3D SHAPE ESTIMATION IN COMPUTER VISION

José R.A. Torreão

Instituto de Computação
Universidade Federal Fluminense
24210-240 Niterói RJ, BRAZIL
jrat@caa.uff.br

Abstract: 3D shape estimation is an essential capability of human vision, which several computer vision processes try to emulate. The goal of this article is to present some of these processes, starting with very basic material, but also providing pointers to more recent and sophisticated approaches. We treat stereoscopy, structure from motion, shape from shading and photometric stereo – a choice of topics that does not evidently exhaust the subject, but which should give the novice a fair introduction to this thriving area of research.

1 INTRODUCTION

The goal of computer vision research can be easily stated as that of creating machines which are able to see. A little less easy would be to define what such seeing ability exactly means. For our purposes here, it suffices to state that it entails the capacity of extracting, from images, the information that would allow the correct interpretation of the depicted scene. In the most general case of three-dimensional vision, such interpretation necessitates the recovery of the 3D scene structure, that is to say, of the attributes of shape, location, pose, and perhaps movement, of all the imaged entities. There are a host of computer vision techniques which aim at the estimation of these intrinsically recorded attributes. The purpose of this paper is to present some of them in a tutorial fashion, giving emphasis to the question of 3D shape estimation. We start by considering stereoscopy, in the following section. Next, we present some basic material on optical flow and structure from motion. Finally, we treat in greater length the monocular processes of shape from shading and photometric stereo.

2 STEREOSCOPY

Stereoscopy is a process which allows the determination of the position of a point in 3D space, from its projections in two distinct image planes. It is thus a binocular vision process, founded on geometric transformations that can be easily described. Given two cameras, to which we associate the reference systems r_1 and r_2, with $z-$directions pointing along the respective optical axes, the goal is to obtain, from its projections in the two image planes, the position of a 3D point, $P = (X, Y, Z)$, relative to a reference system r fixed on the scene. If P has coordinates (X_i, Y_i, Z_i), $i = 1, 2$, relative to the camera systems, such projections, assuming perspective transformations, are given by (Shirai, 1987)

$$x_i = \frac{f_i X_i}{Z_i} \text{ and } y_i = \frac{f_i Y_i}{Z_i},\ i = 1, 2 \tag{1}$$

where f_i is the focal length of the $i-$th camera.

Now, through the geometric transformations relating the camera frames, r_1 and r_2, to the scene reference system, r, we can express X_i, Y_i and Z_i, in (1), in terms of the scene coordinates of P, thus obtaining four equations which give these coordinates as functions of the observed projections:

$$x_i = x_i(X, Y, Z) \text{ and } y_i = y_i(X, Y, Z),\ i = 1, 2 \tag{2}$$

Since there are four equations and only three unknowns, namely, X, Y and Z, a relation can be found between the projection coordinates of P in the two image planes. Such relation, which can be written, for instance, as $y_2 = y_2(x_1, y_1, x_2)$, is the equation of the so-called *epipolar line* (Shirai, 1987). Its meaning is that, given the projection of a 3D point in the first image plane, the corresponding projection in the second image must be found along a line whose orientation is determined by the 3D point position and by the centers of projection of the two cameras.

A particular case of interest is that of two cameras with parallel optical axes. The transformations between the reference systems r, r_1 and r_2 are then just translations, the epipolar lines are horizontal, and the depth coordinate of the observed point P is a function only of the difference in projection along the $X-$direction, the so-called *disparity*. For instance, when the camera focal lengths are equal, we find

$$Z = \frac{fd}{D} \tag{3}$$

where $D \equiv x_1 - x_2$ is the disparity, and d is the distance between the centers of projection of the two cameras.

2.1 Stereo Matching

Although the formulation of stereoscopy is thus simple, its proper implementation is a hard issue, object of ongoing research. The foregoing discussion assumed that the pairs of corresponding projections of each point on the scene were known, in order for its 3D position to be determined. But the problem of stereoscopic correspondence, or stereo matching, is not a trivial one. Ways must be devised for finding pairs of points, one in each image, which are projections of the same point in space. The epipolar geometry constrains the search of matching pairs to be performed along straight lines, but, for reasonably sized images, there would still be many possible candidate matches.

