Convex Optimization Based Low-Rank Matrix Completion and Recovery for Photometric Stereo and Factor Classification

Lun Wu, *Student Member, IEEE,* Arvind Ganesh, *Student Member, IEEE,* Boxin Shi, Yasuyuki Matsushita, *Senior Member, IEEE,* Yongtian Wang, *Member, IEEE,* and Yi Ma *Senior Member, IEEE*

Abstract—We present a new approach to robustly solve the photometric stereo problem. We cast the problem of recovering surface normals from images taken under multiple lighting conditions as a one of recovering a low-rank matrix from missing and corrupted observations, which model many different non-Lambertian effects such as shadows and specularities. Unlike previous approaches that use least-squares or heuristic robust techniques, our method uses an advanced convex optimization technique that is guaranteed to find the correct low-rank matrix by simultaneously fixing its missing and erroneous entries. Extensive experimental results demonstrate that our method achieves unprecedentedly accurate estimates of surface normals in the presence of significant amount of shadows and specularities. The new technique can be used to improve virtually any existing photometric stereo method. We not only showcase its effectiveness in the case of photometric stereo, but also extend our method to perform classification of photometric factors. In contrast to existing robust techniques, our method is computationally more efficient and provides theoretical guarantees for robustness to large errors.

Index Terms—Photometric stereo, Robust PCA, Low-rank matrix, Sparse errors, Photometric factor classification

1 Introduction

PHOTOMETRIC stereo [1], [2] is the classical computer vision problem of estimating surface orientations from a set of photographs taken from a fixed viewpoint under different lighting directions. It is well-known that when a Lambertian surface is illuminated by at least three known lighting directions, the surface orientation at each visible point can be uniquely determined from its shading variations. Since photometric stereo can produce a dense normal field at the level of detail that cannot be achieved by other triangulation-based approaches, it has generated a lot of interest for accurate shape reconstruction.

- *Lun Wu is with the School of Optics and Electronics, Beijing Institute of Technology, Beijing, China. E-mail: lun.wu@hotmail.com*
- *Arvind Ganesh is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign. E-mail: abalasu2@illinois.edu*
- *Boxin Shi is with the Institute of Industrial Science, The University of Tokyo, Japan. E-mail: shi@cvl.iis.u-tokyo.ac.jp*
- *Yasuyuki Matsushita is with Microsoft Research Asia, Beijing, China. Email: yasumat@microsoft.com*
- *Yongtian Wang is with the School of Optics and Electronics, Beijing Institute of Technology, Beijing, China. E-mail: wyt@bit.edu.cn*
- *Yi Ma is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, and Microsoft Research Asia, Beijing, China. E-mail: yima@illinois.edu*

From different perspectives, it has long been shown that if there are no shadows, the appearance of a convex Lambertian scene illuminated from different lighting directions span a three-dimensional subspace [3] or an illumination cone [4]. Basri and Jacobs [5] and Georghiades *et al*. [6] have further shown that the images of a convex-shaped object with cast shadows can also be well-approximated by a low-dimensional linear subspace. These works indicate that there exists a degenerate structure in the appearance of Lambertian surfaces under variation in illumination, and this is the key property that almost all photometric stereo methods harness to determine the surface normals. However, this structure is rarely observed in real images due to complex factors such as specular reflections and shadows which render most existing algorithms unsuitable for practical purposes.

One of the most popular approaches in photometric stereo for Lambertian surfaces is to use a *Least Squares* solution to a set of linear equations that relate the observations and known lighting directions, or equivalently, to identify the underlying low-dimensional subspace using conventional Principal Component Analysis (PCA) [7]. Such a solution is known to be optimal if the measurements are corrupted by *i.i.d*. Gaussian noise of small magnitude. Unfortunately, in reality, photometric measurements rarely obey such a simplistic noise model. The intensity values at some pixels can be severely affected by specular reflections (deviation from the basic Lambertian assumption), sensor saturations, or shadowing effects. As a result, the Least Squares solution normally ends up with incorrect estimates of surface

orientations in practice. Most existing algorithms deal with this problem by either treating these corruptions as small noise, or by removing them using a global threshold on the pixel intensity value.

For years, researchers have been exploring robust methods to handle the above mentioned problems in images. In the context of uncalibrated photometric stereo, Hayakawa [8] used a matrix factorization framework to deal with multiple images with shadows. That algorithm estimates the surface normal and reflectance in shadowed regions by starting with a partial solution obtained from a sub-matrix with no shadowed pixels and then gradually estimates the surface normals in the shadowed regions. The shadowed regions can also be refilled using linear regression [9], [10], or removed by techniques from robust statistics using a steepest descent algorithm [11]. A significant limitation of these methods is that they require an intensity matrix without large magnitude errors barring shadows, which amounts to assuming that there are no specularities in the observed images.

More recently, researchers have explored various robust approaches to eliminate such deviations by treating the corrupted measurements as outliers, *e.g.*, using a Random Sample Consensus (RANSAC) scheme [12], [13], or a median-based approach [14]. To identify different types of corruptions in images more carefully, Mukaigawa *et al*. [15] proposed a method based on RANSAC, which can estimate the surface normals with high accuracy. Although these methods are shown to work well, their computational cost is rather high for practical applications.

To achieve the goal of both accuracy and efficiency, we develop a new robust photometric stereo method based on recent advances in the area of low-rank matrix recovery [16], [17]. We recast the photometric stereo problem as one of recovering and completing a low-rank matrix in the presence of sparse, gross errors that account for corrupted and missing pixels. Unlike previous heuristic methods, under fairly broad conditions, the new method is guaranteed to correctly recover the low-rank Lambertian diffuse component from the highly corrupted and incomplete observations. Based on advanced convex optimization tools for nuclear norm and ℓ_1-norm minimization, the new method can efficiently obtain highly accurate estimates of surface orientations. Our method can be used to improve virtually any existing photometric stereo method, including uncalibrated photometric stereo [8], where traditionally, corruption in the data (*e.g.*, shadow and specularity) is either neglected or ineffectively dealt with conventional heuristic robust estimation methods. In addition, we show that the proposed technique can be efficiently used for photometric factor classification that can classify diffuse, specular, attached shadows and cast shadows. Experimental results on both synthetic and real data show that photometric factors can be identified with very high accuracy by our approach. Preliminary results of our approach were presented in [18].

In contrast to previous robust approaches, our method is computationally more efficient and provides theoretical guarantees for robustness to large errors. More importantly, our method uses all the available information simultaneously for obtaining the optimal result, instead of discarding informative measurements, *e.g.*, by either selecting the best set of illumination directions [13] or using the median estimator [14].

The remainder of this paper is organized as follows. In Section 2, we provide a brief description of the Lambertian image formation model and the resulting low-rank matrix structure in photometric stereo. In Section 3, we propose an efficient and scalable convex optimization algorithm to accurately recover the surface normal map. In Section 4, we explain how the recovered solution can be used for photometric factors classification. We showcase the efficacy of our method through experiments in Section 5. In Section 6, we discuss some potential extensions of this work.

2 Photometric Stereo as Low-Rank Matrix Recovery with Sparse Errors

In this section, we formulate the problem of estimating the normal map as a rank minimization problem. We first review the basic Lambertian image formation model, and then discuss how to model large deviations like shadows and specularities. In the following discussion, we make a few assumptions:

- The relative position of the camera and object is fixed across all images.
- The object is illuminated by a point light source at infinity.
- The sensor response is linear.

2.1 Lambertian Image Formation Model

The appearance I of a point in a Lambertian scene observed under a lighting direction $\mathbf{l} \in \mathbb{R}^3$ is described by the inner product:

$$I = \rho\, \mathbf{n} \cdot \mathbf{l}, \tag{1}$$

where ρ is the diffuse albedo, and $\mathbf{n} \in \mathbb{R}^3$ is the surface normal at the point. Suppose that we are given n images $I_1, \ldots, I_n$ of a scene under different lighting conditions. Let the region of interest be composed of m pixels in each image.[1] We order the pixel locations with a single index k, and let $I_j(k)$ denote the observed intensity at pixel location k in image I_j. With this notation, we have the following relation about the observation $I_j(k)$:

$$I_j(k) = \rho_k\, \mathbf{n}_k \cdot \mathbf{l}_j, \tag{2}$$

where ρ_k is the albedo of the scene at pixel location k, $\mathbf{n}_k \in \mathbb{R}^3$ is the (unit) surface normal of the scene at pixel location k, and $\mathbf{l}_j \in \mathbb{R}^3$ represents the normalized lighting direction vector corresponding to image I_j.[2]

1. Typically, m is much larger than the number of images n.

2. The convention here is that the lighting direction vectors point from the surface of the object to the light source.

We assume that the light intensity is constant across images to simplify the discussion, although the proposed method is not limited to such a condition.

