respect to the energy function. Their idea was to allow the binary variables to vary continuously between 0 and 1 and to introduce terms in the energy function that forced them in the final solution to be 0 or 1. Koch *el al* [1986], following Hopfield's work, described an analog network for surface reconstruction.

Analog networks, in contrast to stochastic optimization methods, provide solutions instantaneously. However, as one might expect, there are difficulties using analog networks to solve complex problems. It is difficult to ensure solutions are valid (0's and 1's) and globally optimal (with respect to the original discrete problems) [Wilson and Pawley 1988].

The Highest Confidence First estimation described in this thesis can be used to compute decisions in optimization problems. It is deterministic and intrinsically serial. It searches for a solution in an augmented space rather than in the solution space (as in simulated annealing) or the interior of a hypercube (as in using analog networks). Only local optimality is guaranteed. This local optimality might be desirable because, from an engineering viewpoint, only simple tasks have cost functions that truly reflect the goodness of solutions. Usually cost functions involve

many subjectively chosen weightings of the various contributions. Globally optimizing such functions may ignore important information conveyed in observable data. HCF attempts to find solutions most consistent with input data while satisfying *a priori* constraints (Chapter 4).

Chapter 3 Probabilistic Information Fusion

3. Probabilistic Information Fusion and the Labeling Problem

Most research on evidence combination has focused on updating the "belief" in a given hypothesis about an individual entity when bodies of new evidence become available [Pearl 1986] [Shafer 1976] [Reynolds et al. 1986]. When dealing with problems involving large number of entities with local interactions (e.g. the labeling problem), the approach - maintaining "marginal belief" - requires the effects of updating local "belief" to be propagated to other sites. It is interesting to compare the class of algorithms called "probabilistic relaxation", also aiming at the labeling problem, with the approach proposed in this thesis.

Probabilistic relaxation was inspired partially by the discrete relaxation labeling and partially by the Bayesian-probability formalism [Rosenfeld, Hummel, and Zucker 1976] [Davis and Rosenfeld 1981] [Hummel and Zucker 1983] [Shvaytser and Peleg 1985] [Peleg 1980] [Kohler 1984]. In their formulation, each site has a set of weights attached to the labels. The weights are nonnegative and sum to unity, reminiscent of a probability distribution of the labels, and thus the name "probabilistic" relaxation was adopted. The *a priori* contextual knowledge, being inexact, is about how compatible are pairs of labels associated with adjacent sites, usually represented in terms of conditional probabilities or statistical correlations. A local updating operator computes how much supporting evidence provided by the neighbors in accordance with their weights and compatibilities, and adjusts the label weights accordingly. Ideally, the weights evolve in a local parallel fashion, and eventually converge to a configuration when a label with high weight emerges and is considered as the correct label of the site.

There are two drawbacks with the use of probabilistic relaxation for solving the labeling problem with multiple sources of knowledge. First, the updating rules are rather ad hoc and heuristic. It is difficult, if not impossible, to interpret the weights after a couple of iterations from a Bayesian decision point of view [Kittler and Foglein 1986] [Haralick 1983]. Experimental evidence also suggests that only the first few iterations "improve" the initial assignments of the weights. Subsequent iterations tend to lead the results away from the observations. Consistent

combination of observable evidence and contextual knowledge is thus impossible under this formulation. The second drawback is the computations required for propagating the effects of bodies of new evidence. A body of evidence bearing directly upon an individual site influences the belief distributions of the rest of sites. Usually many iterations through the sites are necessary (if possible) before converging to a state in which the set of marginal beliefs becomes stable. We believe that an information fusion mechanism should constantly maintain a representation of knowledge to reflect the total information available, except possibly for transient periods of time for aggregating evidence locally. This requirement is also desirable, if not necessary, for implementing vision systems in distributed environments.

Instead of updating marginal beliefs, our approach maintains joint probability distributions of the sites by decoupling the notion of external evidence and *a priori* knowledge. Since bodies of evidence based on local image observations bears directly upon individual sites, they can be combined locally without having to interfere with other sites. On the other hand, *a priori* knowledge is mostly about the interactions among the sites. It is best described by a joint probability distribution of all variables. When combined with external evidence using Bayes' rule, the resultant distribution reflects the *a posteriori* belief in the global configurations. Inference methods can thus be applied to find the true labeling based on the *a posteriori* beliefs. A global view of the proposed approach is provided in Figure 3.1.

This chapter proposes the representations for the two sources of imperfect knowledge -- the external evidence and the *a priori* knowledge, the combination mechanism that combines these two sources of knowledge, and the criteria for evaluating the goodness of the estimates of the true labeling based on the available knowledge. Only finite, unordered labels are concerned here. In Chapter 6, we describe a method for reconstructing and segmenting depth maps using both symbolic and numerical labels.

In Section 3.1, we limit our attention to individual sites. We describe the use of likelihoods and likelihood ratios as the representation of external evidence, and a procedure that consistently and coherently aggregates evidence for labels whose

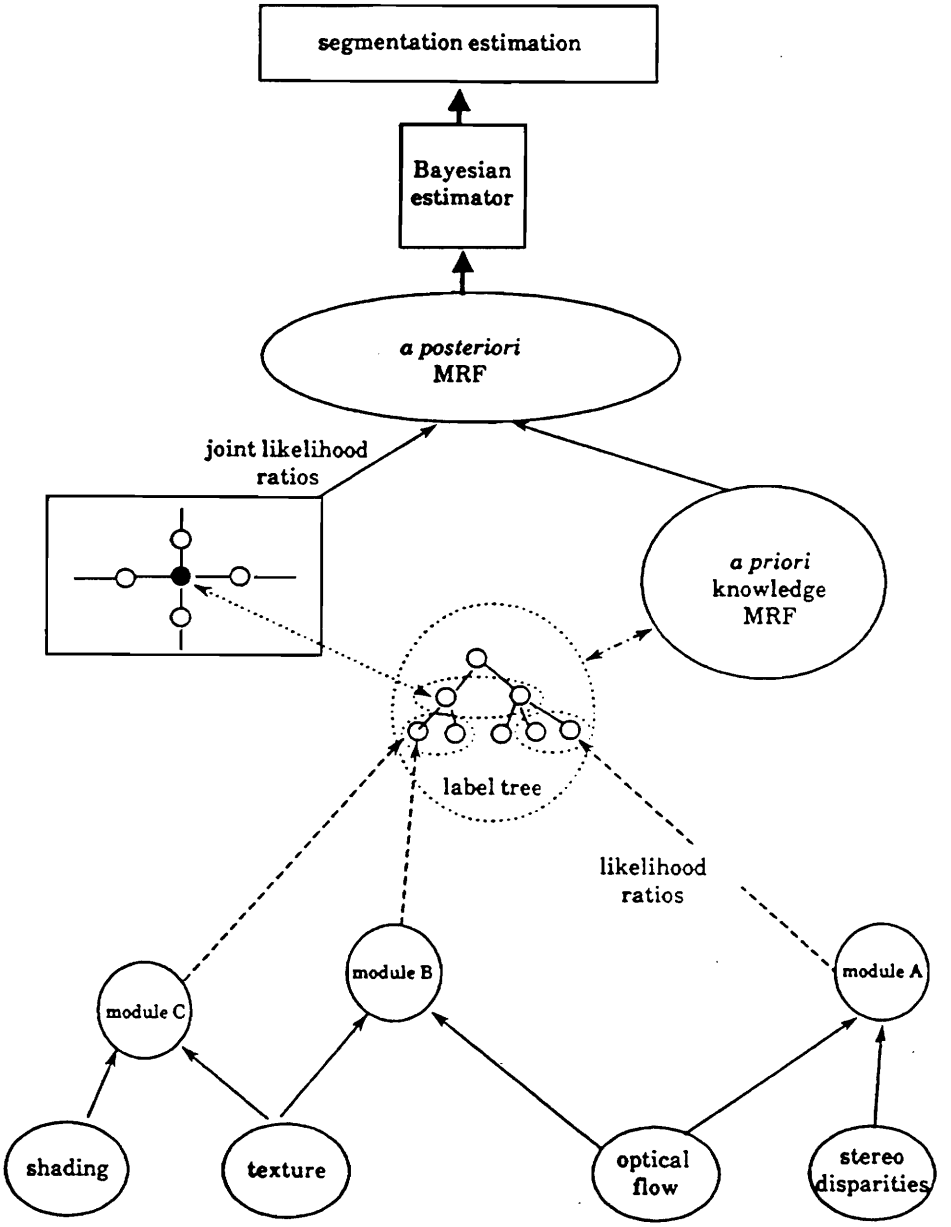

Figure 3.1. System Diagram

semantics are hierarchically related. Markov Random Fields and their properties are introduced in Section 3.2. They are used to quantify the inexact *a priori* knowledge about the spatial interactions between sites. In Section 3.3, the two sources of

knowledge are combined by Bayes' theorem. The resultant *a posteriori* knowledge is thus used to estimate the true labeling using Bayesian decision criteria (Chapter 4).

3.1. Hierarchical Integration of Early Visual Observations

Unordered symbolic labels are commonly used for the high level representations of a scene. Each label corresponds to some event directly or indirectly observable in the scene. Labeling a site has the semantics of hypothesizing the occurrence of the corresponding event (what) at the corresponding location (where). Since the existence and uniqueness of the label assignments are assumed, the corresponding set of events must be mutually exhaustive and exclusive. Frequently, certain subsets of the events have semantic interest. Under judicious treatment, they can be organized as a hierarchically structured tree. Each internal node in the tree represents the disjunct of its son events. The partial ordering defined by the father-son links reflects the causal relations between the events.

Figure 3.2 shows one example of how we may organize the knowledge about various types of image edges. An abrupt change of the image irradiance (edge) may be caused by the variations of underlying scene structure (geometrical edge) or surface reflectance (photometric edge), and the changes of surface reflectance may be due to the variation of surface albedo (e.g. texture edge) or the amount of incident light (e.g. shadow edge). It is desirable to organize various types of edges into such trees. For example, it is absolutely vital to have the finest-level descriptions of the edges for object recognition tasks. However, when obstacle avoidence is the primary concern, only depth discontinuities are of interest. Furthermore, a visual module such as an intensity edge detector might not be able to tell geometrical edges from photometric edges, but might be capable of indicating occurrences of discontinuities of any sort.

One utility of using hierarchically structured trees, as pointed out by Gordon and Shortliffe [Gordon and Shortliffe 1985] and later by Pearl [Pearl 1986], is the ability to represent a particular piece of knowledge at whatever level of abstraction is

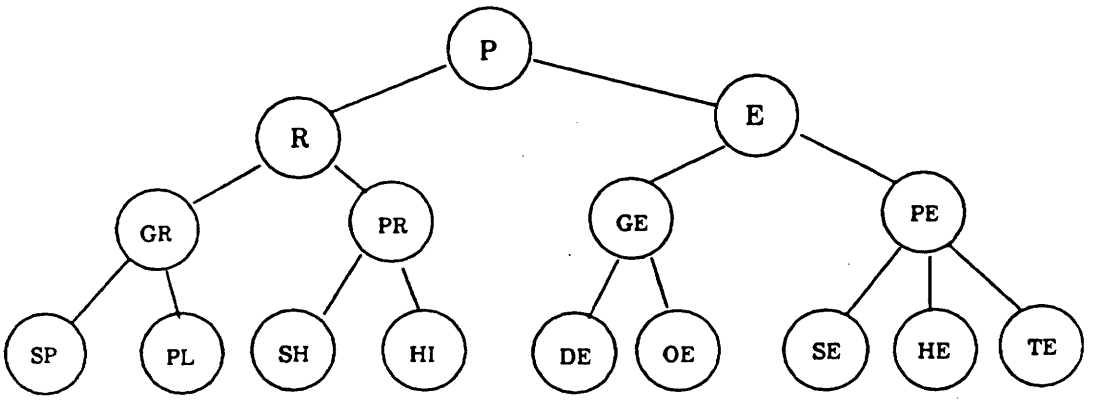

Figure 3.2. A Label Tree
E - edge, R - region, PE - photometrical edge, GE - geometrical edge
PR - photometrical region, GR - geometrical region, TE - texture edge
HE - highlight edge, SE - shadow edge, OE - orientation edge, DE - depth edge,
HI - highlight region, SH - shadow region, PL - planar region, SP - spherical region.

appropriate. Also, in practice, it is easy to independently design and maintain modules that are experts at detecting particular events. Their opinions about the label events can then be pooled according to the known semantic relations. In addition, many visual tasks can simultaneously share the knowledge accumulated in a label tree. Since events on different branches of the tree are mutually exclusive, every cross-section corresponds to a mutually exclusive and exhaustive set of events. Instances of the labeling problem can thus be defined with respect to appropriate cross-sections in a label tree dependent on the particular goals of the tasks.

In this section, we limit our attention to individual sites. We assume that the interesting labels for each site are organized as a tree. We describe a new method, in terms of message passing between nodes of a label tree, for combining bodies of evidence represented as label likelihood ratios. Probabilistic justifications for this methods are also provided.

Some notation has to be introduced here. Represent an image as a set of sites indexed by the set $S = \{s_1, s_2, \ldots, s_N\}$. Without loss of generality, assume all sites

have the same set of interesting labels orgainized as a hierarchically structured tree H (e.g. Figure 3.2). Node H_l denotes the hypothesis that s can be labeled l. Each internal node stands for the disjunct hypothesis of its sons. For convenience purposes, let γ_l denote the set of sons of H_l. Let $L = \{l_1, l_2, \ldots, l_Q\}$ be a mutually exclusive and exhaustive set of labels of H. Let $X_s \in L$ be a random variable associated with $s \in S$. A labeling ω of the image with respect to L is a realization of the set of random variables $X = \{X_s, s \in S\}$. Let $\omega_s = \omega(X_s) \in L$ represent the label attached to s_i according to the labeling ω, and Ω be the set of all labelings -- the admissible solution space of the labeling problem with respect to S and L.

3.1.1. Weighing of External Evidence

We consider early visual computations as the computations performed by a set of independent modules. A module encodes a piece of knowledge that relates image observations to some label events. The input for these modules is noise-corrupted visual data, such as image irradiance, texture, stereo disparities, etc. When making an opinion about site s, a module may restrict its consideration of input to some spatial region dependent on s. Typically, the region will include s and its spatially adjacent sites. Also, a module might not use all the input available for a given site, that is, it might use only irradiance when other data are also available. The subset of input used to make an opinion about s is observation O_s.

In our treatment, the *opinions* of the modules are presented in terms of likelihood ratios. For example, module A is an expert on a label l. After observing O_s, the module reports its opinion on l as a likelihood ratio :

$$\lambda_l = \frac{P(O_s \mid l)}{P(O_s \mid \neg l)}$$

The semantics of likelihood ratios are well-known [Good 1950] [Duda, Hart, and Nilsson 1976] [Bolles 1976]. Confirmation of l based on O_s is encoded by $\lambda_l > 1$, and disconfirmation by $\lambda_l < 1$.

It is possible that a module knows more than one label. In such cases, the module distributes the support of the evidence to the set of labels and its complement.

More precisely, if module A is an expert on the subset $L_A \subseteq L$, it reports one likelihood ratio for each label in L_A. A *likelihood ratio* is the probability of the observation given that one label truly applies divided by the probability of the observation should none of the labels in L_A apply. For example, the likelihood ratio reported for label $l_i \in L_A$ is:

$$\lambda_i = \frac{P(O_s \mid l_i)}{P(O_s \mid \neg(\bigcup_{l \in L_A} l))}$$

Note that we define $P(O \mid \neg(\bigcup_{l \in L_A} l))$ to be 1 when L_A is exhaustive.

The values of likelihoods and likelihood ratios can be derived from stochastic models of relationships between the label events and the observations [Sher 1987a] [Bolle and Cooper 1984] [Bolle and Cooper 1986], or estimated from statistical data [Bolles 1977]. Sometimes they are subjectively assigned by human experts based on their experience [Duda, Hart, and Nilsson 1976]. An example of computing likelihoods for intensity edges based on Sher's work is given in Chapter 5.

For a purpose that will soon become clear, we impose the following assumption of conditional independence between spatially distinct observations:

$$P(O^A \mid \omega) = \prod_{s \in S} P(O_s^A \mid \omega_s) \tag{3.1}$$

where the superscript A indicates the observations of the module A. This assumption has been used implicitly in numerous applications [Besag 1986] [Derin and Cole 1986] [Marroquin 1985] [Bolle and Cooper 1984] [Bolle and Cooper 1986]. The assumption is not always valid. For example, the noise of an ultra-sound image may be spatially dependent given the true scene due to the change in conductance. There are also cases in which O_s may contain information not only from site s but also from its adjacent sites. In such cases, the spatial independence assumption is still valid, but (3.1) needs to be modified according to point-spread functions to take into account the blurring effects. For conceptual and notational convenience, we do not consider these conditions in this thesis.

3.1.2. Hierarchical Aggregation of Evidence

We maintain a number for each node of the label tree H, except for the root. Unlike other approaches, the numbers here do not indicate the states of belief of the hypotheses but rather the degrees of hypothesis confirmation or disconfirmation provided by the collected evidence.

Let α_l denote the current degree of confirmation/disconfirmation for node H_l. The probabilistic interpretations for the α's will be given in the next section. Initially, all α's are set to unity indicating "neither confirmed nor disconfirmed" before observing any evidence. Besides α, each internal node H_l keeps one value, w_i^l, for each node $H_i \in \gamma_l$. Initially, w_i^l is set to the *a priori* probability of $X_s = i$ given $X_s = l$. Obviously, the w_i^l's of each node sum up to unity.

The Bayesian evidence aggregation on the label tree can be best described in terms of local updating and message passing between the nodes of H. Suppose a module A reports its opinion as a set of likelihood ratios $\{\lambda_l^A \mid l \in L_A\}$ where L_A is a set of mutually exclusive labels contained in H described in the previous section. The α's are updated according to the following rules:

Local updates:

For each node H_l where $l \in L_A$, α_l is updated from its current value to

$$\alpha_l \leftarrow \lambda_l^A \alpha_l \qquad \forall l \in L_A. \tag{3.2}$$

The effect of this update has to be propagated throughout the label tree to maintain the coherence of the α's. Node H_l sends a message, $m^+ = \lambda_l^A$, to its father and each of its sons.

Downward propagation:

Any node H_k that receives a message m^- from its father executes the following two steps: (1) passing on $m^+ = m^-$ to each of its sons, and (2) replacing α_k by

$$\alpha_k \leftarrow m^- \alpha_k. \tag{3.3}$$

Upward propagation:

Any node H_j that receives a message m^- from one of its sons (say H_i), updates

w_i^j to

$$w_i^j \leftarrow m^- w_i^j, \tag{3.4a}$$

and sends a message m^+ to its father, where

$$m^+ = \sum_{H_i \in \gamma_j} w_i^j \tag{3.4b}$$

then updates α_j and all the w^j's according to

$$\alpha_j \leftarrow m^+ \alpha_j, \tag{3.4c}$$

$$w_k^j \leftarrow \frac{w_k^j}{m^+} \quad \forall k \in \gamma_j. \tag{3.4d}$$

Note that the combination and propagation procedures are commutative and associative (Appendix 3.A). This property enables an asynchronous parallel network implementation.

3.1.3. Probabilistic Justification

The above method fits in a Bayesian formalism if we maintain two notions of conditional independence. First, a piece of evidence O that bears directly on the hypothesis $X_s = l$ says nothing about the descendants of H_l:

$$P(O \mid l,i) = P(O \mid l), \quad i \text{ descendant of } l, \tag{3.5}$$

$$P(O \mid \neg l,j) = P(O \mid \neg l) \quad j \text{ descendant of } \neg l.$$

Second, the observations of different modules are conditionally independent.

$$P(O \mid l) = \prod_A P(O^A \mid l), \quad \text{and} \tag{3.6}$$

$$P(O \mid \neg l) = \prod_A P(O^A \mid \neg l),$$

where the products on the right-hand side are over a set of modules and O is the union of their observation O^A's.

Equation (3.5) states, as suggested by Pearl [Pearl 1986], that when the observation O^A is a unique property of H_l, common to all its descendants, once we know $X_s = l$ is true/false, the identity of the descendants H_i or H_j does not make O^A more or less likely. Equation (3.6) states that each piece of evidence observed by the early modules provides conditionally independent information about a label. In other words, knowing that $X_s = l$, the observation of O by a module will not make the observation O' of another module more or less likely. We believe that the disparate types of image clues in vision applications satisfy this assumption.

Define *consistent states* of α's as the states in which for each available opinion, all of the α's are either updated according to rules (3.2) to (3.4), or none of the α's have been changed with respect to this opinion. We say that a set of opinions *yields* a consistent state if all opinions in this set, and no other opinions, have been used to update the α's. The following theorem relates the α's to the likelihood probabilities at consistent states.

Theorem 3.1: Let α_l^t denote the α value associated with H_l at the consistent state t, and $P(O_t \mid l)$ be the probability of O_t given $X_s = l$, where O_t denotes the union of those observations that form the set of opinions that yields the state t. If $O_t \neq \varnothing$, then

$$\alpha_l^t = c_t P(O_t \mid l) \quad \forall H_l \in H \tag{3.7}$$

where c_t is a constant depending only on t, given the conditional assumptions (3.5) and (3.6) .

The proof of Theorem 3.1 is given in Appendix A at the end of the chapter. This theorem states that the supports from the pooled evidence for the label events are properly weighted, following the evidence combination procedure of Section 3.1.2. To see how it could be used, we start with an *a priori* distribution P_0 of a set of mutually exclusive and exhaustive labels L. Assume that subsets of L are hierarchically related and organized as a tree H, and the w's associated with the internal nodes are initialized by the corressponding *a priori* conditional probabilities in accordance with P_0. For example, if $i \subset L$ and $j \in i$, w_j^i can be computed by

$\frac{P_0(j)}{\sum_{k\in i} P_0(k)}$. The following corollary demonstrates the utility of the theorem.

Corollary 3.2: The posterior probability $P_t(l)$ of $l \in L$ at the consistent state t is

$$P_t(l) = \frac{\alpha_l^t P_0(l)}{\sum_{l\in L} \alpha_l^t P_0(l)}.$$

As mentioned previously, it is difficult to maintain marginal belief for each variable associated with the sites. Corollary 3.2 suggests that it is possible to combine bodies of evidence according to the *a priori* probabilities. The resulting combination can then be merged with the original *a priori* probabilities.

To summarize: We have developed an evidence combination method for a hierarchy of hypotheses based on the notions of conditional independence given by equations (3.5) and (3.6). This scheme, besides having all the desirable characteristics listed in [Pearl 1986], has the following advantages:

(1) The computations involved are extremely simple. Simpler and fewer messages must be passed, comparing with Pearl's procedure [1986]. Normalizations are never needed since relative degrees of confirmation/disconfirmation are maintained instead of probabilities (Theorem 1).

(2) This scheme decouples the notion of evidence and belief. That is, the evidence can be collected and combined consistently and coherently without having to maintain probability distributions. In the next section we show this characteristic is very helpful when the prior knowledge is represented as an MRF.

3.2. Spatial Priors and Markov Random Fields

Markov Random Fields have been used for image modeling in many applications for the past few years [Besag 1974] [Hassner and Slansky 1980] [Cross and Jain 1983] [Marroquin, Mitter, and Poggio 1985] [Geman and Geman 1984]

[Derin and Cole 1986] [Murray and Buxton 1987] [Cohen and Cooper 1987] [Amblard, Cooper, and Cernuschi-Frias 1986] [Szeliski 1987] [Derin et al. 1984] [Besag 1986] [Geman and Graffigne 1986] [Poggio 1985] [Drumheller and Poggio 1986]. In this section, we review the properties of MRF's and discuss how to encode prior knowledge in this formalism. We refer the reader to [Kindermann and Snell 1980] for an extensive treatment of MRF's.

3.2.1. Noncausal Markovian Dependency

Let E be a set of unordered pairs (s_i, s_j)'s representing the "connections" between the elements in S. The semantics of the connections will become clear shortly. E defines a symmetric and non-reflexive neighborhood system $\Gamma = \{N_s \mid s \in S\}$, where N_s is the neighborhood of s in the sense that

(1) $s \notin N_s$, and

(2) $r \in N_s$ if and only if $(s, r) \in E$.

X is a *Markov Random Field* with respect to Γ and P, where P is a probability function, if and only if

$$(\textit{positivity}) \quad P(X=\omega) > 0 \text{ for all } \omega \in \Omega \tag{3.8}$$

$$(\textit{Markovianity}) \quad P(X_s=\omega_s \mid X_r=\omega_r,\ r \in S,\ r \neq s) = P(X_s=\omega_s \mid X_r=\omega_r,\ r \in N_s) \tag{3.9}$$

The set of conditional probabilities on the left-hand side of (3.9) is called the *local characteristics* that characterizes the random field. It can be shown that the joint probability distribution $P(X=\omega)$ of any random field satisfying (3.8) is uniquely determined by these conditional probabilities [Besag 1974]. An intuitive interpretation of (3.9) is that the contextual information provided by $S-s$ to s is the same as the information provided by the neighbors of s. Thus the effects of members of the field upon each other is limited to local interaction as defined by the neighborhood. Notice that any random field satisfying (3.8) is an MRF if the neighborhoods are large enough to encompass all the dependencies.

3.2.2. Encoding Prior Knowledge and Gibbs Distributions

The utility of the MRF concept for image labeling problems is that the prior knowledge about spatial dependencies among the image entities can be adequately modeled with neighborhoods that are small enough for practical purposes. Very often, the image entities are regularly structured and prior distributions on the image are homogeneous and isotropic. In such cases, the number of parameters needed to specify the priors is just a fraction of Q^M, where M is the size of the neighborhoods. This is a significant saving over Q^N - the number of possible configurations, especially when M is small.

There are difficulties, as stated in [Geman and Geman 1984], associated with using the MRF formulation by itself:

(1) The joint distribution of the X_s is not apparent;

(2) It is extremely difficult to spot local characteristics, *i.e.*, to determine when a given set of functions are conditional probabilities for some distribution on Ω.

(1) is not a serious problem for some special classes of MRF models such as *Markov Mesh* (MM) processes [Abend, Harley, and Kanal 1965][Kanal 1980], since their joint distributions can be represented in a recursive formulation due to the casual dependency assumed. For (2), parametric probability distributions such as Gaussian and binomial, have been been used in the literature [Cross and Jain 1983] [Cohen and Cooper 1987]. Using such distributions further simplifies the encoding of the local characteristics and has shown some impressive results on modeling and generating texture patterns. However, whether these kinds of simplifications preserve the power of MRF's for modeling spatial knowledge remains questionable.

Fortunately, these difficulties vanished when the following property of MRF's was realized.

Hammersley-Clifford Theorem: A random field X is an MRF with respect to a neighborhood system Γ if and only if there exists a function V such that

$$P(\omega) = \frac{e^{-\frac{1}{T}U(\omega)}}{Z} \quad \forall\ \omega \in \Omega \tag{3.10}$$

where T and Z are constants and

$$U(\omega) = \sum_{c \in C} V_c(\omega). \tag{3.11}$$

C denotes the set of totally connected subgraphs (cliques) with respect to Γ. Z is a normalizing constant and is called the *partition function*.

