Obtaining the Shape of a Moving Object with a Specular Surface

Atsuto Maki
atsuto.maki@crl.toshiba.co.uk
Roberto Cipolla
cipolla@cam.ac.uk

Toshiba Research Europe
Cambridge Research Laboratory
Department of Engineering
University of Cambridge

Abstract

This paper addresses the basic problem of recovering the 3D surface of an object that is observed in motion by a single camera and under a static but unknown lighting condition. We propose a method to establish pixelwise correspondence between input images by way of depth search by investigating optimal subsets of intensities rather than employing all the relevant pixel values. The thrust of our algorithm is that it is capable of dealing with specularities which appear on the top of shading variance that is caused due to object motion. This is in terms of both stages of finding sparse point correspondence and dense depth search. We also propose that a linearised image basis can be directly computed by the procudure of finding the correspondence. We illustrate the performance of the theoretical propositions using images of real objects.

1 Introduction

Recovering the 3D surface of an object from shaded images has been long investigated in computer vision, based on different assumptions on reflectance models and illumination conditions [8]. Photometric stereo [21] uses multiple images of a static object taken under different illumination conditions in order to remove the ambiguity inherent in a single image. As an alternative, a technique of using the shading variance due to object motion, rather than varying illumination, has been recently studied for the same purpose and attracts attention, see for example [10, 11, 12, 16, 22]. In this paper, we also concern ourselves with this technique because it allows us to stay with the simplest possible setup for obtaining the shape of an object. The basic scheme of the technique is to employ the framework of multi-view stereo, which first requires correspondence of several sample points to recover the extrinsic camera parameters, and then to exploit the shading variance due to object pose for dense matching. Among others the recent work of [10] provides a state-of-the-art solution for a final surface which they compute by further incorporating the technique of photometric stereo.

Notwithstanding the recent advances, one of the limitations of the previous techniques is concerned with the object surface, which is commonly assumed to have Lambertian reflectance with a certain number of distinctive feature points [11]. That is, if the surface takes on specularities it will be a problem for the first stage to accurately compute point correspondence. At the second stage, in which the linear subspace constraint [15] is imposed, specular reflections will be even more problematic as they produce non-linear highlights [9, 14].

The motivation of this paper is to overcome such a limitation at each stage in the scheme. This paper is most relevant to the work of [10] and [12] in that we also investigate the linear subspace constraint. However, it differs from them in the challenge that we propose techniques to deal with specular components both in feature correspondence and dense depth search. Moreover, as a direct application we show that we can acquire a linear image basis of target object by using the consequence of dense matching and removal of specular components.

1.1   Overview

Our assumption in the following is that the target object is illuminated by a distant unknown light source. The surface may or may not be textured, and is approximated by a Lambertian component plus a specular component but without shadow.

Point correspondence   We first compute sparse feature correspondence across the input images which we use for estimating light source parameters as well as external camera parameters. In order to eliminate the spurious matching due to specularities at this stage, we compute outliers by minimising the LMedS error in terms of photometric measures and exclude them before computing the camera parameters and light source parameters.

Dense depth search   We propose to recover the shape of an object by a dense depth search while evaluating the correct matching with the linear subspace constraint as investigated in [10, 12]. In order to deal with specularities, in our case, we choose to employ a minimum of five images as inputs. This is because we first need three images, if no specularities are present, to estimate the surface normal (up to an ambiguity) at a hypothesised surface location as well as the fourth image to verify that the surface indeed passes through that point. The fifth image is to detect the existence of a specularity by checking the consistency with the other four inputs.

This detection of specularity is inspired by the principle of 4-source photometric stereo [1, 3, 17]. That is, among five or more input images we interchangeably utilise an optimal *intensity subset* consisting of four projected image intensities to find the best matches. The strategy is motivated by the fact that the specularities travel on the surface thanks to the object motion and therefore on the assumption that they are non-overlapping.

Linearising the basis   We also show that a linear image basis of the target object becomes available as a consequence of the dense matching despite the specularities. Although a basis can be generated by re-aligning each pixel in the input images, non-linear artifact would then appear in the by-product of simple re-alignment. Nevertheless, we factor out the specular components as residuals with respect to a linear combination of an intensity subset and compute linearised pixel intensities, which directly achieves a linearised image basis. Linear image bases are known to be useful for example to synthesise different appearances of an object as learning data for the problem of object recognition [4, 13].