There are essentially two distinct approaches for performing the stereo matching: the intensity-correlation approach and the feature-based approach. The intensity-based matching relies on the assumption that the scene points project with approximately the same intensities in the two stereo images. Thus, the search for corresponding points is guided by the similarity between the intensities found in regions of the two images. An appropriate measure must be introduced to quantify the similarity. It can be based, for instance, on the intensity differences across windows centered on the points being checked in the two images, compensated, if required, for brightness and contrast variations between the cameras, as the measure M, below, illustrates (Shirai, 1987):

$$M = \sum_{j=1}^{n} \sum_{i=1}^{m} [(I_1(i,j) - \mu_1) - (I_2(i,j) - \mu_2)]^2 / \sigma_1 \sigma_2 \qquad (4)$$

Here, the I_i's denote the image intensities, the summation is taken across $n \times m$ windows, and the σ_i's and μ_i's are the standard deviations and mean values of the intensities found therein.

Given a point in I_1, the corresponding point in I_2 can be found as that for which the measure M is minimized. A basic question in the process is, of course, the choice of window size, which affects not only its computational overhead, but also its accuracy. Smaller windows allow more precision, but are sensitive to noise, while larger windows provide a more robust, though only global, matching. Also, since there are points for which the matching is ambiguous, and since the cost of determining the correspondence for the whole image is very high, one usually restricts the matching to a subset of points, employing interpolation techniques to recover the whole three-dimensional scene. The candidate points for correspondence search are chosen as those around which the image intensities vary appreciably, such as in regions of high intensity gradient.

In order to avoid the problems associated with the choice of window size and with the sparse matching, one can resort to a multiscale process, whereby low-resolution versions of the input images, obtained by taking block-averages of their intensities, are employed for a dense, point-by-point correspondence, whose result is then used to guide the matching of the higher-resolution pairs. Since the search, at each resolution, is restricted to a small window around the matching position found at the previous step, the computational cost can be drastically reduced (Shirai, 1987).

The alternative approach to stereo correspondence employs, as matching tokens, not the intensity values, but certain features previously extracted from the images. This necessarily yields sparse matches and requires interpolation, but also represents a faster approach, and one less sensitive to photometric variations. The most common implementations seek the correspondence of edge elements, such as the ones based on the zero-crossings of the signal obtained by taking the laplacian of the image function convolved with a gaussian kernel (Grimson, 1981). By employing gaussians with different variances, contours arising from intensity gradients at different spatial scales can be found, which serve as basis for a multiscale stereo matching strategy. The idea is, at each resolution, to match contours in the two images which have approximately the same orientation, and across which the intensity changes have the same sign, with the search region being defined around the matching position found at the previous scale.

3 STRUCTURE FROM MOTION

Structure from motion (SFM) is the process of recovery of 3D information from the displacement of features registered in a dynamic image sequence (Horn, 1986). The dynamic aspect here arises from the movement of the camera relative to the scene, and is the distinctive feature of SFM, which, like stereoscopy, relies on geometric information – feature displacements – for its implementation.

The input to SFM is the optical flow, which is the observed velocity of the intensity pattern over the image plane, due to the camera/scene relative movement. Such velocity is subject to photometric influences, since a point's recorded intensity usually changes as it moves. The common approach is, nevertheless, to disregard such variations, and to obtain the optical flow through an intensity conservation relation,

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t) \qquad (5)$$

where t denotes an instant in time, δt is a short time interval, and u and v are the $x-$ and $y-$ components of the optical flow.

