Journal of Electronic Imaging 8(1), 92–97 (January 1999).

Visualization of the three-dimensional topography of the optic nerve head through a passive stereo vision model

Juan Manuel Ramirez
Universidad de las Américas
Department of Electrical Engineering
Cholula Puebla 72820, Mexico
E-mail: mramirez@udlapvms.pue.udlap.mx

Sunanda Mitra
Texas Tech University
Department of Electrical Engineering
Computer Vision and Image Analysis Laboratory
Lubbock, Texas 79400

Jose Morales
Texas Tech University Health Sciences Center
Department of Ophthalmology and Visual Sciences
Lubbock, Texas 79409

Abstract. *This paper describes a system for surface recovery and visualization of the three-dimensional topography of the optic nerve head, as support of early diagnosis and follow up of glaucoma. In stereo vision, depth information is obtained from triangulation of corresponding points in a pair of stereo images. In this paper, the use of the cepstrum transformation as a disparity measurement technique between corresponding windows of different block sizes is described. This measurement process is embedded within a coarse-to-fine depth-from-stereo algorithm, providing an initial range map with the depth information encoded as gray levels. These sparse depth data are processed through a cubic B-spline interpolation technique in order to obtain a smoother representation. This methodology is being especially refined to be used with medical images for clinical evaluation of some eye diseases such as open angle glaucoma, and is currently under testing for clinical evaluation and analysis of reproducibility and accuracy.* © *1999 SPIE and IS&T.*
[S1017-9909(99)01101-0]

1 Introduction

A physical consequence of elevated intraocular pressure in an eye afflicted with glaucoma is progressive atrophy of the optic nerve head. In addition to some other cues involving loss of the visual field and pain under pressure, change in cupping of the optic disk represents a valuable indicator for the ophthalmologist to diagnose and monitor the disease.[1,2] The emphasis of this work is surface recovery from sparse two-dimensional (2-D) stereo data. Stereo vision allows the recovery of depth information of three-dimensional (3-D) scenes by triangulation of corresponding points in a pair of stereoscopic images. Resembling the human visual system, two cameras sense the three-dimensional scene from different perspectives, providing a pair of stereo images which can provide the range information. Many different stereo systems for range determination using binocular stereo have been developed. The different approaches can be compared by considering their camera modeling, feature acquisition, and matching techniques. Image matching is clearly dependent on the choice of feature primitives. When the elements to be matched are low level and dense, such as the image intensity at each pixel in a neighborhood, the matching strategy is called an area-based process, while for sparse and usually more abstract and high-level features, such as edges or zero crossings, the process is referred as feature-based. Area-based techniques have been used by Hanna[3] and Cochran and Medioni[4] using cross correlation for texture matching in the image pair to obtain an initial disparity estimate. Matching more abstract features are less sensitive to noise, although some interpolation should be included since only sparse data are obtained. Marr and Pogio proposed a computational model of the human stereo vision system using zero crossings from the Laplacian as the matching feature at different filter sizes.[5] The disparity arrays created for each filter are combined into a single disparity description. Grimson[6] implemented an improved version of this model with good results when there is a dense set of features. He included variational methods for surface reconstruction from sparse data. Otha and Kanade[7] explored some alternatives of dynamic programming for edge matching using interscanline search for finding an op-

Paper BIR-03 received Aug. 3, 1998; accepted for publication Sept. 17, 1998.
1017-9909/99/$10.00 © 1999 SPIE and IS&T.

