SECUR

AD-A248 471

REPORT DOCUMENTATION PAGE

1a. RE	1b. RESTRICTIVE MARKINGS None
2a. SE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NONE	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 92 0186
6a. NAME OF PERFORMING ORGANIZATION Worcester Polytechnic Institute / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Dr. Abraham Waksman/NM (202)767-5027
6c. ADDRESS (City, State, and ZIP Code) Worcester Polytechnic Institute, EE Dept. 100 Institute Road Worcester, Massachusetts 01609	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM, Building 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR / 8b. OFFICE SYMBOL (If applicable) AFOSR/NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-..
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM, Building 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A7; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Integration of Stereo Vision and Optical Flow Using Energy Minimization

12. PERSONAL AUTHOR(S)

Nasrabadi, Nasser Mohammadi

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final Report	FROM 10/15/89-10/14/91	To 92, Jan., 30	18

16. SUPPLEMENTARY NOTATION

NONE

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Stereo Vision, Computer Vision, Robotics, Stereo-Matching, Relaxation, Hough Transform, Binocular Stereo.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A cooperative motion-stereo method is proposed where image intensity (brightness) and optical flow information are integrated into a single stereo technique by modeling the input data as coupled Markov Random Fields (MRF). The Baysian probabilistic estimation method and the MRF-Gibbs equivalence theory are used to integrate the optical flow and the gray level intensity information to obtain an energy function which will explicitly represent the depth discontinuity and occlusion constraints on the solution. This energy function involves the similarity in intensity (or edge orientation) and the optical flow between corresponding sites of the left and right images as well as the smoothness constraint on the disparity solution. If a simple MRF is used to model the data, the energy function will yield a poor disparity by smoothing across object boundaries, particularly when occluding objects are present. We exploit optical flow information to indicate object boundaries (depth discontinuities) and occluded regions, in order to improve the disparity solution in occluded regions. A stochastic relaxation algorithm (Simulated Annealing) is used to find a favourable disparity solution by minimizing the energy equation.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Abraham Waksman	22b. TELEPHONE (Include Area Code) (202) 767-5027 / 22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

FINAL REPORT SUBMITTED TO

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)

Prepared by the Principal Investigator

Nasser M. Nasrabadi

January 25 1992

INTEGRATION OF STEREO VISION AND OPTICAL FLOW USING ENERGY MINIMIZATION APPROACH

Worcester Polytechnic Institute
Electrical Engineering Department
100 Institute Road
Worcester, MA, 01609.
(716) 636-2427

PUBLICATIONS CITING THE AFOSR GRANT

JOURNAL PUBLICATIONS

[1] Nasrabadi, N. M., Clifford, S., and Liu, Y., "Integration of stereo vision and optical flow," J. Opt. Soc. Am. A, Vol. 6, No. 6, pp. 900-907, June 1989.

[2] Nasrabadi, N. M., and Liu, Y., "Stereo vision correspondence using a multi-channel graph matching technique," J. Image and Vision Computing, Vol. 7, No. 4, pp. 237-245, Nov. 1989.

[3] Nasrabadi, N. M., "Stereo Correspondence using Curve-segments and Relaxation", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 14, No. 3, March 1992.

[4] Nasrabadi, N. M., and Choo, Y. C., "Hopfield Network for Stereo Vision Correspondence," IEEE Trans. on Neural Networks, Vol. 3, No. 1, pp. 1-9, Jan. 1992.

[5] Nasrabadi, N. M., and Li, E., "Object Recognition by a Hopfield Neural Network," IEEE Trans. on Systems, Man and Cybernetics, Vol. 21, No. 6, Jan-Dec 1992.

[6] Nasrabadi, N. M., Lin, S., and Feng, Y., "Interframe Hierarchical Vector Quatization," J. of Optical Engineering, Vol. 28, No. 7, pp. 717-725, July

[7] Nasrabadi, N. M., and Feng, Y., "A Multi-Layer Address-Vector Quantization", IEEE Trans. on Circuits and Systems, Vol. CAS-37, No. 7, pp. 912-921, July 1990.

[8] Nasrabadi, N. M. and Feng Y., "Image compression using Address-Vector Quantization", IEEE Trans. on Communications, Vol. 38, No. 11, pp. 2166-2173, Dec. 1990.

[9] Nasrabadi, N. M., Roy, J. U. and Choo, C. Y., "An Interframe Hierarchical Address-Vector Quantization" to appear in IEEE Trans. on Communications, May 1992.

CONFERENCE PUBLICATIONS

[1] Nasrabadi, N. M., W. Li, and C. Y. Choo, "Object Recognition by a Hopfield Neural Network," in IEEE Third Int. Conf. Computer Vision (ICCV'90), Osaka, Japan, 4-7 Dec. 1990, pp. 325-328.

[2] Li, W. and N. M. Nasrabadi, "Object Recognition Based on Graph Matching Implemented by a Hopfield-Style Neural Network," IEEE Int. Joint Conf. on Neural Networks, IJCNN'89, June 18-22, 1989, vol. II, pp. 287-290.

[3] Nasrabadi, N. M., and Y. Liu, "Integration of Intesity-Based and Edge-Based Algorithms Using Markov Random Fields," IEEE International Conf. on Image Processing, ICIP'89, Singapore, 5-8 September 1989, pp. 207-211.

