Robust Photometric Stereo using Sparse Regression

Satoshi Ikehata[1*]   David Wipf[2]   Yasuyuki Matsushita[2]   Kiyoharu Aizawa[1]
[1]The University of Tokyo, Japan   [2]Microsoft Research Asia

Abstract

This paper presents a robust photometric stereo method that effectively compensates for various non-Lambertian corruptions such as specularities, shadows, and image noise. We construct a constrained sparse regression problem that enforces both Lambertian, rank-3 structure and sparse, additive corruptions. A solution method is derived using a hierarchical Bayesian approximation to accurately estimate the surface normals while simultaneously separating the non-Lambertian corruptions. Extensive evaluations are performed that show state-of-the-art performance using both synthetic and real-world images.

1. Introduction

Photometric stereo recovers surface normals of a scene from appearance variations under different lightings. The early work of Woodham [15] has shown that when a Lambertian surface is illuminated by at least three known lighting directions, the surface orientation can be uniquely determined from the resulting appearance variations using least-squares. In reality, however, the estimation may be disrupted by common non-Lambertian effects such as specular highlights, shadows, and image noise. Photometric stereo approaches to dealing with these problems are categorized into two classes: one uses a more sophisticated reflectance model than a simple Lambertian model to account for non-Lambertian effects [10, 6, 4, 1], and the other uses statistical approaches to eliminate the non-Lambertian effects [9, 17, 8, 11, 16]. Our work is more related to the latter class of *robust* approaches.

The robust approaches generally attempt to eliminate non-Lambertian observations by treating them as outliers, *e.g.*, a RANSAC (random sample consensus) scheme [9, 5], a Big-M approach [17], median filtering [8], and a graph-based approach [11]. While effective, these methods often require extensive computations to derive a stable solution. Our work is motivated by an alternative photometric stereo formulation by Wu *et al.* [16] that explicitly accounts for

sparse corruptions using R-PCA (Robust Principal Component Analysis). However, unlike Wu *et al.*'s approach where rank minimization is employed, our method explicitly uses a rank-3 Lambertian constraint and solves the problem using efficient sparse regression tools. This eliminates the need for specifying a shadow mask, which was needed in [16], and achieves significantly more accurate estimation of surface normals.

The primary contributions of this work are twofold. First, we present a practical photometric stereo method designed for highly corrupted observations. We cast the problem of photometric stereo as a well-constrained problem of sparse regression. By introducing a sparsity penalty that accounts for non-Lambertian effects, our method uniquely decomposes an observation matrix of stacked images under different lighting conditions into a Lambertian, rank-3 component and a sparse error component without having to explicitly minimize rank as in previous methods. Secondly, we perform analytical and empirical studies of computational aspects of sparse regression-based photometric stereo. Namely, we assess two different deterministic algorithms to solve the photometric stereo problem, a convex ℓ_1-norm based relaxation [2], and a hierarchical Bayesian model derived from a SBL (Sparse Bayesian Learning) framework [12, 14]. Our detailed discussion and extensive experiments show that SBL has various advantages over the ℓ_1-norm-based relaxation relevant to photometric stereo. Furthermore, we show that our method does not require careful tuning of parameters.

2. Photometric stereo via sparse regression

In this section, we formulate the estimation of surface normals using photometric stereo as a particular sparse linear inverse problem. Henceforth we rely the following assumptions:
(1) The relative position between the camera and the object is fixed across all images.
(2) The object is illuminated by point light sources at infinity whose positions are known.
(3) The camera view is orthographic and the sensor response function is linear.
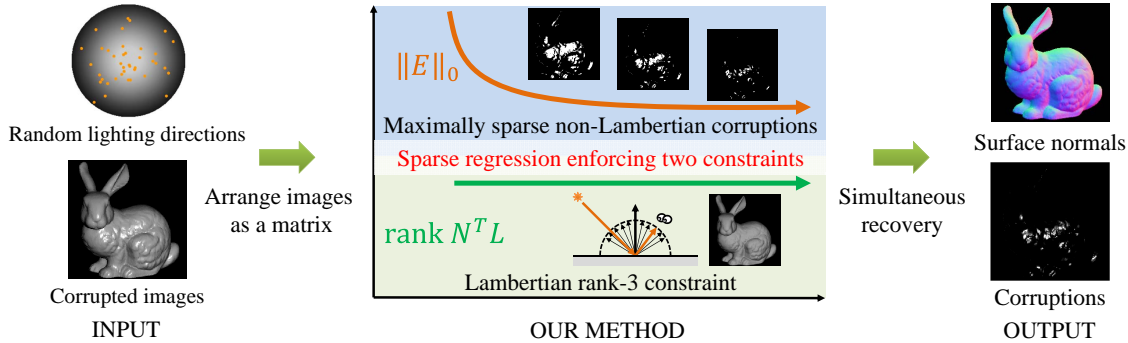
Figure 1. Illustration of our approach. Our method simultaneously recovers both surface normals and corruptions as the solution of a well-constrained problem of sparse regression which enforces both Lambertian rank-3 constraint and sparsity of corruptions.

2.1. Lambertian image formation model

Woodham *et al.* [15] revealed that the intensity I of a point in a Lambertian scene under a lighting direction $\boldsymbol{l} \in \mathbb{R}^3$ is expressed as follow,

$$I = \rho\, \boldsymbol{n}^T \boldsymbol{l}, \tag{1}$$

where $\rho \in \mathbb{R}$ is the diffuse albedo, and $\boldsymbol{n} \in \mathbb{R}^3$ is the surface normal at the point. Given n images with m pixels, we define an observation matrix by aligning each image as a vector:

$$O \triangleq [\text{vec}(I_1)|\dots|\text{vec}(I_n)] \in \mathbb{R}^{m \times n}, \tag{2}$$

where $\text{vec}(I_k) \triangleq [I_k(1), \dots, I_k(m)]^T$ for $k = 1, \dots, n$, and $I_k(j) = \rho_j\, \boldsymbol{n}_j^T \boldsymbol{l}_k$.

