NAIST-IS-TR94001
ISSN 0919-9527

Surface Reconstruction from Stereo Images Preserving Discontinuities[1]

Naokazu Yokoya[2]

January 1994

ABSTRACT

This paper describes a method of surface reconstruction from binocular stereo images. The problem addressed is formulated as the minimization of an energy functional that integrates a similarity term and a smoothness term realizing a controlled-continuity constraint. The minimization reduces to solving a large system of non-linear equations which is a discretized version of Euler-Lagrange equation. A disparity map, which is a real-valued function of the image coordinates, is obtained by solving the system of equations through scale hierarchy. Initially assuming a flat (constant disparity) surface, the proposed method first finds a local minimum of energy functional at a coarse scale and then proceeds to finer scales by using the previous result as an initial estimate at the current scale. At the finest scale, discontinuities and occluded areas are detected from a computed disparity surface. The proposed method can compute a dense disparity map directly from stereo images preserving discontinuities.

Keywords: computer vision, binocular stereo vision, surface reconstruction, regularization, energy minimization.

[1]This work was supported in part by the Real World Computing Program from the Ministry of International Trade and Industry and also by the Foundation for Nara Advanced Institute of Science and Technology.

[2]Information Technology Center, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-01, Japan
E-Mail: yokoya@itc.aist-nara.ac.jp

1 Introduction

In computer vision, it is important to obtain 3D information of a visible surface, because some problems in 3D object description and recognition can be solved easily. Approaches to the problem of 3D data acquisition can be classified into two categories: *acitive* and *passive* ranging methods. Binocular stereo [2], which is a typical among passive methods, has the advantage that absolute distance can be obtained without affecting the scene. However, conventional stereo methods usually suffer from the difficulty in solving the so-called *correspondence problem*. Many approaches to this problem have appeared in the literature: e.g., correlation-based matching [5]; segment-based matching [7]; dynamic programming [8]. In addition to the difficulty of the correspondence problem, most existing methods cannot obtain a dense depth map. Depth information can only be obtained sparsely at feature points or segments, so that interpolation techniques such as surface reconstruction [4, 12] are needed to reconstruct a surface.

Recently, some researchers have attempted to obtain a dense disparity map directly from stereo images by formulating the stereo correspondence problem as the minimization of global measures in the context of *regularization* [9]. For example, March [6] developed a regularization-based approach using an energy functional which involves the controlled-continuity constraint [11], and applied his method to synthetic data. Barnard [1] regarded the pixel-to-pixel stereo matching as a combinatorial optimization problem and employed a stochastic algorithm based on simulated annealing, where discontinuities were not treated explicitly.

This paper describes a method of surface reconstruction from binocular stereo images. The method integrates the stereo matching and the surface reconstruction. The problem addressed is formulated as the minimization of an energy functional that combines a similarity term and a smoothness term. The energy functional involves the controlled-continuity constraint [11] and a visibility function specifying occlusions. The minimization of the energy functional reduces to solving a large system of non-linear equations which is a discretized version of the Euler-Lagrange equation for each pixel. A disparity map is computed by solving the system of equations using a coarse-to-fine approach. Initially assuming a flat (constant disparity) surface, the proposed method first finds a local minimum of the energy functional at a coarse scale and then proceeds to finer scales by using the previous result as an initial estimate at the current scale. A multistage reconstruction algorithm is then carried out at the finest scale detecting occlusions and discontinuities. Our method can directly compute a dense disparity map of real values preserving discontinuities. The proposed method has been successfully applied to both synthetic and real stereo images.

2 Regularization of Stereo Correspondence Problem

2.1 Formulation of Energy Minimization Problem

We assume that stereo images are obtained by two identical TV cameras with parallel optical axes and thus the epipolar lines are parallel to the x axis. The left and right images are denoted by $L(x, y)$ and $R(x, y)$, respectively. The disparity is referred to as $d(x, y)$ at a point (x, y) in the right image.

The problem here is to determine the disparity function $d(x, y)$ which is assumed to be real-valued. It is difficult to determine the function uniquely only by considering the similarity criterion and thus the problem is *ill-posed* [9]. In order to overcome this difficulty, we adopt a framework of *regularization* [9, 10]. The stereo correspondence problem can be formulated as that of finding a function $d(x, y)$ which minimizes

$$E(d) = P(d) + \lambda S(d), \tag{1}$$

where $P(d)$ is a *penalty functional* which measures the difference in image features between corresponding points, and $S(d)$ is a *stabilizing functional* by which a *smoothness constraint* is imposed upon the solution. The *regularization parameter* λ controls the strength of the smoothness constraint.

We assume that a set of n feature images are available for the left and right images and are represented by $L(x, y) = \{L_1(x, y), ..., L_n(x, y)\}$ and $R(x, y) = \{R_1(x, y), ..., R_n(x, y)\}$, respectively. The components $L_p(x, y)$ and $R_p(x, y)$ $(p = 1 \sim n)$ are the intensity arrays of a stereo pair, and their simple transformations such as derivatives. We define the penalty and stabilizing functionals as follows:

$$P(d) = \int\int \alpha(x, y)\Big\{\sum_{p=1}^{n} w_p \left\{L_p(x + d(x, y), y) - R_p(x, y)\right\}^2\Big\}dxdy, \tag{2}$$

$$S(d) = \int\int \beta(x, y)\{d_x(x, y)^2 + d_y(x, y)^2\}dxdy. \tag{3}$$

In Eq. (2), w_p is a weight for the p-th feature image. Subscripts on d in Eq. (3) denote partial derivatives. Note that the *membrane spline* is employed to embed the smoothness constraint above. We use the *continuity control function* [11] and introduce a visibility function. The *visibility function* $\alpha(x, y)$ is a binary function which specifies the visibility in the left image for each right image point (x, y), and is defined as:

$$\alpha(x, y) = \begin{cases} 1; & \text{if } (x, y) \text{ is visible in the left image} \\ 0; & \text{otherwise (occluded area)} \end{cases}. \tag{4}$$