[4] Nasrabadi, N. M., S. E. Lin and Y. Feng, "Vector Quantization of Image Sequence Using a Regular Decomposition Quadtree," IEEE International Conf. on Image Processing, ICIP'89, Singapore, 5-8 September 1989, pp. 55-59.

[5] Nasrabadi, N. M., W. Li, B. G. Epranian, and C. A. Butkus, "Use of Hopfield Network for Stereo Vision Correspondence," IEEE Int. Conf. on Systems, Man, and Cybernetics, Cambridge, MA, 14-17 November 1989, pp. 429-432.

[6] Li, W., and N. M. Nasrabadi, "Object Recognition by a Global Technique Based on Graph-Matching by the Hopfield Network," SPIE Advances in Intelligent Robotics Systems, Intelligent Robots and Computer Vision VIII: Algorithms and Techniques, Philadelphia, Pennsylvannia, Vol. 1192, 6-10 Nov. 1989, pp. 425-442.

[7] Feng, Y. and N. M. Nasrabadi, "A Dynamic Finite State-Vector Quantizer Scheme," in the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Albuquerque, New Mexico, 3-6 April, 1990, pp. 2261-2264.

[8] Li, W. and N. M. Nasrabadi, "Invariant Object Recognition Based on a Neural Network of Cascaded RCE Nets," in IEEE Int. Joint Conf. on Neural Networks, San Diego, CA, 17-21 June, 1990.

Table of Contents

INTEGRATION OF STEREO VISION AND OPTICAL FLOW USING ENERGY MINIMIZATION APPROACH

Nasser M. Nasrabadi

Computer Vision Research Group
Electrical Engineering Department
Worcester Polytechnic Institute
Worcester, Massachusetts 01609
Tel: (716) 636-2427

ABSTRACT

A cooperative motion-stereo method is proposed where image intensity (brightness) and optical flow information are integrated into a single stereo technique by modeling the input data as coupled Markov Random Fields (MRF). The Bayesian probabilistic estimation method and the MRF-Gibbs equivalence theory are used to integrate the optical flow and the gray level intensity information to obtain an energy function which will explicitly represent the depth discontinuity and occlusion constraints on the solution. This energy function involves the similarity in intensity (or edge orientation) and the optical flow between corresponding sites of the left and right images as well as the smoothness constraint on the disparity solution. If a simple MRF is used to model the data, the energy function will yield a poor disparity by smoothing across object boundaries, particularly when occluding objects are present. We exploit optical flow information to indicate object boundaries (depth discontinuities) and occluded regions, in order to improve the disparity solution in occluded regions. A stochastic relaxation algorithm (Simulated Annealing) is used to find a favorable disparity solution by minimizing the energy equation.

1. INTRODUCTION

Integration of visual cues have recently been investigated by many researchers [1]- [2]. Stereo vision [3]-[4] and optical flow [5] information have been used to obtain 3-dimensional structure about the environment. The major problems encountered in stereo matching are identifying the occluded areas and the depth discontinuities. Our objective in this research is to develop a motion-stereo matching algorithm that is minimally affected by occlusion and depth discontinuities. This is down by integrating the edge information and short range optical flow data into our stereo matching algorithm.

In this report a motion-stereo vision technique is investigated where short range optical flow information [5] is integrated into an intensity-based stereo vision technique. A Bayesian model is used to derive the Maximum A Posteriori (MAP) stereo matched solution for a motion-stereo intensity matching algorithm. The input data and the disparity solution are modeled as Markov Random Fields (MRFs). The MRF-Gibbs Distribution equivalence [6] reduces the MAP problem to that of finding an appropriate energy function that describes the constraints on the solution. Stereo algorithms based on low-level information (intensity, edges for example) have difficulty with occlusion as well as depth discontinuities because they have to make decisions about object boundaries in order to find the correct disparity for a region that is not visible in both images. If an object boundary is not pronounced, the disparity solution for the occluded region will usually be a smoothly changing function between the disparity of the occluded object and the visible object. In this report the optical flow data is used to flag the potentially occluded regions, and when a region is identified as occluded, a discontinuity in the disparity solution is assumed. However, the resulting energy function is no longer convex so a stochastic relaxation technique is employed to find the minima of the function. In stereo images occlusion occurs when some points are visible in one image but not in the other due to the displacement between the two cameras. For instance, this will happen when an object looks like it is behind another one from the angle of the camera. Figure 1 shows the mismatch resulting from a stereo matching algorithm which does not consider occlusion. Occlusion problem has been tackled by other researchers who have used the rate of change of disparity [7], or the "ordering constraint" [8] to flag a potentially occluded region.

-In Section 2 of this report we describe our algorithm in detail, and presents properties of MRFs and their Gibbs equivalence distribution, as well as the used energy functions. In section 3, the effect of occlusion and depth discontinuities are reported with experimental results.

2. DESCRIPTION OF THE MOTION-STEREO MATCHING ALGORITHM

The approach of our stereo matching algorithm is to find an estimate of the "optimal" disparity solution for a pair of stereo images. Optimality is defined as the Maximum A Posteriori probability (MAP) solution. The MRF model allows us to integrate the low level visual cues as well as reducing the stereo matching problem to that of finding an appropriate energy function. We minimize the energy function and obtain a MAP estimate by using a well established stochastic optimization technique. In the following subsections we introduce MRFs and the Coupled MRFs, MRF-Gibbs distribution equivalence, energy functions representing the constraints on the solutions.