Therefore, the observations in a Lambertian scene can be expressed via the rank-3 expression

$$O = N^T L, \tag{3}$$

where $N = [\rho_1 \boldsymbol{n}_1 \mid \dots \mid \rho_m \boldsymbol{n}_m] \in \mathbb{R}^{3 \times m}$ and $L = [\boldsymbol{l}_1 \mid \dots \mid \boldsymbol{l}_n] \in \mathbb{R}^{3 \times n}$.

2.2. Non-Lambertian effects as sparse errors

In real scenes, various effects beyond the Lambertian formulation are observed, *e.g.*, specularities, shadows, image noise and so on. We can interpret them as additive corruptions $E \in \mathbb{R}^{m \times n}$ applied to an otherwise ideal Lambertian scene leading to the image formation model as [16]

$$O = N^T L + E. \tag{4}$$

Given observed images O and lighting directions L, our goal is to recover surface normals N as a part of the Lambertian diffusive component $N^T L$ in the presence of non-Lambertian corruptions E. However, this is an under-constrained problem since the number of unknowns exceeds the number of linear equations. While most previous methods recover surface normals from the Lambertian component purified by traditional outlier removal techniques [9, 17, 8] , we try to recover N without explicitly removing corruptions in a separate step. The overview of

the proposed method is illustrated in Fig. 1. An essential ingredient is a sparsity penalty applied to E, whose minimization disambiguates the infinity of feasible solutions to Eq. (4). This penalty quantifies the reasonable observation that non-Lambertian effects emerge primarily in limited areas of each image. For example, specularities surround the spot where the surface normal is oriented halfway between lighting and viewing directions, while shadows are created only when $L^T N \le 0$ (attached shadow) or when a non-convex surface blocks the light (cast shadow). Strictly speaking, we assume that the optimal feasible solution to Eq. (4) produces a sparse error matrix. Reflecting this assumption, our estimation problem can be formulated as

$$\min_{N,E} \|E\|_0 \ \text{ s.t. } \ O = N^T L + E. \tag{5}$$

Here, $\|\cdot\|_0$ is an ℓ_0-norm penalty, which counts the number of non-zero entries in the matrix. To reiterate, Eq. (5) builds on the assumption that images are captured under known lighting conditions and any non-Lambertian corruptions have sparse structure. If these assumptions are not true (*e.g.*, because of imperfect lighting calibration, non-sparse specularities, etc.), then the hard constraint in Eq. (5) is no longer appropriate. To compensate for more diffuse modeling errors, we relax the hard constraint via an additional model mismatch penalty giving

$$\min_{N,E} \|O - N^T L - E\|_2^2 + \lambda\|E\|_0, \tag{6}$$

where λ is a nonnegative trade off parameter balancing data fit with sparsity. Note that in the limit as $\lambda \to 0$, the two problems are equivalent (the limit must be taken outside of the minimization). Since Eq. (6) decouples, we can consider instead an equivalent series of separate, pixel-wise optimizations problems of the canonical form

$$\min_{\boldsymbol{x},\boldsymbol{e}} \|\boldsymbol{y} - A\boldsymbol{x} - \boldsymbol{e}\|_2^2 + \lambda\|\boldsymbol{e}\|_0\,. \tag{7}$$

where the column vector $\boldsymbol{y}$ denotes an arbitrary transposed row of O, $A \triangleq L^T$, $\boldsymbol{x}$ is the associated unknown normal vector, and $\boldsymbol{e}$ is the sparse error component (we have omit-

ted pixel-wise subscripts for simplicity). Eq. (7) entails a difficult, combinatorial optimization problem that must be efficiently solved at every pixel. Here we consider two alternatives to brute force exhaustive search. First, in the machine learning and statistics literature, it is common to replace the discontinuous, non-convex ℓ_0 norm with the convex surrogate ℓ_1 norm.[1] In certain situations the resulting estimate will closely match the solution to Eq. (5) and/or Eq. (6); however, in the context of photometric stereo this substitution may not always be adequate (see Section 2.4 for more details). Secondly, we can apply a simple hierarchical Bayesian approximation to estimate $\boldsymbol{x}$ while simultaneously accounting for $\boldsymbol{e}$. This formulation, called SBL, is described in detail next.

2.3. Recovery of normals and corruptions via SBL

SBL [12] assumes the standard Gaussian likelihood function for the first-level, diffuse errors giving

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{e}) = \mathcal{N}(\boldsymbol{y}; A\boldsymbol{x} + \boldsymbol{e}, \lambda I). \tag{8}$$

We next apply independent, zero-mean Gaussian prior distributions on both $\boldsymbol{x}$ and $\boldsymbol{e}$:

$$p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{0}, \sigma_x^2 I), \qquad p(\boldsymbol{e}) = \mathcal{N}(\boldsymbol{e}; \boldsymbol{0}, \Gamma). \tag{9}$$

σ_x^2 describes the prior variance of the unknown normal vector; it is fixed to a large value to convey our lack of *a priori* certainty about $\boldsymbol{x}$. Thus the prior on $\boldsymbol{x}$ will be relatively uninformative (the value of σ_x^2 will be discussed further below). In contrast, $\Gamma \triangleq \mathrm{diag}[\boldsymbol{\gamma}]$ is a fully-parameterized, diagonal matrix, where $\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_k]^T$ is a non-negative vector of variances in one-to-one correspondence with elements of $\boldsymbol{e}$. A large variance γ_i indicates that the corresponding e_i is free to reflect the data, compensating for non-Lambertian effects, while a small or zero-valued variance implies that the associated error term is constrained near zero. Combining the likelihood and prior using Bayes' rule leads to the posterior distribution $p(\boldsymbol{x}, \boldsymbol{e}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{e})p(\boldsymbol{x})p(\boldsymbol{e})$. To estimate the normal vectors $\boldsymbol{x}$, we may further marginalize over $\boldsymbol{e}$ to give

$$p(\boldsymbol{x}|\boldsymbol{y}) = \int p(\boldsymbol{x},\boldsymbol{e}|\boldsymbol{y})d\boldsymbol{e} = \mathcal{N}(\boldsymbol{x}; \mu, \Sigma), \tag{10}$$

with mean and covariance defined as

$$\mu = \Sigma A^T (\Gamma + \lambda I)^{-1} \boldsymbol{y}, \tag{11}$$
$$\Sigma = \left[\sigma_x^{-2} I + A^T (\Gamma + \lambda I)^{-1} A\right]^{-1}.$$