This function is intended to prevent the inclusion of the similarity term from the total energy calculation at occluded points where right image points have no corresponding left image

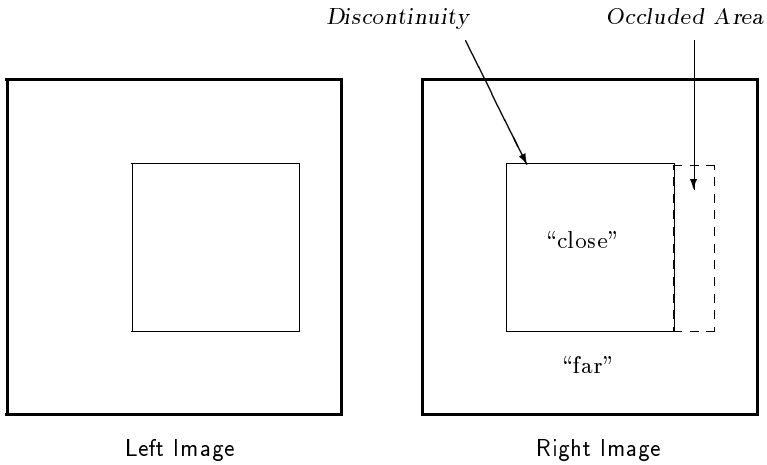


Figure 1: Discontinuities and occlusions in a stereo pair.

points (see Fig. 1). The function $\beta(x, y)$, which locally controls the continuity of the disparity surface, is a binary version of the continuity control function [11] defined as follows:

$$\beta(x,y) = \begin{cases} 1; & \text{if } (x,y) \text{ is a continuous point} \\ 0; & \text{otherwise (discontinuity)} \end{cases} . \tag{5}$$

2.2 Numerical Solution of Energy Minimization Problem

The *calculus of variations* tells us that for continuous surface areas ($\beta(x, y) = 1$) the necessary condition of the minimization of Eq. (1) is the following Euler-Lagrange equation [13]:

$$\begin{aligned} &\alpha(x,y)\Big\{\sum_{p=1}^{n} w_p\{L_p(x+d(x,y),y) - R_p(x,y)\} \cdot L_{p_x}(x+d(x,y),y)\Big\} \\ &-\lambda\left\{d_{xx}(x,y) + d_{yy}(x,y)\right\} = 0, \end{aligned} \tag{6}$$

where L_{p_x} denotes the partial derivative of the p-th left feature image L_p with respect to the x coordinate.

By discretizing Eq. (6) at each pixel (i, j), we obtain a Gauss-Seidel relaxation method:

$$\begin{aligned} d^{(k)}(i,j) \;=\; &\frac{1}{4}\Big[\{d^{(k-1)}(i+1,j) + d^{(k)}(i-1,j) + d^{(k-1)}(i,j+1) + d^{(k)}(i,j-1)\}\Big] \\ &-\frac{t^2}{\lambda}\alpha(i,j)\Big\{\sum_{p=1}^{n} w_p\{L_p(i+d^{(k-1)}(i,j),j) - R_p(i,j)\} \cdot L_{p_x}(i+d^{(k-1)}(i,j),j)\Big\}\Big], \end{aligned} \tag{7}$$

where t represents the distance between two adjacent pixels and the superscript k denotes the iteration step.

Since $d(i, j)$ is assumed to be a real-valued, a point $(i + d(i, j), j)$ in the left image does not necessarily correspond to an integer grid, and hence we have to calculate image feature

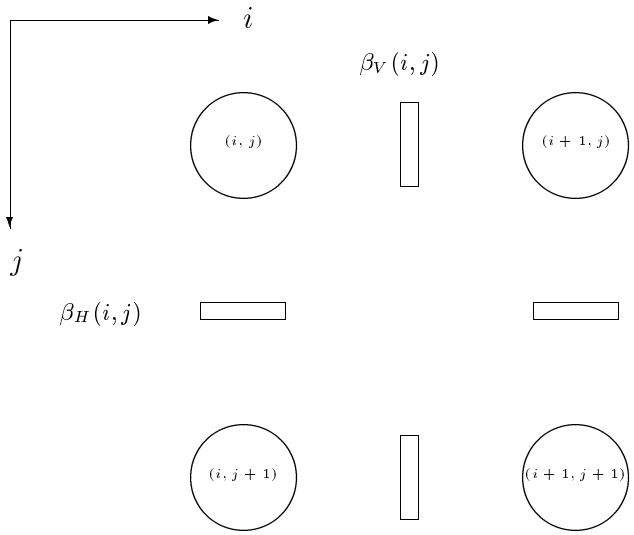


Figure 2: Edge element configuration.

values at subpixel locations. We estimate image feature values at subpixel locations by fitting cubic B-spline curves [3] to the feature images along the epipolar line (x axis).

