Progress in the Adaptation of Shape-from-Shading to Human Body Shape Measurement

Harvey MITCHELL*

School of Engineering, University of Newcastle, Newcastle, Australia

Abstract

Shape-from-shading (SFS) is a process of determining an object's three-dimensional surface shape by using nothing more than the radiometric levels in an image of the object. SFS is feasible only with objects which have even optical texture and whose shapes are free from sharp curvature and discontinuities. Human body surfaces typically satisfy both these optical and physical texture requirements. SFS is also simple and cheap. Its major disadvantage is that, if only a single image of an object is used, it is impossible to quantify 3D shape variation using SFS – unless some approximations or assumptions are made. This paper is a report into an investigation which seeks to adapt single-image SFS to reliable human body shape measurement. An approximated mathematical model used to convert images to three-dimensional shapes aims to minimise the use of integration in the determination of surface shape. Since earlier reports on the project, the errors which occur in this process have been quantified. The paper outlines some theoretical concepts. The theory has been verified on simulated surfaces, and tested with real images of geometric shapes and real human body measurement. It is concluded that SFS techniques should be possible for the measurement of some body surfaces which have low levels of gradients and only a small degree of curvature.

Keywords: shape-from-shading, radiometric levels, digital imagery

1. Background to Shape-from-Shading and its relevance to Body Surface Measurement

Shape-from-shading (called 'SFS' here) is a process of determining an object's three-dimensional surface shape by using nothing more than the radiometric levels recorded in digital imagery of the object. In the method discussed here, SFS is based on single images. (This approach can be distinguished from methods which use two or perhaps more images, in a technique which is often called 'photometric stereo', e.g. [1]).

SFS is attractive for human body shape measurement, because SFS depends on objects having even optical texture, preferably of a lighter shade. As well, shapes which are free from discontinuities or even high curvature are imperative. Human body surfaces typically satisfy these optical and physical texture requirements.

Another crucial benefit and dominant characteristic of shape-from-shading (SFS) is that the data collection – the imaging – is simple and cheap. The imaging requires nothing more than an ordinary camera, normally a flash camera, although other external lighting is feasible. The camera can even be a mobile phone camera, although, not surprisingly, better cameras (in terms of both radiometric and geometric characteristics) facilitate higher accuracy. A second advantage of single-image SFS is that, for the measurement of humans, who continuously move, even imperceptibly, the use of a single image means that the object does not have to be held perfectly still during the measurement, in contrast to some scanners with a longer scan period.

Given the appeal of SFS for body surface measurement, the writer has investigated whether single-image SFS can be used for reliable and accurate body measurement. It is assumed that simple (and cheap) imaging is a valuable feature of SFS, and it would therefore be appropriate for the foreseeable applications that processing should be equally straightforward.

A serious predicament to SFS usage exists: if only a single image of an object is used, it is theoretically impossible to find two parameters which define the direction of the normal at a surface element given just one image reflectance level obtained at the corresponding pixel in the image – unless some approximations or assumptions are made. A solution to the problem of deducing the surface shape can be sought by assuming that the directions of surface normals do not vary randomly, but that the directions of normals at nearby pixels vary only gradually. Finding a workable approach is not easy.

* harvey.mitchell@newcastle.edu.au; + 61 2 4921 5000; http://www.newcastle.edu.au/

Even if the directions of the surface normal (and hence the surface gradients in two directions) can somehow be deduced from the image reflectances, a significant difficulty remains: the integration of the gradients (in two directions) accumulates the numerical errors in the calculations of the surface normal directions plus the effects of radiometric noise in the image, which is often substantial. This project centres on limiting the use of integration. An approximated mathematical model is used to relate the image reflectance levels to the object's geometry. The quantification of the errors which occur in this process may enable its limitations to be assessed. Calculation of the error represents the progress made since earlier reports on the project, [2, 3].