We now have a closed-form estimator for $\boldsymbol{x}$ given by the posterior mean. The only issue then is the values for the unknown parameters Γ. Without prior knowledge as to the locations of the sparse errors, the empirical Bayesian approach to learning Γ is to marginalize the full joint distribution over all unobserved random variables, in this case $\boldsymbol{x}$

---

[1]The ℓ_1 norm of a vector $\boldsymbol{z}$ is given by $\sum_i |z_i|$, which constitutes the tightest convex approximation to the ℓ_0 norm.

and $\boldsymbol{e}$, and then maximize the resulting likelihood function with respect to Γ [12]. Equivalently, we will minimize

$$
\begin{aligned}
L(\Gamma) &\triangleq -\log \int p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{e})p(\boldsymbol{x})p(\boldsymbol{e})d\boldsymbol{x}d\boldsymbol{e} \\
&\equiv \log|\Sigma_y| + \boldsymbol{y}^T \Sigma_y^{-1} \boldsymbol{y} \\
&\text{with} \quad \Sigma_y \triangleq \sigma_x^2 AA^T + \Gamma + \lambda I,
\end{aligned} \tag{12}
$$

with respect to Γ. While $L(\Gamma)$ is non-convex, the cost function from Eq. (12) is composed of two terms which are concave with Γ and the element-wise inverse of Γ respectively. Therefore, optimization can be accomplished by adopting a majorization-minimization approach [13], which builds on a basic convex analysis that any concave function can be expressed as the minimum of a set of upper-bounding hyperplanes whose slopes are parameterized by auxiliary variables. Thus, we can re-express both terms of Eq. (12) as a minimization over hyperplanes, where $\boldsymbol{u}$ are associated with the first term, and $\boldsymbol{z}$ are associated with the second. When we temporarily drop the respective minimizations, we obtain a rigorous upper bound on the original cost function Eq. (12) parameterized by $\boldsymbol{u}$ and $\boldsymbol{z}$. For fixed values of $\boldsymbol{z}$ and $\boldsymbol{u}$, there exists a closed form solution for Γ that incorporates the tighter bound. Likewise, for a fixed value of Γ, the auxiliary variables can be updated in closed form to tighten the upper bound around the current Γ estimate. While space precludes a detailed treatment, the resulting update rules for the $(k+1)$-th iteration are given by

$$\gamma_i^{(k+1)} \rightarrow \left(z_i^{(k)}\right)^2 + u_i^{(k)}, \forall i, \quad \Gamma^{(k+1)} = \mathrm{diag}[\boldsymbol{\gamma}^{(k+1)}]$$
$$\boldsymbol{z}^{(k+1)} \rightarrow \Gamma^{(k+1)} \left(S^{(k+1)}\right)^{-1} \boldsymbol{y} \tag{13}$$
$$\boldsymbol{u}^{(k+1)} \rightarrow \mathrm{diag}\left[\Gamma^{(k+1)} - \left(\Gamma^{(k+1)}\right)^2 \left(S^{(k+1)}\right)^{-1}\right],$$

where $S^{(k+1)}$ is computed via

$$
\begin{aligned}
S^{(k+1)} &= D - DA\left[\sigma_x^{-2}I + A^T DA\right]^{-1} A^T D \\
&\text{and} \quad D \triangleq (\Gamma^{(k+1)} + \lambda I)^{-1}. \tag{14}
\end{aligned}
$$

These expressions only require $O(n)$ computations and are guaranteed to reduce $L(\Gamma)$ until a fixed point Γ_* is reached. This value can then be plugged into Eq. (11) to estimate the normal vector. We denote this point estimator as $\boldsymbol{x}_{sbl}$. If the variances Γ_* reflect the true profile of the sparse errors, then $\boldsymbol{x}_{sbl}$ will closely approximate the true surface normal. This claim will be quantified more explicitly in the next section.

We have thus far omitted details regarding the choice of λ and σ_x^2. The former can be reasonably set according to our prior expectations regarding the magnitudes of diffuse modeling errors, but in practice there is considerable flexibility here since some diffuse errors will be absorbed into $\boldsymbol{e}$. In contrast, we can realistically set $\sigma_x^2 \rightarrow \infty$, which implies zero precision or no prior information about the normal vectors and yet still leads to stable, convergent update

rules. However we have observed that on certain problems a smaller σ_x^2 can lead to a modest improvement in performance, presumably because it has a regularizing effect that improves the convergence path of the update rules from Eq. (13) (perhaps counterintuitively, in certain situations it does not alter the globally optimal solution as discussed below). It is also possible to learn σ_x^2 using similar updates to those used for Γ, but this introduces additional complexity and does not improve performance.

