Shadows in three-source photometric stereo

Carlos Hernández[1]   George Vogiatzis[1]   Roberto Cipolla[2]

[1]Computer Vision Group, Toshiba Research Europe, Cambridge, UK
[2]Dept. of Engineering, University of Cambridge, Cambridge, UK

Abstract. Shadows are one of the most significant difficulties of the photometric stereo method. When four or more images are available, local surface orientation is overdetermined and the shadowed pixels can be discarded. In this paper we look at the challenging case when only three images under three different illuminations are available. In this case, when one of the three pixel intensity constraints is missing due to shadow, a 1 dof ambiguity per pixel arises. We show that using integrability one can resolve this ambiguity and use the remaining two constraints to reconstruct the geometry in the shadow regions. As the problem becomes ill-posed in the presence of noise, we describe a regularization scheme that improves the numerical performance of the algorithm while preserving data. We propose a simple MRF optimization scheme to identify and segment shadow regions in the image. Finally the paper describes how this theory applies in the framework of color photometric stereo where one is restricted to only three images. Experiments on synthetic and real image sequences are presented.

1   Introduction

Photometric stereo is a well established 3d reconstruction technique based on the powerful shading cue. A sequence of images (typically three or more) of a 3d scene are obtained from the same viewpoint and under varying illumination. From the intensity variation in each pixel one can estimate the local orientation of the surface that projects onto that pixel. By integrating all these surface orientations a very detailed estimate of the surface geometry can be obtained. As any other reconstruction method, photometric stereo faces several difficulties when faced with real images. One of the most important of these difficulties is the frequent presence of shadows in an image. No matter how careful the arrangement of the light sources, shadows are an almost unavoidable phenomenon, especially in objects with complex geometries. This paper investigates the phenomenon of shadows in photometric stereo with three light sources.

Shadows in photometric stereo have been the topic of a number of papers [1–3]. Most papers assume we are given four or more images under four different illuminations. This over-determines the local surface orientation and albedo (3 degrees of freedom) which implies that we can use the residual of some least squares solution, to determine whether shadowing has occurred. However when we are only given three images there are no *spare* constraints against which to

test our hypothesis. Therefore the problem of detecting shadows becomes more difficult. Furthermore, when a pixel is in shadow in one of the three images most methods simply discard it. We show how one can use the remaining two image intensity measurements to estimate the surface geometry inside the shadow region. The solution we propose is based on enforcing (1) integrability of the gradient field, as well as (2) smoothness in the recovered intensity of the missing channel.

Using photometric stereo on just three images may seem like an unreasonably hard restriction. There is however a particular situation when only three images are available. This technique is known as color photometric stereo [4] and it uses three light sources with different light spectra. When the scene is photographed with a color camera, the three color channels capture three different photometric stereo images. Because shape acquisition is performed on each frame independently, the method can be used on video sequences without having to change illumination between frames [5]. In this way we can capture the 3D shape of deforming objects such as cloth, or human faces. Since the method is constrained to operate only on three images it is an ideal application of the theory we present here. Summarizing, the two main contributions of this paper are the following:

- We show how to exploit image regions in photometric stereo where one of the three images is in shadow. A geometric formulation of the problem is given where a set of point-to-point and point-to-line distances are minimized under the integrability condition.
- We develop a regularization scheme that makes the optimization problem well posed while not suppressing the data. This scheme is successfully validated within the color photometric stereo technique of [5].

1.1   Previous work

A vast literature exists on the topic of photometric stereo. Its applications range from 3D reconstruction [6], medical imaging [7] or cloth modeling [5]. One of the main limitations of photometric stereo is the number of different lights required and how the algorithm copes with highlights or shadows.

A minimum of 3 lights is required to perform photometric stereo with no extra assumptions [6], and only 2 lights with the additional assumption of constant albedo [8]. Whenever more lights are available, the light visibility problem becomes a labeling problem where each point on the surface has to be assigned to the correct set of lights in order to successfully reconstruct the surface.

For objects with constant albedo, [3] used a Rank-2 constraint to detect surfaces illuminated by only 2 lights. In the case of general albedo, every point on the surface has to be visible in at least 3 images. A 4-light photometric stereo setup was proposed in [9], where light occlusion was detected by checking the consistency of all the possible triplets of lights. The work by [10] was able to detect light occlusions in a 4-light setup and simply treat them as outliers. In [1]

a similar algorithm to [9] is presented using a 4-light colored photometric stereo approach.

In the recent work by [2], an iterative MRF formulation is proposed for detecting light occlusion and exploiting it as a surface integration constraint. However, the algorithm also requires a minimum of 4 lights and is targeted for setups with a large number of lights.

In this paper we propose a novel solution for 3-light photometric stereo with shadows and varying albedo. We are able to detect and exploit photometric stereo constraints with only two lights while the constant albedo constraint is relaxed into a more practical smoothly varying albedo constraint.