The probability distribution defined by (3.10) and (3.11) is called a *Gibbs distribution* with respect to Γ. The class of Gibbs distributions has been extensively applied to model physical systems, such as ferromagnets, ideal gases, and binary alloys. When such systems are in a state of *thermal equilibrium* , the fluctuations of their configurations follow a Gibbs distribution. In statistical mechanics terminology, U is the *energy* function of a system. The V_c functions represent the *potentials* contributed to the total energy from the local interactions of the elements of clique c. T, the *temperature* of the system, controls the "flatness" of the distribution of the configurations.

Gibbs distributions, and therefore MRF's, possess a property that appears to be desirable for modeling - when constrained by a fixed expected value of some sufficient statistic of the random field, the *maximum entropy* distribution among the class of distributions compatible with the constraint is a Gibbs distribution.

The MRF-Gibbs equivalence not only relates the local conditional probabilities to the global joint probabilities, but also provides us a conceptually simpler way of specifying MRF's - specifying potentials. The importance of the joint probabilities will become evident in the next section. Based on (3.9) and the Hammersley-Clifford theorem, the local characteristics can be computed from the potential function through the following relation:

$$P(X_s=\omega_s \mid X_r=\omega_r, r\neq s) = \frac{e^{-\frac{1}{T}\sum_{c\in C_s} V_c(\omega)}}{\sum_{\omega'} e^{-\frac{1}{T}\sum_{c\in C_s} V_c(\omega')}} \tag{3.12}$$

where C_s is the set of cliques that contain s, and ω' is any configuration of the field that agrees with ω everywhere except possibly s.

There has been some work that applies statistical estimation methods to estimate parameters used for specifying MRF's. The standard approach is the maximum likelihood estimation: Choose the parameter values that maximize the likelihood of some realizations. Brute force search for the maximum is computationally intractable, and stochastic methods have been shown to work well for this problem [Geman and Graffigne 1986] [Hinton and Sejnowski 1983]. There are more efficient methods based on a "pseudolikelihood" measure proposed by Besag [Besag 1975]. The pseudolikelihood function is the product of the conditional likelihoods (local characteristics) over all sites. The computations involved in maximizing pseudolikelihoods are shown to be much simpler than in maximizing the real likelihood functions. Cross and Jain [1983] applies a coding scheme [Besag 1974] to estimate the parameters in their binomial distribution models. The idea is to partition the sites into "codes" so that no two in the same code are neighbors. An estimate of the paramters is chosen to maximize the psudolikelihood of a code, and the average of different coding estimates is taken to be the final estimate. Elliott and Derin [Elliott and Derin 1984] uses a least-square-fit method to estimate potential functions in the Gibbs distributions of their texture models. These methods are good when many uncorrupted realizations are available, such as in the case of natural texture modeling. When such data are difficult to acquire, choosing the clique potentials on an *ad hoc* basis has been reported to produce promising results [Geman and Geman 1984] [Marroquin, Mitter, and Poggio 1985]. Our experiments (Chapters 5 and 6) have also shown good results. These results are not surprising since the notion of clique potentials provides a simple mapping from "qualitative" spatial knowledge to numeric values of the parameters specifying the

MRF's.

3.3. A Posteriori Markov Random Fields

In this section, we move our attention to the relationships of segments of the image S. Recall that in Section 3.1, each site is associated with a set of α's, maintaining the opinions of the early visual modules. The updating of α's requires the knowledge of prior probabilities of individual labels conditioned on their parents' identities -- the initial values of w's. In the last section, we review Markov Random Fields and their properties. We mention that MRF's and thus the class of Gibbs distributions are suitable to represent our prior knowledge about local interactions between images sites. The problems remained are (1) computing the conditional probabilities required for the evidence aggregation procedure from the *a priori* Gibbs distribution, and (2) combining the α's with the *a priori* Gibbs distribution.

For (1), since the exact evaluation of MRF statistical moments is computationally intractable, we propose to use a stochastic Monte Carlo procedure to generate sample configurations of the prior MRF at equilibrium. Local statistics of label events can thus be collected to estimate the required prior conditional probabilities. More details of stochastic sampling and estimation are described in Chapter 4.

For (2), let β_s denote the set of α's associated with $s \in S$, and $\beta_s(l)$ be the α value for label l in β_s. Define a *global consistent state* to be a state of the β's at which each β_s is in a consistent state.

Assume that the prior knowledge about the image is represented as an MRF X over S, $X_s \in L$ - a mutually exclusive and exhaustive label set in H, with respect to a neighborhood system N. The Gibbs measure that characterizes the prior MRF is:

$$P_0(\omega) = \frac{e^{-\frac{1}{T}U_0(\omega)}}{Z_0} \tag{3.13}$$

where

$$U_0(\omega) = \sum_{c \in C} V_c(\omega),$$

and Z_0 is a normalizing constant. Given (3.1), Theorem 3.1, and Bayes' rule, the *a posteriori* Gibbs measure of a configuration ω at a global consistent state t can be computed as:

$$P_t(\omega) = \frac{e^{-\frac{1}{T}U_t(\omega)}}{Z_t} \tag{3.14}$$

where the *a posteriori* energy is

$$U_t(\omega) = \sum_{c \in C} V_c(\omega) - T \sum_{s \in S} \ln(\beta_s^t(\omega_s)) \tag{3.15}$$

It is easy to see that the *a posteriori* Gibbs measure characterizes a MRF over S with respect the neighborhood system N with the local characteristics:

$$P_t(X_s=\omega_s \mid X_r=\omega_r, r \neq s) = \frac{e^{-\frac{1}{T}\sum_{c \in C_s} V_c(\omega)+\ln(\beta_s^t(\omega_s))}}{\sum_{\omega'} e^{-\frac{1}{T}\sum_{c \in C_s} V_c(\omega')+\ln(\beta_s^t(\omega'_s))}} \tag{3.16}$$

Note that, based on (3.15) and (3.16), only simple local operations are involved in updating the energy measure and local characteristics as new opinions from the early visual modules become available. Therefore, inference methods depending only upon these measures, such as stochastic MAP and MPM estimations, can easily be implemented in the proposed framework. We investigate various decision criteria in the next chapter.

A. Appendix

In this appendix, we show that the evidence aggregation procedure described in Section 3.1 is commutative and associative (Lemma A.1). We then use it to prove Theorem 3.1. Corollary 3.2 immediately follows.

Lemma A.1: The updating and propagation rules described in Section 3.1.2 are commutative and associative.

Proof:

We first show the following: (1) The impact of two incoming messages (or likelihood ratios in the case of local updating) from the same source, say m_1^- and m_2^-, on the α and w's associated with a node is the same as the impact of a single message $m^- = m_1^- \, m_2^-$. (2) The two outgoing messages, m_1^+ and m_2^+ corresponding to m_1^- and m_2^- respectively, have the product of $m^+ = m_1^+ \, m_2^+$, same as the outgoing message corresponding to m^-. Therefore, the influence of m_1^- and m_2^- can be properly propagated by sending the two messages m_1^+ and m_2^+ in any order, or a single message m^+.

(1) and (2) are trivially true for the local updating and downward propagation rules, because only multiplications are involved in these rules. To see they are also true for the upward propagation rule, let us suppose node H_l receiving two messages m_1^- and m_2^- in this order, from a son H_i. For convenience purposes, we use superscripts 0, 1, and 2 to represent the values of α and w's prior to the arrival of m_1^-, after the arrival of m_1^- but before the arrival of m_2^-, and after the arrival of m_2^- respectively. Also, we drop the subscripts and superscripts l in α's and w's.

According to Equations (3.4a) - (3.4d), we have

$$m_1^+ = m_1^- \, w_i^0 + \sum_{j \neq i, j \in \gamma_l} w_j^0,$$

$$\alpha^1 = m_1^+ \, \alpha^0,$$

and

$$w_j^1 = \frac{m_1^- \, w_j^0}{m_1^+} \quad \text{if} \quad j = i,$$

$$= \frac{w_j^0}{m_1^+} \quad \text{otherwise.}$$

Upon receiving m_2^-,

$$m_2^+ = m_2^- \, w_i^1 + \sum_{j \neq i, j \in \gamma_l} w_j^1 = \frac{m_2^- \, m_1^- \, w_i^0 + \sum_{j \neq i, j \in \gamma_l} w_j^0}{m_1^+},$$

$$\alpha^2 = m_2^+ \, \alpha^1 = (m_2^- \, m_1^- \, w_i^0 + \sum_{j \neq i, j \in \gamma_l} w_j^0) \, \alpha^0$$

and

$$w_j^2 = \frac{m_2^- \, m_1^- \, w_j^0}{m_2^- \, m_1^- \, w_i^0 + \sum_{j \neq i, j \in \gamma_l} w_j^0} \quad \text{if} \quad j = i,$$

$$= \frac{w_j^0}{m_2^- \, m_1^- \, w_i^0 + \sum_{j \neq i, j \in \gamma_l} w_j^0} \quad \text{otherwise.}$$

It is easy to see that (1) is satisfied. Also we have

$$m_1^+ \, m_2^+ = (m_1^- \, m_2^-) \, w_i^0 + \sum_{j \neq i, j \in \gamma_l} w_j^0 = m^+.$$

Therefore, (2) is also true.

In case when m_1^- and m_2^- are sent by two different sons, say H_i and H_j respectively, it is easy to check, using a similar derivation as the one above, that the order of handling the two messages is irrelevant to the final values of α and w's. Also,

$$m_1^+ \, m_2^+ = m_1^- \, w_i^0 + m_2^- \, w_j^0 + \sum_{k \neq i, j; k \in \gamma_l} w_k^0$$

That is, m_1^- and m_2^- can be simultaneously processed by H_l, and the net effect can be summarized in one outgoing message. ■

This lemma suggests that the evidence propagation procedure can be implemented as an asynchronous tree-structured network. Early modules can simultaneously provide their opinions to individual nodes. Each node may process the incoming messages one at a time, or it can merge several messages together to minimize the number of messages to pass on. Now, we show that this procedure correctly combines the opinions based on the Bayesian formalism.

Theorem 3.1: Let α_l^t denote the α value associated with H_l at the consistent state t, and $P(O_t \mid l)$ be the probability of O_t given $X_s = l$, where O_t denotes the union of those observations that form the set of opinions that yields the state t. If $O_t \neq \varnothing$, then

$$\alpha_l^t = c_t P(O_t \mid l) \quad \forall \; H_l \in H \tag{A.1}$$

where c_t is a constant depending only on t, given the conditional assumptions (3.5) and (3.6) .

Proof:

Assume that at state t, there have been exactly m opinions processed. According to Lemma A.1, we can assume that they are processed one by one, generating consistent states $t_1, t_2, \ldots, t_m = t$. For the clearity of the proof, assume each opinion bears directly on a single node. It is easy to generalize the proof to the cases when the opinions consist of more than one likelihood ratios as described in 3.1.1.

Before we go on further, the semantics of w's need to be explained. Suppose H_i is a son of H_l, we shall show that at consistent state t_k, w_i^l is the probability of i conditioned on l after observing O_k. That is

$$_k w_i^l = P_l(i \mid O_k). \tag{A.2}$$

First let us prove that both (A.1) and (A.2) are correct at t_1. Assume the opinion bears directly on H_l. Let $\lambda_1 = \dfrac{P(O^1 \mid l)}{P(O^1 \mid \neg l)}$ represent this opinion. The nodes other than H_l and the root H can be classified into three groups: the descendants of H_l, the ancestors of H_l, and everyone else. Let us call the last group the

"cousins" of H_l. Recall that the α's are initialized to unity and the w's are initialized to the *a priori* conditional probabilities, e.g. $P_l(i)$.

Let $c_1 = \frac{1}{P(O^1 \mid \neg l)}$. According to the local update rule,

$$\alpha_l^1 = \lambda_1 = c_1 P(O^1 \mid l).$$

Let H_i be a descendant of H_l. According to (3.5) and the downward propagation rules, we have

$$\alpha_i^1 = \lambda_1 = c_1 P(O^1 \mid l) = c_1 P(O^1 \mid i).$$

The α^1's of the cousins remain 1. If H_i is a cousin of H_l, i has to be in $\neg l$. According to (3.5),

$$P(O^1 \mid i) = P(O^1 \mid \neg l) = \frac{1}{c_1}.$$

Therefore, $\alpha_i^1 = c_1 P(O^1 \mid i)$. Now let us look at the ancestors of H_l. Let H_j be the father of H_l. According to (3.4),

$$\alpha_j^1 = \lambda_1 \, {}_0w_l^j + \sum_{i \neq l; i \gamma_j} {}_0w_i^j = c_1 \left(P(O^1 \mid l) P_j(l) + P(O^1 \mid \neg l) P_j(\neg l)\right) = c_1 P(O^1 \mid j).$$

Using Bayes rule and (3.4d),

$${}_1w_l^j = \frac{c_1 P(O^1 \mid l) P_j(l)}{c_1 P(O^1 \mid j)} = P_j(l \mid O^1),$$

$${}_1w_i^j = \frac{P(O^1 \mid \neg l) P_j(i)}{P(O^1 \mid j)} = P_j(i \mid O^1) \quad \text{for} H_{\ i} \in \gamma_j, i \neq l.$$

Notice that $P(O^1 \mid \neg j) = P(O^1 \mid \neg i)$ according to (3.5) and the fact that $\neg j \subset \neg i$, also that the message to H_j's father has the value $c_1 P(O^1 \mid j)$. The above derivation recursively applys to all ancestors of H_l.

Now let us assume equations (A.1) and (A.2) are true at state t_k with respect to c_k, and $\lambda_{k+1} = \frac{P(O^{k+1} \mid l)}{P(O^{k+1} \mid \neg l)}$. Let O_k represent the union of the observations up to t_k. We would like to show they are also true at t_{k+1} with respect to

$$c_{k+1} = \frac{c_k}{P(O^{k+1} \mid \neg l)}.$$

Given (3.6),

$$\alpha_l^{k+1} = \lambda_{k+1}\, \alpha_l^k = \frac{c_k\, P(O^{k+1} \mid l)\, P(O_k \mid l)}{P(O^{k+1} \mid \neg l)} = c_{k+1}\, P(O_{k+1} \mid l)$$

Similar derivations apply to the descendants and the cousins of H_l. For the father,

$$\alpha_j^{k+1} = (\lambda_{k+1}\, {}_k w_l^j + \sum_{i \neq l; i \gamma_j} {}_k w_i^j)\, c_k\, P(O_k \mid j)$$

$$= \frac{c_k\, P(O_k \mid j)}{P(O^{k+1} \mid \neg l)} (P(O^{k+1} \mid l)\, P_j(l \mid O_k) + P(O^{k+1} \mid \neg l)\, P_j(\neg l \mid O_k)) = c_{k+1}\, P(O_{k+1} \mid j)$$

$${}_{k+1} w_l^j = \frac{P(O^{k+1} \mid l)\, P_j(l \mid O_k)}{P(O^{k+1} \mid j, O_k)} = P_j(l \mid O_{k+1})$$

For $H_i \in \gamma_j, i \neq l$,

$${}_{k+1} w_i^j = \frac{P(O^{k+1} \mid \neg l)\, P_j(i \mid O_k)}{P(O^{k+1} \mid j, O_k)} = P_j(i \mid O_{k+1})$$

Finally, the outgoing message:

$$m^+ = \frac{P(O^{k+1} \mid j, O_k)}{P(O^{k+1} \mid \neg l)} = \frac{P(O^{k+1} \mid j)}{P(O^{k+1} \mid \neg j)}$$

due to (3.5). ■

Chapter 4 Highest Confidence First Estimation

4. Highest Confidence First Estimation

At various levels of a visual hierarchy, estimations (inferences) must be made based on the available knowledge to extract more condensed, symbolic information and to reduce the amount of data passed between levels of visual tasks. In this chapter, we describe how to make label inferences, using Bayesian decision rationale, based on imperfect knowledge represented as *a posteriori* Gibbs distributions described in the previous chapter.

The goodness of a labeling $\hat{\omega}$, following the Bayesian formalism, is evaluated in terms of its expected loss,

$$Loss(\hat{\omega}) = \sum_{\omega \in \Omega} loss(\hat{\omega},\omega) P(\omega) \tag{4.1}$$

where $loss(\hat{\omega},\omega)$ is a penalty associated with the estimate $\hat{\omega}$ while the "truth" is ω, and $P(\omega)$ is the (*a posteriori*) probability of ω.

One question concerning the applicability of (4.1) is which loss function should be used for a given task. Except for few simple cases, the answer to this question usually relies on subjective judgements. One popular choice is assigning the same penalty to incorrect estimates: $loss(\hat{\omega},\omega)$ equals to a constant (positive) value whenever $\hat{\omega} \neq \omega$, and 0 otherwise. Using this loss function, the configuration minimizing (4.1) maximizes the *a posteriori* probability $P(\omega \mid O)$, and therefore minimizes the *a posteriori* energy (3.15) in the MRF formalism. This Maximum A Posteriori (MAP) criterion has been widely applied to the labeling problem [Feldman and Yakimovsky 1974] [Geman and Geman 1984] [Derin and Cole 1986] [Murray and Buxton 1987] [Cohen and Cooper 1987]. Marroquin et al [1985], also Besag [1986], suggest that the number of mislabeled image entities of an estimation is a better loss measure for the labeling problem. They derive the Maximizer of the a Posteriori Marginals (MPM) estimation - choosing the configuration $\hat{\omega} = (\hat{\omega}_{s_1}, \cdots, \hat{\omega}_{s_N})$ such that

$$\hat{\omega}_s = \max_{l \in L} P_s(l \mid O) \quad \forall s \in S,$$

where $P_s(l \mid O)$ denotes the *a posteriori* marginal probability of l on s. In their experiments the MPM estimator is shown to be superior to the MAP criterion when the signal to noise ratio is low. The idea of maximizing *a posteriori* marginal probabilities is not new, however. It has been extensively applied in probabilistic relaxation as described at the beginning of Chapter 3.

There are three problems with the MAP and MPM estimations for the labeling problem. The first problem is related to the cost of calculation of the estimates. Notice that the rationale of minimizing the loss function in (4.1) does not take the cost of computation into account, despite the fact that computational cost is usually a primary consideration in image understanding applications because of their immense configuration spaces. A sub-optimal estimator with an effective computation procedure would be much more useful than an optimal estimator that no one could ever compute. It is believed that the exact evaluation of MRF statistical moments, and therefore (4.1), is generally impossible since no analytic solutions exist [Hassner and Slansky 1980] [Geman and Geman 1984]. MAP and MPM can not be exactly determined for the same reason, except for some simple energy functions.

The second problem has to with the large-scale characteristics induced by using MRF's to model local dependencies among the neighboring sites. Besag [1986] pointed out that even relatively simple MRF's, such as binary Ising model, exhibit positive correlations over arbitrarily large distances when adjacent sites have high probability to be the same label. Thus there is a strong tendency to form infinitely large patches of a single label. This phenomenon has also been observed repeatedly in our experiments with Monte Carlo simulations of MRF's.

The third problem concerns the match between the image of interest and the prior model. It is possible for the true labeling of an instance of the problem to be very unlikely according to the *a priori* distribution. Therefore, the *a posteriori* probability of the true labeling and the corresponding marginal probabilities are relatively low in such cases, even though at some image sites there may be external evidence that strongly supports or refutes certain labels.

It seems desirable to devise an inference method that is computationally inexpensive, is unaffected by the large-scale characteristics of the MRF's, and incorporates spatial priors judiciously.

In this chapter we describe a new inference method -- the Highest Confidence First estimation, based on the *a posteriori* Gibbs distribution described in the previous chapter. This method aims directly at the above problems of MAP and MPM estimates. It is deterministic with local computations and global scheduling, and is shown to be efficient and robust by two sets of applications discussed in Chapter 5 and Chapter 6.

4.1. Background and Motivation

There exist several methods for the approximate evaluations of the MAP and MPM estimations in the MRF formalism. One class of methods, known as *stochastic relaxation* [Geman and Geman 1984], concern the stochastic behavior of the MRF's. They rely on stochastic sampling procedures to approximate MAP and MPM. Those methods asymptotically guarantee arbitrary accuracy of the results, but it is difficult, or impossible, to predict their behavior in a predetermined finite time interval. The second class of methods use strictly deterministic computations. Their concerns are mostly computational: How to (trivially) find a good starting point in the configuration space so that a simple energy descent procedure would lead to a reasonable approximation. We review these two classes of methods in the rest of the section.

4.1.1. Stochastic Relaxation Methods

One method that has been successfully used to analyze the behavior of complex systems is generating sample configurations of a given system through stochastic simulations. Briefly, the Monte Carlo method of estimating the ensemble average of a variable $Y(\omega)$,

$$\langle Y \rangle = \int_{\Omega} Y(\omega) dP(\omega),$$

is averaging its values over a set of samples $\{ \omega_1, \cdots, \omega_R \}$ drawn from Ω. If the sampling of ω's follows the distribution P, then $<Y>$ can be approximated by

$$<Y> \approx \frac{1}{R} \sum_{r=1}^{R} Y(\omega_r).$$

We are interested in sampling procedures that generate configurations according to Gibbs distributions in the form of (3.10). With such procedures, the sample frequencies of the realizations of X_s can be used as approximations for the marginal probabilities, *i.e.*, MPM can be estimated; the configurations with higher probabilities are more likely to be sampled, and therefore MAP estimation becomes possible. Several procedures exist for this purpose. The basic idea of these procedures is to construct a regular Markov chain whose states correspond to the configurations of the system with the limiting distribution being the desired Gibbs distribution. That is, construct P_C - the transition matrix of the chain - in such a way that the following condition holds.

$$\pi P_C = \pi, \tag{4.2}$$

where π is the desired Gibbs measure. At equilibrium, the system's configurations are distributed according to π since π is the unique invariant measure of the constructed Markov chain [Kemeny and Snell 1960] .

Consider each state transition of the Markov chain involving only the change of the state of a single site in the system. To fulfill the requirement of the chain being regular, the procedure must continue to "visit" every site. Let $s(t)$ be the site being visited at time t. The change of $X_{s(t)}$ would result in a change of the system energy by the amount specified by the configurations of those cliques that contain $s(t)$ according to (4.2). Stochastic sampling procedures reminiscent of "relaxation" can be designed in the sense that the state transition of the site being visited is stochastically decided by the states of the neighboring entities and itself. We will describe two of the stochastic relaxation procedures, namely the Metropolis algorithm [Metropolis et al. 1953] and the Gibbs sampler [Geman and Geman 1984], for their representativeness. Other variations basically follow the same principle and serve

special purposes [Hassner and Slansky 1980] [Cross and Jain 1983] [Hinton and Sejnowski 1983] .

4.1.1.1. The Metropolis Algorithm and the Gibbs Sampler

Let $X(t)$ denotes the state of the system at time step t. The state transition from step t to $t+1$ of the Markov chain generated by the Metropolis sampling algorithm consists of two basic steps:

(1) Randomly select a new configuration ω' (randomly visit a site s and choose a new state ω'_s), and compute the energy change $\Delta E = U(\omega') - U(X(t))$, where U is the energy function of the system as in (3.15).

(2) If $\Delta E<0$, set $X(t+1)=\omega'$. Otherwise, set $X(t+1)$ to ω' or $X(t)$ with probabilities $\frac{\pi(\omega')}{\pi(X(t))} = e^{-\Delta E/T}$ and $1 - e^{-\Delta E/T}$ respectively.

Allowing transitions with energy increases, a common characteristic of all stochastic relaxation procedures, prevents the sampling process from getting stuck at states of local energy minimum - an undesirable property of every deterministic hill-climbing procedure. In contrast to the explicit use of the energy difference in the Metropolis algorithm, the Gibbs sampler uses the local characteristics to construct a Markov chain. A state transition of the Gibbs sampler also consists two steps:

(1) Visit a site s.

(2) Randomly select the new state ω'_s for $X_s(t+1)$ following the distribution $\pi(X_s(t+1)=\omega_s \mid X_r(t), r\neq s)$. Having the form in (3.12), this distribution is generally easy to compute because only local operations are involved.

For binary systems, the Gibbs sampler is equivalent to the widely used "Heat Bath" algorithm - changing the state with probability $\frac{1}{1+e^{\Delta E/T}}$. Like other relaxation methods, the above procedures suggest the use of a parallel implementation since "updating" the X_s's requires propagating information only among neighboring computing units. Extra caution must be paid to the updating patterns of synchronous machines. For the Metropolis and Heat Bath algorithms, using any prescribed

updating order may result in the Markov chain not converging to the desired Gibbs distribution π [Marroquin 1985]. Our experiments (Chapter 5) use the Gibbs sampler exclusively because it guarantees the coincidence of π with the invariant measure of the chain as long as neighboring entities are not updated simultaneously.

4.1.1.2. The Monte Carlo and Simulated Annealing Methods

The stochastic relaxation scheme can be used to approximate the *a posteriori* marginal probabilities for the MPM estimation by simulating the equilibrium behavior of the *a posteriori* MRF. Since the Markov chain constructed by either the Metropolis algorithm or the Gibbs sampler leads to the desired limiting distribution regardless of its initial state, the law of large numbers suggests the marginal probability $P_s(l \mid O)$ be approximated by the sample frequency of $X_s = l$ at equilibrium, that is,

$$P_s(l \mid O) = \frac{1}{n-k} \sum_{t=k}^{n} \delta(X_s(t)-l) \tag{4.3}$$

where $\delta(0) = 1$, and 0 elsewhere. k is the number of steps for the chain to reach equilibrium, and n is the total number of steps of the simulation. Practically, experimentation is needed to determine how large n and k should be to achieve a desirable approximation accuracy given an arbitrary MRF. Cross and Jain [1983] have observed that in less than 10 iterations (full sweeps over the image entities), their texture modeling system becomes "stable" when sampled by a variation of the Metropolis algorithm. In general, in the order of hundreds of iterations are needed for the MPM estimation.

The system temperature - T in (3.10) - also plays an important role in MRF simulations. With low temperatures, the Gibbs distribution strongly favors the low energy configurations, but the time required for the system to reach equilibrium may be long. The system may reach equilibrium faster at higher temperatures, but the configurations are more evenly sampled; *i.e.*, it may require more samples to make accurate MPM estimations. The idea of *simulated annealing* [Kirkpatrick, Gelatt, and Vecchi 1983], obviously inspired by physical annealing, is to reach the

minimum energy states of a system by starting the system at a high temperature and gradually reducing it. In doing so the system tends to respond to large energy differences at the beginning, and is likely to find a good minimum energy state independent of its starting state. As the temperature decreases, the system tends to respond to small energy differences, and ideally settles at the lowest energy states ever encountered. The decreasing sequence of temperatures, called the annealing schedule, decides the effectiveness of this process. If the time spent at each temperature is not enough, the system may not converge to the global minimum states. On the other hand, it is often computationally prohibitive to use a slowly decreasing schedule. Geman and Geman [84] have derived an upper bound for the annealing schedules so that the schedules slower than this bound are guaranteed to converge to the global minimum energy states. However, this bound is very difficult to decide in practice since it relates to the range of energy values of the system.

Simulated annealing and Monte Carlo estimation have been applied in many computer vision tasks that involve optimization over exponential spaces, including image restoration [Geman and Geman 1984] [Besag 1986], stereo fusion [Barnard 1987], and depth and motion flow reconstruction [Hutchinson et al. 1988] [Gamble and Poggio 1987] [Koch, Marroquin, and Yuille 1986]. One major concern of using the stochastic relaxation scheme is its efficiency: at what cost can this scheme deliver satisfactory results? As mentioned previously, it is difficult to decide a tight bound on the computations required *a priori*. Usually one has to commit a computational cost that is intolerable for many applications to achieve reliable approximations. Together with the large-scale characteristics of MRF's and the possible modeling errors involved, it seems that the stochastic approach to approximate MAP or MPM is not suitable for vision problems with immense state spaces and imperfect knowledge sources.

