A new approach to shape from shading

Vladimir A. Kovalevsky

1 Introduction

Most of the works on Shape from Shading [e.g. 1, 2] consider the problem as the restoration of the spatial shape of a smooth continuous surface when a continuous function representing the brightness at each point of a plane projection of the surface is given. In order to solve the problem, properties of nonlinear partial differential equations are investigated. The ultimate solution is then performed by some numerical methods. As seen e.g. from the recent review [3], this approach has not led to a practically usable solution. The author sees the reason of the lack of success in the discrepancy between continuous models and the theory of differential equations on the one hand and the digital nature of the images as well as the use of numerical methods on the other hand. Thinking first about continuous functions and differential equations, and then coming back to a finite numerical representation and a numerical solution, is irrational.

The author suggests therefore an essentially digital approach: the surface to be reconstructed is considered as a polyhedron. The image contains then a finite number of plane polygonal regions each having a constant grey value. The problem consists in finding the heights of the vertices of the polyhedron. The heights must satisfy a system of quadratic equations.

In the present paper the conditions for the existence and the uniqueness of the solution of the system of equations are investigated. It turns out that the system is strongly overdetermined. Therefore the solution must be performed by minimizing the sum of the squared discrepancies. An improved nonlinear least-squares method of solution is suggested and some results of computer experiments are reported.

2 Problem statement

2.1 The simplest version

Given is a digitized grey value image in a rectangular region of the plane. We consider it as the result of viewing a polyhedron surface being illuminated by a single infinitely distant light source, e.g. the sun. In the simplest case we suppose that the surface obeys the Lambertian law of reflection and has a constant albedo (reflectance ability). We also suppose (in the simplest case) the direction of a vector pointing to the sun to be known. Thus each face of the polyhedron must have a constant grey value. Correspondingly, we subdivide the given image into regions with constant grey values and approximate the boundaries of the regions by polygonal lines. In this way we get the X- and Y-coordinates of their vertices. The problem consists in finding such values of the Z-coordinates of the vertices that the grey values calculated on the basis of these Z-coordinates be equal (or approximately equal) to the given grey values. The formal problem statement looks as follows:

Given a subdivision of a (rectangular) region of the plane into polygons, a grey value of each polygon and a vector pointing to the infinitely distant light source (the sun)
Find the height (i.e. the *Z*-coordinate) of each vertex of the given polygons such that the grey value of each polygon, calculated from the heights, be equal to the given grey value of the corresponding polygon.

2.2 More complicated versions

More complicated versions of the problem statement include also finding an unknown sun vector, or two vectors of two light sources, and/or the albedo of certain parts of the surface. It is also possible to look for the best matching of the grey values rather then for an exact one.

3 The solution

3.1 Existence of the solution

Consider first the simplest case when all given polygons are triangles. In the general case it is possible to subdivide polygons into triangles. Consider the dependence of the grey value of the surface of a triangle in the 3D space from the heights of their vertices. It is well known that according to the Lambertian law the grey value *GV* (or the brightness) of a surface is proportional to the cosine of the angle between the sun vector and the normal to the surface. A value proportional to the cosine may be calculated as the scalar product of that vectors divided by the norm of the normal. Division by the norm of the sun vector is not necessary since (in the case of an infinitely distant light source) this vector does not depend on the heights of vertices while we are only interested in this dependence.
Let us denote the coordinates of the vertex *P0* by:

$$P0.X,\ P0.Y \text{ and } P0.Z,$$

and similarly for other vertices and vectors. Then the components of the normal $\boldsymbol{N}$ to the surface of the triangle (*P0, P1, P2*) are:

$$\begin{aligned} N.X &= (P1.Y-P0.Y)\cdot(P2.Z-P0.Z) - (P1.Z-P0.Z)\cdot(P2.Y-P0.Y); \\ N.Y &= (P1.Z-P0.Z)\cdot(P2.X-P0.X) - (P1.X-P0.X)\cdot(P2.Z-P0.Z); \\ N.Z &= (P1.X-P0.X)\cdot(P2.Y-P0.Y) - (P1.Y-P0.Y)\cdot(P2.X-P0.X). \end{aligned} \quad (1)$$

We are seeking for the *Z*-coordinates of all three vertices.