2   Three-source photometric stereo with shadows

In classic three-source photometric stereo we are given three images of a scene, taken from the same viewpoint, and illuminated by three distant light sources. The light sources emit the same light frequency spectrum from three different non-coplanar directions. We will assume an orthographic camera (with infinite focal length) for simplicity, even though the extension to the more realistic projective case is straightforward [11]. In the case of orthographic projection one can align the world coordinate system so that the xy plane coincides with the image plane while the z axis corresponds to the viewing direction. The surface in front of the camera can then be parameterized as a height function $Z(x,y)$. If Z_x and Z_y are the two partial derivatives of Z one can define the vector

$$\mathbf{n} = \frac{1}{\sqrt{Z_x^2 + Z_y^2 + 1}} \begin{pmatrix} Z_x & Z_y & -1 \end{pmatrix}^\top$$

that is locally normal to the surface at (x,y). For $i = 1 \ldots 3$ let $c_i(x,y)$ denote the pixel intensity of pixel (x,y) in the i-th image. We assume that in the i-th image the surface point $\begin{pmatrix} x & y & Z(x,y) \end{pmatrix}^\top$ is illuminated by a distant light source whose direction is denoted by the vector $\mathbf{l}_i$ and whose spectral distribution is $E(\lambda)$. We also assume that the surface point absorbs incoming light of various wavelengths according to the reflectance function $R(x,y,\lambda)$. Finally, let the response of the camera sensor at each wavelength be given by $S(\lambda)$. Then the pixel intensity $c_i(x,y)$ is given by

$$c_i(x,y) = \left(\mathbf{l}_i^\top \mathbf{n}\right) \int E(\lambda) R(x,y,\lambda) S(\lambda) \, d\lambda. \tag{1}$$

The value of this integral is known as the surface *albedo* of point $\begin{pmatrix} x & y & Z(x,y) \end{pmatrix}^\top$. We can define the albedo-scaled normal vector

$$\mathbf{b} = \mathbf{n} \int E(\lambda) R(x,y,\lambda) S(\lambda) \, d\lambda$$

so that (1) becomes a simple dot product

$$c_i = \mathbf{l}_i^\top \mathbf{b}. \tag{2}$$

Photometric stereo methods use the linear constraints of (2) to solve for $\mathbf{b}$ in a least squares sense. From this they obtain the partial derivatives of the height function which are integrated to produce the function itself. In three-source photometric stereo, when the point is not in shadow with respect to all three lights, we measure three positive intensities c_i each of which gives a constraint on $\mathbf{b}$. If we write $L = \begin{bmatrix} \mathbf{l}_1 & \mathbf{l}_2 & \mathbf{l}_3 \end{bmatrix}^\top$ and $\mathbf{c} = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}^\top$ then the system has exactly one solution which is given by

$$L^{-1}\mathbf{c}. \tag{3}$$

If the point however is in shadow, say in the i-th image, then the measurement of c_i cannot be used as a constraint. Since each equation (2) describes a 3D plane, the intersection of the two remaining constraints is a 3D line. If $\mathbf{e}_i$ is the i-th column of the 3×3 identity matrix and D_i is the identity matrix with a zero in the i-th position in the diagonal, the possible solutions for $\mathbf{b}$ are

$$L^{-1}D_i\mathbf{c} + \mu L^{-1}\mathbf{e}_i. \tag{4}$$

where μ is a scalar parameter. This parameter represents the value that c_i would have, had the point not been in shadow in the i-th image.

3 Projection to 2D space

At this point it is useful to project the scaled normals $\mathbf{b}_j$ into 2D space with coordinates p and q. We define the following operator: $\mathcal{P}[\mathbf{x}] = (x_1/x_3, \; x_2/x_3)$. Let the scaled normal $\mathbf{b}$ of surface point $(x, y, Z(x, y))$ project to point $\mathcal{P}[\mathbf{b}] = (p, q)$. Then the coordinates p and q are equal to the two partial derivatives Z_x and Z_y respectively. According to the image constraints and assuming no noise in the data, we can have one of the following three cases:

1. The surface point is in shadow in two or more images. In this case there is no constraint in $\mathcal{P}[\mathbf{b}]$ from the images.
2. The surface point is not in shadow in any of the three images. In this case (p, q) coincides with $\mathcal{P}\left[L^{-1}\mathbf{c}\right] = (P, Q)$.
3. The surface point is in shadow in exactly one image, say the i-th. In this case (p, q) must lie on the line that joins $\mathcal{P}\left[L^{-1}D_i\mathbf{c}\right] = \left(\overline{P}, \overline{Q}\right)$ and $P\left[L^{-1}\mathbf{e}_i\right] = \left(P_e^{(i)}, Q_e^{(i)}\right)$. We call this line the *shadow line* of the shaded pixel.

Note that for all pixels j that are occluded under the i-th image, the corresponding points $\left(\overline{P}_j, \overline{Q}_j\right)$ are on the line described by equation $\mathbf{l}_i^\top \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = 0$. Also, the shadow lines of all those pixels intersect at the point $\left(P_e^{(i)}, Q_e^{(i)}\right)$.