Employing a Taylor-series expansion on the right-hand side, and assuming that the interval δt is infinitesimal, we easily find the optical flow equation (the subscripts denote differentiation)

$$I_x u + I_y v + I_t = 0 \qquad (6)$$

which represents a constraint on the values of u and v: at each image point, the projection of the optical flow along the local direction of the image gradient is given by the rate of change of the intensity pattern there ($\nabla I \cdot (u, v) = -I_t$). The flow component perpendicular to the gradient is not fixed, leaving the determination of u and v still an underscontrained problem. Algorithms for optical flow estimation must therefore consider, together with relation (6), additional constraining information, such as a smoothness requirement. This is illustrated by Horn & Schunck's algorithm (Horn and Schunk, 1981), which obtains the flow field as the minimizer of a two-term functional, or error measure, of the form

$$\int\int dxdy[(u_x^2 + u_y^2 + v_x^2 + v_y^2) + \lambda(I_x u + I_y v + I_t)^2] \qquad (7)$$

where λ is an empirical weighting factor. The second term in (7) comes from the optical flow constraint, while the first one expresses the smoothness condition.

The minimization of (7) yields a converging iterative scheme for the estimation of u and v (Horn and Schunk, 1981). Once such estimates have been found, they can be used for the recovery of relevant 3D information, as, for instance, the position and velocity of the scene objects. Let us consider here a global least-squares approach to this structure from motion problem. Such approach is also based on the minimization of a functional, now of the type

$$e = \int\int dxdy[(u - u_{mod})^2 + (v - v_{mod})^2] \qquad (8)$$

where u and v represent the observed flow, while u_{mod} and v_{mod} are functions of the 3D parameters, obtained by appropriately modeling the movement in the scene (Horn, 1986). As an example, let us assume the case of rigid-body movement with uniform velocity. A 3D point, P, at position $\mathbf{R} = (X, Y, Z)$, moving with angular velocity $\Omega = (A, B, C)$ and linear velocity $\mathbf{V^t} = (U, V, W)$, has total velocity given by

$$\frac{d\mathbf{R}}{dt} = \mathbf{V^t} + \Omega \times \mathbf{R} \tag{9}$$

Such velocity gives rise to an optical flow which can be obtained from (9), by using the perspective projection equations (1). Taking the time derivatives of such equations, we get, $u_{mod} = u_t + u_r$ and $v_{mod} = v_t + v_r$, where

$$u_t = (U - xW)/Z \text{ and } v_t = (V - yW)/Z \tag{10}$$

are the translational components of the flow, and

$$\begin{aligned} u_r &= -Axy + B(1 + x^2) - Cy \text{ and} \\ v_r &= -A(1 + y^2) + Bxy - Cx \end{aligned} \tag{11}$$

are the corresponding rotational components.

With such u_{mod} and v_{mod} substituted into equation (8), we can proceed with a traditional least-squares estimation. Let us illustrate this with the example of translational movement only, for simplicity. The integrand in (9) becomes a function of Z, in this case, allowing us to look for an estimate of depth at each image point, besides trying to find U, V and W. Differentiating with respect to $Z(x, y)$, and equating to zero, we get, from this integrand,

$$Z(x, y) = \frac{\alpha^2 + \beta^2}{\alpha u + \beta v} \tag{12}$$

where $\alpha = U - xW$ and $\beta = V - yW$. Plugging such Z back into (8), we next proceed to look for the velocity components, by minimizing the resulting integral with respect to U, V and W. There result only two independent non-linear equations, which can be numerically solved to yield the translational velocity up to a multiplicative constant. Accordingly, when employed back in equation (12), this velocity estimate allows recovery of $Z(x, y)$ up to a scaling factor.

4 PHOTOMETRIC STEREO AND SHAPE FROM SHADING

Differently from stereoscopy (also known as geometric stereo) and from structure from motion, whereby the scene attributes are recovered from geometric information – the stereoscopic disparities in one case and the optical flow in the other –, shape from shading (SFS) and photometric stereo (PS) are two computer vision shape estimation processes which rely essentially on photometric data, and on the physics of image formation. Both are monocular processes, but while SFS is a single-image technique, PS employs two or more images acquired under different illuminations. Their formulation requires the introduction of the concept of the reflectance map function (Horn, 1977).