2. Outline of Theory of Shape-from-Shading for Body Surface Measurement

The digital reflectance level at any pixel can easily be converted to a quantity, called $\cos(\theta)$ here, which is the dot product of two vectors, the first being the normal to the surface, the second being the vector from pixel on the image to the image's perspective centre. (We consider the image to be created on the object side of the perspective centre, rather than being produced by the real sensor which is behind the lens). The angle θ is the angle between the vectors of course. The determination of $\cos(\theta)$ from the image reflectances is so trivial that it is simpler to assume that the raw data is the set of $\cos(\theta)$ values, rather than an image of reflectance levels.

The value of $\cos(\theta)$ recorded at a certain pixel is related to the gradients of the surface dz/dx and dz/dy (which are dimensionless) at the surface element which is imaged by that pixel, by

$$\cos(\theta) = \frac{\left(-\tfrac{1}{2}.x.\frac{dz}{dx} - \tfrac{1}{2}y.\frac{dz}{dy} + f\right)}{r_x . r_s} \tag{1}$$

where

x , y , z are coordinates of a pixel, in units of pixels, relative to an origin which is at the lens centre, z is along the lens axis; but is taken as positive towards the object, in contravention of the conventional right-handed rule adopted for x and y, which are effectively coordinates in the image plane;

f is the focal distance of the camera, i.e. the focal length of the lens amended to be the lens-to-image distance when the lens is focussed, (a distance typically called "principal distance" in photogrammetry), measured from the origin of x and y , and is essentially the z coordinate of the image plane;

r_x is the distance from the lens perspective centre to the pixel, in units of pixels;

r_s is a function of the gradients, and is of the order of 1 in magnitude, and is dimensionless.

Despite the assertions above that objects cannot be measured by SFS, if an object is a plane, whose gradients are small enough to let r_s be set the value of 1.00, the two gradients in the x and y directions are easily determinable. Moreover, integration can be avoided with such a model. A surface which is not a plane can be re-created with an error whose magnitude depends how much the surface deviates from the low-gradient plane, i.e. the error is a function of the magnitude of the gradients and on the curvatures of the surface. One challenge is to evaluate the error which occurs, to estimate the tolerable magnitude of deviation from a plane which still enables the surface shape to be determined to a satisfactory accuracy.

More specifically, if the object shape is plane, we can depict the surface by a quantity which we call here z_{approx}, (which has units of pixels) and which is derived from Equation (1), and involves $\cos(\theta)$ and f, as well as r_x and r_s. If the object shape is not planar but close to it, we may be able to approximate the surface by the same z_{approx} quantity which does not involve integration.

An error arises because this calculation of z_{approx} requires knowledge of r_s, which is a function of the gradients, which are not normally known. However, r_s is close to 1, and we can approximate r_s as 1.00 if the gradients are small. For example, if both gradients are equal to 0.05, then r_s is 1.0006. The effect on z of this error at any pixel is then only 0.06%.

There is a second source of error: a difference between the approximation given by z_{approx} and the real z. It can be estimated by recognising that:

$$z \quad = \quad \int \frac{dz}{dx}.dx \ + \ \int \frac{dz}{dy}.dy \tag{2}$$

(with units of pixels), whereas

$$z_{approx} \ = \ \tfrac{1}{2}x.\frac{dz}{dx} + \ \tfrac{1}{2}\,y.\frac{dz}{dy} \tag{3}$$

(in which both sides of both equations are in units of pixels). The difference between Equations (2) and (3) is the error, called Δz , which it can be shown, is given by:

$$\Delta z = \int x.\frac{d^2z}{dx^2}\,dx + \int y.\frac{d^2z}{dy^2}.dy + \int(x+y).\frac{d^2z}{dx.dy}.dy \qquad (4)$$

Equation (4) explains the earlier assertion that Δz is a function of the curvature of the surface. The effect of the error is to exaggerate the curvature in the direction away from the origin. Equation (4) reveals furthermore that the second derivatives are multiplied by the image coordinates, so the effect of the curvature is magnified substantially. The error can thus become prohibitively large, especially at larger distances from the origin. Equation (4) also involves integration of errors caused by numerical estimates of curvatures, emphasizing that avoidance of integration is highly desirable. Consequently, the methods involving Equation (3) to estimate z may be erroneous unless the curvature is small or unless the integration is executed over a short interval.