2.1 Markov and Coupled Markov Random Fields

The stereo matching problem is an ill-posed problem in the sense defined by Hadamard [9-10], which can be solved by using a standard regularization method [11-13]. The standard regularization theory produces a convex quadratic energy function (a cost functional) thus with a unique solution. However, in order to include the occlusion process and to deal effectively with the integration of visual cues, a non-convex energy function is derived. Therefore, in contrast to standard regularization methods, we represent the a priori knowledge in terms of probability distributions rather than constraints on the actual solution space. Specifically, we maximize the a posteriori probability of the disparity solution given the degraded (i.e., noisy) images by modeling the disparity solution as a Markov Random Field D over a lattice V_N. Assuming that in general the disparity is smooth over rigid objects except at depth discontinuities the MRF property can be stated by Equation (1) which says that probability of a disparity value d_{ij} at a pixel location (i, j), given the disparity of all other pixels in the image, is the same as the probability of that value given only the disparity of the neighbors of site (i, j)

$$Prob(D = d_{ij} \mid d_{kl},\ (k,l) \neq (i,j)) = Prob(D = d_{ij} \mid d_{kl},\quad (k,l) \in N_d(i,j)) \tag{1}$$

where $N_d(i,j)$ specifies the pixel indices of those defined to belong to the neighborhood system of (i,j).

The disparity data with the depth discontinuity process or occlusion process can be combined together to form a coupled MRF [6]. For example, define the depth discontinuity as a binary MRF $\Delta\rho_{ij}(i_N,j_N)$ defined over a lattice U_N as

$$\Delta\rho_{ij}(i_N,j_N) = \begin{cases} 1, & \text{if } [\, \rho^r(i,j) - \rho^r(i_N,j_N)\,]^2 \geq T_\rho \\ 0, & otherwise \end{cases} \tag{2}$$

where $i_N, j_N \in N_\rho(i,j)$ represents the eight nearest-neighborhood system, T_ρ is a pre-defined threshold, and $\rho^r(i, j)$ represents the magnitude of the optical flow. This binary process is very similar to the line process introduced by Geman and Geman [6]. We now associate this binary process with the disparity process to form a coupled MRF defined as $D = (D, \Delta\rho)$ over a lattice $H_N = V_N \cup U_N$ where there is a line process between each pair of disparity sites.

Each input image can also be considered to be a combination of two MRF's coupled together to represent the corresponding intensity and optical flow information. For example, let the intensity image be a MRF X on a lattice Z_N such that,

$$X = I(i, j) \text{ if } (i,j) \in Z_N \ ; \ Z_N = \left\{ (i,j) : 1 \leq (i,j) \leq N \right\} \tag{3}$$

where $I(i, j)$ is the gray level of the image of size $N \times N$.

The neighborhood system ζ_x for each pixel $I(i, j)$ is defined as its eight nearest-neighboring pixels. Similarly the optical flow information can also be considered as a MRF O on a lattice W_N such that,

$$O = \rho(i,\ j) \ \text{ if } \ (i,j)\ \varepsilon\ W_N \ ; \ \ W_N = \left\{ (i,j) : 1 \le (i,j) \le N \right\} \tag{4}$$

Where $\rho(i,\ j)$ represents the magnitude of the optical flow at location $(i,\ j)$. The neighborhood system ζ_o for each pixel is defined as the eight nearest-neighbors.

Now each input image can be considered as a coupled MRF $F = (X,\ O)$ a collection of an intensity process X and optical flow process O. The values of the field F is defined as:

$$F = \begin{cases} X_{i,j} = I(i,\ j) & \text{if} \quad i,j\ \varepsilon\ Z_N \\ O_{i,j} = \rho(i,\ j) & \text{if} \quad i,j\ \varepsilon\ W_N \end{cases} \tag{5}$$

where F is defined as a MRF on a lattice $S_N = Z_N \cup W_N$ the components of F assumes values among the allowable gray level and optical flow values. The interaction between the intensity process X and the optical flow process O will become clear in section 2.3 when correspondence problem (energy expression) is considered. We can define a pair of stereo images as two coupled MRFs $F^l = (X^l,\ O^l)$ and $F^r = (X^r,\ O^r)$ representing the left and the right image data respectively. Each coupled MRF can be considered as an observation or an external input to the stereo system.

2.2 MRF-Gibbs Distribution Equivalence

It has been shown [14-17] that if an image measure (such as disparity) can be modeled by a MRF, then the probability distribution of the measure can be represented as a Gibbs distribution (Hammersley-Clifford theorem). Equation (6) defines the probability of arriving at a solution configuration ω, with a Gibbs distribution given by

$$P\,(D{=}\omega) = \frac{e^{-U(\omega)/kT}}{Z} \tag{6}$$

where Z, is the normalizing constant such that the sum of $P(\omega)$ for all configurations is unity:

$$Z = \sum_{\omega} \exp\left[-U(\omega)/kT\right] . \tag{7}$$

T is analogous to the temperature of the system that yields the configuration ω, and k is a universal constant. $U(\omega)$ is called the energy function and is of form,

$$U(\omega) = \sum_{C} U_c(\omega) \tag{8}$$

where $U_c(\omega)$ is called a potential function and is defined over the neighborhood's cliques. A clique is a set of sites (pixels) such that for a defined neighborhood system every pair of sites belonging to this set are neighbors. An example of neighborhood system and possible potential functions used in this proposal is given below:

1) The potential function for a simple disparity MRF with no discontinuity process at a pixel location (i,j) with a neighborhood system $N_d(i,j)$ is defined by

$$U_c(d_{ij}) = \sum_{i_N, j_N} [\, d(i,\ j) - d(i_N,\ j_N)]^2 \qquad i_N, j_N \ \varepsilon\ N_d(i,\ j) \tag{9a}$$

where the neighborhood system $N_d(i,\ j)$ represents only the eight nearest-neighbors.

