A THREE CAMERA APPROACH FOR THE RECONSTRUCTION OF A VISIBLE SURFACE USING CORNER, EDGE & SHADING INFORMATION

Ignatios Vakalis
Computer Science Department
Western Michigan University

Kalamazoo, Michigan 49008
U.S.A.

ABSTRACT

In this paper we investigate the problem of the reconstruction of visible surfaces(s) using a trinocular stereo approach. Information from matched corner triplets, zero-crossings, intensity edges as well as SFS, is used for a faster and more accurate solution.

At first, the method detects corners by operating directly on the three gray-tone images and does not rely on prior segmentation. Then the zero-crossings of a : $\nabla^2 G$ operator applied in each image are detected. After the images have been smoothed by a Gaussian filter, a directional operator : $\partial^2/\partial n^2$ is applied, for the detection of intensity edges. A multiresolution matching algorithm is invoked for finding the correct correspondance of the above features in the three stereo images. Three disparity arrays in different resolutions are maintained. Matching of different features is done in parallel. With the triangulation technique, 3-D distances are obtained and disparity information as well as information from intensity edges is used to detect depth discontinuties.

The depth maps obtained from matching different type features, are superimposed, and a three-stage smooth surface interpolation is invoked. The reconstructed surface "favors" the depth values which correspond to matched corner points. Needle maps obtained using classical SFS methods and the shape of the smooth surface is adjusted according to the output of the SFS module. For areas of depth discontinuities, the surface patches are locally reconstructed again, by using depth information from the finest resolution.

INTRODUCTION

Vision is one of the most important subsystems of an intelligent robot and deals with the problem of deriving a description of a three-dimensional scene from its two-deimensional images. It attempts to solve the inverse of the image formation problem which is well posed. As it is true with the majority of inverse problems, vision is faced with an ill-posed problem in the sense of Hadamard [1]. Computational studies have provided in recent years promising but far from complete theories of the processes necessary to solve the ill-posed problem. It becomes clear that a single module is not sufficient to solve this problem. Recently a lot of effort is made towards the integration and fusion of information provided by different visual cues.

One of the most important "shape-from" modules which have been used in the literature is the stereo module. Many different matching techniques have been proposed to solve the correspondance problem and minimize the ambiguity which occurs during the matching phase [2], [3], [4], [5]. In this paper a three camera approach is considered where the correspondance problem is reduced to a simple verification, at a precise location, in the third image. The location is determined by the intersection of two epipolar lines [6], [7]. The trinocular stereo vision method also overcomes the problem of the conventional binocular stereo in matching horizontal edge elements. Faster and more reliable matches of the different types of symbolic descriptors are obtained and therefore more accurate depth values are calculated.

The next "shape-from" module, which alone cannot solve the problem, is the shape-from-shading module. Many different approaches can be found in the literature, but the results are niether accurate nor robust since it is ambiguous to reconstruct the surface shape from one image. Horn [8] has pioneered the research towards the solution of the SFS problem by solving a FOPDE. Ikeuchi and Horn [9] used calculus of variations. Brooks and Horn [10] tried a similar approach by enforcing the integrability constraint but failed to develop a convergent iterative scheme. In our approach needle maps obtained from the three images are transformed to correspond to a global coordinate system and used to enhance the shape of the reconstructed surface in the first stage of the interpolation process.

INTEGRATION

A crucial interaction that will be examined in this paper is the combination or integration of information from the two visual cues which have been mentioned above. Prior vision research was concentrated in solving the reconstruction problem using a single module. The output of these modules provide only sparse information and are consistent with more than one surface definition.

Recently a lot of effort is made in combining different shape-from modules which, in a first approximation could be considered independent. Grimson [11] tried to integrate the depth information from a binocular stereo system with the difference in shading from the two images. His method does not work well in practice in areas of specularities. Moerdler and Boult [12] suggested a two level integration process combining information from two shape-from-texture methods with depth values from stereo. Their algorithm constructs only smooth surfaces without any discontinuity representation and is applied to textured surfaces. Numerous papers have been published exploring the integration of binocular stereo and the structure-from-motion module. Sandini and Tistorelli [13] combined depth maps obtained from the two modules so that more accurate 3-D distances could be calculated.

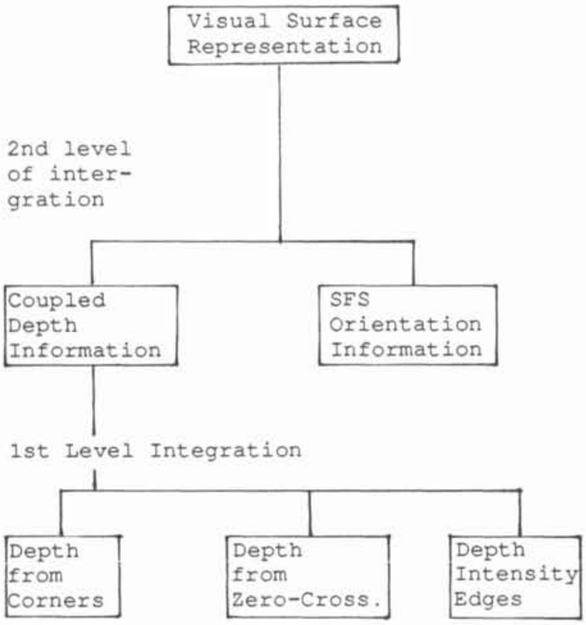

Figure 1.