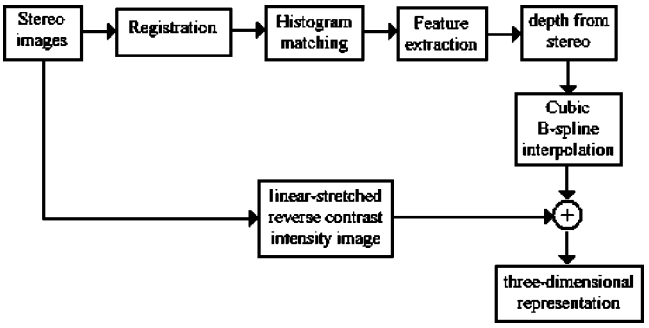

Fig. 1 Block diagram of the implemented method.

timal matching surface. Hoff and Ahuja[8] proposed the combination of feature matching, contour detection, and surface interpolation in one process. Weng[9] explored the use of windowed Fourier phase profiles as the major matching primitive, and Jepson and Jenkin[10] described some methods based on the output phase behavior of bandpass Gabor filters. However, phase information as a primitive to match has not offered, at this point, a better solution for false target correspondence, featureless areas, noise sensitivity, or occlusion problems. A stereo-motion model with a matching technique based on cepstrum was developed to be used with a sequence of nine images with small disparities between consecutive images.[11] The cepstrum transform was then proven to be very noise tolerant and accurate in the matching procedure. An increase in the base line while decreasing the size of the matching window provided the fine resolution in the disparity map corresponding to the original sequence of nine images. In this work, this concept is extended to be used with only one pair of stereo images, which is the usual in the medical images that were considered. An interpolation scheme was incorporated into the system[12,13] with the ultimate goal of obtaining a complete system for surface recovery of the smooth surface corresponding to the topography of the optic nerve head.

2 Methodology

The work presented in this paper consists of an improved stereo vision model with several preprocessing techniques and an interpolation scheme, as represented in Fig. 1. The method of registration used in this work is based on spectral and cepstral transform domain manipulations. In the next step, a correction for intensity variations based on histogram matching is performed. Feature extraction using a Sobel operator and further binarization provides the edge information to be used for matching in the stereo algorithm. The sparse depth map is obtained by using the stereo algorithm based on a hierarchical window matching and cepstral analysis. A cubic *B*-spline interpolation provides the smooth range map corresponding to the excavation of the optic nerve head. A pseudointensity image obtained from a linear stretching transformation is incorporated in the last step, in order to provide the location of the blood vessels without adding false depth to the topography map. These operations are further explained in detail.

3 Disparity Detection by Cepstrum Transform

Cepstrum transformation was initially developed in seismic and acoustic signal processing for echo detection. In this work, the cepstrum transformation is used for finding disparity between corresponding areas in every step of the depth-from-stereo algorithm described. The power cepstral transformation is defined as

$$P\{i(x,y)\}=|\mathcal{F}\{\ln|\mathcal{F}\{i(x,y)\}|^2\}|^2, \tag{3.1}$$

where $\mathcal{F}$ is the notation for the Fourier transform and $i(x,y)$ is the given image function. Let us consider a composed image formed by adding the two corresponding blocks of the stereo images, which are supposed to have a small translational difference given by x_0,

$$i(x,y)=w(x,y)+w(x-x_0,y). \tag{3.2}$$

The Fourier transform of $i(x,y)$ is

$$\mathcal{F}\{i(x,y)\}=W(u,v)+W(u,v)e^{-j2\pi ux_0}, \tag{3.3}$$

where u and v represent spatial frequencies and $W(u,v)$ is the Fourier transform of $w(x,y)$; the Power spectrum is obtained as

$$\begin{aligned}|\mathcal{F}\{i(x,y)\}|^2&=|W(u,v)|^2|1+e^{-j2\pi ux_0}|^2\\&=|W(u,v)|^2|1+\cos 2\pi ux_0-j\sin 2\pi ux_0|^2\\&=|W(u,v)|^2((1+\cos 2\pi ux_0)^2\\&\quad+\sin^2 2\pi ux_0)\\&=|W(u,v)|^2(2+2\cos 2\pi ux_0).\end{aligned} \tag{3.4}$$

When the logarithm function is applied to the power spectrum of $i(x,y)$, the multiplicative terms are separated as

$$\ln|\mathcal{F}\{i(x,y)\}|^2=\ln|W(u,v)|^2+\ln(2+2\cos 2\pi ux_0). \tag{3.5}$$

Using the logarithm series expansion, the second term can be expanded into a convergent infinite series, the application of the power spectrum according to the definition of power spectral transformation yields

$$\begin{aligned}P\{i(x,y)\}&=P\{w(x,y)\}+A\delta(x,y)+B\delta(x\pm x_0,y)\\&\quad+C\delta(x\pm 2x_0,y)+\cdots.\end{aligned} \tag{3.6}$$