2.2 Low-rank Matrix Structure

Consider the matrix $O \in \mathbb{R}^{m \times n}$ constructed by stacking all the vectorized images, denoted by $\mathrm{vec}(I_k)$, as

$$O = [\mathrm{vec}(I_1) \mid \cdots \mid \mathrm{vec}(I_n)], \tag{3}$$

where $\mathrm{vec}(I_j) = [I_j(1), \ldots, I_j(m)]^T$ for $j = 1, \ldots, n$. It follows from Eq. (2) that O can be factorized as follows:

$$O = NL, \tag{4}$$

where $N \doteq [\rho_1 \mathbf{n}_1 \mid \cdots \mid \rho_m \mathbf{n}_m]^T \in \mathbb{R}^{m \times 3}$, and $L \doteq [\mathbf{l}_1 \mid \cdots \mid \mathbf{l}_n] \in \mathbb{R}^{3 \times n}$. Suppose that the number of images $n \geq 3$. Clearly, irrespective of the number of pixels m and the number of images n, the rank of the matrix O is at most 3.

2.3 Modeling Corruptions as Sparse Errors

The low-rank structure of the observation matrix O described above is seldom observed with real images. This is due to the presence of shadows and specularities in real images.

- **Shadows** arise in real images in two possible ways. Some pixels are not visible in the image because they face away from the light source. Such dark pixels are referred to as *attached shadows* [19]. In deriving Eq. (4) from Eq. (2), we have implicitly assumed that all pixels of the object are illuminated by the light source in each image. However, if the pixel faces away from the light source, then the relation no longer holds. Mathematically, this implies that Eq. (2) must be rewritten as follows:

$$I_j(k) = \max\{\rho_k \, \mathbf{n}_k \cdot \mathbf{l}_j, 0\}. \tag{5}$$

  Shadows can also occur in images when the shape of the object's surface is not convex: parts of the surface can be occluded from the light source by other parts. Even though the normal vectors at such occluded pixels may form a sharp angle with the lighting direction, these pixels appear entirely dark. We refer to such dark pixels as *cast shadows*. Irrespective of the type, all shadows occur in images as dark pixels with very small, if not zero, intensity values.
- **Specularities.** Specular reflection arises when the object of interest is not perfectly diffusive, *i.e.*, when the surface luminance is not purely isotropic. Thus, the intensity of reflected light depends on the viewing angle, and light is reflected in a mirror-like fashion accompanied by a specular lobe when viewed from certain angles. This gives rise to some bright spots or shiny patches on the surface of the object that significantly deviate from the Lambertian assumption.

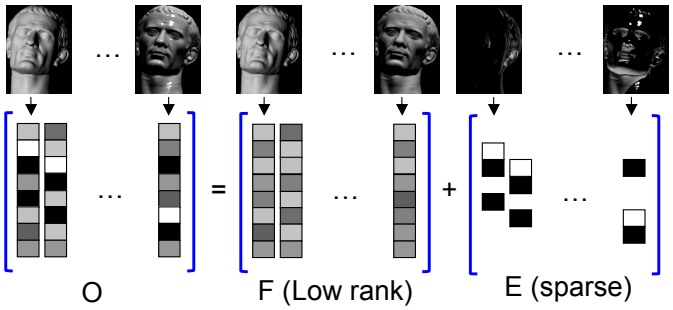


Fig. 1. **Illustration of our approach**. All vectorized images are stacked into matrix with higher rank as observation. The problem is to find a sparse matrix E such that matrix F has the lowest possible rank.

Suppose that we represent all these deviations from the ideal low-rank diffusive model Eq. (4) by an error matrix $E \in \mathbb{R}^{m \times n}$. Thus, instead of Eq. (4), the image measurements should be modeled as

$$O = NL + E, \tag{6}$$

where the matrix E accounts for corruption by shadows or specularities. Now suppose that only a small fraction of the pixels in each image exhibit strong specular reflectance and that a large majority of the pixels are illuminated by the light source. Then, most pixels in the input images obey the low-rank diffusive model given by Eq. (4), and hence, most entries in the error matrix E will be zero, *i.e.*, E is a sparse matrix. If the matrix L of lighting directions is known, then we can compute the surface normals, provided that we can decompose O as the sum of a low-rank matrix and a sparse error matrix, as illustrates in Fig. 1. Thus, the problem can be stated more formally as follows:

> *Let $I_1, \ldots, I_n$ be n images of an object under different illumination conditions. If $O \in \mathbb{R}^{m \times n}$ is defined as given in Eq. (3), then find a sparse matrix E such that the matrix $F \doteq O - E = NL$ has the lowest possible rank.*

Using a Lagrangian formulation, we can write the above problem as the following optimization problem:

$$\min_{F,E} \ \mathrm{rank}(F) + \gamma \|E\|_0 \quad \text{s.t.} \quad O = F + E, \tag{7}$$

where $\|\cdot\|_0$ denotes the ℓ_0-norm (number of non-zero entries in the matrix), and $\gamma > 0$ is a parameter that trades off the rank of the solution F versus the sparsity of the error E. Let $(\hat{F}, \hat{E})$ be the optimal solution to Eq. (7).

Now, if the lighting directions L are given,[3] we can easily recover the matrix N of surface normals from $\hat{F}$ as:

$$N = \hat{F} L^{\dagger}, \tag{8}$$

3. If the lighting directions are accurately given, our formulation reduces to a simpler robust estimation problem: $\min_N ||O - NL||_1$. However, the low-rank minimization formulation applies to cases when the directions are unknown or not accurate.

where $L^\dagger$ denotes the Moore-Penrose pseudo-inverse of L. The surface normals $\mathbf{n}_1, \ldots, \mathbf{n}_m$ can be estimated by normalizing each row of N to have unit norm.

While Eq. (7) follows from our formulation, it is not tractable since both rank and ℓ_0-norm are non-convex and discontinuous functions. Solving this optimization problem efficiently will be the topic of discussion in the next section.

3 Efficient Solution via Convex Programming

As discussed above, the optimization problem given in Eq. (7) is extremely difficult (NP-hard in general) to solve. In this section, we propose to solve it efficiently based on recent advances in algorithms for matrix rank minimization [16], [17], [20].

3.1 Convex relaxation and modification

Recently, Candès *et al*. [16] and Chandrasekaran *et al*. [17] have proposed that the problem in Eq. (7) can be solved efficiently by replacing the cost function with its convex surrogate, provided that the rank of the matrix F is not too high and the number of non-zero entries in the matrix E is not too large. This convex relaxation, dubbed *Principal Component Pursuit* (PCP) in [16], replaces $\mathrm{rank}(\cdot)$ with the *nuclear norm* (sum of the singular values of the matrix) and the ℓ_0-norm with the matrix ℓ_1-norm (sum of the absolute values of all entries of the matrix). Under quite general conditions, it has been proved in [16], [17] that the following optimization problem can recover the low-rank matrix F from corrupted observations O:

$$\min_{F,E} \|F\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad O = F + E, \tag{9}$$

where $\|\cdot\|_*$ and $\|\cdot\|_1$ represent the nuclear norm and ℓ_1-norm, respectively, and $\lambda > 0$ is a weighting parameter. Theoretical considerations in [20], [16] suggest that λ must be of the form $C/\sqrt{\max\{m,n\}}$, where C is a constant, typically set to unity. It is interesting to note that the equivalence between Eq. (7) and Eq. (9) is not affected by the magnitude of the singular values of the solution F or by the magnitude of the non-zero entries of the error matrix E.

In the framework of PCP, the locations of the non-zero entries of the sparse matrix E are assumed to be unknown. But if the locations of some of the corrupted entries are known, then we can incorporate that information into the recovery procedure and make the problem somewhat easier to solve. This is similar in spirit to the matrix completion problem [21], [22], [23]. Notice that although both shadows and specularities corrupt the low-rank matrix, they have different characteristics. While the locations of the specular pixels are hard to detect, especially that of pixels in specular lobes, it is relatively easy to detect the location of shadows in an image (*e.g.*, by a simple thresholding of the pixel values). Thus, we have more information about the shadows than specularities, and such information can greatly help finding the correct solution. So mathematically, we have a problem of recovering a low-rank matrix with both missing entries (shadows) and unknown corrupted entries (specularities).

We denote by Ω the locations of missing entries in the observed matrix O, defined in Eq. (3), that correspond to shadows in the input images. By a slight abuse of notation, we also denote by Ω the linear subspace of $m \times n$ matrices with support in Ω. Let π_Ω represent the orthogonal projection operator corresponding to the subspace Ω. Thus, we modify the PCP problem in Eq. (9) to the following one which does both matrix completion and error correction:

$$\min_{F,E} \|F\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad \pi_{\Omega^c}(O) = \pi_{\Omega^c}(F + E), \tag{10}$$

where Ω^c denotes the linear subspace orthogonal to Ω, and π_{Ω^c} is the associated projection operator. The above problem is almost identical to the PCP problem (Eq. (9)), except that the linear equality constraint is now applied only on the set Ω^c of pixels that are not affected by the detected shadows. It is not difficult to show that the above problem is equivalent to

$$\min_{F,E} \|F\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad \pi_{\Omega^c}(O) = \pi_{\Omega^c}(F) + E. \tag{11}$$

This is because if there exists an optimal solution $\hat{E}$ that has a non-zero component in Ω, we could set it to zero and achieve a reduction in the cost function. The feasibility constraint is not affected by this operation since it does not alter any entry in Ω^c.