In Section 2 we first introduce the procedure of computing light source parameters using the sparse point correspondence across input images. Section 3 proposes a core method for dense depth search in the presence of specular reflections on the surface, followed by Section 4 which describes how we can apply the result to generate a linearised image basis. We show the results of our experiments with real objects in Section 5. Section 6 concludes the paper.

2 Computing Light Source Parameters

For estimating camera parameters in the camera projection matrix, P, we solve the well known affine structure from motion problem using SVD [18]. We then need a measurement matrix containing the coordinates of some number of corresponding sample points throughout input images. This is generally considered possible with standard techniques for matching corners [5] and removing outliers unless the artifacts due to specularities (or some intensities not fulfilling the assumed conditions, e.g. self-shadowing) distort the solution. Accurate point correspondences under appearance change is also important because the intensities of the corresponding pixels parametrise the light source parameters, and thereby linear subspace constraint, which we will utilise for dense depth search at the subsequent stage.

2.1 The Light Source Matrix

Let us consider the intensity $I_i(j)$ of the ith point on the surface of a moving object projected into the jth image. Given the correspondences for n sample points through m images, we record the corresponding pixel intensities, $I_i(j)$, in an $n \times m$ matrix as $\mathscr{I} = (I_i(j))$, which we call the *illumination matrix*.

For Lambertian surface, we can express $I_i(j)$ in terms of the image formation equation process so that $I_i(j) = \max(\mathbf{n}_i^\top \mathbf{l}(j), 0)$. The 3-vector $\mathbf{n}_i$ is defined at the first frame to be the product of the albedo with the inward facing unit normal for the ith point whereas the 3-vector $\mathbf{l}(j)$ is the product of the strength of the light source with the unit vector for its direction. Note that

$$\mathbf{l}(j) = \mathbf{R}^\top(j)\,\mathbf{l}(1), \tag{1}$$

where the 3×3 matrix, $\mathbf{R}(j)$, is the rotation of the object from the first frame[1], which is taken to be the reference, to the jth frame. Multiplication of $\mathbf{R}^\top(j)$ represents virtually inverse rotation of the light source.

Then, we can form a matrix equation,

$$\mathscr{I} = NL \tag{2}$$

where N is an $n \times 3$ matrix containing the rows $\mathbf{n}_i^\top$, and L is a $3 \times m$ matrix containing the columns $\mathbf{l}(j)$. We call matrix L (as well as $\check{L}$ that appears in the sequel) *light source matrix* for convenience of explanation. The familiar form for solution by singular value decomposition to obtain a rank 3 approximation to the matrix $\mathscr{I}$ is such that

$$\mathscr{I} = \check{N}\check{L} = (\check{N}A^{-1})(A\check{L})\ . \tag{3}$$

As is well known, the solution is unique up to an arbitrary invertible 3×3 transformation A which transforms $\check{L}$ into L by $L = A\check{L}$, or for its each column $\check{\mathbf{l}}(j)$ into $\mathbf{l}(j)$ by $\mathbf{l}(j) = A\check{\mathbf{l}}(j)$ where $\check{\mathbf{l}}(j)$ denotes a column of $\check{L}$.

2.2 Computing The Light Source Matrix by LMedS

As stated above, we estimate the light source matrix, L, based on the sample point that are in correspondence. However, computing point correspondences on a moving object is not

---

[1]The rotation matrix is computed by solving for the structure from motion problem.

so trivial when the surface is with specular component that may cause spurious matching. A known solution to this problem in the literature would be by removing outliers in terms of geometric errors [19].

For an accurate computation of light source matrix, L, we carry this out by taking the photometric measures into account [2] in order to ensure that the artifacts which are caused by specularities do not distort the correct solution. In practice, we choose to employ the LMedS method without a particular need of a threshold. Namely, choosing a set of three sampled rows in the illumination matrix, $\mathscr{I}_3$, we can compute a putative light source matrix which we call $\check{L}_3$. We then evaluate the consistency of each remaining row of the illumination matrix, e.g. $\mathbf{I}_i = \{I_i(1), I_i(2), ..., I_i(m)\}$, with $\check{L}_3$ by computing the residual error

$$\varepsilon = |\mathbf{I}_i - \mathbf{I}_i \check{L}_3^\top (\check{L}_3 \check{L}_3^\top)^{-1} \check{L}_3|, \tag{4}$$

and record the median of ε^2. Note that the arbitrary transformation matrix A will disappear by substituting $\check{L}_3 = A^{-1} L_3$ to this definition and therefore the putative matrix $\check{L}_3$ can be directly used here. Repeating this procedure while picking up many different sample sets of three rows of illumination matrix, we select a combination of three rows (points) which minimises the median of the residual errors and compute the light source matrix accordingly with the inliers.

