

Depth recovery using Markov Random Fields

S KAPOOR, P Y MUNDKUR and U B DESAI

Department of Electrical Engineering, Indian Institute of Technology,
Powai, Bombay 400 076, India

MS received 5 December 1991; revised 7 October 1992

Abstract. This paper deals with the problem of depth recovery or surface reconstruction from sparse and noisy range data. The image is modelled as a Markov Random Field and a new potential function is developed to effectively detect discontinuities in highly sparse and noisy images. No use of any line process is made. Interpolation over missing data sites is first done using local characteristics and simulated annealing is then used to compute the maximum a posteriori (MAP) estimate. Results of software simulations carried out on actual range images along with details of the simulations are presented.

Keywords. Depth recovery; Markov Random Field; clique model; simulated annealing; surface reconstruction; early vision.

1. Introduction

Depth recovery is one of a set of early vision modules that deal with the recovery of properties of three-dimensional surfaces from their two-dimensional images. This has been a vibrant area in the recent past and there have been many papers in the literature dealing with surface reconstruction. We cite a few that have provided the motivation for this work: Boulton & Kender (1986), Marroquin *et al* (1987), Shao *et al* (1987, pp. 530–5), Choi & Kender (1988, pp. 189–94), Frankot & Chellappa (1988), Terzopoulos (1988), Horn & Brooks (1989, and the papers therein), Prasad *et al* (1990) and Geiger & Giroso (1991). For a recent review on three-dimensional surface reconstruction, see Bolle & Vemuri (1991). Most of the work focuses on recovering three-dimensional surfaces from one- or two-dimensional intensity images, while some (Boulton & Kender 1986, pp. 68–76; Choi and Kender 1988, pp. 189–94) deal with surface reconstruction from sparse and noisy range data. In this paper, we consider the latter problem.

It is well known that problems in early vision are usually ill-posed, i.e. a solution may not exist, may not be unique, or may not depend continuously on the data. Ill-posed problems are usually handled by using methods from regularization theory (Bertero *et al* 1988), which however may fail to give *good* results in the presence of discontinuities. To overcome this, one often uses probabilistic approaches which usually involve representing *a priori* knowledge in terms of an appropriate probability

distribution of an image model. If the statistics of the noise are known, then given the observation, the posterior distribution of the solution can be found using Bayes theory.

An extensively used stochastic model is the Markov Random Field (MRF) model (Besag 1974; Geman & Geman 1984), whose equivalence to the Gibbs Random Field can be used to exploit the Gibbs energy to specify the image model and the relation between the parameters of the model. Discontinuities in the image are modelled by selecting an appropriate clique model for the Gibbs energy function. Another advantage of such models is the existence of globally convergent algorithms like simulated annealing (Kirkpatrick *et al* 1983; Aarts & Korst 1989) to compute the maximum *a posteriori* (MAP) estimate of an image. An overview of probabilistic methods for ill-posed computational vision problems is given in Marroquin *et al* (1987). Faster algorithms, though not proven to be globally convergent, have been developed using mean field annealing (for example, Geiger & Girosi 1991 and Zerubia & Chellappa 1990).

1.1 Depth recovery

We consider the problem of depth recovery in the MRF framework. The basic objective is to determine a depth map of an image, given corrupted range data. Range data is taken to be sparse which is the typical case. This means that depth observations at some points of the image are not known or are undefined due to some reason. Thus, one must be able to recover the depth at these points given only sparse noisy observations. The main motivation for using the MRF framework is to make this recovery process capable of retaining edges or discontinuities, since these often contain most of the information present in the image.

The problem of depth recovery or surface reconstruction is encountered in many areas of computer vision, e.g. shape from stereo (where edges represent object boundaries and sparseness may be due to the limited number of points at which the matching or correspondence problem has been solved to give the depth from disparity), and applications like geophysical information processing (where edges represent geological faults and sparseness may occur due to scanning of the surface at a limited number of points). Sparseness may also be due to the loss of data caused by transmission over communication channels prone to fading or blackouts.

2. The MRF framework

We model the depth map as a Markov Random Field \mathcal{D} over a finite lattice of N depth sites. We denote a particular configuration or realization of \mathcal{D} by \mathbf{d} , the random variable at a depth site i by \mathcal{D}_i , and the value of \mathcal{D}_i by \mathbf{d}_i , the i th component of \mathbf{d} . The value \mathbf{d}_i represents the depth value at a point i in the depth map. The depth values belong to a finite set Ω . The dependence of a value at a depth site on the values at other depth sites is given by a neighbourhood structure, which specifies for each site the set of sites that are its neighbours. The neighbourhood structure is assumed to be translation invariant, except at the boundary, where the free boundary assumption is used, i.e. the set of neighbours of a boundary site is the intersection of the translation invariant neighbourhood with the finite lattice.