(None of this theory takes perspective effects into account. The perspective effects decrease as the depth of the object decreases relative to the distance of the object from the camera lens.)

3. Numerical Test with a Simulated Surface

In order to examine the above theory, a simple numerical example has been developed, by using a simulated three-dimensional surface. Simulated surfaces have the advantage that there is no noise. Figure 1 shows the simulated surface, which is modelled by:

$$z = 0.005\,(x-2)^2 + 0.0025\,y^2 \qquad (5)$$

i.e., it is a surface which is parabolic in both the x and y directions. The example surface comprises only (9 x 9) points in order to make the concepts as apparent as possible. The z values range from 0.00 to 0.22 pixels, with the lowest level of the surface at the point (2, 0) relative to the origin.
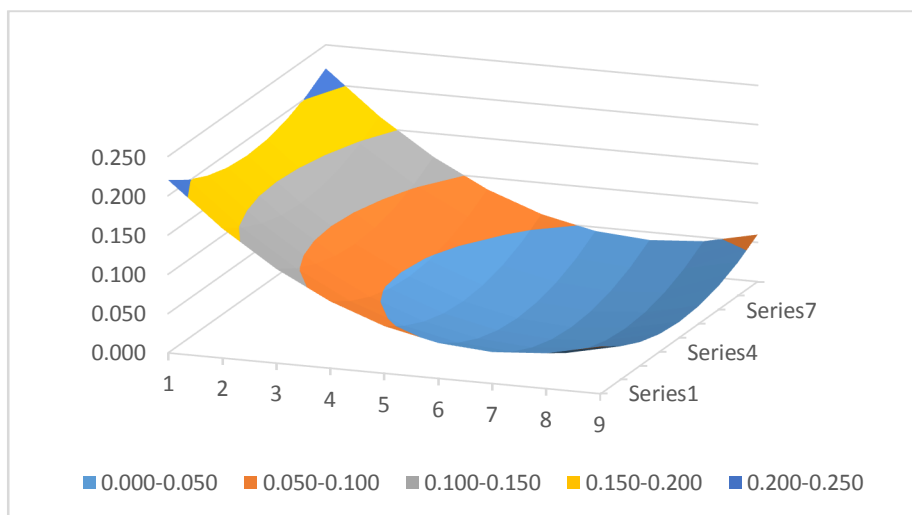

Fig. 1: The simulated surface shape over a (9 x 9) area. Figure created by Microsoft Excel spreadsheet.
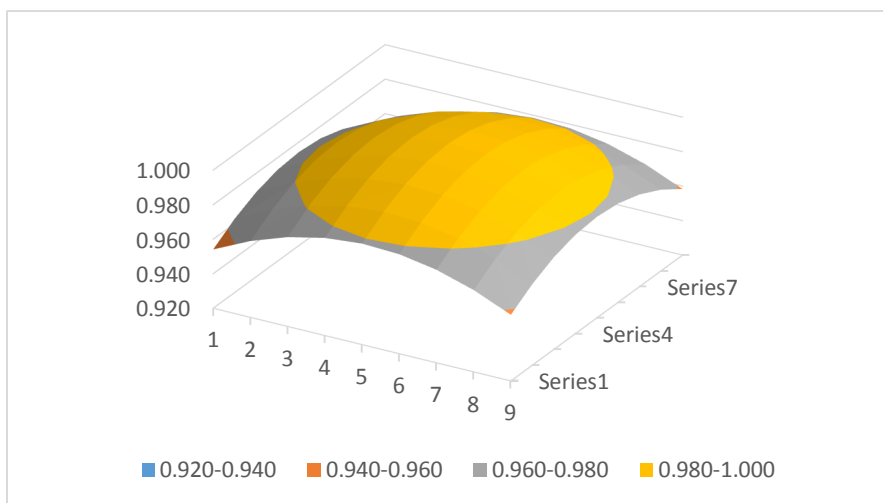

Fig. 2: The $cos(\theta)$ values which would exist for the example (9 x 9) simulated surface if imaged using a camera having a focal distance of 20 pixels. Figure created by Microsoft Excel spreadsheet.