The reflectance map is a function which embodies the physics of image formation, relating the intensity value at each point in the image plane (the *image irradiance*, measured as incident power per unit area) to the light radiated by the corresponding point on the scene (the *scene radiance*, measured as power emitted per unit area per unit solid angle) (Horn and Sjoberg, 1979). The intensity observed at a given image point is, to a reasonable approximation, directly proportional to the light emitted by the corresponding point on the scene, and this, in turn, depends on the direction in which this point is illuminated, on the direction in which it is observed, on the local orientation of the surface of which it is part, and on the reflective properties of this surface.

The reflectance map is the function that models such dependences, allowing the expression of the image formation process through the so-called *image irradiance equation*, as below

$$I(x, y) = R(\hat{n}, \hat{s}, \hat{v}) \tag{13}$$

where R is the reflectance map, $\hat{n}$ is the unit vector normal to the imaged surface at each point – which represents the local orientation –, and $\hat{s}$ and $\hat{v}$ are unit vectors denoting the illumination and observation directions, respectively. It is usual to assume that both the light source and the observer are far away from the scene, so that $\hat{s}$ and $\hat{v}$ can be taken as constants, and the relevant dependence of R thus reduces to the orientation vector, $\hat{n}$. This vector can be expressed, in terms of the gradient components of the imaged surface, as

$$\hat{n} = \frac{1}{\sqrt{1 + p^2 + q^2}}(-p, -q, 1) \tag{14}$$

where (using lower-case z to denote relative depth)

$$(p, q) = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right) \tag{15}$$

The reflectance map is therefore also commonly given as $R(p, q)$.

The simplest example of a reflectance map is that of a perfectly diffusing surface, which reflects light equally well in all directions. This so-called *lambertian* surface is characterized by a reflectance function of the form

$$R_l = \hat{n} \cdot \hat{s} = \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2}\sqrt{1 + p_s^2 + q_s^2}} \tag{16}$$

for $\hat{n} \cdot \hat{s} \geq 0$. On the right-hand side of (16), the vector $\hat{s}$ has been expressed in a form similar to $\hat{n}$ in (14), for a gradient (p_s, q_s) corresponding to the surface orientation which points directly towards the light source.

The reflectance of glossy surfaces is usually modelled by combining a term like (16) with one corresponding to a quasi-specular reflectance component, to give (Shirai, 1987)

$$R_g = \rho_l \hat{n} \cdot \hat{s} + \rho_s \{[2(\hat{n} \cdot \hat{s})\hat{n} - \hat{s}] \cdot \hat{v}\}^m \tag{17}$$

The expression under the square brackets denotes the direction of perfect specular reflection, $\hat{v}_s$, for given surface orientation and illumination directions. The quasi-specular component in R_g thus decreases as $\hat{v}$ moves away from $\hat{v}_s$, in a rate that depends on the parameter m. The factors ρ_l and ρ_s represent the lambertian and the quasi-specular reflectivities – or *albedos* – of the surface, that is, the fractions of the incident light which are diffusely and specularly reflected by it.

4.1 Shape from Shading

The goal, here, is to recover the shape of a surface, given a single image of it. One must thus solve an image irradiance equation, like (13), for the orientation map, $\hat{n}$, or, equivalently, the surface gradient, (p, q), at each image point. It is assumed that the underlying surface has uniform or slowly varying reflecting properties, so that the produced irradiance corresponds to a pattern of shading, and not to a textured image.

The most common approach to this problem is to try to minimize an error measure, introduced so as to quantify the departure of a given solution from the ideal one, similarly to what was done for optical flow in equation (7). For instance, one of the earliest proposed functionals had the form (Horn and Brooks, 1986)

$$\int\int dxdy[(I(x,y) - R(p,q))^2 + \lambda(p_x^2 + p_y^2 + q_x^2 + q_y^2)] \quad (18)$$

The first term in the integrand clearly derives from the image irradiance equation, and should ideally be made to vanish by the correct orientation map. The second one, multiplied by the weighting factor λ, corresponds to a regularization factor (Poggio et al., 1985), and tends to favor smooth solutions.