The probability distribution of the configurations of a Markov Random Field can

be shown to be equivalent to the Gibbs distribution via the Clifford–Hammersley theorem (Besag 1974; Geman & Geman 1984). Using this equivalence, we can write the probability distribution of the field configurations as

$$P\{\mathcal{D} = \mathbf{d}\} = \frac{\exp(-U(\mathbf{d})/T)}{Z}, \quad (1)$$

where U , the Gibbs free energy, is written as

$$U(\mathbf{d}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{d}), \quad (2)$$

where \mathcal{C} represents the set of all cliques defined by the neighbourhood structure, V_c is the potential at a clique $c \in \mathcal{C}$ whose value depends only on the components of \mathbf{d} that lie in c , and Z is the normalization constant given by

$$Z = \sum_{\mathbf{d} \in \Omega^N} \exp(-U(\mathbf{d})/T). \quad (3)$$

2.1 Constructing the potentials V_c

In the literature on image and texture modelling using Markov Random Fields, numerous clique models and associated potential functions have been considered (Geman & Geman 1984; Geman & Graffigne 1987; Lakshmanan & Derin 1989; Geiger & Girosi 1991). Most models use an explicit line process to detect edges or discontinuities. For our problem, we have chosen the potentials and the neighbourhood structure with the express purpose of retaining discontinuity information *without* the use of a line process. The basic motivation behind our choosing the potentials in the manner described below is the observation that, in the continuous case, the double partial derivative of an image $\partial^2 I(x, y)/\partial x \partial y$ vanishes at flat edges oriented in the coordinate directions.

We illustrate the procedure for a vertical edge in a discrete image, where partials are approximated by differences. Consider the lattice of depth values in figure 1. Assume there is a vertical edge E_v between the two columns of sites $\{5, 10, 14\}$ and $\{6, 0, 15\}$. Due to the edge, the depth values at sites $\{5, 10, 14\}$ will be higher than those at the corresponding sites $\{6, 0, 15\}$ by an amount that does not change very much with the position of the pair of corresponding sites in the columns. So if we

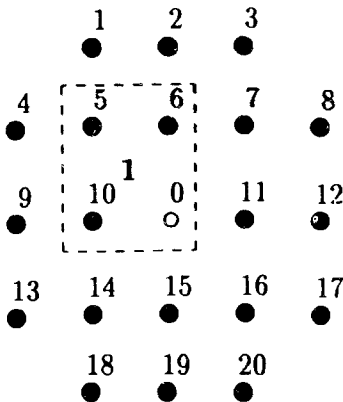


Figure 1. The neighbourhood of a depth site.

consider the differences of depth values in the two columns of sites, they will remain approximately the same down the column. This is the case even if the depth values at these sites are part of a smooth surface without any edges.

To express the above in equations, we first define:

$$a_{j,i} = |\mathbf{d}_j - \mathbf{d}_i|.$$

Then, the values $\{a_{6,5}, a_{0,10}, a_{15,14}\}$ are approximately the same. Note that these terms are the absolute values of the partials of the depth map with respect to x at particular sites, i.e., $a_{6,5}$ is the discrete counterpart of the absolute value of $\partial I(x, y)/\partial x$ evaluated at the site 5.

One way to examine the fluctuation in these values is to compare the difference between $(a_{6,5} - a_{0,10})$ and $(a_{0,10} - a_{15,14})$. That is, consider the quantity

$$V(E_v) = ((a_{6,5} - a_{0,10}) - (a_{0,10} - a_{15,14}))^2. \quad (4)$$

The term $(a_{6,5} - a_{0,10})$ is analogous to $\partial^2 I(x, y)/\partial x \partial y$ evaluated for a discrete I at site 10, but with absolute values of differences instead of just the differences themselves. This quantity will be small, and roughly equal to $\partial^2 I(x, y)/\partial x \partial y$ at site 14, i.e. $(a_{0,10} - a_{15,14})$. So the potential $V(E_v)$ will be small for the vertical edge E_v .

Expanding (4), we get

$$V(E_v) = (a_{6,5} - a_{0,10})^2 + (a_{0,10} - a_{15,14})^2 - 2(a_{6,5} - a_{0,10})(a_{0,10} - a_{15,14}) \quad (5)$$

We denote each of these terms by $V_v(a_{6,5} - a_{0,10})$, $V_v(a_{0,10}, a_{15,14})$ and $2V_{vv}(a_{6,5}, a_{0,10}, a_{15,14})$. Each constitutes a potential defined over some clique. We now examine these cliques. $V_v(a_{6,5}, a_{0,10})$ depends on or is defined on the sites $\{0, 5, 6, 10\}$, i.e. a 2×2 clique. Similarly, $V_v(a_{0,10}, a_{15,14})$ is defined on the 2×2 clique $\{0, 10, 14, 15\}$. $V_{vv}(a_{6,5}, a_{0,10}, a_{15,14})$ is formed by the product of two terms which are defined over overlapping 2×2 cliques $\{0, 5, 6, 10\}$ and $\{0, 10, 14, 15\}$, and can be seen to be defined on a 2×3 clique $\{0, 5, 6, 10, 14, 15\}$ that is a union of these two cliques (figure 2). Cliques that are formed by the union of two overlapping 2×2 cliques i and j will be denoted by $i \diamond j$.

Similarly, for the horizontal edge E_h between the two rows of sites $\{5, 6, 7\}$ and $\{10, 0, 11\}$, we will have the quantity

$$V(E_h) = (a_{7,11} - a_{6,0})^2 + (a_{6,0} - a_{5,10})^2 - 2(a_{7,11} - a_{6,0})(a_{6,0} - a_{5,10}) \quad (6)$$

$$= H_h(a_{7,11}, a_{6,0}) + H_h(a_{6,0}, a_{5,10}) + 2H_{hh}(a_{7,11}, a_{6,0}, a_{5,10}), \quad (7)$$

where $H_h(a_{7,11}, a_{6,0})$ and $H_h(a_{6,0}, a_{5,10})$ are potentials defined over 2×2 cliques, and $H_{hh}(a_{7,11}, a_{6,0}, a_{5,10})$ is defined over a 3×2 clique.