In order to implement the controlled-continuity stabilizer numerically, we modify the continuity control function using a notion of *edge element* (see Fig. 2). Horizontal and vertical edge specifiers are denoted by β_H and β_V, respectively, and have values as follows:

$$\beta_H(i,j) = \begin{cases} 1; & \text{no jump edge between } (i,j) \text{ and } (i,j+1) \\ 0; & \text{otherwise (horizontal discontinuity)} \end{cases}, \tag{8}$$

$$\beta_V(i,j) = \begin{cases} 1; & \text{no jump edge between } (i,j) \text{ and } (i+1,j) \\ 0; & \text{otherwise (vertical discontinuity)} \end{cases}. \tag{9}$$

Using the edge specifiers $\beta_H(i,j)$ and $\beta_V(i,j)$ above, we rewrite Eq. (7) as follows:

$$\begin{aligned} d^{(k)}(i,j) \;=\; & \frac{1}{\beta^*}\Big\{\beta_V(i,j)d^{(k-1)}(i+1,j) + \beta_V(i-1,j)d^{(k)}(i-1,j) \\ & +\beta_H(i,j)d^{(k-1)}(i,j+1) + \beta_H(i,j-1)d^{(k)}(i,j-1)\Big\} \\ & -\frac{t^2}{4\lambda}\alpha(i,j)\Big\{\sum_{p=1}^{n} w_p\Big\{L_p(i+d^{(k-1)}(i,j),j) - R_p(i,j)\Big\}\cdot L_{p_x}(i+d^{(k-1)}(i,j),j)\Big\}, \end{aligned} \tag{10}$$

where

$$\beta^* = \beta_V(i,j) + \beta_V(i-1,j) + \beta_H(i,j) + \beta_H(i,j-1).$$

By using the iterative formula in Eq. (10) involving prespecified functions $\alpha(i,j)$, $\beta_H(i,j)$ and $\beta_V(i,j)$, we can estimate the disparity function $d(i,j)$ from a certain initial estimate. However we encounter a difficulty in this iterative computation; that is, we cannot expect to

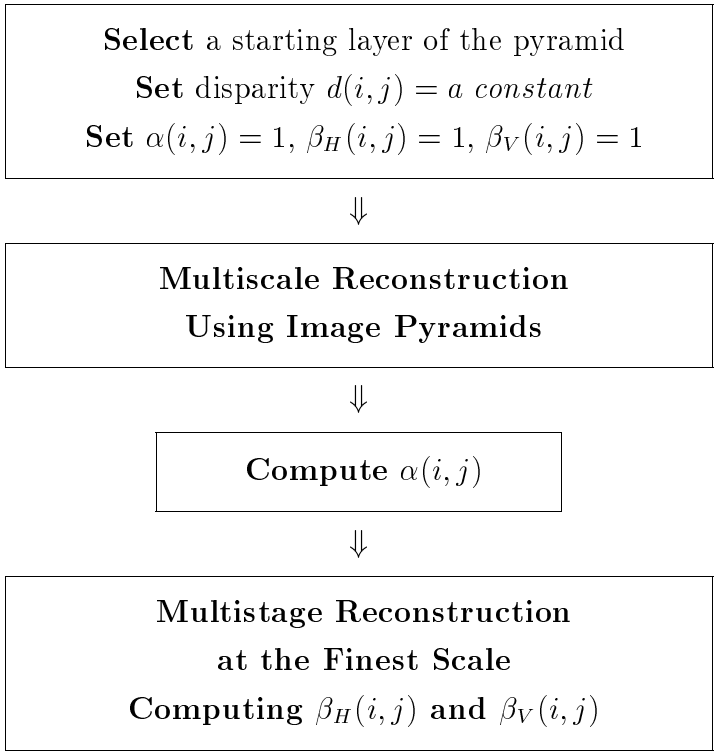


Figure 3: The overall control flow of the multiscale-multistage stereo surface reconstruction algorithm.

obtain a global minimum of the energy in Eq. (1) from an arbitrarily selected initial estimate of disparity map, because Eq. (1) generally have a lot of local minima. In the next section, we propose a coarse-to-fine approach which can relax the difficulty. Also described will be practical methods for determining $\alpha(i,j)$, $\beta_H(i,j)$ and $\beta_V(i,j)$.

3 Iterative Estimation of Disparity Map over Scale

3.1 Multiscale-Multistage Stereo Surface Reconstruction

In order to avoid meaningless local minima of Eq. (1), We employ a multiscale algorithm using stereo pairs of pyramidal feature images. The pyramid consists of $L+1$ layers of dimension $2^{L-l} \times 2^{L-l}$ stacked one upon the other, with the finest image of dimension $2^L \times 2^L$ at the bottom (level 0) and the $2^0 \times 2^0$ single point image at the apex (level L). Thus each pixel on level l has its four predecessors on level $l-1$. In our implementation, the level l of the pyramid is the Gaussian filtered version of an original feature image with the characteristic size $\sigma = 2^l$.

The control flow of the algorithm is show in Fig 3. A stereo surface reconstruction algorithm goes down through the pyramids from a starting (coarsest) layer to the bottom (finest) layer, when the algorithm computes a disparity map using Eq. (10) at each scale without

considering discontinuities and occlusions. In addition to avoiding the local minima, such a multiscale technique will significantly improve efficiency of the relaxation method. After completing the reconstruction at the finest scale, the visibility function is computed from the disparity function obtained. A multistage reconstruction algorithm is then performed at the finest scale preserving discontinuities. The multiscale-multistage stereo reconstruction is described in more detail in the following steps:

(1) [Initialization]

First, we select a starting layer of the pyramids, and then set the disparity function, visibility function, and the continuity control functions: $d(i,j) = constant$, $\alpha(i,j) = 1$, and $\beta_H(i,j) = \beta_V(i,j) = 1$ for all the image points.

(2) [Multiscale Reconstruction]

(2-a) At the start of each scale, the disparity function obtained at the previous scale is projected down to the current scale.

(2-b) The stereo surface reconstruction algorithm then computes the smoothest disparity surface $d(i,j)$, using Eq. (10).

(2-c) The computation at (2-a) and (2-b) is iterated until the finest scale is reached.

(3) [Detection of Occlusions]

After the completion of the multiscale reconstruction, the visibility function $\alpha(i,j)$ is updated from $d(i,j)$ obtained at the finest scale.