Now in the presence of noise in the data $\mathbf{c}$, cases 2 and 3 above do not hold exactly as points (P, Q) and $\left(\overline{P}, \overline{Q}\right)$ are corrupted: The point (p, q) is slightly different from (P, Q) for unoccluded pixels, and $\left(\overline{P}, \overline{Q}\right)$ is not exactly on the line
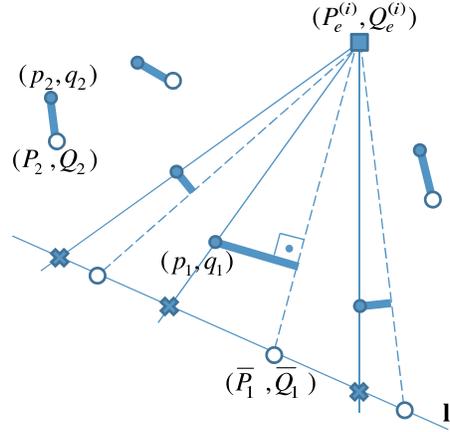
Fig. 1. Geometry of shadowed pixels. The points (p_j, q_j) (dark dots) represent the partial derivatives of the height function at pixel j. For each point (p_j, q_j) there is a corresponding data point (white dot). Pixel 2 is unoccluded and hence (p_2, q_2) must be as close as possible to its data point (P_2, Q_2). Pixel 1 however is occluded so (p_1, q_1) must be as close as possible to its *shadow line*. This is the line joining its data point $(\overline{P}_1, \overline{Q}_1)$ and the intersection point $\left(P_e^{(i)}, Q_e^{(i)}\right)$.

joining (p, q) and $\left(P_e^{(i)}, Q_e^{(i)}\right)$. Figure 1 shows the configuration for six pixels where pixel 1 is under shadow in the i-th image while pixel 2 is not in shadow in any image.

3.1   Integrability condition

Note that the (p, q) coordinates are not independent for each pixel as they represent the partial derivatives of a scalar field and as a result they must satisfy an integrability condition. By assuming that the height function is a continuous piecewise polynomial, the integrability condition takes the form of a convolution between p and q. If for example $Z(x, y)$ is a linear interpolation between height values at pixel centers $Z(i, j)$ then we can express integrability as

$$p(i + 1, j) - p(i, j) = q(i, j + 1) - q(i, j).$$

Our strategy is to obtain values (p, q) for each pixel that minimize the discrepancy between the noisy data and the model while satisfying the integrability condition. This is discussed in the following section.

4   Integrating in the shadow regions

As mentioned, under noise in the image data $\mathbf{c}$, the 2D points (P, Q) and $(\overline{P}, \overline{Q})$ are not perfectly consistent with the model. For non-shadowed pixels, difference

between model and data can be measured by the point-to-point square difference term

$$\mathcal{E} = (p - P)^2 + (q - Q)^2 .$$

In the case of the shadowed pixels however we have a choice of possible ways to quantify the non-collinearity of (p, q), $(\overline{P}, \overline{Q})$ and $\left(P_e^{(i)}, Q_e^{(i)} \right)$. Ideally, since $(\overline{P}, \overline{Q})$ is the point that contains the noise, we should be measuring the distance from $(\overline{P}, \overline{Q})$ to the intersection of the line joining (p, q) and $\left(P_e^{(i)}, Q_e^{(i)} \right)$ with the line $\mathbf{l}_i^\top \begin{pmatrix} p \\ q \\ 1 \end{pmatrix} = 0$. This leads to the following distance term

$$\left\| (\overline{P}, \overline{Q}) - \mathcal{P} \left[\left(I - \mathbf{m}_i \mathbf{l}_i^\top \right) \mathbf{b} \right] \right\|^2$$

where $\mathbf{m}_i$ is the i-th vector of L^{-1}. The expression inside the square is non linear with respect to p and q. We therefore choose to minimize the distance from (p, q) to the line joining $(\overline{P}, \overline{Q})$ and $\left(P_e^{(i)}, Q_e^{(i)} \right)$. This distance is shown in the red dotted line in figure 1. The term this corresponds to is

$$\mathcal{E}_o^{(i)} = \frac{\left(\left(\overline{Q} - Q_e^{(i)} \right) \left(p - P_e^{(i)} \right) - \left(\overline{P} - P_e^{(i)} \right) \left(q - Q_e^{(i)} \right) \right)^2}{\left(\overline{P} - P_e^{(i)} \right)^2 + \left(\overline{Q} - Q_e^{(i)} \right)^2}$$

and here the quantity being squared is linear with respect to p and q.

