COMBINING SOURCES OF INFORMATION IN VISION
I. COMPUTING SHAPE FROM SHADING AND MOTION

John (Yiannis) Aloimonos

Center for Automation Research
University of Maryland
College Park, MD 20742

ABSTRACT

Most of the basic problems in computer vision, as formulated, admit infinitely many solutions. But vision is full of redundancy and there are several sources of information that if combined can provide unique solutions for a problem. Furthermore, even if a problem as formulated has a unique solution, in most cases it is unstable. Combining information sources in this case leads to robust solutions. In this paper, we combine shading and motion to uniquely recover the light source direction and the shape of the object in view, and we describe how one could combine information sources present in the image in order to achieve uniqueness and robustness of low level visual computations.

0. Prolegomena

Several problems in human or machine vision, as formulated, admit infinitely many solutions, because the available constraints that relate image properties to the unknown parameters are not enough to guarantee uniqueness. One approach to the solution of these problems is to regularixe them by requiring that the desired parameters satisfy some criteria (basically smoothness). Poggio et al. ([l], [2], [3]) have very successfully shown that several problems in early vision may be "solved" by standard regularization techniques.

Another approach to the solution of ill-posed problems would be to look for more information sources in order to augment the number of physical constraints and achieve uniqueness of the parameters to be computed. For example, we can combine shading with retinal motion to compute the illuminant source and the shape of the object in view, combine shading with stereo, stereo with motion, texture with contour, and so on, in order to compute three-dimensional scene properties. The results of such an approach are summarized in Table 1 for reference. In the ellipses (top) are the different image cues (we take the liberty of calling stereo or motion a cue). By. a cue we mean a source of information, either coming from the image(s) or from the particular set-up or condition of the visual system (stereo, motion). In the squares are the results that can be obtained (in terms of propositions) when information is combined from two or more cues. Two or more different cues are combined with arcs which lead to small circles containing pluses. Then, an arc leads from the "plus" to <u>a square containing the result</u> obtained from this combination. The interested reader is referred to [4]. IN this paper we study only the combination of shading and motion for the computation of the lighting direction and shape.

1. Process of image formation

The ability to obtain three-dimensional shape from two dimensional intensity images is an important part of vision. The human visual system in particular is able to use shading cues to infer changes in surface orientation fairly accurately, with or without the aid of texture or surface markings. The direction of illumination is required to be known in order to obtain accurate three-dimensional surface shape from two dimensional shading because changes in image intensity are primarily a function of changes in surface orientation relative to the illuminant, for many of the surfaces in our visual environment. For the purposes of this paper we use the following simple and universally accepted model. Assuming orthographic projection, if $\vec{n}$ is the surface normal at a point on the imaged surface, *T* is the illuminant direction, and f is the flux emitted towards the surface, and we assume a Lambertian reflectance function for the surface, the image intensity function is given by $i = \rho f(\vec{n} \cdot \vec{l})$, where *p* is the albedo of the surface, a constant depending on the surface, and "•" denotes the vector inner product.

2. Motivation and Previous Work

Most of the work in shape from shading assumes that the albedo of the surface in view and the illuminant direction are known *a priori*. The only work done on illuminant direction determination is due to Pentland [5], Brooks and Horn [6], Brown and Ballard (7), and Lee and Rosenfeld [8].

It is easy to prove that the illuminant direction cannot be recovered from only one intensity image of a Lambertian surface without other assumptions. In the sequel, we prove that from two intensity images (moving object or moving observer), with the correspondence between them established, we can uniquely recover the illuminant direction, and thence the shape. In this paper we do not deal with the computation of correspondence (retinal motion). We assume that it has been obtained by some method such as those described in the recent literature. See (4) for detailed references.

3. Technical Prerequisites

In this section we develop two technical results, one concerning the relationship among shape, intensity, displacements, and the lighting direction, and the other concerning the parameters of a small motion (small rotation) of the shape. We begin with the following theorem.