4.1.2. Deterministic Relaxation Methods

For vision systems that require predictable results in reasonable time periods, using suboptimal estimation criteria (with respect to (4.1)) and/or heuristics in

searching for solutions seems to be a reasonable alternative to the stochastic relaxation scheme. In [Derin and Cole 1986], MAP estimations are performed on narrow strips of the image. The strips are limited to at most four rows wide so that MAP can be exact computed for each strip by a dynamic programming algorithm at feasible cost. For each estimation, only the estimate of the first row of a strip is kept. It serves as the boundary condition for the next strip consisting of the rest of the rows and a new one. Though limiting the extent of the (column-wise) interactions, the texture segmentation results appear to be impressive. Here we examine deterministic iterative relaxation methods for estimating the true labelings. The results of these methods are local minima of the *a posteriori* energy function (3.15), therefore they can be considered as approximation methods for MAP.

4.1.2.1. Iterative Energy Minimization

A simple version of deterministic iterative relaxation methods for energy minimization is the Metropolis algorithm without randomness: Start with any initial configuration. At each iteration through the sites, the state of each site is either changed to the state that yields maximal decrease of the energy, or is left unchanged if no energy reduction is possible. The process stops when no more changes can be made. This algorithm is guaranteed to find a local minimum of the energy function since each iteration strictly decreases the energy value and there are only a finite number of different values of the energy function. In terms of conditional probabilities, based on (3.15) and (3.16), the energy difference between two states of a site is proportional to the ratio of the conditional probabilities. That is

$$U(\omega) - U(\omega') \propto \frac{P(X_s=\omega_s \mid O,\ X_{S-s})}{P(X_s=\omega_s' \mid O,\ X_{S-s})}$$

where ω and ω' differ only at site s. Assume at time $t+1$, site s is visited. The new $X_s(t+1)$ maximizes the *a posteriori* local characteristic of s, that is

$$X_s(t+1) = \max_l P(X_s=l \mid O,\ X_{S-s}(t)).$$

Since every step of the energy reduction attempts to maximize a local conditional

probability, Besag [1986] named this method the Iterative Conditional Modes (ICM) estimation. Note that the right hand side has the form of Equation (3.16). The computations involved are local -- suitable to be executed simultaneously at every site, providing the neighboring sites are not updated at the same time.

Unavoidably, the local minimum obtained by the above algorithm may be far from optimal. Two enhancements are apparently helpful:

(1) Start with a better initialization of the MRF. One possibility is to use the *maximum likelihood estimates* (MLE) -- $X_s(0)=\omega_s$ if $\max_{l \in L} P(O_s \mid l)=P(O_s \mid \omega_s)$ [Besag 1986]. Such initializations do not rely on the spatial priors. The idea assumes that MLE's are mostly correct in large-scale characteristics, by require local adjustments to be consistent with the priors. Since the energy minimization procedure is monotonic, the undesirable large-scale characteristics of the *a posteriori* MRF are thus ignored. Hopefully the energy value of the true labeling is the local minimum of the valley where the value of the initial configuration lies.

(2) Escape from shallow valleys. By changing the states of more than one site at once, the new configuration may lead to a better local minimum. In a procedure described in [Cohen and Cooper 1987], the sites with small preferences of the current states over the others are assigned new states when a local minimum is reached. The relaxation restarts with the new configuration as the initialization. At each convergence, the magnitude of the local minimum is estimated, The procedure halting when no significant change of the magnitudes is observed. The hope is that the deepest valley will be found in this process.

Using (1) is not adequate to achieve robust estimations. The error rate of local MLE's can be low only when the likelihood function correctly models the relation between the label hypotheses and the observations, and, more importantly, there must be significant differences among the likelihoods of the hypotheses. Frequently these conditions cannot be met, resulting initial configurations far away from the true

labeling in the energy space. Moreover, since the energy space is usually very rough -- full of local minima, significantly different estimates may result under different visiting orders to the sites, even with the same initial estimate. It is impossible to decide the best ordering *a priori*, and thus a predetermined (or random) order must be programmed, resulting unpredictable results. Cohen and Cooper's procedure with reasonable initializations has shown good results. It uses extra computations in exchange for better performance, and is compatible with the proposed method as a postprocessing step.

4.2. Estimation with Highest Confidence First

It has become apparent that an estimation method should possess the following properties:

Efficiency: The cost of computation meets the demands of the visual tasks. It is desirable for a procedure to maximally improve the estimates progressively so that it can provide the "best" current estimate upon requests.

Predictability: The final estimate depends only on the inputs and the chosen *a priori* distribution. Thus the user is free from the considerations concerned by the existing methods such as the choices of initial estimate, annealing schedule, and visiting order.

Robustness: The estimates degrade gracefully with the increase of noise and the modeling error. Also, they are unaffected by the large-scale characteristics of the chosen MRF's.

The Highest Confidence First (HCF) method introduced in this section satisfies the above requirements. It blends the initialization into the estimation process. Instead of stepping through the configuration space, this method constructs a configuration with a local minimal energy measure by following a path, suggested by the observations, in a augmented space. Observable evidence and spatial prior knowledge are combined in the process of the construction, resulting in robust estimates and better efficiency. The details of this method are described next.

4.2.1. Augmented Search Space

To see how this algorithm works, some terminology needs to be introduced. Recall that $L = \{l_1, \cdots, l_Q\}$ is a set of mutually exclusive and exhaustive labels with respect to which the labeling problem is defined, and Ω denotes the corresponding configuration space. The *a posteriori* knowledge about the labelings is represented by a Gibbs distribution; the *a posteriori* probability of a labeling ω, according to (3.14), is

$$P(\omega|O) = \frac{e^{-\frac{1}{T}\left(\sum_{c\in C} V_c(\omega) - T\sum_{s\in S}\ln(P_s(\omega_s|O_s))\right)}}{Z} \tag{4.4}$$

Let $\bar{L} = L \cup \{l_0\}$ denote the *augmented label set*, where l_0 is the null label corresponding to the "uncommitted" state in the construction. Let $\bar{\Omega} = \{\omega = (\omega_1, \cdots, \omega_N) \mid \omega_s \in \bar{L}, \forall s \in S\}$ denote the *augmented configuration space*. The basic idea of this algorithm is to construct a sequence of configurations $\omega^0, \omega^1, \cdots$ of $\bar{\Omega}$ with the starting configuration $\omega^0 = (l_0, \cdots, l_0)$, and a terminal configuration -- the final estimate $\omega^f \in \Omega$, where the energy measure $U_O(\omega^f)$ is a local minimum with respect to Ω.

We say a site s has *committed* to a label $l \in L$ at step t of the construction if $\omega_s^t = l$, and it is uncommitted if $\omega_s^t = l_0$. We impose a rule that states once a site has committed to a label, it can not nullify its commitment, but it is allowed to *change* its commitment to other labels of L. The rationale behind this rule will soon become clear.

Define the *augmented a posteriori local energy* of $l \in L$ with respect to $s \in S$ and a configuration $\omega \in \bar{\Omega}$ as

$$E_s(l) = \sum_{c:\, s\in c} V_c'(\omega') - T \ln P(O_s|l), \tag{4.5}$$

where $\omega' \in \bar{\Omega}$ is the configuration that agrees with ω everywhere except $\omega_s' = l$, and V_c' is 0 if $\omega_r = l_0$ for any r in c, otherwise it is equal to V_c - the potential function. This measure quantifies the goodness of a label with respect to the current configuration of

the neighbors. It is related to the conditional probabilities (local characteristic) with respect to an MRF that has the same potential functions as the prior MRF, but consists of only the committed sites. Notice that only committed neighbors contribute to this measure, thus uncommitted sites do not actively influence others' commitments based on this measure. However, an uncommitted site always takes into account of the states of the active neighbors when making a commitment.

4.2.2. Local Stability Measures

To ensure the quality of the resulting estimate - ω^f, we impose the following rule that decides the "updating" order: At each step of the construction, only the least "stable" site is allowed to change/make its commitment. We define the *stability* of s with respect to the current configuration ω as follows.

$$G_s(\omega) = \min_{k \in L, k \neq \omega_s} \Delta E_s(k, \omega_s) \quad \text{if} \quad \omega_s \in L \tag{4.6a}$$

$$G_s(\omega) = -\min_{k \in L, k \neq j} \Delta E_s(k, j) \quad \text{if} \quad \omega_s = l_0, \tag{4.6b}$$

where j in (4.6b) satisfies: $j \in L$ s.t. $E_s(j) = \min_{k \in L} E_s(k)$, and $\Delta E_s(j,k) = E_s(j) - E_s(k)$ with respect to ω.

The stability defined above is a combined measure of the observable evidence and the *a priori* knowledge about the preferences of the current state over the other alternatives. A negative value of G, that is always true for uncommitted sites (4.6b), indicates a more stable - less energy - configuration will result from an alternative commitment. The magnitude of the measure corresponds to how much energy would be lost/gained by changing the current state: the larger the negative value of G, the more confidence we have in changing its state. Since a site has no effect on its neighbors unless it has committed, the sites with large likelihood ratios of one label over the others - strong external evidence in favor of a label - will be visited early in the construction sequence. Observe that when no neighbor is active, the augmented local energy measure reduces to the local likelihood of the label. Therefore, commitments under such circumstances are equivalent to the local maximum

likelihood estimates. The sites with little idea from the observations will take the neighbors' configuration into account when making their commitments. As mentioned previously, such commitments are equivalent to the conditional modes. An early commitment will be altered if the neighbors' later commitments are strongly against it, thus an error estimate based on local evidence can be corrected when more contextual information becomes available.

A short summary of HCF: Every step of this construction makes a maximal progress in reducing the energy measure (4.4) based on the current knowledge about the field - the G's. The initial configuration is a constant, and the visiting order is implicitly decided by the external observations and the priors. This method is unaffected by the large-scale characteristics of the chosen MRF's, and the results degrades gracefully in accordance with the noise and modeling error, due to the highest confidence first construction of the estimates (Section 4.3). In Chapter 5, several estimation methods based on the MRF modeling are applied in the domain of boundary detection. Experimental results are strongly in favor of HCF for both efficiency and correctness.

4.2.3. Serial Implementation

The Highest Confidence First estimation method can be implemented serially with a heap (priority queue) maintaining the visiting order of the construction according to the values of G's in such a way that the *top* of the heap is the site with the smallest G value. Updating the *top*'s decision will cause the changes of its neighbors' G-values, and therefore the structure of the heap. The following is the pseudo code for the Highest Confidence First algorithm:

```
ω = (l_0, ···, l_0);
top = Create_Heap(ω);
while (G_top < 0) {
   s = top;
   Change_State(ω_s);
   Update_G(G_s );
   Adjust_Heap(s);
   foreach (r ∈ N_s) {
      Update_G(G_r);
      Adjust_Heap(r);
   }
}
return(ω);
```

Change_State(ω_s) changes the current state ω_s of s to the state l such that $\Delta E_s(l, \omega_s) = \min_{k \in L, k \neq \omega_s} \Delta E_s(k, \omega_s)$ if $\omega_s \in L$, or $E_s(l) = \min_{k \in L} E_s(k)$ if $\omega_s = l_0$. Upon this change taking place, the stability of s changes to positive. Update_G is called for every site that is affected by this change, namely the neighbors of s according to (5.1), to update their stability measures with respect to the new configuration. Adjust_Heap(r) maintains the heap property by moving r up or down according to its updated G-value.

4.2.4. Convergence Properties

Several desirable properties of this procedure can easily be verified:

(1) Termination: This procedure always returns in finite time. To see this property, let us consider the two types of Change_State - making and changing a commitment - separately. The procedure can make at most N commitments, one for each site, since nullifying commitments is impossible. Let $D = (S_D, S - S_D)$ be a *partition* of S such that S_D is the set of sites that have made commitments. Let $\overline{\Omega_D} = \{\omega \in \overline{\Omega} \mid \omega_s \in L \ \forall s \in S_D, \ \wedge \ \omega_s = l_0 \ \forall s \in S - S_D\}$. Since, by (4.5) and (4.6a), changing the commitment of $s \in S_D$ strictly decreases the function $U_D : \overline{\Omega_D} \rightarrow R$,

$$U_D(\omega) = \sum_c V'_c(\omega) - T \sum_{s \in S_D} \ln P(O_s \mid \omega_s),$$

the procedure can make only a finite number of changes with respect to a fixed partition D. There are only a finite number of partitions, therefore the total number of commitment changes is finite.

(2) Feasibility: The returned configuration is in Ω - the space of feasible solutions. For if otherwise, there exists an s such that $\omega_s = l_0$. From (4.6b), $G_s < 0$. This violates the heap invariant property since it requires $G_{top} \geq 0$ to exit the while loop.

(3) Optimality: The returned configuration has the locally minimal energy measure with respect to Ω. That is, changing the commitment of any single site can not decrease the *a posteriori* energy measure U_O. As above, this property can easily be derived from (4.6a) and the heap properties.

This implementation takes O(N) comparisons to create the heap and O(log(N)) to maintain the heap invariance for every visit to a site, provided the neighborhood size is small relative to N -- the number of sites. The overheads of heap maintenance are well repaid since the procedure makes progress for every visit, in contrast to the iterative relaxation procedure that may make only few changes per iteration (N visits). Our edge detection experiments (Chapter 5) show that on the average, less than one percent of the sites are visited more than once using the proposed algorithm while the deterministic relaxation procedure takes around 10 iterations to reach a local minimum. This advantage becomes more evident as the number of sites gets larger.

4.3. Discussion and Possible Extensions

The HCF estimation method meets two important design principles for visual procedures suggested by Marr [Marr 1982], namely, the principle of graceful degradation and the principle of least commitment. A common effect of lowering signal to noise ratio is the decrease of the feature saliency. For instance, a sharp edge in a badly degraded image may appear rather weak. As suggested previously, the

search is guided by salient features; the utilization of contextual information increases as the degree of saliency decreases. Therefore, the results degrade gracefully with the increase of the noise level. Also, spatial priors are used when external evidence is weak, thus model noise and large-scale characteristics are less likely to affect the estimates at the sites with apparent answers. There are two more advantages of delaying the commitments of the sites with weak observations: (1) To minimize the possibility of *undoing* previous commitments: it is likely for those sites to make incorrect commitments without enough contextual information, therefore they should commit late. Principle of least commitment thus follows. (2) To reduce the chance of misinforming other sites. A site can do better without the incorrect information of a neighbor.

The concept of highest confidence first can be used as a heuristic search strategy for large state-space optimization or a rule for the nodes of a cooperative network to reach mutual agreements. It can be extended in many directions to achieve, perhaps, better results. Let us look more closely at the construction process of the HCF estimate. At each stage, S_D consists of a set of isolated clusters. A cluster is a set of spatially connected (with respect to Γ) sites. We say two clusters are isolated from each other if none of the sites of a cluster is a neighbor of any site of the other cluster. Each cluster corresponds to an MRF with free boundaries in our formalism. When a site makes a commitment, a cluster is created or expanded, or clusters are merged. When a site changes a commitment, the energy of the corresponding MRF is reduced. Eventually, all the clusters are merged and the final estimate corresponds to a local minimum configuration of the corresponding MRF.

The notion of growing clusters suggests a natural partition of the image. At any instant, the sites belonging to the same cluster are tightly related, but they are independent of the members of other clusters. The addition of a new member to a cluster may change the commitments of the old members, but the changes are expected to be small due to the way the clusters are constructed. Therefore, it makes sense to compute the MAP estimates exactly for small clusters early in the construction to reduce the possibility of early mistakes without been affected by the

large-scale characteristics of the fields. We believe that by doing so the results would be better than the results using the horizontal strip partition as in [Derin and Cole 1986].

The process of growing clusters is similar to annealing in the sense that it responds to large energy differences earlier than small ones. Nondeterminism can be introduced to those sites that stay "unstable" - the sites on or exterior to the border of the clusters - late in the process, since more contextual information is required for them to reach a globally satisfactory agreement. Cohen and Cooper's postprocessing procedure [1987] described in Section 4.1.2 can similarly be incorporated.

The Highest Confidence First estimation can be implemented with a set of cooperative computing units. Consider a *winner-take-all* network where each unit corresponds to a site of the image [Feldman and Ballard 1981]. Only the units with the smallest stability measures can "fire" at one instant; each unit maintains the knowledge about the neighboring units so that its stability measure can be updated immediately should any neighbor change its state. The parallelism gained, however, is limited due to the sequential firing order.

The strict sequentiality of HCF can be relaxed. One possibility is the use of a global stability threshold. The (negative) threshold values increases in time and is broadcasted to all sites. Any site with a stability less than the threshold is allowed to change its state. The computation stops when the threshold reaches zero. It is not yet clear how this modification would affect the resulting estimates, and how to choose an appropriate schedule to increase the threshold.

It is interesting to note that Koch et al [Koch, Marroquin, and Yuille 1986] and Besag [Besag 1986] have independently observed that better estimates can be resulted by using a sequence of weaker fields on previous cycles. In the case of surface reconstruction, Koch et al strongly penalize the formation of lines at the beginning, and slowly decrease the penalty as the computation proceeds. Thus lines are formed only at very steep disparity gradients early in the process, and surface can break at smaller depth disparity gradients by paying a smaller price later. Similarly,

Besag decreases the contribution from the prior field for the task of image restoration after each iteration of ICM. Their ideas are related to HCF in the sense they all try to avoid committing too early based on the unreliable initial estimate. However, the explicit uses of the uncommitted state and the stability measures by HCF have the advantages of efficient computation (least commitment), robust results (least commitment and graceful degradation), and easy implementation (no need to choose a proper schedule).

Chapter 5 Probabilistic Boundary Detection

5. Probabilistic Boundary Detection

We have chosen to tackle the well studied problem of intensity edge detection using MRF's as the underlying formalism. The labeling problem in this context is to assign to each site a label from the set $\{EDGE, NON\text{-}EDGE\}$, based on discrete intensity measures on the pixels of a square lattice-structured image.

Since the primary purpose of edge detection is to locate sharp changes (discontinuities) of certain parameters such as surface depth and orientation in the scene, the semantics of the labels are defined in terms of the events of the scene rather than the events of the image (Chapter 3). The intensity at a single pixel says nothing about how an edge site should be labeled, but the variations among pixel intensities provide a strong clue to the true answer. To account for corruptions due to random noise, sampling, and quantization error in the imaging process, usually the intensity measures of many pixels are required to extract the information about the label of a site. Such information, although important, by no means captures completely all we know about the scene discontinuities.

Most work in the past on edge detection has concentrated on locating large intensity gradients. The usual approach is to analyze the intensity function defined by a predetermined window (set of spatially adjacent pixels) surrounding the site of interest. Then, probably incorporating with knowledge such as smoothness of the window intensity function, the noise process, and the point-spread function of the imaging device, a number is calculated reflecting the edgeness of a possible edge location [Roberts 1965] [Hueckel 1971] [Prewitt 1970] [Kirsch 1971] [Nevatia and Babu 1980] [Haralick 1984] [Canny 1983] [Nalwa 1984] [Sher 1987b]. This number is then used to decide how the site should be labeled based on local thresholding, non-maximum suppression, and linking processes.

While the class of local edge detection techniques has produced reasonable results in many applications, it falls short of the goal of edge detection for the following reasons. First of all, it is difficult to choose a proper size for windows *a priori*. Using windows too small may lead to poor detection due to noise in the

image, and using windows too large may fail to localize the edges accurately due too the possible interactions from nearby edges. In practice, it usually requires using windows of several sizes to overcome such problems. Secondly, it is difficult to incorporate knowledge about spatial interactions among the edges. In most cases, the scene consists of well-defined objects. The edge contours corresponding to the boundaries of the objects tend to be smooth and connected. Most local edge detecting schemes ignore such knowledge. Sophisticated thresholding and linking techniques exist, but they are rather ad hoc and not adequate [Nevatia and Babu 1980] [Canny 1983]. A classic example is the detection of the boundaries of a block standing in front of a slow varying background. Any local scheme will fail to detect part of segments due to lack of contrast, however such boundaries are obvious to a human observer. Lastly, the intensity measures only provide a piece of partial evidence of edges. Oftenly, other image cues such as depth, orientation, and texture are available. It is important that such cues can coherently be integrated with the intensity data and the *a priori* spatial knowledge.

An alternative to local edge detection is to detect discontinuities through the process of reconstructing global intensity functions. *A priori* knowledge is encoded via global energy functions (see Chapter 2 and 3). Edges are identified at the locations where the connections between pixels should be broken in order to reduce the energy measure related to the intensity configuration. Such schemes have demonstrated superior results in both robustness and localization [Blake and Zisserman 1987]. However, the computation usually involves optimization of non-convex functions with large state spaces. Since the reconstructed intensity data are, at least for now, of little use for higher-level tasks, the cost/benefit ratio of the method seems too high for our needs.

Our approach to edge detection, based on the probabilistic framework proposed in Chapter 3 and 4, has the advantages of the above local and global schemes but none of the associated disadvantages. It uses the outputs of a set of local operators that relate the intensity observations to edge labels, and ignores the intensity values afterwards. Thus the global optimization is performed over a space much smaller

than the ones for full reconstruction. On the other hand, the results are robust against local noise and consistent with spatial priors due to the use of a global energy measure. In addition, our approach is probabilistic; it is easy to incorporate other image clues. Finally, our approach seems consistent with the human visual system: They both use a set local operators in the early stage, and are able to detect weak but connected contours..

5.1. Local Edge Models

The edge sites are considered to be situated on the boundary between two pixels (see Fig. 5.2). We adopt a step-edge with white Gaussian noise model to compute the local likelihoods of a site s being $EDGE$ or $NON-EDGE$ -- $P(O_s \mid \omega_s=EDGE)$ and $P(O_s \mid \omega_s=NON-EDGE)$. As mentioned previously, local edge detection must use windows that are large enough to tolerate local noise yet small enough not to involve multiple edges. We choose to use a 1×4 cr 4×1 window of brightness observations surrounding s to be the observation of the site s - O_s. This window of intensity values is assumed to be a realization of one of the possible events depicted in Figure 5.1, corrupted by independent Gaussian noise. Figure 5.1.a shows the event E_1 of an edge occurring at the center of the window. 5.1.b indicates that the window corresponds to a uniform region (E_2), and 5.1.c depicts the events (E_3 and E_4) that the window consists of two regions, however the boundary between the two regions is one pixel off (right or left) the center of the window. Events of 5.1.b and 5.1.c constitutes the $NON-EDGE$ event of interest.

There is no doubt that the above edge model can greatly be improved in many directions. The obvious ones include the modeling of different types of edges such as roof, line, and peaks, the use of more information by employing circular windows, and the modeling of image blurrs. Since one primary goal of this work is to study the robustness of MRF modeling and HCF estimation in the presence of errors in sensory modeling, we feel that the above edge model serves our intention well, and further improvements can only result better estimates. The computation of the likelihoods given a window of intensity observations and the above edge model is described next.

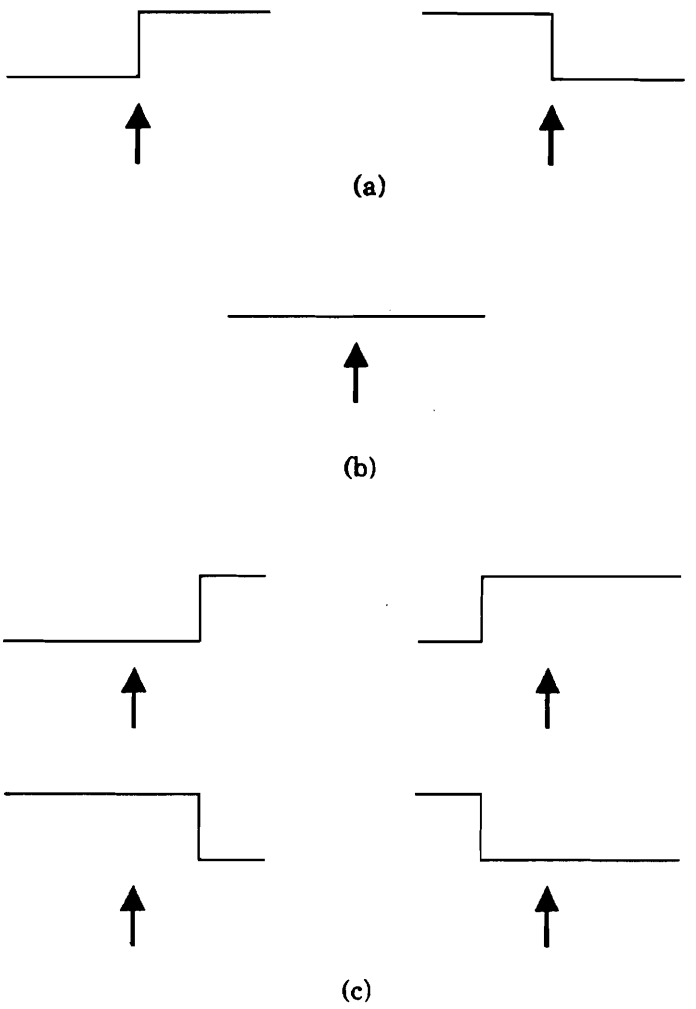

Figure 5.1. Step Edge Model
Image events in a 4×1 window.
(a) Edge occuring at center of window,
(b) Homogenous region: no edge occurs,
(c) No edge at center, offset edge occurs.
(Arrow indicates center of window)

It is based on the work of Sher [Sher 1987b]. The reader is referred to [Sher 1987a] for a complete treatment of probabilistic local edge detection.

5.1.1. Computing Edge Likelihoods

The computation of the likelihood of E_2 is relatively straight forward. Let W denote the vector of window observations (w_1, w_2, w_3, w_4), where the index indicates the spatial order of the pixels in the window. Since the noise is considered

to be independent Gaussian additive, with zero mean and variance σ^2,

$$P(W \mid E_2) = \sum_{r=0}^{255} P(r)\, P(W \mid r\, T_0) = \frac{1}{(2\pi\sigma^2)^2} \sum_{r=0}^{255} P(r)\, e^{-\frac{1}{2\sigma^2}(W-rT_0)(W-rT_0)^t},$$

where T_0 is vector consists of all 1's, and $P(r)$ is the prior probability of a pixel having intensity r. In our experiments, we assume that all intensity levels are equally likely, therefore $P(r) = 1/256\ \forall r$. More complex distributions may easily be incorporated. One possibility is to use a normalized intensity histogram of the input image as the prior distribution.