So we now assume we are given a labeling of pixels into all the possible types of shadow. Let $\mathcal{S}$ contain all non-shadowed pixels while $\mathcal{S}_i$ contains pixels shaded in the i-th image. Our cost function becomes

$$\sum_{j \in \mathcal{S}} \mathcal{E}_j + \sum_{j \in \mathcal{S}_1} \overline{\mathcal{E}}_j^{(1)} + \sum_{j \in \mathcal{S}_2} \overline{\mathcal{E}}_j^{(2)} + \sum_{j \in \mathcal{S}_3} \overline{\mathcal{E}}_j^{(3)}$$

which is a set of quadratic terms in p and q. Finding the minimum of this quantity subject to the integrability condition is a constrained linear least squares problem that can be solved by a variety of methods [12].

Figure 2 shows this idea applied in practice on synthetic data. This experiment indicates that the overall geometry seems to be reconstructed quite well in shadowed regions, provided that these are surrounded by unshaded pixels. The latter act as boundary conditions for the shadowed regions and give the problem a unique solution. Furthermore it also provides evidence that in its present form the problem is ill-conditioned, especially in larger shadowed regions. The following section sheds more light on this and describes our proposed remedy.

4.1   Regularization

The linear least squares optimization framework described in section 2 when executed in practice shows signs of ill-posedness in the presence of noise. This is

demonstrated in the synthetic case of figure 2 where three images of a sphere have been generated. Three shadow regions corresponding to each of the three lights have been introduced. Even though the overall shape of the object is accurately captured some characteristic 'scratch' artifacts are observed. These are caused by the point-to-line distances which do not introduce enough constraints in the cost function. The point (p, q) can move significantly in a direction parallel to the corresponding shadow line only to gain a slight decrease in the overall cost. This results in violent perturbations in the resulting height function that manifest themselves as deep scratches running perpendicular to the shadow lines. The solution to this is some type of regularization on the space of solutions. We have two main requirements on the choice of regularizing criterion:

- The scheme must be consistent with the linear least squares framework. No non-linear constraints can be enforced.
- It must suppress noise while preserving as much of the data as possible.

One possible choice for a regularization criterion is minimizing the Laplacian of the height field $\nabla^2 z$. This is known to have good noise reduction properties and to produce smooth well behaved surfaces with low curvature. However, the Laplacian is isotropic so it tends to indiscriminately smooth along all possible directions. See [13] for a good discussion of anisotropic alternatives to Laplacian filtering in the context of gradient field integration. In the case of our problem, we would like to enforce regularization in (p, q) space along the direction of the shadow line for each shadowed pixel. Fortunately there is an efficient way of achieving this that satisfies both our goals. For this we need to modify our point to line distance term to the following:

$$\hat{\mathcal{E}}^{(i)} = \left(p - w\overline{P} - (1 - w)\,P_e^{(i)}\right)^2 + \left(q - w\overline{Q} - (1 - w)\,Q_e^{(i)}\right)^2. \qquad (5)$$

This introduces a new variable w per shaded pixel, that specifies a location along the shadow line of that pixel. The term is still quadratic with respect to p, q and w but this now allows us to regularize the solution in a meaningful way. The variable w is related to parameter μ of (4) by

$$w = \frac{\mathbf{e}_3^\top L^{-1} D_i \mathbf{c}}{\mathbf{e}_3^\top L^{-1} D_i \mathbf{c} + \mu \mathbf{e}_3^\top L^{-1} \mathbf{e}_i}. \qquad (6)$$

As we mentioned, μ represents the value of c_i that would have been measured had the pixel not been in shadow in that image. We propose putting a cost on the length of ∇w inside the shaded regions. As w is a proxy for μ, this corresponds to introducing smoothness in $\mathbf{l}_i^\top \mathbf{b}$. We can therefore eliminate the scratch artifacts while letting $\mathbf{b}$ have variability in the directions perpendicular to $\mathbf{l}_i$. Figure 2 shows that this scheme works quite well in practice.

Throughout all of the previous discussion we have assumed knowledge of labeling of pixels according to shadows. The next section discusses how we propose to segment shadow regions in the image.

4.2   Segmenting shadow regions

It is known [1] that in photometric stereo with four or more images one can detect shadows by computing the scaled normal that satisfies the constraints in a least squares sense. If the residual of this least squares calculation is high, this implies that the pixel is either in a shadow or in a highlight. With three images however this becomes impossible as the three constraints can always be satisfied exactly, leaving a residual of zero. Recently, [2] proposed a graph-cut based scheme for labeling shadows in photometric stereo with four or more images. Based on the constraint residual, they compute a cost for assigning a particular shadow label to each pixel. This cost is then regularized in an MRF framework where neighboring pixels are encouraged to have *similar* shadow labels. We would like to use a similar framework but we must supply a different cost for assigning a shadow label. The basic characteristic of a shadow region is that pixel intensities inside it are dark. However this can also occur because of dark surface albedo. To remove the albedo factor we propose to divide pixel intensities with the magnitude of the intensity vector $\mathbf{c}$. Our cost for deciding that a pixel is occluded in the i-th image is $c_i / \|\mathbf{c}\|$. This still leaves the possibility that we mistakenly classify a pixel whose normal is nearly perpendicular to the i-th illumination direction $\mathbf{l}_i$. However in that case the pixel is close to being in a self shadow so the risk from misclassifying it is small. The cost for assigning a pixel to the non-shadowed set is given by

$$\sqrt{3} - \min_i \frac{c_i}{\|\mathbf{c}\|}.$$

We regularize these costs in an MRF framework under a Potts model pairwise cost. This assigns a fixed penalty for two neighboring pixels being given different shadow labels. The MRF is optimized using the Tree Reweighted message passing algorithm [14]. Figure 4 shows an example of applying our shadow region segmentation to a real image.

