Optimal Algorithms for Image Understanding:

Current Status and Future Plans

D. Lee

Department of Computer Science
Columbia University
New York, N. Y. 10027

April 1985

This research was supported in part by the National Science Foundation under Grant DCR-82-14322 and in part by an IBM graduate fellowship.

Abstract

We discuss four image understanding problems: 2 1/2 D sketch, three-dimensional reconstruction from projections, shape from shading, and optical flow. We point out how known general optimality results may be applied to the first three problems. We indicate some preliminary results and work in progress, concerning the numerical solution of all four problems. Algorithms which differ from those currently used in practice are proposed.

1. Introduction

Information-based complexity is a field whose goal is to solve problems optimally with limited or contaminated information. We apply the optimality results of the theory to *image understanding* problems. We also discuss numerical algorithms for these problems. We first review briefly the relevant part of the theory and then discuss the following problems in turn: *2 1/2 D sketch, three-dimensional reconstruction from projections, shape from shading,* and *optical flow.*

We approximate an element f in a normed space F, with norm $\|\cdot\|$. We have *information* about f: $N(f) = \{L_1(f), \cdots, L_n(f)\}$, where L_i is a functional on F, and N: $F \rightarrow R^n$. An *algorithm* ϕ uses this information to construct an approximation $\phi(N(f)) \in F$, where ϕ: $R^n \rightarrow F$ is an arbitrary mapping. Since any $f^\wedge$, such that $N(f^\wedge) = N(f)$, could be the element we want to approximate, and since in practice, we only consider $f^\wedge$ in a subset of **F**, F_0, the *worst case algorithm error* is defined as

$$e(\phi,N,f) = \sup_{\substack{f^{\sim} \in F_0 \\ N(f^{\sim})=N(f)}} \|f^{\sim} - \phi(N(f))\|, \quad f \in F. \tag{1}$$

We seek a *strongly optimal algorithm*, which minimizes the algorithm error, among all algorithms, for each f in F.

In the following discussion of image understanding problems, we use the notation and terminology given above.

2. 2 1/2 D Sketch

2 1/2 D sketch is to recover the surface based on a finite set of depth values, obtained from direct ranging, binocularity etc., see Grimson [81].

We assume that the class of real world surfaces F is smooth and is viewed from a position free of accidental alignments. Formally, $F = \{f: D \to R,\ D \subseteq R^2$, f and its first and second order partial derivatives are all square integrable$\}$. Information is depth data:

$$N(f) = [L_1(f), \cdots, L_n(f)] = [f(x_1,y_1), \cdots, f(x_n,y_n)]. \tag{2}$$

A natural constraint from practice is the *surface consistency constraint*, which states that the surface cannot change in a radical manner between known data. The change of a surface is quantified by its variation θ, defined as

$$\theta(f) = \{\int_D \{ (f_{xx})^2 + 2(f_{xy})^2 + (f_{yy})^2 \}dx\ dy \}^{1/2}. \tag{3}$$

We confine ourselves to the class of surfaces F_0, which has bounded surface variation. Without loss of generality, we assume that the bound is 1, i.e., $F_0 = \{f \in F: \theta(f) \leq 1\}$.

Grimson [81] further explored the surface consistency constraint and proposed *spline interpolation*, which interpolates the data and minimizes θ. This is the *spline algorithm*. It is known (see e.g. Traub and Woźniakowski [80] ch.2 and 4.5) that

Proposition 2.1 The spline algorithm ϕ^s is strongly optimal and linear. It has the form

$$\phi^s(N(f)) = \sigma_{N(f)} = \sum_{i=1}^{n} f(x_i, y_i)\, \sigma_i \,, \tag{4}$$

where σ_i is a *basis spline*, i.e., a function of minimal surface variation such that $\sigma_i(x_j, y_j) = \delta_{i,j}$, and $\delta_{i,j}$ is the Kronecker delta.

Constructing the spline is important in practice and has been the subject of much work, see Grimson [81] and Terzopoulos [84].