For the cases of edge occurrence, let r_1 and r_2 be the assumed intensities of the two regions on the left and right of the edge respectively. The window of assumed intensities can be represented as the vector $T = r_1 T_1 + r_2 T_2$, where T_1 and T_2 are the template vectors for the left and right regions. For example, for the region on the left of three-pixel wide (E_3), $T_1 = (1,1,1,0)$, and the corresponding $T_2 = (0,0,0,1)$. The likelihood of such cases (an edge present at a particular location in the window) can be computed by

$$\sum_{r_1,r_2} P(r_1,r_2)\, P(W \mid T) = \frac{1}{(2\pi\sigma^2)^2} \sum_{r_1,r_2} P(r_1,r_2)\, e^{-\frac{1}{2\sigma^2}(W-r_1T_1-r_2T_2)(W-r_1T_1-r_2T_2)^t}.$$

Again we assume r_1 and r_2 are independently and uniformly distributed in the experiments. The *EDGE* event E_1 corresponds to the case with the pair of templates (1,1,0,0) and (0,0,1,1). The likelihood of *NON–EDGE* is

$$\begin{aligned} P(O \mid NON\text{–}EDGE) &= P(O \mid E_2 \vee E_3 \vee E_4) \\ &= P(O \mid E_2)\, P(E_2 \mid NON\text{–}EDGE) + P(O \mid E_3)\, P(E_3 \mid NON\text{–}EDGE) \\ &\quad + P(O \mid E_4)\, P(E_4 \mid NON\text{–}EDGE). \end{aligned}$$

We set $P(E_2 \mid NON\text{–}EDGE) = 0.8$ and $P(E_3 \mid NON\text{–}EDGE) = P(E_4 \mid NON\text{–}EDGE) = 0.1$ in our experiments.

The calculation of edge likelihoods is efficiently carried out with a set of local convolutions followed by a table-lookup operation [Sher 1987b].

Based on the fact that scaling $P(O_s \mid l)$ for every $l \in L$ by a constant factor for fixed s in Equation (4.4) does not change the *a posteriori* distribution, we can use the log likelihood ratios -- $\log P(O_s \mid \omega_s = EDGE) - \log P(O_s \mid \omega_s = NON-EDGE)$ -- as the only input data, thus simplifying the computation of the stability measures (5.2). Thresholding the log likelihood ratios by the logarithm of prior (local) odds -- $\frac{P(NON-EDGE)}{P(EDGE)}$ -- of a site results in the *thresholded log likelihood ratio* (TLR) configuration. This configuration can be considered as an MAP estimate obtained without using contextual information, because

$$\frac{P(EDGE \mid O)}{P(NON-EDGE \mid O)} = \frac{P(EDGE)}{P(NON-EDGE)} \frac{P(O \mid EDGE)}{P(O \mid NON-EDGE)},$$

therefore,

$$P(EDGE \mid O) > P(NON-EDGE \mid O) \Leftrightarrow$$

$$\log \frac{P(O \mid EDGE)}{P(O \mid NON-EDGE)} - \log \frac{P(NON-EDGE)}{P(EDGE)} > 0.$$

In our experiments, we use TLR's as the initial estimates whenever possible.

5.2. Markov Random Field of Line Process

The MRF model used is similar to the "Line Process" MRF used both by Geman *et al* [1984] and Marroquin *et al* [1985]. Each edge site is modeled as a random variable of the field. The field is binary, with $2(N^2-N)$ entities where the image is a $N \times N$ rectangular pixel array. Notice one major difference between our setup and the existing MRF segmentation work: the line process of the latter is implicit. There are no external observations directly associated with it, and the formation of the lines depends on the configurations of a coupled MRF of the intensity process. In our setup, the intensity values are used only to calculate the local likelihoods for the edge sites, and the likelihoods constitute the input of the stand-alone line process.

5.3. Construction of Potential Functions to Encode Prior Knowledge

The spatial relationships between edge sites we wish to encourage have the following effects:

(1) To encourage the growth of continuous line segments,

(2) To discourage abrupt breaks in line segments,

(3) To discourage close parallel lines (competitions) and

(4) To discourage sharp turns in line segments.

A second order neighborhood turns out to be sufficient to encode all the relationships we want. In this neighborhood system, each MRF element is adjacent to eight others (see Fig.s 5.2 and 5.3).

The second order neighborhood has cliques of sizes 1 through 4 (see Fig. 5.4). The potential values we assign to various configurations of these cliques are shown in Fig. 5.5. These values form the specification of the potential functions. Therefore

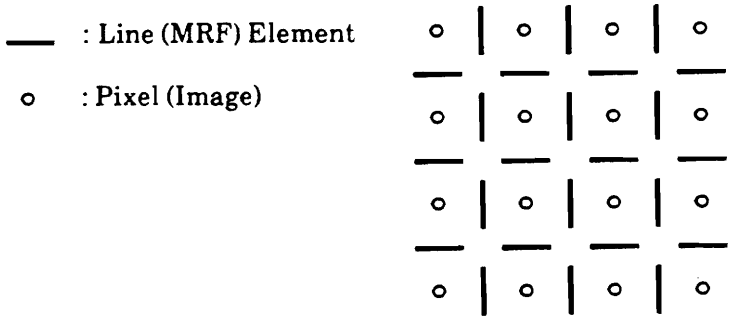

Figure 5.2. Pixels and Edges
Relationship between MRF (edge) sites and pixels.

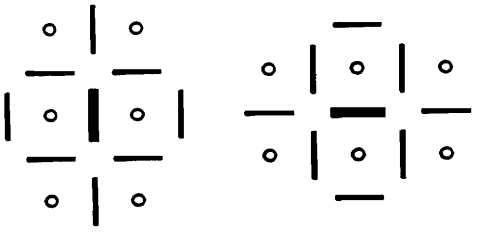

Figure 5.3. Edge Neighoborhood
The second order neighborhood system.

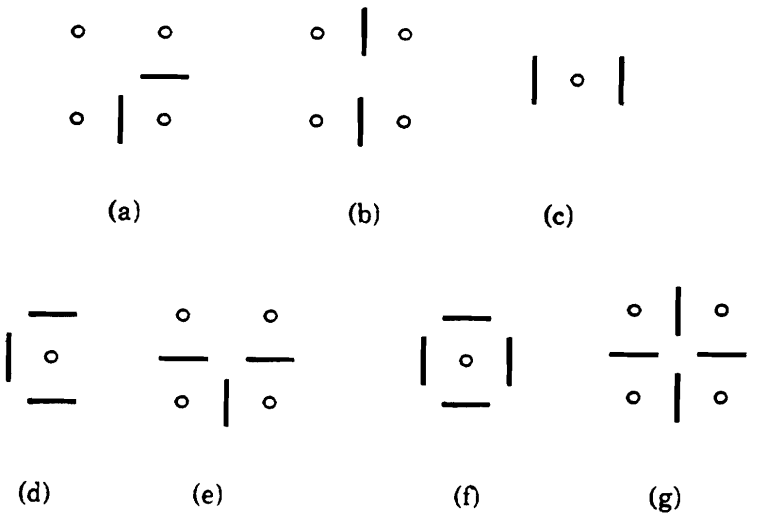

Figure 5.4. Edge Cliques
Cliques in neighborhood system, of size greater than 1.

potential functions can be seen to be specified by about 10 parameters, which are currently assigned in an *ad hoc* manner. The rules of thumb that are used to assign values to these parameters are:

Determine Structure Enforcers For each clique, attempt to determine what kind of structural relation it is uniquely capable of enforcing.

Encode Prior Structural Knowledge By assigning "high" potential values to undesirable configurations of the cliques and "low" values to desirable ones, we attempt to ensure that the final estimate will contain as few of the undesirable ones as possible.

Encode Statistical Prior Knowledge We use the clique consisting of the singleton node to bring the first order statistics (*e.g.* the density of *EDGE*s) of the MRF into line with what we already know. The potential of the clique when the MRF entity is an *EDGE* is set to our estimate of the log of the (local) odds of an entity being an *EDGE* over a *NON–EDGE*, and is set to 0 when it is a *NON–EDGE*.

A point to be noted is that some of these parameter values are interdependent. For example, increasing the energy for "break" (Fig. 5.5b) and "continuation" (Fig. 5.5c) configurations simultaneously would be of little use, as the increases would

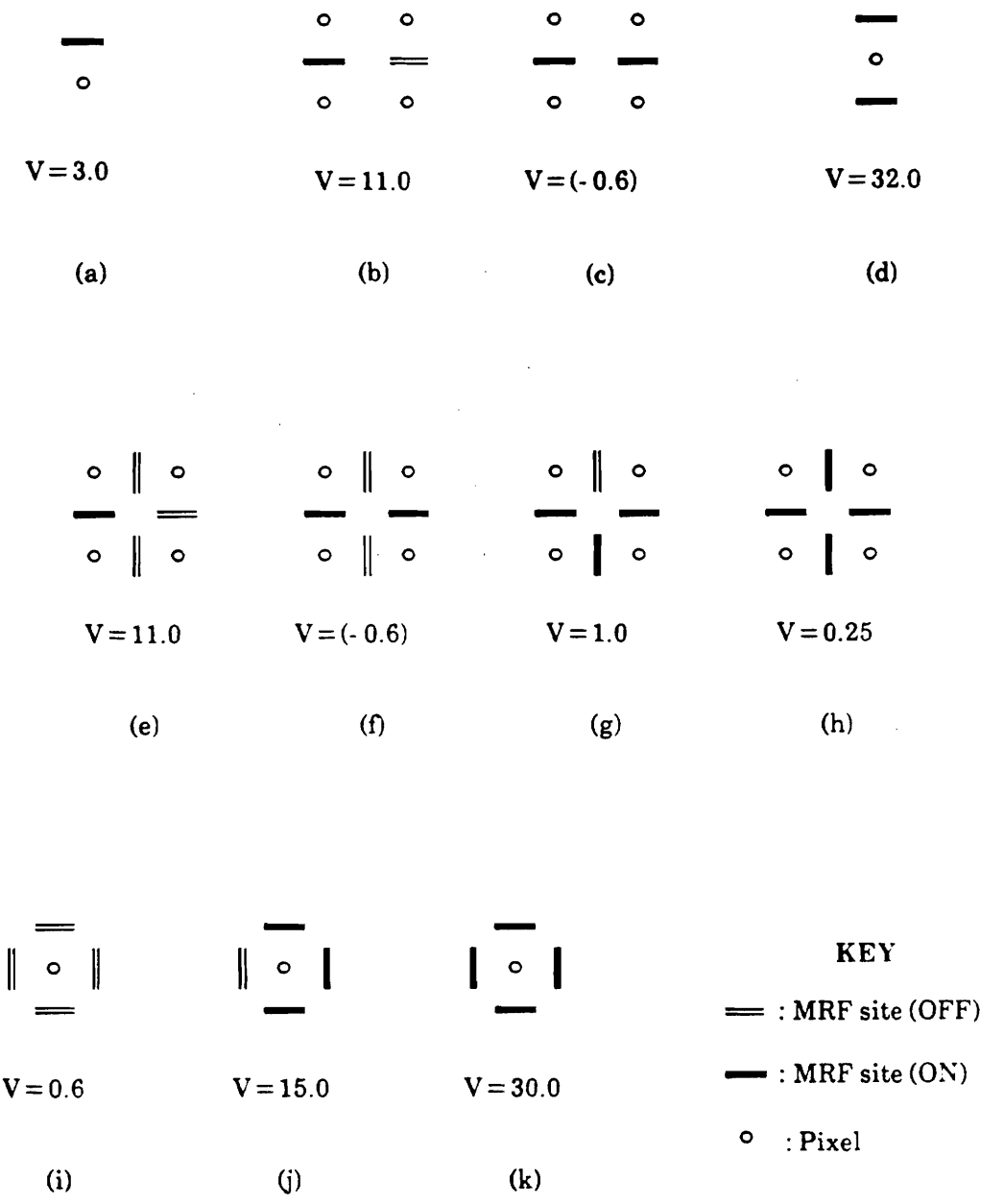

Figure 5.5. Potential Assignments
(Configurations not shown have 0 potential values)

tend to cancel each other out.

The sensitivity of the results obtained to changes in the parameters specifying the potential functions depends upon the parameter in question. Our experience is that changing the potential function associated with the 1-clique had the greatest effect on the final result, followed by the 2-clique and 4-clique potential functions, in

that order. This could be because the singleton clique controls first order statistics and the larger cliques higher order statistics, which are known to be less important in distinguishing images [Julesz 1981].

5.4. A General Purpose MRF Simulator

Our experiments use an interactive general-purpose MRF simulator package with extensive graphics and menu-driven control (Fig. 5.6). This package takes the description of the MRF and the likelihood ratios as input and simulates the state transitions of the sites comprising the the MRF. The user can specify the estimation algorithm to be used and also the initialization of the MRF - each site can be initially set to either a NON-EDGE or to its TLR estimate. Except for HCF, the user can also choose to use either a scan-line or a random visiting order. The input MRF is constrained to be a homogeneous one - uniform spatial connectivity and clique potential functions, so as to make the time and space needed to run simulations reasonable.

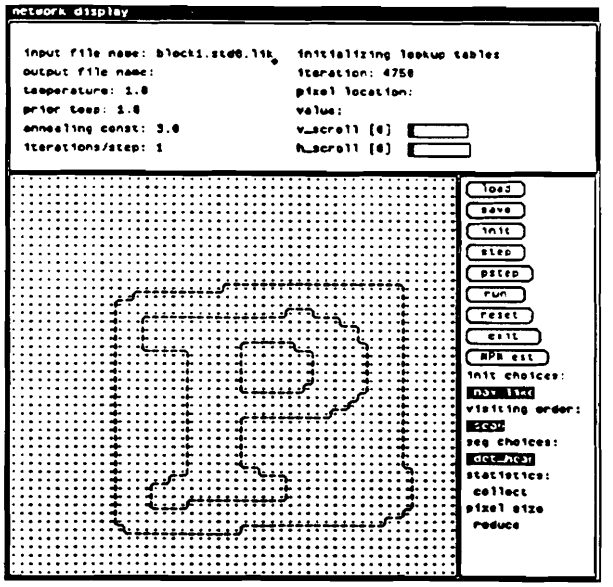

Figure 5.6. Interactive MRF Simulator

The user provides the description of the MRF to the simulator in a file. This file contains a specification of the sites comprising each clique, as well as the potential function associated with it. The user specifies all cliques that a site could belong to in the most general case, including all instances of a particular clique type that contain the site. (*i.e.*, even if all the cliques containing a site are instances of the same clique type, the user specifies each instance separately). The sites forming a clique are specified by their coordinates relative to the site of interest, which is defined to be at relative coordinates (0, 0). Boundary conditions, as in the case of sites near the border of the MRF, are taken care of by the simulator. The potential function is specified as a function that takes as input a configuration vector (a vector of states of sites of the MRF) and returns a potential value. The potential function is associated with the clique description, and the ordering of the site states in the configuration vector passed it is the same as the order the sites are specified in the description of the clique itself.

The simulator performs certain preprocessing actions on the description of the MRF provided by the user, to promote run-time efficiency. The first is to store each potential function as a table indexed into by a configuration vector. This is done so as to avoid run-time calling of the user's potential function code , which can be quite complex, replacing it instead with a simple table lookup. The other is "clique containment", which is based on the observation that if one clique completely contains the other, then a configuration vector of the sites in the larger clique contains implicitly the configuration vector for the smaller clique. This suggests that by judiciously "adding" together the potential functions for the clique in the preprocessing stage, we can avoid run-time evaluation of the potential function for the smaller clique. This simplifies the state transition energy evaluation by reducing the number of terms to be summed up. If floating-point arithmetic is costly, this can save considerable computational effort. The preprocessing needs to be done just once, and can be performed off-line.

The run time operations in a simulation involve reading in the input log likelihood ratios, initializing the MRF configuration according to the user's

specification (all "uncommitted" for HCF), and calculating local characteristics, local energy functions, or the stability measures used by the Gibbs sampler, ICM, and HCF. For Monte Carlo MPM estimations, the simulator collects the statistics about edge occurrence and performs the estimation based on those values after the number of iterations specified by the user.

5.5. Experimental Results

The simulator described above has been used for a series of experiments aimed at comparing the performances of various estimation algorithms with respect to the goodness of final estimates and rate of convergence. We focus upon algorithms based on MRF modeling, including the proposed Highest Confidence First (HCF) (Chapter 4), Iterative Conditional Modes estimation (ICM) [Besag 1986], stochastic MAP (simulated annealing with Gibbs Sampler [Geman and Geman 1984]), and stochastic MPM (Monte Carlo approximation to the MPM estimate [Marroquin, Mitter, and Poggio 1985]). The edge results obtained by applying 3×3 Kirsch operators with non-maximum suppression are also presented for the sake of completeness of comparisons. The annealing schedule for the stochastic MAP follows the one suggested in [Geman and Geman 1984], *i.e.* $T_k = \frac{c}{\log(1+k)}$ where T_k is the temperature for the k^{th} iteration, with $c = 4.0$. The stochastic MAP was run for 1000 iterations and the stochastic MPM for 500 (300 to reach equilibrium, 200 to collect statistics).

5.5.1. Comparison of Estimates

Here we show the results of three sets of experiments (Fig.s 5.7 through 5.9). The figures for each set contain the original image, the result from the Kirsch operators, the TLR configuration and the results obtained by using stochastic MAP, stochastic MPM, ICM (scan-line visiting order), ICM (random visiting order) and HCF algorithms. Except in the case of the HCF algorithm, where the MRF is initialized to all null (uncommitted) states, the MRF is initialized to the TLR configuration. The MRF specification is the same throughout.

Since the final estimates of the stochastic MAP, MPM, and the ICM with random visiting order technically depend on the "seeds" for a pseudo random number generator in addition to the input images, we have observed large variations among the results in repeated runs using the same setup. Theoretically, this should not be true for the stochastic methods because their results asymptotically converge to the true MAP and MPM values. In practice, the results are highly dependent on the configurations of the early iterations. Once a large-scale region or line segment is formed, it is seldom altered in a limited period of time. We have subjectively chosen to show the most typical results in our figures. There are better and worse ones as one might expect.

Fig. 5.7a shows a synthetic 50 pixel square "checkerboard" pattern. Each of patches is 10 pixels across, with an intensity chosen randomly from between 0 and 255. The image has been degraded by independently adding to each pixel Gaussian noise with a mean of 0 and a standard deviation of 16. The results of HCF and stochastic MPM (Figs. 5.7g and 5.7d) are the same, and have completed most of the desired edges. The ICMs (Figs. 5.7d and 5.7e) have incomplete edges and the stochastic MAP has some undesired edges and incomplete desired edges (Fig. 5.7c). The Kirsch operator result is not shown as the edges in this image are always located exactly in between pixels, while the Kirsch operator assumes edges to be at pixel locations, and so a comparison would be unfair to the Kirsch operators.

Fig. 5.8a shows a 50 pixel square natural image of a wooden block with the letter "P" on it. The MAP estimate has several undesirable lines (Fig. 5.8d). The MPM estimate performs poorly on the right edge of the block and the inner ring of the "P". The ICM scheme (serial scan) (Fig. 5.8f) performs better than the random scan version (Fig. 5.8g), but is less than satisfactory on the leg of the "P" and the right edge of the block. The HCF estimate (Fig. 5.8h) does not suffer from the above flaws, producing clean, connected edges.

Fig. 5.9a shows a 100×124 natural image of 4 plastic blocks with the letters "U", "R", "C" and "S" on them. Again, the HCF algorithm produces superior results (Fig. 5.9g). It has the clearest letter outlines and also is alone in detecting the entire bottom

edge of the "R" block. The MAP estimate partially detects the bottom edge of the "R" block, but generates redundant lines (Fig. 5.9c). The MPM estimate has clear letter outlines but does poorly on the outlines of the left blocks (Fig. 5.9d). The ICM scheme (scan-line) does well on the letter outlines but poorly on the block outlines while the random scan version does poorly on both (Figs. 5.9e and 5.9f).

To test the *robustness* of the algorithms, we conduct further experiments using a likelihood generator with a less complete edge model. Since offset edges (Fig. 5.1c) are not considered here, multiple responses become significant as can be seen from the TLR configuration shown in Fig. 5.10a. This change adversely affects the estimates produced by all the algorithms except the HCF, as can be seen from comparing corresponding pictures in Fig. 5.9 and Fig. 5.10.

5.5.2. Rates of Convergence

We restrict ourselves to comparisons between deterministic schemes, as stochastic schemes do not have any convergence criterion *per se* - the point of convergence is dependent upon our judgement as to when equilibrium has been reached, and as to when we have gathered enough statistics to estimate the joint (or marginal) probabilities accurately (typically several hundred iterations are needed). The deterministic algorithms (HCF and ICM (scan-line)) have been timed on images of various sizes using a Sun 3/260 with floating point acceleration. The results are shown in Table 5.1.

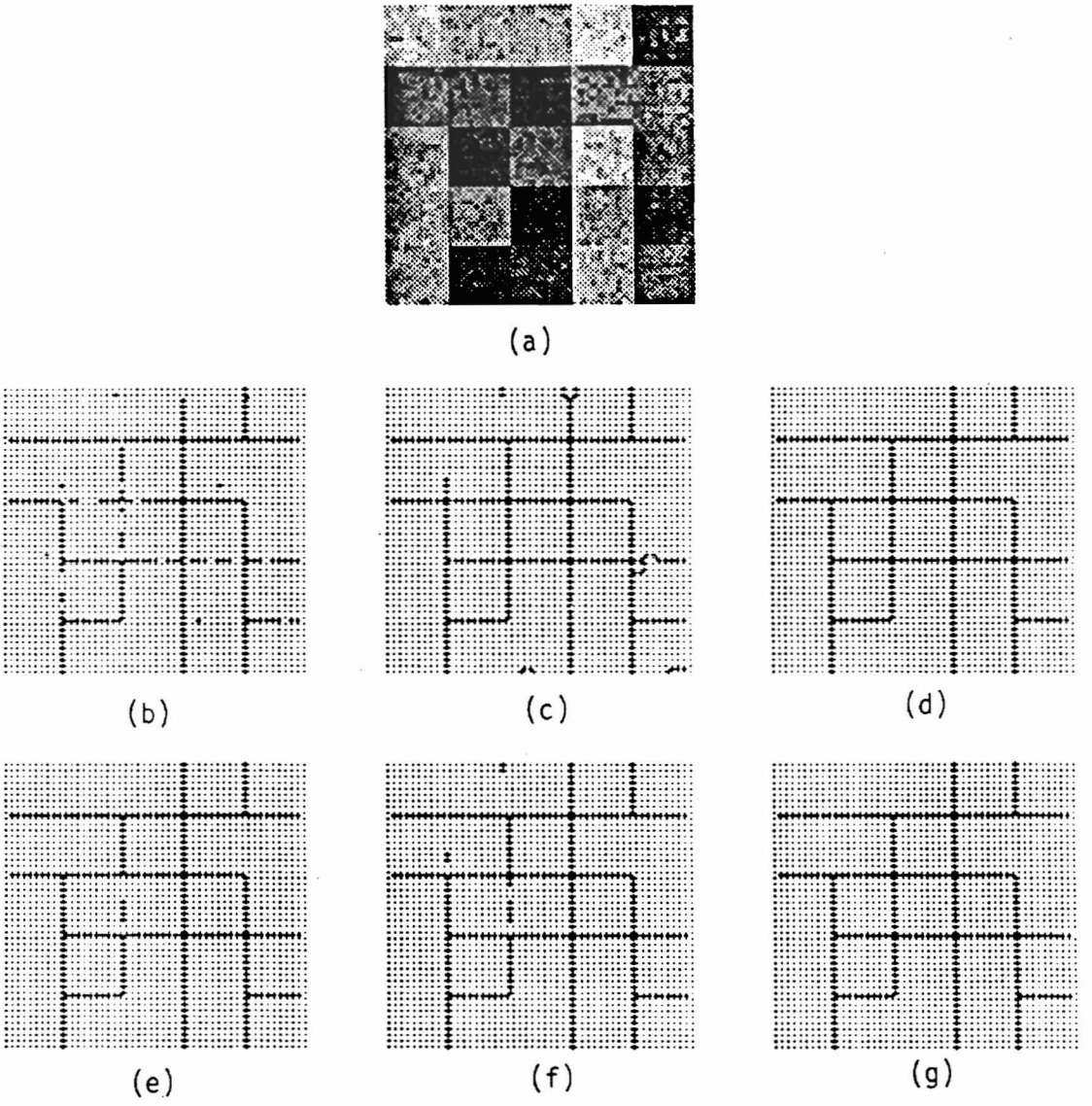

Figure 5.7. Boundary Detection Experiment Set (I)

(a) Synthetic 50×50 "checherboard" image corrupted by independent unbiased Gaussian noise with std. 16. (b) TLR configuration. (c) Stochastic MAP estimate. (d) Stochastic MPM estimate. (e) ICM (scan-line visiting order) estimate. (f)ICM (random visiting order) estimate. (g) HCF result.

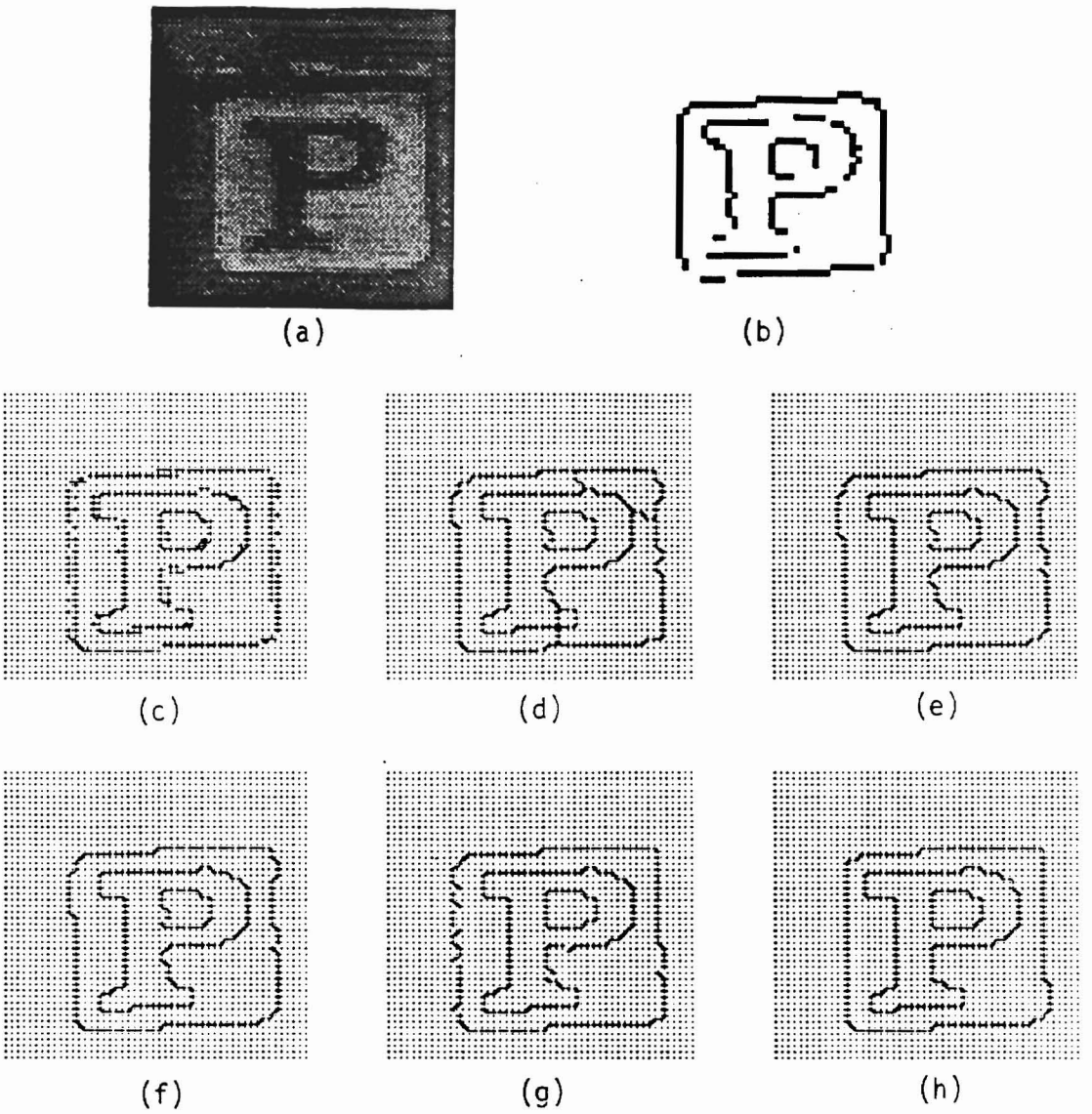

Figure 5.8. Boundary Detection Experiment Set (II)

(a) Natural 50×50 image of wooden block. (b) Thinned and thresholded ouptut of Kirsch operators. (c) TLR configuration. (d) Stochastic MAP estimate. (e) Stochastic MPM estimate. (f) ICM (scan-line visiting order) estimate. (g)ICM (random visiting order) estimate. (h) HCF result.

