Towards shape from shading under realistic photographic conditions

Frederic Courteille Alain Crouzil Jean-Denis Durou Pierre Gurdjos
IRIT - équipe TCI
Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 4, France
{courteille,crouzil,durou,gurdjos}@irit.fr

Abstract

This paper describes a new modeling of the shape from shading problem taking perspective projection into account, and proposes a method of resolution for the new equation. An application is proposed, which consists in correcting the defects of photographs of skew i.e., non flat, documents.

1 Introduction

The shape from shading (SFS) problem consists in recovering the shape of a scene from a single grey-level image, by means of the analysis of the shading. The craze for SFS in the past seems to have subsided, probably because of rather disappointing results on real images [14]. Nevertheless, several recent works [10, 12, 2] have (independently) attempted to modelize SFS in a more realistic way, in particular by considering perspective projection. The present work fits into this scheme; we outline a new modeling of the SFS problem and validate it through a practical application. Our final purpose is to design a system that "unwarps" the image, taken by a digital camera, of a non flat page of an open book, called "skew document" in the sequel, as if it was digitized using a conventional flat-bed scanner. Therefore, there is no pressure on the page to force it flat. It is shown that SFS on its own allows to compute the shape of the pages, without using any prior knowledge about the page content (text, *etc.*) or the outline of the book. The paper is organized as follows. In section 2, we provide the usual assumptions of SFS and the modeling that follows. In section 3, we reappraise the SFS problem, from the orthogonal to perspective projection models. A method of resolution of the obtained equation is proposed in section 4. Finally, the application is described in detail and results are shown in section 5.

2 Image irradiance equation

The SFS problem is known to be difficult to solve in the most general form, so it is usual to make some assumptions: (H1) the shape has neither edge nor hidden part, so the outgoing unit normal $\mathbf{n}(P)$ at a point P is defined without any ambiguity; (H2) the secondary reflections are neglected; (H3) the image is sharp and the aberrations of the lenses are neglected; (H4) the shot angle α (*cf.* Fig. 1) is negligible; (H5) the photosensitive receiver is linear; (H6) the material is lambertian; (H7) the light source emits a parallel and uniform beam, which is described, in direction and density, by vector $\mathbf{S}$. Under H1 to H5, SFS is modelized by the "image irradiance equation" [5]:

$$I(Q) = r\,t\,\frac{\pi\,d^2}{4\,f^2} L(P), \tag{1}$$

where $L(P)$ is the luminance of point P in the direction of the optical center C, $I(Q)$ is the grey-level of point Q conjugated with P, r is the response factor of the receiver, t is the transmission factor of the lenses, d is the diameter of the entrance pupil and f is the focal length. The assumptions H6 and H7 allow to write $L(P) = -[\rho/\pi]\,\mathbf{S}\cdot\mathbf{n}(P)$, where ρ is the albedo. By rewriting (1), we obtain:

$$I(Q) = -r\,t\,\rho\,\frac{d^2}{4\,f^2}\,\mathbf{S}\cdot\mathbf{n}(P). \tag{2}$$

Eq. (2) can be solved only if the relation between P and Q is known *i.e.*, if the projection model is specified. In this work, it is assumed that: (H8) $\mathbf{S} = (0, 0, -S)$. Generalization to any $\mathbf{S}$ is straightforward.

3 Eikonal equation

Under orthogonal projection, points Q and P_o are conjugate (*cf.* Fig. 1). This means that the orthogonal projection P_Ω of point P_o on the "average plane" Ω of the scene is the image of Q by the central projection of center C on Ω. We attach to the camera a three-dimensional coordinate system $(Cxyz)$, such that axis Cz coincides with the optical axis and Cxy is parallel to the image plane Ω. We introduce a function i, such that $i(x, y) = I(Q)$, where Q is the point on Π with coordinates (x, y, f). The unknown of the problem is the function u_o, such that $u_o(x, y)$ is the height of P_o. Whereas the coordinates of P_o are $(x/g, y/g, u_o(x, y))$, where $g = f/z_\Omega$ is the transverse magnification and z_Ω the

IEEE COMPUTER SOCIETY

height of Ω, $\mathbf{n}(P_o)$ depends on the coordinates of $\nabla u_o(x,y)$:

$$\mathbf{n}(P_o) = \frac{(-g\,\partial_x u_o(x,y), -g\,\partial_y u_o(x,y), 1)}{\sqrt{g^2\,\|\nabla u_o(x,y)\|^2 + 1}}. \qquad (3)$$