We now briefly discuss two different approaches. From (4), the spline is a linear combination of basis splines σ_i , and the coefficients are known depth values. The basis splines are data independent and can be precomputed. However, for large n, it may not be feasible to compute and store all basis splines because of time and space limitations. To implement the spline algorithm using this approach, one has to further explore efficient ways of storing and retrieving the basis splines.

Another approach is to use the *reproducing kernel*, where one has to solve a large system of linear equations. The coefficient matrix is dense, but regularly structured. For a regular grid, it is a *block Toeplitz* matrix. Based on the work by Meinguet [83], Boult [85] is seeking an efficient numerical solution.

The information in (2) is depth data. This is *nonadaptive information*, since the sampling location (x_i, y_i) of the *i*th depth value does not depend on the previously computed (i-1) depth values. If the *i*th sampling location *does* depend on the (i-1) depth values obtained, then we call it *adaptive information*. More precisely, adaptive information is defined as

$$N^a(f) = z = [z_1, \cdots, z_n], \tag{5}$$

where $z_i = f(x_i, y_i)$ and

$$x_i = x_i(z_1, \cdots, z_{i-1}),\ y_i = y_i(z_1, \cdots, z_{i-1}), \quad i = 2, \cdots, n.$$

Adaptive information has a richer structure than nonadaptive information. One might hope that the previously computed (i-1) depth values supply additional information for determining where to sample for the *i*th depth value. Counter-intuitively, adaptive information *cannot* aid 2 1/2 D sketch and some other image understanding problems, see Traub and Woźniakowski [80] ch.2 and Traub, Wasilkowski and Woźniakowski [83] ch.4. Therefore, in seeking the best places to sample, we can confine ourselves to nonadaptive information only, which is simple and favorable for *parallel* or *distributed* computation.

3. Three-dimensional Reconstruction from Projections

The problem of reconstructing three-dimensional objects from a set of two-dimensional projected images has arisen and been studied independently in fields ranging from medicine and electron microscopy to holographic interferometry. By using a source of radiation external to the object, we obtain a transmission picture of projection of the three-dimensional object onto a two-dimensional surface such as the film of an ordinary electron micrograph or x-ray. The reconstruction problem is: given a set of projections of an object, estimate its internal density distribution. Much work has been done, see Gordon and Herman

[74] for a survey and also Logan and Shepp [75]. We propose using a different algorithm, which is provably optimal, and we also briefly discuss its implementation.

We assume that the class of density distributions F consists of smooth functions supported on the unit disc D in the *xy*-plane, i.e., F = {f: D → R, the second order partial derivatives of f are square integrable}. As in Section 2, we observe the consistency constraint by confining ourselves to a subset of elements in F, i.e., $F_0 = \{f \in F: \theta(f) \leq 1\}$, where θ is given in (3). Information is given by the projections of f along the line: $x \cos\xi + y \sin\xi = t$. More precisely, $N(f) = \{L_{i,j}(f)\}_{i,j}$, where the linear functional $L_{i,j}$ is

$$L_{i,j}(f) = \int_{-1}^{1} f(t_j \cos\xi_i - s \sin\xi_i \, , \, t_j \sin\xi_i + s \cos\xi_i) \, ds, \tag{6}$$

where $0 \leq \xi_i < \pi$, $-1 < t_j < 1$, $i = 1, \cdots, m$, $j = 1, \cdots, n$.

As in Section 2, it is known that

Proposition 3.1 The *spline algorithm* ϕ^s, which interpolates the data and minimizes θ, is strongly optimal and linear. It has the form

$$\phi^s(N(f)) = \sigma_{N(f)} = \sum_{i=1}^{m} \sum_{j=1}^{n} L_{i,j}(f) \, \sigma_{i,j} \, . \tag{7}$$

where $\sigma_{i,j}$ is a *basis spline*, i.e., a function of minimal variation such that $L_{i,j}(\sigma_{k,l}) = \delta_{i,k} \, \delta_{j,l}$, and $\delta_{i,j}$ is the Kronecker delta.

Here, as in Section 2, there are a number of alternatives for constructing the spline. One could use precomputation of the basis splines. Then the discussion of Section 2 applies.