2.4. Analytical evaluation

Previously we discussed two tractable methods for solving Eq. (7): a convex ℓ_1-norm-based relaxation and a hierarchical Bayesian model called SBL. This section discusses comparative theoretical properties of these approaches relevant to the photometric stereo problem. To facilitate the analysis, here we consider the idealized case where there are no diffuse modeling errors, or that λ is small. In this situation, the basic problem from Eq. (7) becomes

$$\min_{\boldsymbol{x},\boldsymbol{e}} \|\boldsymbol{e}\|_0 \text{ s.t. } \boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{e}, \tag{15}$$

which represents the pixel-wise analog of Eq. (5). If the lighting directions and sparse errors are in general position (meaning they are not arranged in an adversarial configuration with zero Lebesgue measure), then it can be shown that the minimizer of Eq. (15) denoted $\boldsymbol{x}_0$ is guaranteed to be the correct normal vector as long as the associated feasible error component $\boldsymbol{e} = \boldsymbol{y} - A\boldsymbol{x}_0$ satisfies $\|\boldsymbol{e}\|_0 < n - 3$. Therefore, a relevant benchmark for comparing photometric stereo algorithms involves quantifying conditions whereby a candidate algorithm can correctly compute $\boldsymbol{x}_0$.

In this context, recent theoretical results have demonstrated that any minimizer $\boldsymbol{x}_1$ of the ℓ_1 relaxation approach will equivalently be a minimizer of Eq. (15) provided $\|\boldsymbol{e}\|_0$ is sufficiently small relative to a measure of the structure in columns of the lighting matrix A [2]. However, for typical photometric stereo problems the requisite equivalency conditions often do not hold (*i.e.*, $\|\boldsymbol{e}\|_0$ is required to be prohibitively small) both because of structure imposed by the lighting geometry and implicit structure that emerges from the relatively small dimensionality of the problem (meaning we do not benefit from asymptotic large deviation bounds that apply as n becomes large). Fortunately, SBL offers the potential for improvement over ℓ_1 via the following result.

Theorem: For all $\sigma_x^2 > 0$ (and assuming $\lambda \to 0$), if Γ_* is a global minimum of Eq. (12), then the associated estimator $\boldsymbol{x}_{sbl}$ will be a global minimum of Eq. (15). Moreover, for σ_x^2 sufficiently large it follows that: (i) Any analogous locally minimizing SBL solution is achieved at an estimate $\boldsymbol{x}_{sbl}$ satisfying $\|\boldsymbol{y} - A\boldsymbol{x}_{sbl}\|_0 \leq n - 3$, (ii) SBL can be implemented with a tractable decent method such that convergence to a minimum (possibly local) that produces an $\boldsymbol{x}_{sbl}$ estimator as good or better than the global ℓ_1 solution

is guaranteed, meaning $\|\boldsymbol{y} - A\boldsymbol{x}_{sbl}\|_0 \leq \|\boldsymbol{y} - A\boldsymbol{x}_1\|_0$.

The basic idea of the proof is to transform the cost function using block-matrix inverse and determinant identities, as well as ideas from [2], to extend SBL properties derived in [14] to problems in the form of Eq. (15). First, we consider the limit as σ_x^2 becomes large. It is then possible using linear algebra manipulations to generate an equivalent cost function where Σ_y is now equal to $B\Gamma B^T + \lambda I$ for some B with $B^T A = 0$. We then apply known results for SBL with this revised Σ_y and arrive at the stated theorem.

We may thus conclude that SBL can enjoy the same theoretical guarantees as the ℓ_1 solution yet boosted by a huge potential advantage assuming that we are able to find the global minimum of Eq. (12) (which will always produce an $\boldsymbol{x}_{sbl} = \boldsymbol{x}_0$, unlike ℓ_1). There are at least two reasons why we might expect this to be possible based on insights drawn from [14]. First, the sparse errors $\boldsymbol{e}$ will likely have substantially different magnitudes depending on image and object properties (meaning the non-zero elements of $\boldsymbol{e}$ will not all have the same magnitude), and it has been shown that in this condition SBL is more likely to converge to the global minimum [14]. Secondly, as discussed previously, A will necessarily have some structure unlike, for example, high dimensional random matrices. In this environment, SBL vastly outperforms ℓ_1 because it is implicitly based on an A-dependent sparsity penalty that can compensate, at least in part, for structure in A. To clarify this point, we note that the SBL procedure can equivalently be recast as the minimization of a function in the same form as the canonical sparse optimization problem given by Eq. (15), but with a different sparse penalty function that is directly dependent on A. Thus the matrix A appears twice with SBL, once explicitly in the constraint specification and once embedded in the penalty function. It is this embedding that helps compensate for structure in the A-dependent constraint surface, a claim that has been verified by numerous experiments beyond the scope of this paper.

3. Experiments

Experiments on both synthetic and real data were carried out. All experiments were performed on an Intel Core2 Duo E6400 (2.13GHz, single thread) machine with 4GB RAM and were implemented in MATLAB. For the SBL- and ℓ_1-based methods we used $\lambda = 1.0^{-6}$ in the synthetic experiments with no diffuse modeling errors (Section 3.1), and $\lambda = 1.0^{-2}$ for the other cases (Sections 3.2 and 3.3). We set $\sigma_x^2 = 1.0^6$ for all experiments.

3.1. Quantitative evaluation with synthetic images

We generate 32-bit HDR gray-scale images of two objects called *Bunny* (256×256) and *Caesar* (300×400) with foreground masks under different lighting conditions
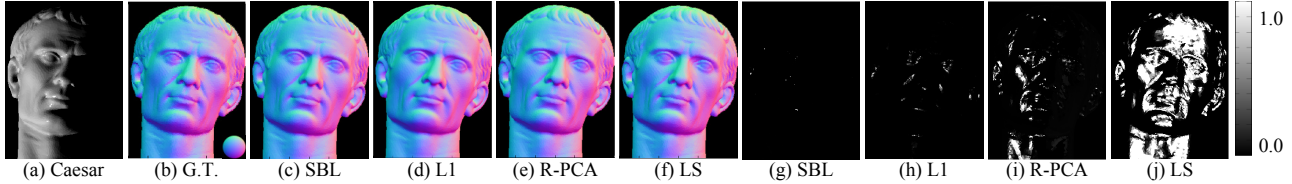
Figure 2. Recovery of surface normals from 40 images of *Caesar* dataset (300×400) with explicit shadow removal (Exp.3-1(a)). (a) Input, (b) Ground truth, (c)-(j) Recovered surface normals and Error maps (in degrees).