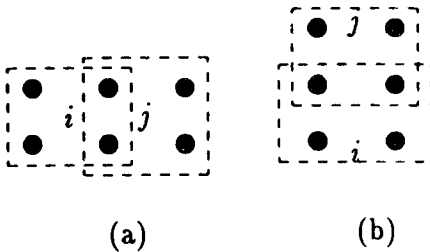


Figure 2. Unions of 2×2 cliques to give (a) a 2×3 clique and (b) a 3×2 clique.

Adding $V(E_v)$ and $V(E_h)$ with suitable weights β_v and β_h , we get the potential $V(E_v, E_h)$ for both the edges as

$$V(E_v, E_h) = \beta_v V(E_v) + \beta_h V(E_h). \quad (8)$$

This gives an expression for the potentials for the two edges E_v and E_h . After grouping terms in (7) according to the various clique types in E_v and E_h that they are defined on, we have

$$\begin{aligned} V(E_v, E_h) = & \sum_{i \in \mathcal{C}_{(2,2)}(E_v, E_h)} \gamma_i^{(2,2)}(E_v, E_h) + \sum_{i \diamond j \in \mathcal{C}_{(2,3)}(E_v, E_h)} \gamma_{i \diamond j}^{(2,3)}(E_v, E_h) \\ & + \sum_{i \diamond j \in \mathcal{C}_{(3,2)}(E_v, E_h)} \gamma_{i \diamond j}^{(3,2)}(E_v, E_h) \end{aligned} \quad (10)$$

where

$$\begin{aligned} \sum_{i \in \mathcal{C}_{(2,2)}(E_v, E_h)} \gamma_i^{(2,2)}(E_v, E_h) = & \beta_v (V_v(a_{6,5}, a_{0,10}) + V_v(a_{0,10}, a_{15,14})) \\ & + \beta_h (H_h(a_{7,11}, a_{6,0}) + H_h(a_{6,0}, a_{5,10})), \end{aligned} \quad (11)$$

$$\sum_{i \diamond j \in \mathcal{C}_{(2,3)}(E_v, E_h)} \gamma_{i \diamond j}^{(2,3)}(E_v, E_h) = \beta_{vv} V_{vv}(a_{6,5}, a_{0,10}, a_{15,14}), \quad (12)$$

and

$$\sum_{i \diamond j \in \mathcal{C}_{(3,2)}(E_v, E_h)} \gamma_{i \diamond j}^{(3,2)}(E_v, E_h) = \beta_{hh} H_{hh}(a_{7,11}, a_{6,0}, a_{5,10}), \quad (13)$$

and where $\beta_{vv} = 2\beta_v$ and $\beta_{hh} = 2\beta_h$. $\mathcal{C}_{(2,2)}(E_v, E_h)$, $\mathcal{C}_{(2,3)}(E_v, E_h)$ and $\mathcal{C}_{(3,2)}(E_v, E_h)$ are respectively the sets of 2×2 , 2×3 and 3×2 cliques for the sites included in E_v or E_h specific to figure 1.

We note that these terms take into account only edges present in the specified positions. If we sum these terms over all possible cliques of these types, thus taking into account all the sites of the lattice, we arrive at a Gibbs energy for the entire field which takes into account vertical and horizontal edges anywhere in the lattice. Thus, we have

$$U(\mathbf{d}) = \sum_{i \in \mathcal{C}_{(2,2)}} \gamma_i^{(2,2)}(\mathbf{d}) + \sum_{i \diamond j \in \mathcal{C}_{(2,3)}} \gamma_{i \diamond j}^{(2,3)}(\mathbf{d}) + \sum_{i \diamond j \in \mathcal{C}_{(3,2)}} \gamma_{i \diamond j}^{(3,2)}(\mathbf{d}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{d}), \quad (14)$$

where $\mathcal{C}_{(2,2)}$, $\mathcal{C}_{(2,3)}$ and $\mathcal{C}_{(3,2)}$ are the sets of all cliques of the corresponding type, and \mathcal{C} is the set of all cliques.

This model takes into account only vertical and horizontal edges. But edges oriented differently may be approximated closely by an appropriate zigzag combination of vertical and horizontal edges of small length. This is especially the case in discrete images, where one cannot have continuous straight edges oriented at an angle.

Each depth site is a member of four 2×2 cliques, six 2×3 cliques and six 3×2 cliques. The neighbourhood of an interior depth site is a 5×5 window, centred at the site and excluding the site and the four corners, and is thus of the fourth order (figure 1). As mentioned earlier, the neighbourhood is modified at the boundary sites in accordance with the free boundary assumption.

To generalize the model, we may use different weighting parameters for the potentials for the 2×3 and 3×2 cliques, i.e. we may use for β_{vv} and β_{hh} values different from the ones that arise out of the above derivation.