(4) [Multistage Reconstruction at the Finest Scale]

(4-a) At the start of each stage, the continuity control functions $\beta_H(i,j)$ and $\beta_V(i,j)$ are updated according to the disparity function obtained at the previous stage.

(4-b) Then the algorithm improves the disparity map $d(i,j)$ using Eq. (10), treating $\alpha(i,j)$, $\beta_H(i,j)$, and $\beta_V(i,j)$ as fixed functions.

(4-c) The computation at (4-a) and (4-b) is repeated until $\beta_H(i,j)$ and $\beta_V(i,j)$ converge.

In step (1), a starting layer of the pyramid must be determined according to a range of disparity in the input stereo pair; we suggest that level l_0 is selected such that $2^{l_0} \geq$ *maximum disparity* as the starting layer when the zero disparity surface is used as an initial estimate. It should be noted in step (2-b) that the distance t between two adjacent pixels included in Eq. (10) is set to 2^l at level l of the pyramid.

3.2 Detection of Occlusions and Discontinuities

(1) Detection of Occluded Areas

At occluded areas where right image points have no corresponding left image points, the first partial derivatives of the disparity function in the direction of the epipolar line are

expected to be negative and have large magnitudes of their values. Thus we set the visibility function $\alpha(i,j) = 0$ at the points where the following condition is satisfied:

$$\frac{1}{t}\{d(i,j) - d(i-1,j)\} < C_\alpha, \tag{11}$$

where C_α is a negative constant (usually, -0.5).

(2) Localization of Discontinuities

Disparity discontinuities should be localized between two adjacent pixels as shown in Fig. 2. At discontinuities in depth, the gradient of the disparity function will be high and have a local maximum of its magnitude in the direction perpendicular to the edge element. Thus we set the continuity control function $\beta_H(i,j) = 0$ at the points where the following two conditions are satisfied:

$$\frac{1}{t}|d(i,j+1) - d(i,j)| > C_\beta, \tag{12}$$

$$\begin{aligned} |d(i,j+1) - d(i,j)| > \max\{&|d(i,j+2) - d(i,j+1)|, \\ &|d(i,j) - d(i,j-1)|\}. \end{aligned} \tag{13}$$

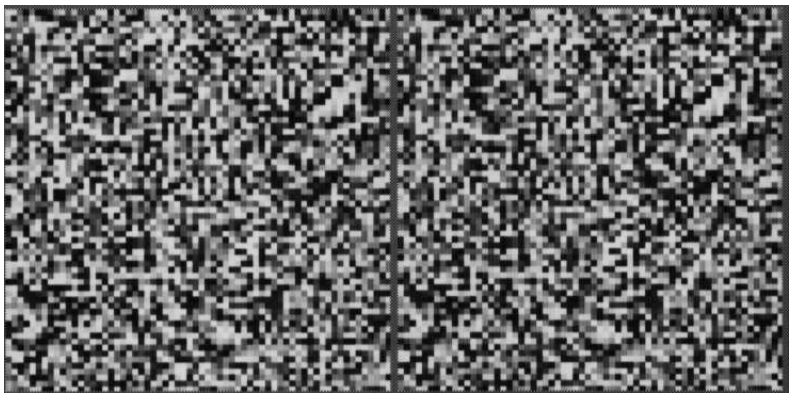
The function $\beta_V(i,j)$ is also determined according to the similar criteria. In Eq. (12), C_β is a threshold for the magnitude of the first partial derivative of disparity in the direction perpendicular to the edge elements.

4 Experimental Results and Discussions

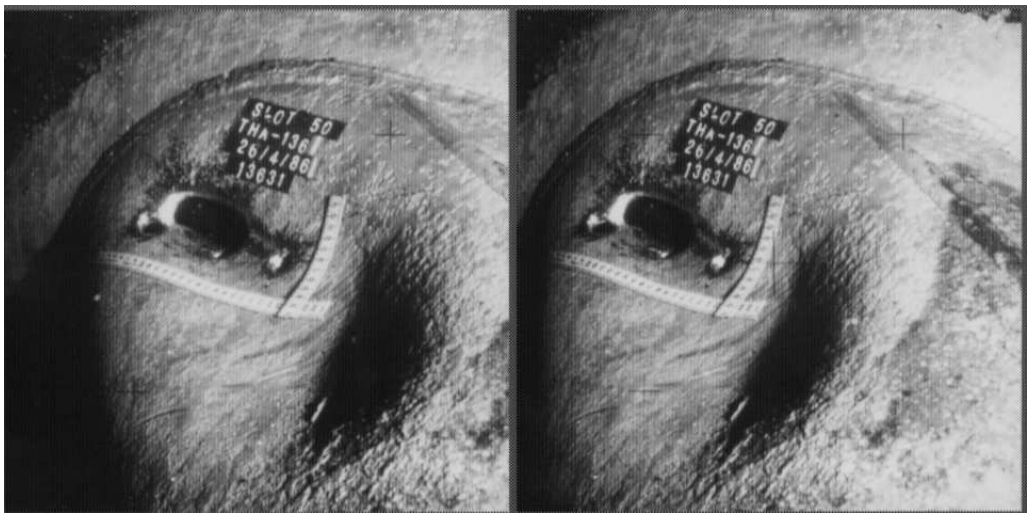
We have experimented with both synthetic and real stereo images. At present, we employ an intensity image and its partial derivatives in the direction of the epipolar line (x axis) as feature images. Figure 4(a) is a 256×256 synthetic random-texture stereogram where four concentric square regions in the right image have disparities of 4, 8, 12, and 16 $pixels$, respectively. A three-dimensional relief of the disparity surface is shown in Fig. 5. Figure 4(b) shows a pair of real stereo images with the size of 336×336, which are acquired by digitizing photographic prints from color transparencies obtained with an underwater 35mm stereo camera. In the experimental results given below, level 4 of the pyramid was selected as the starting layer of the algorithm. The termination of the reconstruction algorithm at each scale/stage was decided according to the decreasing rate of total energy changes in consecutive iteration steps:

$$\frac{E_{k-1} - E_k}{E_k} < 0.0001 \tag{14}$$

Figure 6 shows a sequence of reconstructed disparity surfaces at five scales and the final result for TEXTURE, obtained by starting from a flat zero disparity surface. Figure 7



(a) Random Texture (TEXTURE)



(b) Damaged Underwater Oil Platform (UNDERWATER)

Figure 4: Examples of stereo images.