Another possible approach is to use the reproducing kernel. It is known, Atteia [70], that there exists a reproducing kernel $\omega(x,y \; ; \; u,v)$, defined on DxD, which is useful for representation of splines. An analytic form of the reproducing kernel is derived in Atteia

[70], which also includes an ALGOL60 program for its numerical computation. The spline is then of the form

$$\sigma_{N(f)}(u,v) = \sum_{k,l} a_{k,l} \, L_{k,l} \, (\omega(\cdot \, , \, \cdot \, ; \, u \, , \, v)). \tag{8}$$

Therefore, we only have to compute the coefficients $a_{i,j}$ in (8). Applying $L_{i,j}$ to both sides of (8); since $\sigma_{N(f)}$ interpolates the projection data, we have

$$\sum_{k,l} L_{i,j}(g_{k,l}) \, a_{k,l} = L_{i,j}(f), \tag{9}$$

where $g_{k,l}(u,v) = L_{k,l}(\omega(\cdot \, , \, \cdot \, ; \, u \, , \, v))$. Thus the problem is reduced to solving a system of linear equations in $a_{k,l}$. If the number of projection data is not large, then this system can be easily solved. If it is very large, standard numerical methods are not feasible due to their high cost. Since the coefficient matrix is a well structured Gram matrix, it might be possible to devise an efficient numerical algorithm for solving (9).

4. Shape from Shading

Research in shape from shading explores the relationship between image brightness and object shape. A great deal of information is contained in the image brightness values, since image brightness is related to surface orientation. Information can also be obtained from *occluding boundaries* and other *boundary conditions*, see Ikeuchi and Horn [81]. To determine surface orientations, Ikeuchi and Horn used the spline-smoothing approach and reduced the problem to solving a system of non-linear equations, and proposed an iterative algorithm for solving it. We discuss the numerical solution of this system of non-linear equations.

The relation between the surface orientation and brightness is specified by the *image-irradiance equation*:

$$R(\xi,\eta) = E(x,y), \tag{10}$$

where $\xi = \xi(x,y)$ and $\eta = \eta(x,y)$ represent the surface orientation at the point (x,y), E(x,y) is the brightness measured at the point (x,y), and $R(\cdot\ ,\ \cdot)$ can be determined experimentally or theoretically, see Horn and Sjoberg [79], and Nicodemus et al. [77]. The goal is to recover ξ and η.

Ikeuchi and Horn used *spline-smoothing*, see Laurent [72], and seek ξ and η, which minimize

$$\int_D \{ [(\xi_x)^2 + (\xi_y)^2 + (\eta_x)^2 + (\eta_y)^2] + \lambda[R(\xi,\eta) - E(x,y)]^2 \} \, dxdy, \tag{11}$$

where D is the image domain of the object and λ is a penalty parameter.

After discretization, with mesh size h, and applying the method of Lagrange multipliers, we have a system of non-linear equations in $\xi_{i,j}$ and $\eta_{i,j}$. An iterative method was proposed in Horn and Schunck [81] for solving this system of equations. The initial values are supplied by the boundary conditions, i.e., $\xi_{i,j}$ and $\eta_{i,j}$ are known if (i,j) belongs to the occluding boundaries or other boundary points. The existence and uniqueness of the solution remain a problem. Furthermore, the convergence of the iterative method has not been established.

We will use a new iterative algorithm, which, for a range of λ, converges to the unique solution of the system, see Lee [85]. For arbitrary λ, the convergence of the algorithm needs further study.

Let $N = h^{-2}$, where h is the mesh size. To implement the iterative algorithm, one has to multiply a dense matrix by a vector, with cost $O(N^2)$, using conventional matrix multiplication. However, we can use Fast Fourier Transforms (FFT) to reduce the cost to $O(N \log N)$. If the error bound is $O(h)$, then one performs ($\log$ N) iterative steps, with the total cost $O(N (\log N)^2)$.