Writing $i_{\text{max}} = r\,t\,\rho\,d^2\,S/(4\,f^2)$, (2) can be rewritten as:

$$\frac{f^2}{z_\Omega{}^2}\,\|\nabla u_o(x,y)\|^2 = \frac{i_{\text{max}}{}^2}{i(x,y)^2} - 1. \qquad (4)$$

This equation, so-called "eikonal equation", is a first order non-linear partial differential equation. It is the most frequently found in literature concerning SFS. But, it suffices to observe the photograph of Fig. 3 to be convinced that the assumption of orthogonal projection is not realistic with regard to our application, that is, the correction of photographs of skew documents, since the text lines are not parallel. The consideration of perspective projection in SFS has been the object, until recently, of only very few works [9, 6, 3]; moreover, none of these works have proposed a new modeling of SFS. Recently, three groups of authors have simultaneously established a new modeling for perspective SFS [10, 12, 2]. With a pinhole camera model, the conjugate point of Q is now P_p, obtained by central projection of center C. Point P_p has coordinates $(x\,u_p(x,y)/f, y\,u_p(x,y)/f, u_p(x,y))$, and the function u_p is the new unknown of the problem. It follows that vector $\mathbf{n}(P_p)$ depends on $u_p(x,y)$ and $\nabla u_p(x,y)$:

$$\mathbf{n}(P_p) = \frac{(-\hat{g}(x,y)\,\partial_x u_p(x,y), -\hat{g}(x,y)\,\partial_y u_p(x,y), 1)}{\sqrt{\hat{g}(x,y)^2\,\|\nabla u_p(x,y)\|^2 + 1}}, \qquad (5)$$

where $\hat{g}(x,y) = f/\widehat{u_p}(x,y)$ and $\widehat{u_p}(x,y) = u_p(x,y) + x\,\partial_x u_p(x,y) + y\,\partial_y u_p(x,y)$. Eq. (2) can be rewritten as:

$$\frac{f^2}{\widehat{u_p}(x,y)^2}\,\|\nabla u_p(x,y)\|^2 = \frac{i_{\text{max}}{}^2}{i(x,y)^2} - 1. \qquad (6)$$

This equation, that we call "perspective eikonal equation", is very similar to (4), but it deals with the unknown itself, and not only with its gradient. We can expect the resolution of (6) to be more complicated than that of (4). In [10], (6) is solved through the search for its viscosity solutions. A numerical scheme approximating these solutions, as well as the proof of convergence of this scheme, are stated, which is a generalization of results in [7]. Our contribution is to tackle the resolution of (6) from a different point of view.

4 Pseudo-eikonal equation

Like many authors [4], we add the normal $\mathbf{n}(P)$ as a new unknown. A unit vector having two degrees of freedom, this comes down to introducing two unknown functions, denoted here by p and q, such as:

$$\mathbf{n}(P) = \frac{(-p(x,y), -q(x,y), 1)}{\sqrt{p(x,y)^2 + q(x,y)^2 + 1}}. \qquad (7)$$

Eq. (2) can then be rewritten as:

$$p(x,y)^2 + q(x,y)^2 = \frac{i_{\text{max}}{}^2}{i(x,y)^2} - 1. \qquad (8)$$

This equation, which is similar to (4) and (6), is not a partial differential equation. For this reason, we propose to call it "pseudo-eikonal equation". Its resolution is obviously an ill-posed problem, since at each (x,y), there is one equation for two unknowns, but various strategies can be used if it is to be well-posed [6]. Within the framework of our application, we will see that knowledge on the shape of the scene enables to calculate p and q without any ambiguity. The resolution of the problem then consists in recovering the shape, starting from p and q, through integration. Under the assumption of orthogonal projection, we deduce from (3) and (7) that $\nabla u_o(x,y) = (p(x,y)/g, q(x,y)/g)$. Under the assumption of perspective projection, we deduce from (5) and (7) that $\nabla u_p(x,y) = (p(x,y)/\hat{g}(x,y), q(x,y)/\hat{g}(x,y))$. This last relation cannot be used just as it is, since it is equivalent to the following system which is linear w.r.t. the components of $\nabla u_p(x,y)$ (the dependances in (x,y) are omitted):

$$\begin{cases} [x\,p - f]\;\partial_x u_p + y\,p\,\partial_y u_p = -p\,u_p, \\ x\,q\,\partial_x u_p + [y\,q - f]\;\partial_y u_p = -q\,u_p. \end{cases} \qquad (9)$$