Theorem It Suppose that two views (under rigid motion) of the same Lambertian, locally planar surface are given; let I_1

A part of this work wss done when the author was with the University of Rochester, NY, under an IBM predoctoral fellowship. The work was completed at the Center for Automation Research of the University of Maryland under Grant DAAB07-86-K-F073 of the Defence Advanced Research Projects Agencyand the U.S. Army Night Vision and Electro-Optics Laboratory. The comments of Professors A. Rosenfeld, C. Brown, D. Ballard, J. Feldman and L. Davis are highly appreciated.

and I_2 be the two intensity functions. Suppose also that the displacement vector field $\left(u(x,y),\, v(x,y)\right)$, $(x,y)\in\Omega$ is known, where Ω is the domain of the image, i.e. a point (x,y) in the first image moves to the point $\left(x+u(x,y),\, y+v(x,y)\right)$ in the second image. If the lighting direction is $\bar{l} = (l_1,l_2,l_3)$ and the gradient of the surface point whose image is the point (x,y) is (p,q), then the following relation holds:

$$A_1p^2+B_1q^2+C_1pq+D_1p+E_1q+F_1 = 0, \qquad (1)$$

with

$$A_1 = \left(l_2\Delta^y u - l_1(1+\Delta^y v)\right)^2 - r^2 l_1^2$$

$$B_1 = \left(l_1\Delta^x v - l_2(1+\Delta^x u)\right)^2 - r^2 l_2^2$$

$$C_1 = 2\left[\left(l_1\Delta^x v - l_2(1+\Delta^x u)\right)\left(l_1(1+\Delta^y v) - l_2\Delta^y u\right) - 2r^2 l_1 l_2\right]$$

$$D_1 = 2l_1l_3r\left(r - \left((1+\Delta^x u)(1+\Delta^y v) - \Delta^x u\,\Delta^y v\right)\right)$$

$$E_1 = 2l_2l_3r\left(r - \left((1+\Delta^x u)(1+\Delta^y v) - \Delta^y u\,\Delta^x v\right)\right)$$

$$\begin{aligned} F_1 = &-\left((1+\Delta^x u)(1+\Delta^y v) - \Delta^y u\,\Delta^x v\right)^2 \\ &+ l_1^2\left((\Delta^x v)^2 + (1+\Delta^y v)^2\right) + l_2^2\left((\Delta^y u)^2 + (1+\Delta^x u)^2\right) \\ &+ l_1l_2\left((1+\Delta x^x)\Delta^x v + \Delta^y u(1+\Delta^y v)\right) \\ &- r^2l_3^2 - 2rl_3^2\left((1+\Delta^x u)(1+\Delta^y v) - \Delta^y u\,\Delta^x v\right) \end{aligned}$$

where

$$r = \frac{I_2\left(x+u(x,y),\, y+v(x,y)\right)}{I_1(x,y)} \text{ and}$$

$$\Delta^x u = u(x+1,y) - u(x,y)$$

$$\Delta^y u = u(x,y+1) - u(x,y)$$

$$\Delta^x v = v(x+1,y) - v(x,y)$$

$$\Delta^y v = v(x,y+1) - v(x,y).$$

It is clear that equation (1) is local, i.e. it involves the gradient at a point (x,y) and the displacements around the point (x,y), along with the global direction of lighting.

Proof: See [4].

We now proceed to state the following theorem.

Theorem 2: Suppose that the surface S (locally planar) is moving with a rigid motion, and the camera model is the one described in the previous theorem. Let the gradient of the surface (with respect to the first frame) be $\left(p(x,y), q(x,y)\right)$. It is known that this motion can be considered as a translation (dx,dy,dz) plus a rotation by an angle θ about an axis $(n_1^2+n_2^2+n_3^3 = 1)$. If the rotation angle θ is small, then the following relations hold:

(a) The displacement vector field $\left(u(x,y), v(x,y)\right)$ is given by

$$u(x,y) = dx + Bz(x,y) - Cy$$

$$v(x,y) = dy + Cx - Az(x,y)$$

where $A = n_1\theta$, $B = n_2\theta$, $C = n_3\theta$, and $z(x,y)$ is the depth of the surface point whose projection is the image point (x,y).

(b) $p(x,y) = \dfrac{1}{B}\Delta^x u$

$q(x,y) = \dfrac{1}{A}\Delta^y v$

$$\frac{A}{B} = \frac{\Delta^y v}{\Delta^x u}\,\frac{\Delta^x v - \Delta^y u + \sqrt{(\Delta^y u + \Delta^x v)^2 - 4\Delta^x u\,\Delta^y v}}{2}\,\frac{\Delta^x v}{\Delta^y v}$$

Proof: See [4].