This derivation considers only horizontal displacements because according to the concept of stereo vision, only horizontal disparities between corresponding blocks are expected. After removing the power cepstrum of $w(x,y)$, the translational difference between the two corresponding windows can be obtained by inspecting the remaining impulse train. In other words, the translational difference corresponds to the distance between the origin and the location of the first peak in the cepstrum plane.

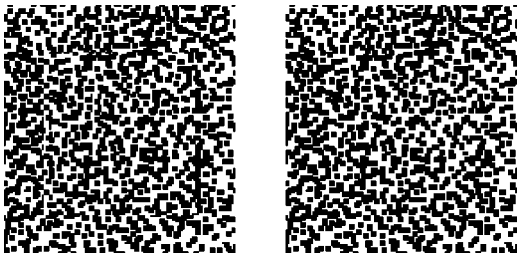

Fig. 2 Random dot stereogram; three-layer cake.

4 Hierarchical Search of Disparities

In a nonconvergent stereo vision model, the depth information is directly derived from the disparity between corresponding points. Random dots stereograms have been widely used to test the performance of stereo vision systems. In a random dot stereogram, the disparities are artificially generated by introducing horizontal displacements in a computer generated image with a pattern of random dots. A stereogram of a three-layer cake is shown in Fig. 2. When a stereoscope is used to analyze these images the three-dimensional effect can be perceived. The goal of any automated depth-from-stereo algorithm is to find the disparities between corresponding points in the stereo pair using some matching technique. Establishing these corresponding points is, however, the most critical step in the estimation of depth from stereo.

The ideal situation is finding disparities between every individual pixel in both images. However, it is obvious that the intensity value of a single pixel is not enough for finding corresponding points, so a collection of pixels in a neighborhood has to be used to match. The approach used in this work is the technique of successive refinement of parallax based on hierarchical coarse-to-fine resolution steps.[10] The algorithm starts with a partition of the stereo images in windows of size 32×32. The two-dimensional cepstrum is used to obtained disparity between corresponding windows in both images. A second partition by two is performed in each window in order to obtain a correction value of the disparity in the corresponding subwindow. This quadrant subdivision for coarse-to-fine search of disparities is represented in Fig. 3.

The coarse disparity assigned initially to the first window is modified in every step to determine the most detailed disparity information. This procedure continues until the size of the window is so small that the matching technique cannot be applied. In the implementation described in this work this point was reached usually at a window size of 4×4 pixels. This accumulation of disparities is represented in Fig. 4.

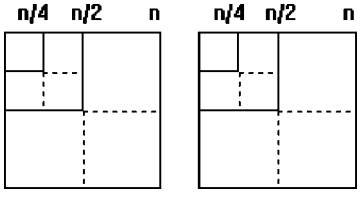

Fig. 3 Quadrant division for coarse-to-fine search.

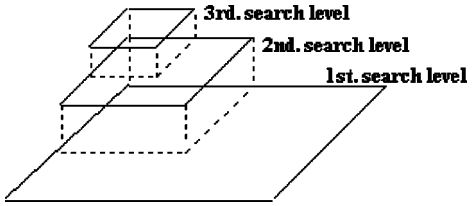

Fig. 4 Schematic representation of the disparity accumulation.