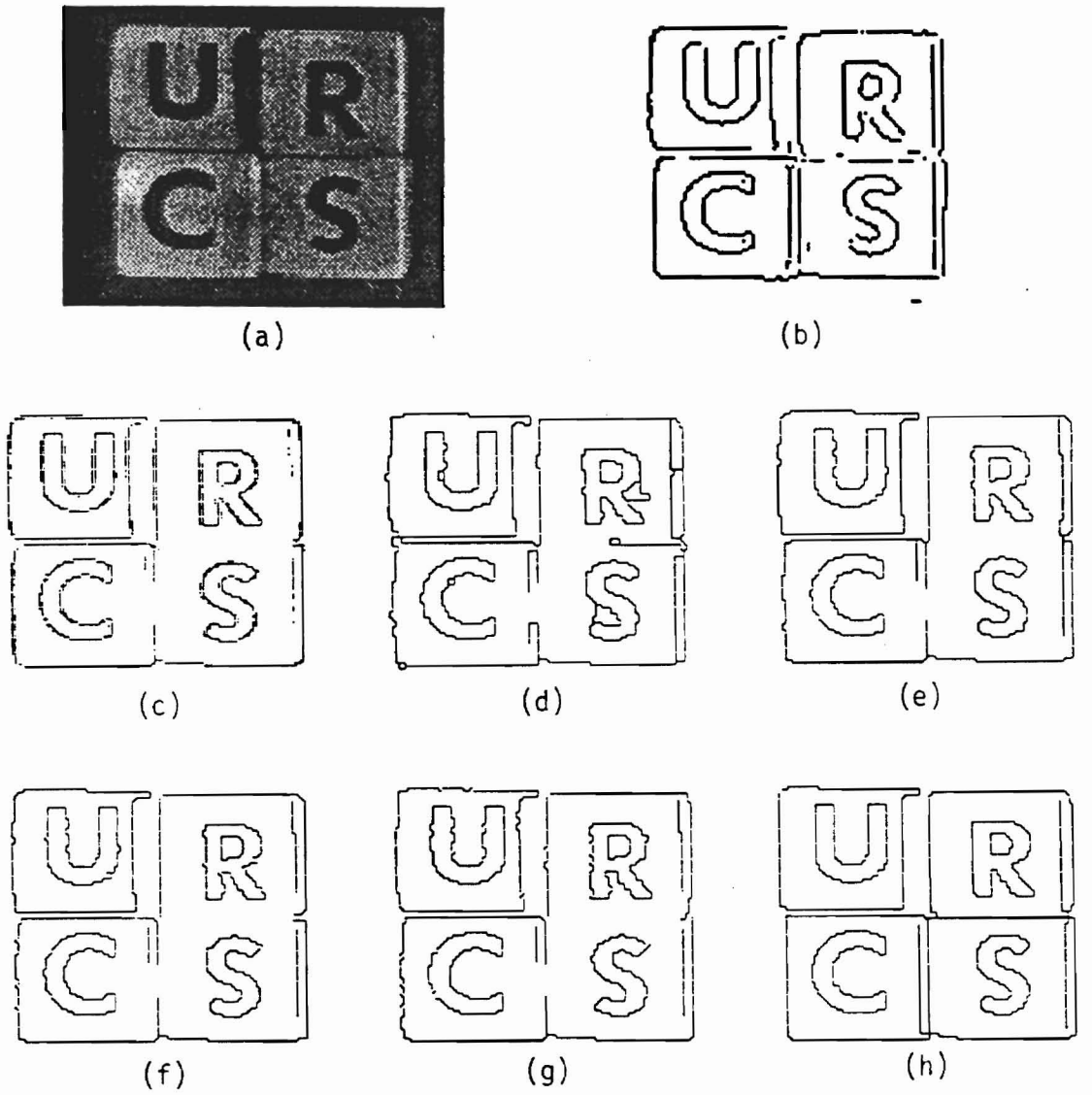

Figure 5.9. Boundary Detection Experiment Set (III)

(a) Natural 100×124 image of four plastic blocks. (b) Thinned and thresholded ouptut of Kirsch operators. (c) TLR configuration. (d) Stochastic MAP estimate. (e) Stochastic MPM estimate. (f) ICM (scan-line visiting order) estimate. (g)ICM (random visiting order) estimate. (h) HCF result.

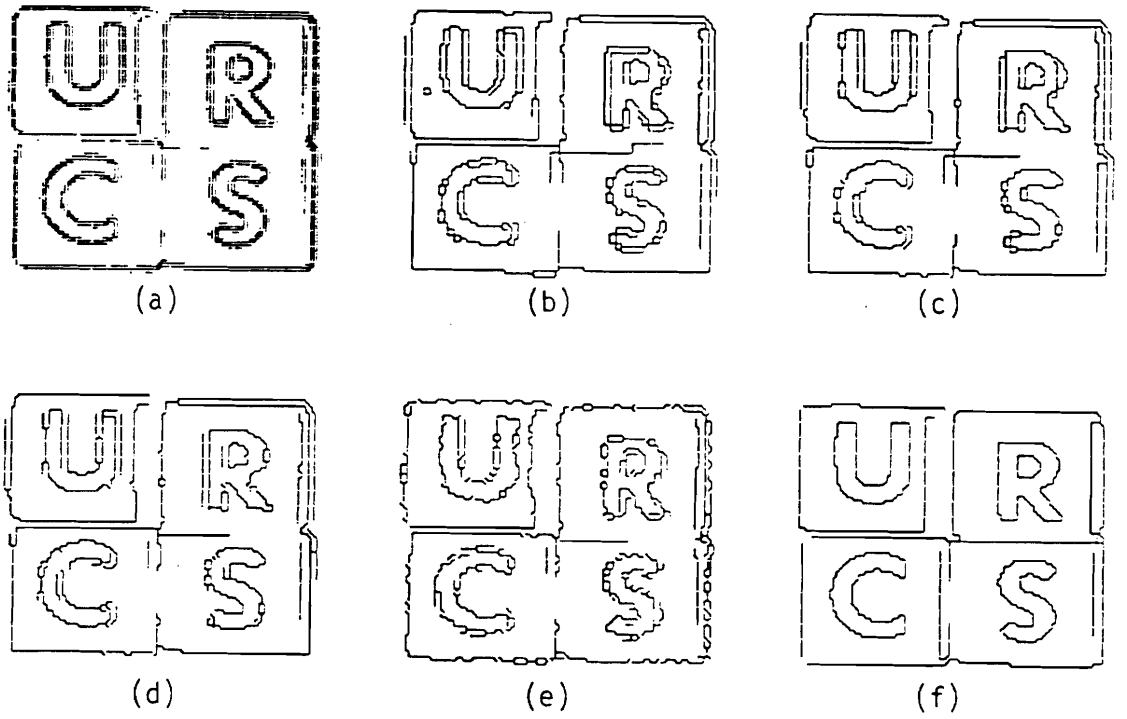

Figure 5.10. Boundary Detection Experiment Set (IV)

Experiments with incomplete edge model - original image in Fig. 5.9.a. (a) TLR configuration. (b) Stochastic MAP estimate. (c) Stochastic MPM estimate. (d) ICM (scan-line visiting order) estimate. (e)ICM (random visiting order) estimate. (f) HCF result.

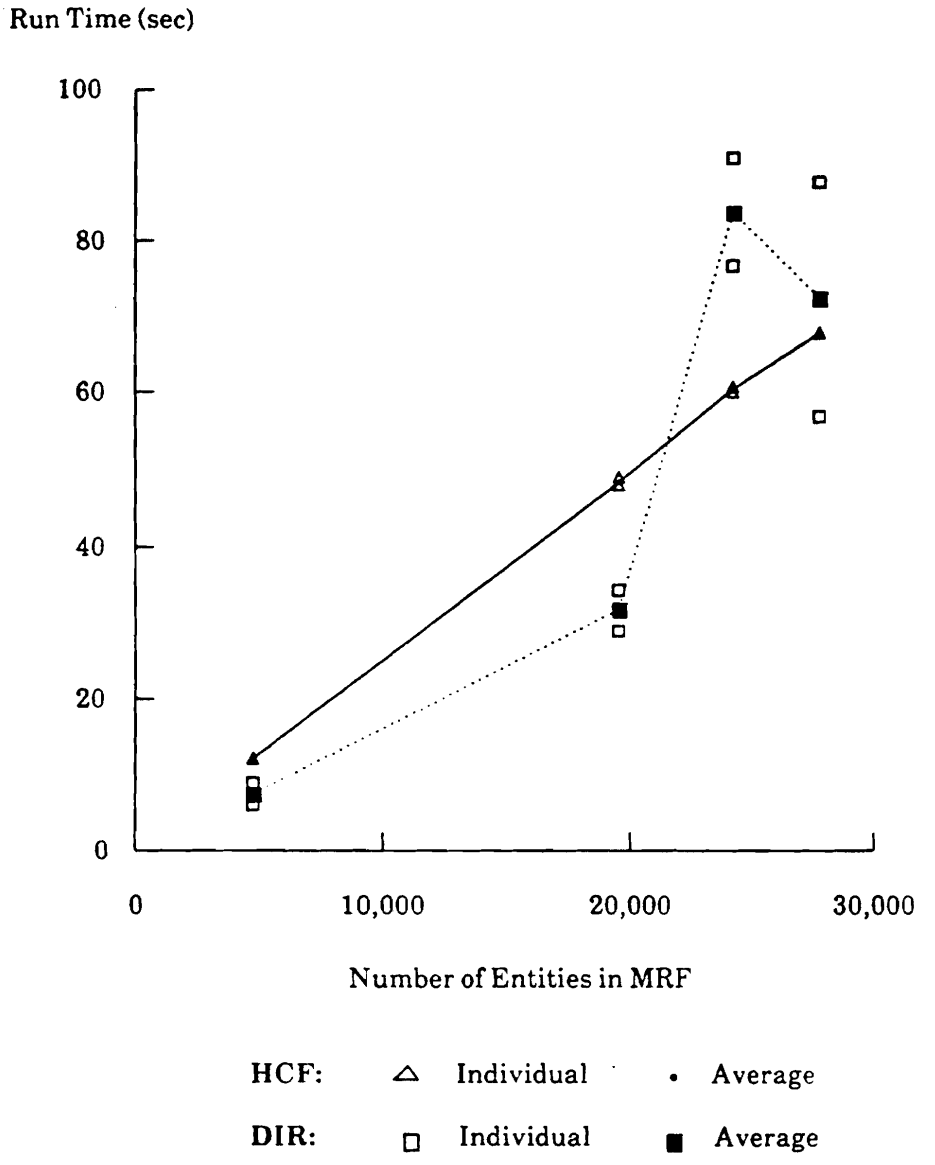

Table 5.1. Timing Test Results

The HCF and ICM algorithms are each run on two images of the same size, for four image sizes. Individual and average run-times are shown.

5.6. Analysis of Experimental Results

Goodness of Estimates

(1) The HCF algorithm repeatedly outperforms all other algorithms, giving superior results both with synthetic and real image data. The common characteristics of the results we have obtained from using this algorithm are that they all fit well in our model of the world, which consists of smoothly continuous boundaries, and that they are consistent with the observations.

(2) The HCF algorithm also appears to be robust, in that it produces an estimate consistent with the observations even when the MRF model used is inadequate, as in the experiment using the less sophisticated edge detector. Since our MRF model does not take into account multiple responses, the MAP criterion may not lead to the "best" results. In this case, the local minimum found by the HCF algorithm is clearly better than the results produced by other methods as it is based on the strength of external evidence.

(3) The ICM algorithm performs inconsistently and its results depend to a large extent upon the initialization of the MRF and the visiting order. It is also not clear which, if any, of the visiting orders studied is better than the other. The scan-line visiting order performs better in some of our experiments, but it is due to the horizontal and vertical characteristics of the boundaries. HCF does not rely on any predefined order, thus is not biased for any boundary shape.

(4) The stochastic MAP algorithm with simulated annealing gets stuck in undesirable local minima, suggesting that our annealing schedule might have lowered the temperature too fast. However, an appropriate annealing schedule seems hard to obtain *a priori*. We have conducted further experiments with simulated annealing with varying annealing constant c for 1000 iterations each (Figure 5.11). It appears that starting with temperature too high will destroy the TLR initial estimate, resulting in estimates inconsistent with the input data. The Monte Carlo MPM estimates are more reliable than simulated annealing results in most cases. However, we

occasionally observed large-scale mistakes.

(5) In addition to the qualitative comparisons, we have evaluated the results in terms of energy measures (Table 5.2). Recall that, as discussed in Chapter 4, the energy measures may not reflect the correctness of the estimates in the presence of significant modeling errors. The comparisons based on these measures serve the purposes of verifying the validity of our potential assignments, and, more importantly, identifying the effectiveness of HCF as an energy minimization strategy for similar applications. It is worth mentioning that in our many trials, HCF consistently found better local minima in all but one case when the Monte Carlo MPM beat HCF by 0.1%.

Convergence Times

(1) The HCF algorithm makes a perhaps surprisingly small number of visits before converging. Clearly, due to the initialization, it must visit every site at least once. What is surprising is that it visits each site on the *average* less than 1.01 times before converging. What this implies is that the first decision made by a site is nearly always the best one. Also, the HCF algorithm takes

Fig.	TLR	MAP	MPM	ICM(s)	ICM(r)	HCF
5.7	-3952	-4282	-4392	-4364	-4334	-4392
5.8	-572	-680	-723	-693	-715	-740
5.9	4785	-349	-503	-503	-513	-629
5.10	59719	-5303	-5296	-4954	-3728	-9587

Table 5.2. Energy Values

Quantitative comparisons between results of simulated annealing (MAP), Monte Carlo (MPM), ICM (scan-line order), ICM (random order), and HCF. Column Fig. lists the figure numbers of the input images. Some of the values (of MAP, MPM. ICM's) are the averages of the results from several runs. (The smaller the energy the better the estimate.)

almost the same time on different images of the same size.

(2) The convergence times of the ICM algorithms are unpredictable - they vary with visiting order, MRF initialization and even upon the particular image given as input.

(3) The time taken by the HCF algorithm includes the time taken to set up the heap initially. This may, in some circumstances, be a little unfair. For instance, if one has to process data online from various information sources (Chapter 3) [Poggio 1985], the heap setting up cost can be treated as a preprocessing cost rather than a run-time one. In theory, the time taken by the HCF algorithm should be given by $c_1 N + c_2 V \log_2 N$, where c_1 and c_2 are positive constants, N the number of sites to be labeled and V the number of visits. V here is at least N and we conjecture that on the average it is cN for some small ($1<c<2$) constant c. Since the latter term should dominate, one would expect to see a nonlinear curve in a plot of run time *vs.* number of sites. However, the curve is very nearly a straight line. which indicates either that the constant c_2 is very small, or that the changed stability values do not propagate very far up the heap on the average. The former does not appear to be true, as our experiences suggest that the initial heap construction takes far less time than the rest of the algorithm.

5.7. Discussion

In this chapter, we have applied the Markov Random Field formalism to the boundary detection problem, and compared the results of several estimation methods based on the MRF formalism. The experiments strongly support the claims that the probabilistic approach to the labeling problem is appropriate, and that HCF performs better in both robustness and efficiency than previous methods. There is much left to do, however, to build a boundary detection system based on this approach that is capable of dealing data from various domains in unfamiliar environments.

The 1×4 and 4×1 edge likelihood operators are certainly not adequate. Designing different size and shape operators in accordance with different feature

types, resolutions, and noise levels will certainly improve the resulting likelihood estimates. Sher [Sher 1987a] has discovered that by judiciously combining different operators aimed at different domains, one can obtain results better than the ones by using any single operator based on wrong domain assumptions, and nearly as good as the optimal one. We have not yet incorporated this result in our system, but we are confident that work along this direction will increase the robustness of the system.

Another interesting topic of obvious utility is to estimate or "learn" the line MRF parameters from (noise corrupted) realizations of the intensities. The current choices of the neighborhood and the potential functions are by no means optimal, and they should be adjusted according to the domain characteristics. It would probably be practical to start with the current setup, which has demonstrated good results, and devise a method to improve the potential assignments given new realizations. Ultimately, the assumptions of homogeneity and isotropy may not be valid, especially when the lattice-structured MRF model is extended to general graphs for higher-level applications.

The success of HCF in energy minimization is interesting. It seems quite possible that similar successes could be achieved by applying the HCF heuristic to other domains involving combinatorial optimization.

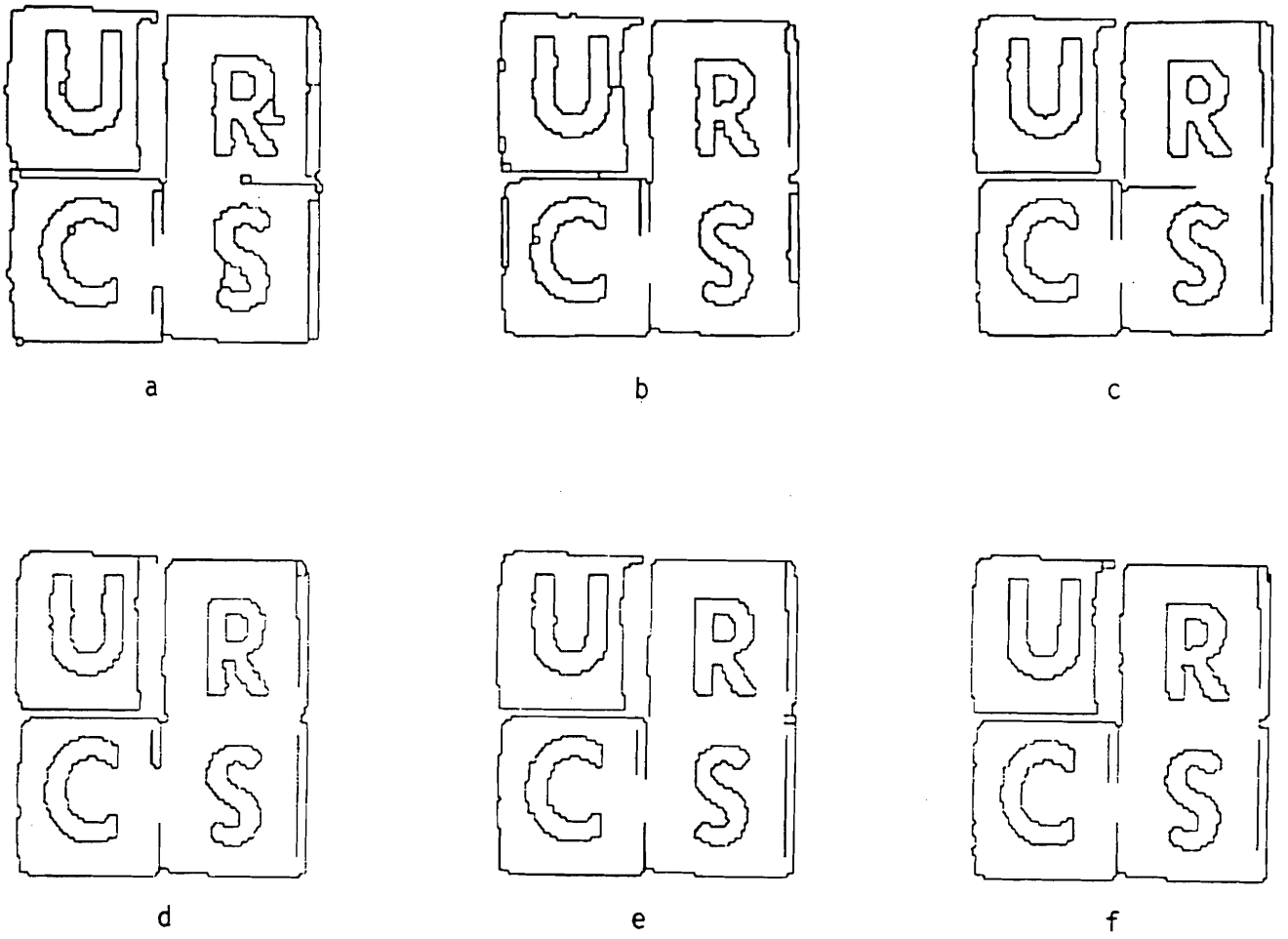

Figure 5.11. Annealing Results

Experiments with simulated annealing (MAP) procedure with different annealing constants (schedules). (a) c = 4.0 (b) c = 3.5 (c) c = 3.0 (d) c = 2.5 (e) c = 2.0 (f) c = 1.5.

Chapter 6 Depth Map Reconstruction

6. Segment/Reconstruct Depth Maps by Incorporating Intensity Edge Information with Sparse Depth Measures

It has been proposed that visual integration occurs naturally during the reconstruction of the visible surfaces [Marr 1982], and is best performed at the locations of discontinuities [Gamble and Poggio 1987]. For example, one problem may be to reconstruct a depth map and to detect its discontinuities simultaneously from a (sparse) set of corrupted depth observations, possibly with the aid of other sources of information. This chapter reports the work in incorporating intensity discontinuity observations to reconstruct and segment such a depth map.

The reconstruction of three-dimensional scene parameters (intrinsic images) from visual information is often accomplished using a smoothness assumption to regularize the computation. Smoothing is not wanted across object boundaries, and reliable reconstruction can not be achieved without the detection of the discontinuities [Stuth, Ballard, and Brown 1983]. On the other hand, discontinuities are best described as boundaries between surface patches defined by the corresponding scene parameters, thus can not be detected directly from sparse, noisy data. The cooperation of reconstruction and discontinuity detection has been of interest for some time: The challenge is to develop a unified treatment for reconstruction and segmentation. The mechanism we use is *coupled MRF's*, in which MRF's, one for the depth process and one for the discontinuity process, work in parallel and interact.

In fusing depth and intensity information, Gamble and Poggio [1987] use the intensity edges detected with the Canny operator [Canny 1983] to constrain the locations of the depth discontinuities while reconstructing a depth map. Their rule is that no depth discontinuity is allowed without a corresponding intensity discontinuity. The results of combining the two information modalities are encouraging and better than either modality operating alone, but the uncompromising relation between depth and intensity discontinuities means that depth discontinuities within regions of little intensity variation will be lost even if the depth information is good. The problem is thus how to assign a general *a priori* relation between depth

and intensity information.

Another purpose of this work is to study whether the concept of Highest Confidnence First can be applied to complex problems with both numerical and symbolic labels. To use the HCF algorithm to fuse depth and intensity based on coupled MRF's, three issues arise: how to specify the energy functional for the continuous depth label, how to specify the "confidence" or "stability" of a site for the HCF algorithm, and how to modulate the depth and discontinuity estimations.

In the remainder of this chapter, we give technical details of the formulation of the fusion problem and present some experimental results and directions of future work.

6.1. Coupled Markov Random Fields

Represent a pixel image $S = \{s_1, s_2, \ldots, s_N\}$ as a set of lattice-structured sites, and the discontinuity image, D, as the set of sites placed midway between each vertical and horizontal pair of pixel sites. Let $F=\{f_s, s \in S\}$ be the set of random variables indexed by S, with $f_s \in R$ representing the depth value at location s, and $L=\{l_d, d \in D\}$ be the set of random variables indexed by D, with $l_d \in \{0,1\}$ representing the absence or presence of a depth discontinuity at site d. F and L correspond to the depth process and the line (depth discontinuity) process respectively of the coupled Markov Random Fields introduced by Geman and Geman [Geman and Geman 1984] and used in [Marroquin, Mitter, and Poggio 1985], [Gamble and Poggio 1987], and this thesis. A configuration of (F,L) corresponds to an admissible solution to our problem.

6.1.1. Observation Models

Early Depth Measurements: The depth measurements are considered sparse and independently measured. Denote by $\hat{S}$ the set of sites in S at which depth measurements are available and $G=\{g_s, s \in \hat{S}\}$ the measurement process. We assume

$$P(G=g \mid F=f) = \prod_{s \in S} P_s(g_s \mid f_s) \tag{6.1}$$

where g_s denotes the measurement at pixel site s, and g represents these measurements. Often the noise can be adequately modeled by unbiased Gaussian distributions. That is

$$P_s(g_s \mid f_s) = \frac{1}{z_s} e^{\frac{-(f_s - g_s)^2}{2\sigma_s^2}} . \tag{6.2}$$

Discontinuity Observations Based on Intensity: Instead of treating intensity edges as constraints on the locations of depth discontinuities, we consider them as partial evidence supporting or refuting the hypotheses about depth discontinuities. The motivation is simple. The intensity images are the results of many confounding factors - lighting, surface geometry, surface reflectance, and camera characteristics. Intensity discontinuities may reflect sudden changes of depth values, but depth discontinuities do not necessarily imply large intensity variations. Figure 6.1 shows

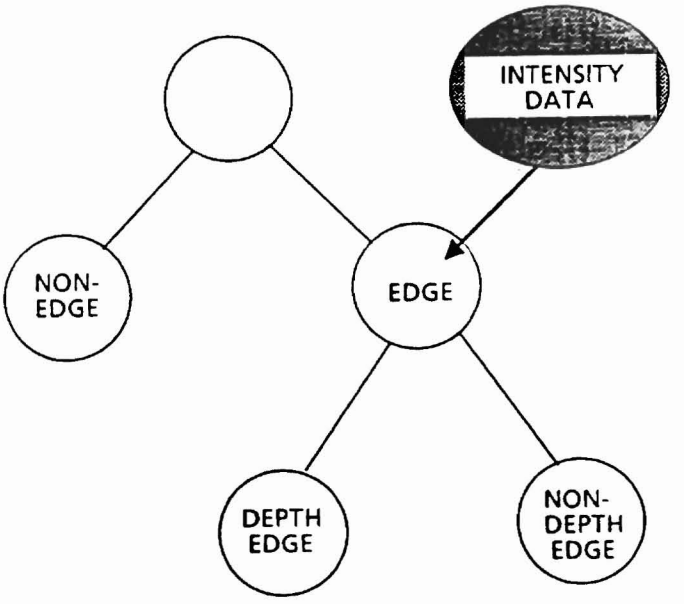

Figure 6.1. Relation between Intensity and Depth Edges

the conceptual hierarchy that consists of the interesting events involved here. At the first level, only *EDGE* or *NON–EDGE* is of concern. Node *EDGE* represents the event that the site of interest corresponds to some sort of discontinuity in the world; *NON–EDGE* represents the event that the site is within a homogeneous region. At the next level, whether a particular intensity discontinuity is due to depth discontinuity becomes interesting. Intensity observations provide information about the events in the first level, but say nothing about the events in the second level, which are important to the depth segmentation problem. Our approach is to incorporate prior experience and knowledge, represented in terms of conditional probabilities, to infer the amount of support, represented as likelihood ratios, to the events of depth discontinuities provided by the intensity observations. To be more precise, we are interested in computing the likelihood ratio of a site being a *DEPTH–EDGE* given the numbers α_0, α_1, and α_2 constantly proportional to the likelihoods $P(O \mid NON\text{–}EDGE)$, $P(O \mid DEPTH\text{–}EDGE)$, and $P(O \mid EDGE\text{–}NON\text{–}DEPTH)$ (Chapter 3), and the conditional probability $p = P(NON\text{–}EDGE \mid NON\text{–}DEPTH\text{–}EDGE)$, where *NON–DEPTH–EDGE* stands for the joint event $NON\text{–}EDGE \vee EDGE\text{–}NON\text{–}DEPTH$. We have

$$\frac{P(O \mid DEPTH\text{–}EDGE)}{P(O \mid NON\text{–}DEPTH\text{–}EDGE)} = \frac{\alpha_1}{p\,\alpha_0 + (1-p)\,\alpha_2} \tag{6.3}$$

In the rest of the paper, λ_d denotes the likelihood ratio of site d given the intensity observation O_d, where $d \in D$.

$$\lambda_d(l_d) = \frac{P(O_d \mid l_d)}{P(O_d \mid \neg l_d)} . \tag{6.4}$$

Again, we consider the spatially distinct intensity observations are conditionally independent:

$$P(O \mid l) = \prod_{d \in D} P_d(O_d \mid l_d) , \tag{6.5}$$

where O denotes the collection of intensity observations.