When the data are noisy, the spline-smoothing approach is appropriate. However, when the data are relatively precise, the *interpolating spline* approach is preferable. In that approach, one seeks a spline, which interpolates the data and minimizes the first part of (11). This is an *interpolatory algorithm* and is therefore *almost strongly optimal*, i.e., strongly optimal within a factor of 2, see Traub and Woźniakowski [80] ch.1. The uniqueness of the spline and its construction need further investigation.

5. Optical Flow

Biological systems typically move relatively continuously through the world, and the images projected on their retinas vary essentially continuously while they move. Such continuous flow of the imaged world across the retina is called *optical flow*. The optical flow assigns to every point on the visual field a two-dimensional "retinal velocity", at which it is moving across the visual field. We study approximation of the optical flow, or *velocity field*, based on a sequence of images.

Assume that D is a finite image domain of interest. We denote the image brightness, projected by a point on a moving object at time t, by $E(x,y,t)$, and the velocity field by $(u(x,y,t),v(x,y,t))$. Then we have (Cornelius and Kanade [83], Horn and Schunck [81]):

$$w(x,y,t) = p(x,y,t)\, u(x,y,t) + q(x,y,t)\, v(x,y,t) + r(x,y,t), \tag{12}$$

where $p = \partial E/\partial x$, $q = \partial E/\partial y$ and $r = \partial E/\partial t$ can be computed directly. The function w is the total rate of change of brightness, which is not known.

From (12) alone, one cannot determine u, v and w uniquely. Assume that the partial derivatives of u, v and w are square integrable. In addition to requiring that (12) be satisfied, a *consistency constraint* is imposed in Cornelius and Kanade [83] and Horn and Schunck [81], which is the minimization of the following:

$$\int_D [(u_x)^2 + (u_y)^2 + (v_x)^2 + (v_y)^2]\, dx\, dy \quad \text{and} \int_D [(w_x)^2 + (w_y)^2]\, dx\, dy. \qquad (13)$$

Then u, v and w are uniquely determined. We discuss two approaches for approximating u, v and w in Subsections 5.1 and 5.2, respectively: *spline-smoothing* and *interpolating spline*. We approximate the velocity field at an arbitrary instance t_0 , and we omit the time factor t in the following discussion.

5.1. Spline-smoothing

This approach was used by Horn and Schunck [81] (where $w \equiv 0$), and by Cornelius and Kanade [83]. They seek (u,v,w), which minimizes

$$\int_D \{ \lambda[(u_x)^2 + (u_y)^2 + (v_x)^2 + (v_y)^2] + \mu[(w_x)^2 + (w_y)^2] \qquad (14)$$

$$+ [pu + qv + r - w]^2 \}\, dx\, dy,$$

where λ and μ are penalty parameters.

After discretization, with mesh size h, to solve the minimization problem, one has to solve a system of linear equations. To solve this system of linear equations, the Gauss-Seidel iterative method was proposed in Cornelius and Kanade [83] and Horn and Schunck [81]. The convergence of the algorithm remains to be analyzed, since the coefficient matrix is not symmetric. Furthermore, even if the Gauss-Seidel iterative method converges for this case, it converges slowly.

We propose using the *conjugate gradient* iterative method, see Hageman and Young [81] ch.7, based on the following observations. With appropriate algebraic manipulations, without increasing the condition number, we can reduce the problem to solving a system of linear equations with symmetric and positive definite coefficient matrix, which is sparse, see Lee [85]. The conjugate gradient method converges much faster than Gauss-Seidel and is *strongly optimal*, see Traub and Woźniakowski [84].

On the other hand, algorithms with *simple, local* and *parallel* operations are preferable in image understanding, since they are suitable for parallel computation and feasible for a biological system. Conjugate gradient method requires global interaction, which might not be desirable.

We estimate the minimal and maximal eigenvalues of the matrix, and its condition number, see Lee [85]. We can use the *Chebyshev method*, which also converges much faster than Gauss-Seidel and involves only simple, local and parallel operations. Chebyshev method is also optimal, see Traub and Woźniakowski [84]. For Chebyshev methods, see Hageman and Young [81] ch.4 - 6 and Appendix A, which includes FORTRAN subroutines.

5.2. Interpolating Splines

In this approach, we seek (u,v,w), which satisfy the information constraint (12) exactly and minimize

$$\int_D \{ [(u_x)^2 + (u_y)^2 + (v_x)^2 + (v_y)^2] + \lambda[(w_x)^2 + (w_y)^2] \} \, dx \, dy, \tag{15}$$

where λ is a chosen parameter. (There is a short discussion of this approach, for $w \equiv 0$, in Horn and Schunck [81]).

After discretization, and applying the method of Lagrange multipliers, we have to solve a system of linear equations. The coefficient matrix is neither symmetric nor positive definite. However, we can reduce the problem, without increasing the condition number of the coefficient matrix, to solving a system of linear equations, with symmetric and positive definite coefficient matrix. Therefore, the conjugate gradient iterative method can be used. We also estimate the minimal and maximal eigenvalues of the matrix, and its condition number, see Lee [85]. Therefore, as before, the Chebyshev method can be used.

The coefficient matrix is dense, and each iterative step requires multiplying this matrix by a vector, which could cost $O(N^2)$, where $N = h^{-2}$. However, we can use the FFT, to reduce it to $O(\ N\ (\log N)\ )$.

Acknowledgements.

My thanks are to J. R. Kender for his encouragement and advice during this work. I appreciate very much valuable comments from G. W. Wasilkowski.

References.

Atteia, M., (1970), Fonctions (spline) et noyaux reproduisants D'Aronszajn- Bergman, Rev. Francaise Informat, Rechereche Operationnelle 4E Annee R-3, 31-43.

Boult, T., (1985), Reproducing Kernel for Visual Surface Interpolation, Department of Computer Science Rep., Columbia University.

Cornelius, N., and Kanade, T., (1983), Adapting Optical-flow to Measure Object Motion in Reflectance and X-ray Image Sequences, Proceeding of Image Understanding Workshop, DARPA, July.

Gordon, R., and Herman, G. T., (1974), Three-dimensional Reconstruction from Projections: a Review of Algorithms, International Review of Cytology 38, 111-151.

Hageman, L. A., and Young, D. M., (1981), Applied Iterative Methods, Academic Press, New York.

Horn, B. K. P., and Sjoberg, R. W., (1979), Calculating the Reflectance Map, Applied Optics 18, 1770-1779.

Ikeuchi, K., and Horn, B. K. P., (1981), Numerical Shape from Shading and Occluding Boundaries, Artificial Intelligence 17, 141-184.

Horn, B. K. P., and Schunck, B. G., (1981), Determining Optical Flow, Artificial Intelligence 17, 185-203.

Laurent, P. J., (1972), Approximation et optimisation, Hermann, Paris.

Lee, D., (1985), Contributions to Information-based Complexity, Image Understanding and Logic Circuit Design, Ph. D Thesis, Department of Computer Science Rep., Columbia

University.

Logan, B. F., and Shepp, L. A., (1975), Optimal Reconstruction of a Function from its Projections, Duke Math. J. 42, 4, 645-659.

Meinguet, J., (1983), Surface Spline Interpolation: Basic Theory and Computational Aspects, Institut de Mathematique Pure et Appliquee, Universite Catholique de Louvain, Rapport n° 35.

Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T., (1977), Geometrical Considerations and Nomenclature for Reflectance, NBS Monograph 160, U. S. Department of Commerce, National Bureau of Standards.

Terzopoulos, D., (1984), Multiresolution Computation of Visible-surface Representations, Ph. D. Thesis, MIT Artificial Intelligence Lab.

Traub, J. F., and Woźniakowski, H., (1980), A General Theory of Optimal Algorithms, Academic Press.

Traub, J. F., Wasilkowski, G. W., and Woźniakowski, H., (1983), Information, Uncertainty, Complexity, Addison-Wesley, Reading, MA.

Traub, J. F., and Woźniakowski, H., (1984), On the Optimal Solution of Large Linear Systems, J. of ACM 31, 3, 545-559.