3   Dense Depth Search for a Specular Surface

For dense surface reconstruction we search for the depth to the object at each pixel in a reference image. We essentially need a minimum of three more input images (and thus four in total) to impose the three-dimensional linear subspace constraint as suggested in [10, 12]. Choosing the reference image to be the canonical (as seen in [6]), we search for the depth, Z, while comparing the set of image intensities, $I(j; P_j(\mathbf{x}, Z))|_{j=1,...,m; m \geq 4}$, at the coordinates determined by $\mathbf{x}$ in the first canonical image, guess of depth Z, and the projection matrix P_j in each frame. The task is now to impose the geotensity constraint [12], i.e., to evaluate the consistency between the referred set of intensities.

The linear subspace constraint[3] violates in the presence of non-linear specular reflection. Thus, we present an algorithm to deal with specularities by adding extra input images for detecting the existence of a specularity when evaluating the consistency. That is, we choose to employ five (or more) images as inputs, and interchangeably utilise an optimal *intensity subset* consisting of four projected image intensities to find the best matches. The idea can be well explained as an analogue to 4-source photometric stereo [3] and thus we brief the notion of it before describing our algorithm.

Given the estimate of depth at each pixel in a reference image, a dense depth map can be computed by updating with any well researched algorithm, for example, by minimising the multiview constraint using graphcuts [10].

3.1   The 4-source Photometric Stereo

In the conventional 3-source photometric stereo, if a point on the surface is subject to specular reflection from one of the three sources, the computed normal vector will be incorrect due

---

[2]Light source estimation is also viewed as robust model fitting in multiview reconstruction of textureless object [20] where the inliers are the contour generator points of visual hull.
[3]It is available for regions with or without texture.

to an elevated intensity value. By adding a fourth source, it becomes possible to compute a set of four surface vectors based on the four possible combinations of three light sources. This redundancy allows tagging and removal of the specular source [3]. Namely, for a Lambertian surface the four albedos would be identical, but a specularity in one light source will make three of the four albedos higher while the albedo computed without using that light source remains low. As a method to eliminate specular effects, Coleman and Jain define a relative deviation in the surface albedo R_{dev} at each point on the surface,

$$R_{dev} = \frac{1}{4R_{min}} \sum_{j=1}^{4} |R_j - R_{mean}| \quad , \tag{5}$$

where R_j is each estimated albedo and R_{min} is the smallest of these. Then, they regard R_{dev} greater than a threshold as indication of a specular component, and choose the combination of the three intensity values which have the smallest albedo for computing the surface normal. An improved scheme to determine a statistically more meaningful threshold is also proposed using the variance of the camera's intensity response [17].

3.2 The 5-source Algorithm for Dense Depth Search

Analogously to the 4-source photometric stereo, we add a fifth image for tagging and removal of specular source. That is, by adding a fifth source, we compute a set of five surface vectors, $\hat{\mathbf{n}}$ (although it is determined only up to an invertible 3×3 transform), at each guess of depth based on the five possible combinations of four virtual light sources. Then, instead of using all the elements of $I(j; P_j(\mathbf{x}, Z))|_{j=1,...,m;m=5}$, for an appropriate evaluation of referred set of image intensities we define an intensity subset by selectively skipping such $I(k; P_k(\mathbf{x}, Z))$ that is the most likely to be specular among the five different intensities. We denote such an intensity subset as $I_{\bar{k}}(j; P_j(\mathbf{x}, Z))|_{j=1,...,m, j \neq k}$. For this selection to be possible, we compute five different linear subspaces in advance.