3.2 Fast Algorithm using Augmented Lagrangian Methods

The optimization problem in Eq. (11) can be recast as a semidefinite program and solved using interior-point methods. Although interior-point methods have excellent convergence properties, they are not very scalable for large problems. Fortunately, there has been a flurry of work recently on developing scalable algorithms for high-dimensional nuclear-norm minimization [24], [25], [26]. In this section, we show how one such algorithm, the Augmented Lagrange Multiplier (ALM) method [24], [27], can be adapted to efficiently solve Eq. (11).

The basic idea of the ALM method is to minimize an augmented Lagrangian function instead of directly solving the original constrained optimization problem. For our problem Eq. (11), the augmented Lagrangian is given by

$$\begin{aligned}\mathcal{L}(F, E, Y, \mu) = \|F\|_* \quad + \quad & \lambda\|E\|_1 + \langle Y, \pi_{\Omega^c}(O - F) - E\rangle \\ & + \frac{\mu}{2}\|\pi_{\Omega^c}(O - F) - E\|_F^2,\end{aligned} \tag{12}$$

where $Y \in \mathbb{R}^{m \times n}$ is a Lagrange multiplier matrix, $\mu > 0$, $\langle \cdot, \cdot \rangle$ denotes the matrix inner product,[4] and $\|\cdot\|_F$ denotes the matrix Frobenius norm. With an appropriate choice

4. $\langle X, Y \rangle \doteq \mathrm{trace}(X^T Y)$.

Algorithm 1 (Matrix Completion and Recovery via ALM).

INPUT: $O \in \mathbb{R}^{m\times n}$, $\Omega \subset \{1,\ldots,m\}\times\{1,\ldots,n\}$, $\lambda > 0$, $\rho > 1$
Initialize $F_1 \leftarrow 0$, $E_1 \leftarrow 0$, $Y_1 \leftarrow 0$.
while not converged ($k = 1, 2, \ldots$) **do**
  $F_{k,1} \leftarrow F_k$, $E_{k,1} \leftarrow E_k$
  while not converged ($j = 1, 2, \ldots$) **do**
    $E_{k,j+1} = \text{shrink}\left(\pi_{\Omega^c}(O - F_{k,j}) + \frac{1}{\mu_k}Y_k, \frac{\lambda}{\mu_k}\right)$
    $t_1 \leftarrow 1$, $Z_1 \leftarrow F_{k,j}$, $F_{k,j,1} \leftarrow F_{k,j}$
    while not converged ($i = 1, 2, \ldots$) **do**
      $(U_i, \Sigma_i, V_i) \leftarrow \text{svd}\left(\frac{1}{\mu_k}Y_k + \pi_{\Omega^c}(O) - E_{k,j+1} + \pi_\Omega(Z_i)\right)$
      $F_{k,j,i+1} \leftarrow U_i \,\text{shrink}\left(\Sigma_i, \frac{1}{\mu_k}\right) V_i^T$
      $t_{i+1} \leftarrow 0.5\left(1 + \sqrt{1 + 4t_i^2}\right)$
      $Z_{i+1} \leftarrow F_{k,j,i+1} + \frac{t_i - 1}{t_{i+1}}(F_{k,j,i+1} - F_{k,j,i})$
      $F_{k,j+1} \leftarrow F_{k,j,i+1}$
    end while
    $F_{k+1} \leftarrow F_{k,j+1}$, $E_{k+1} \leftarrow E_{k,j+1}$
  end while
  $Y_{k+1} \leftarrow Y_k + \mu_k\left(\pi_{\Omega^c}(O - F_{k+1}) - E_{k+1}\right)$
  $\mu_{k+1} \leftarrow \rho \cdot \mu_k$
end while
OUTPUT: $(\hat{F}, \hat{E}) = (F_k, E_k)$.

of the Lagrange multiplier matrix Y and sufficiently large μ, it can be shown that the augmented Lagrangian function has the same minimizer as the original constrained optimization problem in Eq. (11) [27]. The ALM algorithm iteratively estimates both the Lagrange multiplier and the optimal solution.

The basic ALM iteration is given by

$$\left\{\begin{array}{rcl}(F_{k+1}, E_{k+1}) & = & \text{argmin}_{F,E}\ \mathcal{L}(F, E, Y_k, \mu_k)\\ Y_{k+1} & = & Y_k + \mu_k\,\pi_{\Omega^c}(O - F_{k+1}) - E_{k+1}\\ \mu_{k+1} & = & \rho\cdot\mu_k.\end{array}\right. \tag{13}$$

where $\{\mu_k\}$ is a monotonically increasing positive sequence ($\rho > 1$).

We now focus our attention on solving the non-trivial first step of the above iteration. Since it is difficult to minimize $\mathcal{L}(\cdot)$ with respect to both F and E simultaneously, we adopt an alternating minimization strategy as follows:

$$\left\{\begin{array}{rcl}E_{j+1} & = & \text{argmin}_E\ \lambda\|E\|_1 - \langle Y_k, E\rangle\\ & & +\frac{\mu_k}{2}\|\pi_{\Omega^c}(O - F_j) - E\|_F^2\\ F_{j+1} & = & \text{argmin}_F\ \|F\|_* - \langle Y_k, \pi_{\Omega^c}(F)\rangle\\ & & +\frac{\mu_k}{2}\|\pi_{\Omega^c}(O - F) - E_{j+1}\|_F^2.\end{array}\right. \tag{14}$$

The above minimization problems in Eq. (14) can be solved as described below.

We first define the *shrinkage* (or soft-thresholding) operator for scalars as follows:

$$\text{shrink}(x, \alpha) = \text{sign}(x)\cdot\max\{|x| - \alpha, 0\}, \tag{15}$$

where $\alpha \geq 0$. When applied to vectors or matrices, the shrinkage operator acts element-wise. Then, the first step in Eq. (14) has a closed-form solution given by

$$E_{j+1} = \text{shrink}\left(\pi_{\Omega^c}(O - F_j) + \frac{1}{\mu_k}Y_k, \frac{\lambda}{\mu_k}\right). \tag{16}$$

Since it is not possible to express the solution to the second step in Eq. (14) in closed-form, we adopt an iterative strategy based on the Accelerated Proximal Gradient (APG) algorithm [28], [26], [25] to solve it. The iterative procedure is given as:

$$\left\{\begin{array}{rcl}(U_i, \Sigma_i, V_i) & = & \text{svd}\left(\frac{1}{\mu_k}Y_k + \pi_{\Omega^c}(O) - E_{j+1} + \pi_\Omega(Z_i)\right)\\ F_{i+1} & = & U_i\,\text{shrink}\left(\Sigma_i, \frac{1}{\mu_k}\right)V_i^T\\ Z_{i+1} & = & F_{i+1} + \frac{t_i - 1}{t_{i+1}}(F_{i+1} - F_i).\end{array}\right. \tag{17}$$

where $\text{svd}(\cdot)$ denotes the Singular Value Decomposition operator, and $\{t_i\}$ is a positive sequence satisfying $t_1 = 1$ and $t_{i+1} = 0.5\left(1 + \sqrt{1 + 4t_i^2}\right)$. The entire algorithm to solve Eq. (11) has been summarized as Algorithm 1.

In our experience, it is not necessary to solve the innermost loop of Algorithm 1 exactly, but an approximate solution is sufficient. More specifically, we found that restricting the innermost loop to just one iteration does not significantly affect the convergence of the algorithm. Although the convergence of the ALM method in Algorithm 1 has been well established in the optimization literature, we currently know of no proof that its approximation described here converges too. The main difficulty comes from the fact that we have a π_{Ω^c} operator in each iteration of the alternating minimization. The projection operator can be equivalently replaced with a separate variable term in the equality constraint. The case without any projection operator between two terms

has been studied extensively as the *alternating direction method of multipliers* in the optimization literature and its convergence has been well established for various cases [29], [30], [31]. In particular, the convergence for the Principal Component Pursuit problem, given in (9), has been established in [24]. Recently, [32] obtained a convergence result for a certain three-term alternation scheme applied to the noisy Principal Component Pursuit problem (see also [33]). However, [32] reflects a very similar theory-practice gap – the three-term alternation for which convergence has been established is slower in practice than an alternation used in this paper, for which a rigorous proof of convergence remains elusive.

