IAPR Workshop on CV - Special Hardware and Industrial Applications OCT.12-14, 1988, Tokyo

USING TRINOCULAR SHADING IN THE RECONSTRUCTION OF A VISIBLE SURFACE

Vakalis, Ignatios
Department of Mathematics
Western Michigan University

Kalamazoo, Michigan 49008
U.S.A

ABSTRACT

This paper describes a new approach for integrating information provided by two important visual cues : a) stereo and b) SFS. A three camera stereo is considered whose image correspondance is established among three images taken from triangularly configured viewpoints. By using trinocular stereo the search problem of candidate matches, zero-crossings, is greatly simplified. The method also obtains depth information along horizontal edge elements and deals with occlusion. To improve the accuracy of the process, generating the depth map from the stereo module, shading information obtained from the three views is integrated in the above process. An algorithm is described which couples depth and shading imformation from the three cameras, and obtains an accurate dense depth map of the object in the scene. Supplying additional information in the reconstruction process, results in an improvement of the computed surface shape. The overall conceptual formulation of the solution is such that other visual cues, i.e., structure-from-motion, shape-from-texture, could be integrated too. The current method could be applied in the design of autonomous industrial robots as well as ground vehicles for planetary exploitations.

INTRODUCTION

One of the most important subsystems of an intelligent robot is vision. Without vision, a robot can repeat only a predetermined task sequence without any tolerance for even slight disturbances.

Vision has been explored by many researchers and the problem of obtaining a 3-D surface shape has been attacked using binocular stereo, SFS, shape-from-texture, shape-from-contours, structure-from-motion methods [1]. Output from the above methods, which provide only sparse information, are consistent with more than one surface definition. Rrestrictions to the space of admissable solutions has been achieved by imposing global variational principles, stated as smoothness constraints and expressed as regularizing terms [2], [3]. The problem with this technique, is the lack of detection and explicit representation of discontinuities which are very important for the reconstruction process. A complete analysis is presented by Terzopoulas [2].

It is known that "shape-from" processes provide only partial information about a scene. A crucial interaction is the integration of constraints from multiple early visual cues. Even though the initial information provided by each module does not determine the shape of surfaces it contributes to constraining the surface shape. With integration more accurate information from one process could compensate for less accurate information from another. Also, multiple visual processes may cooperate in infering the surface shape in places where no initial estimates where available.

In the approach described here, sparse local estimates of surface depth obtained by the trinocular stereo module, are allowed to interact with local estimates of surface orientation obtained from the SFS module.

REVIEW OF LITERATURE

This section briefly describes previous approaches towards the solution of the SFS problem. Also it gives an analysis of the three camera stereo as well as the matching process and the determination of depth.

The monocular stereo problem has been addressed extensively in the literature [6], [7], [8], but the results are neither accurate nor robust, since it is ambiguous to reconstruct the surface shape from one image. Horn [9], used the "characteristic strip" approach to solve the FOPDE but his method is inherently sequential and presents difficulties with noisy data. Ikeuchi and Horn [7] used the calculus of variations for the analysis of the problem and regularization term in the functional to be minimized. Brooks and Horn [6] tried a similar approach by enforcing the integrability constraint i.e., $Z_{XY} = Z_{YX}$, but they failed to develop a convergent iterative scheme.

In this paper the ambiguity of the SFS is resolved by using three images. Needle maps obtained from the images of the stereo device are combined and coupled with depth maps for a more accurate surface reconstruction. The integrability constraint is also enforced.

The other "shape-from" module which is used in the integration process is the stereo module. The most difficult problem of a conventional binocular stereo system is the correspondance problem. In the literature there are various hierarchical and global matching techniques to avoid ambiguous matches [11], [12], [13], [14]. Many different constraints for the search process are used, but the proposed algorithms are complicated and time consuming. A solution to the correspondance problem and aneffective measurement of the 3-D distance from a surface is achieved by using a trinocular stereo system. The problem is solved by adding a third camera [15]. Figure 1 shows the geometry of the trinocular system.