Computing the light source matrices (multiple candidates) Let us consider that five images are given as input so that $m = 5$, and first assume that specular reflection occurs mostly in either of them according to the object motion. Given an $n \times m$ illumination matrix as $\mathscr{I} = (I_i(j))$, we define five different $n \times 4$ *illumination submatrices*, $\mathscr{I}_{\bar{k}}|_{k=1,...,5}$, by skipping the kth column, $I(k)$. We then recall equation 3 and compute a rank 3 approximation to each $\mathscr{I}_{\bar{k}}$, so that

$$\mathscr{I}_{\bar{k}} = \check{N}_{\bar{k}} \check{L}_{\bar{k}} = (\check{N}_{\bar{k}} A^{-1})(A\check{L}_{\bar{k}}), \tag{6}$$

where $\check{L}_{\bar{k}}|_{k=1,...,5}$ in this case is a 3×4 light source matrix containing the columns $\check{\mathbf{l}}(j)|_{j=1,...,4}$. It should be noted that we apply the LMedS method introduced in Section 2.2 to compute each candidate of light source matrix.

Selection of optimal intensity subset Given multiple representations of linear subspaces, $\check{L}_{\bar{k}}|_{k=1,...,5}$, at each pixel, $\mathbf{x}$, and at each guess of depth, Z, we compute a set of five surface vectors $\hat{\mathbf{n}}_{\bar{k}}$ as

$$\hat{\mathbf{n}}_{\bar{k}}^{\top}|_{k=1,...,5} = I_{\bar{k}}(j; P_j(\mathbf{x}, Z))\check{L}_{\bar{k}}^{\top}(\check{L}_{\bar{k}}\check{L}_{\bar{k}}^{\top})^{-1} \quad . \tag{7}$$

Using those estimates of surface vectors, as was the case in the 4-source photometric stereo,

we derive a relative deviation in the surface albedo, $|\hat{\mathbf{n}}_{\bar{k}}|$, such that

$$|\hat{\mathbf{n}}|_{dev} = \frac{1}{5|\hat{\mathbf{n}}|_{min}} \sqrt{\sum_{k=1}^{5} (|\hat{\mathbf{n}}_{\bar{k}}| - |\hat{\mathbf{n}}|_{mean})^2} \quad , \tag{8}$$

where $|\hat{\mathbf{n}}|_{mean}$ is the mean value of $|\hat{\mathbf{n}}_{\bar{k}}|$ and $|\hat{\mathbf{n}}|_{min}$ is the smallest of these, respectively. We then select the combination of the four intensity values which give the smallest estimate of albedo.

Computing the correspondence If a specularity is judged to be occurring when $k = h$, we pick $\check{\mathbf{I}}(j)$ from $\check{L}_{\bar{h}}$ to compute the estimated values of the intensities as,

$$\hat{I}_{\bar{h}}(j; P_j(\mathbf{x}, Z))|_{j=1,\dots,5, j \neq h} = \hat{\mathbf{n}}_{\bar{h}}^{\top} \check{\mathbf{I}}(j) \quad . \tag{9}$$

We then define the error function to evaluate the set of intensities $I(j; P_j(\mathbf{x}, Z))$ without being influenced by the suspected specularity as

$$e(\mathbf{x}, Z) = \sum_{j=1, j \neq h}^{5} (I(j; P_j(\mathbf{x}, Z)) - \hat{I}(j; P_j(\mathbf{x}, Z)))^2 \quad . \tag{10}$$

We measure the error, e, at regular small intervals of depth, Z. It is clear that as the depth parameter is varied the location of the corresponding points in each image will trace out the corresponding epipolar line in each image. When the depth is correct we expect the error to approach zero and we store a primary depth hypothesis that minimises the error $e(\mathbf{x}, Z)$ as well as several other depth candidates where $e(\mathbf{x}, Z)$ is locally minimised as has been studied in [2]. We then choose such candidate of depth Z that is most consistent with the primary depth hypotheses in the surrounding pixels.

4   Linearising the Image Basis

Based on the pixelwise correspondence between input images of an object in motion, we can generate images of the object as if they were captured in an identical pose as seen in the reference frame but under varying lighting. This is done by re-aligning the pixels in the input images. We can either employ a set of three images from the generated images directly as an image basis, or apply the PCA to the re-aligned images and form an orthogonal basis by the three eigenvectors corresponding to the highest eigenvalues [13]. In either case, however, non-linear artifacts will naturally appear in the resulting basis as long as specular reflections are present in the input images even when the correspondence is accurately computed despite the specularities. Since the linear functionality of the basis will then deteriorate, we wish to replace the specularities with some values that linearly agree to each other, i.e., linearise the appearance of image basis.