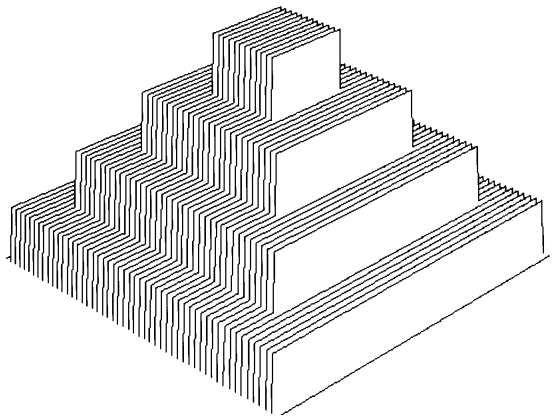
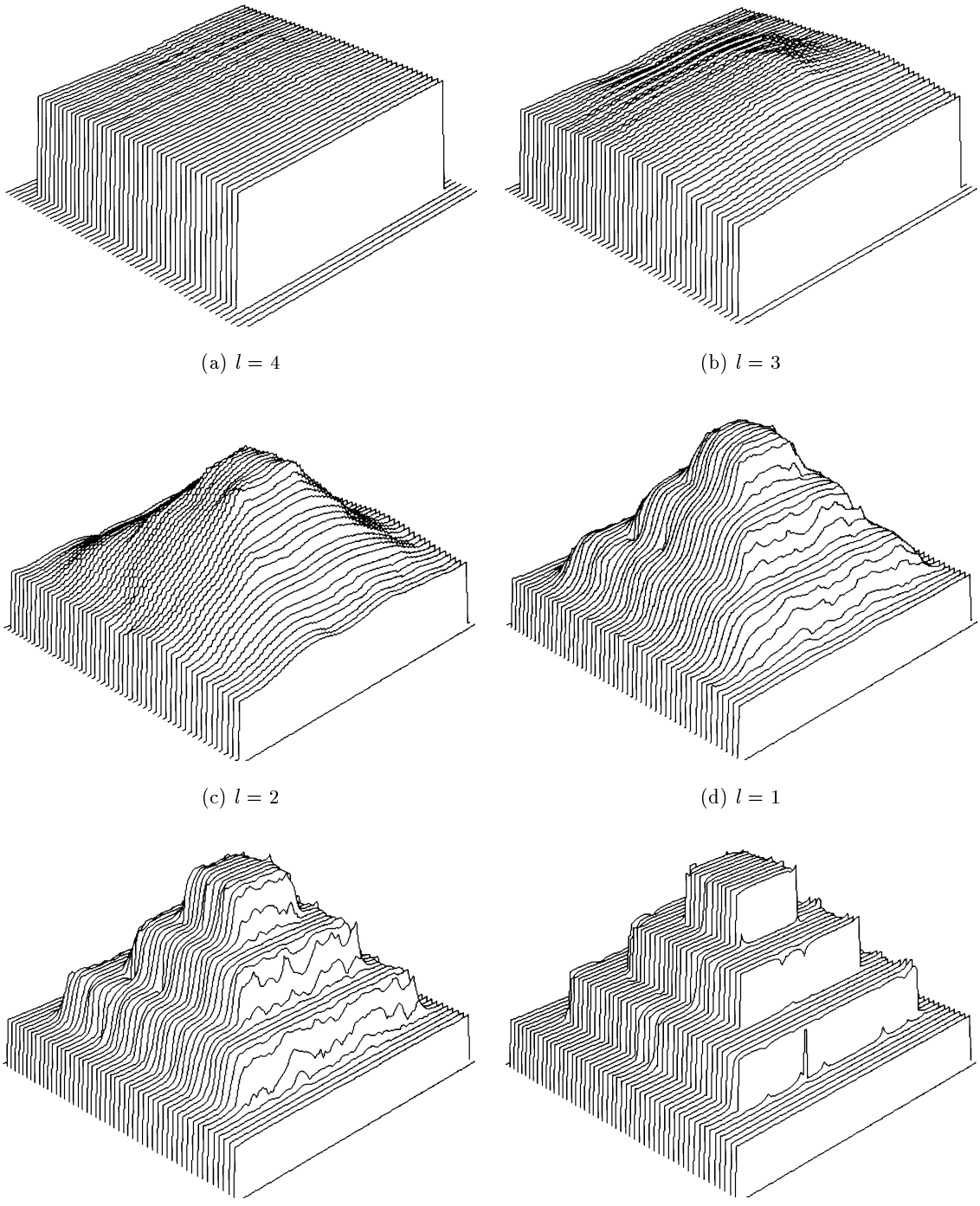


Figure 5: Disparity surface of TEXTURE.