5 Cubic *B*-Spline Interpolation

Early vision is a research field in which many of the problems encountered are inverse ill-posed problems. In many cases no unique solution of the problem exists unless additional constraints are imposed. Surface recovery from sampled data is a typical case in which most passive techniques provide sparse depth information of the surface to be reconstructed. In that situation, constraints like smoothness or allowed discontinuities transform the problem into a well-posed one. Formal analysis of this ill-posed problem uses the Tikhonov regularization method to make the problem well posed. *A priori* knowledge of the type of reconstruction that is desired is used to determine an appropriate stabilizer based on invariant characteristics. In our case, reconstruction of a smooth surface is required, and first- and second-order stabilizers lead to linear and spline approximations, respectively. In the first case, the reconstructed surface will be piecewise planar while the second case will provide a smoother reconstruction. There is at present no automated computational system able to perform the task of dense and precise depth perception as human beings perform this function. Passive methods, and specifically stereo-based systems, rely on the availability of enough information in the scenes to perform the matching, the corresponding disparity detection, and finally, depth extraction from triangulation. Even when random-dot stereograms are used for experimentation, and the information to match has a controlled and uniform density, the obtained depth map is sparse with a density proportional to the minimum size of the dots used to form the stereogram. Those stereograms, however, give a vivid and uniform depth impression, when they are fused stereoscopically by a human being.

It is apparent that in every case, only sparse data of the corresponding surface are available. Invariant surface recovery from sparse data in 3-D space is then required. The basic problem in surface recovery is to compute a complete, dense, and consistent representation of 3-D shapes, when only scattered points on the surface are available as initial input. Formal analysis of this ill-posed problem uses Tikhonov regularization to make the problem well posed. *A priori* knowledge of the type of reconstruction that is desired is used to determine an appropriate stabilizer based on invariant characteristics. Invariant surface recovery is an inverse mathematical problem. Visual reconstructions are inverse problems which tend to be ill posed in the sense that existence, uniqueness, and stability of the solution are not guaranteed when additional constraints are not considered. Constraints such as smoothness have been useful *a priori* information for possible solutions. The obtained representation should be independent of the viewpoint for a

stable shape description. Because of their formulations, the surface reconstruction techniques can be considered as extreme problems in which a functional minimization is involved. The classic problem of finding some value of the independent variable where the function is minimized belongs to traditional calculus, but the problem of minimizing a definite integral is considered as part of the calculus of variations.

Consider the direct problem of finding the function y given z as $Az=y$, where z corresponds to the original, total, and dense representations of some surface, and the function y is the sparse data of the surface obtained when the mapping A is applied. The inverse problem of finding z from y is an ill-posed problem, which can be solved using regularization theory. The main idea is to restrict the admissible solution space by introducing *a priori* knowledge as previously discussed. The solution can be found, then, as the function that minimizes some functional. This functional can be represented as

$$\Theta(z)=|Az-y|^2+\lambda Pz_1. \tag{5.1}$$

This functional can be interpreted as an energy or cost function, which provides a measurement of how close the solution is to the data in the first term, and how well it follows the *a priori* constraints of the desired function to be recovered in the second term. λ is the regularization parameter which controls the compromise between the degree of regularization of the solution and closeness to the data, whereas P is the regularization function. The problem is then finding a function z which minimizes that expression.

Stevenson and Delp[14] developed a mathematical analysis for invariant surface recovery from the constraint data based on regularization theory. From parametric representations of curves and surfaces, they construct a stabilizer which not only makes the problem well posed, but which is also based on invariant surface characteristics. First, they analyze the problem of invariant curve recovery from sparse data, and then they extended the approach to the problem of surface reconstruction. In their work, they use vector analysis for the surface representation. For the specific part of surface recovery, their approach is the use of a model of an ideal thin flexible plate of elastic material, and as the corresponding stabilizer, a measure of the strain energy of the deformed plate. The potential-energy density is expressed as

$$\Psi=A\frac{k_1^2(u)+k_2^2(u)}{2}+Bk_1(u)k_2(u), \tag{5.2}$$

or in terms of Gaussian and mean curvatures given by $K(u)=k_1(u)k_2(u)$, and $H(u)=k_1(u)+k_2(u)$, respectively, as

$$\Psi=2AH(u)^2-(A-b)K(u). \tag{5.3}$$

After a further simplification with $A=1$ and $B=O$, where A and B are constants of the material, they define an invariant stabilizer by integrating this energy density over the surface area as

$$\Omega[r(u)]=\int_U (2H^2(u)-K(u)dA. \tag{5.4}$$

A common approximation is to use the assumption of small values of z_x and z_y, in such a way that the Gaussian and mean curvatures can be approximated by

$$H(x,y)=\frac{z_{xx}+z_{yy}}{2}, \tag{5.5}$$

$$K(x,y)=z_{xx}z_{yy}-z_{xy}^2,$$

and the stabilizer becomes

$$\Omega[z]J=\frac{1}{2}\int\int_U (z_{xx}^2+2z_{xy}^2+z_{yy}^2)dx\ dy. \tag{5.6}$$