Let us consider the case that m images are given as input and that for simplicity we should typically replace an intensity, $I(h)$, that is most likely to be specular among a corresponding intensity set, $I(j)|_{j=1,\dots,m}$. For each corresponding set of pixels in a set of re-aligned images, we have a tag of the specular source such that the hth element was removed in the process of depth search, and the utilised intensity subset was

$$\hat{I}_{\bar{h}}(j; P_j(\mathbf{x}, Z))|_{j=1,\dots,m, j \neq h} \quad . \tag{11}$$

That is, we have used the intensity subset excluding the hth element for computing the correspondence at the correct depth. Since it should be consistent with the corresponding 3×4 light source matrix, $\check{L}_{\bar{h}}$, containing the columns $\check{\mathbf{I}}(j)|_{j=1,\dots,m-1}$ (excluding $\check{\mathbf{I}}(h)$), we can estimate a linearised value of $I(h)$ by

$$\hat{I}(h) = \hat{\mathbf{n}}_{\bar{h}}^{\top} \check{\mathbf{I}}(h) \ , \tag{12}$$

where $\hat{\mathbf{n}}_{\bar{h}}^{\top}$ is computed by equation 7 where $k = h$. We can then factor out the specular component of $I(h)$ as a residual with respect to the linear estimate by

$$I(h) - \hat{I}(h) \ . \tag{13}$$

The residual should be zero when no specularity is present, given the re-alignment is accurate. If some residual is observed, we can linearise the corresponding set of pixels by directly replacing $I(h)$ with $\hat{I}(h)$. This operation at each pixel in a set of re-aligned images allows us to linearise the entire image basis.

5 Experiments

We illustrate the performance of the proposed scheme using a few examples of real objects: a human face and a water pitcher, whose surfaces are composites of Lambertian and specular components. See Figure 1. Each of them shows three input images out of five that were used, respectively. They were captured while the object pose varies under a point light source whose position is close to the viewing direction, but unknown. Ambient light was also present in capturing the human face. As can be observed, weak or strong specularities appear on different parts of surfaces according to the object pose relative to the light source. Since a light source is placed beside the camera, in each image, parts of the surface where the orientation is close to the direction toward the camera tend to take on specular reflection. For example, the face in the leftmost image takes on relatively strong specular components in the left part of forehead and cheek according to the face direction.

Figure 1 also include resulting depth maps of water pitcher and human face, computed for the reference input images in the midst, respectively. The lighter, the closer to the camera. In both examples we tracked about 15 sample points for correspondence but a few of them were removed by the LMedS method with the photometric measure. For the dense depth search in each case, we employed a 15×15 template to suppress the error arising from image noise (but not to over-smooth the resulting surface). The depth maps essentially represent the accuracy of the computed dense correspondence between the input images, and we also display the reconstructed surfaces viewed from different directions. The surfaces are well recovered considering the fact that only five images are used, and uninfluenced by the partial specularities. The results on the upper part of the water pitcher is erroneous because that part is not visible throughout all the input images and is also with strong inter-reflections.

Figure 2 shows the reference image of water pitcher (same image that is in the midst of Figure 2 (left-top)) with annotations of two sample locations, (a) $\mathbf{x} = (160, 100)$ and (b) $\mathbf{x} = (200, 100)$ and illustrate the errors at hypothesised depths. Indicated as 'b' is an example point that takes on a specular reflection as the water pitcher rotates (see the leftmost-top image of Figure 2) whereas 'a' is a point that does not across the five input images. It should be noted that those two points are sampled at different locations but at an identical distance from the camera and therefore their depths should be estimated identically.
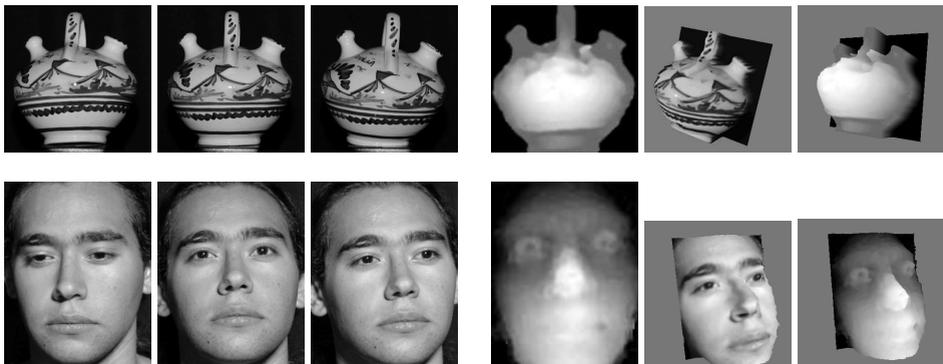
Figure 1: Left: Three examples out of five input images of a water pitcher (top) and a human face (bottom) captured in different poses under a point light source whose direction is unknown, respectively. Right: A depth map estimated referring the input image in the midst (the lighter, the closer) and the recovered surface with and without texture shown from different directions.