4. Development of the Motion/Illumination Constraint

If we let $1/A = a$ and $1/B = b$ and $B/A = k$ and use part (b) of Theorem 2 to substitute in equation (1) for p and q, we get the following equation:

$$\begin{aligned} &(\Delta^x u)^2 b^2\left[\left(l_2\Delta^y u - l_1(1+\Delta^y v)\right)^2 - r^2 l_1^2\right] + \\ &+ 2\Delta^x u\,\Delta^y v\,K\,b^2\left[\left(l_1\Delta^x v - l_2(1+\Delta^x u)\right)\left(l_1(1+\Delta^y v) - l_2\Delta^y u\right) - 2 \right. \\ &\left. r^2 l_1 l_2\right] + (\Delta^y v)^2 K^2 b^2\left[\left(l_1\Delta^x v - l_2(1+\Delta^x u)\right)^2 - r^2 l_2^2\right] - \\ &- 2(\Delta^x u) b\, l_1 l_3 r\left(r - (1 + \Delta^x u + \Delta^y v + \Delta^x u\Delta^y v - \Delta^y u\Delta^x v)\right) \\ &- 2(\Delta^y v) K\, b\, l_2 l_3 r\left(r - (1 + \Delta^x u + \Delta^y v + \Delta^x u\Delta^y v - \Delta^y u\Delta^x v)\right) - \qquad (2) \\ &- (1 + \Delta^x u\Delta^y v + \Delta^x u + \Delta^y v - \Delta^y u\Delta^x v) + \\ &+ l_1^2\left((\Delta^x v)^2 + (1+\Delta^y v)^2\right) + l_2^2\left((\Delta^y u)^2 + (1+\Delta^x u)^2\right) + \\ &+ l_2 l_1\left((1+\Delta^x u)\Delta^x v + \Delta^y u(1+\Delta^y v)\right) - r^2 l_3^2 + \\ &+ 2r\, l_3^2(1 + \Delta^x u + \Delta^y v + \Delta^x u\Delta^y v - \Delta^y u\Delta^x v) = 0 \end{aligned}$$

Equation (2) is a polynomial in l_1,l_2,l_3 and b. Considering equation (2) at four points we get a polynomial system of four equations in four unknowns. These equations are independent. This can be formally proved using the Jacobian of the system. A simple but tedious calculation of the Jacobian of this system shows that the Jacobian is non-zero (except in degenerate cases). This means (inverse function theorem), that the function defined by the equations of the system is locally an isomorphism, which means that its zeros are isolated. But from Whitney's theorem, the set of zeros of this algebraic system is an algebraic set and as such it has finitely many components. The conclusion is that the set of solutions of the system is finite (uniqueness). If we now consider equation (2) at five points, we get a system of five equations in four unknowns, which barring degeneracy will have at most one solution.

It is clear from the above discussion that two intensity images of a Lambertian surface, with the correspondence between them established, gives the illumination direction uniquely. In the next section we present an algorithm for the recovery of the light source direction, based on the constraint developed in this section.

5. The Algorithm for Finding Illuminant Direction

We use the Gaussian sphere formalism (azimuth-elevation) to represent the vector that denotes the light source direction. More specifically we set

$l_1 = \cos\phi\cos\theta$
$l_2 = \sin\theta$
$l_3 = \sin\phi\cos\theta$

where θ and ϕ are the azimuth and elevation. Now we consider equation (2) at n points in the images, and we get n equations $eq_1, eq_2, \ldots, eq_n$ in the unknowns b, θ, ϕ. Then the following algorithm gives a solution:

```
for all 9
    for all $\phi$
        {get n quadratic equations in b. Check if
        they have a common solution. If yes, output $\theta,\phi$.}
```

We have implemented the above algorithm and it works successfully for synthetic images. Due to lack of space we do not describe them. The interested reader is referred to [4].

6. Computing Shape from Shading and Motion

This section discusses the problem of uniquely determining shape from shading and motion. Before we proceed, we need some technical preliminaries.