Conditional Independence between Intensity and Depth Observations: We assume that the depth and intensity observations are only related through the geometry of the surfaces in view. They are conditionally independent in the following sense:

$$P(\boldsymbol{g},\boldsymbol{O} \mid \boldsymbol{f},\boldsymbol{l}) = P(\boldsymbol{g} \mid \boldsymbol{f},\boldsymbol{l})\, P(\boldsymbol{O} \mid \boldsymbol{f},\boldsymbol{l})\,. \tag{6.6a}$$

We further assume that the knowledge of depth discontinuities contributes no information to make one prefer the observation of $\boldsymbol{g}$ over others once the true depth values are known:

$$P(\boldsymbol{g} \mid \boldsymbol{f},\boldsymbol{l}) = P(\boldsymbol{g} \mid \boldsymbol{f})\,, \tag{6.6b}$$

and that the knowledge of surface depth does not make $\boldsymbol{O}$ more or less likely once the depth discontinuities are known:

$$P(\boldsymbol{O} \mid \boldsymbol{f},\boldsymbol{l}) = P(\boldsymbol{O} \mid \boldsymbol{l})\,. \tag{6.6c}$$

The scene depth, in many circumstances, affects the observed intensity values. The assumption (6.6c) is reasonable, however, since it is the indirect observations of intensity discontinuities but not the magnitude of the intensity that are actually used in this work. Thus in this work we discard intensities after computing likelihoods of discontinuities. An interesting research problem would be to use the intensity information (perhaps through the irradiance-orientation constraint [Horn 1975]) more directly.

Summarizing (6.1) - (6.6), we assume

$$P(\boldsymbol{g},\boldsymbol{O} \mid \boldsymbol{f},\boldsymbol{l}) = \prod_{s \in S} P_s(g_s \mid f_s) \prod_{d \in D} P_d(O_d \mid l_d)\,. \tag{6.7)}$$

6.2. Markov Random Fields and Energy Measures

Within each F and L, spatially adjacent variables tend to have similar values. That is, surfaces and boundaries tend to be continuous and smooth. MRF's corresponding to F and L can be separately defined to model these properties. Chapter 5 has demonstrated some promising edge detection results using an MRF for the line process alone. The depth and line processes, however, are not independent of

each other. The presence of a line at an edge site breaks the connection between the two variables at the adjacent pixel sites; a small change in the values of two adjacent depth variables suggests the absence of a discontinuity in between. This interdependence is the basis for the concept of coupled MRF's - an unified treatment of reconstruction and segmentation. Figure 6.2 shows a neighborhood system Γ of the MRF consisting of the depth and line processes. In addition to the depth and line processes, the concept of coupled MRF's can also be applied to model many other interdependent processes corresponding to various intrinsic parameters [Poggio 1985].

(F,L) is an MRF with respect to a neighborhood system Γ if and only if, according to Hammersley-Clifford theorem, the joint probability distribution of the variables is a Gibbs distribution. That is,

$$P(f,l) = \frac{1}{Z} e^{\frac{-U(f,l)}{T}}, \tag{6.8a}$$

where the energy functional

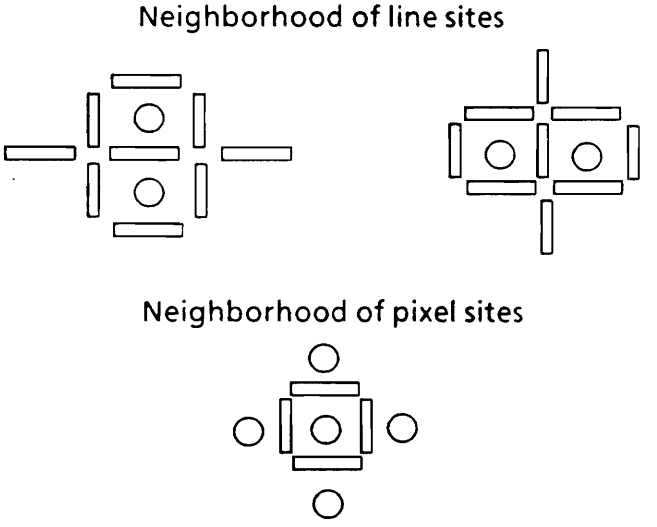

Figure 6.2. Neighborhood System for Coupled MRF's

$$U(f,l) = \sum_{c \in C} V_c(f,l) \,. \tag{6.8b}$$

where C is the set of cliques defined by Γ. Continuous surfaces can be modeled by setting the potential energy V for the cliques consisting of two adjacent depth sites, say i and j, and the line site in between them , say ij, proportional to $(1-l_{ij})(f_i-f_j)^2$ [Marroquin, Mitter, and Poggio 1985] [Gamble and Poggio 1987]. Using this potential function, minimizing the energy measure has the effect of fitting membrane patches to the lattice. Higher-order spline surfaces can similarly be encoded with larger neighborhood systems to account for higher-order derivatives. Since only depth discontinuties are concerned here, we use the neighborhood system depicted in Fig. 6.2 and the above potential function throughout our experiments. Other types of cliques that have non-zero potential functionals used in our experiments consist only of line sites. They are same as the ones described in Chapter 5.

6.3. A Posteriori Energy

Bayes' rule combines the *a priori* knowledge and the early visual observations to derive the *a posteriori* belief.

$$P(f,l \mid g,O) = \frac{P(f,l)P(g,O \mid f,l)}{\sum_{f,l} P(f,l)P(g,O \mid f,l)} \,.$$

Note, from (6.1) and (6.5), that scaling all of the likelihoods for a fixed site by a constant does not change the posterior distribution of (F,L). From (6.7) and (6.8), and assuming (6.2), the posterior distribution is a Gibbs distribution, with the *a posteriori* energy functional

$$U(f,l \mid g,O) = \sum_{c \in C} V_c(f,l) + T\Big(\sum_{s \in \hat{S}} \frac{(f_s-g_s)^2}{2\sigma_s^2} - \sum_{d \in D} \log \lambda_d(l_d) \Big) \,. \tag{6.9}$$

6.4. HCF: Coping with Continuous Variables

Let ζ denote the uncommitted state, and $\bar{R} = R \cup \{\zeta\}$, $\bar{L} = \{\zeta, 0, 1\}$ denote the *augmented state spaces* for the depth and line processes respectively. Based on (6.9), define the *augmented local energy* measures with respect to an augmented

configuration (f,l) as:

$$E_s(f) = \sum_{c:s\in c} V'_c(f',l) + T\,\frac{(f_s-g_s)^2}{2\sigma_s^2} \quad \text{for } s\in\hat{S}\,, \tag{6.10a}$$

$$E_s(f) = \sum_{c:s\in c} V'_c(f',l) \qquad \text{for } s\in S-\hat{S}\,, \tag{6.10b}$$

and

$$E_d(l) = \sum_{c:d\in c} V'_c(f,l') - T\log\lambda_d(l) \quad \text{for } d\in D\,, \tag{6.10c}$$

where (f',l') agree with (f,l) everywhere except $f'_s = f$ and $l'_d = l$, with $(f,l)\in(R,L)$. $V'_c = 0$ if there is an uncommitted site in c, otherwise it is equal to V_c. Thus the cliques containing uncommitted sites have no effect on the augmented energy measures. Since the only cliques involved in (6.10a) and (6.10b) are those consisting of two neighboring pixel sites and a line site in between, the terms $\sum_{c:s\in c} V'_c(f',l)$ can be written as $\sum_{r\in N_s} \beta(1-l_]) (f_r-f_s)^2$, where N_s is the pixel neighborhood of s. The augmented local energy measures for pixel sites thus are quadratic; the shape of each quadratic depends on the constant parameter β, the number of active neighbors, and the variances of the noise in the early measurements. The temperature T is set to 1 throughout our experiments, therefore β decides the degree of smoothness relative to the magnitude of noise.

6.5. Stability Measures

The confidence of a site in a configuration (f,l) is evaluated in terms of the following *stability* measures:

$$\begin{aligned} G_s(f,l) &= \Delta E_s(f_{\min}, f_{\min}+\alpha) \quad \text{if } f_s=\zeta\,, \\ &= \Delta E_s(f_{\min}, f_s) \qquad \text{otherwise}\,, \end{aligned} \tag{6.11a}$$

where $E_s(f_{\min}) = \min_{f\in R} E_s(f)$, and

$$G_d(f,l) = \max_{l\in L, l\neq l_{\min}} \Delta E_d(l_{\min},l) \quad \text{if } l_d=\zeta,$$
$$= \Delta E_d(l_{\min},l_d) \qquad \text{otherwise}, \tag{6.11b}$$

where $E_d(l_{\min}) = \min_{l\in L} E_d(l)$. The term $\Delta E_r(k,j)$ is defined as $E_r(k) - E_r(j)$ with respect to (f,l). It represents the change in local energy measure of r, thus the global energy (e.g. (6.9)), if r should switch its state from j to k. The stability of a site is nonnegative only when it is in its minimal energy state (i.e. $l_{\min}$ or $f_{\min}$) with respect to its current local energy measure. A large negative stability value signals high confidence in making a state change to the minimal energy state. The constant α determines the stability of uncommitted pixel sites: it gives the price paid in energy for remaining uncommitted. α has the semantics of offset along R from state of minimal energy. Using it, the stability measure for uncommitted states has the semantics "how much energy could be lost by committing." Large α encourages quicker commitment.

6.6. Convergence Properties

The HCF network behaves as follows. Every site starts in the uncommitted state. At any instant, only the sites with the highest confidence in changing their states, i.e., the least stable ones with respect to the current configuration, are allowed to change their states. The identities of the sites, pixel or discontinuity, are ignored in the process of comparing the stability measures. Thus the reconstruction and segmentation processes proceed simultaneously. Eventually, the network settles at a configuration when no further reduction of the global energy measure can be made at each site; i.e., when all local stability measures are nonnegative.

The convergence property can be easily verified (Chapter 4). It is possible, however, that the final configuration contains uncommitted depth sites, due to the sparseness of the depth data. In the initial stage of the computation, $E_s(f) \equiv 0$ if $s\in S-\hat{S}$. This local energy measure, and thus the stability measure, will remain zero until one of the pixel neighbors becomes committed and the line process indicates there is no discontinuity in between. If a region of pixel sites, consisting only of

members of $S-\hat{S}$, was surrounded with discontinuities before any of them has nonzero stability measure, all of them will stay uncommitted. In the extreme case, if there is no depth measurement at all, this network will not produce an estimate of the depth map. This is an advantageous feature since in such degenerate cases, there are infinite number of configurations that have the minimal energy measure. It is important for the low-level process to indicate the lack of information to higher-level processes so that attention can be directed to acquire more information. This feature can be turned off, if desired, by assigning *a priori* estimates (e.g. expected range of the scene) to those sites.

An enhanced version of the graphical simulator (Chapter 5) was built for experiments on coupled depth and intensity fields with HCF optimization. We implemented the HCF network on a serial machine using a binary heap to decide the visiting order of the sites. At any instant, the top element of the heap is the site with the smallest stability measure. A state change made by a site in general changes the stability measures of the sites in its neighborhood. The number of comparisons in maintaining the heap property for each change is limited by the height of the heap -- $\simeq \log_2(3N)$ where N is the number of pixels. Ideally, the computation terminates when the top element has nonnegative stability since no more energy reduction would be possible afterwards. In practice, a small (negative) threshold is used to force termination without noticeable degradation of depth value (see below). Some threshold would have been necessary in any case because of limited precision in the calculations.

6.7. Experiments and Results

6.7.1. Synthetic Scenes

The enhanced HCF algorithm, which reconstructs depth and finds depth discontinuities from a pair of depth and intensity images, is demonstrated on two synthetic scenes. Each scene consists of a range image and an irradiance image. Noise of a particular description is added independently to each image. The range image is sampled either at full resolution or randomly at reduced resolution (this

section reports experiments with 100%, 80%, and 50% of the original full-resolution data points kept in the sparse data.) We assume that 95% of *NON–DEPTH–EDGE* events are *NON–EDGE*.

The first sequence shows experiments with images created by utilities in the PADL-2 solid modeling system [Brown 1982] Figures 6.3a and b show original full-resolution intensity and depth images, and Figs. 6.3c and d show the images with added noise. The intensity image has a range of pixel values in [0,255], and is perturbed by zero-mean, signal-independent Gaussian-distributed noise $G(0, \sigma)$ with σ=16. Perturbed values less than 0 are set to 0, greater than 255 are set to 255. The range image has an unusual noise pattern. Part of the motivation is to test the effects of spatially nonuniform noise (the noise model affects the stability calculations of elements). Another motivation is to reflect a range imaging system whose values are more accurate near its optic axis, perhaps as an effect of reduced resolution (averaging) in the periphery. The noise distribution for the synthetic range image is radially symmetric around the center of the image, with standard deviation of the additive, signal-independent, zero-mean Gaussian noise at a point increasing exponentially as the distance of the point from the center of the image. The exponential is scaled so the maximum noise has σ=20 (in the corners of the image.)

Figures 6.4a and b show the reconstructed depth and the depth discontinuities found with a high setting of the parameter β, which increases sensitivity to small depth discontinuities. Here orientation changes in the cube and noise in the sphere boundary have given rise to a spurious segmentation. In Figs. 6.4c and d, a smaller β avoids the problems.

Figures 6.5 and 6.6 illustrate the effect of sparse depth data on the final depth reconstruction and segmentation, and also show the beneficial effects of incorporating intensity information. In these four experiments α and β are held constant. Figs. 6.5a and b show the reconstruction and segmentation using just 80% of the depth information. It is interesting that the results are no worse; however Figs. 6.5c and d show the improvement gained by allowing the MRF access to the irradiance image as well. Fig. 6.6 is precisely analogous to Fig. 6.5, only the depth

density is down to 50%.

Figures 6.7a and b show the "depth" and "irradiance" parts of an artificial scene of the sort used by Sher [Sher 1987a]. Such scenes have the advantage of having well-specified right and wrong locations for edges. In this case both depth and intensity images have spatially-independent, zero-mean, signal-independent, additive Gaussian noise, and the perturbed image values are clipped to the range [0, 255]. For the depth image, σ=16, and for the intensity image σ=20. Fig. 6.7c shows the depth edges recovered using only the depth image (Fig. 6.7a), and Fig. 6.7d shows the depth edges recovered using depth (full resolution) and intensity. As expected, 6.7c only shows and edge structure related to the brightness differences of squares in Fig. 6.7a. Fig. 6.7d has clearly incorporated information from the intensity image Fig. 6.7b. The ideal desired is of course that all lines of the checkerboard are in evidence. It can be seen that the lines are missed when the evidence in both images is weak.

Figure 6.8 shows the effects of sparse depth in the domain of Fig. 6.7. Fig. 6.8a gives a glimpse into the inner state of the MRF algorithm. It shows the initial thresholded likelihoods (line elements) used by the intensity discontinuity detector, overlaid on the points where sparse (50%) depth information is available. Figs. 6.8b and c show the reconstructed depth and depth discontinuities, respectively. The areas of bad performance correlate with areas of weak or missing information.

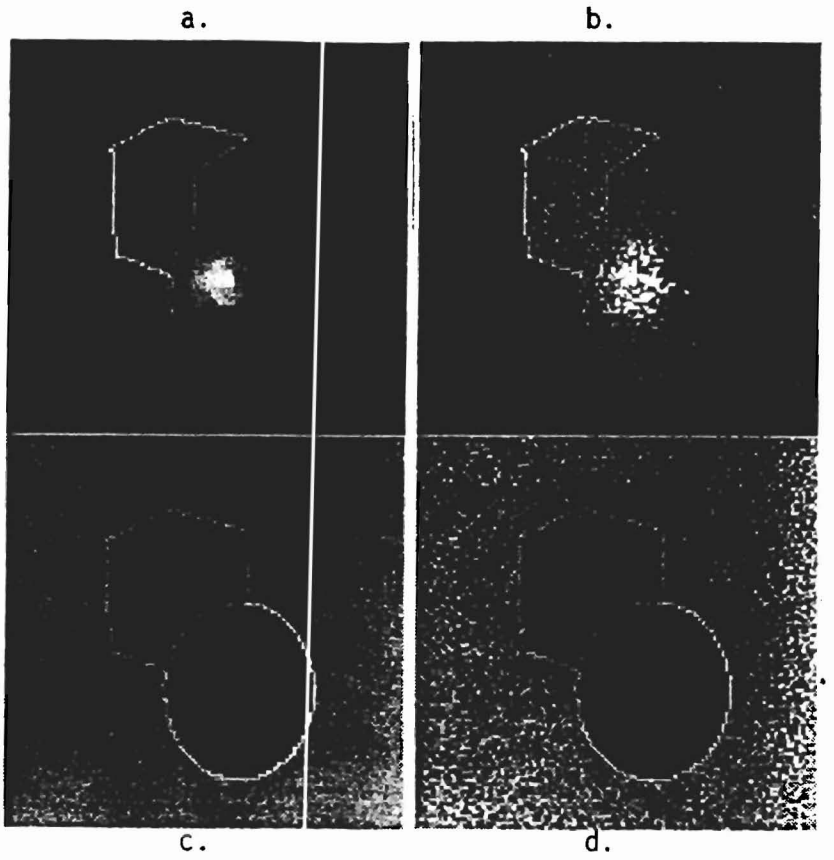

Figure 6.3. Synthetic Intensity and Range Data

a) Original intensity image. b) Original depth image. c) Intensity image with G(0,20) additive noise, clipped to range [0,255]. d) Depth image with spatially-varying Gaussian noise, maximum standard deviation of 20 (see text).

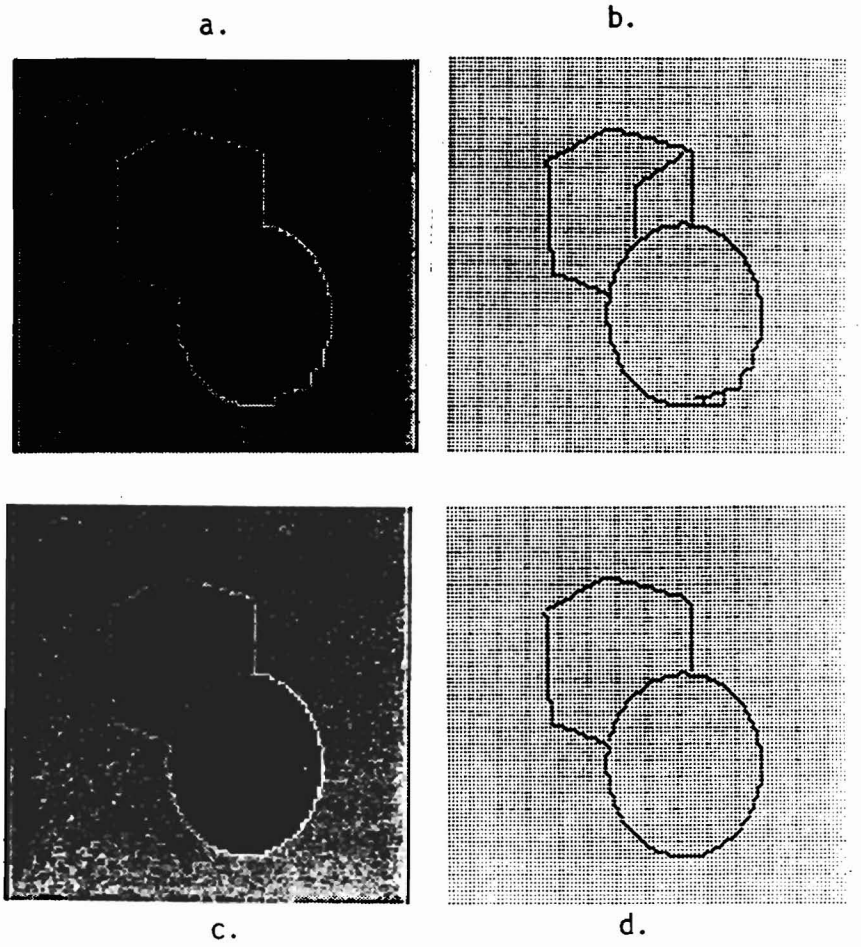

Figure 6.4. Results with Synthetic Data (I)

a) Reconstructed depth with α=10 and β=0.01. b) Depth edges with α and β as in a). c) Reconstructed depth with α=10 and β=0.001. d) Depth edges with α and β as in c).

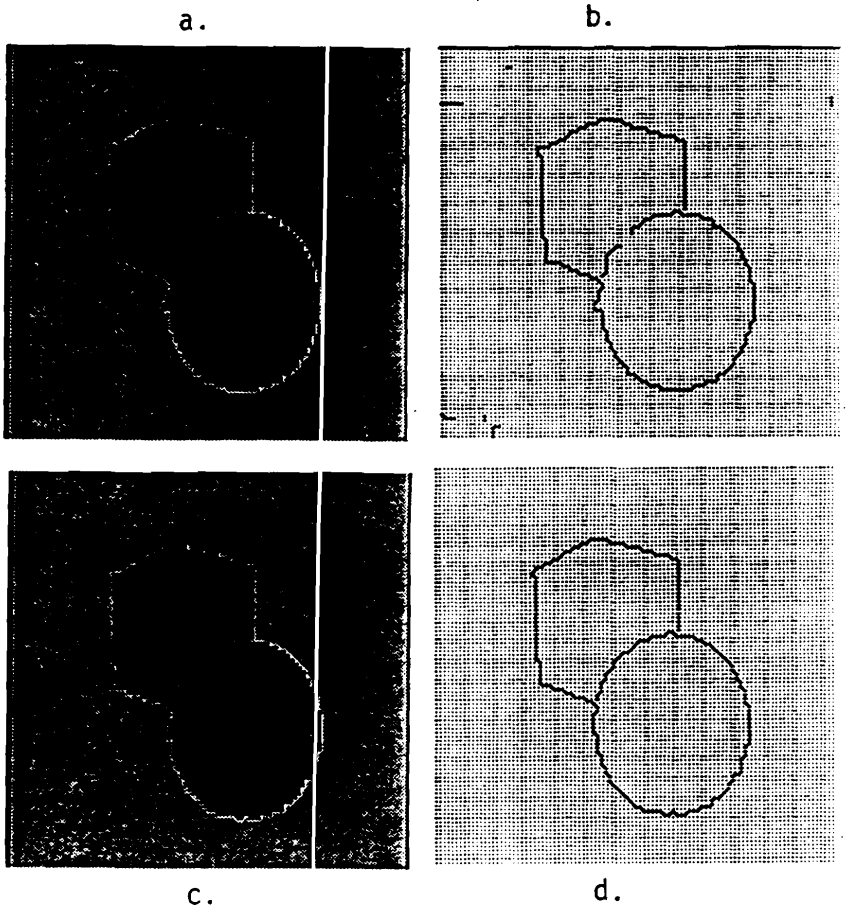

Figure 6.5. Results with Synthetic Data (II)

a) Reconstructed depth with α=50 and β=0.001, with 80% depth data randomly sampled and no intensity input. b) Depth edges with conditions of a). c) Reconstructed depth with α and β and 80% depth sampling as in a), but using intensity input in MRF. d) Depth edges with conditions of c).

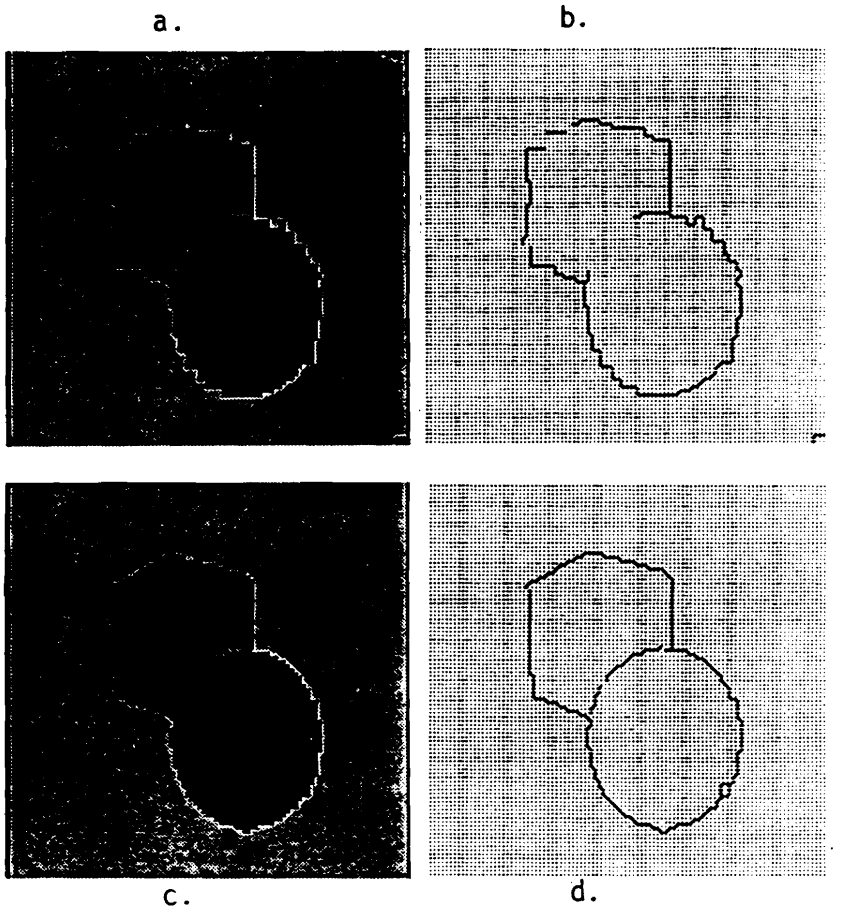

Figure 6.6. Results with Synthetic Data (III)

a), b), c), d) as in Fig. 6.5 except that depth sampling density is 50%.