The method that will be presented here is an initial stage of an on-going research towards the design of a system which will incorporate many "shape-from" modules. We present a method which integrates 3-D depth values obtained from matched corner, zero-crossing and intensity edges. At a second level of integration, SFS information is coupled for a more accurate reconstruction. Depth discontinuities for the reconstructed surface(s) are detected by integrating information from disparity arrays and intensity edges. Local surface patches around discontinuities are recomputed, using fine resolution depth values. The overall, two stage integration process is depicted in Figure 1.

METHOD

In this section we describe in an algorithmic form the steps towards the reconstruction of visible surfaces preserving depth discontinuities.

(A) Convolve each of the three stereo images with three different size Gaussian filters : $G_\sigma(X,Y) = \sigma^2 e^{-(x^2+y^2)/2\sigma^2}$. The sizes correspond to coarse, medium and fine resolution. Obtain the zero-crossings in the output of the convolved images by seeking points (X_o, Y_o) such that $f(X_o, Y_o, \sigma) = 0$ where $f(X,Y,\sigma) = \nabla^2[G_\sigma(X,Y) * E(X,Y)]$.

(B) Apply the directional differentiation operator $\partial^2/\partial n^2$ in each image to localize the intensity edges. In the above operator n denotes the direction of the intensity gradient. A directional operator has been used since it is better than an isotopic operator i.e., ∇^2, in the reduction of noise effects and so give better localization. This projects to small errors in the computed depth values.

(C) The new symbolic descriptor which is considered in the matching process is what we define as corner points. These are points where one-sided directional derivatives exist but have different values in pairs of opposite directions. Corners have not been used in the past for the stereo problem even though they are significant information conveyors. Their position and type implies important constraints about the shape of the object(s) in the scene. The motivation behind the usage of these points is that corners are more sparse tokens than "edges" in the images so they are easily and more reliably matched among the three images. Thus, more accurate depth imformation is obtained. The method developed by Haralick [14] is used for the detection of corners in the images.

Currently new approaches are investigated for more accurate localization of corner points in different resolutions.

(D) The matching process is executed by applying a multiresolution technique. Initial estimations of the disparity value from coarse resolution are used to drive the matching in finer resolutions. With the trinocular stereo the matching phase is very simple and mainly verifies the location of the third homologous point in the third image [6]. Edges of any orientation can be matched with the three camera stereo and so the mismatch difficulty of horizontal edges in a conventional binocular stereo is eliminated. The matching of the different type features is done in parallel.

(E) Three depth maps, for each resolution, corresponding to matched corners, zero-crossings and intensity edges are calculated using the camera parameter from the calibration phase. For every resolution the three depth maps are superimposed and so a blended depth map is created containing more accurate 3-D distances, for the physical points which are depicted as corner points in the images.

(F) A three stage approximation method is described in this step which accomplishes an initial representation of the visual surfaces. The surface(s) is fitted through the blended depth map obtained from the medium resolution depth values. The first stage sets up a gridded problem consisting of points (u_i, v_i). A uniform grid is considered which encloses all data points (X_k, Y_k). In approximating the function and its partial derivatives a weighted local least square fit is made to the nearest 15 points for every grid point. The weights are selected as $w_k = (1-d_k/d)^2$ [14]. The weighted least squares problem :

$$\min[\sum_{k=1}^{15} w_k [q(X_k,Y_k) - Z(X_k,Y_k)]^2]$$

is solved, where $q(X_k,Y_k)$ is a cubic polynomial and $Z(X_k,Y_k)$, the depth values encoded in the blended depth map. The second stage performs a piecewise bicubic Hermite interpolant to the gridded data generated from the first stage [15]. For point (X,Y) inside a grid square, the function is defined as :

$$S(X,Y) = \sum_m \sum_n a_{mn} b_m(X) C_n(Y)$$

where $b_m(X)$, $C_n(Y)$ are standard cardinal Hermite basis functions. The third stage involves a modified Shepard's method so that the surface is forced to interpolate only through the depth values which correspond to corner points as being more accurate [16].

(G) Assuming that the surfaces are Lambertian, and that a single light source is present, the Ikeuchi-Horn algorithm [17] is used to obtain needle maps from each stereo image. The needle maps are combined and a correction is made so that they refer to the same coordinate system. The surface shape is adjusted so that its partial derivatives will be compatible with the information of this step.