0.1. The Motion Constraint

Theorem 3: With the assumptions and notation of Theorem 1, the gradient (p,q) of a surface point whose image is the point (x,y) with displacement vect (u,v) isfies the constraint

$$A_2p^2+B_2q^2-2C_2pq+D_2=0 \tag{3}$$

with

$$A_2=\left(\Delta^y u(x,y)\right)^2+\left(\Delta^y v(x,y)\right)^2+2\Delta^y v(x,y)$$

$$B_2=\left(\Delta^x u(x,y)\right)^2+\left(\Delta^x v(x,y)\right)^2+2\Delta^x u(x,y)$$

$$C_2=\Delta^y u(x,y)+\Delta^y u(x,y)\Delta^x u(x,y)+\Delta^x v(x,y)\Delta^y v(x,y)$$

$$D_2=C_2^2-A_2B_2.$$

Proof: See [4|.

6.2. How to Utilise the Constraints

Now we show how to utilize the constraints to recover the three-dimensional shape of the object in view, using shading and motion. It is assumed that the illuminant direction has already been computed using the algorithm described in the previous sections. Up to now, we have developed two constraints on shape that also involve the displacements and the illuminant direction. The *motion/illumination* constraint is a conic in p and *q* with coefficients that depend on the relative intensities, the displacements, and the illuminant direction (eq. 1). The *motion* constraint is again a conic in *p* and *q,* with coefficients that depend on the displacement vectors (eq. 2). Finally, the image irradiance equation

$$I=\rho\vec{n}\cdot\vec{l} \tag{11}$$

which determines the intensity $I(x,y)$ at a point (x,y) of the image as a function of the shape n of the world point whose image is the point (x,y), the illuminant direction $\vec{l}$, and the albedo ρ of the surface, is another constraint on p and *q* that is also a conic. In the sequel we will call the above constraint *the image irradiance* constraint. It is worth summarizing at this point that we have at our disposal three constraints on the shape *p,q* of a point whose image is the point (x,y); each of these constraints is a conic in gradient space. Figure 1 gives a geometrical description of the constraints.

6.3. Computing Shape When the Albedo is Known

When the albedo is known, we have at our disposal three constraints at every image point for the computation of shape: the *motion/illumination* constraint, say $F_1(p,q)=0$; the motion constraint, say $F_2(p,q)=0$; and the *image irradiance* constraint, say $F_3(p,q)=0$. These are three equations, each of degree two, and their system, barring degeneracy, will have at most one solution.

The solution can be easily found by solving the system of equations $F_i(p,q)=0,\ i=1,2,3.$ But if the input is noisy, since this method is local, i.e. shape is computed at every point separately, we might get results corrupted by noise. To avoid this, we adopt an iterative technique, at the expense of assuming that the surface in view is smooth. We prove, however, that our method will converge to a unique solution. We assume that we know the shape at the occluding boundaries. Because of the fact that gradient space is unbounded, we use stereographic shape coordinates (ξ,n). In this case, all possible orientations are contained in a disc of radius two and (ξ,n) represents an orientation at the occluding boundary iff $\xi^2+n^2=4$. Furthermore, assuming that orientations (ξ,n) with $\xi^2+n^2=4$ occur only at the boundaries and substituting the values of p and q in terms of ξ and n from

$$p=\frac{4\xi}{4-\xi^2-n^2}$$

and

$$q=\frac{4n}{4-\xi^2-n^2}$$

in $F_i(p,q)=0,\ i=1,2,3,$ we get equations i (ξ,n) a t are polynomials in £ and n. Let us denote them by $G_i(\xi,n,x,y)=0,\ i=1,2,3,$ where for i = 1 we have the motion/illumination constraint, for $i=2$ the motion constraint and for t = 3 the image/irradiance constraint. The reason why *G₁* is a function of *x* and *y* as well (where x,y are the image coordinates) is that the coefficients in $G_i=0$ for $i=1,2$ are functions of position in the image plane.