Using the array of z values, the values of $\cos(\theta)$ which would occur on an image given by camera with a (9 x 9) image with a focal distance of 20 pixels have been calculated; see Figure 2. (The depth variation of 0.22 pixels (as in Figure 1 is therefore about 1% of the focal distance.)

The $\cos(\theta)$ values have then been used to re-create the surface of approximate z values shown in Figure 3 by using Equation (3). It should be noted that the cosine values have a small range, only 0.954 to 1.000, and the calculation of surface shape is generally sensitive to the $\cos(\theta)$ values. This emphasises the influence that radiometric noise can have on calculated shapes.
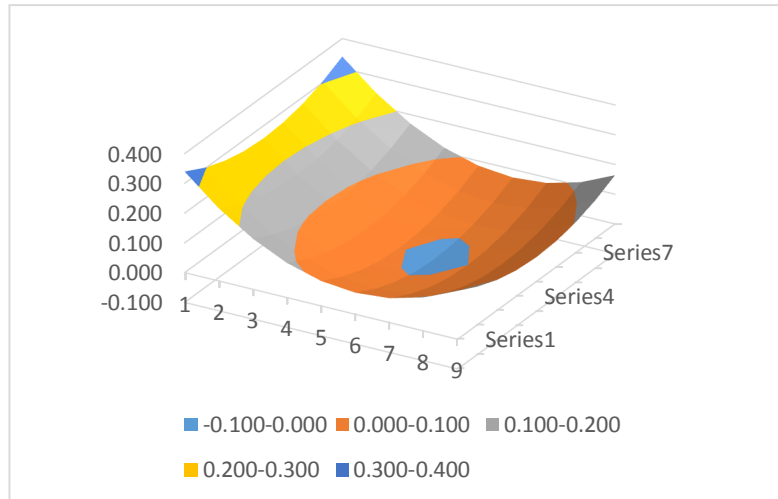

Fig. 3: The shape of the approximated surface as given by z_{approx}. Figure created by Microsoft Excel spreadsheet.

The z values in Figure 3 vary from -0.009 to 0.340, giving a range of 0.349, compared with the range for the original surface of 0.220. The errors, i.e. the differences between the surface shown in Figure 1 and that shown in Figure 3, range from -0.120 to 0.018, i.e. a range of 0.138.

Corrections, calculated using Equation (4), have then been applied to the surface shown above. The revised surface shape is very similar to the original one shown in Figure 1. The z values across the new surface vary from a minimum -0.030 to a maximum 0.190, i.e. a range of 0.220, in agreement with the range of the z values of the simulated surface. The surfaces are not identical, because of numerical errors, but the standard deviation of the errors of the surface is only 0.007. The average difference of 0.031 between the original and re-created surfaces across the 81 points represents a bias due to the impossibility of knowing the original height at the origin.

The correction calculated using equation (4) uses values derived from the real surface, i.e. derived from a differentiating of Equation (5), but these real surface values would not be known in practical cases. Furthermore, the calculation in this simulation have made use of known values of r_s, which would not be available in practical cases either.

4. Results from Real Imagery

The simulation example confirms the validity of the theory, but it does not confirm the validity of the assumptions used to develop that theory. The latter aspect has therefore been assessed via practical tests, which have been carried out on objects with geometric shapes and human body measurement.