5   Color photometric-stereo

It may seem that a photometric stereo scheme with three images is unnecessarily restrictive. The overall cost in practicality of acquiring one more image is small compared to the rest of the process (calibration, darkening the environment, changing the illumination etc). In this section we examine color photometric stereo [4]. This is a setup where it is not possible to obtain more than three images. The key observation is that in an environment where red, green, and blue light is simultaneously emitted from different directions, a Lambertian surface will reflect each of those colors simultaneously without any mixing of the frequencies. The quantities of red, green and blue light reflected are a linear function of the surface normal direction. A color camera can measure these quantities from a single RGB image. Recently [5] it was shown how this idea can be used to obtain a reconstruction of a deforming object. Because color photometric stereo

is applied on a single image, one can use it on a video sequence without having to change illumination between frames. In [5] shadowed pixels were detected and discarded. Here we show how to improve that method by incorporating shadow regions into the reconstruction. In color photometric stereo each of the three camera sensors can be seen as one of the three images of classic photometric stereo. The pixel intensity of pixel (x, y) for the i-th sensor is given by

$$c_i(x, y) = \sum_j \left(\mathbf{l}_j^\top \mathbf{n}\right) \int E_j(\lambda) R(x, y, \lambda) S_i(\lambda) \, d\lambda. \tag{7}$$

Note that now the sensor sensitivity S_i and spectral distribution E_j are different per sensor and per light source respectively. To be able to determine a unique mapping between RGB values and normal orientation we need to assume a monochromatic surface. We therefore require that $R(x, y, \lambda) = \alpha(x, y) \rho(\lambda)$. Where $\alpha(x, y)$ is the monochromatic albedo of the surface point and $\rho(\lambda)$ is the characteristic chromaticity of the material. Let

$$v_{ij} = \int E_j(\lambda) \rho(\lambda) S_i(\lambda) \, d\lambda$$

and

$$\mathbf{v}_j = \left(v_{1j} \ v_{1j} \ v_{1j} \right)^\top.$$

Also define the scaled normal to be

$$\mathbf{b} = \alpha(x, y) \, \mathbf{n}.$$

Then the vector of the three sensor responses at a pixel is given by

$$\mathbf{c} = \left[\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \right] \left[\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3 \right]^\top \mathbf{b}.$$

Essentially each vector $\mathbf{v}_j$ provides the response measured by the three sensors when a unit of light from source j is received by the camera. If matrix $\left[\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \right]$ is known, then we can compute

$$\hat{\mathbf{c}} = \left[\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \right]^{-1} \mathbf{c}.$$

The values of $\hat{\mathbf{c}}$ can be treated in exactly the same way as the three gray-scale images of section (2). The next section describes a simple process for calibrating color photometric stereo for handling shadows.

5.1 Calibration

In [5] the authors propose a simple scheme for calibrating objects that can be flattened and placed on a planar board. The system detects special patterns on the board, from which it can estimate its orientation relative to the camera. By measuring the RGB response corresponding to each orientation of the material they estimate the entire matrix

$$M = \left[\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \right] \left[\mathbf{l}_1 \ \mathbf{l}_2 \ \mathbf{l}_3 \right]^\top$$
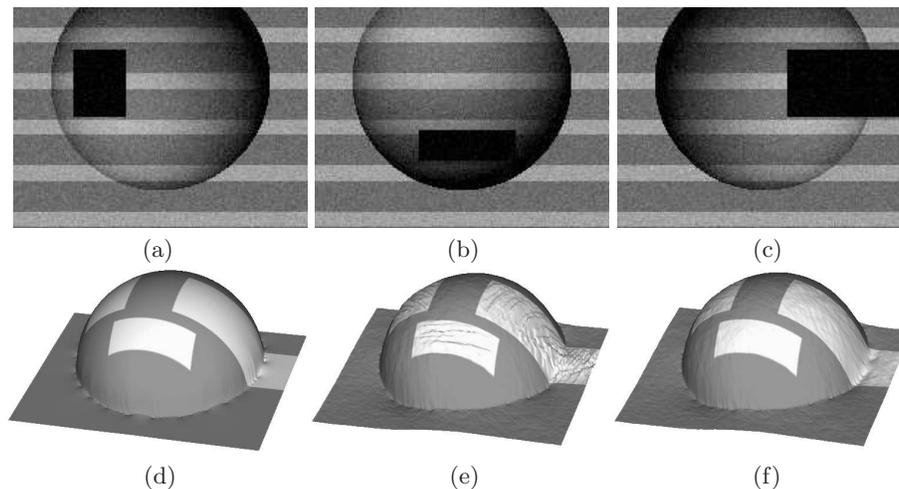
Fig. 2. Sphere sequence. In this experiment we validate our regularization scheme on a synthetic sphere with varying albedo and 10% Gaussian noise. This object is illuminated from three directions with an occluded black region (a-c). The image in (d) shows the reconstruction in the absence of noise and regularization. The image in (e) shows the effect of optimizing the surface using integrability alone. Note the characteristic 'scratch' artifacts. The image in (f) shows the resulting surface after regularization. The artifacts have been suppressed while the data has been preserved unsmoothed.