From the variational calculus (Courant and Hilbert, 1987), one finds that the functions p and q which minimize the measure (18) must satisfy Euler-Lagrange equations of the form

$$\lambda \nabla^2 f = -(I - R)R_f \quad (19)$$

where f here stands for p or q, and the subscript denotes differentiation, as usual.

Using a discrete approximation for the laplacian,

$$(\nabla^2 f)_{kl} \approx \frac{\kappa}{\epsilon^2}(\overline{f}_{kl} - f_{kl}) \quad (20)$$

where the overline indicates local average, $\{kl\}$ denotes a position in the discretized image, ϵ is the interpixel distance and κ is a constant depending on the size of the neighborhood considered, one finds the iterative scheme

$$f_{kl}^{(n+1)} = \overline{f}_{kl}^{(n)} + \frac{1}{\kappa\lambda'}(I - R^{(n)})R_f^{(n)} \quad (21)$$

where n denotes the iteration step and $\lambda' = \lambda/\epsilon^2$.

Once p and q have been obtained as above, one may look for the depth map of the recovered surface. This may be given by the minimization of a second functional (Horn and Brooks, 1986),

$$\int\int dxdy[(z_x - p)^2 + (z_y - q)^2] \quad (22)$$

which corresponds to the Euler-Lagrange equation

$$\nabla^2 z = p_x + q_y \quad (23)$$

whose solution can be iteratively found as

$$z_{kl}^{(n+1)} = \overline{z}_{kl}^{(n)} - \frac{\epsilon^2}{\kappa}(\{p_x\}_{kl}^{(n)} + \{q_y\}_{kl}^{(n)}) \quad (24)$$

Recently, iterative approaches based on functionals which couple gradient and depth reconstruction have been proposed (Horn, 1990; Szeliski 1991). Converging schemes have been found, for instance, for the Euler-Lagrange equations obtained from the measure

$$\begin{aligned}\int\int dxdy[(I(x,y) - R(p,q))^2 + \lambda(p_x^2 + p_y^2 + q_x^2 + q_y^2) + \\ + \mu((z_x - p)^2 + (z_y - q)^2)] \quad (25)\end{aligned}$$

when linear approximations to the reflectance map were used. In this case, one expresses

$$R(p,q) \approx k_0 + k_1 p + k_2 q \quad (26)$$

where $k_0 = R(p_0, q_0) - k_1 p_0 - k_2 q_0$, $k_1 = R_p(p_0, q_0)$ and $k_2 = R_q(p_0, q_0)$. For a lambertian reflectance map, and for $p_0 = q_0 = 0$, one finds $k_0 = \cos\sigma$, $k_1 = \sin\sigma\cos\tau$ and $k_2 = \sin\sigma\cos\tau$, where σ and τ are the slant and tilt, respectively of the illumination vector. k_0 is thus the component of $\hat{s}$ along the optical axis (z-component), and (k_1, k_2) is the illumination component along the image plane.

Alternative approaches to SFS have been proposed which are also based on the linearization of the reflectance map. The linear shape from shading by Pentland (Pentland, 1990), for instance, is a frequency-domain technique which allows direct depth estimation from the Fourier transform of the input image. Starting with the linearized image irradiance equation

$$I(x,y) = k_1 p + k_2 q \quad (27)$$

where the DC term, k_0, has been neglected, and taking the Fourier transform of both sides, one easily finds

$$z(x,y) = \mathcal{F}^{-1}\left\{\frac{F_I(\omega,\theta)}{2\pi\exp\{i\pi/2\}(k_1\cos\theta + k_2\sin\theta)}\right\} \quad (28)$$

where $F_I(\omega,\theta) = \mathcal{F}\{I(x,y)\}$, and $\mathcal{F}^{-1}$ denotes the inverse Fourier transform.