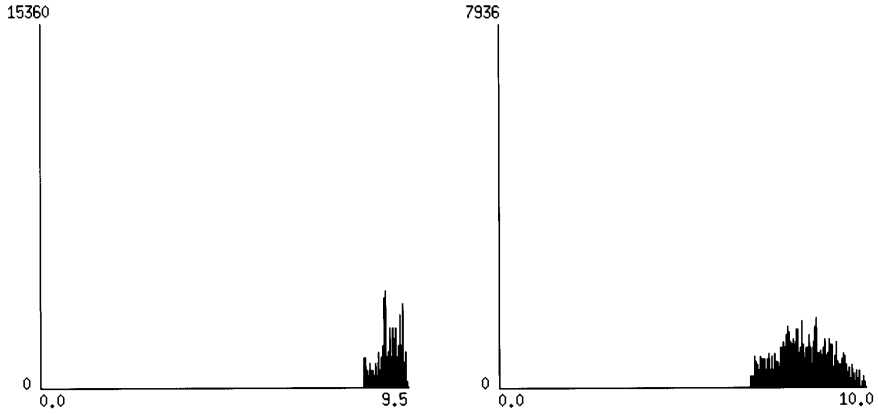


(a) $l = 4$ (b) $l = 3$

(c) $l = 2$ (d) $l = 1$

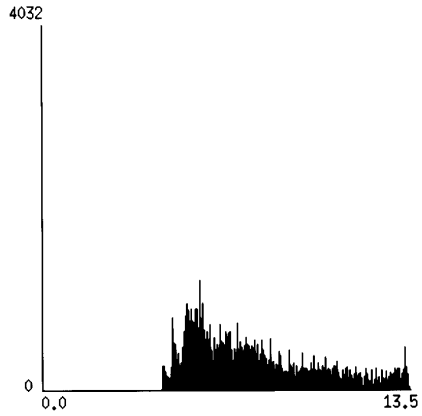
(e) $l = 0$ (f) Final Result

Figure 6: Disparity surface of TEXTURE obtained at each stage of the algorithm.

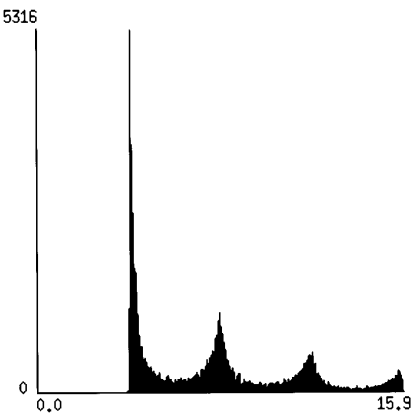


(a) $l = 4$

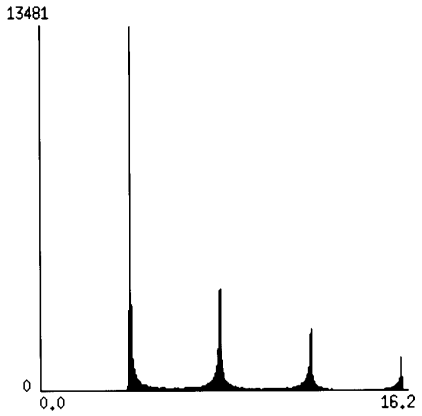
(b) $l = 3$



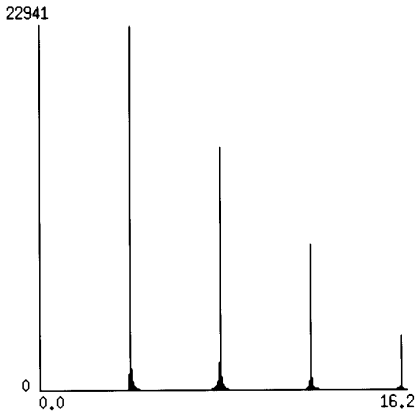
(c) $l = 2$



(d) $l = 1$



(e) $l = 0$



(f) Final Result

Figure 7: Histograms of disparities for TEXTURE computed at each stage of the algorithm.

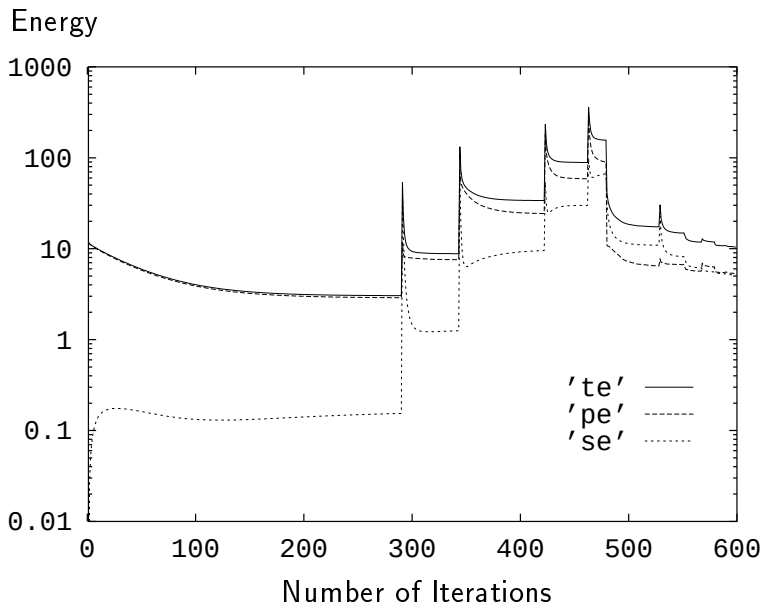


Figure 8: Change of energy in iterative disparity estimation for TEXTURE; 'te': total energy (E), 'pe': penalty term (P), 'se': smoothness term (λS).