that links the scaled normals to RGB triplets. We propose a two-step process. Firstly, we use a mirror sphere to estimate light directions $\mathbf{l}_1$, $\mathbf{l}_2$ and $\mathbf{l}_3$. This is a standard process which has been applied in a number of photometric stereo methods. Secondly, we capture three sequences of the object moving in front of the camera. In each sequence, we switch on only one of the three lights at a time. In the absence of noise and if the monochromatic assumption was satisfied, the RGB triplets we acquired would be multiples of $\mathbf{v}_j$ when light j was switched on. We therefore do a least squares fit to the three sets of RGB to get the directions of $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. To get the relative lengths of the three vectors we can use the ratios of the lengths of the RGB vectors. The length of $\mathbf{v}_j$ is set to the maximum length in RGB space, of all the triplets when light j was switched on.

6   Experiments

We present one synthetic experiment and two real experiments on a Frog sequence and a Face sequence.

In figure 2 we study the effect of the proposed framework to automatically detect and correct light occlusions on a half sphere with varying albedo. Figure 2d shows a perfect reconstruction of the sphere in the absence of noise. The algorithm is capable of segmenting the shadowed regions and recovering the

shape without any type of regularization on the albedo. However, as soon as we add noise (see Fig. 2e), the recovered shape shows some characteristic artifacts due to an almost unconstrained variation of w in eq.(5) along the shadow line. These artifacts basically show that the recovered shape and albedo are coupled and integrability constraints on their own are not enough to separate them when one intensity constraint is missing. Introducing the regularization term of section 4.1 adds a prior on the intensity of the missing channel. This helps recover the correct shape without loosing any information (see Fig. 2f). In terms of quantitative results, we have compared the reconstructed normal maps with the ground truth sphere in terms of angle difference, the results being as follows:

algorithm	shadows present	error RMS
ignoring shadows	yes	29.52 degrees
just integrability (Fig. 2e)	yes	15.79 degrees
proposed method (Fig. 2f)	**yes**	**8.67** degrees
data without shadows (ideal case)	no	8.30 degrees

As shown by the table, the proposed method performs almost as well as in the ideal case without any shadows. We obtain an improvement factor of 2 with respect to just using integrability, and a factor of 4 with respect to ignoring the fact that shadows are present.

As a first experimental validation with real data, we use the Frog dataset of [2] which consists of 5 photographs of an object with varying albedo illuminated from 5 different directions. In order to demonstrate the effectiveness of out technique we select a subset of only three images shown in Fig. 3 top. The normalized images $\frac{c_i}{\|\mathbf{c}\|}$ in the middle of Fig. 3 allow the algorithm to easily detect and segment the shadows and to obtain a very accurate shape reconstruction (see Fig. 3 bottom). Note how these images are almost completely invariant to albedo changes in the object surface.

Finally, we have performed a second experiment with video data of a white-painted face illuminated by three colored lights in a similar way as in [5] (see supplemental video and Fig. 4 left). The setup is calibrated as described in section 5.1. The automatic shadow segmentation results in Fig. 4 right, demonstrate the accuracy of the shadow detection algorithm in Section 4.2. Figure 5 shows the reconstruction of 3 different frames of the video sequence without taking the shadows into consideration (top) and after detecting and adding the additional constraints to the linear solver (middle). We can appreciate how the nose reconstruction is dramatically improved when correctly processing the shadows, even though only two lights are visible in the shadowed regions (bottom).