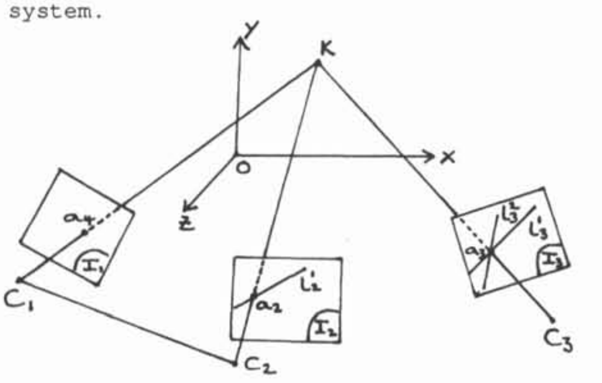

Figure 1.

The three cameras are modeled by optical centers C_i and image planes I_i where i=1,2,3. A physical point K has an image a_i on I_i plane. One very important constraint which has been used to find the corresponding point of a_1 on I_2, is the epipolar constraint which reduces the search from $O(N^2)$ to $O(N)$. Let l_3^1 be the projection of C_1a_1 on I_3 and l_3^2 be the projection of C_2a_2 on I_3. If a_1, a_2 are true corresponding points then the image of K on I_3 must exist at the intersection of l_3^2 and l_3^1, which is unique.

The symbolic descriptors which have been chosen to be matched among the three images are the zero-crossings of a $\nabla^2 G$ operator applied to each image. For the calculation of the epipolar lines the calibration technique developed by Feugeras and Toscani [16] could be used.

To establish the corresponding point of a_1, all possible candidates are detected lying on l_2^1. For each of them, the epipolar lines l_3^{2i} in I_3 as well as the intersections with l_3^1 are calculated. If a zero-crossing is detected in a small neighbourhood around the intersection of l_3^1 and l_3^{2j}, then l_3^{2j} uniquely determines the correct match of a_2^j in I_2.

To disambiguate multiple matches, the following constraints are used : a) the zero-crossings must have similar orientation and b) the same contrast sign. This critieria is used by the binocular stereo process for matching points lying on occluding boundaries, visible from only two cameras. Note also that for the three camera stereo, zero-crossings could have any orientation with horizontal lines.

In practise, the lines C_1a_1, C_2a_2, C_3a_3 will not intersect and so point K will be located at a point for which the distances from all three lines is a minimum [17].

CONTINUOUS FORMULATION

It is known that the general fusion problem is beyond current abilities of the artificial intelligence field and so in this work, an effort is made to combine only two shape-from modules. Figure 2 describes the camera setup.

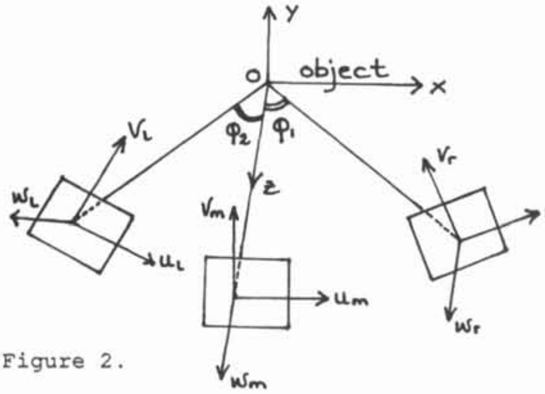

Figure 2.

The surface normal at point (X,Y,Z) is $\hat{n} = (-p(X,Y), -q(x,Y), 1)$. For the middle image plane the gradient is $p^m = p$, $q^m = q$. The values of p^r, q^r in the system (X,Y,Z) are :

$$p^{*r} = \frac{C_{1R} p^r - C_{2R}}{C_{1R} + C_{2R} p^r}, \quad q^{*r} = \frac{q^r}{C_{1R} + C_{2R} p^r} \qquad (1)$$

where $C_{1R} = \cos\varphi_1$, $C_{2R} = \sin\varphi_2$. Similarly the values for p^L, q^L are :

$$p^{*L} = \frac{C_{1L} p^L - C_{2L}}{C_{1L} + C_{2L} p^L}, \quad q^{*L} = \frac{q^L}{C_{1L} + C_{2L} p^L} \qquad (2)$$

For the continuous formulation of the problem, we assume that the segmentation process has already taken place so the process of reconstructing a single surface is described. We assume that the surface is Lambertian and the object is far away so orthographic projection can be used. The position of a single light source is estimated using techniques in [10],[21].