illustrates histograms of estimated disparities at each scale/stage. In this experiment, we used only intensities as image features, and the regularization parameter λ was set to 1000. Note that the vertical scale is magnified to emphasize the relief. It can be observed that the obtained disparity surfaces in Figs. 6(a) - (e) are gradually converging to a correct disparity surface going down through scale. However, the result (e) at the final scale is deformed in the neighborhood of discontinuities. Figure 6(f) is the final result of the multistage reconstruction with $C_\alpha = -0.5$ and $C_\beta = 0.4$. Comparing Fig. 6(f) with 6(e), it is obvious that the proposed method has excellently estimated the disparities preserving discontinuities. Figure 7 also shows that the estimated disparity surfaces are gradually converging to a correct one through scale. Figure 8 illustrates the change of energy in iterative disparity estimation for TEXTURE. It should be noted that in Fig. 8 discontinuities in energy correspond to the change of scale. The calculation time was about 2 minutes for the multiscale reconstruction and 5 minutes for the multistage reconstruction at the finest scale on a UNIX workstation (TITAN 3000). Table 1 shows the mean and variance of absolute errors of estimated disparities from known values at each scale and at the final stage for five different values of the parameter λ; λ = 300, 500, 1000, 2000, and 3000. This also depicts the effectiveness of automatically computing the visibility and continuity control functions. The experiment with TEXTURE data has also shown that disparity estimation errors are distributed in the neighborhood of discontinuities (see Fig. 9).

Figure 10 illustrates a reconstructed disparity surface of UNDERWATER. It can be seen

Table 1: Absolute errors of estimated disparities for TEXTURE with five different values of the regularization parameter λ.

Parameter λ \ Scale		Scale					Final
		$l = 4$	$l = 3$	$l = 2$	$l = 1$	$l = 0$	Result
$\lambda = 300$	mean	3.258	2.349	0.948	0.489	0.400	0.216
	variance	3.511	2.463	1.072	0.829	0.709	0.335
$\lambda = 500$	mean	3.253	2.483	1.035	0.455	0.309	0.175
	variance	3.382	2.608	0.965	0.697	0.499	0.290
$\lambda = 1000$	mean	3.247	2.648	1.197	0.485	0.283	0.125
	variance	3.190	2.794	0.882	0.619	0.397	0.322
$\lambda = 2000$	mean	3.226	2.768	1.412	0.590	0.310	0.117
	variance	3.194	3.107	0.943	0.580	0.365	0.291
$\lambda = 3000$	mean	3.200	2.804	1.574	0.662	0.342	0.107
	variance	3.320	3.357	1.091	0.562	0.347	0.272

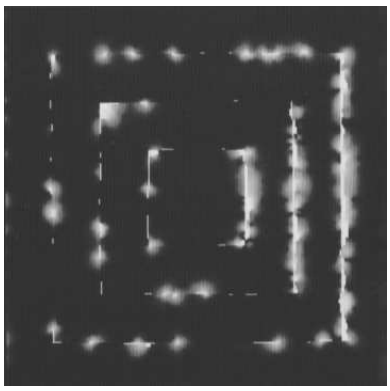


Figure 9: Distribution of disparity estimation errors for TEXTURE.

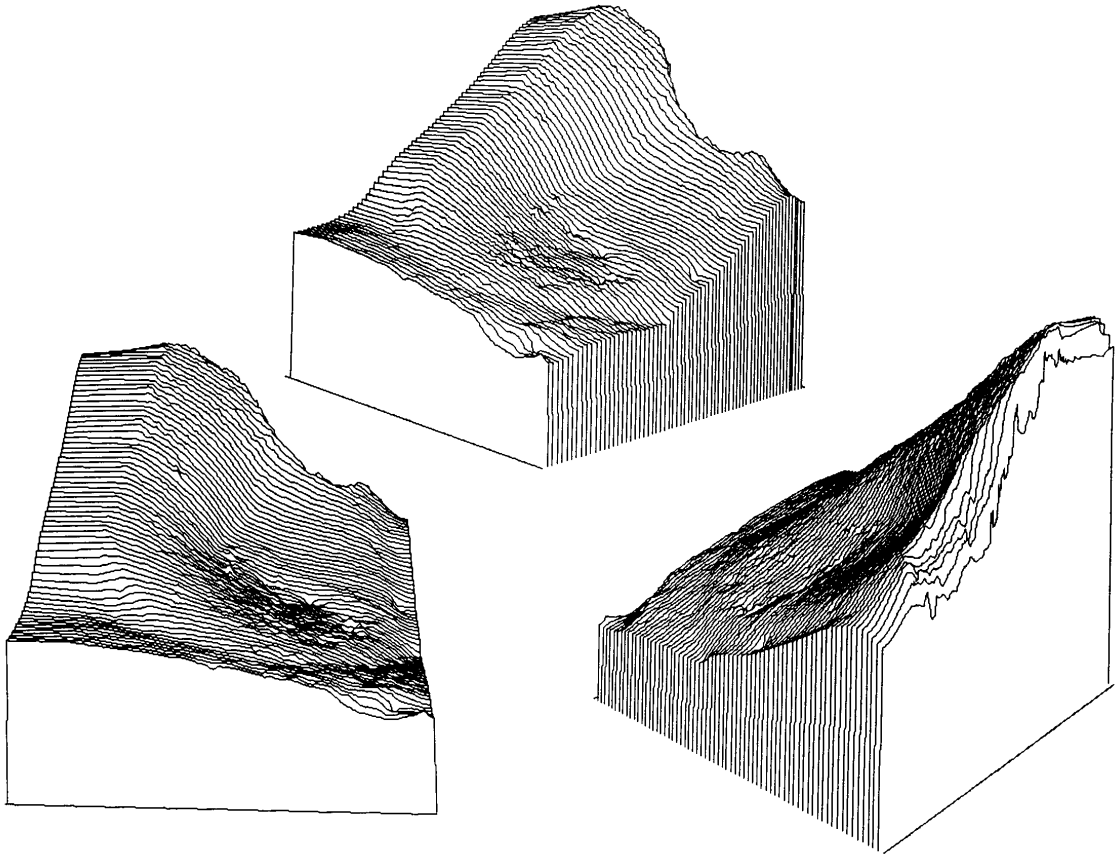


Figure 10: Reconstructed disparity surface of UNDERWATER viewed from three different positions.