The surface reconstructed by minimization using this stabilizer is referred as the thin plate spline approximation or minimal energy B-spline approximation. The complete expression directly related to the last equation of the functional to be minimized is

$$\Theta(\nu)=\sum_{(x_i,y_i)\in\Omega}[v(x_i,y_i)-c(x_i,y_i)]^2 + \int\int_\Omega \frac{1}{2}(\Lambda^2\nu)^2-(1-\sigma)(\nu_{xx}\nu_{yy}-\nu_{xy}^2)dx\ dy, \tag{5.7}$$

where the roughness of the obtained function is measured by the energy in the functional of the second term, and the first term minimizes least-squared error at the discrete points. The equivalent analysis in one dimension is the minimization of the functional:

$$\Theta(v)=\sum_{x_i\epsilon(a,b)}[s(x_i)-c(x_i)]^2+\lambda\int_a^b |s''(x)|^2\ dx. \tag{5.8}$$

The solution of this variational problem is referred to as the minimal energy B splines.

The functional is called the energy functional and the cubic spline corresponds to the minimum energy configuration of a flexible wire. The spline interpolation minimizes the least-square error of the function values and its derivatives at the values of the discrete input data or points of interpolation. A B spline of degree n is obtained[15,16] as the following piecewise polynomial:

$$B_n(x;x_0,x_1,\ldots,x_{n+1}) =(-1)^{n+1}\sum_{k=0}^{n+1}\frac{(x-x_k)^n U(x-x_k)}{w(x_k)}, \tag{5.9}$$

where

$$w(x_k)=\prod_{\substack{j=0\\ j\neq k}}^{n+1}(x_k-x_j). \tag{5.10}$$

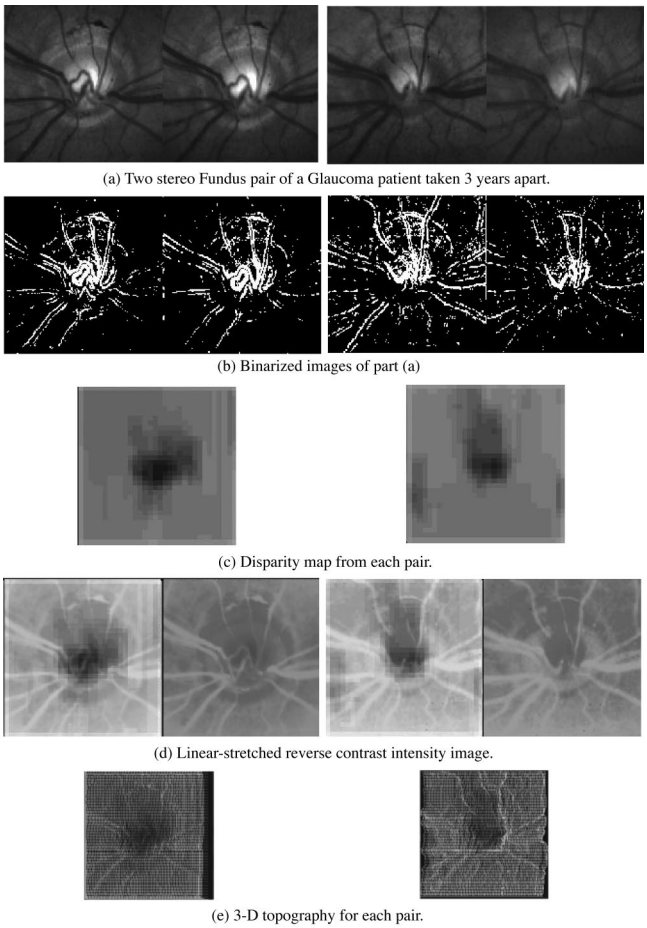

Fig. 5 Image processing steps.