From (28), one sees that those Fourier components of $z(x,y)$ for which $k_1\cos\theta + k_2\sin\theta = 0$ can not be recovered, and must be obtained by other means. For a lambertian reflectance map, these correspond to the frequency components perpendicular to the projection of the illumination vector on the image plane.

Pentland later extended this linear shape from shading by proposing the photometric motion process, in which motion-induced photometric variations are used for the estimation of shape from image sequences of dynamic scenes (Pentland, 1991).

In a somewhat different vein, Tsai and Shah (Tsai and Shah, 1994) have introduced another linearized shape from shading, now performing the linear expansion directly in terms of the depth function, $z(x,y)$. They use $p \approx z(x,y) - z(x-1,y)$ and $q \approx z(x,y) - z(x,y-1)$, and express the image irradiance equation as

$$I(x,y) = R(z(x,y) - z(x-1,y); z(x,y) - z(x,y-1)) \quad (29)$$

which can be written as

$$f(I(x,y), z(x,y), z(x-1,y), z(x,y-1)) = 0 \quad (30)$$

Now, taking the linear expansion of the function f around an approximation z^{n-1} of the surface function, they arrive at a sparse system of linear equations in the z values at the image-lattice sites, whose iterative solution, by Jacobi's method (Press et al., 1986), can be easily found as

$$\begin{aligned} z^n(x,y) = z^{n-1}(x,y) - \\ - \frac{1}{\left(\frac{\partial f}{\partial z}\right)_{(n-1)}} f(I(x,y), z^{n-1}(x,y), z^{n-1}(x-1,y), \\ z^{n-1}(x,y-1)) \quad (31)\end{aligned}$$

This method, too, has been extended, now to deal with a kind of continuous form of photometric stereo, called photomotion

(Zhang et al., 1996), whereby monocular images obtained under continuously varying illumination are taken as input to a Kalman filter estimator. The basic image irradiance equations are there treated similarly to what was done in the linear shape from shading case above described.

A third linear approach to shape from shading is the one introduced by Lee and Kuo (Lee and Kuo, 1992), where the surface function is approximated as a combination of triangular elements, $\phi_i(x,y)$, over a uniform lattice in the image. Expressing $z(x,y)=\sum_i z_i\phi(x,y)$, we get

$$p=\sum_i z_i\frac{\partial\phi_i}{\partial x} \text{ and } q=\sum_i z_i\frac{\partial\phi_i}{\partial y} \tag{32}$$

For a linearized reflectance map, there results

$$R(p,q)=k_0+k_1p+k_2q=\sum_i z_i\Phi_i(x,y)+k_0 \tag{33}$$

where

$$\Phi_i(x,y)=k_1\frac{\partial\phi_i}{\partial x}+k_2\frac{\partial\phi_i}{\partial y} \tag{34}$$

Since the functions ϕ_i are known, the surface estimation process reduces to obtaining the node values, z_i. These can be found by minimizing the error functional

$$Q=\int\int dxdy[(I(x,y)-R(p,q))^2+ \\ +\frac{\lambda}{2}(z_{xx}^2+2z_{xy}^2+z_{yy}^2)] \tag{35}$$

With $R(p,q)$ and the derivatives of z expressed in terms of the triangular functions, this leads to a quadratic form on the z_i's, which can be solved, for instance, by gradient descent (Press et al., 1986).

This approach bears some resemblance to the one by Jones and Taylor (Jones and Taylor, 1994), where a kind of *scale-space tracking* (Shirai, 1987) is performed. At each scale, the surface function is expressed as $z(x,y;\sigma_k)=\sum_{i=1}^N\Psi(x,y;\sigma_i)$, with $\Psi(x,y;\sigma_k)=z(x,y;\sigma_k)-z(x,y;\sigma_{k+1})$, where the Ψ's are given by a sum of gaussian basis functions positioned on a regular grid:

$$\Psi(x,y;\sigma_k)=\sum_{ij}a_{ij}\exp\{-[(x-i)^2+(y-j)^2]/2\sigma_k^2\} \tag{36}$$

The procedure here is to obtain the coefficients a_{ij} at each scale, and employ the recurrence relation

$$z(x,y;\sigma_k)=z(x,y;\sigma_{k+1})+\Psi(x,y;\sigma_k) \tag{37}$$

from $z(x,y;\sigma_N)=0$ up to $z(x,y;\sigma_0)$, which is the surface estimate.