2.2 Posterior Gibbs distribution

We assume that the noise process \mathcal{N} corrupting the depth process \mathcal{D} to give the observation process $\mathcal{G} = \mathcal{D} + \mathcal{N}$ is white Gaussian, with mean μ and variance σ^2 , and is independent of \mathcal{D} . Then the degradation model is a conditional probability distribution for \mathcal{G} given \mathcal{D} , namely,

$$P\{\mathcal{D} = \mathbf{d} | \mathcal{G} = \mathbf{g}\} = \frac{P\{\mathcal{G} = \mathbf{g} | \mathcal{D} = \mathbf{d}\} P\{\mathcal{D} = \mathbf{d}\}}{P\{\mathcal{G} = \mathbf{g}\}},$$

where

$$\begin{aligned} P\{\mathcal{G} = \mathbf{g} | \mathcal{D} = \mathbf{d}\} &= P\{\mathcal{D} + \mathcal{N} = \mathbf{g} | \mathcal{D} = \mathbf{d}\} \\ &= P\{\mathcal{N} = \mathbf{g} - \mathbf{d}\} \\ &= (1/2\pi\sigma^2)^{N/2} \exp\left\{-1/2\sigma^2 \sum_{i=1}^N (\mathbf{g}_i - \mathbf{d}_i - \mu)^2\right\}, \end{aligned}$$

$$P\{\mathcal{D} = \mathbf{d}\} = \frac{\exp(-U(\mathbf{d})/T)}{Z},$$

and

$$P\{\mathcal{G} = \mathbf{g}\} = \text{a constant.} \quad (15)$$

Analogous to the approach in Geman & Geman (1984), it can be shown that the degradation model is also Gibbsian, with the same neighbourhood structure as \mathcal{D} ,

$$P\{\mathcal{D} = \mathbf{d} | \mathcal{G} = \mathbf{g}\} = \frac{\exp(-U_p(\mathbf{d})/T)}{Z}. \quad (16)$$

The posterior Gibbs energy $U_p(\mathbf{d})$ in this case will be given by

$$U_p(\mathbf{d}) = U(\mathbf{d}) + \|\mathbf{g} - \mathbf{d} - \mu\|^2/2\sigma^2. \quad (17)$$

The MAP estimate of \mathcal{D} corresponds to that configuration \mathbf{d} for which the posterior energy $U_p(\mathbf{d})$ is minimum. The minimization of the posterior energy is done using the simulated annealing algorithm (Kirkpatrick *et al* 1983; Aarts & Korst 1989).

2.3 Initialization

Ideally, one can start the simulated annealing algorithm with any initial configuration \mathbf{d} and converge to the global minima with probability 1. However, this requires infinite time. In practice, a finite time approximation for the algorithm is carried out, and a ‘‘good’’ estimate of \mathcal{D} is used as the starting configuration. The simplest way of generating an initial estimate from the given data is by filling in missing observations by the average of the neighbouring data values (assuming that the latter are present). However, it has been our experience that this gives poor results, which may be due to the fact that this approach does not utilize the clique model. Very good initial estimates can be generated with the clique model by using ‘‘local’’ MAP estimates of the depth values at sites at which observations are not present. These estimates are taken with respect to the local characteristics (conditional distributions) of the depth values. The local characteristic of \mathcal{D}_i describes the probability distribution of the

value of site i , given the (fixed) values at all the other sites, and is given by

$$P\{\mathcal{D}_i = \alpha | \mathcal{D}_j = \mathbf{d}_j, j \neq i\} = \frac{\exp(-U(\mathbf{d})/T)|_{\mathbf{d}_i = \alpha}}{\sum_{\hat{\alpha} \in \Omega} (\exp(-U(\mathbf{d})/T)|_{\mathbf{d}_i = \hat{\alpha}})}, \quad (18)$$

where $\alpha \in \Omega$. We emphasize that the notation $|_{\mathbf{d}_i = \alpha}$ indicates that the expression immediately to the left is evaluated with the random variable \mathcal{D}_i at site i set to α , with the other random variables \mathcal{D}_j kept at the conditioning values \mathbf{d}_j .

We can simplify this by expressing the posterior Gibbs energy as a sum of terms that depend on \mathbf{d}_i and terms that do not:

$$U(\mathbf{d}) = \sum_{c \in \mathcal{C}_i} V_c(\mathbf{d}) + \sum_{c \in \bar{\mathcal{C}}_i} V_c(\mathbf{d}),$$

where \mathcal{C}_i is the set of cliques that contain the site i , and $\bar{\mathcal{C}}_i$ denotes the complement of \mathcal{C}_i . Noting that the latter terms are constant and can be cancelled from the numerator and denominator, we arrive at the reduced expression

$$P\{\mathcal{D}_i = \alpha | \mathcal{D}_j = \mathbf{d}_j, j \neq i\} = \frac{\exp(-\sum_{c: i \in c} V_c(\mathbf{d})/T)|_{\mathbf{d}_i = \alpha}}{\sum_{\hat{\alpha} \in \Omega} (\exp(-\sum_{c: i \in c} V_c(\mathbf{d})/T)|_{\mathbf{d}_i = \hat{\alpha}})} \quad (19)$$

The local MAP estimate for \mathcal{D}_i is that value of \mathbf{d}_i for which $\sum_{c \in \mathcal{C}_i} V_c(\mathbf{d})$ is minimum.

During initialization, or computing the initial estimates for sites with missing observations, it may happen that the neighbourhood of those sites contain sites at which observations are absent. In such cases, the potentials at cliques which contain such sites are set to zero. Since such terms drop out of the summation, the other terms are correspondingly scaled up.