4 Photometric Factor Classification

In this section, we show that our robust photometric stereo approach can be applied to the problem of classifying photometric factors, such as diffuse reflection, specular reflection, and shadows. The classification of photometric factors is important for various vision tasks including image segmentation [34], shape recovery from shadows [35], and shape from specularity [36]. Furthermore, the classification of photometric factors can help in object recognition by enabling the removal of non-diffuse components of the image, as well as provide a more compact representation of objects in an image [37], [38], [39], [40].

In the context of photometric stereo, specularities and shadows are common factors that make accurate estimation of surface normals a difficult problem. There have been various techniques to identify and remove these photometric factors. Wolff and Boult [41] presented a polarization reflectance model to segment material surfaces according to varying levels of relative electrical conductivity. Ikeuchi and Sato [42] proposed a method for determining reflectance properties based on the depth and brightness observations. Lin *et al*. [34] proposed a method for diffuse-specular separation based on color analysis and multi-baseline stereo. Chandraker *et al*. [43] proposed an algorithm for Lambertian photometric stereo in the presence of shadows by graph cuts. Mukaigawa *et al*. [15] proposed a method to classify photometric factors based on photometric linearization. Their method used a RANSAC approach to linearize input images and classified pixels into diffuse, specular, and shadowed pixels with high accuracy.

Using a classification criteria similar to the one suggested in Mukaigawa *et al*. [15], we show that our rank minimization approach can be used for classifying photometric factors with much less computational cost than using a RANSAC based approach. Each iteration of Algorithm 1 requires a SVD computation that has a complexity of $O(mn^2)$. However, to speed up the algorithm, we compute only partial SVDs in each iteration since we expect the optimal solution to have rank at most 3. Thus, the complexity of each iteration reduces to $O(lmn)$, where l is the number of singular vectors computed in each iteration. Typically, we set l to be much smaller than m and n. Although there is no good theoretical bound on the number of iterations of the ALM method used in Algorithm 1, in practice, the algorithm converges after a few hundred iterations. The RANSAC-based method has a complexity of $O(kmn)$, where k is the number of iterations. Although the RANSAC-based approach has a complexity similar to our method, in reality, the number of iterations k is typically quite large. Moreover, the RANSAC criterion is implemented twice in the algorithm proposed in [15], for the base images and for the calculation of coefficients, which further increases the computation time. On the other hand, our method computes the linearized images in a holistic fashion and comes with theoretical guarantees for recovery.

Unlike Mukaigawa's method, which linearizes images by random sampling, our method works in a global manner. Our method recovers a rank-3 matrix F from the observed image matrix O and corrects errors caused by corruptions (specularities, cast shadow and attached shadow) via an efficient convex program. Therefore, the ideal diffuse images, *i.e.*, linearized images in [15], can be naturally obtained.

Borrowing some notation from [15], we define five different classes of photometric factors - cast shadow ($\mathcal{C}$), attached shadow ($\mathcal{A}$), diffuse reflection ($\mathcal{D}$), specular reflection ($\mathcal{S}$), and undefined regions ($\mathcal{U}$). Our goal is to classify the pixels in each image, or equivalently each entry of O, into one of these five classes.

Suppose that we apply our algorithm described in the previous section to the observation matrix O, and let F be the recovered matrix of rank 3. Let us order the entries of O and F with a single index k, and let the k^{th} entry of O and F be denoted by O_k and F_k, respectively. We now use the Lambertian image formation model described in Section 2.1 to define a simple procedure to classify the image pixels.

We have seen that if a point in the scene is illuminated by the light source, then the angle between the surface normal $\mathbf{n}$ at the point and lighting direction $\mathbf{l}$ is acute, *i.e.*, $F_i = \mathbf{n} \cdot \mathbf{l} > 0$. If this point is not corrupted by shadows and specularities, we have $O_i = F_i$. On the other hand, this point could be corrupted by cast shadows or specularities. In the case of cast shadow, we would have $O_i = 0$, whereas in the case of a specularity, we would typically have $O_i > F_i$. The other possibility is that the point does not receive any light from the source. In this case, we have $F_i = \mathbf{n} \cdot \mathbf{l} \leq 0$. Equivalently, this point is part of an attached shadow, and we have $O_i = 0$. Thus, each case can be distinguished by studying the corresponding O_i and F_i values.

Mathematically , there could be cases when the values of O_i and F_i do not satisfy any of the above criteria, although such cases may not be possible physically. We label such pixels as undefined. Furthermore, in practice, shadowed regions seldom have perfectly zero pixel values due to inter-reflections and imaging noise in the sensor. In addition, due to observation noise and

numerical errors, there may not be a case where $F_i = O_i$ holds exactly. Thus, we relax the equalities in the criteria described above with thresholds $T_1, T_2 \in [0,1]$ that are chosen empirically. We summarize below the revised classification criteria:

$$\begin{cases} \mathcal{C} &= \{i \,|\, F_i > 0, O_i \leq T_1\} \\ \mathcal{A} &= \{i \,|\, F_i \leq 0, O_i \leq T_1\} \\ \mathcal{D} &= \{i \,|\, |F_i - O_i| < T_2 \cdot O_i, O_i > T_1\} \\ \mathcal{S} &= \{i \,|\, O_i - F_i > T_2 \cdot O_i, O_i > T_1\} \end{cases} \quad (18)$$

5 Experiments

In this section, we verify the efficacy of the proposed method using both synthetic and real-world images. We first compare our results with a simple Least Squares (LS) approach [1], which assumes the ideal diffusive model given by Eq. (4). However, we do not use those pixels that were classified as shadows (the set Ω). In practice, we set the threshold for determining shadowed pixels. Thus, the LS method can be summarized by the following optimization problem:

$$\min_N \; \|\pi_{\Omega^c}(O - N\,L)\|_F. \quad (19)$$

We also compare our method with a RANSAC-based method proposed in [15]. We also demonstrate the efficacy of our method in photometric factors classification through extensive experimental results.

We first test our algorithm using synthetic images whose ground-truth normal maps are known [44]. In these experiments, we quantitatively verify the correctness of our algorithm by computing the angular errors between the estimated normal map and the ground-truth. We then test our algorithm on more challenging real images. Throughout this section, we denote by m the number of pixels in the region of interest in each image, and by n the number of input images (typically, $m \gg n$).

5.1 Quantitative evaluation with synthetic images

In this section, we use synthetic images of three different objects (see Fig. 2(a)-(c)) under different scenarios to evaluate the performance of our algorithm. Since these images are noise-free, we use a pixel threshold value of zero to detect shadows in the images. Unless otherwise stated, we set $\lambda = 1/\sqrt{m}$ in Eq. (11).

a. Specular scene. In this experiment, we generate images of an object under 40 different lighting conditions, where the lighting directions are chosen at random from a hemisphere with the object placed at the center. The images are generated with some specular reflection. For all our experiments, we use the Cook-Torrance reflectance model [45] to generate images with specularities. Thus, there are two sources of corruption in the images – attached shadows and specularities.

A quantitative evaluation of our method and the Least Squares approach is presented in Table 1. The estimated normal maps are shown in Fig. 3(b),(c). We use the RGB channel to encode the three spatial components (XYZ) of the normal map for display purposes. The error is measured in terms of the angular difference between the ground truth normal and the estimated normal at each pixel location. The pixel-wise error maps are shown in Fig. 3(d), (e). From the mean and the maximum angular error (in degrees) in Table 1, we see that our method is much more accurate than the LS approach. This is because specularities introduce large magnitude errors to a small fraction of pixels in each image whose locations are unknown. The LS algorithm is not robust to such corruptions while our method can correct these errors and recover the underlying rank-3 structure of the matrix. The row on the extreme bottom of Table 1 indicates the average percentage of pixels in each image (averaged over all images) that were corrupted by shadows and specularities, respectively. We note that even when more than 30% of the pixels are corrupted by shadows and specularities, our method can efficiently retrieve the surface normals.

(a) Sphere (b) Caesar (c) Elephant (d) Caesar (With texture)

Fig. 2. Synthetic images used for experiments.

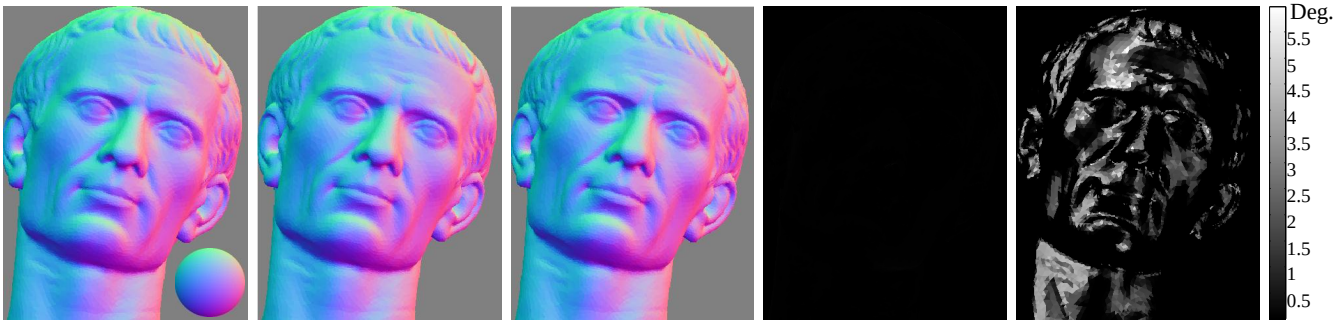
TABLE 1
Specular scene.

Object		Sphere	Caesar	Elephant
Mean error (in degrees)	LS	0.99	0.96	0.96
	Our method	**0.0051**	**0.0014**	**0.0087**
Max error (in degrees)	LS	8.1	8.0	8.0
	Our Method	**0.20**	**0.22**	**0.29**
Avg. % of corruption	Shadow	18.4	20.7	18.1
	Specularity	16.1	13.6	16.5

Statistics of angle error in the normals for different objects. In each case, 40 images were used. In the bottom row, we indicate the average percentage of pixels corrupted by attached shadows and specularities in each image.

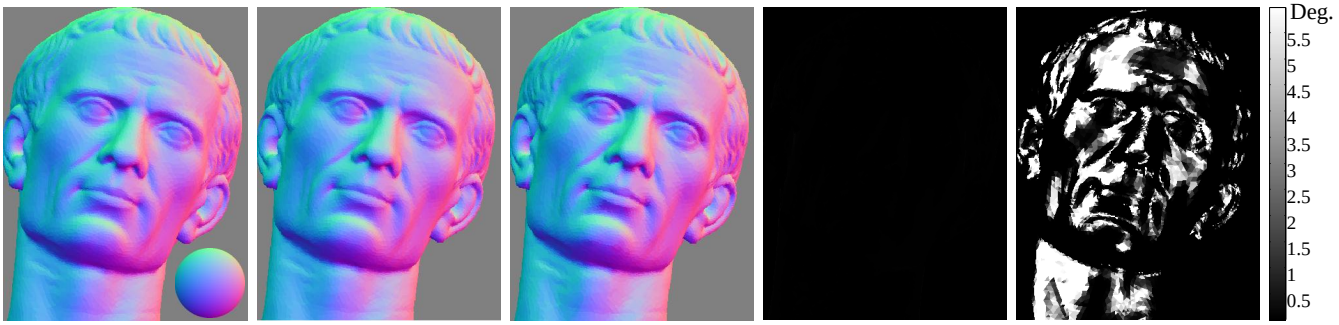
b. Textured scene. We also test our method using a textured scene. Like the traditional photometric stereo approach, our method does not have a dependency on the albedo distribution and works well on such scenes.

We use 40 images of Caesar for this experiment with each image generated under a different lighting condition (see Fig. 2(d) for example input image). The estimated normal maps as well as the pixel-wise error maps are shown in Fig. 4. We provide a quantitative comparison in Table 2 with respect to the ground-truth normal map. From the mean and maximum angular errors, it is evident that our method performs much better than the LS approach in this scenario.



(a) Ground truth (b) Our method (c) Least Squares (d) Error map (our method) (e) Error map (LS)

Fig. 3. **Specular scene.** 40 different images of Caesar were generated using the Cook-Torrance model for specularities. (a) Ground truth normal map with reference sphere. (b) and (c) show the surface normals recovered by our method and LS, respectively. (d) and (e) show the pixel-wise angular error w.r.t. the ground truth.



(a) Ground truth (b) Our method (c) Least Squares (d) Error map (our method) (e) Error map (LS)

Fig. 4. **Textured scene with specularity.** 40 different images of Caesar were generated with texture, using the Cook-Torrance model for specularities. (a) Ground truth normal map with reference sphere. (b) and (c) show the surface normals recovered by our method and LS, respectively. (d) and (e) show the pixel-wise angular error w.r.t. the ground truth.

TABLE 2
Textured scene with specularity.

Object	Mean error (in degrees)		Max error (in degrees)	
	LS	Our method	LS	Our method
Caesar	2.2	**0.014**	36.5	**0.19**

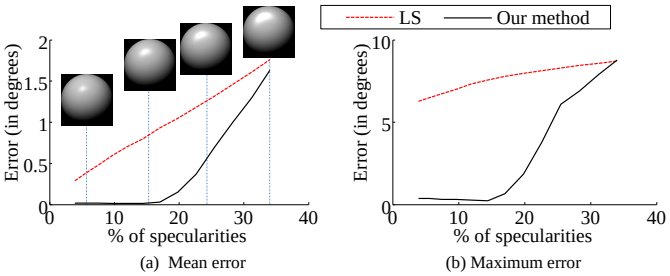
Statistics of angle errors. We use 40 images under different illuminations.

TABLE 3
Effect of number of input images.

No. of images	Mean error (in degrees)		Max error (in degrees)	
	LS	Our method	LS	Our method
5	**4.5**	15.1	88.2	127.9
10	0.52	**0.23**	**34.5**	56.6
15	0.51	**0.036**	**13.7**	25.6
20	0.53	**0.026**	9.0	**5.8**
25	0.62	**0.015**	8.4	**0.42**
30	0.59	**0.019**	7.6	**0.48**
35	0.59	**0.017**	7.6	**0.37**
40	0.57	**0.013**	7.0	**0.37**

We use synthetic images of Caesar under different lighting conditions. The number of illuminations is varied from 5 to 40. The angle error is measured with respect to the ground truth normal map. The illuminations are chosen at random, and the error has been averaged over 20 different sets of illumination.

c. Effect of the number of input images. In the above experiments, we have used images of the object under 40 different illuminations. In this experiment, we study the effect of the number of illuminations used. In particular, we would like to find out empirically the minimum number of images required for our method to be effective. For this experiment, we generate images of Caesar using the Cook-Torrance reflectance model, where the lighting directions are generated at random. The mean percentage of specular pixels in the input images is maintained approximately constant at 10%. The angular difference between the estimated normal map and the ground truth is used as a measure of accuracy of the estimate.



(a) Mean error (b) Maximum error

Fig. 5. **Effect of increasing size of specular lobes.** We use synthetic images of Caesar under 40 randomly chosen lighting conditions. (a) Mean angular error, (b) Maximum angular error w.r.t. the ground truth. The illuminations are chosen at random, and the error has been averaged over 10 different sets of illumination. (a) contains illustrations of increasing size of specular lobe.

We present the experimental results in Table 3. We observe that with five input illuminations, our method is worse than LS, although, in this scenario, the estimates of both algorithms are very inaccurate. When the number of illuminations is larger than ten, we observe that the mean error in the LS estimate is higher than that of our method. Upon increasing the number of images further, the proposed method consistently outperforms the LS approach. If the number of input images is less than 20, then the maximum error in the LS estimate is smaller than that of our method. However, our method performs much better when at least 25 different illuminations are available. Thus, the proposed technique performs significantly better as the number of input images increases.

d. Varying amount of specularity. From the above experiments, it is clear that the proposed technique is quite robust to specularities in the input images when compared to the LS method. In this experiment, we empirically determine the maximum amount of specularity that can be handled by our method. We use the Caesar scene under 40 randomly chosen illumination conditions for this experiment. On an average, about 20% of the pixels in each image are corrupted by attached shadows. We vary the size of the specular lobe in the input images (as illustrated in Fig. 5(a)), thereby varying the number of corrupted pixels. We compare the accuracy of our method against the LS technique using the angular error of the estimates with respect to the ground-truth.

The experimental results are illustrated in Fig. 5. We observe that our method is very robust when up to 16% of all pixels in the input images are corrupted by specularities. The LS method, on the other hand, is extremely sensitive to even small amounts of specularities in the input images. The angular error in the estimates of both methods rises as the size of the specular lobe increases.

e. Enhancing performance by better choice of λ. We recall that λ is a weighting parameter in our formulation given by Eq. (11). In all the above experiments, we have fixed the value of the parameter $\lambda = 1/\sqrt{m}$, as suggested in [16]. While this choice promises a certain degree of error correction, it may be possible to correct larger amounts of corruption by choosing λ appropriately, as demonstrated in [46] for instance. Unfortunately, the best choice of λ depends on the input images, and cannot be determined analytically.

TABLE 4
Handling more specularities by appropriately choosing λ.

C	1.0	0.8	0.6	0.4
Mean error (in degrees)	1.42	0.78	0.19	0.029
Max error (in degrees)	8.78	8.15	1.86	0.91

We use 40 images of Caesar under different lighting conditions specularities and shadows, and set $\lambda = C/\sqrt{m}$.