The desired coefficients, a_{ij}, must be such that the solution at each scale, $z(x,y;\sigma_k)$, explains the image irradiance at that scale, $I_k(x,y)$, where I_k is an appropriately blurred version of the input image. The proposed blurring algorithm is the one of (Peleg and Ron, 1990), which is adequate for approximately circularly symmetric reflectance maps. In order to produce the image that would result from the surface approximation at a certain scale, one blurs the field $\sqrt{p^2+q^2}=R^{(-1)}(I(x,y))$.

Given the blurred image at scale σ_k, Jones and Taylor try to minimize the simple error functional

$$Q=\int\int dxdy[R(p,q)-I_k(x,y)]^2 \tag{38}$$

where

$$p=\frac{\partial(z_{k+1}+\Psi_k)}{\partial x} \text{ and } q=\frac{\partial(z_{k+1}+\Psi_k)}{\partial y} \tag{39}$$

This, too, leads to a quadratic form, which can be solved via gradient descent techniques.

4.2 Photometric Stereo

Photometric stereo was introduced by Woodham as a practical means of solving the shape from shading problem in controlled situations (Woodham, 1980). It uses two or more monocular images, acquired under different illuminations, to produce an orientation map of the scene. Under certain conditions, the set of image irradiance equations thus obtained may allow a closed-form solution for the gradient function at each image point.

An academic example of interest is the reconstruction of a lambertian surface of unkown albedo, from three photometric stereo inputs (Shirai, 1987). The corresponding image irradiance equations are given by

$$I_k=\rho(\hat{s}_k\cdot\hat{n}) \tag{40}$$

for $k=1,2,3$, where the albedo, ρ, and the orientation map, $\hat{n}$, are the unknowns. In matrix form, (40) can be recast as

$$\mathbf{I}=\rho\mathbf{S}\mathbf{n} \tag{41}$$

where $\mathbf{I}=[I_1,I_2,I_3]^T$, $\mathbf{n}=[n_1,n_2,n_3]^T$, and the matrix $\mathbf{S}$ has entries of the form s_{ij}, meaning the $j-$th component of the $i-$th illumination vector.

From (42), one thus easily gets $\rho=|\mathbf{S}^{-1}\mathbf{I}|$ and $\mathbf{n}=\mathbf{S}^{-1}\mathbf{I}/\rho$, as long as $\mathbf{S}$ is non-singular, which requires the vectors $\hat{s}_1$, $\hat{s}_2$ and $\hat{s}_3$ not to be simultaneously coplanar.

The main advantage of PS is the possibility of its implementation as a simple look-up process, in controlled environments. For instance, in a manufacturing plant dealing with parts made up of a specific material, a calibration procedure may yield the associated reflectance map in tabular form, avoiding the need for its analytical determination. A sphere of the same material as the manufactured parts may be produced, which, illuminated from different directions (usually three or more illuminations are required), generate a set of irradiance values easily associated, at each point, to the known local orientation. The table thus constructed, whose entries are the surface gradient components and their corresponding irradiances, constitute a reflectance map representation which may be used for the inverse process of surface reconstruction, when the parts are imaged under the same set of illuminations as the calibration sphere was.

