A NEW CONCEPTUALLY ATTRACTIVE & COMPUTATIONALLY EFFECTIVE APPROACH TO SHAPE FROM SHADING

B. Cernuschi-Frias, R. M. Bolle and D. B. Cooper

Brown University, Division of Engineering, Providence, RI 02912

ABSTRACT

A conceptually new, computationally simple approach to shape from shading is presented. It is assumed that objects of interest in 3-D space can be approximated by chunks of spheres, cylinders and planes, and that the formation of the true image can be modelled as the scattering from a Lambertian surface of light from a distant point source or distributed source. The observed image is modelled as this true image plus a white Gaussian perturbation. To permit parallel processing for computational speed, an image is partitioned into many small square windows (image patches) that can be processed simultaneously. It is assumed that a chunk of only one object type (plane, cylinder, sphere) is seen in a patch. Two kinds of results are presented. The first result is an algorithm for the recognition of the object type seen within a patch. This algorithm uses constrained 2-D quadric polynomial approximations to the picture function to implement true Bayesian recognitition. The second result is an algorithm for the estimation of the 3-space location (and orientation) of the recognized object seen within a patch. The image is thresholded with many pairs of thresholds and lines constrained to be parallel or ellipses constrained to have the same shapes and orientations are fit to the resulting swaths of data. Three-D location and orientation parameters are estimated from these statistics. The two algorithms are derived from a general formal statistical formulation of the shape from shading problem.

I. OVERVIEW

The ultimate goal of this work is the recognition of a manufactured object and the estimation of its location and orientation in 3-D space, based on the use of a single 2-D image. A priori information of object structure is used, as well as appropriate functions for image formation based on knowledge of camera location and perhaps of the illuminating light source. Our recognition and information extraction approach is based largely on data generation modelling and then the application of Bayesian recognition and roughly maximum likelihood estimation. The approach is robust to deviations from the assumed models and is computationally attractive. In this paper, we concentrate on the subproblem of object type recognition and object 3-D parameters estimation based on an image patch—the data in a single window. The patches are sufficiently large to contain important structural information, but are sufficiently small to permit simple processing. The information extracted simultaneously from many such patches must then be used to make reliable inferences concerning the object that is composed of the chunks of planes, cylinders and spheres seen. Section II describes experimental results on extracting information from a patch. Section III contains a brief introduction to our general formulation, and points out how the algorithms for extracting information from a patch come from this general formulation.

This work was partially supported by Army Research office grants #DAAG29-81-G-0124 and DAAG29-81-K-0167 and by the National Science Foundation grant #ECS-8119676.

II. EXPERIMENTAL RESULTS

The two results briefly described are typical of those observed in experimentation with the algorithms. Figure la is an image of a portion of a sphere taken with a vidicon camera in a room illuminated by standard ceiling mounted fluorescent tubes behind diffusers. Three data swaths can be seen in Figure Ib. Each swath consists of pixels having image intensities lying between a specific pair of thresholds. An ellipse has been fit to each swath. The elliptic shape of the data swaths occurs because the contours of constant image intensity associated with the sphere are ellipses when the light source is a point source, and are approximately ellipses for a distributed source such as the ceiling fixtures commonly encountered. A new twist arising here is that the ellipse fitting has been constrained such that all ellipses have the same shape and orientation. In practice, many pairs of thresholds and associated data swaths and fitted ellipses are involved.

The elliptic curve fitting is computationally simple and involves constrained least squares curve fitting. The location of the projection of the sphere center on the image plane can be estimated from the ellipse parameters. Though less computation is required if the sphere radius and source direction are known, these parameters can be a priori unknown and everything can be estimated from the fitted ellipses. For the image of a cylinder, contours of constant image intensity are parallel straight lines, and an analogous constrained least squares line fitting algorithm can be used to fit parallel lines to data swaths resulting from thresholding the image.

Figure 2a is an image of a cylinder (a can) taken with the same set-up as used for the sphere. Figure 2b shows the image partitioned into patches along with the 3-D object shape-type classifications. The symbols p, c, s, m, stand for plane, cylinder, sphere, and mixed. Mixed patches are those which are views of two or more surfaceshape-types. The decisions are based on partially unknown 3-D surface shape parameters, because object orientation and location are a priori unknown. Each image patch in Figure 2b is a constrained 2-D polynomial fit to the raw image data in the window. The polynomial fit is constrained to be appropriate for data arising from a 3-D planar surface, a cylindrical surface, a spherical surface, or two surfaces, respectively. The 3-D cylinder viewed is slightly crushed where it is held by a clamp.