A variational approach to the SFS problem is explored and the smooth surface to be reconstructed must minimize the following functionals :

$$E_1 = \iint\left[(z_x - p^l)^2 + (z_y - q^l)^2\right]dx\,dy + \lambda\iint\left\{\left[p^{*l}(x^*, y^*) - p^m(x,y)\right]^2 + \left[q^{*l}(x^*, y^*) - q^m(x,y)\right]^2\right\}dx\,dy$$

$$E_2 = \iint\left[(z_x - p^r)^2 + (z_y - q^r)^2\right]dx\,dy + \lambda\iint\left\{\left[p^{*r}(x^{**}, y^{**}) - p^m(x,y)\right]^2 + \left[q^{*r}(x^{**}, y^{**}) - q^m(x,y)\right]^2\right\}dx\,dy$$

where (X^*, Y^*), (X, Y), (X^{**}, Y^{**}) are corresponding points in I_1, I_2, and I_3 image planes.

ALGORITHM

A) Establish correspondance between zero-crossings from left-middle as well as from middle-right images. Obtain depth information.

B) Find orientation at matched contours. The derivatives of depth and image intensities are used to provide the orientation values [22].

C) 1. Use Ikeuchi-Horn [6] iterative scheme for each image separately. Boundary conditions are provided by step **B**.

$$f_{ij}^{n+1} = \overline{f_{ij}}^n + \frac{h^2}{\lambda}\left(E_{ij} - R(f_{ij}^n, g_{ij}^n)\right)\frac{\partial R(f_{ij}^n, g_{ij}^n)}{\partial f_{ij}^n}$$

$$g_{ij}^{n+1} = \overline{g_{ij}}^n + \frac{h^2}{\lambda}\left(E_{ij} - R(f_{ij}^n, g_{ij}^n)\right)\frac{\partial R(f_{ij}^n, g_{ij}^n)}{\partial g_{ij}^n}$$

An improvement has been incorporated where a nine point approximation for the Laplacian is used. Thus, $(\nabla^2 f)_{ij} \simeq \frac{1}{h^2}(\overline{f_{ij}} - f_{ij})$ where :

$$\overline{f_{ij}} = \frac{1}{5}(f_{i+1,j} + f_{i,j+1} + f_{i-1,j} + f_{i,j-1}) + \frac{1}{20}(f_{i+1,j+1} + f_{i+1,j-1} + f_{i-1,j-1} + f_{i-1,j+1})$$

2. Using a mapping of (f,g) to (p,q) space, derive values for p_{n+1}^l, q_{n+1}^l, p_{n+1}^r, q_{n+1}^r, p_{n+1}^m, q_{n+1}^m

D) 1. Reconstruct the depth map (middle image) by solving :

$$\nabla^2 z_{n+1}^m = (q_v^m + p_u^m) + \lambda\left[p^m - p^{*l}\right]\frac{\partial p^{*l}}{\partial x} + \lambda\left[q^m - q^{*l}\right]\frac{\partial q^{*l}}{\partial x}$$

The above equation is derived from the Euler equation of E_1 on the previous section. The poisson equation is solved by using an iterative technique described in [20] with an improvement of a none point approximation to the Laplacian.

2. Reconstruct a depth map from left as well as from right image by solving two poisson equations :

$$\nabla^2 z_{n+1}^l = (q_v^l + p_u^l) + \lambda\left(p^l - \frac{C_{1l}p^m + C_{2l}}{C_{1l} - C_{2l}p^m}\right)\frac{\partial p^m}{\partial u} + \lambda\left(q^l - \frac{q^m}{C_{1l} - C_{2l}p^m}\right)\frac{\partial q^m}{\partial v}$$

A similar equation holds for the right image.

E) 1. From the solutions of steps **D1** and **D2**, obtain values for : $\overline{p_{n+1}^l}$, $\overline{q_{n+1}^l}$, $\overline{p_{n+1}^m}$, $\overline{q_{n+1}^m}$, $\overline{p_{n+1}^r}$, $\overline{q_{n+1}^r}$ by using a five point difference approximation to the depth values on a grid.

2. Replace the values of p_{n+1}^l, q_{n+1}^l, p_{n+1}^m, q_{n+1}^m, p_{n+1}^r, q_{n+1}^r with the above estimates.

F) Return to step **C** until no significant changes occur.

CONCLUSIONS - FUTURE WORK

A three camera approach to the stereo problem was considered so that problems like, matching horizontal edge elements, occlusion, false targets, e.t.c., disappear. Consequently, faster and more accurate depth values were calculated. Better accuracy in estimation of the Laplacian with nine points was used. The method forced the integrability constraint and coupled from an earlier stage, the needle and depth maps for each image.