2) The potential function for a coupled MRF (disparity with discontinuity) at pixel location (i,j) with a neighborhood system $M_d(i,\ j)$ is defined by

$$U_c(d_{ij}) = \sum_{i_M, j_M} [\, d(i,j) - d(i_M, j_M)\,]^2 \cdot (1 - \Delta\rho_{ij}(i_M, j_M)) + \Delta\rho_{ij}(i_M, j_M) \cdot c_\rho; \quad i_M, j_M \ \varepsilon\ M_d(i,\ j) \tag{9b}$$

where $M_d(i,\ j)$ is simply the 8-nearest neighbors and c_ρ is a constant. Equation (6), with the potential function given by expression (9a) or (9b) is said to represent the prior distribution of the disparity solution. It is easily seen that low energy configurations represent a higher probability than high energy configurations. The $\beta = kT$ factor affects the equilibrium state of the solution. At high temperatures, the solution's equilibrium state will be very random. At low T, however, the equilibrium state is more regular and dependent on its initial configuration.

The advantage of representing each probability in the Gibbs distribution form is that it can be formulated with an energy function $U(\omega)$ which suitably describes the constraints of the solution. These constraints are similar (if not identical) to those used in a standard regularization method, but the interactions between the depth discontinuity and disparity smoothness processes can now be explicitly included in the energy function.

2.3 The Energy Function for the Proposed Motion-Stereo Matching Algorithm

The energy function describes the constraints on the desired solution in terms of local characteristics [18]. For example, there should be a strong similarity in the intensity (or orientation if an edge) between two corresponding sites in the stereo solution. Likewise, the disparity solution should be smooth over the same object.

The MRF model allows us to reduce the stereo matching problem to that of finding an expression for the a posteriori probability $P(D \mid F^l, F^r)$ in terms of the a priori probability $P(D)$ and the conditional probability of observations. Then using the MRF-Gibbs (or Boltzmann) distribution equivalence, one can determine an appropriate energy function representing the corresponding MAP.

The left and the right image data are defined as two observations corresponding to random fields F^l and F^r respectively. Our goal is to determine the most likely solution $D = (d_{i,j})$ given the above observations. Using the Bayesian rule the a posteriori probability can be written as

$$P(D \mid F^l, F^r) = \frac{P(F^l, F^r, D)}{P(F^l, F^r)} = \frac{P(F^l \mid D, F^r) \cdot P(D, F^r)}{P(F^l, F^r)} = \frac{P(F^l \mid D, F^r) \cdot P(D \mid F^r)}{P(F^l \mid F^r)}. \tag{10}$$

If the disparity D is independent of the observation F^r then $P(D \mid F^r) = P(D)$ which can be replaced by its corresponding Gibbs distribution given by expression (6). The conditional probability $P(F^l \mid F^r)$ is a constant since it is independent of D. The key term in expression (10) is the conditional probability $P(F^l \mid D, F^r)$ which has to be written in Gibbs distribution form. From the stereo camera geometry we have $F^l(i, j) = F^r(i, j+d_{ij})$ in absence of any noise or occlusion.

However, assuming F^l was degraded by a Gaussian random noise N independent of D and F^r with zero mean then,

$$P(N=x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-x^2}{2\sigma^2}\right] \tag{11}$$

$$P(F^l \mid D, F^r) = \prod_{i,j \,\varepsilon\, Z_N} \frac{1}{\sqrt{2\pi}\sigma} \exp - \left[\frac{(I^l(i, j) - I^r(i, j+d_{ij}))^2}{2\sigma^2}\right] \tag{12}$$

Expression (12) can be written in Gibbs distribution form as

$$P(F^l \mid D, F^r) = \frac{1}{Z} \exp\left\{-\frac{1}{kT} \sum_{i,j \in Z_N} U_{ij}(F^l \mid D, F^r)\right\} \tag{13}$$

where

$$U_{ij}(F^l \mid D, F^r) = \lambda \left[(I^l(i, j) - I^r(i, j+d_{ij}))^2\right] \tag{14}$$

with $\lambda = \frac{kT}{2\sigma^2}$.

The overall local energy contribution from a single site (i, j) can be written as

$$U_{ij} = U_{ij}(F^l \mid D, F^r) + U_c(d_{ij}) \tag{15}$$

where the first term represents data term the similarity in the intensity (or orientation if an edge) between two corresponding sites in the stereo solution as given by expression (14). The second term represents the smoothness of the solution given by expression (9a) or (9b) when depth discontinuity is present. The total energy is determined by totalling the energy contribution from each site.