Although time consuming, it was observed that the local MAP estimate approach gives far better depth recovery than the neighbourhood averaging approach.

3. Simulation results

Simulations were carried out on 240×240 range data images. The depth image chosen for display here is that of a collection of pipes (figure 3a). The intensity of a pixel was proportional to the height of the corresponding point over a fixed datum, and was quantized to lie from 0 to 511. Noisy depth images were generated by adding zero mean ($\mu = 0$) white Gaussian noise to the original image. To obtain sparse noisy depth images, a specified percentage of data points at random positions of the noisy depth images were made invalid. Figures 3b, 3e, 4b and 4e show the noisy sparse images for various degrees of noise and sparseness as given in table 1. In all cases, the inverse log cooling schedule was used for annealing. The results were found to be relatively insensitive to the starting temperature in the range 1 to 4. The clique parameters were estimated by trial and error. The parameters chosen were $\beta_h = \beta_v = 1.5$ and $\beta_{hh} = \beta_{vv} = 3.0$. During simulated annealing, the generating mechanism picked up depth values from a Gaussian distribution. At each site, the mean of the distribution was the current depth value at that site, and the variance decreased linearly with the temperature, i.e.

$$\text{variance} = \text{scale factor} \times \text{temperature},$$

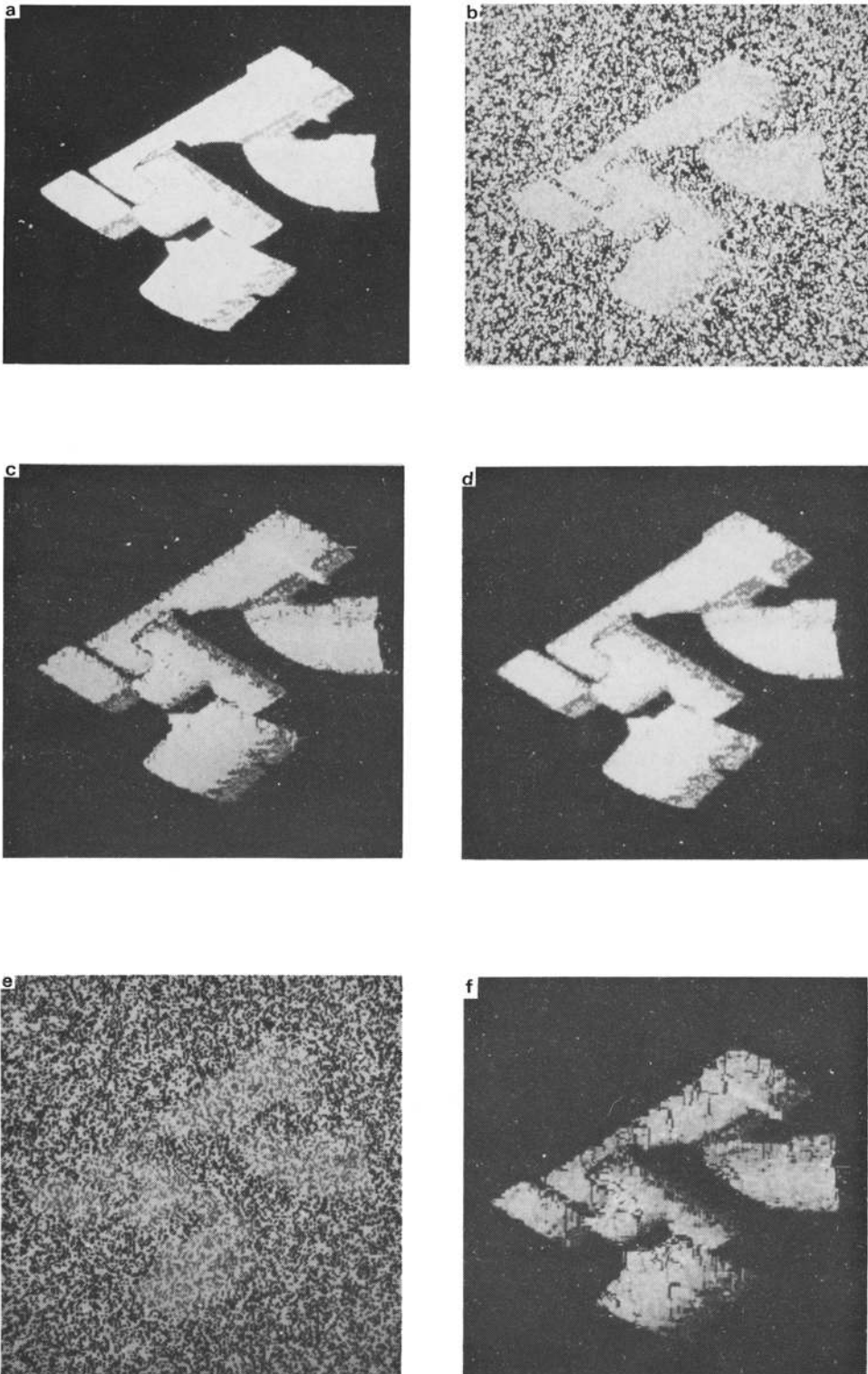


Figure 3. (Continued)

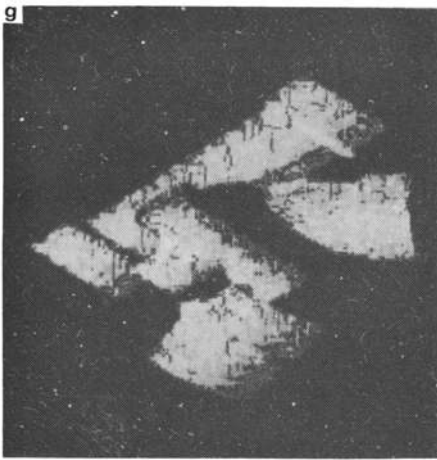


Figure 3. (Continued) (a) Original image; (b) 40% data loss; (c) initial estimate; (d) after annealing; (e) 60% data loss; (f) initial estimate; (g) after annealing.