III. PICTURE FUNCTION ASSOCIATED WITH A PATCH OF 3-D OBJECT SURFACE

The true picture function $g(x,y)$, at pixel (x,y), for the 3-D quadric surfaces of interest satisfies Eq. (1) when no noise or other perturbations are present.

$$g^2 + \beta_1 xg + \beta_2 yg + \beta_3 g + \beta_4 x^2 + \beta_5 xy + \beta_6 y^2 + \beta_7 x + \beta_8 y + \beta_9 = 0 \qquad (1)$$

There will be constraints among the β_i depending on the 3-D surface shape being viewed. The β_i are functions of the 3-D surface shape parameters, of the angle and intensity of the incident illumination, of the 3-D surface reflection properties, and of the viewer to object distance and direction. Denote Plane, Cylinder, and Sphere as 3-D shape type 1, 2, 3, respectively. Then, for example, Eq. (1) takes the form [1]

$$\left\{g - [K/(1-v_3^2)](u_1v_2 - u_2v_1)(v_2x - v_1y)\right\}^2 \cdot - [K/(1-v_3^2)]^2[-u_1v_1v_3 - u_2v_2v_3 + u_3(1-v_3^2)]^2 \cdot \cdot [(1-v_3)R^2 - (v_2x - v_1y)^2]$$

for shape type 2(cylinder). Here, $K = I\rho$ with I the intensity of the incident illumination and ρ the surface reflection coefficient. v_1, v_2, v_3 are the components of a unit column vector $\underline{v}$ lying along the cylinder axis, and u_1, u_2, u_3 are the components of the unit vector $\underline{u}$ pointing from the distant point source to the cylinder. If the observed image is this picture function plus additive white Gaussian noise having mean 0 and variance σ^2, the joint likelihood of an image patch and that the patch is a view of a 3-D surface of type k is p(patch, type k)

$$= P_k \int \prod_{x,y} (\sigma^2 2\pi)^{-1/2} \exp\left\{-(1/2\sigma^2)[g(x,y) - m_k(x,y)]^2\right\} \cdot p_k(\underline{\alpha}_k)\, d\underline{\alpha}_k$$

Here, $m_k(x,y)$ is a solution of (1) for $g(x,y)$ and is the mean value function for $g(x,y)$ under the assumption that 3-D shape type k is seen; $\underline{\alpha}_k$ is the vector of unknown parameters, e.g., for $k = 3$, $\underline{\alpha}_k$ consists of sphere center location parameters and any other unknowns such as perhaps unknown source-direction. The β_i in (1) and $m_k(x,y)$, are partially unknown coefficients that are determined by known parameters and by the vector $\underline{\alpha}_k$ of unknown parameters; $p_k(\underline{\alpha}_k)$ is an a priori probability density function for $\underline{\alpha}_k$, and P_k is the a priori probability that the image patch is a view of 3-D shape type k. Bayesian recognition of object shape type is made by choosing that k for which (3) is a maximum. A maximum a posteriori estimate of the unknown vector parameter $\underline{\alpha}_k$ given the image patch and the fact that it is a view of a chunk of an object of shape type k is that value of $\underline{\alpha}_k$ for which the integrand of (3) is a maximum.

The integral in (3) is impossible to compute exactly, but in general if there are many pixels in the image patch, the exponential in (3) is a Gaussian function of $\underline{\alpha}_k$ about the maximum likelihood estimate $\hat{\underline{\alpha}}_k$, the value of $\underline{\alpha}_k$ at which the exponential achieves a maximum. Using this approximation, a simple closed form expression can be derived analytically for the integral (3). Let $\underline{W}^T = (1, \beta_1, \beta_2, \ldots, \beta_9) = (1, \underline{\beta})$ and $\underline{U}^T = (g^2, gx, \ldots, y, 1) = (g^2, \underline{V}^T)$. Eq. (1) is $\underline{W}^T\underline{U} = 0$. An approximation to the MLE estimator $\hat{\underline{\alpha}}_k$ for (3) can be had by solving for the value of $\underline{\alpha}_k$ for which (4) is a minimum

$$e(\underline{\beta}) = \sum_{x,y,g(x,y)\in \text{Patch}} (\underline{W}^T\underline{U})^2 \qquad (4)$$

It can be shown that (4) can be expressed as

$$e(\underline{\beta}) = e(\bar{\underline{\beta}}) + (\underline{\beta} - \bar{\underline{\beta}})^T R(\underline{\beta} - \bar{\underline{\beta}}) \qquad (5)$$

where $R = \sum_{x,y,g(x,y)\in \text{Patch}} \underline{V}\underline{V}^T$ and $\bar{\underline{\beta}}$ is the coefficient vector for the unconstrained least squares approximation and is given by $\bar{\underline{\beta}} = -R^{-1}\sum_{x,y,g\in \text{Patch}} g^2(x,y)\underline{V}$.