The interpolated function obtained when the third-order B spline is applied consists of a sequence of third degree piecewise polynomials, which join at successive points continuously together with their slopes and curvatures. This interpolation is easily extended to the two-dimensional case, and incorporated in the whole procedure of surface recovery for obtaining the three-dimensional topography of the optic nerve head, according to the block diagram described.

6 Results

A series of synthesized as well as natural stereo images were tested to evaluate the depth-from-stereo algorithm described in this paper. In the first case, a random-dot stereogram whose underlying structure is a half sphere, was tested yielding good results. If a maximum error of ± 1 pixel is tolerated, the percentage of correct values obtained in the case of random-dot stereograms was 95% on average. This figure can be estimated in the case of random-dot stereograms because they are computer generated and we know in advance the correct values expected. This is different when real stereo imagery is processed. The algorithm was applied to the obtention of three-dimensional topography of the optic nerve head from a stereo pair of images. A cubic B-spline interpolation was performed in the last step in order to obtain a smoother surface. Surface recovery

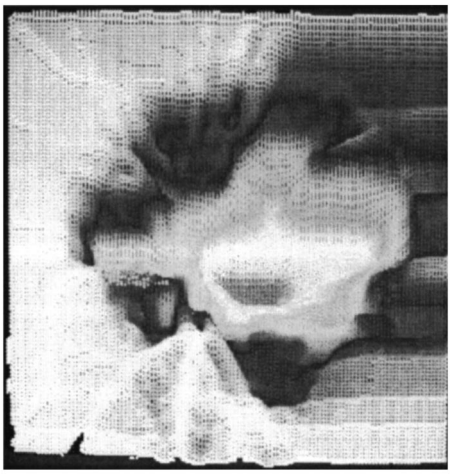

Fig. 6 3-D reconstruction of the deformation of the optic nerve head in a glaucomatous retina.

from sparse data is an inverse and ill-posed problem, which is currently a research area. Typical solutions involve regularization techniques, but in one form or another the final solution yield to some type of interpolation. Figure 5 shows the image processing steps and Fig. 6 shows the three-dimensional representation of the topography of a typical fundus image and the optic nerve head, obtained with the described algorithm.

7 Concluding Remarks

This paper presented a depth-from-stereo algorithm based on disparity detection by the cepstrum transform as the required matching procedure. In this context, the cepstrum transform represents a good alternative in terms of robustness, accuracy, and execution time, to find disparity between corresponding blocks. This technique is embedded in a coarse-to-fine search strategy, which provides the control between the various scales at which the operator is to be applied. The performance of the matching procedure was tested using computer generated stereograms formed of random dots as well as natural stereo images with good results. In every case, and especially when natural stereo imagery is used, the range image obtained consists of sparse information with density according to the available number of features which can be matched. A cubic B-spline interpolation was used in the last step as the final refinement of the range images, and it was shown to be a good alternative to get a dense representation from the available sparse depth data obtained from the depth-from-stereo algorithm. The whole system has been used for obtaining a three-dimensional representation of the optic nerve head topography, as medical support for the early diagnosis and follow up of glaucoma. This methodology is under testing for clinical evaluation and analysis of reproducibility and accuracy.