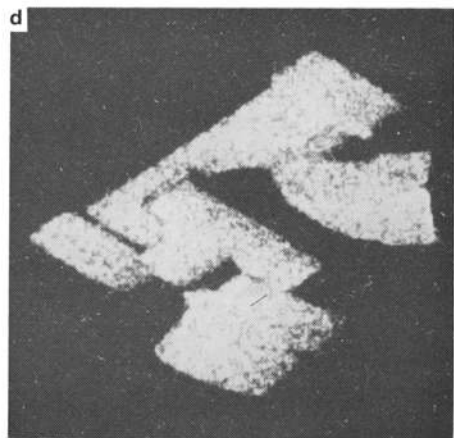
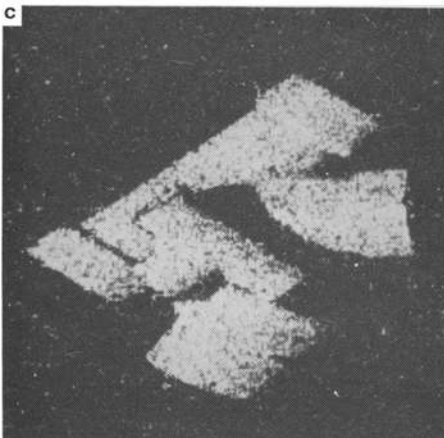
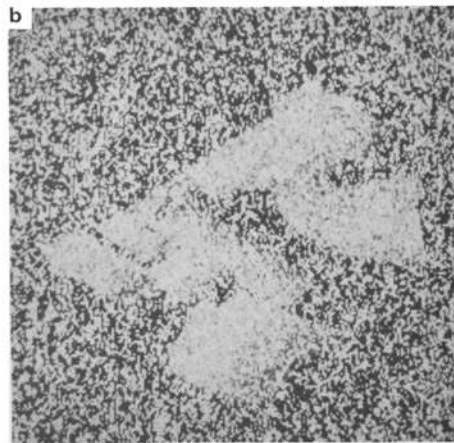
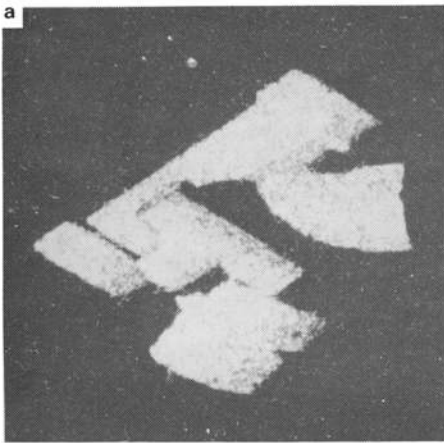


Figure 4. (Continued)