In order to include depth discontinuities as well as occlusion process we assume that for each site the optical flow estimate $\vec{\rho}(i,j)$ is available for the left and the right images. There are a number of standard techniques that can be used to measure the optical flow information [5]. Optical flow is represented by a vector random field where $\vec{\rho}(i,j) = (u_{ij}, v_{ij})$, the two components of the optical flow. The difference in optical flow between the two corresponding sites is thresholded to indicate a potentially occluded region ($\phi_\rho = 1$) or a visible region ($\phi_\rho = 0$), only the magnitude of the optical flow is considered.

$$\phi_{\rho_{ij}} = \begin{cases} 1, & \text{if } [\ |\rho^l(i,j)| - |\rho^r(i+d_{ij},j)|\ ]^2 \geq T_\phi \\ 0, & otherwise \end{cases} \tag{16}$$

An occluded region is then processed differently than visible regions; heavily weighting the smoothness term over the data term. As shown in Equation (17), the data term D_{ij} does not use the measure of similarity of the corresponding pixels in an occluded region as they will not match for the correct disparity value. Instead, the data term is assigned a small fixed energy c_ϕ causing the smoothness term to have more weight in determining the correct disparity estimate. The non-zero cost (energy) associated with this process inhibits non-occluded areas from settling into a falsely-indicated occluded state.

$$D_{ij} = (1 - \phi_{\rho_{ij}}) \cdot [\; c_A[\, I^l(i,j) - I^r(i + d_{ij}, j)]^2 \;+\; c_B\,[\; |\rho^l(i,j) - |\rho^r(i + d_{ij}, j)|\;]^2\;] + \phi_{\rho_{ij}} \cdot c_\phi \quad (17)$$

where $(i,j) \,\varepsilon\, S_N$. The smoothness term S_{ij} is not changed significantly except that we now use the optical flow gradient to indicate potential discontinuities in the disparity solution. If the difference in optical flow of two neighboring sites exceeds a pre-specified threshold T_ρ then there is a discontinuity in the disparity and smoothness measure should not be performed across this discontinuity. Instead a small energy cost c_ρ is included, in this way we are introducing a local minima in our cost function.

$$S_{ij} = \sum_{i_M, j_M} [\; d(i,j) - d(i_M, j_M)\;]^2 \cdot (1 - \Delta\rho_{ij}(i_M, j_M)) + \Delta\rho_{ij}(i_M, j_M) \cdot c_\rho; \qquad i_M, j_M \,\varepsilon\, M_d(i,j) \quad (18)$$

where

$$\Delta\rho_{ij}(i_M, j_M) = \begin{cases} 1, & \text{if } [\; |\rho^r(i,j)| - |\rho^r(i_M, j_M)|\;]^2 \geq T_\rho \\ 0, & \textit{otherwise} \end{cases} \quad (19)$$

If the calculated optical flow is reliable it can be segmented into different regions representing each object and their boundaries [19] or a cooperative technique could be designed that can do the segmentation as the optical flow is calculated.

3. EXPERIMENTAL RESULTS

We work with motion-stereo images generated by two cameras situated in a parallel axis geometry yielding epipolar lines which are parallel to the scan lines. Thus, corresponding image points lie along the same horizontal scan line limiting the search space to one dimension. For the initial algorithm validation, we simulated the intensity and optical flow vector for stereo images of size 64 × 64 pixels. Our simulated image model includes random errors in both the image intensity and the optical flow estimate. Figure 2 shows a pair of stereo images simulated for experimental evaluation of the proposed technique.

For the preliminary experimental investigation it is assumed that optical-flow values for the motion stereo images are already estimated from a motion estimation technique. Although there are problems with calculating reliable optical flow data, but by means of post-processing, the optical flow information can be smoothed and segmented into regions representing different objects and can be used to locate the occluded regions as discussed in this report.

To maximize the probability given by Equation (10), and thus obtain the a MAP estimate, we minimize the total energy by achieving an equilibrium state at low temperature. We use a stochastic relaxation method called Simulated Annealing (SA) which is described in detail in [20-21]. The physical process of annealing brings materials to low energy states by *gradually* lowering their temperature. As in chemical annealing, if the temperature is dropped too quickly the solution does not have sufficient time to reach equilibrium states, resulting in configurations which correspond to local minima of the energy function; these constitute errors. The ideal temperature schedule which guarantees convergence to a minimum energy state, or the maximum a posteriori (MAP) configuration is given in [6] if the Gibbs sampler schedule is used; however, this schedule is impractically slow. For the preliminary evaluation of the algorithm we have used a non-ideal, monotonically decreasing temperature schedule and some result are presented. The Simulated Annealing algorithm can efficiently be implemented with parallel processing units, such as neural networks, due to its localized nature. The stochastic characteristic of the SA algorithm however, requires a random parameter to simulate the effect of lowering the temperature.

Figure 3 illustrates the disparity solution when no occlusion process is employed. The energy function used is a combination of the solution smoothness and intensity similarity constraints given by expressions (9a) and (14) respectively. Disparity error occurs in the occluded areas when this energy function is used. Here the solution has matched to the wrong pixel in an attempt to minimize the data term. Figure 4 shows the result of applying Equations (17)-(18) to images similar to those in Figure 2. The solution is much improved and yields a better disparity estimate in the occluded region. Figure 5 shows disparity solution at three different temperatures, $T = 50, 10,$ and 1 for the intensity-based stereo technique incorporating optical flow data. The temperature was decreased by 10% when there was no significant change in the energy value (system in an equilibrium state). The system is said to be in an equilibrium state whenever the number of jumps in the energy increase and decrease are about the same. Although there is no guarantee of convergence to a minimum energy state due to our monotonic temperature schedule, but a close approximation to the optimal solution is expected whenever the system is kept in equilibrium at each temperature.

5. REFERENCES

[1] E. Gamble, and T. Poggio, "Visual integration and detection of discontinuities: the key role of intensity edges," MIT AI Lab. Memo No. 970, (1987).