4.1. Cylinder

A cylinder was chosen as a useful example because it has curvatures in one direction only, and numerical values for its gradients and curvatures are easily determined. Figure 7 shows an image of a cylindrical shape obtained with a mobile phone camera (Apple iphone 4S), with illumination by the camera's own flash. It and other mobile phone cameras have the distinct advantage that the flash is located very close to the lens. The original colour image, at 1936 x 2592 pixels, has been converted to monochrome, rotated, and reduced to 27 x 36 pixels (to fit the spreadsheet being used for this work), giving a focal length assessed to be, at that scale, 30 pixels. The cylinder has a diameter of 105 mm, and is covered in white paper to give it an appropriate light and even colour, with two marks separated by a 70 mm distance around the circumference to provide scale. The phone camera was 150 mm from the closest point of the cylinder.

Fig 4: Image, at 1936 x 2592 pixels, of a cylinder of 110 mm diameter as obtained using a mobile phone camera with illumination by the camera flash.

The image was then converted to the approximate three-dimensional shape given by the quantity, z_{approx}; see Figure 5. The resulting surface has been smoothed by a (3 x 3) box filter. Irrelevant regions to the left and right sides of the above image have been excluded from Figure 5, so that the figure shown covers only 34 x 19 pixels. The cross-section that contains the surface markers is shown in Figure 6.
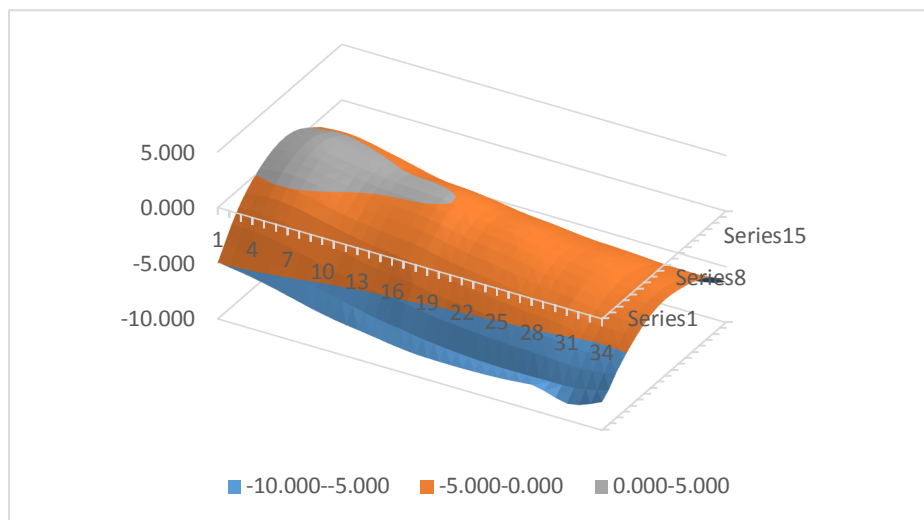
Fig 5: The three-dimensional shape derived from the image in Figure 4 by using Equation (3). Figure created by Microsoft Excel spreadsheet.
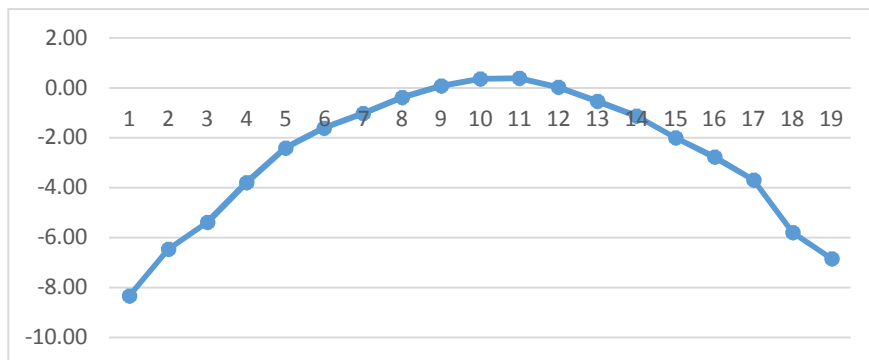

Fig. 6: Cross-section of the approximated cylinder shape, through the arc which contains marked points, which are detectable at pixels 3 and 18. Four points at each edge of the column have been deleted. Figure created by Microsoft Excel spreadsheet.