Trying to find a solution that satisfies $G_i=0$ for i = 1, 2, 3 and is smooth, we wish to minimize the functional

$$e=\int_D\!\!\int\left\{\lambda(G_1^2+G_2^2+G_3^2)+(\xi_x)^2+(n_x)^2+(\xi_y)^2+(n_y)^2\right\}dx\,dy$$

where *D* is the unit square region in the $x-y$ plane (where the image appears) with mesh size m; the surface in view is assumed to have derivatives that are square integrable; and λ is a constant (regularization) that weights the relative importance of the constraints and smoothness. We follow a standard numerical analysis technique, widely used in the area of partial differential equations [9, 10]. We discretize e by using difference operators instead of differential ones and summations instead of integrals. Let $n=k^2$ and $k+1=1/m$. The desired surface is the one that minimizes

$$e=\sum_{ij}(s_{ij}+\lambda g_{ij})$$

where

$$s_{ij}=\frac{1}{m^2}\Big[(\xi_{i+1,j}-\xi_{i,j})^2+(\xi_{i,j+1}-\xi_{i,j})^2+(n_{i+1,j}-n_{i,j})^2+(n_{i,j+1}-n_{i,j})^2\Big]$$

$$g_{ij}=G_1^2(\xi_{ij},n_{ij},i,j)+G_2^2(\xi_{ij},n_{ij},i,j)+G_3^2(\xi_{ij},n_{ij})$$

and where ξ_{ij}, n_{ij} represent the surface orientation at the regular grid point (im,jm). This minimization is subject to boundary conditions, i.e. ξ_{ij}, n_{ij} are known if (im,jm) belong to the boundary. We assume that the surface normal at a boundary point (i,j) is parallel to the image plane (i.e. $\xi_{i,j}^2 + n_{i,j}^2 = 4$). We have assumed for simplicity and without loss of generality that the boundary is square. The analysis for a nonsquare boundary is very involved (using finite element methodology), and it will not be presented here.

We can prove that if we choose λ such that

$$\lambda m^2 \left[2\pi^2 m^2 \left(1 - \frac{\pi_2 m^2}{24} \right)^2 \right]^{-1} \mu < 1,$$

then there exists at most one solution for the minimization problem.

6.3.1. Learning the regularised algorithm

In the previous section we gave an algorithm for computing shape, using the actual constraints and the smoothness of the surface in view. The results depend on the choice of the parameter λ, and even though any choice of λ in an interval will result in a convergent algorithm, it does not mean that the solutions we will get will reflect the reality (physical plausibility). To achieve this, we would have to train the system with examples, so that the best λ is chosen. A method that can be used for this is the cross-validation technique [12]. Another can be found in [13].

0.4. Computing Shape When the Albedo is Not Known See |4].

0.5. Implementation and Experiments

Due to lack of space we do not describe our experimental results. The interested reader is refered to [4].

7. Conclusions and Future Directions

We have presented a theory of how to compute the illuminant direction and shape uniquely from shading and retinal motion. Our input consists of two intensity images in a dynamic sequence, with the correspondence between the two frames established. In the past, only the work of Grimson [15] has combined shading with another source of information, in particular stereo.

It is clear from the preceding sections that a similar theory can be developed when very general (i.e. not only Lambertian) reflectance maps are used. It is one of our future goals to extend our theory to reflectance maps that model the illumination due to the sun and sky.

References

1 Poggio, T and C Koch, "An analog model of computation for the ill-posed problems of early vision", Massachusetts Institute of Technology, Artificial Intelligence Laboratory Memo 783, C B I P paper 002, 1984

2 Poggio, T, V Torre, and C Koch, "Computational vision and regulanzation theory", *Nature* 317, 314-319, 1985

3 Poggio, T , "Early vision from computational structure to algorithms and parallel hardware", *Computer Vision, Graphics and Image Processing* 31, 139-155, 1985

4 Aloimonos, J, "Computing intrinsic images", Ph D thesis, Dept of Computer Science, University of Rochester, 1986

5 Pentland, A, "Finding the illuminant direction", j *Optical Society of America* 71, 448-455, 1982

6 Brooks, M and B K P Horn, "Shape and source from shading", MIT AI Lab Memo, 1985

7 Brown, C M , D Ballard, and 0 A Kimball, "Constraint interaction in shape from shading algorithms", *Proc. Image Understanding Workshop,* 79-89, 1982.