7   Discussion

This paper investigated the problem of shadows in the context of three-source photometric stereo. This is a particularly challenging case because the surface orientation is under-determined inside shadow regions. This is because one of
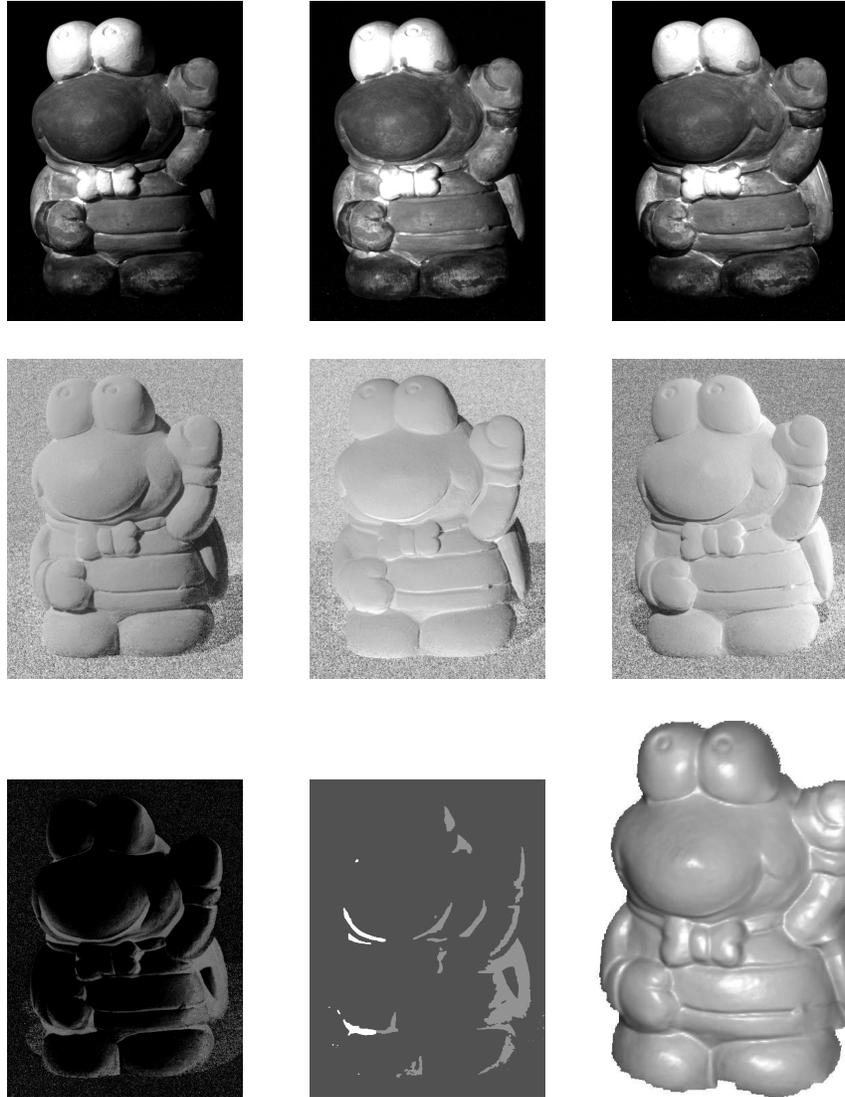
Fig. 3. Frog sequence courtesy of Manmohan Chandraker. In this experiment we use three images of the Frog sequence (top). The middle row shows the normalized images $\frac{c_i}{\|\mathbf{c}\|}$. Note that they do not show any traces of the albedo variation shown in the actual images. In the third row from left to right: The non-shadowing cost $\sqrt{3} - \min_i \frac{c_i}{\|\mathbf{c}\|}$. The shadow segmentation obtained by our proposed scheme. The final surface reconstruction.
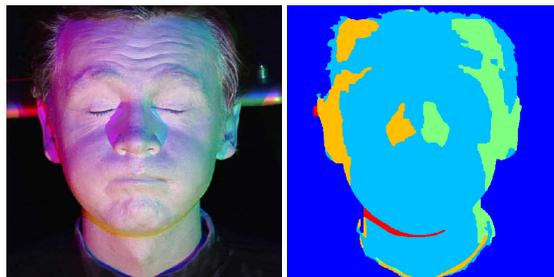
Fig. 4. Shadow segmentation. This experiment shows the result of our shadow region segmentation. On the left is the input image and on the right is the mask with the resulting shadow labels.

the three necessary constraints is missing due to the shadow. We have shown however that by exploiting integrability, one can still use the remaining two constraints to estimate the surface orientation. In its pure form the problem is ill posed even in the presence of some noise in the data. We provided a remedy to this in the form of a regularization scheme that does not suppress the data of the problem. To detect and segment shadows the paper described a simple MRF optimization scheme, based on the relative darkness of pixel locations. Finally we showed how the ideas in this paper can be applied to the interesting acquisition setup of color photometric stereo.