(a) The error $e(\mathbf{x}, Z)$ at 'a'. (b) The error $e(\mathbf{x}, Z)$ at 'b'.
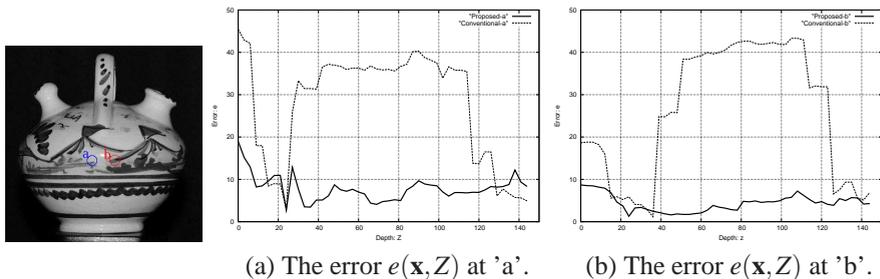
Figure 2: Left: The 320×320 reference image of water pitcher example (the same image as in the midst of Figure 2 (left)). Annotated as 'a' and 'b' are two sample locations for detailed analyses. Right: The errors $e(\mathbf{x}, Z)$ with respect to hypothesised depths, Z, are plotted for two sampled points. In both examples, the solid line indicates the errors by the proposed method and the dashed line by a conventional method [12].

In Figure 2 (a), with the solid line we plot the error e computed by equation 10 at different hypothesised depths for the sample point 'a'. Smaller values of Z correspond to smaller (closer) distances to the viewer. The error is minimised at $Z = 24$ and it is consistent with the true depth although other local minimum are also present. With the dashed line, for a comparison, we superimpose the error e computed by a conventional method [12] but using all the five input images. The minimum is found at the same correct depth, which is as expected since point 'a' does not take on specular reflection across the input images. The error, e, obtained by the proposed method turned out to be generally lower than those by the conventional ones. Recalling that it is defined in terms of grey scale values, we can explain the reason by the fact that (the square of) the highest pixel value among the referred set tends to be excluded from the computation in equation 10. Analogously, Figure 2 (b) shows the error e computed for point 'b'. In this case, the errors by the two methods are not minimised at the same depth. The result by the conventional method is obviously influenced by a specularity and the minimum error is realised at a shifted wrong depth. However, we can observe that the proposed method still minimises the error at the same correct depth as found for point 'a' (the two sample points are at identical distance).