We demonstrate the effect of the weighting parameter λ on a set of 40 images of Caesar used in the previous experiments. In this set of images, approximately 20% of the pixels are corrupted by attached shadows and about 28% by specularities. We choose $\lambda = C/\sqrt{m}$, and vary the value of C. We evaluate the results using angular error with respect to the ground-truth normal map. We observe from Table 4 that the choice of C greatly influences the accuracy of the estimated normal map. For real-world applications, where the data is typically noisy, the choice of λ could play an important role in the efficacy of our method.

f. Computation. The core computation of our method is solving a convex program of Eq. (11). For the specular Caesar data (Fig. 2(b)) with 40 images of 288×213 resolution, and with single-core MATLAB implementation of our method takes about 68 seconds on a PC with a 3.0 *GHz* Core4 i7 processor and 16 GB memory, as against 12 seconds taken by the LS approach. While our method is slower than the LS approach, it is much more accurate in a wide variety of scenarios and is more efficient than other existing methods (*e.g.*, [14]).

g. Comparison with RANSAC-based method. We also compare with the RANSAC-based method proposed in [15], where the assumption is that an image under any lighting direction can be expressed by a linear combination of three basis images. The photometric linearization is then regarded as a problem of finding deviations from this linear combination. Mukaigawa *et al.*'s method used RANSAC for calculating coefficients and identifying the basis images. In our approach, instead of using sampled observations, we use *all* the observations for achieving robust estimation without increasing the computational cost. Despite this, our method runs much faster than the RANSAC-based approach as demonstrated below.

In this experiment, 40 images with specularities were taken from varying lighting conditions. The first three images are selected as bases for the RANSAC-based method, in which average 39% of the pixels are corrupted. The number of samples is determined by the fraction of outliers in the data and their distribution for the RANSAC-based method. Since we have ground

TABLE 5
Theoretical number of samples of RANSAC-based method

No. of samples	Process of.	Max	Min	$Mean$
	Coefficients Calculation	11256	732	2256
	Base images Linearization	106	1	14

We use 40 images of Caesar under different lighting conditions specularities and shadows, specularities and shadows with size of 288 by 213. Here list the maximum, minimum and mean number of samples for coefficients calculation and base image linearization, respectively.

TABLE 6
Performance with theoretical number of samples of RANSAC-based method

RANSAC	Max error (in degrees.)	0.0376
	Mean error (in degrees.)	0.0037
	Computation time (in Sec.)	456
Our method	Max error (in degrees.)	0.0695
	Mean error (in degrees.)	0.005
	Computation time (in Sec.)	82

We use 40 images of Caesar under different lighting conditions specularities and shadows, specularities and shadows with size of 288 by 213. Both our method and RANSAC-based method achieve a higher degree of accuracy.

truth, thus, the optimal number of samples for each image can be easily obtained. In Table 5, we have listed the theoretical number of samples for coefficients calculation and base images linearization, respectively. The number can ensure that the results are correct with a probability of 99%. The experimental result is presented in Table 6. We observe both our method and RANSAC-based method achieve a higher degree of accuracy. While our method takes just 82 seconds as against 456 seconds taken by the RANSAC-based approach (All experiments in this section are running on the same PC described in f. Computation.).

Unfortunately, for the RANSAC-based method, it is hard to estimate a proper number of samples due to fraction of outliers in the data and their distribution are not known exactly a priori, although in some cases they can be approximated. In practice, one can limit the number of iterations, however, this could lead to sub-optimal solutions, as illustrated in Table 7. In this experiment, we fixed 110 samples [5] for base images linearization and test their method by with 800, 1,200 and 1,600, 2,000 samples for coefficients calculation, respectively. We observe from Table 7 that for the RANSAC-based method, the angular error is reduced as the number of samples increases. However, the results show some degree of randomness and uncertainty. For instance, when only 1200 samples are available.

5. The number can ensure the base images linearization is correct with high probability.

TABLE 7
Performance with varying number of samples of RANSAC-based method

No. of samples	800	1200	1600	2000
Max error (in degrees.)	0.7585	15.3252	0.3368	0.0417
Mean error (in degrees.)	0.2297	6.1671	0.1145	0.0059
Computation time (in Sec.)	996	1036	1082	1125

We use 40 images of Caesar under different lighting conditions specularities and shadows, specularities and shadows with size of 288 by 213. Inaccurate estimation of number of samples lead to sub-optimal solutions.

5.2 Qualitative evaluation with real images

We now test our algorithm on real images. We use a set of 40 images of a toy Doraemon and Two-face taken under different lighting conditions (see Fig. 6(a), (d)). A glossy sphere was placed in the scene for light source calibration when capturing the data. We used a Canon 5D camera in the RAW image mode without gamma correction. These images present new challenges to our algorithm. In addition to shadows and specularities, there is potentially additive noise inherent to the acquisition process as well as possible deviations from the idealistic Lambertian model illuminated by distant lights. In this experiment, we use a threshold of 0.01 to detect shadows in images.[6] We also found experimentally that setting $\lambda = 0.3/\sqrt{m}$ works well for these images.

Since the ground truth normal map is not available for these scenes, we compare our method and the LS approach by visual inspection of the output normal maps shown in Fig. 6(b), (c), (e), (f). We observe that the normal map estimated by our method appears smoother and hence, more realistic. This can be observed particularly around the necklace area in Doraemon and nose area in Two-face (see Fig. 6) where the LS estimate exhibits some discontinuity in the normal map.

5.3 Photometric Factor Classification

In this section, we show that our method can identify photometric factors with high accuracy on both synthetic and real data. We first demonstrate our results on synthetic data. For this experiment, the thresholds T_1 and T_2, defined in Eq. (18), are set to 10^{-7} and 10^{-3}, respectively. We use a set of 40 synthetic images of Caesar as input that are corrupted by attached shadows, cast shadows, and specularities as shown in Fig. 7.

We summarize our classification results in Table 8. As defined earlier, the labels $\mathcal{C}$, $\mathcal{A}$, $\mathcal{D}$, $\mathcal{S}$ and $\mathcal{U}$ indicate cast

6. All pixels are normalized to have intensity between 0 and 1.

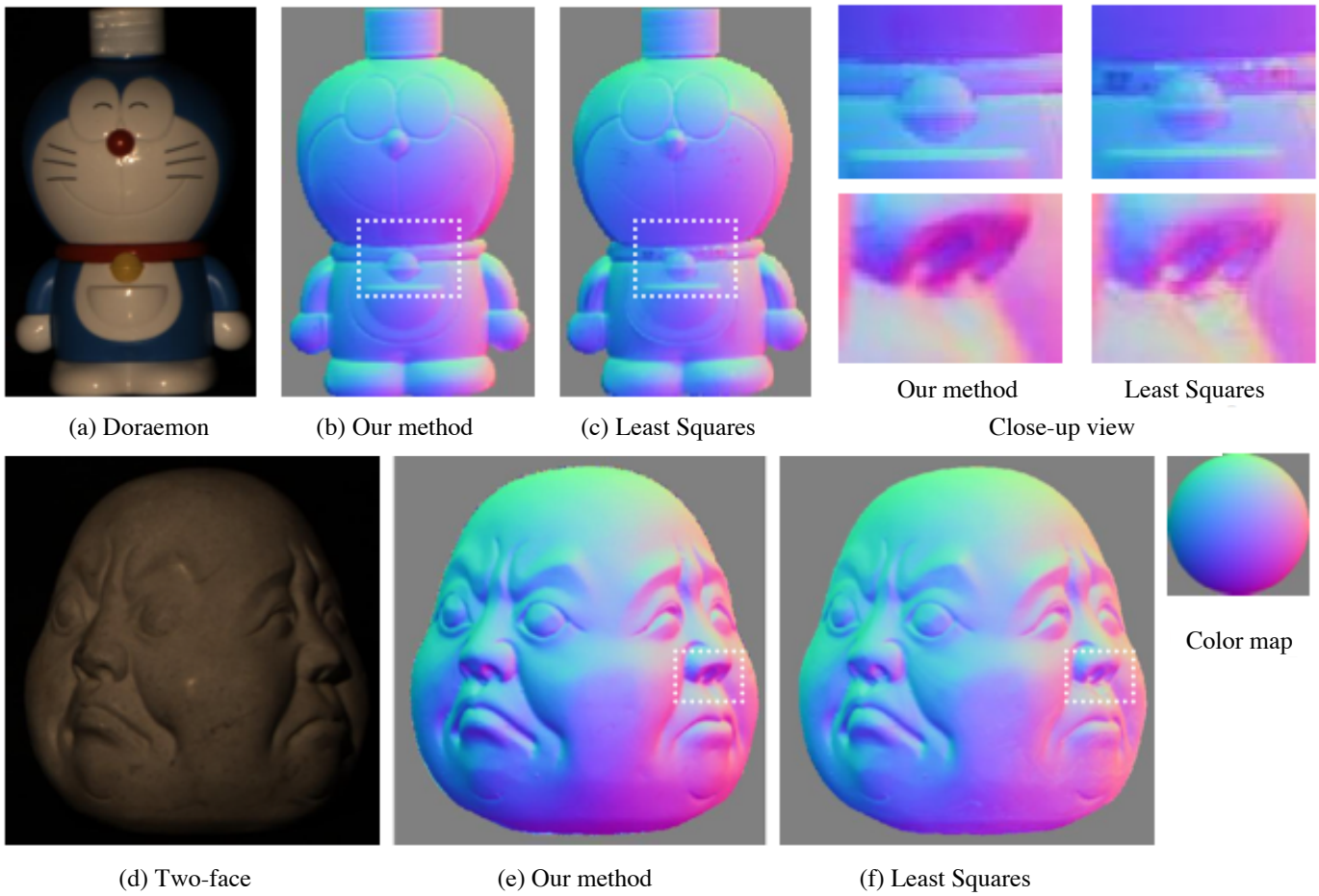


Fig. 6. **Qualitative comparison on real data.** We use images of Doraemon and Two-face taken under 40 different lighting conditions to qualitatively evaluate the performance of our algorithm against the LS approach. (a), (d) Sample input images. (b), (e) Normal map estimated by our method. (c), (f) Normal map estimated by Least Squares. Close-up views of the dotted rectangular areas (top-right) where the normal map estimate of our method is much more smoother and realistic than that of Least Squares.