The determinant of (9) is equal to $f(f - x\,p - y\,q)$. Using the notations of Fig. 1, the scalar product $\mathbf{n}(P)\cdot\overrightarrow{CQ}$ is equal, on the one hand, to $(f - x\,p - y\,q)\cos\theta$, and on the other hand, to $f\,\cos\psi/\cos\alpha$, thus $f - x\,p - y\,q = f\,\cos\psi/(\cos\alpha\,\cos\theta)$, that vanishes only if $\psi = \pi/2$. If not *i.e.*, if Q is not found on a "silhouette", then the solution of (9) is:

$$\nabla u_p(x,y) = \frac{u_p(x,y)\;\,(p(x,y), q(x,y))}{f - x\,p(x,y) - y\,q(x,y)}. \qquad (10)$$

This new formulation shows that any method of integration existing for orthogonal projection is generalizable for perspective projection. It is worthy of mention that orthogonal projection will provide the shape up to a constant, whereas perspective projection will provide the shape up to a scale factor. This difference, which is visible in Eq. (4) and (6), was already noticed in [10, 12].

5 Application to the correction of photographs of skew documents

Correcting photographs of skew documents like books is not a new problem. In [1], a specific system is described, but the SFS model and the method of resolution used are classical. It was necessary to take the effects of perspective into account, which are very visible on photographs taken at close range.

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

IEEE COMPUTER SOCIETY

Fig. 2 shows the image I_0, of size 2440×3500, of a scanned document (Agfa SnapScan 600, 600 dpi). Fig. 3 shows the image I_1 of the same document, of size 1320×1750, obtained with a digital camera (Canon EOS 300D, 6.5 Mp) under the following conditions: the camera is located at approximately 400 mm from the document, so that the left and right edges of the document coincide with columns of the image; focal length f and the position of principal point O are estimated by a classical calibration method; the scene is lit by the flash of the camera; the document lies on a gauge which simulates the shape of the left page of an open book. The point in proceeding thus, and not really with an opened book, is that the secondary reflections are quasi-non-existing, in accordance with assumption H2. There exist methods making it possible to take the possible secondary reflections into account [8], but we preferred to focus on taking perspective into account. In order to verify assumption H3, the objective is open as little as possible, which minimizes the blur, and the distortion is estimated at the time of calibration (on image I_1, the distortion was already corrected). On the other hand, assumptions H4, H5, H7 and H8 are not realistic. A very simple (and even naive) way to offset this problem consists in taking a photograph of a flat white page of the same paper as the document, then to divide I_1 by this reference image I_{ref} (after correction of the distortion), which provides a new image I_2. Obviously, assumption H6 is not realistic as well, but we will see that despite this, the results are good. Lastly, it is obvious that the albedo is not uniform: the albedo ρ_0 of the non-inked paper is close to 1, whereas the inked zones have albedos which can take all the values in the interval $[0, \rho_0[$. However, the equations of SFS are based on the implicit assumption of a uniform albedo. It is thus necessary to extract from image I_2, using an adaptive thresholding, the image I_3 of the non-inked zones: the grey-level of the inked pixels is calculated by interpolation, from the grey-levels of the non-inked neighbouring pixels. It is I_3, and not I_2, that is used for the computation of the shape of the document through SFS.

As said earlier, the computation of functions (p, q) is made easier by the way that the shape of the document is a cylinder *i.e.*, two points on the same column in the image correspond in the scene to two points having the same height, which implies that q is null and that functions p and u_p depend only on x. Eq. (8) allows thus to calculate $p(x)^2$ according to the average of $i(x, y)$ w.r.t. y, denoted by $j(x)$: $p(x)^2 = {j_{\max}}^2/j(x)^2 - 1$, where $j_{\max}$ is the maximal value of $j(x)$, reached at $x = x_{\max}$. Moreover, the shape being convex, p is positive if and only if $x \geq x_{\max}$. Then the method of integration of Wu and Li [13] is used to calculate u_p, knowing that (10) holds.