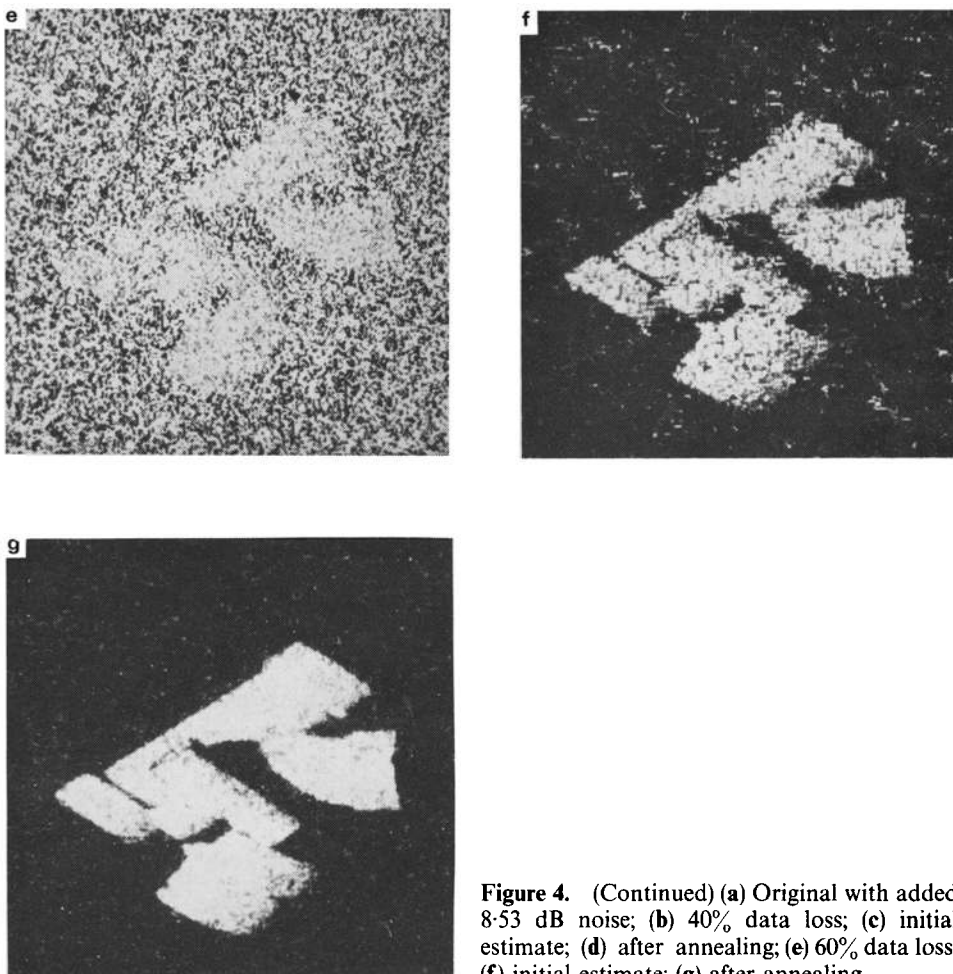


Figure 4. (Continued) (a) Original with added 8.53 dB noise; (b) 40% data loss; (c) initial estimate; (d) after annealing; (e) 60% data loss; (f) initial estimate; (g) after annealing.

where scale factor was chosen to be 10–20 for images with high signal to noise ratio (SNR) (8 dB or more), and about 40–50 for lower SNR images.

During initialization, the sparse image was scanned in raster scan fashion to check for missing data points. At each such point, the posterior Gibbs energy (19) was computed for every fifth depth value in the range of all possible depth values (in our case 0 to 511). The depth value for which it was minimum was chosen as the local MAP estimate. The sampling of the range of depth values was done to reduce the

Table 1. Performance time.

Figure set	SNR (dB)	Data loss (%)	Interpolation time (s)	Annealing time (s)
3	No noise	40	618	9199
	No noise	60	908	9244
4	8.53	40	786	9057
	8.53	60	942	9101

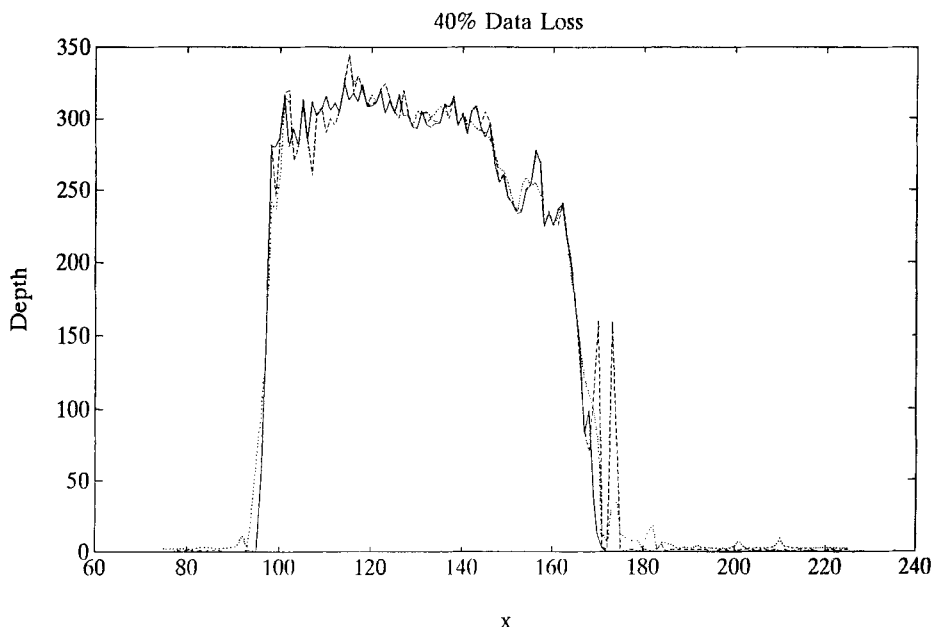


Figure 5. Depth values in the 100th row (— continuous; --- initial estimate; after 200 iterations).

computation involved in initialization (by a factor of five). Figures 3c, 3f, 4c and 4f show the interpolated starting estimates for the different cases.

For about 40 to 50% data loss, excellent recovery was obtained (figures 3d and 4d). For 60 to 75% data loss, the recovery was partial but encouraging (figures 3g and 4g). For noisy images, good results were obtained when the SNR was over about 5 dB. Lower signal to noise ratios required larger number of iterations of the algorithm. From the figures (3c, 3f, 4c, 4f), it can be seen that the local estimate of the depth map is itself a good estimate of the original image, even for high sparseness and noise levels. Thus, the time-consuming process of annealing can be avoided for applications where the initial estimate itself provides sufficient information for decision making.

Figure 5 shows the performance of the algorithm across a discontinuity in one of the rows of the depth map. The plots show the depth values in the 100th row in the original depth map, the initial estimate and the final output. As can be seen, for even large amounts of sparseness, the clique model (10)–(13) is able to recover the discontinuities very well.

All simulation results were done on a 33 MHz 80386 based PC-AT. Table 1 tabulates the CPU times taken by various stages of the algorithm. The annealing times are for 200 iterations of the simulated annealing algorithm.