[2] Nasrabadi, N. M., Clifford, S. P., and Liu Y., "Integration of stereo vision and optical flow by using an energy-minimization approach," J. Opt. Soc. Am. A/Vol. 6, No.6, June 1989.

[3] U. R. Dhond and J. K. Aggarwal, "Structure from Stereo-A Review," IEEE Trans. on Systems and Cybernetics, vol. 19, no. 6, pp. 1489-1510, Nov/Dec. 1989.

[4] S. Barnard and M. Fishler, "Computational stereo," ACM Computing Surveys. 14, 553-572 (1982).

[5] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artificial Intell., 17, 185-203 (1981).

[6] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the baysian restoration of images," IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-6, 721-741 (1984).

[7] S. B. Pollard, E. W. Mayhew, and J. P. Frisby "PMF: A stereo correspondence algorithm using a disparity gradient limit," Perception 14, 449-470 (1985).

[8] H. H. Baker and T. O. Binford, "Depth from edge and intensity based stereo," in proc. of Seventh Int. Joint Conf. Artificial Intell., 631-636 (1981).

[9] Hadamard, J., "Lectures on the Cauchy problem in linear partial differential equations," Yale U. Press, New Haven, Conn., 1923

[10] Tikhonov, A. N. and Arsenin, Y. Z., "Solution of ill-posed problems," Winston, New York, 1977.

[11] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," Proceedings of the IEEE Vol. 76, 869-889 (1985).

[12] T. Poggio, V. Torre and C. Koch, "Computational vision and regularization theory," Nature Vol. 317, 314-319 (1985).

[13] T. Poggio, C. Koch, "Ill-posed problems in early vision: from computational theory to analogue networks," Proc. R. Soc. Lond. B 226, 303-323 (1985).

[14] E. Ising, "Beitrag sur theorie des ferromagnetismus," Zeit. fir Physik, Vol. 31, 253-258, (1925).

[15] R. Kindermann, and J. L. Snell, "Markov random fields and their applications," (Prov. RI, AMS, 1980).

[16] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," Journal of the American Statistical Association, Vol. 82, No. 397, 76-89 (1987).

[17] J. Besag, "On the Statistical Analysis of Dirty Pictures," J. Royal Statist. Soc., Vol. 48, Series B. 259-279, 1986.

[18] A. Yuille, "Energy functions for early vision and analog networks," MIT AI Lab. Memo No. 987, (1987).

[19] Thompson, W. B., Mutch, K. M., and Berzins, V. A., "Dynamic occlusion analysis in optical flow fields", IEEE PAMI-Vol. 7, no. 4, 1985.

[20] N. Metropolis and et. al., "Equation of state calculations by fast computing machines," J. Phys. Chem. 21, 1087-1091 (1953).

[21] S. Kirkpatrick, C. D., Jr. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science, 220, 671-680 (1983).

[22] C. Koch, J. L. Marroquin, and A. Yuille, "Analog 'neuronal' networks in early vision," Proc. Nat'l. Acad. Sci., Vol. 83, 4263-67 (1986).

[23] Hopfield, J. J., and Tank, D. W., "Neural Computation of Decisions in Optimization Problems", Biological Cybernetics, 52, PP. 141-152, 1985.

[24] Rumelhart, D. E., McClelland, J. L., "Parallel Distributed Processing", The MIT Press, Vol. 1 and Vol. 2, 1986.