The knowledge of the shape of the document is useful on two accounts: it must allow, of course, to simulate the photograph of the flattened document, but also to calculate the image I_5 of the albedos. With this intention, we calculate, thanks to (6), the image I_4 of the shading corresponding to the estimated shape u_p and we obtain I_5 by dividing I_2 by I_4. Each point (x, y) thus corresponds to an albedo and to a point of the surface of the document, which has $(x\, u_p(x, y)/f, y\, u_p(x, y)/f, u_p(x, y))$ as coordinates. To simulate the photograph I_6 of the flattened document, we must then flatten the skew surface. Regarding the method of interpolation, we used `xmorph` [1], a free software, that implements the algorithm of Smythe [11].

Image I_6, which is the final result, is presented on Fig. 4. The text lines are quasi-rectilinear, without use of this information as *a priori* knowledge. Fig. 5 shows three sub-images extracted from I_0, I_1 and I_6, corresponding to a common zone of text. Visually, the obtained correction seems quite satisfactory. To validate our result more rigorously, we ran `gocr` [2], a free OCR software, on these three sub-images. The results, which are reported in Table 1, show that the correction enables the OCR to decide more often and in a more reliable way.

	I_0	I_1	I_6
success rate	81 %	2 %	45 %
failure rate	13 %	68 %	36 %
abstention rate	6 %	30 %	19 %

Table 1. OCR results.

6 Conclusion

In this paper, a new model for SFS that takes perspective into account has been described and discussed; it has also been proposed a method of resolution adapted to the obtained equation. The performance of the resulting algorithm has been illustrated on a practical application with real images. We have shown that SFS on its own can correct the images of skew documents. We can foresee that this algorithm becomes a part of a more sophisticated one, integrating shape from texture and/or contour. As for improving the SFS algorithm itself, it would be worthy of experimenting various lightings, like for instance an annular flash.

References

[1] S. I. Cho, H. Saito, and S. Ozawa. A Divide-and-conquer Strategy in SFS Problems. In *CVPR 1997*, pages 413–419.

[2] F. Courteille, A. Crouzil, J.-D. Durou, and P. Gurdjos. SFS en conditions réalistes d'acquisition photographique. In *RFIA 2004*, pages 925–934 (vol. II). (in french).

[3] J. K. Hasegawa and C. L. Tozzi. Shape from Shading with Perspective Projection and Camera Calibration. *Computer and Graphics*, 20(3):351–364, May 1996.

[1] http://xmorph.sourceforge.net/

[2] http://jocr.sourceforge.net/

[4] B. K. P. Horn and M. J. Brooks. The Variational Approach to Shape From Shading. *CVGIP*, 33(2):174–208, Feb. 1986.

[5] B. K. P. Horn and R. W. Sjoberg. Calculating the Reflectance Map. *Applied Optics*, 18(11):1770–1779, June 1979.

[6] K. M. Lee and C.-C. J. Kuo. Shape from Shading with Perspective Projection. *CVGIPIU*, 59(2):202–212, Mar. 1994.

[7] P.-L. Lions, E. Rouy, and A. Tourin. SFS, Viscosity Solutions and Edges. *Num. Mathematik*, 64:323–353, 1993.

[8] S. K. Nayar, K. Ikeuchi, and T. Kanade. Shape from Interreflections. *IJCV*, 6(3):173–195, 1991.

[9] M. A. Penna. A SFS Analysis for a Single Perspective Image of a Polyhedron. *PAMI*, 11(6):545–554, June 1989.

[10] E. Prados and O. Faugeras. Perspective Shape from Shading and Viscosity Solutions. In *ICCV 2003*, pages 826–831.

[11] D. B. Smythe. A Two-Pass Mesh Warping Algorithm for Object Transformation and Image Interpolation. Technical Report 1030, Comp. Graph. Dept., Lucasfilm Ltd., 1990.

[12] A. Tankus, N. Sochen, and Y. Yeshurun. A New Perspective [on] Shape-from-Shading. In *ICCV 2003*, pages 862–869.

[13] Z. Wu and L. Li. A Line-Integr. Based Meth. for Depth Recov. from Surf. Normals. *CVGIP*, 43(1):53–66, July 1988.

[14] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from Shading: A Survey. *PAMI*, 21(8):690–706, Aug. 1999.