8 Lee, C H and A Rosenfeld, "Improved methods of estimating shape from shading using the light source coordinate system", *Artificial Intelligence* 26, 125-143, 1986

9 Lee, D, "A provably convergent algorithm for shape from shading", Proc *Image Understanding Workshop,* 489-496, Miami Beach, FL, December 1985

10 Smith, G D , *Numerical Solution of Partial Differential Equations Finite Difference Methods"*Oxford University Press, 1978

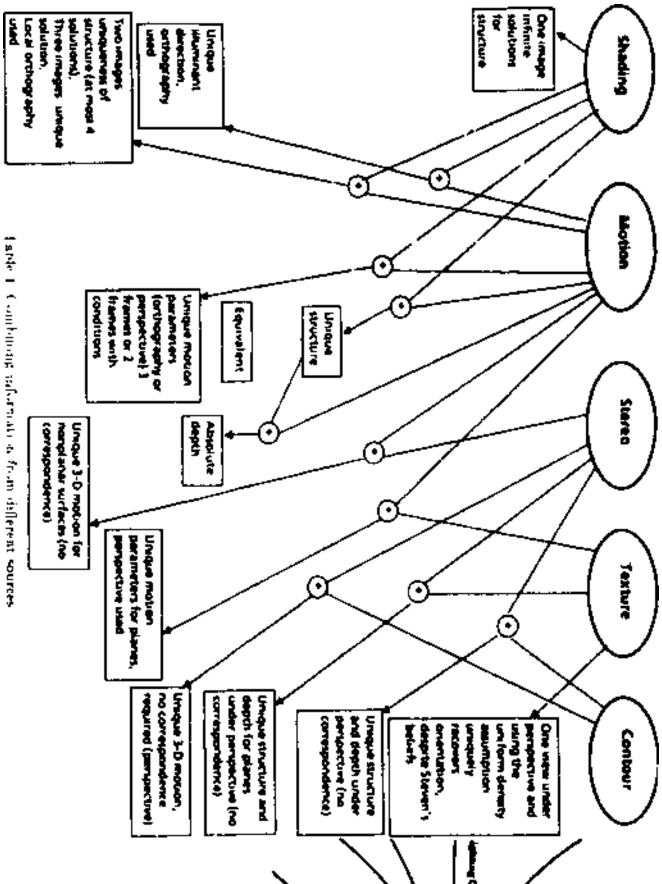
11 Morozov, V A , *Methods for Solving Incorrectly Posed Problems,* Springer, New York, 1984

12 Wahba, G , "Cross-validated spline methods for the estimation of functions from data on functionals", in HA and HT David (Eds), *Statistics: An Appraisal,* Iowa State University, 1984

13 Aloimonos, J and Shulman, D, "Learning shape computations", forthcoming Technical Report, Center for Automation Research, University of Maryland

14 Tichonov, AN and VY Arsenin, *Solution of Ill-Posed Problems,* Winston and Wiley, Washington, D C , 1977

15 Grimson, E, "Binocular shading and visual surface reconstruction", *Computer Vision, Graphics and Image Processing* 28, 19-44, 1984



Shading
Motion
Stereo
Texture
Contour
One image infinite solutions for structure
Unique illuminant direction, orthography used
Two images uniqueness of structure (at most 4 solutions), Three images unique solution, Local orthography used
Unique motion parameters (orthography or perspective) 3 frames or 2 frames with conditions
Equivalent
Unique structure
Absolute depth
Unique 3-D motion for nonplanar surfaces (no correspondence)
Unique motion parameters for planes, perspective used
Unique 3-D motion, no correspondence required (perspective)
Unique structure and depth for planes under perspective (no correspondence)
Unique structure and depth under perspective (no correspondence)
One view under perspective and using the uniform density assumption uniquely recovers orientation, despite Stevens's beliefs

Table 1 Combining information from different sources

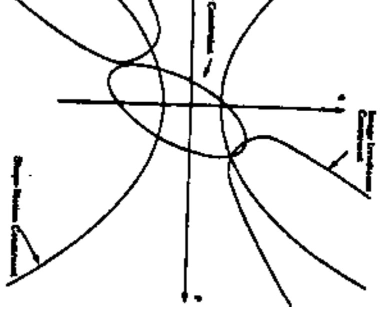


Figure 1 Pictorial description of the constraints