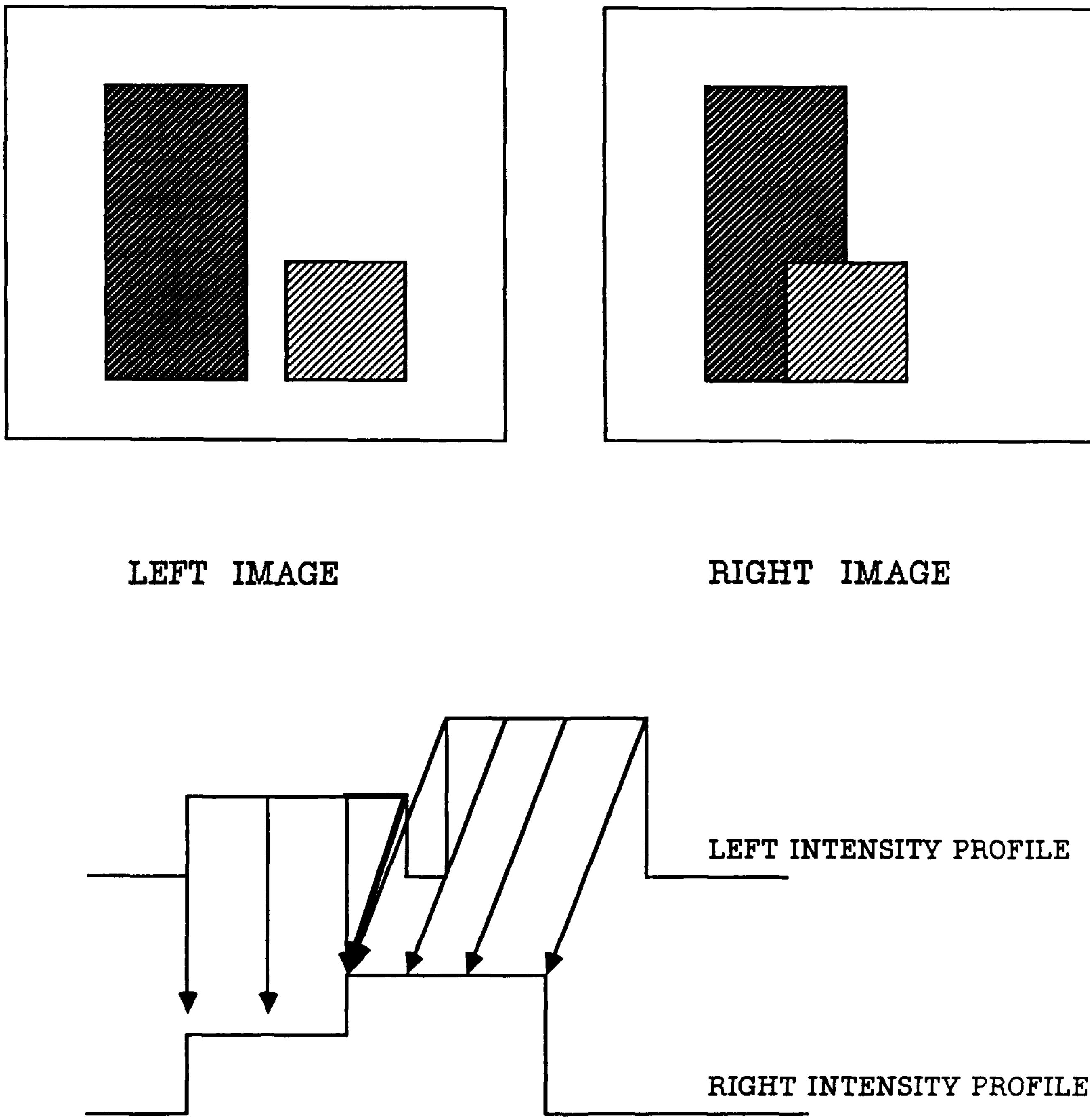

Figure 1. Typical stereo matching error in occluded regions.

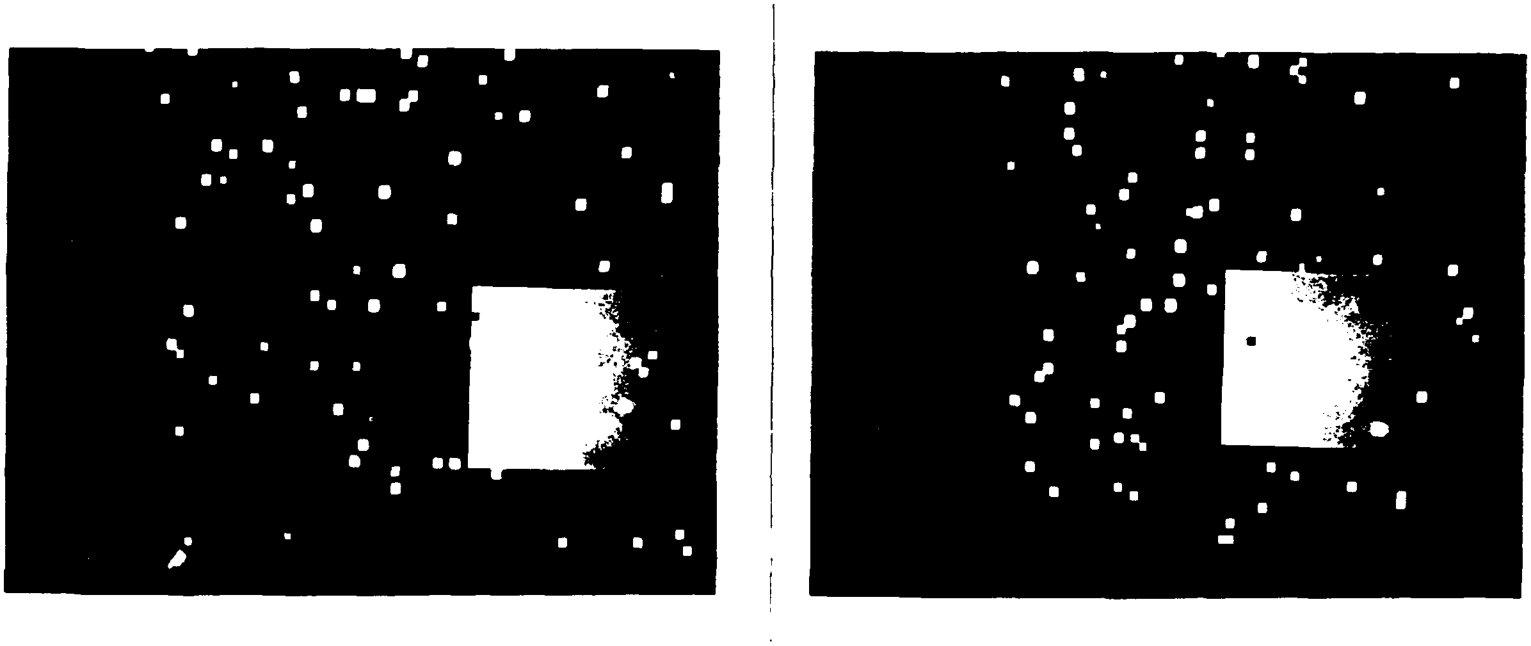

Figure 2. Typical left and right stereo images used to generate results in Figures 3-7, Gaussian noise is also introduced into the data. Disparity of left box =15; of right box = 5.