The surface was then corrected by Equation (4). The advantage of modelling a cylinder is that curvature in the x direction could be assumed to be zero, and the consistent curvature in the y direction could be averaged across the surface to eliminate noise in its calculation, the one curvature value being used for all corrections in that direction.
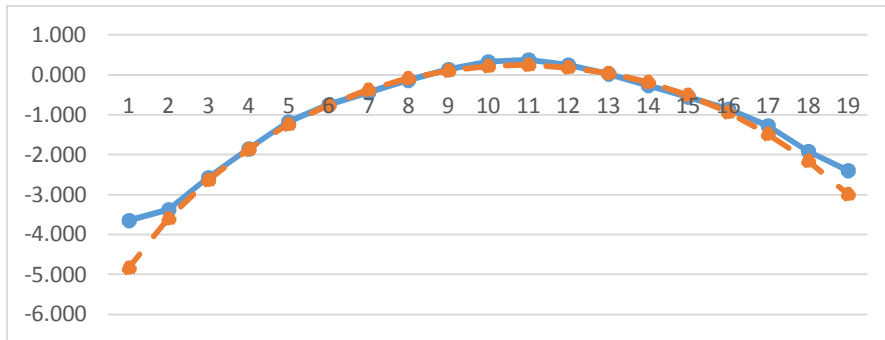

Fig. 7: Cross-section of the approximated cylinder shape, after curvature correction, the arc being through the arc which contains marked points, seen at pixels 3 and 18. Four points at each edge of the column have been deleted. The scale is the same as that in Figure 6. The figure also shows as a broken line the close fit between the result and a circle of radius 12 pixels. Figure created by Microsoft Excel spreadsheet.

In Figure 6, the distance between the marks, which are separated by 65 mm along the chord in reality, is 15 pixels on the image, so the model scale is 4.3 mm / pixel. A chord defined by an arc of 70 mm around a circle of radius 105 mm is separated by 11 mm from the circumference, so, at the calculated scale, this chord-circumference separation of 11 mm should be about 2.5 pixels. The true circle radius of 52.5 mm is scaled to 12 pixels. Figure 8 shows the cross-sectional shape in the centre of the cylinder over the length of the chord defined by the two points which were marked on the surface and which are visible in Figure 7, which is 3 pixels. The result provides some confirms of the concepts in practice. The figure also shows as a broken line the close fit between the result and a circle of radius 12 pixels.

4.2. Human Back

Figure 7 shows as a three-dimensional shape, the image of a human back, obtained with a good quality Canon digital SLR camera, revealing a maximum reflectance level of 250. The original image, at 1728 x 1152 pixels, has been reduced to a 144 x 96 pixels monochrome image. For the calculations, this image has been further reduced to 36 x 24 pixels, and then converted to the approximate three-dimensional shape using z_{approx} : see Figure 8.
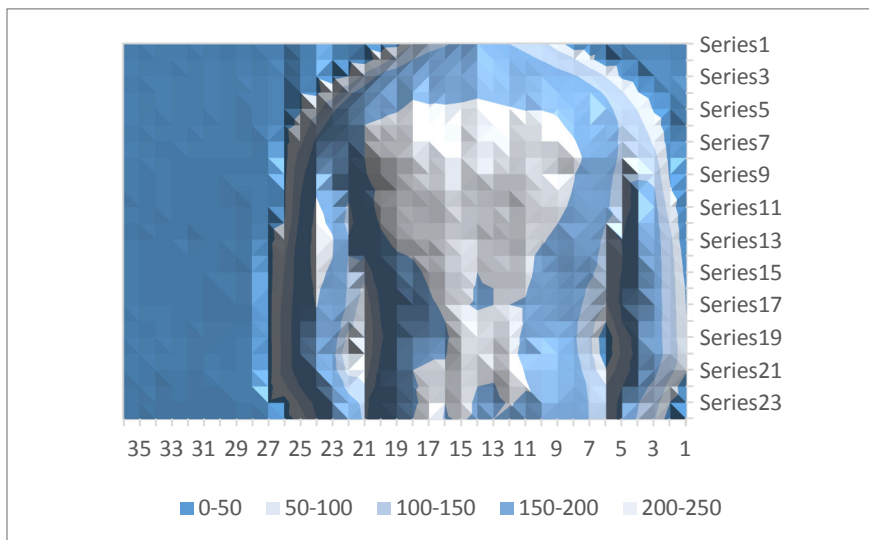