Of course, photometric stereo can also be employed when fewer than three images are available, and also without resorting to calibration schemes, in situations where knowledge of the reflectance map function is incomplete (e.g., when only a linear approximation to R is known). Some of the previously discussed approaches to shape from shading can, for instance, be extended to PS. Lee and Kuo have done so with their iterative SFS (Lee and Kuo, 1992), based on triangular basis functions (see Section 4.1). They have also considered the question of choosing the most appropriate illuminations to be employed, showing that, for a two-image process, the accuracy of the surface reconstruction is optimized when the tilt angles of the illumination vectors differ by 90^o.

Another two-image approach to PS has also been proposed recently, which is based on a parallel between this process and the geometric stereo depth reconstruction. In this so-called disparity-based photometric stereo (Fernandes and Torreão, 1998), the input images are matched as if they were a stereoscopic pair, to yield a disparity map, (D_x, D_y), which encodes shape information. For a matching performed along the direction given by $D_y/D_x = k_2/k_1 \equiv \gamma$, with k_1 and k_2 obtained from the linear expansion of the difference image, $\Delta I \equiv I_1 - I_2 = k_0 + k_1 p + k_2 q$, the resulting disparity field allows the direct recovery of an approximation to the surface function, under the form

$$z(x,y) = \frac{D_x I_2}{k_1} - \frac{k_0(x+\gamma y)}{k_1(1+\gamma^2)} \tag{42}$$

The accuracy of such reconstruction improves, the closer to central the two illuminations are, and the smaller the angle between them (Torreão and Fernandes, 1998).

REFERENCES

Courant, R. and D. Hilbert (1987). *Methods of Mathematical Physics*. John Wiley & Sons, New York.

Fernandes, J.L. and J.R.A. Torreão (1998). Estimating depth through the fusion of photometric stereo images. *Lect. Notes Comp. Sci.* 1351, 64-71.

Grimson, W.E.L. (1981). *From Images to Surfaces*. MIT Press, Boston.

Horn, B.K.P. (1977). Understanding image intensities. *Artif. Intelligence* 8(2), 1-31.

Horn, B.K.P. (1986). *Robot Vision*. MIT Press.

Horn, B.K.P. (1990). Height and gradient from shading. *Intl. J. Comput. Vision* 5(1), 37-75.

Horn, B.K.P. and M. Brooks (1986). The variational approach to shape from shading. *Comput. Vision, Graph & Image Process.*, 33(2), 174-208.

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

Horn, B.K.P. and B.G. Schunck (1981). Determining optical flow. *Artif. Intelligence* 17, 185-203.

Jones, A.G. and C.J. Taylor (1994). Robust shape from shading. *Image & Vision Comput.*, 12(7), 411-421.

Lee, K.M. and C.-C.J. Kuo (1992). Shape reconstruction from photometric stereo. *IEEE Intl. Conf. Computer Vision*, 479-484.

Peleg, S. and G. Ron (1990). Nonlinear multiresolution: a shape from shading example. *IEEE Trans. PAMI*, 12(12).

Pentland, A. (1990). Linear shape from shading. *Intl. J. Comput. Vision* 4, 153-162.

Pentland, A. (1991). Photometric motion. *IEEE Trans. PAMI* 13(9), 879-890.

Poggio, T., V. Torre and C. Koch (1985). Computational vision and regularization theory. *Nature* 317(26), 314-319.

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986). *Numerical Recipes in C*. Cambridge University Press.

Shirai, Y. (1987). *Three-Dimensional Computer Vision*. Springer-Verlag, Heidelberg.

Szeliski, R. (1991). Fast shape from shading. *CVGIP-Image Understanding*, 53(2), 129-153.

Torreão, J.R.A. and J.L. Fernandes (1998). Matching photometric stereo images. *J. Opt. Soc. Am. A* 15(12), 2966-2975.

Tsai, P.-S. and M. Shah (1994). Shape from shading using linear approximation. *Image & Vision Comput.* 12(8), 487-498.

Woodham, R.J. (1980) Photometric method for determining surface orientation from multiple images. *Opt. Engineering* 19, 139-144.

Zhang, R., P.-S. Tsai and M. Shah (1996). Photomotion. *Comput. Vision & Image Underst.* 63(2), 221-231.