The calculated grey value GV is then equal to

$$GV=(N.X\times S.X+N.Y\times S.Y+N.Z\times S.Z)/\sqrt{(N.X^2+N.Y^2+N.Z^2)}, \quad (2)$$

where $S.X$, $S.Y$ and $S.Z$ are the components of the vector $\boldsymbol{S}$ pointing to the sun. The calculated value GV must be equal to the given grey value GG. Thus we arrive at the equation:

$$(N.X\times S.X+N.Y\times S.Y+N.Z\times S.Z)/\sqrt{(N.X^2+N.Y^2+N.Z^2)}=GG.$$

After having squared both sides of it we arrive at the quadratic equation in the components of $\boldsymbol{N}$:

$$(N.X\times S.X+N.Y\times S.Y+N.Z\times S.Z)^2-GG^2\times(N.X^2+N.Y^2+N.Z^2) = 0. \quad (3)$$

If the heights of the vertices $P0$ and $P1$ are supposed to be known, then it is a quadratic equation in $P2.Z$, since the only terms in (3) depending on $P2.Z$ are $N.X$ and $N.Y$, and they depend linearly on $P2.Z$.

Note that (3) contains the square of GG. Thus a solution of (3) may correspond to a negative value of GG which is inadmissible: such a solution corresponds to a surface, not visible from above. The sign of a solution may be checked by means of (2). As we see, equation (3) may have none, or one, or two admissible solutions.

Consider the case when the given image is subdivided into NT triangles. The height of one of the points may be always chosen arbitrarily, since according to (1) and (2) the grey values depend only on the *differences* of the heights. Thus we have a *system of NT equations in $NP-1$ unknowns*, NP being the number of points (i.e. vertices of the triangles). Let us look for the relation between these two numbers. According to the Euler's formula for a closed polyhedron surface (of genus 0):

$$NF-NE+NP=2,$$

where NF is the number of the faces, NE is the number of the edges (triangle sides) and NP is the number of vertices. Our triangle system is however not closed. Nevertheless we can apply Euler's formula if we consider the system as being placed onto a sphere. Then the outside of the system must be considered as one more face of the corresponding polyhedron surface. Thus we have for our triangle system: $NF=NT+1$ and hence:

$$NT+1-NE+NP=2. \quad (4)$$

Now let us establish a relation between NT and NE. Denote NE_e the number of the *exterior* triangle sides which separate a triangle from the outside of our system. The rest of $NT-NE_e$ sides are the *interior* ones each of which is incident with exactly two triangles. Let us introduce the notion of a "half-side": we split each interior side S into two half-sides and assign each half-side to one of the two adjacent triangles incident with the original side S. The exterior sides are not split: each exterior side corresponds to exactly one half-side. It is assigned to their incident triangle. Now every triangle has three half-sides assigned to it and the number NH of the half-sides is equal to $3NT$. On the other hand:

$$NH = NE_e + 2(NE–NE_e),$$

and hence:

$$NH = NE_e + 2(NE–NE_e) = 3NT.$$

After having combined this equation with (4) one obtains:

$$NT = 2NP–NE_e–2. \qquad (5)$$

For example, if we have a square image raster of $N \times N$ pixels and we subdivide each pixel (being considered as a small square) into two triangles, then we have $NP=(N+1)^2$ points, $NE_e=4N$ exterior sides and we obtain $NT=2N^2$ triangles which number is in accordance with (5). Equation (5) is true for any triangulation of a connected plane region.

This knowledge may be useful to consider, to which extent our system of equations is overdetermined. Important is the difference between the number NT of the equations and the number $NP-1$ (see above) of the unknowns. Let us denote the difference OVR. Then we have:

$$OVR = NT–NP + 1.$$

After having substituted (5) into this equation one obtains:

$$OVR = NP–NE_e –1.$$

For example, for the raster of $N \times N$ pixels:

$$OVR = (N+1)^2–4N–1 = (N–1)^2–1.$$

Thus we see that, except of some trivial cases of very small images, the system of equations is *strongly overdetermined*. An exact solution is only possible if the equations are mutually dependent and exactly $NP-1$ equations among them are independent. In practice this is impossible because of inexact measuring of the grey values. Thus one of the possible ways of solution is the *least-squares method*, in which case we must look for the minimum of the sum of squared differences between the measured and the calculated grey values. This problem will be discussed in Section 4.