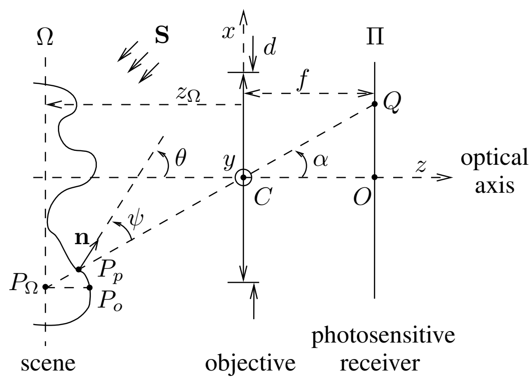

Figure 1. Relation between scene and image.

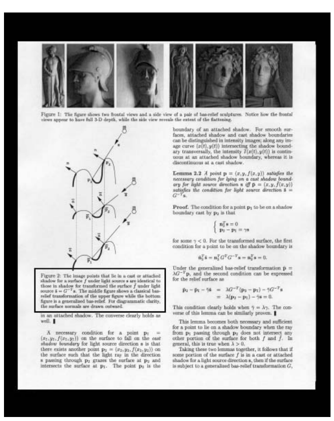

Figure 2. Image I_0 of a scanned document.

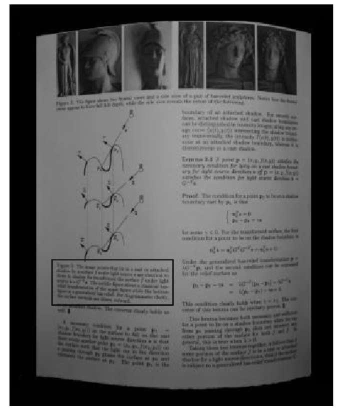

Figure 3. Image I_1 of the skew document obtained by digital camera.

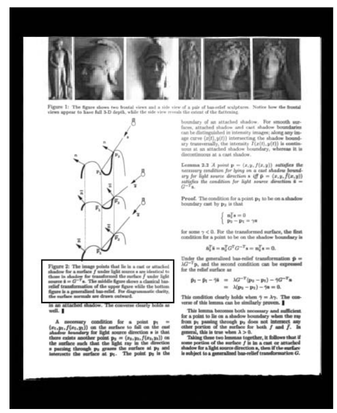

Figure 4. Corrected image I_6 from I_1.

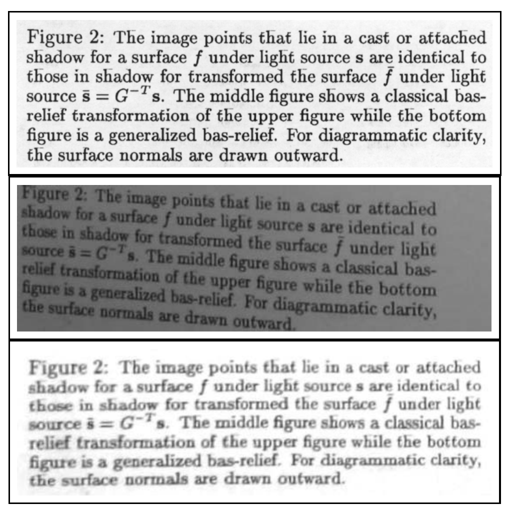
Figure 2: The image points that lie in a cast or attached shadow for a surface f under light source $\mathbf{s}$ are identical to those in shadow for transformed the surface $\bar{f}$ under light source $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$. The middle figure shows a classical bas-relief transformation of the upper figure while the bottom figure is a generalized bas-relief. For diagrammatic clarity, the surface normals are drawn outward.

Figure 2: The image points that lie in a cast or attached shadow for a surface f under light source $\mathbf{s}$ are identical to those in shadow for transformed the surface $\bar{f}$ under light source $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$. The middle figure shows a classical bas-relief transformation of the upper figure while the bottom figure is a generalized bas-relief. For diagrammatic clarity, the surface normals are drawn outward.

Figure 2: The image points that lie in a cast or attached shadow for a surface f under light source $\mathbf{s}$ are identical to those in shadow for transformed the surface $\bar{f}$ under light source $\bar{\mathbf{s}} = G^{-T}\mathbf{s}$. The middle figure shows a classical bas-relief transformation of the upper figure while the bottom figure is a generalized bas-relief. For diagrammatic clarity, the surface normals are drawn outward.

Figure 5. Zoom on images I_0, I_1 and I_6.

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE