Visibility Subspaces: Uncalibrated Photometric Stereo with Shadows

Kalyan Sunkavalli, Todd Zickler, and Hanspeter Pfister

Harvard University
33 Oxford St., Cambridge, MA, USA, 02138
`{kalyans,zickler,pfister}@seas.harvard.edu`

Abstract. Photometric stereo relies on inverting the image formation process, and doing this accurately requires reasoning about the visibility of light sources with respect to each image point. While simple heuristics for shadow detection suffice in some cases, they are susceptible to error. This paper presents an alternative approach for handling visibility in photometric stereo, one that is suitable for uncalibrated settings where the light directions are not known. A surface imaged under a finite set of light sources can be divided into regions having uniform visibility, and when the surface is Lambertian, these regions generally map to distinct three-dimensional illumination subspaces. We show that by identifying these subspaces, we can locate the regions and their visibilities, and in the process identify shadows. The result is an automatic method for uncalibrated Lambertian photometric stereo in the presence of shadows, both cast and attached.

1   Introduction

Photometric stereo seeks to recover the geometry of a scene by analyzing appearance changes under varying illumination. In spite of being based on a crude reflectance model, Lambertian photometric stereo is one approach that is frequent used. One of the reasons for the utility of Lambertian photometric stereo is its support of auto-calibration. In the ideal case, given a set of images under varying, but unknown, directional lighting, it is possible to recover both a surface normal field and the light source directions up to a three-parameter family of solutions [7, 33].

Like any photometric stereo technique, uncalibrated Lambertian photometric stereo relies on inverting the image formation process. It seeks to explain observations using combinations of light sources, surface normals, and surface albedos; and in order to succeed, it must be able to reason effectively about which light sources are visible to each surface point. This problem is deceptively hard because shadowing is a non-local function of surface geometry, and heuristics for shadow detection, such as simple thresholding, are unreliable in the presence of albedo variations and sparse input images.

In this paper, we avoid explicit shadow detection by reasoning about illumination subspaces instead. It is well-known that the set of images of a convex

(a) Input (4 of 8)   (b) Visibilities   (c) Normals   (d) Reconstruction
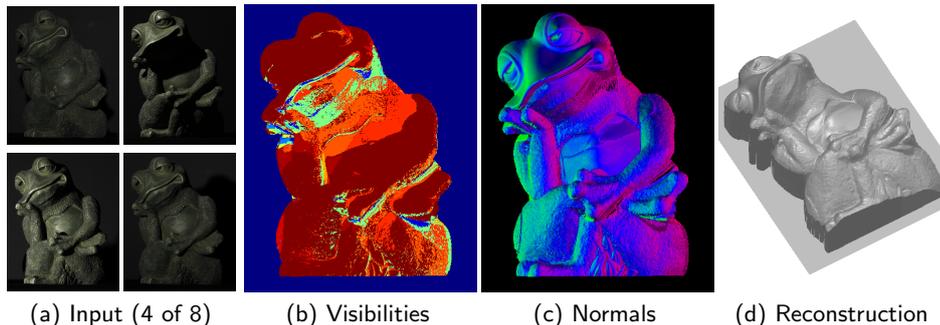
Fig. 1. *Uncalibrated photometric stereo with shadows. From a sparse set of images of a Lambertian scene (a), we identify regions that can see a common set of lights (b) through subspace estimation. This provides per-pixel visibility and allows the recovery of surface normals (c) and light directions up to the standard global linear ambiguity. Integrating these normals produces a reconstruction (d) that is not corrupted by the strong shadowing in the input images.*

Lambertian surface under directional lighting spans a three-dimensional linear subspace. It is also well-known that attached shadows and cast shadows violate this subspace property, so that the image-span of a scene with shadows can grow to a high dimension. What has not been fully exploited is that these high-dimensional spans have useful structure. We show that the image-span of any Lambertian scene captured under a discrete set of light sources with arbitrary shadowing can be decomposed into a set of three-dimensional subspaces. We refer to these as *visibility subspaces* because they correspond to sets of surface points that can see a common set of lights.

Given a sequence of uncalibrated photometric stereo images of a Lambertian object, the visibility subspaces can be automatically identified—without knowledge of the lighting directions—using well-known subspace clustering techniques. We show that once these subspaces are identified, the surface is partitioned, the exact set of lights that is visible to each region can be computed, and the surface and light directions can be reconstructed up to the usual global linear ambiguity.

2   Related Work

Photometric stereo can produce per-pixel estimates of surface normals and is a common technique for scene reconstruction. Originally developed for Lambertian surfaces and calibrated directional lighting [29], photometric stereo has been generalized to handle uncalibrated directional lights [15], specular and glossy surfaces [20, 21, 14], symmetric reflectance functions [1, 19, 25], reflectance mixtures [18], and uncalibrated environment map lighting [4]. Despite these generalizations, Lambertian photometric stereo remains useful because of its simplicity and allowance for uncalibrated acquisition, as well as being an analytical "stepping stone" for developing more comprehensive techniques.

In order to obtain accurate reconstructions with any photometric stereo technique, Lambertian or not, one must identify shadowed regions in the images. Most approaches for isolating shadows rely on using enough light sources such that every surface point is illuminated by at least two or three of them, and then detecting and discarding intensity measurements having low values. The number of images may be as few as three or four [10, 3, 16] but can also be many more [31, 30]. Since these methods detect shadows by analyzing the intensities at individual pixels, they can be unreliable when a surface has texture with low albedo, and when cast shadows prevent some surface points from being illuminated by a sufficient number of lights.

An alternative approach is proposed by Chandraker et al. [8]. They estimate which light sources can be seen by each surface point using a Markov random field in which the per-pixel "data term" is based on Lambertian photometric stereo and the "smoothness term" acts to encourage spatial coherence. This approach requires that the light directions are calibrated and known, and like the methods above, relies on reasoning about the intensities at each pixel. Our approach also derives from Lambertian photometric stereo, but unlike [8], does not require the light sources to be calibrated. Moreover, instead of reasoning about per-pixel intensities, it reasons about illumination subspaces.

Our work is also related to the problem of characterizing the structure of the set of a scene's images. There exist bounds on the dimension of the image-span of convex Lambertian scenes under directional lighting [23] and environment map lighting [5, 22], as well as convex scenes with a single arbitrary reflectance function [6] and mixtures of reflectance functions [13]. All of these bounds assume the scene to be convex so that cast shadows are absent. As a by-product of our analysis, we derive a complimentary bound that accommodates cast shadows and is valid for any Lambertian scene illuminated by a finite set of directional lights.

Finally, our work leverages insight from subspace clustering techniques, such as Generalized Principal Component Analysis (GPCA) [28] and Local Subspace Affinity (LSA) [32], that have been developed for motion segmentation. In our case, we perform subspace clustering using RANdom SAmple Consensus (RANSAC) [12, 26, 27]. This is quite different from a previous use of RANSAC in photometric stereo [17], which was aimed at identifying contour generators within an object's visual hull.

3   Visibility Subspaces

We begin with background and notation. For a Lambertian surface, the radiance from a surface point with normal $N \in \mathbb{S}^2$ and albedo ρ, illuminated with directional lighting L (i.e., with direction $L/||L|| \in \mathbb{S}^2$, and magnitude $||L||$), is given by $I = \max(0, \rho L^T N)$. In the absence of shadows, we know that $L^T N > 0$, and the image observations at m surface points illuminated by n light sources can be arranged as an $n \times m$ data matrix $\mathbf{I}$ that is the product of the $3 \times n$ lighting matrix $\mathbf{L} = [L_1, L_2, \cdots, L_n]$ and the $3 \times m$ albedo-scaled normals matrix

$\mathbf{N} = [\rho_1 N_1, \rho_2 N_2, \cdots, \rho_m N_m]$:

$$\mathbf{I} = \mathbf{L}^T \mathbf{N}. \qquad (1)$$

$\mathbf{L}$ and $\mathbf{N}$ are at most rank-three, and therefore, so is matrix $\mathbf{I}$ [29, 23].

If the scene is imaged under at least three non-coplanar light sources and these sources are calibrated and known, the surface normals can be estimated from noisy image intensities as $\mathbf{N} = (\mathbf{L}^T)^+ \mathbf{I}$, where $(\cdot)^+$ is the pseudo-inverse operator [29]. If the light sources are not calibrated, we can factor $\mathbf{I}$ using singular value decomposition (SVD) to recover the normals and lights using a rank-three approximation [15]:

$$\mathbf{I} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T, \quad \hat{\mathbf{L}}^T \triangleq \mathbf{U}_3 \boldsymbol{\Lambda}_3^{\frac{1}{2}}, \quad \hat{\mathbf{N}} \triangleq \boldsymbol{\Lambda}_3^{\frac{1}{2}} \mathbf{V}_3^T. \qquad (2)$$

This determines the normals up to a linear 3×3 linear ambiguity such that:

$$\mathbf{L}^T = \hat{\mathbf{L}}^T \mathbf{A}, \mathbf{N} = \mathbf{A}^{-1} \hat{\mathbf{N}}. \qquad (3)$$

for some non-singular matrix $\mathbf{A}$. This ambiguity can be resolved if light source intensities or surface albedos are known [15]. It can also be resolved up to the three-parameter generalized bas-relief ambiguity by enforcing an integrability condition on the normal field [7, 33].

Up to this point we have assumed the absence of cast and attached shadows, or equivalently, that every light source is *visible* to every surface normal. Now suppose that shadows exist, and consider the following toy example. A scene is partitioned into two uniform-visibility regions S_1 and S_2 that project to m_1 and m_2 pixels respectively. The scene is imaged under a set of n light directions that can be grouped into two (potentially) overlapping subsets $\mathbf{L}_1$ and $\mathbf{L}_2$, such that all of the lights $\mathbf{L}_1$ are visible to all points in S_1, and all of the lights $\mathbf{L}_2$ are visible to all points in S_2. Let the number of lights in these overlapping subsets be denoted by n_1 and n_2, and since they might overlap, we have $n_1 + n_2 \geq n$.

Now, the data matrix $\mathbf{I}$ can be permuted so that the first m_1 columns correspond to S_1 and last m_2 columns to S_2, and the first n_1 rows correspond to $\mathbf{L}_1$ and last n_2 rows to $\mathbf{L}_2$ with their shared lights lined up in the middle. Then, the observation matrix can be written as two sub-matrices, and if we denote by $\mathbf{N}_k$ the collection of surface normals in region S_k, the matrix can be factored as:

$$\mathbf{I} = [\, \mathbf{I}_1 \mid \mathbf{I}_2 \,] = \begin{bmatrix} \mathbf{L}_1^T & \mathbf{0}_{n-n_2}^T \\ \mathbf{0}_{n-n_1}^T & \mathbf{L}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{N}_1 & \mathbf{0}_{m_2} \\ \mathbf{0}_{m_1} & \mathbf{N}_2 \end{bmatrix}, \qquad (4)$$

with $\mathbf{0}_x$ representing a matrix of zeros with size $3 \times x$. The form of this factorization shows that while the row-space of $\mathbf{I}$ spans six dimensions, it actually consists of two rank-three subspaces corresponding to the two disjoint surface regions with different visibilities.

To generalize this to multiple regions with arbitrarily overlapping visibilities (i.e., sets of visible light sources), we define the *visibility vector* of region S_k to be the binary vector $V_k = [v_{k1}, v_{k2}, \cdots, v_{kn}]$, such that $v_{ki} = 1$ if light source L_i is visible to all the points in S_k and $v_{ki} = 0$ otherwise. The light sources visible

to region S_k can then be expressed (with a slight change in notation from Eq. 4) as

$$\mathbf{L}_k = \mathbf{L} \otimes V_k, \tag{5}$$

where $\otimes$ represents the element-wise Hadamard product applied to every row of the lighting matrix. As above, we can then factor the observation matrix for a scene with s distinct visibility regions as:

$$\mathbf{I} = [\ \mathbf{I}_1 \mid \mathbf{I}_2 \mid \cdots \mid \mathbf{I}_s\ ] = [\ \mathbf{L}_1^T \mid \mathbf{L}_2^T \mid \cdots \mid \mathbf{L}_s^T\ ] \begin{bmatrix} \mathbf{N}_1 & & & \\ & \mathbf{N}_2 & & \\ & & \ddots & \\ & & & \mathbf{N}_s \end{bmatrix}, \tag{6}$$

where $\mathbf{N}_k$ is the surface normal matrix corresponding to region S_k.

Thus, the observation matrix is made up of multiple subspaces, and we call these *visibility subspaces* because they correspond to regions in the scene that each have a consistent set of visible lights. Clearly, each subspace is at most rank-three, and the row space of a scene with s visibility subspaces has dimension at most $3s$. This leads us to the following:

Proposition. *The set of all images of a Lambertian scene illuminated by any combination of n directional light sources lies in a linear space with dimension at most $3 \cdot 2^n$.*

Proof: A scene illuminated by n light sources will have at most 2^n regions with distinct visibility configurations. The images of each region span at most a three-dimensional space, so the dimension of the image-span of the entire scene is at most $3 \cdot 2^n$.

This result is complementary to previous work that has established bounds on the dimensionality of scene appearance. Belhumeur and Kriegman [6] showed that the images of a scene with an arbitrary uniform BRDF, and illuminated by distant (environment map) lighting, lie in a linear space whose dimension is bounded by the number of distinct surface normals in the scene. Garg et al. [13] generalized this to spatially-varying reflectances that can be expressed as a linear combination of basis BRDFs. However, these results apply only to convex scenes without attached or cast shadows. In addition, these results assume that there are a finite number of normals in the scene to derive a bound on the dimensionality of scene appearance under arbitrary directional (environment map) lighting. In contrast, our analysis provides bounds on the appearance of a Lambertian scene with an arbitrary number of normals but illuminated by a finite number of light sources, and allows any form of shadowing.

In general, we do not know the visibility subspaces of a scene *a priori*, and we cannot permute the rows and columns of the observation matrix to directly obtain the factorization in Eq. 6. However, as we show next, we can identify the subspaces automatically using a subspace clustering technique.

4   Estimating Visibility Subspaces

RANSAC [12] is a statistical method for fitting models of known dimensions to data with noise and outliers. While RANSAC is traditionally used to discard outliers from a dataset, we follow [27] and use it to cluster subspaces. In this context, it can be seen as an alternative to other subspace-estimation techniques, such as GPCA [28] and LSA [32].

Each visibility subspace of the scene is contained in a three-dimensional space. If we randomly choose three surface points that happen to be in the same region S_k, the light estimates $\hat{\mathbf{L}}_k$ that we obtain by factoring the image intensities at these three points (using Eq. 2) will accurately explain the intensities for all pixels in S_k. Thus, we expect a large number of "inliers". (Of course, there will be outliers as well because the points in the remainder of the scene will not have the same set of visible lights, and projecting their intensities onto $\hat{\mathbf{L}}_k$ will produce large errors.) Conversely, if we happen to choose three scene points that are in different regions, the light directions obtained by SVD will be unlikely to accurately explain the intensities at many other scene points, and we expect the number of inliers to be small. These observations suggest the following algorithm:

1. Choose three pixels at random and factor their intensities as $\mathbf{I}_3 = \hat{\mathbf{L}}_3^T \hat{\mathbf{N}}_3$.
2. Use lights $\hat{\mathbf{L}}_3$ to estimate the normal at all the surface points as $\hat{N}_i = (\hat{\mathbf{L}}_3^T)^+ I_i$.
3. Compute the per-pixel error of the estimated lights and normals as $E_i = ||I_i - \hat{\mathbf{L}}_3^T \hat{N}_i||^2$.
4. Mark points with error $E_i < \epsilon$ as inliers and recompute the associated optimal lighting $\hat{\mathbf{L}}_k$ using intensities for all inliers.
5. Repeat steps 1 through 4 for t iterations, or until a sufficiently large set of inliers has been found. During these iterations, keep track of the largest set of inliers found.
6. Mark the largest set of points that are inliers as a valid visibility subspace S_k with associated lighting basis $\hat{\mathbf{L}}_k$. Remove these inliers from the point set, and repeat steps 1 to 5 until all visibility subspaces have been recovered.

This procedure samples the points in the scene to find three points that belong to the same visibility subspace. Each time the sampling is successful, as measured by the number of inliers in Step 4, it extracts the subspace and removes it from the set of unlabeled points. The algorithm does not depend on the scene geometry or the lighting directions; it depends only on the rank-three condition of any visibility subspace. The result of the procedure is the set of per-pixel surface normals $\hat{\mathbf{N}}$, the per-pixel subspace labels $\mathbf{S}$, and a redundant (per-subspace) set of estimates for the light directions $\{\hat{\mathbf{L}}_k\}$. Note that in an uncalibrated setting, the set of normals for each subspace and their corresponding lights $\hat{\mathbf{L}}_k$ are defined up to their own linear ambiguity per Eqs. 2 and 3.

In our experiments, we use $t = 1000$ iterations, set the error threshold ϵ according to the noise in the input images, and run the procedure until 99% of the pixels are assigned to a valid visibility subspace. The remaining 1% of pixels are assigned to the subspace that best explains their intensity variation.

4.1   Degenerate Subspaces

The RANSAC-based method described above assumes that all visibility subspaces have rank-three. This is valid for any region having at least three non-coplanar surface normals, and illuminated by at least three non-coplanar light sources. However, in general, scenes may contain rank-deficient subspaces that corrupt the clustering. Under the assumption that every point in the scene sees at least three non-coplanar lights (without which surface normal recovery is ambiguous), a visibility subspace can only be rank-deficient if it has degenerate normals: a region with coplanar normals will have rank two and a planar region will have rank one. Our task, then, is to check our recovered rank-three subspaces to see if they are composed of smaller degenerate subspaces.

Given the form of the observation matrix factorization in Eq. 6, it follows that a rank-three subspace can only be one of the following three types:

1. A region with a single visibility vector and non-coplanar normals (i.e., a true rank-three subspace).
2. Two regions with distinct visibility vectors, where one region has coplanar normals, and the other is planar (i.e., a combination of rank-two and rank-one subspaces).
3. Three regions with distinct visibilities, each of which is planar (i.e., a combination of three rank-one subspaces).

To ensure that our subspaces estimated by RANSAC are not of Type 2 or Type 3, we test every estimated rank-three subspace by searching for embedded rank-two and rank-one subspaces. If the number of pixels corresponding to the smaller embedded subspaces subsume more than a fraction α of the original set ($\alpha = 0.5$ in our experiments) we relabel them as being members of a different rank-deficient subspace.

5   Subspaces to Surface Normals

This subspace clustering identifies surface regions with uniform visibility, but does not provide a clean visibility vector V_k (or accurate shadows) for each region. Put another way, the non-visible entries of each $\hat{\mathbf{L}}_k$ are not necessarily zero-valued. To recover the visibility vectors and refine the light matrices, we separately examine the light estimates in each subspace $\hat{\mathbf{L}}_k = [\hat{L}_{k1}, \hat{L}_{k2}, \cdots, \hat{L}_{kn}]$, and provided that the subspace is not degenerate, we set

$$v_{ki} = ||\hat{L}_{ki}^T|| > \tau, \tag{7}$$

with $\tau = 0.25$ in our experiments. This simple approach succeeds because the normals $\hat{\mathbf{N}}_k$ in each non-degenerate subspace span three dimensions, so the product $I_{ki} \approx \hat{L}_{ki}^T \hat{\mathbf{N}}_k$ can be zero only if the light strength $||\hat{L}_{ki}||$ is zero. Effectively, we are able to recover the visibility for each subspace by reasoning about the magnitude of the subspace lighting—an approach that is independent of scene albedo and is, therefore, not confounded by texture.

To estimate the visibility for degenerate subspaces, we first project the subspace lighting onto the column-space of the subspace normals before thresholding their magnitudes. This removes the component of the lighting orthogonal to the subspace normals that could be arbitrarily large while not contributing to the observed intensities.

Once the visibility vector for each subspace is known, we can recover the surface normals and reconstruct the surface. In the calibrated case, this is quite straightforward. Since the light sources $\mathbf{L}$ are known, they are combined with the visibility vectors using Eq. 5, and then the normals in every subspace are given by:

$$\mathbf{N}_k = (\mathbf{L} \otimes V_k)^+ \mathbf{I}_k \ , \ k = 1 \ldots s. \tag{8}$$