Figure 3. Recovery of surface normals from 40 images of *Bunny* (256×256) dataset without explicit shadow removal (Exp.3-1(b)). (a) Input, (b) Ground truth, (c)-(j) Recovered surface normals and Error maps (in degrees).

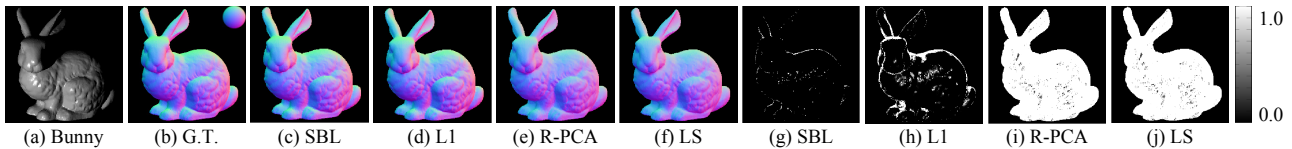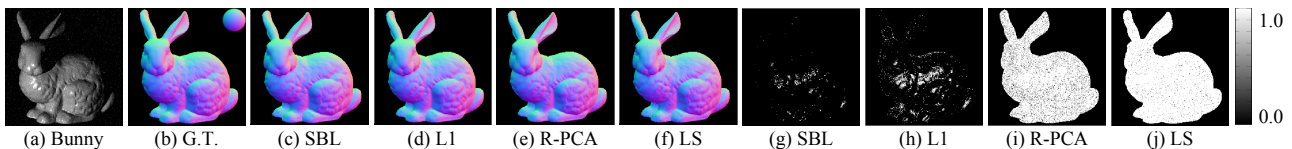Figure 4. Recovery of surface normals from 40 images of *Bunny* dataset with explicit shadow removal and additive Gaussian noises (30%) (Exp.3-1(b)). (a) Input, (b) Ground truth, (c)-(j) Recovered surface normals and Error maps (in degrees).

whose directions are randomly selected from a hemisphere with the object placed at the center. Specular reflections are attached using the Cook-Torrance reflectance model [3] and cast/attached shadows are synthesized under each lighting condition. We change experimental conditions with regard to the number of images, surface roughness (*i.e.*, the ratio of specularities), shadow removal (*i.e.*, whether or not a shadow mask is used to remove zero-valued elements from the observation matrix O), and the presence of additional Gaussian noise. Note that when in use (as defined for each experiment), the shadow mask is applied equivalently to all algorithms. To increase statistical reliability, all experimental results are averaged over 20 different sets of 40 input images. The average ratio of specularities in *Bunny* and *Caesar* are 8.4% and 11.6% and that of cast/attached shadows are 24.0% and 27.8% respectively. For the quantitative evaluation, we compute the angular error between the recovered normal map and the ground truth. In this experiment, our methods via sparse regression are implemented by both SBL and a convex ℓ_1-norm based relaxation (L1). We compare our methods with the R-PCA-based method proposed by Wu *et al.* [16] (using a fixed trade off parameter) and the standard least squares (LS)-based Lambertian photometric stereo [15] estimate obtained by solving

$$\min_{N} \|O - N^T L\|_2^2. \tag{16}$$

(a) Valid number of images for effective recovery

In this experiment, we vary the number of images to estimate the minimum number required for effective recovery when using the shadow mask with fixed surface roughness. Once 40 images are generated for each dataset, the image subset is randomly sampled from a dataset. We illustrate the result in Table 1 and Fig. 2. We observe that the sparse-regression-based methods are significantly more accurate than both R-PCA and LS. We also observe that SBL is more accurate than ℓ_1, although somewhat more expensive computationally.[2] Note that, although not feasible in general, when the number of images is only 5, the most accurate and efficient implementation for regression could be to just systematically test every subset of 3 images (*i.e.*, brute force search only requires 10 iterations at each pixel).

We also compared with a RANSAC-based approach proposed by Mukaigawa *et al.* [9] to empirically verify that our well-constrained formulation contributes to improved performance. The results are presented in Table 2. We have added standard deviations in the table for studying the estimation stability. It shows that although we use a large number of samples for RANSAC (*e.g.*, 2000), the

---

[2]The SBL convergence rate can be slower with fewer images because of the increased problem difficulty. This explains why the computation time may actually be shorter with more images.

Table 1. Experimental results of *Bunny* (left) / *Caesar* (right) dataset with varying number of images (Exp.3-1(a)).

No. of images	Mean error (in degrees)				Median error (in degrees)				Elapsed time (sec)				Mean error (in degrees)				Median error (in degrees)				Elapsed time (sec)			
	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS
5	**6.0**	6.0	15.3	7.0	4.7	**4.3**	10.7	4.8	46.5	13.6	15.7	**4.6**	6.2	6.0	26.2	7.4	4.73	**4.63**	31.0	4.97	106.4	34.2	45.8	**15.9**
10	**0.09**	0.61	3.8	1.9	**0.27**	0.58	0.81	1.8	36.3	13.6	37.8	**5.9**	**0.24**	0.40	10.7	0.94	**0.19**	0.26	1.8	0.93	97.2	34.1	93.7	**19.2**
15	**0.076**	0.16	0.21	1.6	**0.052**	0.13	0.19	1.6	26.8	13.1	55.1	**6.3**	**0.044**	0.11	2.6	0.77	**0.047**	0.083	0.14	0.76	67.0	31.4	153.3	**22.0**
20	**0.033**	0.080	0.11	1.6	**0.022**	0.078	0.11	1.6	24.2	13.5	70.5	**6.9**	**0.018**	0.051	0.079	0.76	**0.015**	0.035	0.065	0.72	60.9	32.7	177.5	**23.3**
25	**0.018**	0.055	0.084	1.6	**0.010**	0.048	0.069	1.6	23.1	14.1	86.0	**7.6**	**0.011**	0.034	0.068	0.76	**0.0081**	0.023	0.059	0.76	57.9	34.3	196.9	**25.1**
30	**0.012**	0.037	0.080	1.7	**0.0048**	0.032	0.065	1.7	23.1	14.2	121.0	**8.4**	**0.0063**	0.018	0.043	0.77	**0.0082**	0.018	0.043	0.77	58.1	33.2	231.7	**27.7**
35	**0.0057**	0.023	0.098	1.6	**0.0029**	0.019	0.093	1.6	22.7	14.6	161.3	**8.5**	**0.0045**	0.012	0.036	0.78	**0.0031**	0.0084	0.033	0.80	58.4	34.5	259.2	**29.4**
40	**0.0039**	0.019	0.12	1.6	**0.0020**	0.015	0.12	1.6	22.6	15.0	200.7	**9.4**	**0.0031**	0.0094	0.037	0.76	**0.0019**	0.0063	0.034	0.78	59.6	35.2	281.2	**31.5**

Table 2. Comparison with RANSAC based approach [9] (Exp.3-1(a)). Table 3. Results of *Bunny* without shadow removal (Exp.3-1(b)).

No. of Images	Mean error		Median error		Standard deviation		Elapsed time	
	SBL	RANSAC	SBL	RANSAC	SBL	RANSAC	SBL	RANSAC
5	**6.0**	6.7	**4.7**	5.4	**4.1**	4.4	**46.5**	52.3
10	**0.09**	0.74	**0.27**	0.38	**0.35**	1.4	**36.3**	544.0
15	**0.076**	0.61	**0.052**	0.12	**0.059**	1.9	**26.8**	958.4
20	**0.033**	0.70	**0.022**	0.058	**0.027**	1.2	**24.2**	1048.9
25	**0.018**	1.0	**0.010**	0.063	**0.020**	2.6	**23.1**	1141.8
30	**0.012**	2.3	**0.0048**	0.042	**0.012**	4.0	**23.1**	1227.8
35	**0.0057**	3.2	**0.0029**	0.051	**0.0064**	9.2	**22.7**	1327.9
40	**0.0039**	2.1	**0.0020**	0.046	**0.0048**	4.3	**22.6**	1430.4

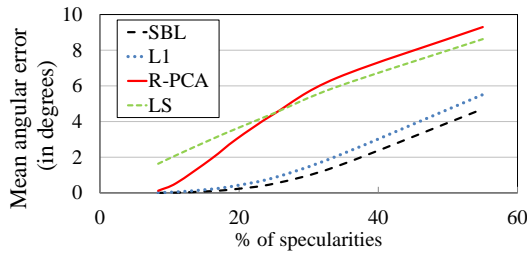| No. of images	Mean error (in degrees)				Median error (in degrees)				Elapsed time (sec)			
	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS
5	**5.2**	11.9	12.1	12.1	**5.0**	12.3	12.5	12.5	213.0	37.0	45.8	**5.1**
10	**2.8**	5.6	10.9	10.9	**2.3**	5.6	11.3	11.3	98.9	33.0	93.7	**6.0**
15	**1.9**	4.0	9.9	10.0	**2.3**	4.0	10.1	10.2	66.8	32.5	153.3	**7.4**
20	**1.2**	2.7	9.4	9.5	**1.0**	2.7	9.6	9.6	52.9	30.0	177.5	**7.6**
25	**0.81**	1.9	8.9	9.0	**0.69**	1.8	8.9	9.0	46.2	31.0	196.9	**9.1**
30	**0.62**	1.6	9.0	9.1	**0.61**	1.5	8.9	8.9	41.1	32.0	231.7	**9.4**
35	**0.59**	1.5	9.1	9.1	**0.58**	1.4	9.3	9.3	41.1	34.4	259.2	**11.0**
40	**0.53**	1.2	8.8	8.9	**0.58**	1.2	9.0	9.1	39.4	33.3	281.2	**10.7**

Figure 5. Experimental results of *Bunny* with varying amount of specularities. The x-axis and y-axis indicate the ratio of specularities and the mean angular error of normal map (Exp.3-1(b)).

RANSAC-based approach cannot always stably find the solution especially when the number of images is large. On the other hand, our method succeeds in finding the solution stably and efficiently.

(b) Robustness to various corruptions

We now set three different conditions for evaluating the robustness of our method to effects of (i) specularities (varying specularities, explicit shadow removal, no image noises), (ii) shadows (fixed specularities, no shadow removal, no image noises), (iii) image noises (fixed specularities, explicit shadow removal, varying amount of image noises). The ability to estimate surface normals without an explicit shadow mask can be important since in practical situations shadow locations may not always be easy to determine a priori. The number of images is 40 in (i), (iii) and varying from 5 to 40 in (ii). We use *Bunny* for evaluation and the ratio of specularities and image noises are varying from 10% to 60% and 10% to 50%, respectively in

Table 4. Experimental results of *Bunny* with varying amount of additive Gaussian noises (Exp.3-1(b)).

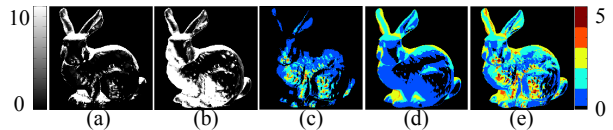| Dens. of noises (%)	Mean error (in degrees)				Median error (in degrees)			
	SBL	L1	R-PCA	LS	SBL	L1	R-PCA	LS
10	**0.0079**	0.040	0.16	3.3	**0.0060**	0.039	0.16	3.3
20	**0.021**	0.11	0.79	4.4	**0.019**	0.099	0.80	4.3
30	**0.068**	0.29	3.6	5.3	**0.060**	0.25	3.2	5.2
40	**0.21**	0.70	9.8	6.2	**0.18**	0.63	9.9	6.1
50	**0.58**	1.5	11.7	7.0	**0.53**	1.4	11.7	6.9

Figure 6. Comparison between SBL and ℓ_1-based method. Errors of (a) SBL and (b) L1 (in degrees). The per-pixel number of (c) specularities, (d) shadows, (e) corruptions (The maximum is 5).

(i) and (iii). Image noise obeys a Gaussian distribution (σ^2 is 0.1). The results are illustrated in Fig. 3, Fig. 4, Fig. 5, and Table 3 and Table 4. In summary, our methods outperform both R-PCA and LS in accuracy and outperform R-PCA in efficiency in the presence of any kind of corruptions. We also observe that SBL is more accurate than ℓ_1 in all conditions but more computationally expensive. As expected, performance of each method degrades as the ratio of specular corruptions increases making the optimal solution more difficult to estimate. Likewise, when the shadow mask is removed additional corrupted pixels (outliers) contaminate the estimation process and each algorithm must try to blindly compensate. SBL performs best since it is able
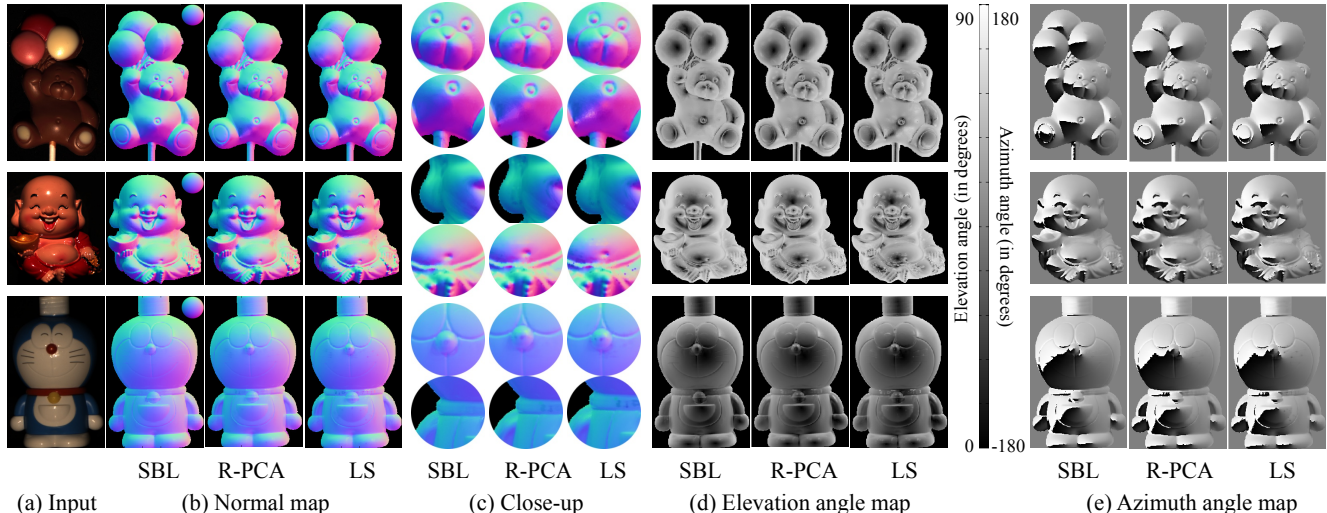
Figure 7. Experimental results with real datasets. We used three kind of datasets called *Chocolate bear* (25 images with 261×421), *Fat guy* (40 images with 293×344) and *Doraemon* (40 images with 269×420). (a) Example of input images (b), (C) Recovered surface normals and close-up views (d) Elevation angles of recovered surface normals (e) Azimuth angles of recovered surface normals

to handle a wider spectrum of sparse errors (see Fig. 3, Table 3, and also Fig. 6 discussed below). Finally, we further compare estimation properties between SBL and L1 using a case from Table 3 where the number of images is 5 for simplicity, and we do not remove shadows. Error maps and per-pixel numbers of corruptions are displayed in Fig. 6. We observe that the ℓ_1 method typically fails when any shadows appear while SBL can find the correct solution in most pixels as long as the number of corruptions is less than 3. Note that this is at the theoretical limit given only 5 images.

3.2. Qualitative evaluation with real images

We also evaluate our algorithm (only the SBL implementation) using real images. We captured RAW images without gamma correction by Canon 30D camera with a 200[mm] telephotolens and set it 1.5[m] far from target object. Lighting conditions are randomly selected from a hemisphere whose radius is 1.5[m] with the object placed at the center. For calibrating light sources, a glossy sphere was placed in the scene. We use a set of 25 images of *Chocolate bear* (261×421), and 40 images each of *Doraemon* (269×420) and *Fat guy* (293×344). We evaluate the performance by visual inspection of the normal maps, elevation angle maps (orientations between normals and a view direction) and azimuth angle maps (normal orientation on the x-y plane) that are illustrated in Fig. 7. We observe that our method can estimate smoother and more reasonable normal maps in the presence of a large amount of specularities.

3.3. Evaluations with model mismatch errors

To explicitly test how deviations from the ideal Lambertian assumption affect the proposed method, we conducted

two additional experiments with (a) various kinds of materials with non-Lambertian diffusions and (b) inaccurate lighting directions. In both cases, the effective corruptions cannot be completely modeled as sparse errors.

(a) Various kind of non-Lambertian materials

In this experiment, we test our sparse-regression-based method on 40 sphere images rendered with one hundred BRDF functions from the MERL database [7] (the image size is 256×256). From the estimation results illustrated in Fig. 8, we observe that for materials whose specular reflections and diffusive reflections are clearly distinguishable (*e.g.*, (10) specular-violet-phenolic), our method outperforms the other two methods even if the diffusive component does not completely obey the Lambertian rule. On the other hand, in a case where a diffusive component is dominant, our sparse-regression-based method seems to have limited advantages over other methods (*e.g.*, (40) black-fabric). We also observe that all methods have difficulty in handling metallic objects which do not obey both the Lambertian rule and the sparsity assumption of corruptions entirely (*e.g.*, (95) black-obsidian), however our method is consistently the most reliable overall.