(H) One of the most important pieces of information which must be recorded and presented, is the locations of the the depth discontinuities. Orientation discontinuitues are very important too, but the current method is unable to detect them. Previous research shows that while it may be possible in certain instances to perform the detection using single sources of discontinuity information in isolation, the only hope for solving the problem in general is to employ multiple discontinuity cues simultaneously. Towards that direction, information from disparity arrays is integrated with information from intensity edges. One way to find depth discontinuities is to identify semi-occluded regions that cannot be matched by the stereo correspondance computation. The matching algorithm fails to match semi-occluded regions but it is capable of signalling these regions. Also potential discontinuities are examined by computing the local depth gradient $\nabla.S(X,Y)$ on the initial surface approximation. Comparing the $\nabla.S(X,Y)$ values at each intensity edge to a threshold level the existence or absence of a depth discontinuity can be established. A double threshold method, similar to [15] is used for the detection. Common regions labelled as discontinuity areas from both methods are established. For only those regions the corresponding surface patches are constructed again using depth information from the finest resolution level. Note that the surface need not be reconstructed at the next level since it is smooth everywhere, so accurately representable by step **G** except for discontinuity areas. With this approach significant computational savings are established.

CONCLUSIONS - FUTURE WORK

In this work a method was described for the complete reconstruction of visible surface(s) detecting and preserving depth discontinuities. The depth map was obtained by trinocular stereo techniques so that the correspondance problem is minimized. A two level integration process was described where an initial surface approximation is established from depth values obtained from matched corners, zero-crossings and intensity edges. Shape-from-shading information was used to adjust the shape of the surface. Integrating information from disparity fields and intensity edges, enable the reconstruction process to exactly locate with high confidence the depth discontinuity regions.

Currently new surface fitting techniques are being investigated so that both depth and orientation discontinuities would be able to be amalgamated in the reconstructed surface(s). Also more "shape-from" modules must be integrated, i.e., shape-from-texture, shape-from-motion, e.t.c., so that a general solution to the fusion problem would be accomplished.

REFERENCES

[1] Hadamard I. (1923) **Lectures on the Cauchy Problems in Linear and Partial Differential Equations**, Yale, University Press

[2] Marr D. & Poggio T. (1976) Cooperative Computation of Stereo Disparity, **Science 194**, pp283-297

[3] Marr D & Poggio T (1979) A Computational Theory of Human Stereo Vision, **Proc. of R. Soc. of Lond. B, 204**, pp301-328

[4] Grimson W. (1981) A Computer Implementation of the Theory of Human Stereo Vision, **Phil. Trans. R. Soc. of Lond. B, 292**, pp217-253

[5] Barnard S. & Fischler M. (1982) Computational Stereo, **Computing Surveys, 14**, 4, pp553-572

[6] Yachida M. (1985) 3-D Data Acquisition by Multiple Views, **Proced. of Third Inter. Symposium of Robotics Research**.

[7] Ayache N. & Lustman F. (1987) Fast and Reliable Passive Trinocular Stereo Vision. **First Inter. Conf. on Computer Vision**, London, England, pp422-427

[8] Horn B (1970) Shape-from-shading : A method of obtaining the shape of a smooth opaque object from one view, **MAC-TR-75, A.I. Lab, M.I.T.**

[9] Ikeuchi K. & Horn B. (1985) Numerical shape from shading and occluding boundaries, **Artificial Intelligence, 17**, pp141-185

[10] Brooks M. & Horn B. (1985) Shape and Source from Shading, **IJCAI-85**

[11] Grimson W. (1984) Binocular Shading and Visual Surface Reconstruction, **Comp. Vision Graphics and Image Processing, 28**, pp19-43

[12] Moerdler M. & Boult T. (1988) The Integration of Information from Stereo and Multiple Shape-from-Texture cues, **Proc. of Computer Vision and Pattern Recognition Conference**, Ann Arbor, MI, pp524-529

[13] Sandini G. and Tistarelli M. (1986) Recovery of Depth Information : camera motion as an integration to stereo. **Workshop on Motion : Representation and Analysis**, South Carolina, pp39-43

[14] Stead S.E. (1984) Estimation of Gradients from Scattered Data, **Rocky Mountain Journal of Mathematics, 14**, pp265-279

[15] Fraux I. & Pratt M. (1979) **Computational Geometry for Design and Manufacture**, New York : Wiley

[16] Bornhill R. (1977) Representation and Approximation of Surfaces, **Proc. of the Math. Soft. Symposium.**, University of Wisconsin

[17] Brooks M. & Horn B. (1985) The Variational Approach to Shape from Shading, **A.I. Memo 813, M.I.T.**

[18] Carny J. (1983) Finding Edges and Lines in Images, **A.I. Memo 720, M.I.T.**