Of great importance in (5) is that the coefficient vector that minimizes (5) and is constrained to be appropriate to the image of an object of type k can be computed by minimizing the last term on the right side of (5), i.e., the quadratic form. This minimization takes place in the 9-dimensional space and does not involve the picture function --- a significant computational simplification.

Even though the computations required by (5) and (3) are now feasible, further approximation is possible and results in algorithms that are both computationally simple and extremely effective for recognition, and location and orientation estimation.

IV. 3-D OBJECT RECOGNITION

If the window size is roughly between one fifth to one half the diameter of the image of a cylinder or a sphere, a quadric polynomial approximation to the mean value function $m_k(x,y)$ is extremely effective. Then Eq. (3) simplifies further and the Bayesian recognition rule becomes: choose that k* for which

$$\|\bar{a} - \hat{a}_{k^*}\|^2 + \sigma^2 \ell_{k^*} \ln(n) \text{ is a minimum.}$$

Here, $\bar{a}$ is the unconstrained least squares error coefficient vector for the approximation of the image patch by a sum of 2-D orthonormal polynomials up through 2nd degree in x and y. $\hat{a}_{k^*}$ is the coefficient vector that is a best fit to a and satisfies a constraint appropriate to the image for object type k. ℓ_{k^*} is an integer constant appropriate to the image constraints associated with object type k, and n is the number of pixels in the image patch. For example, $\ell_1 = 1$, $\ell_2 = 10$, $\ell_3 = 14$. See [1] for more detail.

V. 3-D OBJECT LOCATION AND ORIENTATION ESTIMATION

Computationaly-simple, good, approximate maximum-likelihood estimates of β satisfying the image constraints for object type k can be obtained. We illustrate the approach for a sphere. Eq. (4) can be rewritten as

$$e = \sum_{\substack{I \\ \text{over all levels}}} \left\{ \sum_{\substack{x,y \text{ such that} \\ g(x,y) = I}} [\beta_4 x^2 + \beta_5 xy + \beta_6 y^2 + \beta_{I,1} x + \beta_{I,2} y + \beta_{I,3}]^2 \right\} \quad (6)$$

where $\beta_{I,1} = \beta_7 + \beta_1 I$, $\beta_{I,2} = \beta_8 + \beta_2 I$, $\beta_{I,3} = \beta_9 + I^2$

(Note, that if no additive noise is present, then for each intensity level I the summand in (6) if set equal to 0 is an ellipse; the ellipses for the various levels all have the same orientation and same ratio of major to minor axes; the ellipse centers lie on a line; and ellipse size varies quadratically). If the direction of the light source is known, β_4, β_5, β_6 are known and only the $\beta_{I,1}$, $\beta_{I,2}$ and $\beta_{I,3}$ must be estimated. Otherwise, all parameters must and can be estimated. Though the optimal approach is to estimate all the β_i simultaneously, our suboptimal approach is to estimate the new set of coefficients β_4, β_5, β_6, $\beta_{I,1}$, $\beta_{I,2}$, $\beta_{I,3}$, for one I level at a time ---but subject to the constraint that the estimates for 4, 5, 6 must be the same for all I levels. The require computation is greatly reduced in this way. Additional computational savings can be had by using the data associated with only a subset of the I-levels. Then 3-D object location parameters and other 3-D unknowns can be solved for using the coefficient found in (6). See [2] for more detail.

REFERENCES

[1] R. M. Bolle, D. B. Cooper and B. Cernuschi-Frias, "Three Dimensional Surface Shape Recognition by Approximating Image Intensity Functions with quadric Polynomials," Brown University Division of engineering Tech Report revised March 1983; a short early version appears in Proc. IEEE Conf. on Pattern Recog. and Image Processing, June 1982, pp. 611-617.

[2] B. Cernuchi-Frias, D. B. Cooper and R. M. Bolle, "Location and Orientation Parameter Estimation of Lambertian Spheres and cylinders in 3-D Space Using a Single 2-D Image," Brown University Division of Engineering Tech Report revised January 1983; a short early version appears in Proc. IEEE Conf. on Pattern Recog. and Image Processing, June 1982, pp. 604-610.

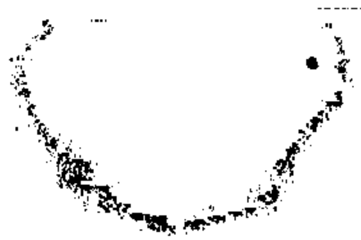

FIG. 1a

FIG. 1b

FIG. 2a

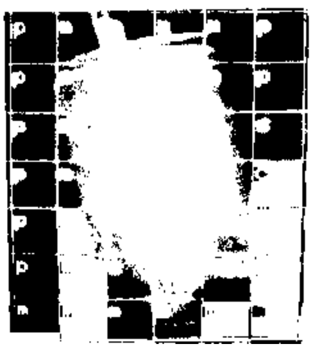

FIG. 2b