One of the important drawbacks of many SFS methods, and this work also, is the lack of explicitly dealing with discontinuities. Future plans will involve the detection and the representation of them concurrent with the reconstruction problem. Strong assumptions made by the SFS module, i.e., lambertian surface, smooth, known albedo, could be relaxed by incorporating similar techniqies used in active vision [23]. Also the integration of many more shape-from modules i.e., shape-from-texture, shape-from-contour, shape-from-motion, is to be considered as well as the fusion of information from different feature and area based stereo methods. Finally, view-point invariant representation of surfaces [24] may be considered in order to eliminate the wobbling effect. Psychophysical experiments may reveal slight variance in the surface perceived by humans viewing sparce ramdom dot stereograms while the dots undergo rigid 3-D transformations and also if these variations are consistent or not with the process of reconstruction.

REFERENCES

[1] Brady I.M. (1982) **Computer Vision**, Amsterdam : North-Holland

[2] Terzopoulos D. (1983) Multilevel computational processes for visual surface reconstruction, **Computer Vision, Graphics, Image Process., 24**, pp52-96

[3] Grimson W. (1983) An implementation of a computational theory of visual surface interpolation, **Computer Vision, Graphics, Image Process., 22**, pp39-69

[4] Marr D. (1982) **Vision : A computational investigation into the human representation and processing of visual information**, San Francisco, CA : Freeman

[5] Nishihara H.K. (1981) Intensity,

visible surfaces and volumetric representations, **Artificial Intelligence, 17**, pp265-284
[6] Horn B. & Brooks M. (1986) The variational approach to shape from shading, **Computer Graphics and Image Processing, 33**, pp174-208
[7] Ikeuch K. & Horn B. (1985) Numerical shape from shading and occluding boundaries, **Artificial Intelligence, 17**, pp141-185
[8] Simchony T. & Chellopa R. (1987) Direct analytical methods for solving poisson equations in computer vision, **IEEE Computer Society Workshop on Computer Vision**
[9] Horn B. (1970) Shape-from-Shading : A method for obtaining the shape of a smooth opaque object from one view, **MAC-TR-79 AI Lab M.I.T.**
[10] Brooks M. & Horn B. (1985) Shape and source from shading, **IJCAI-85**
[11] Marr D. & Poggio T. (1976) Cooperative computation of stereo disparity, **Science 194**, pp283-287
[12] Marr D. & Poggio T. (1979) A computational theory of human stereo vision, **Proc. of R. Soc. of Lond. B, 204**, pp301-328
[13] Grimson W. (1981) A computer implementation of a theory of human stereo vision, **Phil. Trans. R. Soc. Lond. B, 292**, pp217-253
[14] Barnard S. & Fischler M. (1982) Computational Stereo, **Computational Surveys, 14**, 4, pp553-572
[15] Yachida M. (1986) 3-D data acquisition by multiple views, **Robotocs Research : 3rd Intern. Symposium**, p11-18
[16] Feugeras O. & Toscani G. (1986) The calibration problem for stereo, **Proc. of CVPR '86**, pp15-20
[17] Duda R. & Hart P. (1973) **Pattern Classification and Scene Analysis**, John Wiley & Sons
[18] Gerhard A, Platzer H, Steures J. & Leng R. (1986) Depth extraction by stereo triples and a fast correspondance estimation algorithm, **Proc. of 8th ICPR**, pp512-515
[19] Ikeuchi K. (1987) Determining a depth map using dual photometric stereo, **International Journal of Robotics Research, 6**, 1, pp12-31
[20] Ikeuchi K. (1983) Constructing a depth map from images, **A.I. Lab, M.I.T. Memo 744**
[21] Pentland A. (1982) Finding the illumination direction, **Journal of Optical Society of America, 72**, pp4-15
[22] Blake A. (1986) On geometric information obtainable from simultaneous observation of stereo contour and shading, **Report CSR 205-86**, Dept. of CS, University of Edinburgh
[23] Aloimonos J, Weiss I. & Bondyopadhyay A. (1987) Active vision, **Intern, Journal of CV, 1**, 4, pp233-356
[24] Blake A. (1984) Reconstructing a visible surface, **AAAI-84**, pp23-26