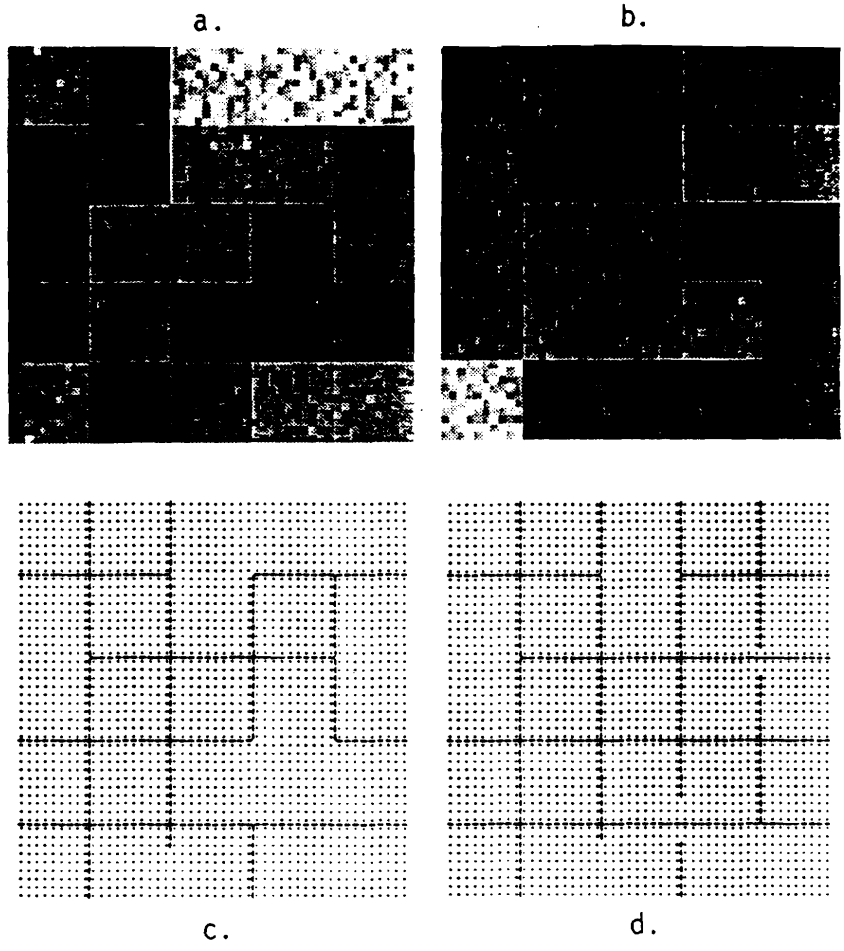

Figure 6.7. Experiments with Checker-Boards

a) Artificial "depth" image, with spatially independent G(0,16) additive noise, clipped to [0,255]. b) Artificial "intensity" image, with G(0,20) noise as in a).a c) Depth edges found with α=50, β=0.001, no intensity input used in MRF. d) Depth edges found under conditions of c) but using intensity input.

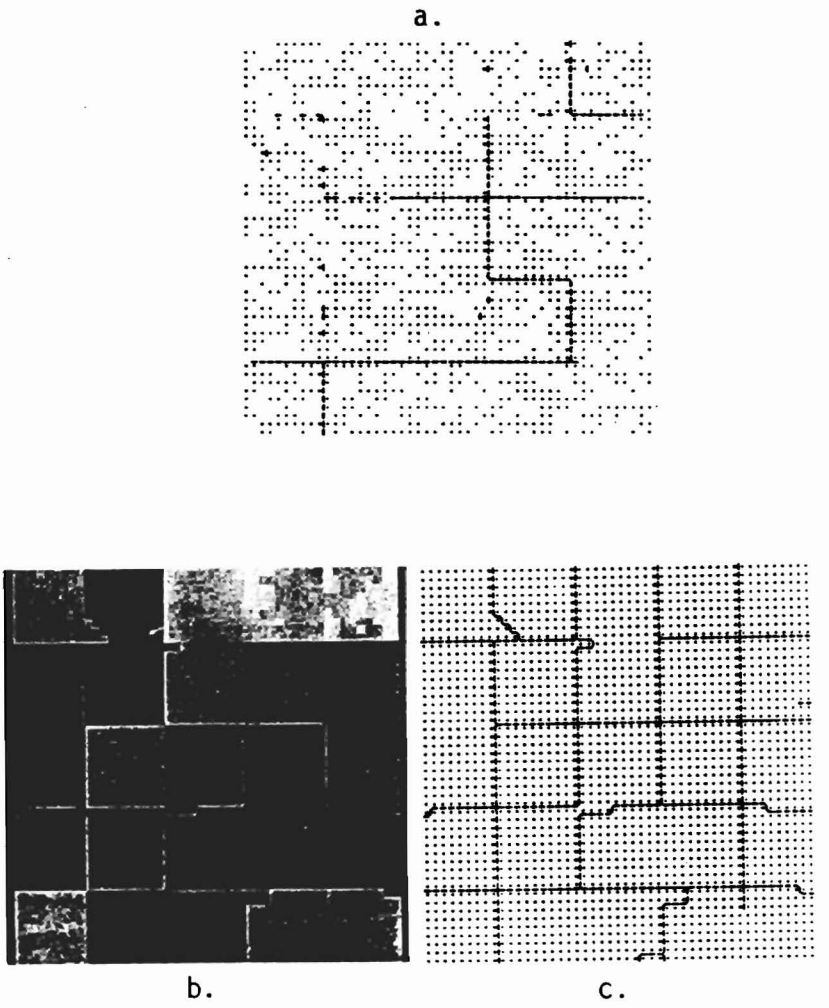

Figure 6.8. Experiments with Checker-Boards

a) Locations of 50% depth sampling overlaid with the initial thresholded likelihoods from intensity-discontinuity detection. α and β as in Fig. 6.7. b) Depth reconstruction. c) Depth discontinuities.

6.7.2. Natural Scenes

Depth segmentation and reconstruction using HCF was performed with stereo disparity data of scenes consisting of table-top objects. The Cooper stereo algorithm [Cooper 1987] yields sparse disparity data associated with intensity contours in the input images. It has been demonstrated to be robust and to work with natural scenes and with structured light. Briefly, it uses a global goodness measure to decide the stereo correspondences between zero-crossing contours of Difference of Gaussian (DOG) using a dynamic programming procedure.

The first scene (Figure 6.9.a) shows a beach ball and a rectangular box sitting in front of a cylindrical object, with a flat background. Vertical (with respect to the epipolar line) light strips were projected onto the scene to create artificial texture needed by the stereo system (Figure 6.9.b). The disparity observations are scaled and rounded to 256 levels, and are assumed to have independent unbiased Gaussian noise with standard deviation of 12. Three pairs of stereo images were taken under different setups of the light projector to increase the density of disparity observations (Figure 6.9.c). Observations at the same pixel location are combined using a conditional independence assumption:

$$P(d_1, d_2 \mid d) = P(d_1 \mid d)\, P(d_2 \mid d)$$

where d_1 and d_2 are two disparity observations obtained from two different setups of the structured light. Figure 6.9.d shows the resulted disparity observations in perspective (0 represents no observation). About 35% of the pixel locations have at least one observation. Figure 6.9.e shows a map of those locations (in black) overlaid with the TLR estimate of the intensity discontinuities (Chapter 5).

A perspective view of the reconstructed disparity map along with the discontinuities detected, with $\alpha = 30$, $\beta = 0.003$, and $P(NON\text{-}EDGE \mid NON\text{-}DEPTH\text{-}EDGE) = 0.95$, are shown in Figures 6.9.f and g. The surfaces corresponding to the sphere, cylinder, and the background planes are smoothly reconstructed, and significant disparity discontinuities are detected. There are two types of mistakes, however, due to the characteristics of the observed

information. First, steep disparity and intensity gradients together near the sphere boundary result in spurious segmentation. Using a smaller prior probability $P(NON\text{-}EDGE \mid NON\text{-}DEPTH\text{-}EDGE)$, as shown in Figures 6.9. h and i, avoids the problem by reducing the assumed coupling between intensity and disparity discontinuities. Intensity discontinuities are thus preferably explained as changes in color rather than depth. (This fact is due to Equation (6.3). If the intensity module supports the edge label, i.e. $\alpha_2 > \alpha_0$, the likelihood ratio of *DEPTH–EDGE* decreases as p becomes smaller.) Second, there are regions that have no (e.g. the table top) or few (e.g. the top of the box) disparity observations. The result is the *leaking* effect: The disparity values of the neighboring regions leak through the holes of weak intensity gradients, resulting in under-segmentation and erroneous disparity estimates. A possible fix is to identify those regions prior to the reconstruction, possibly by building convex hulls of the "adjacent" pixels with disparity measures, and limit the reconstruction process outside of those regions.

The same beach ball is used in the second table-top scene (Fig. 6.10.a), partially occluded by a foam box. The foam box has very rough outlines. Fig. 6.10.b shows the overlay of the disparity data and the TLR intensity edge estimates. Only 16% of the pixels have disparty observations. As expected, the results are less satisfactory due to the lack of information and the ill-defined boundaries (Fig. 6.10.d - e).

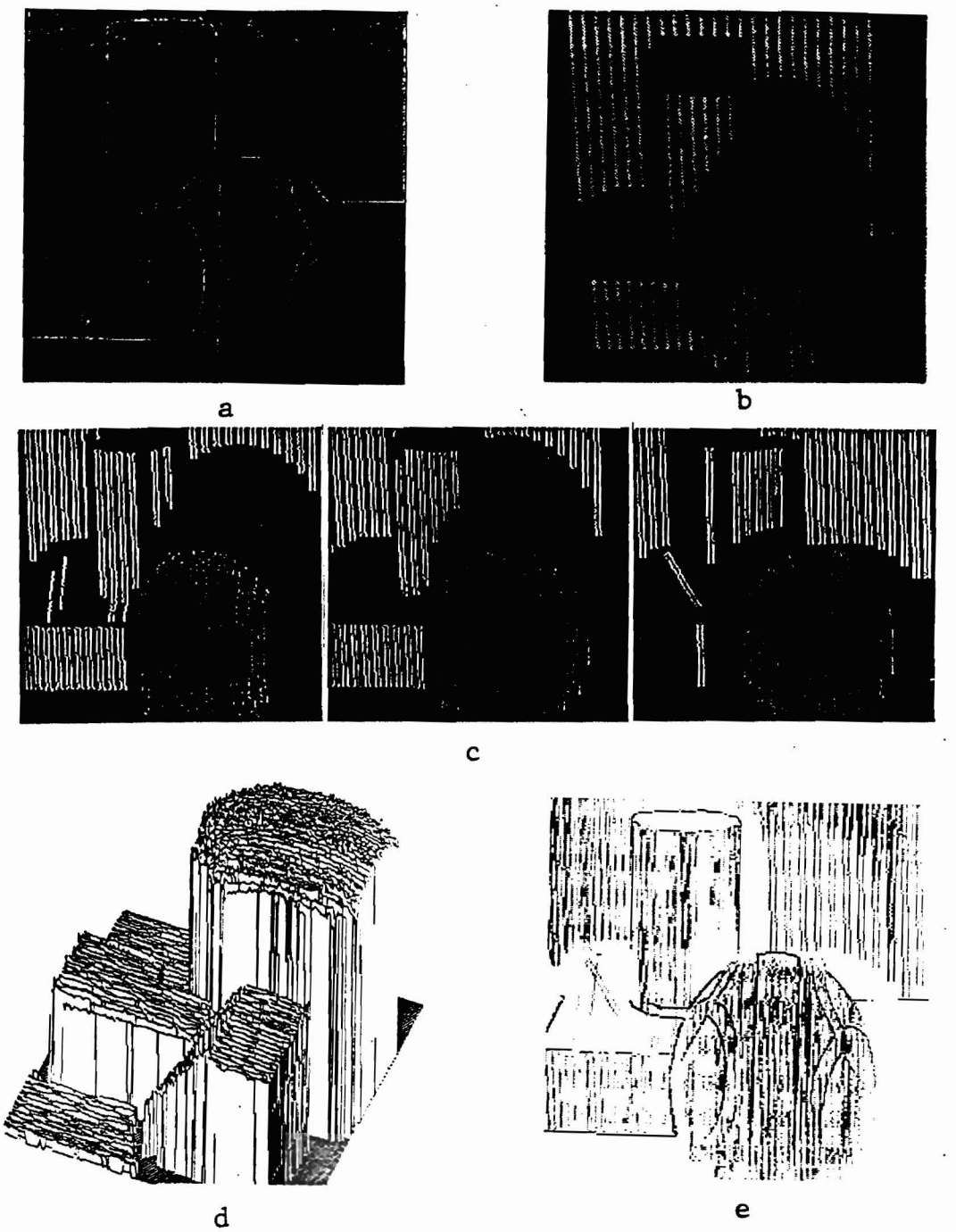

Figure 6.9. Experiments with Stereo Disparity Data (I)

a) 200 by 200 intensity image. b) Scene with projected structured light. c) Three disparity images. d) Perspective view of the combined disparity image. e) Locations of the disparity measurements overlaid with the TLR estimate of the intensity discontinuities.

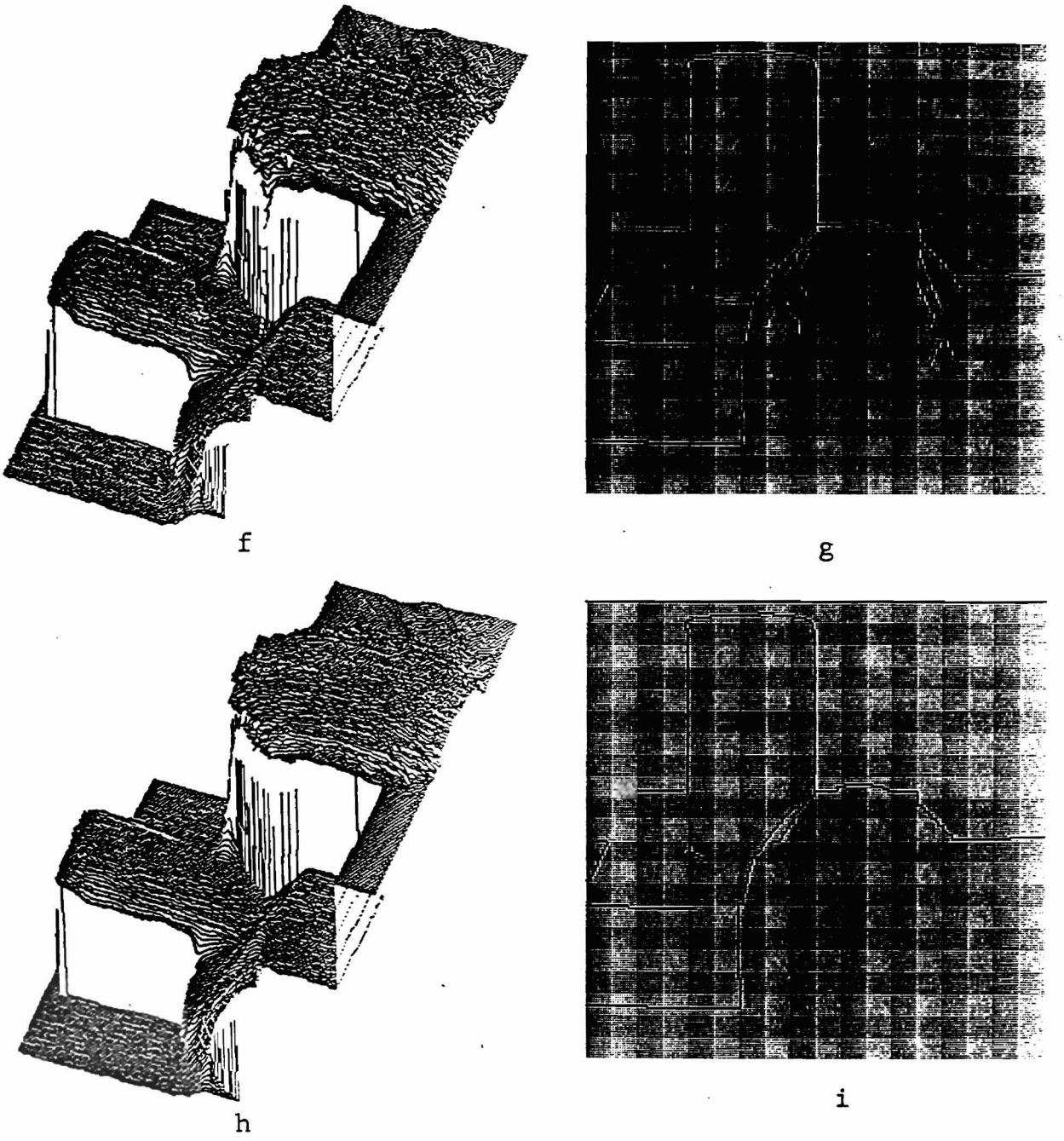

Figure 6.9. (continue)

f) Reconstructed disparity map with $p = 0.95$. g) Disparity discontinuities. h) Reconstructed disparity map with $p = 0.9$. i) Disparity discontinuities.

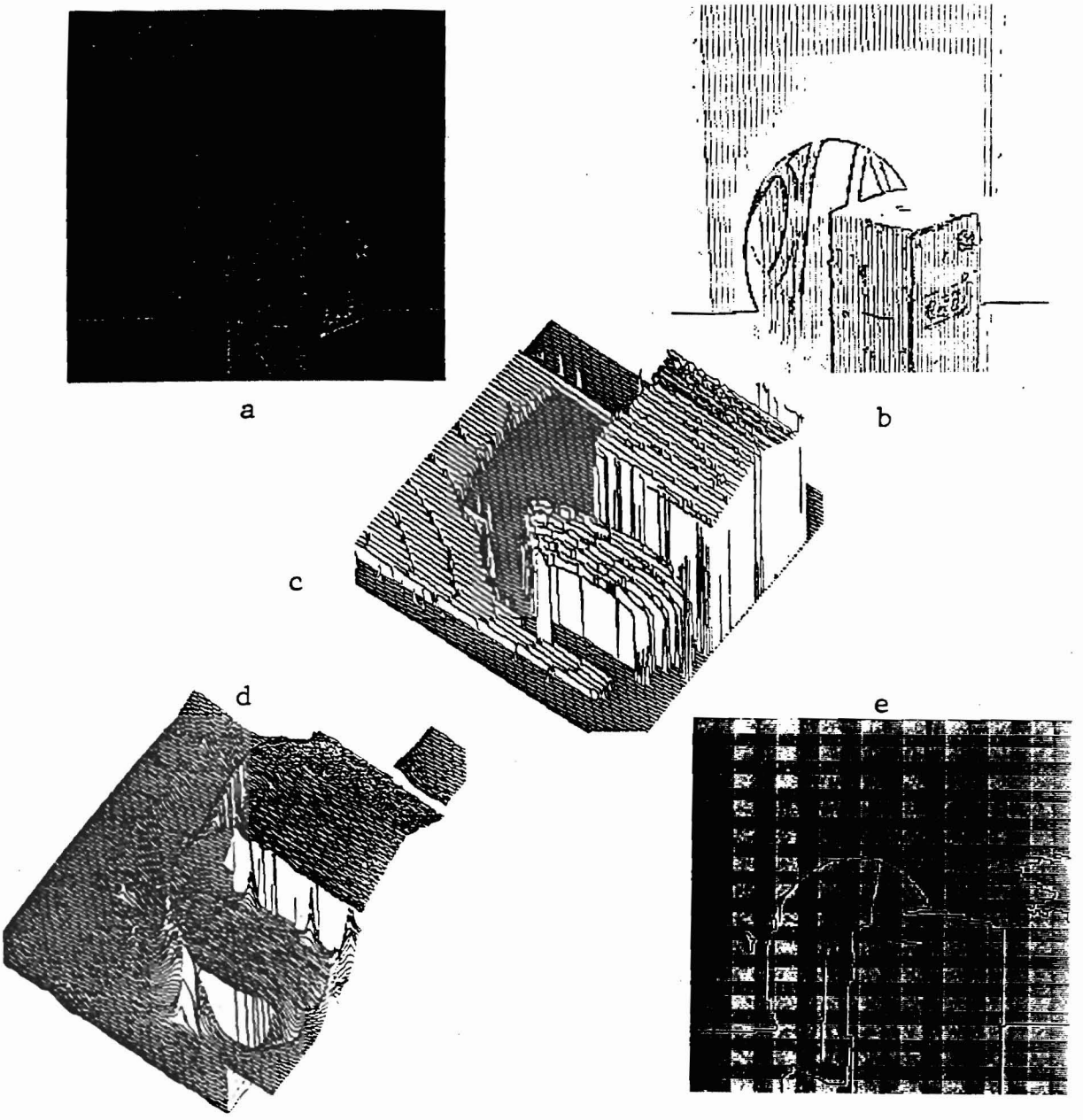

Figure 6.10. Experiments with Stereo Disparity Data (II)

a) 200 by 200 intensity image. b) Locations of the disparity measurements overlaid with the TLR estimate of the intensity discontinuities. c) Input disparity image. d) Reconstructed disparity map. e) Disparity discontinuities.

6.8. Discussion

The Role of Intensity Discontinuities: Successful integration of multi-modal data requires knowledge about the characteristics of scene and the vision modules processing the data. Such knowledge affects the decisions that have to be made when different modalities provide conflicting information about particular events. Fig. 6.3 shows an example: The strong intensity gradients across the cube edges suggest depth discontinuities at the face intersections but the relatively small depth differences at these locations refute such suggestions. Also, the self-shadowed face merges with the background in the intensity image while the depth information indicates clear separation of the two regions. If reliable depth observations are available, e.g., a noise free depth map, it is bad practice to use the intensity observations for clues to depth discontinuities. On the other hand, when the depth observations are sparse and unreliable, the correlation of intensity information with depth information should be recognized and used. The probabilistic integration provided by HCF optimization is one coherent framework for such integration tasks. The HCF scheme, as a deterministic method, finds a local probability maximum. In so doing, its behavior is consistent with the natural evidence weighing described above. In particular, if the depth observations are less reliable than the intensity observations (i.e. they have larger variation from their expected true values), the line sites tend to have larger (negative) stability measures than the depth sites at the early stage of the computation. This means the line sites commit (since they are based on intensity information) earlier than the depth sites, resulting in a final configuration that is more consistent with the intensity discontinuities. At the locations where depth observations are missing (e.g. Figs 6.5, 6.6, 6.8), the intensity discontinuity information helps to localize the depth discontinuities before the depth information can be spatially propagated. Similarly, when the depth observations are more reliable (denser, less noisy), the depth sites commit earlier. The early depth commitments influence the stability measures, and thus the later commitments, of the line sites. The resulting configuration tends to be more consistent with the depth data.

The Computational Advantages of HCF: The performance characteristics of HCF in minimizing the energy of coupled MRFs is consistent with its performance on simple MRFs, incorporating only the line process, reported earlier (Chapter 5). That is, the enhanced HCF algorithm behaves efficiently and predictably. The introduction of the continuous-valued depth process requires more visits to the depth sites than occurred in the optimization using only the binary line process. The line-process only MRF stabilized after fewer than 1.01 visits per site (on the average). Experiments with the checker board images (Figs 6.7), show that it takes on average fewer than 3 visits per site to achieve reasonable estimates in the coupled intensity-depth MRFs. The situation is complicated by the fact that the sizes of the regions affect the speed of convergence: Larger regions require on average more visits per site for results to propagate through them.

Since the energy functional is quadratic given a line configuration, in principle any deterministic minimization method would find the same minimal configuration of the depth process. However, there are some advantages to HCF over iterative relaxation schemes with predetermined visiting orders. HCF always visits the site that can reduce the energy measure the most. Thus early visits are far more important than the later ones with HCF. The rate of stabilizing is always maximized, and the most reliable decisions, which reduce energy most, are made first, and at some point the computation may be terminated with confidence of negligible future improvement. Fixed-order schemes cannot guarantee this property.

Modeling and Aligning Observations: Two technical difficulties require further investigation. First, modeling early depth measurements with (6.1) and (6.2) is not completely satisfactory for the following reasons:

(1) The range of depth values is usually a proper subset of R, depending on the scope of the depth sensor.

(2) The measurement error usually is biased and related to the surface depth and sensory characteristics.

(3) Gross error or "mistakes" [Krotkov 1987] due to, say, mismatching two features (zerocrossing contours) by a stereo system must be considered.

We believe that the decoupling of sensory models and *a priori* knowledge in our MRF formalism and the HCF estimation method provides us enough flexibility to incorporate such complex measurement models.

The second difficulty has to do with the localization (registration) of the measurements. For example, in the experiments with stereo disparity data, the depth measurements are located at zero-crossing contours of the DOG operators, thus at pixel locations. When a zero-crossing contour corresponds to surface locations near an occluding boundary, which projects to a contour of intensity discontinuities, the precise localization of the zero-crossings becomes vital to the success of the integration. We have observed that sometimes a depth measurement is reported at a pixel location just outside of the region (defined by the intensity discontinuities at the line sites) to which it should belong. Such error usually leads to disastrous reconstruction and segmentation. The problem becomes more complex when incorporating data from sensors of different spatial resolution and projection geometry, or using DOG-type filters of various scales. We believe that the success of visual integration at the locations of discontinuities relies on an adequate solution to this problem.

Chapter 7 Summary and Discussion

7. Summary and Discussion

7.1. Summary

This dissertation presents a framework, based on Bayesian-probability theory, for solving the labeling problem. The central issues addressed by the thesis are the representation of knowledge, reasoning procedures for combining distinct bodies of knowledge, and inference methods for using available knowledge to infer scene properties.

Chapter 3 presents a novel approach to knowledge representation and reasoning. The central idea is the decoupling of external evidence and *a priori* knowledge. A hierarchically structured label tree is used to accrue external evidence concerning the labels for each site. A probabilistically justifiable procedure consistently and coherently combines distinct bodies of evidence, represented as label likelihood ratios in the tree. The combination is commutative and associative, and has a simple message-passing implementation. More importantly, this procedure accumulates external evidence in terms of likelihood ratios rather then probability distributions. This feature enables the integration of the *a priori* knowledge, encoded in terms of a joint probability distribution of all sites, with the pooled external evidence in a Bayesian formalism.

The utility of Markov Random Field Models is also discussed in Chapter 3. The image sites are modeled as random variables with noncausal Markovian interactions. *A priori* knowledge can be conveniently encoded in terms of a set of local potential functions that, according to Hammersley-Clifford theorem, decides the *a priori* probability distribution of the random field. The *a posteriori* distribution derived by combining the external evidence and the *a priori* knowledge forms the basis for solving the labeling problem.

Chapter 4 presents a new estimation method -- the Highest Confidence First estimation algorithm. HCF is deterministic and intrinsically serial. Its results depend only on the *a posteriori* distribution. There is no need to choose a updating order or to set any parameter values (temperature, thresholds, etc.). By comparison, previous

approaches are inferior in both efficiency and robustness. Particularly, stochastic methods are computationally expensive, and their results tend to be affected by the undesirable large-scale characteristics associated with MRF models. Previous deterministic methods require good initializations, are sensitive to noise, and their results are partially decided by predetermined updating orders. We have argued that HCF meets the principles of graceful degradation and least commitment. As a consequence, its computational cost is small and the resulting estimates are less sensitive to noise. A priority-heap implementation of HCF is presented, and its convergence properties were discussed. Possible extensions are also sketched.