shadow, attached shadow, diffuse reflections, specular reflection and undefined region, respectively. Table 8 summarizes the results of this experiment. For any two labels $\mathcal{X}$ and $\mathcal{Y}$, the entry in Table 8 corresponding to row $\mathcal{X}$ and column $\mathcal{Y}$ denotes the proportion of pixels of type $\mathcal{X}$ classified as $\mathcal{Y}$ by our scheme. We observe that cast shadows and attached shadows are perfectly classified. 99.65% of diffuse component is correctly classified, while 0.17% of diffuse component is misclassified as specular component and 0.18% of diffuse component is misclassified as undefined. On the other hand, 17.22% of specular component is misclassified as diffuse component, 82.78% of specular component is correctly classified. The RANSAC-based approach [15] can achieve a higher degree of accuracy if sufficient number of samples are provided. However, our method is more efficient than RANSAC-based method as shown in Section 5.1. Besides, our method uses *all* the observations simultaneously for achieving robust estimation, unlike the RANSAC-based scheme that uses a partial sample of the observations.

In Fig. 8, we show the distribution of the various components on the object as well as the classification output at each pixel. We choose the fourth image from our dataset for this illustration since the various classes of pixels are clearly visible in it. Photometric factors are indicated with different intensities, as shown in Fig. 8(b), (c). We observe that the output of our algorithm is very close to the ground truth. From Table 8, we notice that all the misclassified pixels come from diffuse and specular components. This is because specular pixels often occur as specular lobes and the pixels around the boundaries of these lobes appear very close to diffuse pixels. We see from Fig. 8(d)-(g), (h)-(k), that most of the misclassified pixels are either located at the edge of the specular lobes that have very small intensity values. For better clarity of illustration, all intensities are normalized to lie between 0 and 1.

We now test our algorithm on real images. We use two real-world datasets, named Two-face and Cup, with each of them containing 40 images taken under different lighting conditions (see Fig. 9(a) and Fig. 11(a)). The

Fig. 7. **Input images (4 out of 40)**. With all kinds of corruptions: Specularity, cast shadow, attached shadow.

TABLE 8
Accuracy of the classification(%)

	$\mathcal{C}$	$\mathcal{A}$	$\mathcal{D}$	$\mathcal{S}$	$\mathcal{U}$
$\mathcal{C}$	100	0	0	0	0
$\mathcal{A}$	0	100	0	0	0
$\mathcal{D}$	0	0	99.65	0.17	0.18
$\mathcal{S}$	0	0	17.22	82.78	0

Where $\mathcal{C}, \mathcal{A}, \mathcal{D}, \mathcal{S}, \mathcal{U}$ indicate cast shadow, attached shadow, diffuse reflections,specular reflection, undefined, respectively. Each row shows the percentage that how many pixels are classified into the corresponding items in column.

thresholds (T_1, T_2) are set to $(0.1, 0.15)$ for the Two-face data, and $(0.12, 0.3)$ for the Cup data, respectively. Fig. 9(b)-(d), and Fig. 11(b)-(d) show the results of recovered diffuse component, non-diffuse component by our method, and the photometric factor classification results (indicated with different intensities). Although we do not have the ground truth of photometric factors for this data, we can still observe that the photometric factors are classified quite accurately.

We also test our algorithm on some face images from the Extended Yale B database [47]. For this experiment, the thresholds T_1 and T_2 are set to 0.15 and 0.1, respectively. We select 31 images of a subject taken under different lighting conditions. Some sample images from this dataset is illustrated in Fig. 10. We observe that our algorithm successfully removes the specularities and shadows from the face images.

6 Discussion and Future Work

In this paper, we have presented a new computational framework for robust photometric stereo. We have formulated the basic photometric stereo problem as a rank minimization problem that can be solved efficiently by convex optimization. The efficacy of our method is demonstrated using synthetic and real images. The biggest advantage of the proposed technique is its ability to handle shadows, specularities, and other kinds of large-magnitude, non-Gaussian errors in a holistic fashion. We have also shown that a simple extension of this method can be applied to classify photometric factors very effectively.

The new framework also opens up several avenues for future research. In this work, we have assumed that all the input images are noise-free and perfectly aligned with each other at the pixel level. However, in real world scenarios, small noise and misalignment are commonplace in any data acquisition process. It has already been shown in [48] that the low-rank matrix recovery framework used in this paper is stable to additive Gaussian noise. By exploring the low-rank structure described in this work, we believe that the proposed technique can be extended to simultaneously handle small noise and misalignment in the input images.

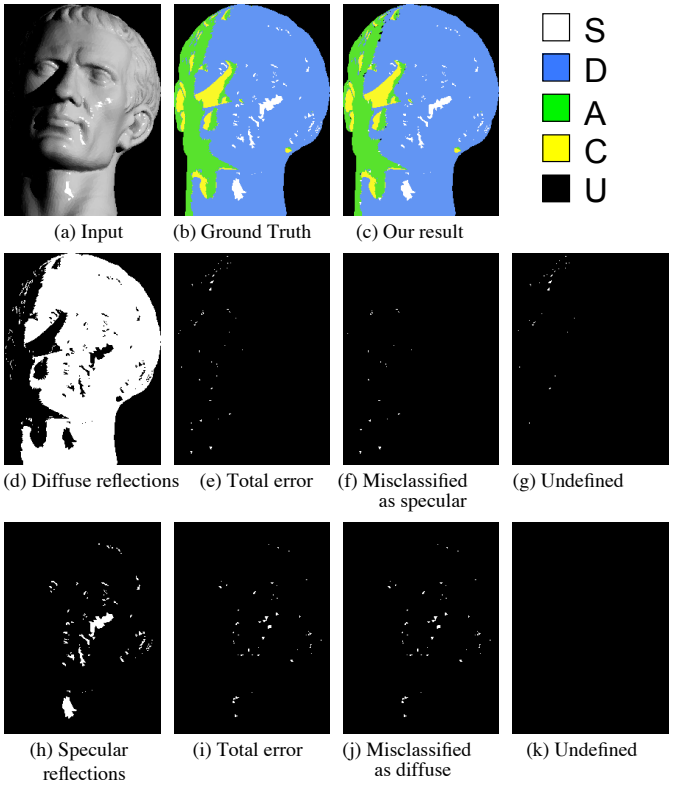


Fig. 8. **Classification results.** All photometric factors are indicated with different intensities: (a) Input synthetic images, (b) Ground truth, (c) Our result,(d) Distribution of diffuse component, (e) Total misclassified distribution for diffuse component, (f) Distribution of pixels which are misclassified as specular, (g) Distribution of pixels which are misclassified as undefined for diffuse component, (h) Distribution of specular component, (i) Total misclassified distribution for specular component, (j) Distribution of pixels which are misclassified as diffuse for specular component, (k) Distribution of pixels which are misclassified as undefined for specular component.

References

[1] R. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.
[2] W. Silver, "Determining shape and reflectance using multiple images," *Master's thesis, MIT*, 1980.
[3] A. Shashua, "Geometry and photometry in 3D visual recognition," *Ph.D dissertation, Department of Brain and Cognitive Science, MIT*, 1992.

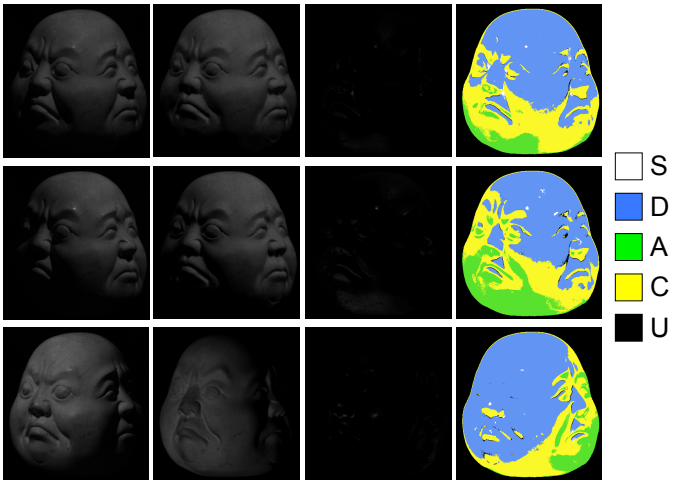


Fig. 9. **Photometric factor classification (Twoface).** (a) Input, (b) Diffuse component, (c) Non-diffuse component, (d) Classification results (indicated with different intensities.)