If the light sources are *not* calibrated, the situation is more complex because the subspace clustering induces a distinct linear ambiguity in each subspace, (i.e., $\mathbf{L}_k^T = \hat{\mathbf{L}}_k^T \mathbf{A}_k, \mathbf{N}_k = \mathbf{A}_k^{-1} \hat{\mathbf{N}}_k, k = 1 \ldots s$). Recovering the entire surface up to a single global ambiguity $\mathbf{A}$, which is the best we can do without additional information, requires that we somehow determine the transformations–one per subspace–that map each set of normals to a common coordinate system. Fortunately, this can be achieved by solving the set of linear equations:

$$\hat{\mathbf{L}} \otimes V_k = \hat{\mathbf{L}}_k \mathbf{A}_k^T \ , \ k = 1 \ldots s, \tag{9}$$

where both the global lights $\hat{\mathbf{L}}$ (i.e., those defined up to a single global ambiguity) and the per-subspace ambiguity matrices $\mathbf{A}_k$ are unknown. This is an over-constrained homogeneous system of linear equations since, for n lights and s subspaces, it contains $3ns$ constraints and $3n + 9s$ unknown variables. To avoid the trivial solution $\hat{\mathbf{L}} = \mathbf{A}_k = 0$ we set the ambiguity matrix for one reference subspace (chosen to be the non-degenerate subspace with the largest number of visible lights) to be the identity matrix. Accordingly, we recover the global lights $\hat{\mathbf{L}}$ and normals $\hat{\mathbf{N}}$ up to a single 3×3 ambiguity, which is that of the reference subspace.

To handle degenerate subspaces in the uncalibrated case, we first solve Eq. 9 using all non-degenerate subspaces, and as long as all of the global lights are visible to at least one of these regions, we can recover all of them. We then use these "auto-calibrated" lights to solve for the normals in the degenerate rank-one and rank-two subspaces using Eq. 8.

As a final step in the uncalibrated scenario, we may reduce or eliminate the global ambiguity using additional constraints, such as integrability of the normal field [7, 33], specular or glossy highlights [14, 11, 25], interreflections [9], or a prior model of object albedo [2, 24]. Then, in either calibrated or uncalibrated conditions, the estimated normals can be integrated to recover scene depth. In this integration process, one may optionally enforce the depth constraints that are induced by the visibility vectors and lights, and an elegant procedure for doing so can be found in [8].

(a) Input images

(b) True subspaces   (c) Our subspaces   (d) Our normals   (e) Reconstruction
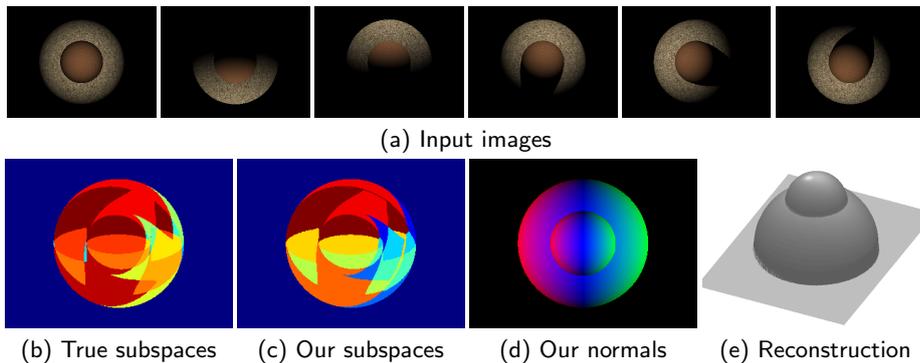
Fig. 2. *Spheres sequence. Attached and cast shadows divide this scene into intricate visibility subspaces (b). We are able to recover them almost perfectly (c), and estimate the surface normals (d) and depth (e) accurately.*

6   Results

We evaluate the uncalibrated instantiation of our approach on two synthetic datasets and two captured datasets. In each case, we automatically cluster subspaces, determine visibility vectors, and compute lights and surface normals up to a global 3×3 linear ambiguity. As mentioned above, there are ways to resolve this ambiguity, and since this is not the focus of this work, we simply do so by manual intervention.

For synthetic examples, we evaluate the recovered normals, lights, and visibility subspaces by comparing them to the ground-truth values that are used to synthesize the input images. For the captured examples, the "true" values for comparison are obtained as follows. First, we acquire a dense set of calibrated photometric stereo images using approximately 50 different light directions. From such a dense set of calibrated images, we can robustly estimate surface albedos, and the image intensities can be reliably thresholded to detect per-pixel shadows and "true" visibilities. Then, we discard the shadowed measurements and recover the "true" normals via calibrated Lambertian photometric stereo. To make a direct comparison between this ground truth and our results, we execute our algorithm using a small subset of the dense input images, with the calibration information held out.

Figure 2 is a synthetic example in which the attached and cast shadows induce intricate visibility subspaces. From the six input images, our approach recovers the visibilities and normals almost perfectly. Figure 3 is a similar example, but in this case, the shadows cast on the back plane create degenerate visibility subspaces. These degenerate rank-one and rank-two subspaces are successfully detected by our approach, and the final visibilities and normals computed from the seven input images are again very close to ground truth. The median angular errors in surface normals for these two examples are 0.49° and 0.51°, respectively. Note that both of these synthetic scenes have high-frequency texture and large