Fig. 7: Flash image of a back, depicted as a three-dimensional shape. Figure created by Microsoft Excel spreadsheet.
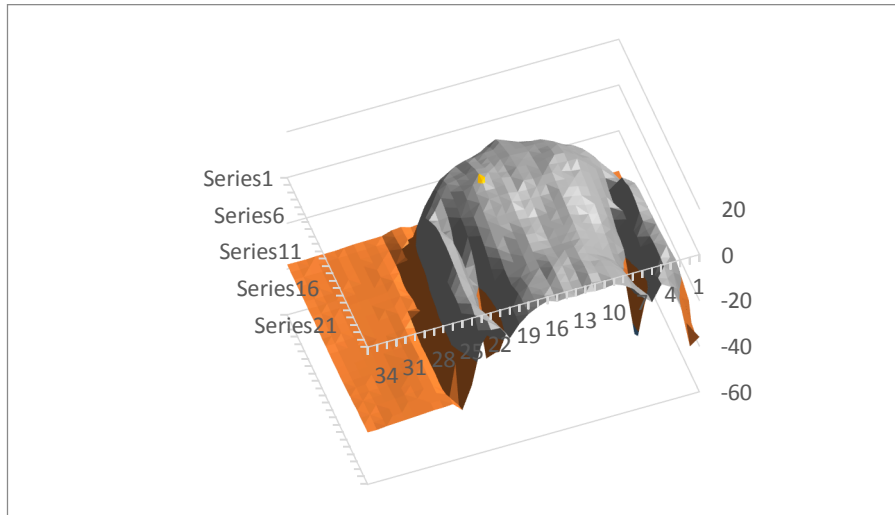
Fig. 8: Shape of the back, deduced using the approximated formula, from the image shown in Figure 7. Units on the axes are pixels. Depths across the central area of the back are seen vary from 0 to about 8 pixels over the central width of about 15 pixels. Figure created by Microsoft Excel spreadsheet.
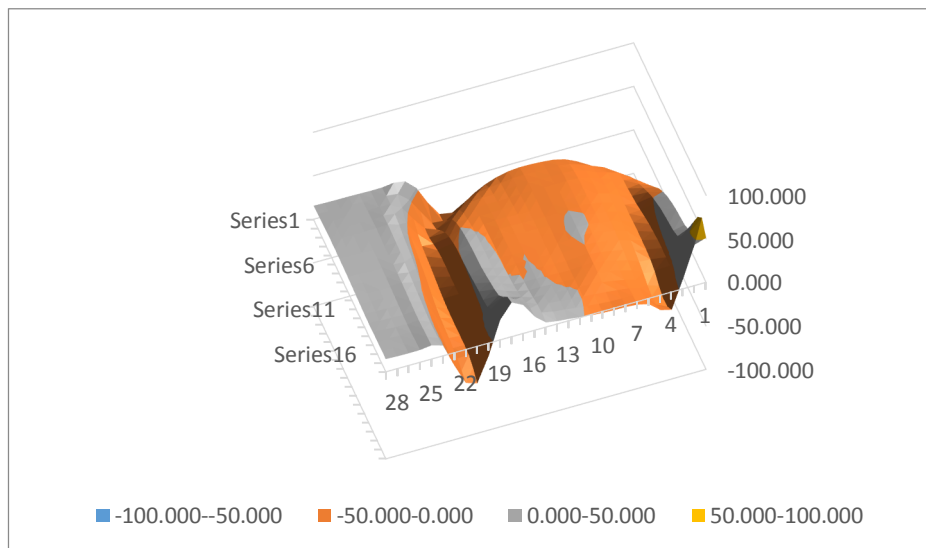