3.2 Uniqueness

It is well known (see e.g. [1]) that in the particular case of a vertical sun vector (the sun stays in zenith) for any solution of the Shape from Shading problem there exists another twinned solution. It contains the same absolute values of the height differences but with opposite signs. This can easily be seen from equations (1) and (2): if we change the sign of all Z-differences, the value of GV will not change, since the terms containing $N.X$ and $N.Y$ in the numerator of (2) disappear, $N.Z$ does not depend on the heights, and the denominator contains only squares of the differences.

In the case of other sun vectors the problem of uniqueness becomes more complicated: we have a system of quadratic equations each of which may have up to two solutions. However, the number of the solutions of the whole system remains open. The author has tried to investigate this question experimentally.

Consider a triangulation of the given image, each triangle having a constant grey value, and a subset *U* of the triangles having all a common vertex. Each triangle of *U* must have exactly two adjacent triangles of *U*. We shall call such a subset of triangles „umbrella“. We use umbrellas since the solution for an umbrella is easier than the general solution, and there is a hope to dissolve the general solution into solutions for single umbrellas.

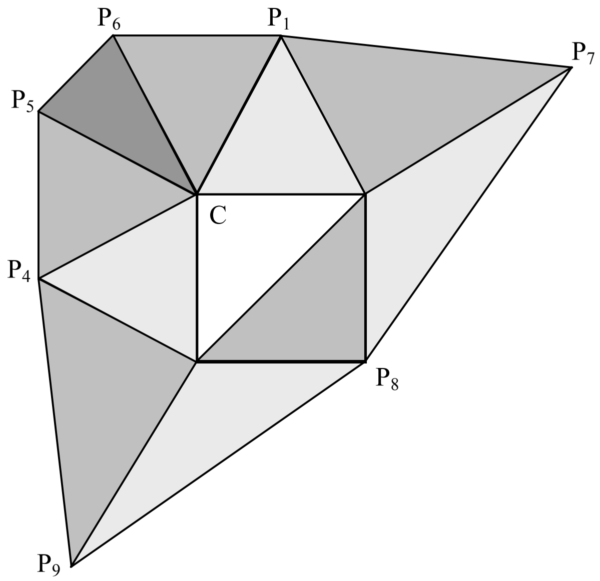


Fig. 1. Illustration to the calculation of the heights discrepancy in an umbrella

As already mentioned, the height of one of the points may be chosen arbitrarily. Let it be the centre point *C* of the umbrella containing the points *P1* to *P6* (Fig. 1). Consider now one of the boundary points *P* of the umbrella. Since the grey values of the triangles are given, it is possible to define the range of the admissible heights of *P*. The idea is as follows.

Consider a triangle *T* of *U* containing *P*. The angle between the normal *N* of *T* and the sun vector is uniquely defined by the grey value of *T*. Let us rotate *N* around the sun vector while preserving the angle. Consider a plane *E* through *C* perpendicular to *N*. It rotates together with *N*. Consider now the vertical line through *P* and the intersection of this line with *E*. At any rotation of *E* a unique value *Z* of the *Z*-coordinate of the intersection point is defined. When having rotated the normal *N* through a full circle, we obtain the full set of the possible values of *Z*. Some of them correspond to negative values of *N.Z*. They must be dropped, since a negative *N.Z* corresponds to a turned over position of the triangle's surface in which case it is invisible. Also a zero value of *N.Z* has no sense: it corresponds to a vertical plane which is only possible under an infinite value of *P.Z*. The author has developed two computer programs: one simulating the rotation of the plane and another one calculating the range of *P.Z* analytically. The results of both programs coincide.

When determining the range of *P.Z* from the two triangles of *U* incident with *P*, the common range (being the intersection of both ranges) may become smaller. Now it is possible to check each admissible value of *P.Z*, whether it may belong to the solution of the equation subsystem for all points of the umbrella *U*. This test works as follows. Denote the boundary points of the umbrella *P1, P2, ..., Pn*. The point *P1* is our point *P*, whose height Z_s we have arbitrarily chosen from the range of admissible values. The point *P2* belongs to a triangle containing the points *C* and *P1*. Their heights are known. Now we can find the possible values of *P2.Z* using the quadratic equation (3). As explained above, there may be none, or one, or two admissible solutions. If there is no solution, the chosen value Z_s of *P1.Z* is not admissible and another value must be chosen. Otherwise the program tests the next boundary point *P3* for all values found for *P2.Z* etc. The point *P2* may have up to 2 solutions, *P3* up to 4 etc., the point $P(n+1)$ up to 2^n. The process stops when the starting point *P1* is reached. The program compares all the 2^n solutions (*n* is the number of points in the umbrella) with the chosen starting value Z_s and looks for the nearest one which we denote Z_n. The discrepancy $DZ(Z_s)=Z_n-Z_s$ is the measure of the quality of the value Z_s for *P1.Z*. The values of *P1.Z* leading to a zero discrepancy are candidates for the global solution for the umbrella *U*. If there is more than one such candidate, all of them must be tested by means of repeating the test described above for all umbrellas having the boundary points of *U* as their centre points. A true solution must lead to all discrepancies equal to zero.

The author has made hundreds of experiments with different triangulations and different values of the sun vector. The results are as follows.

The above introduced function $DZ(Z_s)$ turned out to be discontinuous at some values of its argument. A discontinuity may even happen at a value of Z_s belonging to the true solution (the latter is always known in our experiments since the program generates the image from a given polyhedron by calculating the grey values of the triangles according to equation (2)). The discontinuity of $DZ(Z_s)$ makes the use of the well-known bisection method (or a similar one) for finding the roots of $DZ(Z_s)$ impossible. However, the program finds the true solution nevertheless because, if a value of a height of a point was a discontinuity point of $DZ(Z_s)$ of one umbrella, it always was a

„good“ point in another umbrella. Up to now the program has always found a single solution. Thus we can hazard the conjecture, that under nonvertical sun vectors the solution is unique, if it exists.
As mentioned above, the system of equations is in all practically important cases strongly overdetermined. This means, the unique solution exists only if the given grey values are exactly specified. In our experiments they were double precision floats. In all practical cases, when the given grey values come from a scanner or a similar device, they have a relatively low precision. Only a subset of as many equations as the number of points minus one may have a solution. This is the reason for using an entirely different technique to solve the problem.

4 The Levenberg-Marquardt method

The well-known version of the least-squares method solves the following problem:

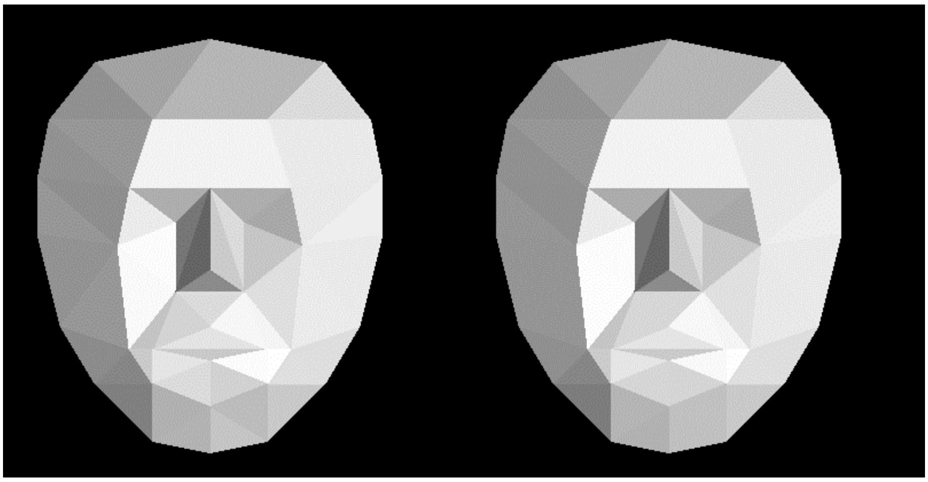
> Find the set $\{a_1, a_2, ..., a_m\}$ of unknown parameters which minimizes the merit function:
>
> $$F = \sum_{i=1}^{N} [y_i - y(x_i; a_1, a_2, ..., a_m)]^2 .$$

In the simplest case the function $y(.)$ is linear in the desired parameters $\{a_1, a_2, ..., a_m\}$ and the solution of the problem can be reduced to solving the well-known normal equations by Gauss. Our problem may be also regarded as a least-squares problem, however our functions are nonlinear. In this case the Newton method may be implemented. This is an iterative method of finding roots of nonlinear functions. The method may be generalized for systems of equations and for multidimensional minimization. In the latter case the equations arise from differentiating the merit function with respect to the desired parameters and demanding that the partial derivatives be zero. The merit function is in our case the sum of squared differences $GV_i–GG_i$, i being the index of a triangle, GG_i its given grey value and GV_i its calculated grey value. The derivatives may be linearized by representing them by Taylor series containing only terms with the first derivatives of y(.). Thus one obtains a system of normal equations, however not in the desired parameters $\{a_1, a_2, ..., a_m\}$ but rather in their increments. The coefficients of the equations are calculated for some initial set of the parameters accepted as the initial approximation. The increments obtained as the solution of the normal equations are then added to the current approximation which gives the next approximation etc.
Unfortunately this method does not ensure a monotone convergence to the minimum. Sometimes it does not converge at all, depending upon the choice of the initial approximation.
An improvement of the Newton method is the Levenberg-Marquardt method [4]. This method represents a combination of the Newton method and the steepest descent method. According to the Levenberg-Marquardt method, the diagonal elements of the matrix of the normal equations are artificially increased by multiplying each diagonal element by $1+\lambda$ while λ is a small positive control parameter. The matrix gets more *diagonally dominant*. The increments of the parameters obtain in this way an

additional value proportional to the components of the gradient of the merit function and the process tends to that of steepest descent. This is appropriate if the merit function *increases* during the current iteration instead of decreasing, what it is supposed to do. In the case when the merit function really decreases, the control parameter λ must be also decreased. The process tends then to that of Newton. This is more appropriate in the vicinity of the desired minimum.

5 Computer experiments

The author has performed hundreds of computer experiments both with solving systems of quadratic equations and with minimizing the merit function. The latter was performed both with the Newton and the Levenberg-Marquardt method. Artificial images computed from reliefs represented in the computer (Fig. 2) have been processed. The best results have been obtained with the Levenberg-Marquardt method.



a b

Fig. 2. Example of an artificial relief used in computer experiments:
a) original; b) reconstructed when starting from constant heights

Experiments with true photographs are now in preparation. For this purpose photographs are scanned, contrasted, smoothed with an edge preserving filter and subdivided into regions with constant grey values. A special new method of encoding grey level images by means of the so-called cell lists is used. A cell list is a data

structure proposed by the author [5, 6]. It contains descriptions of regions, boundary curves (approximated by polygons) and branch points. The cell list contains explicit and complete topological and geometrical information about the image: the image may be reconstructed from the list. On the other hand, the topological information as for example about bounding relations between curves and regions, branch points and curves etc., is directly accessible from the list. Geometrical information is also explicitly represented in the form of coordinates of the branch points and polygon vertices. The image may be subject to geometrical transformations without changing the basic contents of the list: only the coordinates must be changed, and this can be done by means of the well-known matrix multiplication method.
Having obtained the cell list of an image, one may subdivide the regions into triangles and then apply the above described method of calculating the heights of the vertices. In this way the problem of Shape from Shading may be solved not only for artificial images but also for true photographs, e.g. that of human faces. Because of the fact that the equations are strongly overdetermined there is the possibility to calculate the direction to the light source and eventually also the directions to more than one light source. We also consider the possibility of calculating the albedo (i.e. reflectance) of some parts of human faces, of course with some aid of a human operator who must help the computer to identify regions in the face where the albedo may differ from that of other parts of the face.

References

[1] Horn, B.K.P.: Obtaining shape from shading information. In: Horn, B.K.P. and Brooks, M.J. (eds.): Shape from shading. Cambridge, Massachusetts, The MIT Press 1989.

[2] Brooks, M.J. and Choinacki, W.: Direct computation of shape from shading. Rapport de recherche *n*° 2176, INRIA, France, January 1994.

[3] Hildebrand, A.: Von der Photographie zum 3D-Modell. Berlin, Springer 1996.

[4] William H. Press et. all, Numerical recipes in *C*, Cambridge, Massachusetts, Cambridge University Press 1988.

[5] Kovalevsky, V.A.: Finite topology as applied to image processing, CVGIP 46, 141-161, 1989.

[6] Kovalevsky, V.A.: Finite topology and image analysis. In: Hawkes, P. (ed.): Image mathematics and image processing. Boston San Diego New York London Sydney Tokyo Toronto Academic Press 1992 (Advances in Electronics and Electron Physics, vol. 84, pp. 197-259).