Chapter 5 uses the results of Chapters 3 and 4 to address the boundary detection problem. A novel aspect of this work is the use of an MRF to model an explicit line process, and the use of outputs from a set of local edge operators for the external evidence. The edge operators are based on Sher's work [1987]. They generate likelihood ratios about the presence of edges based on a particular step-edge model. Qualitative knowledge such as line coherence is encoded by using an MRF with a second order neighborhood system, and is combined with the edge likelihood ratios. The resulting *a posteriori* probability distribution is used by several estimation methods. The results using synthetic as well as natural input images are compared both qualitatively and in quantitative energy measurements. The comparisons show that HCF outperforms other methods in both robustness and efficiency, and gives better qualitative and quantitative results.

Chapter 6 presents a unified treatment of reconstruction and segmentation of three-dimensional surfaces with sparse depth observations and intensity data. In this work, intensity discontinuity information is incorporated in the process of detecting depth discontinuities. Coupled MRF's are used, one for depth and one for discontinuities, to model the two interacting processes. An extension of HCF successfully implements a solution method for the reconstruction and segmentation problem. Experiments are conducted with both artificially generated data and real disparity data from a feature-based stereo system of table-top scenes. The results are encouraging. They demonstrate the effectness of the approach, illustrate some of its

idiosyncratic characteristics, and suggest areas for improvement.

7.2. Discussion

An issue worth mentioning is the relationship between the HCF algorithm and the more traditional and possibly more familiar approaches to combinatorial optimization. It is also important to clarify the relationship between the process of creating a model and the process of finding a (maximum-likelihood) estimate of the relevant parameters. At the outset we note that the class of functions we are dealing with, though very powerful, is rather restricted. They consist of two terms, one of which is determined by the values of the individual observed variables, and the other is determined by the interaction of the variables.

First, the basic computation performed in the labeling context is one of statistical estimation. The method developed in this research is neither the traditional MAP estimation (which is very difficult to compute) nor the traditional MLE estimation, but rather a new type of estimation that treats individual variables differently in accordance with the relative significance of the variable observations (Chapter 4). The underlying intuition of HCF is simple: in deciding the identities of the variables, the use of contextual information becomes more important as the external observations become less informative. Traditional estimations treat all of the variables equally, thus their results are more likely to be affected by noise and the inaccuracy of prior models.

Second, the statistical estimation computation has the same form as that of traditional function optimization techniques (Equation (3.15)). In the function optimization case the function is given *a priori* and the parameter (in Equation (3.15), T) that controls the proportion of the error terms is chosen by a variety of techniques that are aimed at ensuring finding the correct optimum. In the statistical estimation case the traditional approach has been to use *ad hoc* weights that are dependent on the problem domain. These weights constitute a model of the domain, in the sense that they reflect empirically observed characteristics. Thus in statistical estimation, the behavior of the algorithm is influenced by the character of the domain.

The empirical choice of weights in this work corresponds to picking a model for the domain and instantiating it as *a priori* knowledge (see Sections 5.1 - 5.3 for an example). The attempt is to quantify a set of vague criteria about the goodness of label interpretations for the domain by the use of a (small) set of parameters. Similar approaches have been used in many other places. There have been some efforts on automatic parameter estimation, however their successes have been limited to the domains of texture modeling.

One fact stands out: Prior models, however constructed, usually only reflect intuitions and vague concepts. Estimation procedures must take possible sources of inaccuracy in the models into account. HCF estimation is designed with this principle in mind.

Why is the HCF computation effective in energy minimization? Take the boundary detection problem as an example (Chapter 5). Let $X_s = 1$ denote the presence of an edge at site s and $X_s = 0$ otherwise. The admissible solution space of the problem consists of the corners of the hypercube $[0,1]^N$, where N is number of sites. Many corners (solutions) are locally optimal in energy measure -- no adjacent corners (of distant 1 away) have lower energy values. The computation of existing iterative relaxation methods consists of a sequence of steps through adjacent corners. The resulting path depends on a (predetermined) updating order and the energy values associated with the corners. It is easy for such computations to get stuck at minor local optima. To ensure that the global optimum can be reached, immensely more computations are required, as in the case of using stochastic simulated annealing techniques.

Although global optimality is not guaranteed, HCF effectively avoids minor local optima by a "greedy search" in an augmented space. The augmented space consists of 3^N elements, including the 2^N hypercube corners in the admissible solution space. The elements in the augmented space form $n + 1$ layers. Layer i consists of the elements with exactly i dimensions of values 0 or 1 ("committed"). The elements are conceptually connected in the sense that the computation starts at the 0-th layer (all "uncommitted"), and terminates at some element in the n-th layer --

the admissible solution space. The connections between layers are uni-directional; the computation goes through the layers one by one, from lower-numbered ones to higher-numbered ones. There are also some intra-layer connections that allow the computation to fine tune its directions. The computation is reminiscent of traditional greedy search or steepest descent techniques in that every step is a move to the "best" adjacent element with respect to an augmented energy measure defined over the augmented space. The greedy search in the augmented space has the flavor of "flying over" minor local optima in the "best" directions observed in the air. It is also interesting to note that the use of the augmented space does not impose any significant amount of overhead. In fact, our preliminary implementation of HCF has consistently demonstrated fast and predicatable results.

7.3. Future Directions

We plan to continue to explore the properties of the HCF algorithm and MRF modeling. The following issues deserve further investigation, and can be individually studied.

(1) Systematically learning MRF parameters. The current *ad hoc* assignments of MRF's potential functions are adequate for demonstrative experiments. However, it would be desirable to devise a systematic method to estimate (learn) those parameters. For processes such as intensity and texture for which uncorrupted realizations are available (perhaps from stochastic sampling), the maximum-pseudolikelihood types of estimation (Chapter 3) or unsupervised learning schemes could be helpful. So far, it is not yet clear how to estimate MRF parameters from a set of corrupted images.

(2) Robust likelihood generators. The thesis assumes that likelihood ratios of interesting labels can be computed from image features based on either probabilistic models or statistical data. We have discussed how the edge likelihood operators used in our experiments could be improved (Chapter 5). It is interesting to investigate how to compute likelihoods for symbolic labels other than edges. We are currently studying the computation of likelihoods

for surface types including planar, cylindrical, and spherical from intensity or range observations.

(3) Flexible evidence accumulation and labeling. One question that needs to be addressed is how to deal with non-exclusive labels. For example, it might be desirable to label a site corresponding to an occluding boundary location to be both depth and orientation edges. The current treatment, requiring one label per site, is not completely satisfactory. One obvious solution that can use the current approach is to enhance the label set to include a depth-and-orientation-edge label and require the labels to be disjoint. However, this solution requires many more labels to be considered at the same time. We are investigating how to extend the thesis work to deal with such problems.

(4) Parallel HCF implementation. The HCF algorithm relies on a priority queue of all MRF elements in order of stability. A heaping algorithm is an efficient implementation in a serial model of computation, but we are exploring parallel alternatives both for reasons of speed and because the problem is intrinsically interesting.

(5) Distributed asynchronous data sources. The HCF algorithm so far has only been tested under the condition that it is presented with all the data simultaneously, and thus can correctly find the globally highest-confidence element. Under different circumstances, the data may arrive partially or asynchronously. The question then is how is HCF performance is affected by data arrival that may result in spatial or temporal discontinuities in the MRF element field. A testbed has been constructed on a BBN Butterfly parallel processor by Robert Potter and Drew Asson at the University of Rochester that will allow us to investigate this question.

(6) Spatially nonuniform MRFs and serial data accumulation. We are interested in fields with a peripheral and foveal organization, with a central high resolution area surrounded by a (perhaps progressively) low-resolution one. If we term such an organization a retina, over time, "eyemovements" can

reposition the retinal organization within a larger, uniformly high-resolution MRF representing a stable world, or its projection. The interaction of the elements over time is of interest. If an element is seen first in the periphery and is then foveated, it proceeds through a "coarse to fine" context that may allow it to reach a correct labeling more reliably. There are several technical questions about the implementation of this idea.

(7) Computational advantages. With HCF one can set a threshold on stability measures that will terminate the computation with high confidence that only insignificant changes of, say, the depth configuration would occur if the computation would continue. This property is bought at the cost of maintaining the priority order HCF needs, and raises the obvious question: "Is the gain worth the cost?". We are studying this tradeoff between the overhead paid in deciding the dynamic visiting order of HCF and its computational gain of fewer visits to the sites.

Lastly, we are interested in seeing the thesis work be applied to other domains. Large-scale optimization problems for which HCF could be applied are becoming common in many areas, from signal and speech processing through robotics (e.g. [Barhen, Toomarian, and Protopopescu 1987]). We believe that problems involving large numbers of variables, reasoning with multiple imperfect knowledge sources, statistical estimation, and energy minimization computations can all benefit from the approach presented in this thesis.

Bibliography

A. Bibliography

ABEND, K., T. J. HARLEY, and L. N. KANAL, "Classification of binary random patterns", *IEEE Transactions on Information Theory 11* (1965), 538-544.

ALOIMONOS, J., "Computing Intrinsic Images", TR 198, Computer Science Department, University of Rochester, Aug. 1986.

AMBLARD, F. G., D. B. COOPER, and B. CERNUSCHI-FRIAS, "Estimation by multiple views of outdoor terrain modeled by stochastic processes", *SPIE Intelligent Robots and Computer Vision 726* (1986), 36-45.

AYACHE, N. and O. D. FAUGERAS, "Building, Registering, and Fusion Noisy Visual Maps", 596, INRIA, France, Dec. 1986.

BALLARD, D. H. and C. M. BROWN, in *Computer Vision*, Prentice-Hall Inc., 1982.

BALLARD, D. H., "Parameter Nets", *Artificial Intelligence 22* (1984), 235-267.

BALLARD, D. H., "Eye Movements and Spatial Cognition", TR 218, Computer Science Department, University of Rochester, Oct. 1987.

BALLARD, D. H., "Interpolation Coding: A Representation for Numbers in Neural Models", *Biological Cybernetics 57* (1987), 389-402.

BARHEN, J., N. TOOMARIAN, and V. PROTOPOPESCU, "Optimization of the Computational Load of a Hypercube Supercomputer Onboard a Mobile Robot", *Applied Optics 26*, 23 (Dec. 1987), 5007-5014.

BARNARD, S., "Stereo Matching By Hierarchical, Microcanonical Annealing", *Proceedings: Image Understanding Workshop 2* (Feb. 1987), 792-797.

BARROW, H. G. and J. M. TENENBAUM, "Recovering Intrinsic Scene Characteristics from Images", in *Computer Vision Systems*, HANSON, A. R. and E. M. Riseman (editors), Academic Press, 1978.

BESAG, J., "Spatial interaction and the statistical analysis of lattice systems (with discussion)", *Journal of Royal Statistics Society, series B 36* (1974), 192-326.

BESAG, J., "Statistical Analysis of non-lattic data", *The Statistician 24* (1975), 179-195.

BESAG, J., "On the Statistical Analysis of Dirty Pictures", *Journal of Royal Statistical Society B 48*, 3 (1986), 259-302.

BLAKE, A. and A. ZISSERMAN, *Visual Reconstruction*, MIT Press, 1987.

BOLLE, R. M. and D. B. COOPER, "Bayesian Recognition of Local 3-D Shape by Approximating Image Intensity Functions with Quadric Polynomials", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 4 (July 1984), 418-429.

BOLLE, R. M. and D. B. COOPER, "Optimal Statistical Techniques for Combining Pieces of Information Applied to 3-D Complex Object Position Estimation", in *Pattern Recognition in Practice*, GELSEMA, E. S. and L. N. Kanal (editors), Elsevier Science Publishers B.V. (North-Holland), 1986, 243-253.

BOLLES, R. C., *Verification Vision within a Programmable Assembly System*, Stanford, Sep. 1976. Thesis Draft.

BOLLES, R. C., "Verification Vision for Programmable Assembly", *Proceedings: IJCAI-77*, Aug. 1977, 569-575.

BROOKS, R. A., A. M. FLYNN, and T. MARILL, "Self Calibration of Motion and Stereo Vision for Mobile Robot Navigation", *Proc. Darpa Image Understanding Workshop*, Apr. 1988, 398-410.

BROWN, C. M., "PADL-2: A Technical Summary", *IEEE Computer Graphics and Applications*, Mar. 1982, 69-84.

BROWN, C. M., "Computer Vision and Natural Constraints", *Science 224*, 4655 (June 1984).

CANNY, J. F., "Finding Edges and Lines in Images", AI-TR 720, MIT Artificial Intelligence Laboratory, 1983.

CAVANAGH, P., "Reconstructing the Third Dimension: Interactions between Color, Texture , Motion, Binocular Disparity, and Shape", *Computer Vision, Graphics, and Image Processing 37* (1987), 171-195.

CHOU, P. B. and R. RAMAN, "On Relaxation Methods Based on Markov Random Fields", TR 212, Computer Science Dept., The Univ. of Rochester, July 1987.

CHOU, P. B., C. M. BROWN, and R. RAMAN, "A Confidence-Based Approach to the Labeling Problem", *Proceedings: IEEE Workshop on Computer Vision*, Miami Beach, Florida, Nov. 1987, 51-56.

CHOU, P. B. and C. M. BROWN, "Multi-Modal Segmentation Using Markov Random Fields", *Proceedings: Image Understanding Workshop 2* (Feb. 1987), 663-670.

CHOU, P. B. and C. M. BROWN, "Probabilistic Information Fusion for Multi-Modal Image Segmentation", *Proceedings: IJCAI-87*, Milan, Italy, Aug. 1987.

CHOU, P. B. and C. M. BROWN, "Multimodal Reconstruction and Segmentation with Markov Random Fields and HCF Optimization", *Proceedings: DARPA Image Understanding Workshop 1* (Apr. 1988), 214-221.

CLOWES, M. B., "On Seeing Things", *Artificial Intelligence 2*, 1 (1971), 79-116.

COHEN, F. S. and D. B. COOPER, "Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markovian Random Fields", *IEEE Tansactions on Pattern Analysis and Machine Intelligence PAMI-9*, 2 (Mar. 1987), 195-219.

COOPER, P. R. and S. C. HOLLBACH, "Parallel Recognition of Objects Comprised of Pured Structure", *Proceedings: Darpa Image Understanding Workshop*, 1987, 392-398.

COOPER, P. R., "Order and Structure in Correspondence by Dynamic Programming", TR 216, University of Rochester, June 1987.

COOPER, P., "Structure Recognition by Connectionist Relaxation: Formal Analysis", *Proceedings of Canadian Artificial Intelligence Conference CSCSI'88*, Edmonton, Alberta, 1988.

CROSS, G. R. and A. K. JAIN, "Markov Random Field Texture Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-5*, 1 (Jan. 1983), 25-39.

DAVIS, L. S. and A. ROSENFELD, "Cooperating Processes for Low-Level Vision: A Survey", *Artificial Intelligence 17* (1981), 245-263.

DERIN, H., H. ELLIOTT, R. CRISTI, and D. GEMAN, "Bayes Smoothing Algorithms

for Segmentation of Binary Images Modeled by Markov Random Fields'', *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 6 (Nov. 1984), 707-720.

DERIN, H. and W. S. COLE, ''Segmentation of Textured Images Using Gibbs Random Fields'', *Computer Vision, Graphics, and Image Processing 35* (1986), 72-98.

DRUMHELLER, M. and T. POGGIO, ''On Parallel Stereo'', *Proceedings: IEEE Int. Conf. on Robotics and Automation 3* (1986).

DUDA, R. O., P. E. HART, and N. J. NILSSON, ''Subjective Bayesian Methods for Rule-Based Inference Systems'', SRI Technical Note 124, SRI International, 1976.

ELLIOTT, H. and H. DERIN, ''Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields'', #ECE-UMASS-SE84-1, University of Massachusetts, 1984.

FELDMAN, J. A. and Y. YAKIMOVSKY, ''Decision Theory and Artificial Intelligence: I. A Semantics-Based Region Analyzer'', *Artificial Intelligence 5* (1974), 349-371.

FELDMAN, J. A. and D. H. BALLARD, ''Computing with connections'', TR 72, Computer Science Department, Univ. of Rochester, 1981.

FELDMAN, J. A., ''Dynamic Connections in Neural Networks'', *Biological Cybernetics 46* (1982), 27-39.

FELDMAN, J. A. and D. H. BALLARD, ''Connectionist Models and Their Properties'', *Cognitive Science 6* (1982), 205-254.

FELDMAN, J. A., "Energy and the Behavior of Connectionist Models", TR 155, Computer Science Department, University of Rochester, Nov. 1985.

FELDMAN, J. A., M. A. FANTY, N. GODDARD, and K. LYNNE, "Computing with Structured Connectionist Networks", *Communications of ACM 31* (Feb. 1988), 170-187.

FREUDER, E. C., "Synthesizing Constraint Expressions", *Communications of ACM 21*, 11 (Nov. 1978), 958-965.

GAMBLE, E. and T. POGGIO, "Visual Integration and Detection of Discontinuities: The Key Role of Intensity Edges", MIT A.I. Memo No. 970, October 1987.

GEMAN, S. and D. GEMAN, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 6 (Nov. 1984), 721-741.

GEMAN, S. and C. GRAFFIGNE, "Markov Random Field Image Models and Their Applications to Computer Vision", *Proceedings of the International Congress of Mathematicians*, Berkeley, CA, 1986, 1496-1517.

GIDAS, B., "Non-stationary Markov Chains and Convergence of the Annealing Algorithm", *Journal of Statisticl Physics 39* (1985), 73-131.

GOOD, I. J., in *Probability and the Weighing of Evidence*, Hafner Publishing Company, 1950.

GORDON, J. and E. H. SHORTLIFFE, "A Method of Managing Evidential Reasoning in a Hierarchical Hypothesis Space", *Artificial Intelligence 26* (1985), 323-357.

GRIMSON, W. E. L., *From Image to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, MA, 1981.

GRIMSON, "Binocular Shading and Visual Surface Reconstruction", *CVGIP*, 1984, 19-44.

HADAMARD, J., in *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, 1923.

HARALICK, R. M., "An Interpretation for Probabilistic Relaxation", *Computer Vision, Graphics, and Image Processing 22* (1983), 388-395.

HARALICK, R. M., "Digital Step Edges from Zero Crossing of Second Directional Derivatives", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6*, 1 (Jan. 1984), 58-68.

HASSNER, M. and J. SLANSKY, "The Use of Markov Random Fields as Models of Texture", in *Image Modeling*, ROSENFELD, A. (editor), Academic Press, Inc., 1980, 185-198.

HILDRETH, E. C., *The Measurement of Visual Motion*, MIT Press, Cambridge, MA, 1984.

HINTON, G. E., "Relaxation and Its Role in Vision", PhD Thesis, University of Edinburgh, 1977.

HINTON, G. E. and T. J. SEJNOWSKI, "Optimal Perceptual Inference", *Proceedings: IEEE Conf. CVPR*, 1983, 448-453.

HOPFIELD, J. J. and D. W. TANK, " "Neural" Computation of Decisions in

Optimization Problems'', *Biological Cybernetics 52* (1985), 141-152.

HORN, B. K. P., ''Obtaining shape from shading information'', in *The Psychology of Computer Vision*, WINSTON, P. H. (editor), McGraw-Hill, New York, 1975, 115-155.

HORN, B. K. P. and B. G. SCHUNK, ''Determining Optical Flow'', in *Computer Vision*, BRADY, J. M. (editor), 1981.

HUECKEL, M. H., ''An Operator Which Locates Edges in Digitized Pictures'', *Journal of the ACM 18* (1971), 113-125.

HUFFMAN, D. A., ''Impossible Objects as Nonsense Sentences'', *Machine Intelligence 6* (1971), Edinburgh University Press.

HUMMEL, R. and S. ZUCKER, ''On the Foundation of Ralaxation Labeling Process'', *IEEE PAMI 5* (May 1983), 267-286.

HUTCHINSON, J., C. KOCH, J. LUO, and C. MEAD, ''Computing Motion Using Analog and Binary Resistive Networks'', *Computer 21*, 3 (Mar. 1988), 52-63.

IKEUCHI, K. and B. K. P. HORN, ''Numerical Shape from Shading and Occulding Boundaries'', in *Computer Vision*, BRADY, J. M. (editor), 1981.

JULESZ, B., ''Textons, the elements of texture perception, and their interactions'', *Natural 290 12* (March, 1981), 91 - 97.

KANADE, T., ''A Theory of the Origami World'', *Artificial Intelligence 13* (1980), 279-311.

KANAL, L. N., ''Markov Mesh Models'', in *Image Modeling*, ROSENFELD, A.

(editor), Academic Press, Inc., 1980, 239-243.

KEMENY, J. G. and J. L. SNELL, *Finite Markov Chains*, Van Nostrand, New York, 1960.

KINDERMANN, R. and J. L. SNELL, "Markov Random Fields and their Applications", in *Contemporary Mathematics*, vol. 1 , American Mathematical Society, 1980.

KIRKPATRICK, S., C. D. GELATT, and M. P. VECCHI, "Optimization by Simulated Annealing", *Science 220* (1983), 671-680.

KIROUSIS, L. and C. PAPADIMITRIOU, "The Complexity of Recognizing Polyhedral Scenes", *Proc. 26th Annual Symposium on Foundations of Computer Science*, Oct. 1985.

KIRSCH, R. A., "Computer Determination of the Constituent Structure of Biological Images", *Computers and Biomedical Research 4*, 3 (1971), 315-328.

KITTLER, J. and J. FOGLEIN, "On Compatibility and Support Functions in Probabilistic Relaxation", *Computer Vision, Graphics, and Image Processing 34* (1986), 257-267.

KOCH, C., J. MARROQUIN, and A. YUILLE, Networks in Early Vision"" "Analog "Neuronal" Networks in Early Vision", *Proc. National Academy of Sciences USA 83* (1986), 4263-4267.

KOHLER, R. R., "Integrating Non-Semantic Knowledge into Image Segmentation Process", COINS Technical Report 84-04, 1984.

KROTKOV, E. P., "Exploratory Visual Sensing for Determining Spatial Layout with an Agile Stereo Camera System", Ph.D. Dissertation MS-CIS-87-29, University of Pennsylvania, May 1987.

LIVINGSTONE, M. S., "Art, Illusion and the Visual System", *Scientific American 258*, 1 (Jan. 1988), 78-85.

MACKWORTH, A. K., "Consistency in Networks of Relations", *Artificial Intelligence 8* (1977), 99-118.

MACKWORTH, A. K. and E. C. FREUDER, "The Complexity of Some Polynomial Network Consistency Algorithms for Constraint Satisfaction Problems", *Artificial Intelligence 25* (1985), 65-74.

MARR, D. and T. POGGIO, "Cooperative Computation of Stereo Disparity", *Science 194* (1976), 283-287.

MARR, D., *VISION*, W. H. Freeman and Company, 1982.

MARROQUIN, J., S. MITTER, and T. POGGIO, "Probabilistic Solution of Ill-Posed Problems in Computational Vision", *Proceedings: Image Understanding Workshop*, Dec. 1985, 293-309.

MARROQUIN, J. L., "Probabilistic Solution of Inverse Problems", AI-TR 860, MIT Artificial Intelligence Laboratory, Sep. 1985.

MATTHIES, L., R. SZELISKI, and T. KANADE, "Kalman Filter-based Algorithms for Estimating Depth from Image Sequences", CMU, Pittsburgh, PA-CS-87-185, Dec. 1987.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, and E. TELLER, "Equations of State Calculations by Fast Computing Machines", *Journal of Chemical Physics 21* (1953), 1087-1091.

MURRAY, D. W. and B. F. BUXTON, "Scene Segmentation from Visual Motion Using Global Optimization", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9*, 2 (Mar. 1987), 220-228.

NALWA, V. S., "On Detecting Edges", *Proceedings: Image Understanding Workshop*, 1984, 157-164.

NEVATIA, R. and K. R. BABU, "Linear Feature Extraction and Description", *Computer Graphics and Image Processing 13* (1980), 257-269.

PEARL, J., "On Evidential Reasoning in a Hierarchy of Hypotheses", *Artificial Intelligence 28*, NO. 1 (Feb. 1986), 9-15.

PELEG, S., "A new probabilistic relaxation scheme", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2* (1980), 362.

POGGIO, T., V. TORRE, and C. KOCH, "Computational Vision and Regularization Theory", *Nature 317* (1985), 314-319.

POGGIO, T., "Integrating vision modules with coupled MRFs", Working Paper No. 285, MIT A.I. Lab., Dec. 1985.

PREWITT, J. M. S., "Object Enhancement and Extraction", in *Picture Processing and Psychopictorics*, LIPKIN, B. S. and A. Rosenfled (editors), Academic Press, 1970.

REYNOLDS, G., D. STRAHMAN, N. LEHRER, and L. KITCHEN, "Plausible Reasoning

and the Theory of Evidence'', COINS Technical Report 86-11, April 1986.

ROBERTS, L. G., ''Machine Perception of Three-Dimensional Solids'', in *Optical and Electrical-Optical Information Processing*, TIPPETT, J. T. (editor), MIT Press, 1965, 159-197.

ROSENFELD, A., R. HUMMEL, and S. ZUCKER, ''Scene labeling by relaxation operations'', *IEEE Trans. Syst., Man, Cybern. SMC-6* (1976), 420.

SHAFER, G., *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

SHER, D. B., ''A Probabilistic Approach to Low-Level Vision'', TR 232, University of Rochester, Oct. 1987.

SHER, D. B., ''Advanced Likelihood Generators for Boundary Detection'', TR197, Univ. of Rochester, Computer Science Dpt., Jan. 1987.

SHVAYTSER, H. and S. PELEG, ''A New Approach to the Consistent Labeling Problem'', *CVPR*, June 1985, 320-327.

SIMCHONY, T. and R. CHELLAPPA, ''Stochastic and Deterministic Algorithms for Texture Segmentation'', *To appear*, 1988.

STUTH, B. H., D. H. BALLARD, and C. M. BROWN, ''Boundary conditions in multiple intrinsic images'', *Proceedings: 8th International Joint Conference on Artificial Intelligence*, 1983, 1068-1072.

SWAIN, M. J. and P. R. COOPER, ''Parallel Hardware for Constraint Satisfaction'', *Proceedings of AAAI-88*, 1988.

SZELISKI, R., "Regularization uses Fractal Priors", *Proceedings: AAAI-87*, July 1987.

TERZOPOULOS, D., "Multilevel computational processes for visible surface reconstrucion", *Computer Vision, Graphics, and Image Processing 24* (1983), 52-96.

TERZOPOULOS, D., "Integrating Visual Information from Multiple Sources", in *From Pixels to Predicates*, PENTLAND, A. P. (editor), Ablex Publishing Corp., 1986, 111-142.

TERZOPOULOS, D., "Regularization of Inverse Visual Problems Involving Discontinuities", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8*, NO. 4 (July 1986), 413-424.

TERZOPOULOS, D., "Image analysis using multigrid relaxation methods", *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8* (Mar. 1986), 129-139.

WALTZ, D., "Understanding Line Drawings of Scenes with Shadows", in *The Psychology of Computer Vision*, McGraw-Hill, 1975, 19-91.

WAXMAN, A. M. and J. H. DUNCAN, "Binocular Image Flows: Steps Towards Stereo - Motion Fusion", CS-TR-1494, University of Maryland, May 1985.

WILSON, G. V. and G. S. PAWLEY, "On the Stability of the Travelling Saleman Problem Algorithm of Hopfield and Tank", *Biological Cybernetics 58* (1988), 63-71.