(b) Inaccurate lighting directions

For this experiment, we synthesized 40 *Bunny* images and then used incorrect lighting directions (five degrees of angular error in random directions were added) to recover surface normals. In addition, we also attempt to refine lighting directions by iteratively recovering both surface normals and lighting directions based on the symmetrical structure of Eq. (4). First, we estimate surface normals using the given, errant lighting directions. Then, fixing recovered surface normals, we update the lighting directions using a
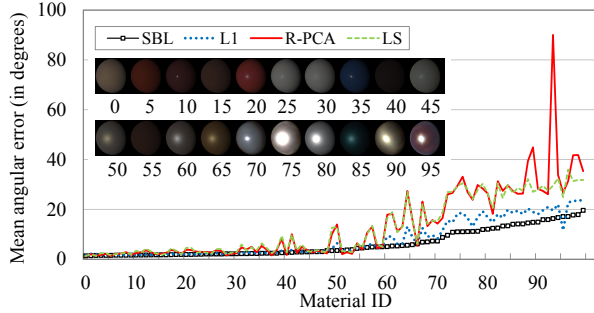
Figure 8. Experimental results of synthesized sphere with MERL BRDF dataset [7] (Exp.3-3(a)). We aligned results in ascending order of mean angular error of SBL. The material names are included in the supplementary.

least squares fit. We continue this process iteratively until convergence. The experimental results are illustrated in Table 5. We observe that our method outperforms the other two methods even without refining the lighting directions; however, optimizing the lighting direction via a few iterations always improves the normal estimates. (Although the exact optimal number of iterations may be difficult to determine, a single iteration always has a substantial benefit).

4. Conclusion

Herein we have demonstrated the superior performance of our sparse regression approach to photometric stereo through extensive analyses and experiments. In particular, our method gives more accurate estimates of surface normals than previous least squares and R-PCA approaches while remaining computationally competitive. Regarding competing sparse regression techniques, SBL is both theoretically and empirically superior to ℓ_1-based estimates but requires a modest increase in computational cost. A current limitation of our method is that we assume the diffusive component is Lambertian. Therefore non-Lambertian diffusive objects can potentially be problematic, although this affect is partially mitigated by the diffuse and sparse error terms built into our model. However, in the future we will refine our model to explicitly handle non-Lambertian diffusive components coupled with sparse corruptions to recover surface normals robustly across a wider range of materials.

References

[1] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proc. of CVPR*, 2008. 1

[2] E. Candés and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005. 1, 4

[3] R. Cook and K. Torrance. A reflectance model for computer graphics. *ACM Trans. on Graph.*, 15(4):307–316, 1981. 5

Table 5. Experimental results of *Bunny* under incorrect lighting directions to the 6-th iteration (Exp.3-3(b)).

No. of iteration	Normal error (in degrees)								Lighting error (in degrees)			
	SBL		L1		R-PCA		LS		SBL		L1	
	Ave.	Med.	Ave.	Med.	Ave.	Med.	Ave.	Med.	Ave.	Med.	Ave.	Med.
1st	8.4	8.3	13.5	17.0	20.7	20.7	22.3	23.5	5.8	5.9	9.7	11.7
2nd	4.6	4.4	8.3	8.6					4.3	4.5	8.7	9.5
3rd	4.2	3.9	8.1	8.6					4.3	4.4	9.0	10.2
4th	4.2	3.9	8.1	8.8					4.3	4.4	9.2	10.1
5th	4.2	3.9	8.2	9.0					4.3	4.4	9.3	10.8
6th	4.2	4.0	8.3	9.1					4.4	4.4	9.4	10.9

[4] D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying BRDFs from photometric stereo. In *Proc. of ICCV*, 2005. 1

[5] C. Hernández, G. Vogiatzis, and R. Cipolla. Multi-view photometric stereo. *IEEE Trans. on Pattern Anal. Mach. Intell.*, 30(3):548–554, 2008. 1

[6] A. Hertzmann and S. Seitz. Shape and materials by example: A photometric stereo approach. In *Proc. of ICCV*, 2003. 1

[7] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. on Graph.*, 22(3):759–769, 2003. 7, 8

[8] D. Miyazaki, K. Hara, and K. Ikeuchi. Median photometric stereo as applied to the Segonko tumulus and museum objects. *International Journal of Computer Vision*, 86(2):229–242, 2010. 1, 2

[9] Y. Mukaigawa, Y. Ishii, and T. Shakunaga. Analysis of photometric factors based on photometric linearization. *Journal of the Optical Soceity of America*, 24(10):3326–3334, 2007. 1, 2, 5, 6

[10] M. Oren and S. K. Nayar. Generalization of the Lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, 1995. 1

[11] K.-L. Tang, C.-K. Tang, and T.-T. Wong. Dense photometric stereo using tensorial belief propagation. In *Proc. of CVPR*, 2005. 1

[12] M. Tipping. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learns. Res.*, 1:211–244, 2001. 1, 3

[13] D. Wipf and S. Nagarajan. Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):317–329, 2010. 3

[14] D. Wipf, B. D. Rao, and S. Nagarajan. Latent variable Bayesian models for promoting sparsity. *IEEE Transactions on Information Theory*, 57(9), 2011. 1, 4

[15] P. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Engg*, 19(1):139–144, 1980. 1, 2, 5

[16] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. of ACCV*, 2010. 1, 2, 5

[17] C. Yu, Y. Seo, and S. Lee. Photometric stereo from maximum feasible Lambertian reflections. In *Proc. of ECCV*, 2010. 1, 2