Figure 3 (left) shows re-aligned images of the water pitcher (top), corresponding to the three input images shown in Figure 1 (left) in the same order. Each of them is also separated to specular components (middle) and a linearised image after removing the residuals (bottom). It can be observed that residuals are extracted in different parts of the surface according to the object motion. It is natural because the orientation of the object changes in such a way that it spans the 3D space so that specular regions travel accordingly. Comparing the top and bottom rows of the figure, it can be noticed that the strong specularity that is present due to the round shape of the object at a close point (viewed from the camera) has been removed in the linearised images. Analogously, Figure 3 (right) shows re-aligned images of the human face, each of which is separated to specular components as the residuals and a linearised image after removing them. Notice that some of the specular components, e.g. in the above mentioned face image are identified as residuals, and then alleviated in the linearised image.
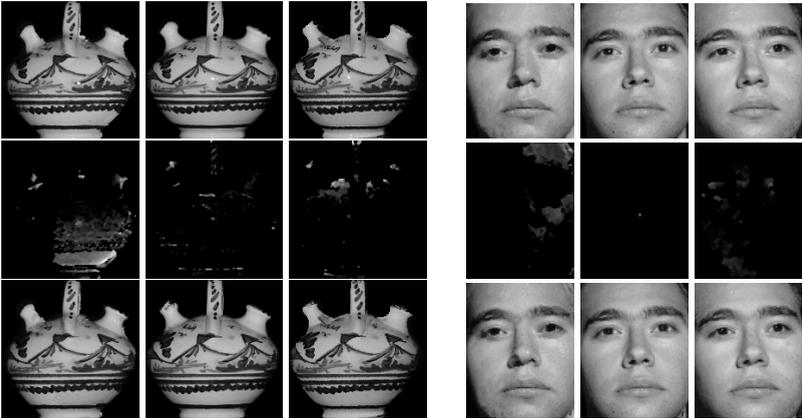
Figure 3: Top: Re-aligned images generated from the central part of pitcher images (left) and face images (right). Middle: Extracted specular components (residuals, stressed by factor 5 for display reason). Bottom: Linearised images after removing the specular components.

6 Summary and Future Work

We have addressed the problem of recovering the surface of a moving object using the shading variance due to object motion. In particular, we have dealt with objects with specular surfaces for which it was previously difficult to find accurate point correspondence and search for dense surface. The key idea for dense point matching is to use an optimal subset of intensities that is the least influenced by specular reflections. To our knowledge it is one of the few trials of finding correspondence in the presence of specularities on the top of intensity variance due to object motion. We have also shown that it is possible to linearise the input images by directly applying our results. It should be stressed that our experimental results were obtained using as few as five input images, with the fifth one to deal with a possible specularity. We are interested in extending our algorithm to accommodate a large number of input images so that it can deal with a more significant specularities. Other interesting potential directions for future work include combinations with different approaches for coping with highlights such as using silhouette cue [20] or shadows [7].

References

[1] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE-PAMI*, 25(10):1239–1252, 2003.

[2] N. D. F. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *10th ECCV*, pages 766–779, 2008.

[3] E.N. Coleman Jr. and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics And Image Processing*, 18:309–328, 1982.

[4] A.S. Georghiades, P.N Belhumeur, and D.J. Kriegman. From few to many: illumination cone models for face recognitionunder variable lighting and pose. *IEEE-PAMI*, 23:6:643–660, 2001.

[5] C.G. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–151, 1988.

[6] R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proc. Applications of Invariance in computer vision*, pages 238–256. Springer-Verlag, 1994.

[7] C. Hernández, G. Vogiatzis, and R. Cipolla. Shadows in three-source photometric stereo. In *ECCV (1)*, pages 290–303, 2008.

[8] B.K.P. Horn. *Robot Vision*. The MIT Press, 1992.

[9] H. Jin, A. J. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. In *3DPVT*, pages 626–631, 2002.

[10] N. Joshi and D.J. Kriegman. Shape from varying illumination and viewpoint. In *11th ICCV*, 2007.

[11] J. Lim, J. Ho, M.-H. Yang, and D. J. Kriegman. Passive photometric stereo from motion. In *10th ICCV*, pages 1635–1642, 2005.

[12] A. Maki, M. Watanabe, and C.S. Wiles. Geotensity: Combining motion and lighting for 3d surface reconstruction. *IJCV*, 48:2:75–90, 2002.

[13] A. Nakashima, A. Maki, and K. Fukui. Constructing illumination image basis from object motion. In *7th ECCV*, pages III:195–209, 2002.

[14] M. Oren and S. Nayar. A theory of specular surface geometry. In *5th ICCV*, pages 740–747, 1995.

[15] A. Shashua. *Geometry and photometry in 3D visual recognition*. PhD thesis, Dept. Brain and Cognitive Science, MIT, 1992.

[16] D. Simakov, D. Srolova, and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *9th ICCV*, pages 1202–1209, 2003.

[17] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE-PAMI*, 18(4):449–454, 1996.

[18] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 9:2:137–154, 1992.

[19] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[20] G. Vogiatzis, C. Hernández, and R. Cipolla. Reconstruction in the round using photometric normals and silhouettes. In *CVPR (2)*, pages 1847–1854, 2006.

[21] R.J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19:139–144, 1980.

[22] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *9th ICCV*, pages 618–625, 2003.