References

1. M. B. Shields, ‘‘The future of computerized image analysis in the management of glaucoma,’’ *Am. J. Ophthalmol.* **108**, 319–323 (1989).
2. A. W. Dreher, P. C. Tso, and R. N. Weinreb, ‘‘Reproducibility of topographic measurements of the normal and glaucomatous optic

nerve head with the laser tomographic scanner,'' *Am. J. Ophthalmol.* **111**, 221–229 (1991).
3. M. Hannah, ''SRI's baseline stereo system,'' *Image Understanding Workshop, Proc. of the DARPA Image Understanding Workshop*, pp. 149–155, Miami Beach, FL (1985).
4. N. Cochran and G. Medioni, ''Accurate surface description from binocular stereo,'' *Proc. Workshop on Interpretation of 3-D Scenes*, pp. 16–23, Austin, TX (1989).
5. D. Marr and T. Pogio, ''A theory of human stereo vision,'' *Proc. R. Soc. London, Ser. B* **204**, 301–325 (1979).
6. W. E. L. Grimson, *From Images to Surfaces, a Computational Study of the Human Early Visual System*, MIT Press, Cambridge, MA (1981).
7. Y. Otha and T. Kanade, ''Stereo by two-level dynamic programming,'' *IEEE Trans. Pattern. Anal. Mach. Intell.* **7**(4), 139–154 (1985).
8. W. Hoff and N. Ahuja, ''Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection,'' *IEEE Trans. Pattern. Anal. Mach. Intell.* **11**(2), 121–136 (1989).
9. J. Weng, ''A theory of image matching,'' *Proc. Third International Conf. on Computer Vision*, pp. 200–209, Osaka, Japan (1990).
10. A. D. Jepson and M. R. M. Jenkins, ''The fast computation of disparity from phase differences,'' *IEEE Trans. Pattern. Anal. Mach. Intell.* **11**, 398–403 (1989).
11. D. J. Lee, S. Mitra, and T. F. Krile, ''Dense depth maps from 2-D cepstrum matching of image sequences,'' *International Workshop on Robust Computer Vision*, Seattle, Washington (1990).
12. M. Ramirez and S. Mitra, ''Cepstrum and Cubic *B*-splines for 3-D surface recovery of smooth surfaces,'' *Canadian Conference on Electrical and Computer Engineering CCECE 92*, Paper No. TM10.6.1, Toronto, Ontario, Canada (1992).
13. S. Mitra, M. Ramirez, and J. Morales, ''Surface topography of the optic nerve head from digital images,'' *Medical Imaging VI, SPIE*, pp. 1652–1667, Newport Beach, California (1992).
14. R. L. Stevenson and E. J. Delp, ''Invariant reconstruction of curves and surfaces with discontinuities with applications in computer vision,'' *Technical Report, TR-EE 90-35*, School of Electrical Engineering, Purdue University, West Lafayette, Indiana (1990).
15. E. Jou and W. Han, ''Elastical and minimal energy splines,'' in *Curves and Surfaces*, Laurent, Mehaute, and Schumaker, Eds., pp. 247–250, Academic, Boston, MA (1991).
16. M. Unser, A. Aldroubi, and M. Eden, ''Fast *B*-spline transforms for continuous image representation and interpolation,'' *IEEE Trans. Pattern. Anal. Mach. Intell.* **13**(3), 277–285 (1991).

Juan Manuel Ramirez received his BS degree from the National Politechnique Institute, Mexico, in 1979, his MS degree from the National Institute of Astrophysics, Optics, and Electronics, Mexico, in 1983, and his PhD degree from Texas Tech University in 1991, all in electrical engineering. Since 1983 he has been with the Department of Electrical Engineering of the Universidad de las Americas, Puebla, Mexico, where he is currently a professor and chairman. His research interests include signal and image processing, neural networks, fuzzy logic, and pattern recognition.

Sunanda Mitra: Biography and photograph appear with the paper ''Special Section on Biomedical Image Representation,'' in this issue.

Jose Morales received his MD degree from the School of Medicine, National Autonomous University of Mexico, in 1979, with a postgraduate residence in the Ophthalmology Hospital at the National Medical Center. He had two fellowship appointments at the Sawelson Eye Center, Miami, Florida, and at the Glaucoma Service, New York Eye and Ear Infirmary, New York, in 1985 and 1988, respectively. Since 1988 he has been with the Health Science Center of Texas Tech University, where he is currently an assistant professor of ophthalmology and director of the Glaucoma Service. He has published over 40 papers in journals and conference proceedings in the area of ophthalmology.