(a) Input images

(b) True subspaces   (c) Our subspaces   (d) Our normals   (e) Reconstruction
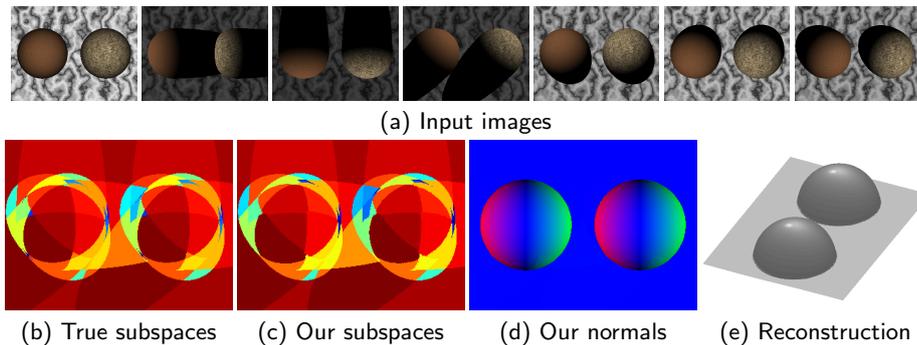
Fig. 3. *Spheres and plane sequence. The shadows cast by the spheres on the plane create degenerate subspaces (b). We are able to disambiguate them and recover the visibility subspaces (c) and surface normals (d), and reconstruct the scene (e).*

variations in albedos. These conditions often lead to poor results when using intensity-based shadow detection from such a small number of images, but this is not the case for the proposed method.

In the two captured datasets we consider – the frog (Fig. 4) and scholar (Fig. 5) sequences – our algorithm was given 8 and 12 input images, respectively. For each of these datasets, we compare to the "true" normals and visibilities obtained from densely-sampled calibrated images as described above. We also compare the normals to those obtained using calibrated Lambertian photometric stereo applied to the same smaller set of (8 and 12) images that are available to our algorithm. We give this algorithm access to both the calibrated light directions as well as the ground truth visibilities. We refer to these normals as the "best calibrated" normals because they can be interpreted as calibrated Lambertian photometric stereo supplied with "perfect" shadow detection, or equivalently, as the best-possible result from a calibrated shadow-detection method, such as [8, 30] applied on this small set of input images.

The input images have significant cast and attached shadows, and they exhibit non-idealities such as mutual illumination and slight specularity. Despite this, our method does reasonably well at locating the visibility subspaces (and shadows) from a small number of images. The median angular errors in the estimated normals (relative to the ground truth) are $7.44°$ and $4.45°$ for the frog and scholar datasets, respectively. The largest errors are made in regions with few non-shadowed measurements and where mutual illumination is most significant. This is not unique to our approach, however, and the errors from calibrated Lambertian photometric stereo with perfect shadow detection have a very similar structure. This suggests that our approach, which automatically handles shadows and is uncalibrated, introduces limited additional errors compared to an ideal calibrated algorithm.

(a) True subspaces   (b) True normals   (c) "Best calibrated" normals   (d) "Best calibrated" angular error

(e) Our subspaces   (f) Our normals   (g) Reconstruction   (h) Our error
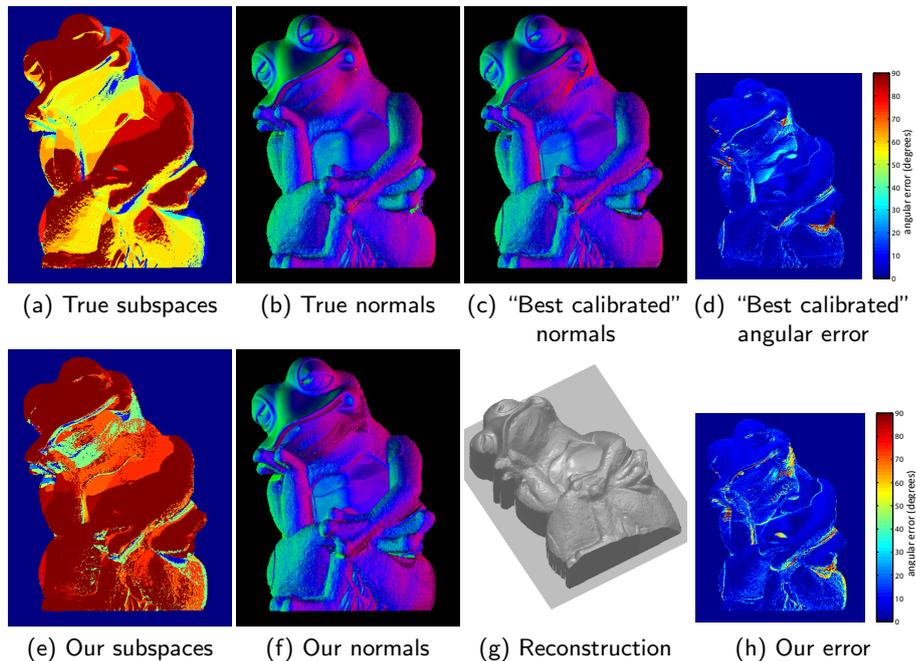
Fig. 4. *Frog dataset. Reconstruction results from sparse input images (shown in Fig. 1). Despite slight specularity and convexities with mutual illumination, our estimated subspaces (e) match the ground truth (a) reasonably well. The angular differences between our normals (f) and ground truth normals (b) are most significant in regions having few non-shadowed measurements (h). For comparison, the normals estimated using calibrated photometric stereo equipped with perfect shadow detection (c) exhibit similar deviations from the ground truth (d).*

7 Conclusion

We formulate shadow-detection in Lambertian photometric stereo as a subspace clustering task. This avoids heuristic reasoning about the intensities at individual pixels, and it allows handling cast and attached shadows in uncalibrated conditions when only a small number of input images are available. In addition, we derive a bound on the dimension of the image-span of a Lambertian scene under a discrete set of lights, and this bound has the rare property of incorporating arbitrary shadowing.