4. Conclusions

The clique model developed was found to be very effective for recovery of depth information from noisy and sparse real depth images. Although the model was especially tailored for recovery of straight line edges, it was also found to be capable of recovering other types of edges found in real depth images.

For simulations reported here, parameters were estimated by a trial and error method. An algorithm that incorporates parameter estimation for the clique model is clearly preferable. Various methods have been proposed for this purpose, based on maximum likelihood estimation (Younes 1988) and maximum pseudo-likelihood estimation (Besag 1977; Geman & Graffigne 1987; Younes 1988; Lakshmanan & Derin 1989). We propose to incorporate into our algorithm a parameter estimation module based on these methods.

Algorithms based on simulated annealing usually are very computationally intensive, which explains the large CPU times in table 1. However, these would be substantially reduced for better machines like an RISC machine, or a massively parallel machine like the connection machine. A useful paradigm for the parallel implementation of this algorithm is the Boltzmann machine, which is a stochastic neural network formulation that uses simulated annealing to minimize an energy function. A Boltzmann machine implementation of the depth model presented in this paper is described in Mundkur *et al* (1992).

The simulated annealing algorithm for the MRF model in this paper gives an initial local estimate of the depth map that, in most cases, is itself a good estimate of the original image. Thus, the time-consuming process of annealing can be avoided for applications where the initial estimate itself provides sufficient information for decision making.

Another approach to speed up annealing is to modify the simulated annealing algorithm to use a small amount of problem-dependent heuristic information. Such an approach has been tried for the problem described in this paper, and has been found to give both excellent speed up as well as much improved recovery results for the initial *and* annealed estimates (Kapoor *et al* 1992).

We would like to thank Prof. S Chaudhuri and Prof. P G Poonacha for their valuable suggestions and comments during various discussions. We would also like to acknowledge the Pattern Recognition and Image Processing (PRIP) Lab at Michigan State University, East Lansing, USA, for making available the range data used in this paper.

This work was supported by the MHRD grants for Laboratory for Artificial Neural Networks and for Digital Communication and Signal Processing.

References

- Aarts E, Korst J 1989 *Simulated annealing and Boltzmann machines* (Chichester: Wiley)
- Bertero M, Poggio T, Torre V 1988 Ill-posed problems in early vision. *Proc. IEEE* 76: 869–889
- Besag J E 1974 Spatial interaction and statistical analysis of lattice systems. *J. R. Stat. Soc. B-36*: 192–236
- Besag J E 1977 On the statistical analysis of dirty pictures. *J. R. Stat. Soc. B-48*: 259–302
- Bolle R M, Vemuri B C 1991 On three-dimensional surface reconstruction methods. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-13: 1–13
- Boult T E, Kender J R 1986 Visual surface reconstruction using sparse depth data. *Proc. IEEE Conf. Comput. Vision Pattern Recog.* (New York: IEEE)
- Choi D J, Kender J R 1988 Solving the depth interpolation problem on a parallel architecture with multigrid approach. *Proc. IEEE Conf. Comput. Vision Pattern Recog.* (New York: IEEE)
- Frankot R T, Chellappa R 1988 A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-10: 439–451

- Geiger D, Giosi F 1991 Parallel and deterministic algorithms from MRFs: Surface reconstruction. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-13: 401–412
- Geman S, Geman D 1984 Stochastic relaxation, Gibbs distributions and the Bayesian distribution of images. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-6: 721–741
- Geman S, Graffigne C 1987 Markov random fields image models and their applications to computer vision. *Proc. Int. Congr. Math.* (ed.) A M Gleason (Providence, RI: Am. Math. Soc.)
- Horn B K P, Brooks M J (eds) 1989 *Shape from shading* (Cambridge, MA: MIT Press)
- Kapoor S, Mundkur P Y, Desai U B 1992 Depth and image recovery using a MRF model. Technical Report, Dept. Elec. Eng., Indian Institute of Technology, Bombay
- Kirkpatrick S, Gelatt G D, Vecchi M P 1983 Optimization by simulated annealing. *Science* 220: 671–680
- Lakshmanan S, Derin H 1989 Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-11: 799–813
- Marroquin J, Mitter S, Poggio T 1987 Probabilistic solution of ill-posed problems in computational vision. *J. Am. Stat. Assoc.* 82: 76–89
- Mundkur P Y, Kapoor S, Desai U B 1992 Boltzmann machines for depth recovery using a MRF model. *Proc. Int. Joint Conf. Neural Networks*, Baltimore
- Prasad K V V, Mammone R J, Yogeshwar J 1990 3-D restoration using constrained optimization techniques. *Opt. Eng.* special issue on *Image Restoration and Reconstruction*
- Shao M, Simchony T, Chellappa R 1987 New algorithms for reconstruction of a 3-D depth map from one or more images. *Proc. IEEE Conf. Comput. Vision Pattern Recog.* (New York: IEEE)
- Terzopoulos D 1988 The computation of visible surface representation. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-10: 417–438
- Younes L 1988 Estimation and annealing for Gibbsian fields. *Ann. Inst. Henri Poincaré* 24: 269–294
- Zerubia J, Chellappa R 1990 Mean field annealing for edge detection and image restoration. *Proc. European Signal Proc. Conf.* Barcelona, Spain