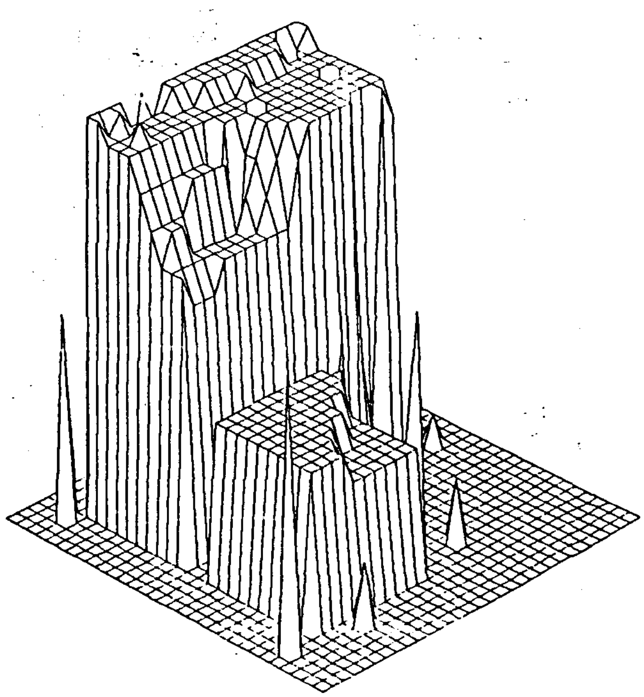

Figure 3. Disparity solution resulting from intensity matching without optical flow data.

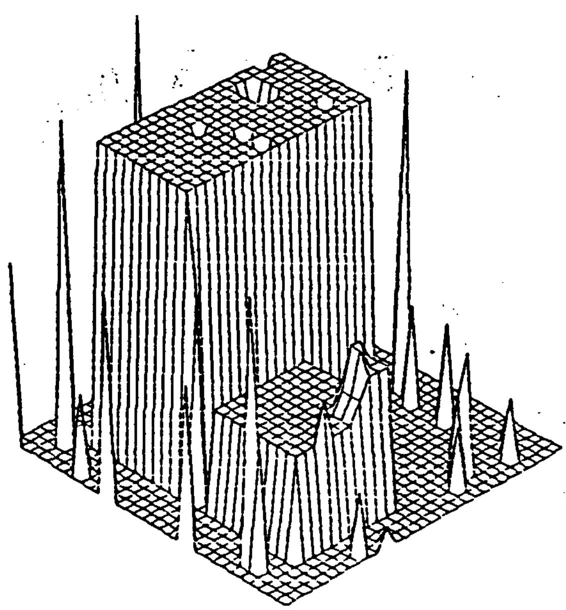

Figure 4. Disparity solution resulting from intensity matching incorporating optical flow data.

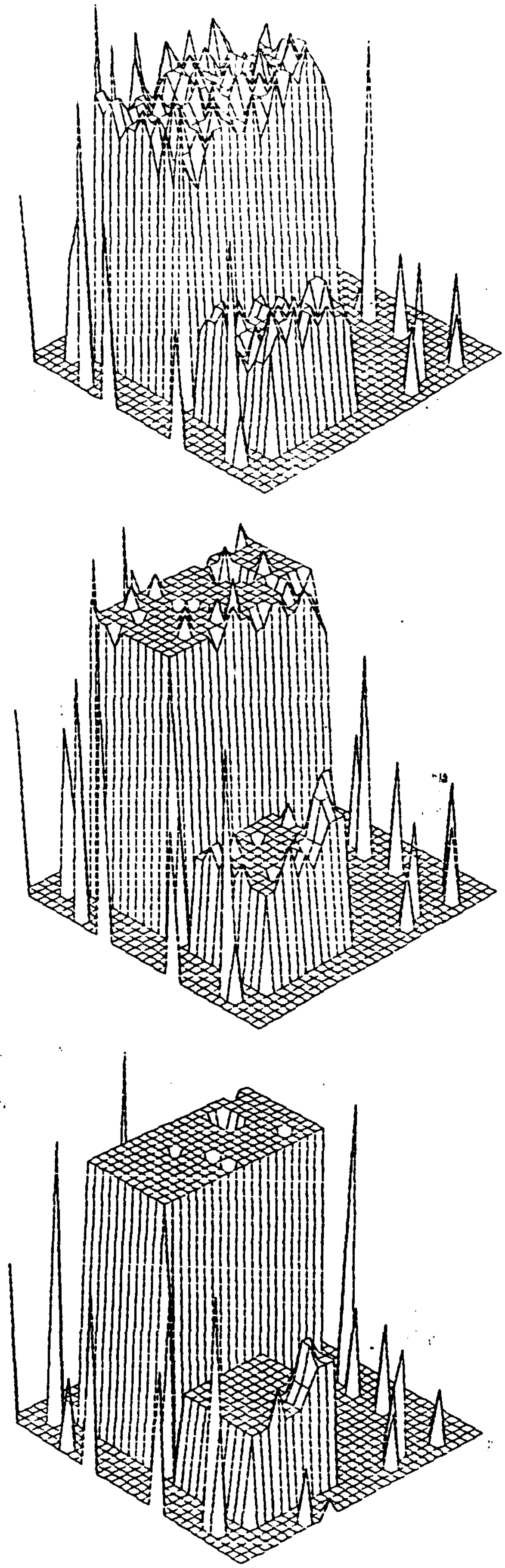

Figure 5. Disparity solution resulting from intensity matching incorporating optical flow data shown at three different tempratures, $T = 50, 30, 1$.