Unlike many previous approaches to shadow detection [8, 16], ours does not impose a preference for spatial coherence while detecting shadow regions. Indeed, we find that subspace clustering naturally leads to relatively coherent regions without this imposition. It is quite likely, however, that incorporating a spatial coherence constraint during subspace clustering could improve the results, especially in the presence of non-idealities like mutual illumination, and this may be a fruitful direction for future research.

Also, we have restricted ourselves to Lambertian scenes illuminated by directional lights, and it is worth considering how this analysis can be extended to handle more general conditions. In particular, one might consider general environment map lighting [4], where a proper consideration of visibility would overcome the current (and severe) restriction to convex surfaces.

References

1. Alldrin, N., Kriegman, D.: Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In: Proc. IEEE Int. Conf. Computer Vision (2007)
2. Alldrin, N., Mallick, S., Kriegman, D.: Resolving the generalized bas-relief ambiguity by entropy minimization. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2007)
3. Barsky, S., Petrou, M.: The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. IEEE Trans. Pattern Anal. Mach. Intell. 25, 1239–1252 (2003)
4. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. Int. Journal of Computer Vision 72(3), 239–257 (2007)
5. Basri, R., Jacobs, D.W.: Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Anal. Mach. Intell. 25(2), 218–233 (2003)
6. Belhumeur, P.N., Kriegman, D.J.: What is the set of images of an object under all possible illumination conditions? Int. Journal of Computer Vision 28(3), 245–260 (1998)
7. Belhumeur, P.N., Kriegman, D.J., Yuille, A.L.: The bas-relief ambiguity. Int. Journal of Computer Vision 35(1), 33–44 (1999)
8. Chandraker, M., Agarwal, S., , Kriegman, D.: Shadowcuts: Photometric stereo with shadows. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2007)
9. Chandraker, M., Kahl, F., Kriegman, D.: Reflections on the generalized bas-relief ambiguity. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2005)
10. Coleman, E., Jain, R.: Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. Computer Graphics and Image Processing 18(4), 309–328 (April 1982)
11. Drbohlav, O., Chaniler, M.: Can two specular pixels calibrate photometric stereo? In: Proc. IEEE Int. Conf. Computer Vision (2005)
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24(6), 381–395 (1981)
13. Garg, R., Du, H., Seitz, S.M., Snavely, N.: The dimensionality of scene appearance. In: Proc. IEEE Int. Conf. Computer Vision (2009)
14. Georghiades, A.: Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. In: Proc. IEEE Int. Conf. Computer Vision (2003)
15. Hayakawa, H.: Photometric stereo under a light source with arbitrary motion. J. Opt Soc. Am. 11(11) (1994)

16. Hernández, C., Vogiatzis, G., Cipolla, R.: Shadows in three-source photometric stereo. In: Proc. European Conf. Computer Vision (2008)
17. Hernández Esteban, C., Vogiatzis, G., Cipolla, R.: Multiview photometric stereo. IEEE Trans. Pattern Anal. Mach. Intell. 30(3), 548–554 (2008)
18. Hertzmann, A., Seitz, S.M.: Example-based photometric stereo: Shape reconstruction with general, varying brdfs. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1254–1264 (2005)
19. Holroyd, M., Lawrence, J., Humphreys, G., Zickler, T.: A photometric approach for estimating normals and tangents. ACM Trans. Graph. 27(5), 1–9 (2008)
20. Ikeuchi, K.: Determining surface orientations of specular surfaces by using the photometric stereo method. IEEE Trans. Pattern Anal. Mach. Intell. 3(6), 661–669 (1981)
21. Nayar, S., Ikeuchi, K., Kanade, T.: Determining shape and reflectance of hybrid surfaces by photometric sampling. IEEE T. Robotics Automation 6(4) (1990)
22. Ramamoorthi, R., Hanrahan, P.: On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object. J. Optical Society of America A 18(10), 2448–2458 (October 2001)
23. Shashua, A.: On photometric issues in 3D visual recognition from a single 2D image. Int. Journal of Computer Vision 31(1), 99–122 (1997)
24. Shi, B., Matsushita, Y., Wei, Y., Xu, C., Tan, P.: Self-calibrating photometric stereo. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2010)
25. Tan, P., Zickler, T.: A projective framework for radiometric image analysis. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2009)
26. Torr, P., Faugeras, O., Kanade, T., Hollinghurst, N., Lasenby, J., Sabin, M., Fitzgibbon, A.: Geometric motion segmentation and model selection [and discussion]. Philosophical Transactions: Mathematical, Physical and Engineering Sciences 356(1740) (1998)
27. Tron, R., Vidal, R.: A benchmark for the comparison of 3-D motion segmentation algorithms. In: IEEE Conf. Computer Vision and Pattern Recognition (2007)
28. Vidal, R., Hartley, R.: Motion segmentation with missing data by powerfactorization and generalized PCA. In: IEEE Conf. Computer Vision and Pattern Recognition (2004)
29. Woodham, R.: Photometric stereo: A reflectance map technique for determining surface orientation from image intesity. In: Proc. SPIE. vol. 155, pp. 136–143 (1978)
30. Wu, T.P., Tang, C.K.: Photometric stereo via expectation maximization. IEEE Trans. Pattern Anal. Mach. Intell. 32(3), 546–560 (2010)
31. Wu, T.P., Tang, K.L., Tang, C.K., Wong, T.T.: Dense photometric stereo: A markov random field approach. IEEE Trans. Pattern Anal. Mach. Intell. 28(11), 1830–1846 (2006)
32. Yan, J., Pollefeys, M.: A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In: Proc. European Conf. Computer Vision (2006)
33. Yuille, A., Snow, D.: Shape and albedo from multiple images using integrability. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (1997)