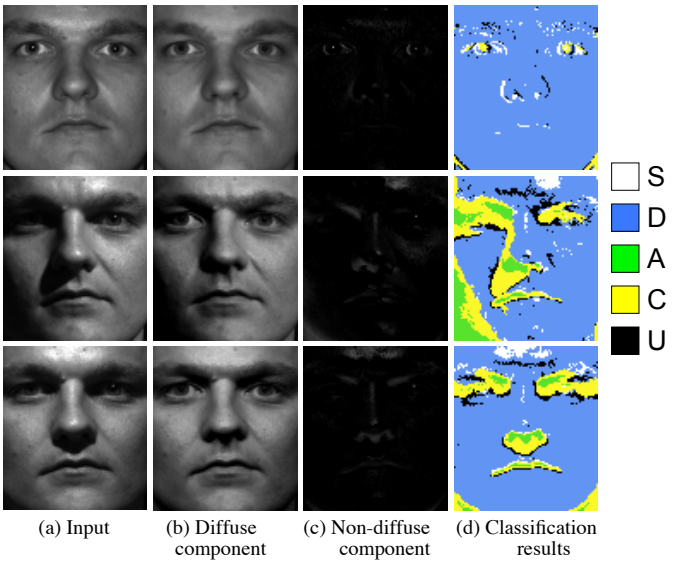


Fig. 10. **Photometric factor classification (Face).** (a) Input, (b) Diffuse component, (c) Non-diffuse component, (d) Classification results (indicated with different intensities).

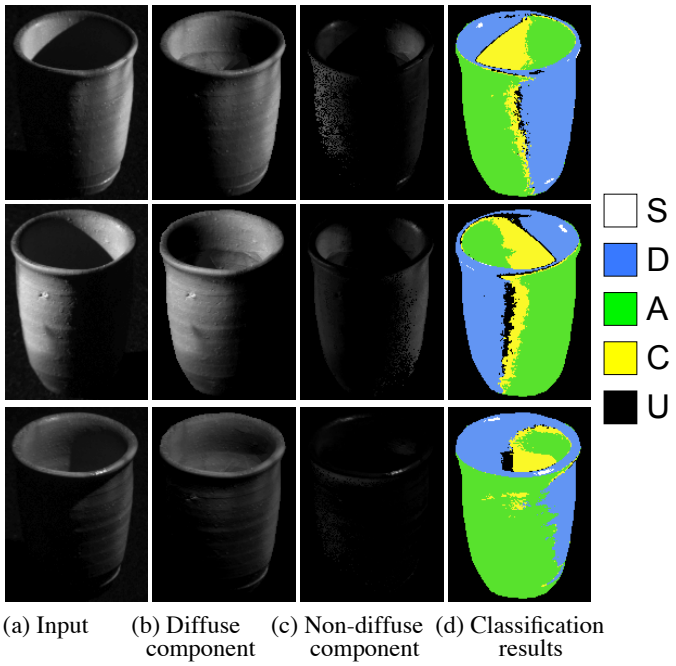


Fig. 11. **Photometric factor classification (Cup).** (a) Input, (b) Diffuse component, (c) Non-diffuse component, (d) Classification results (indicated with different intensities). Data courtesy of Professor Mukaigawa at Osaka University.

[4] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible lighting conditions?" in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 1996.
[5] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.
[6] A. Georghiades, D. Kriegman, and P. Belhumeur, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
[7] I. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.
[8] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *Journal of the Optical Society of America*, vol. 11, no. 11, pp. 3079–3089, 1994.
[9] D. Jacobs, "Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
[10] P. Belhumeur, D. Kriegman, and A. Yuille, "The bas-relief ambiguity," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 33–44, 1999.
[11] A. Yuille, D. Snow, R. Epstein, and P. Belhumeur, "Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 203–222, 1999.
[12] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model-fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[13] C. Hernández, G. Vogiatzis, and R. Cipolla, "Multi-view photometric stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 548–554, 2008.
[14] D. Miyazaki, K. Hara, and K. Ikeuchi, "Median photometric stereo as applied to the segonko tumulus and museum objects," *International Journal of Computer Vision*, vol. 86, no. 2, pp. 229–242, 2010.
[15] Y. Mukaigawa, Y. Ishii, and T. Shakunaga, "Analysis of photometric factors based on photometric linearization," *Journal of the Optical Society of America*, vol. 24, no. 10, pp. 3326–3334, 2007.
[16] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.
[17] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.
[18] 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 Asian Conference on Computer Vision*, 2010.
[19] D. Knill, P. Mamassian, and D. Kersten, "The geometry of shadows," *Journal of the Optical Society of America*, vol. 14, no. 12, pp. 3216–3232, 1997.

[20] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization," in *Proc. of Neural Information Processing Systems*, 2009.
[21] B. Recht, M. Fazel, and P. Parillo, "Guaranteed minimum rank solution of matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
[22] E. Candès and B. Recht, "Exact matrix completion via convex optimzation," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2008.
[23] E. Candès and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2053–2080, 2010.
[24] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *UIUC Technical Report UILU-ENG-09-2215*, 2009.
[25] A. Ganesh, Z. Lin, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast algorithms for recovering a corrupted low-rank matrix," in *Proc. of Computational Advances in Multi-Sensor Adaptive Processing*, 2009.
[26] K. Toh and S. Yun, "An accelerated proximal gradient algorithms for nuclear norm regularized least squares problems," *Pacific Journal of Optimization*, vol. 6, pp. 615–640, 2010.
[27] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2004.
[28] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problem," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2008.
[29] R. Glowinski and A. Marroco, "Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualite, d'une classe de problemes de Dirichlet non lineares," *Revuew Francaise d'Automatique, Informatique et Recherche Operationelle*, vol. 9, pp. 41–76, 1975.
[30] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximations," *Computers and Mathematics with Applications*, vol. 2, pp. 17–40, 1976.
[31] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, vol. 55, pp. 293–318, 1992.
[32] X. Yuan and M. Tao, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *preprint, http://www.optimization-online.org/DB_FILE/2009/12/2510.pdf*, 2010.
[33] B. He, "Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities," *Computational Optimzation and Applications*, vol. 42, no. 2, 2009.
[34] S. Lin, Y. Li, S. Kang, X. Tong, and H. Shum, "Diffuse-specular separation and depth recovery from image sequences," in *Proc. of European Conference on Computer Vision*, 2002.
[35] D. Kriegman and P. Belhumeur, "What shadows reveal about object structure," *Journal of the Optical Society of America*, vol. 18, no. 8, pp. 1804–1813, 2001.
[36] S. Savarese and P. Perona, "Local analysis for 3D reconstruction of specular surfaces," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
[37] M. Hansen, G. Atkinson, L. Smith, and M. Smith, "3D face reconstructions from photometric stereo using near infrared and visible light," *Computer Vision and Image Understanding*, vol. 114, no. 8, pp. 942 – 951, 2010.
[38] R. Woodman, "A Photometric Stereo Approach to Face Recognition," *Master's thesis, University of the West of England*, 2007.
[39] C. Julií, R. Moreno, D. Puig, and M. Garcia, "Shape-based image segmentation through photometric stereo," *Computer Vision and Image Understanding*, vol. 115, pp. 91–104, January 2011.
[40] A. Shashua, "On photometric issues in 3D visual recognition from a single 2D image," *International Journal of Computer Vision*, vol. 21, pp. 99–122, 1997.
[41] L. Wolff and T. Boult, "Constraining object features using a polarization reflectance model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 635–657, 1991.
[42] K. Ikeuchi and K. Sato, "Determining reflectance properties of an object using range and brightness images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 11, pp. 1139–1153, 1991.
[43] M. Chandraker, S. Agarwal, and D. Kriegman, "Shadowcuts: Photometric stereo with shadows," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
[44] "3D meshes research database by INRIA gamma group," http://www-roc.inria.fr/gamma/gamma/download/download.php.
[45] R. Cook and K. Torrance, "A reflectance model for computer graphics," *SIGGRAPH Computer Graphics*, vol. 15, no. 3, pp. 307–316, 1981.
[46] A. Ganesh, J. Wright, X. Li, E. Candès, and Y. Ma, "Dense error correction for low-rank matrices via principal component pursuit," in *Proc. of International Symposium on Information Theory*, 2010.
[47] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
[48] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *Proc. of International Symposium on Information Theory*, 2010.