References

1. Barsky, S., Petrou, M.: The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. PAMI **25**(10) (2003) 1239–1252
2. Chandraker, M., Agarwal, S., Kriegman, D.: Shadowcuts: Photometric stereo with shadows. In: IEEE Conference on Computer Vision and Pattern Recognition. (June 2007)
3. Drew, M.: Reduction of rank-reduced orientation-from-color problem with many unknown lights to two-image known-illuminant photometric stereo. In: IEEE International Symposium on Computer Vision. (1995) 419–424
4. Petrov, A.: Light, color and shape. Cognitive Processes and their Simulation (in Russian) (1987) 350–358
5. Hernández, C., Vogiatzis, G., Brostow, G., Stenger, B., Cipolla, R.: Non-rigid photometric stereo with colored lights. In: Proc. 11th Intl. Conf. on Computer Vision (ICCV). (2007)
6. Woodham, R.: Photometric method for determining surface orientation from multiple images. Optical Engineering **19**(1) (1980) 139–144
7. Lee, S., Brady, M.: Integrating stereo and photometric stereo to monitor the development of glaucoma. Image and Vision Computing **9** (1991) 39 – 44
8. Onn, R., Bruckstein, A.: Integrability disambiguates surface recovery in two-image photometric stereo. Int. J. Comput. Vision **5**(1) (1990) 105
9. Coleman, Jr., E., Jain, R.: Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. Computer Graphics and Image Processing **18**(4) (April 1982) 309–328
10. Yuille, A., Snow, D.: Shape and albedo from multiple images using integrability. In: CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97), Washington, DC, USA, IEEE Computer Society (1997) 158
11. Tankus, A., Kiryati, N.: Photometric stereo under perspective projection. In: Proc. 10th Intl. Conf. on Computer Vision (ICCV), Washington, DC, USA, IEEE Computer Society (2005) 611–616
12. Andersen, E., Roos, C., Terlaky, T.: On implementing a primal-dual interior-point method for conic quadratic optimization. Math. Prog. **95**(2) (2003) 249–277
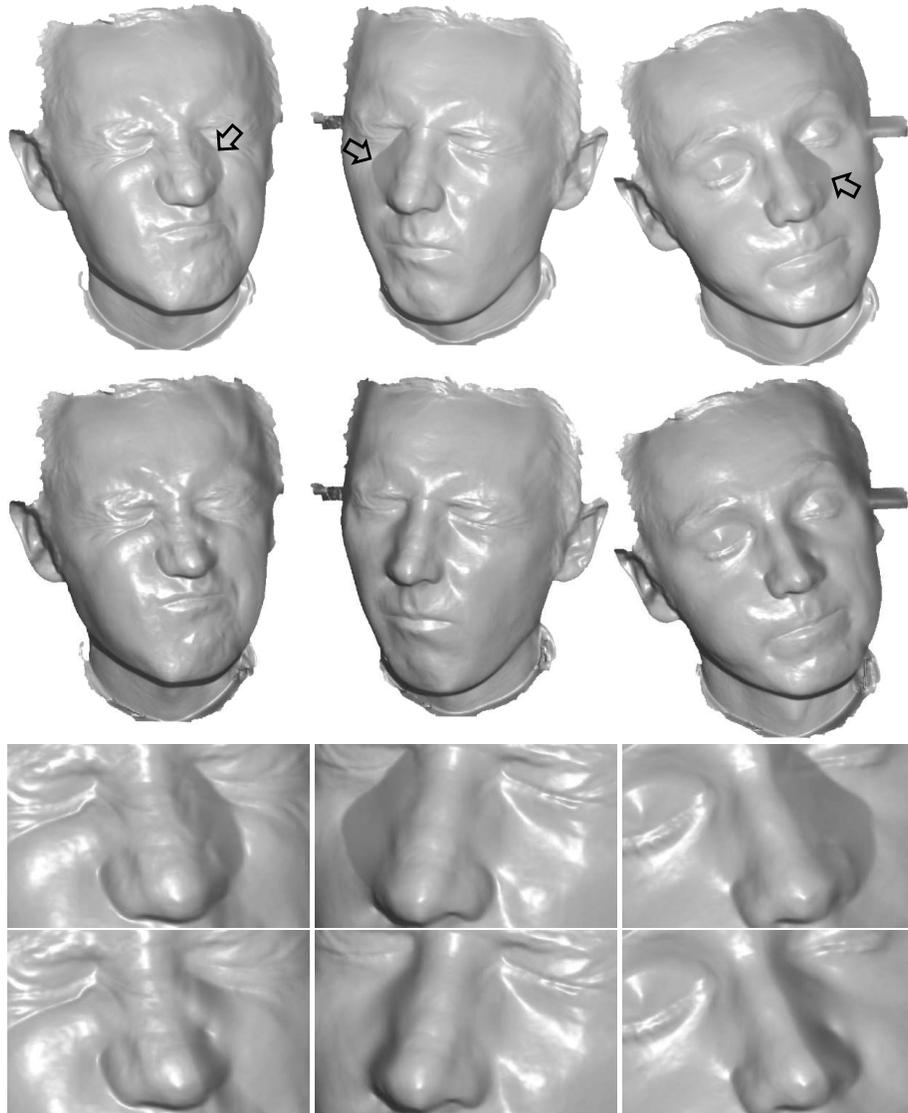
Fig. 5. Face sequence. Three different frames of a video sequence of a face (see supplemental video). The top row shows the reconstruction when shadows are ignored. On the second row are the corresponding reconstructions after detecting and compensating for the shadow regions. Last two rows show a close-up of the face. Note the improvement in the regions around the nose where strong cast shadows appear (see arrows).

13. Agrawal, A., Raskar, R., Chellappa, R.: What is the range of surface reconstructions from a gradient field? In: Proc. 9th Europ. Conf. on Computer Vision (ECCV). (2006)
14. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(10) (2006) 1568–1583