(a) Input images (8 of 12)

(b) True subspaces

(c) Our subspaces

(d) True normals

(e) Our normals

(f) "Best calibrated" normals

(g) Reconstruction

(h) "Best calibrated" angular error

(i) Our angular error
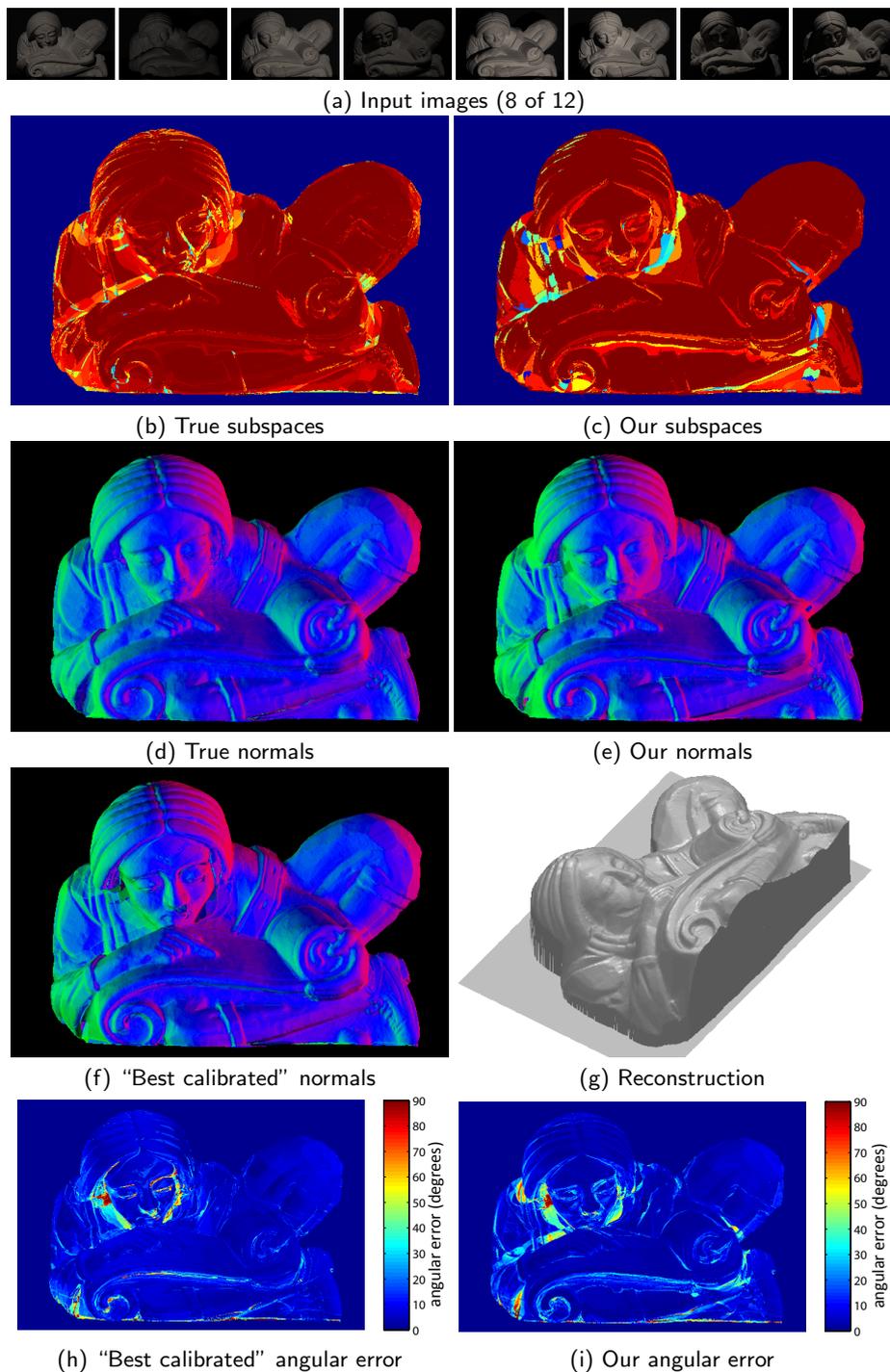
Fig. 5. *Scholar dataset. The left column shows ground truth (b,d) and normals obtained by calibrated photometric stereo applied to sparse input images (f). Our results with the same sparse set of images (a) are shown in the right column (c,e,g). The angular differences between the true normals (d) and our estimates (e) show that most errors are small and that large errors are restricted to small regions with strong inter-reflections (i). For comparison, the calibrated result (f) also exhibits similar deviations (h).*