Fig. 9: Calculated corrections to the approximated shape of the back. Units on the axes are pixels. Figure created by Microsoft Excel spreadsheet.
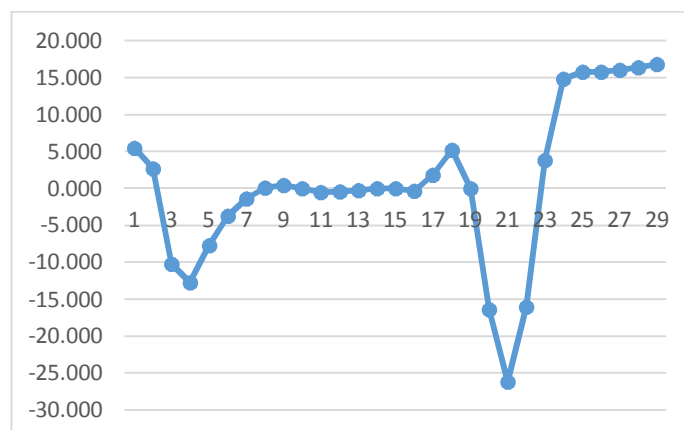

Fig. 10: Calculated corrections across the centre row of the approximated shape of the back. Units on the axis are pixels. Corrections in the flatter regions of the back are seen to be negligible. Figure created by Microsoft Excel spreadsheet.

The figures reveal that errors can, according to the theory, be extreme when curvatures reach high levels, but in the more level regions where shape information is crucial, the errors are small, especially if the areas are also nearer to the centre of the image. The errors may well be tolerable, especially when it is recognised that curvatures have been calculated from the approximate surface and they are larger than those calculated from the real surface. When the curvatures are large, it is true that the calculations of curvature and of the integrals are more likely to be erroneous, as Figure 9 implies. Fortunately, when slopes and curvatures are larger, accurate shape is less important for the user.

5. Conclusions

This project centres on means of limiting the use of integration, plus estimating the errors that occur as a result of the approximated approach.

The simulated image confirms the theory, but it highlights the problem of not knowing first derivatives (i.e. the gradients) and second derivatives (i.e. the curvatures of the surface, not the curvatures of the approximate model). The gradients of a plane over a number of adjacent pixels can easily be deduced if the gradients are not severe. If the shape is close to such a plane, meaning that curvatures are small, then it is feasible to approximate the surface and avoid integration. The order of magnitude of the error may be estimated. Finally, if the surface deviates further from a plane, errors may well be prohibitively large because of the impossibility of estimating the values of both gradients and the curvatures of the surface, and the effects of integration of errors.

For these reasons, the SFS techniques used here may be better suited to the measurement of some body surfaces, i.e. those with lower gradients and smaller curvature, than it is to other areas of the human body. Measurement of the human torso looks to be most practicable. It is also crucial to recognise that the areas where shape is more crucial, i.e. where slopes are lowest, is where corrections are more reliable.

References

[1] A. Sohaib *et al.*, "3D In-Vivo Measurement of Skin Topography Using Photometric Stereo", in *Proc. of 3rd Int. Conf. on 3D Body Scanning Technologies*, Lugano, Switzerland, 2012, pp. 21-29, http://dx.doi.org/10.15221/12.021.

[2] H. Mitchell, "Simple Shape-from-shading for Human Surface Measurement", in *Proc. of 3rd Int. Conf. on 3D Body Scanning Technologies*, Lugano, Switzerland, 2012, pp. 218-226, http://dx.doi.org/10.15221/12.218.

[3] H. Mitchell, "Spread-Sheet Execution of Shape-from-Shading for Human Back Surface Measurement", in *Proc. of 4th Int. Conf. on 3D Body Scanning Technologies*, Long Beach CA, USA, 2013, pp. 203-210, http://dx.doi.org/10.15221/13.203.