that the algorithm has nicely computed a disparity surface from the real stereo data. In this example, we used intensities and their first and second partial derivatives in the direction of the epipolar line (x axis) as image features; the weighting parameter $w_1 = 1$ for intensities, $w_2 = 5$ for first derivatives, $w_3 = 10$ for second derivatives. The regularization parameter λ was set to 1000. The thresholds C_α and C_β for detecting occlusions and discontinuities were set to the same values as those used in Fig. 6.

The experiments including the results given above indicated that the proposed surface reconstruction method can compute dense disparity maps excellently from stereo data with small disparities as well as those with large disparities. The experiment with synthetic data indicated that errors of estimated disparities are concentrated in the neighborhood of sharp disparity discontinuities. These errors are mainly caused by errors in localizing discontinuities. Discontinuity detection should be further investigated. In the experiments, the parameter λ was fixed during the multiscale reconstruction process. This parameter may be adaptively decided so as to have a small value at the coarsest scale and have larger values at finer scales, because the Gaussian filter increasingly reduces noise at higher levels of the pyramid.

5 Conclusions

We have described a new method of reconstructing a disparity surface from binocular stereo images and have presented experimental results. The stereo correspondence problem is formulated as that of energy minimization by means of the regularization principle. The energy functional defined in this paper integrates both the similarity and smoothness terms including the visibility and continuity control functions. By determining these functions in the process of reconstruction, disparity discontinuities are preserved and hence the accuracy of an estimated surface is improved. A coarse-to-fine approarch using Gaussian pyramids is employed for solving the minimization problem avoiding local minima of the functional. The proposed method has the following advantages:

(1) Our method can obtain a dense disparity map directly from stereo images. Moreover, disparity values are computed to subpixel accuracy. This implies that the method works well for stereo data with small disparities as well as those with large disparities.

(2) Depth discontinuities and occluded areas are automatically detected in the process of reconstruction. Using these cues, the disparity surface is reconstructed preserving discontinuities.

(3) The method could be modified for suitably implementing it in massively parallel architectures.

We are now considering the possibilities of additional invariant image features; for example, color information. The proposed stereo surface reconstruction method can be easily modified to relax the eipolar constraint that the epipolar lines strictly coincide with the scan lines by allowing disparities in the y direction. This is practically important because the stereo camera system usually causes some vertical disparity even when the camera is carefully set up.

Acknowledgements

This work was partly carried out at the Electrotechnical Laboratory (ETL), the Ministry of International Trade and Industry. The author thanks Nobuyuki Otsu, Kazuhiko Yamamoto, and Shigeru Muraki of ETL for their support and assistance during this work. We are most grateful to Richard Baldwin of the Polytechnic of East London for providing us with stereo data UNDERWATER.

References

[1] S.T. Barnard: "Stochastic stereo matching over scale", *Int. Journal Computer Vision*, Vol.3, No.1, pp.17-32, 1989.

[2] S.T. Barnard and M.A. Fischler: "Computational stereo", *ACM Computing Surveys*, Vol.14, No.4, pp.553-572, 1982.

[3] B.A. Barsky: "A description and evaluation of various 3D models", *IEEE Computer Graphics and Applications*, Vol.4, No.1, pp.38-52, Jan. 1984.

[4] A. Blake and A. Zisserman: *Visual Reconstruction*, The MIT Press, Cambridge, 1987.

[5] M.D. Levive, D.A. O'Handley, and G.M. Yagi: "Computer determination of depth maps", *Computer Graphics and Image Processing*, Vol.2, No.2, pp.131-150, Oct. 1973.

[6] R. March: "A regularization model for stereo vision with controlled continuity", *Pattern Recognition Letters*, Vol.10, No.4, pp.259-263, Oct. 1989.

[7] G. Medioni and R. Nevatia: "Segment-based stereo matching", *Computer Vision, Graphics and Image Processing*, Vol.31, No.1, pp.2-18, July 1985.

[8] Y. Ohta and T. Kanade: "Stereo by intra- and inter-scanline search using dynamic programming", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.PAMI-7, No.2, pp.139-154, March 1985.

[9] T. Poggio, V. Torre, and C. Koch: "Computational vision and regularization theory", *Nature*, Vol.317, No.6035, pp.314-319, Sep. 1985.

[10] K. Sakaue and N. Yokoya: "Relaxation and regularization", *Journal of IPS Japan*, Vol.30, No.9, pp.1047-1057, Sep. 1989. [in Japanese].

[11] D. Terzopoulos: "Regularization of inverse visual problems involving discontinuities", *IEEE Trans. Pattern Anal. Mach Intell.*, Vol.PAMI-8, No.4, pp.413-424, July 1986.

[12] D. Terzopoulos: "The computation of visible-surface representations", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol.10, No.4, pp.417-438, July 1988.

[13] N. Yokoya: "Surface reconstruction directly from binocular stereo images by multiscale-multistage regularization", *Proc. 11th Int. Conf. Pattern Recognition*, Vol.I, pp.642-646, The Hague, Aug. 1992.

[14] A.P. Witkin, D. Terzopoulos, and M. Kass: "Signal matching through scale space", *Int. Journal Computer Vision*, Vol.1, No.2, pp.133-144, 1987.