Accurate Normal and Reflectance Recovery Using Energy Optimization

Tao Luo, Jianbing Shen, *Senior Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

Abstract—In this paper, we propose a novel energy optimization framework to accurately estimate surface normal and reflectance of an object from an input image sequence. Input images are captured from a fixed viewpoint under varying lighting conditions. In the proposed approach we combine photometric stereo and Retinex constraints into our energy function. To formulate inter-image constraints, shading information is added to the Lambertian model to account for shadows. For intra-image constraints, we moderate the strength of shading smoothness according to shadow mask and normal variations. By minimizing this energy function we are able to recover accurate surface normals and reflectance. Experimental results show that our approach yields more realistic normal map and accurate albedo map than the state-of-the-art uncalibrated photometric stereo algorithms. As for intrinsic image decomposition, results on the real and synthetic scenes show that the proposed approach outperforms previous ones.

Index Terms—Energy optimization, intrinsic images, normal recovery, reflectance.

I. INTRODUCTION

IN THE image formation process, several factors affect the observed appearance of the scene. These factors, including illumination, surface reflectance, shape of the objects, position of the viewer, and so on, are the intrinsic properties of the imaging environment [1]. They are independent to each other while their visual effects are mixed in the captured image. Individually extracting each intrinsic component from photographs gives the middle level descriptions of the scene, which is a core problem in the computer vision and image processing society. High-level vision tasks such as object recognition will benefit from this kind of separation, and image editing applications such as retexturing [2], recoloring [3], [4], and relighting [5]–[8] would become easier with the intrinsic features of the image being separated.

The decomposition of an image into its individual intrinsic characteristics was first presented as the intrinsic images problem in [1]. Because the original definition is severely ill-posed, over time it has been simplified to the problem of separating an image into reflectance and shading components. The reflectance component represents how the surface material reflects light irrespective of incident lighting and viewpoint. The shading component is closely related to imaging geometry, which includes brightness variation on a curved surface, shadows, and inter-reflections. As a matter of fact, typical intrinsic image techniques only extract reflectance as an individual characteristic; other imaging factors are still confounded in the shading map. Another problem related to the separation of intrinsic terms is photometric stereo [9]. It can be thought as a subset of the original intrinsic images problem: recovering surface normals of an object given multiple images taken from a fixed view point under different lighting conditions. Although normal, albedo, and incident lighting can be simultaneously recovered using photometric stereo techniques, only surface normal is emphasized and the resulting albedo is quite imprecise.

As intrinsic image decomposition only separates reflectance from other complex imaging factors and photometric stereo just focuses on the recovery of surface normal, we need an optimization algorithm which accurately solves for both characteristics. The main motivation of our work is to develop a simple, yet powerful optimization framework for accurate normal and reflectance estimation from an image sequence of the same object under varying lighting conditions (Fig. 1). Our energy function consists of three terms. To begin with, the Lambertian image formation model is formulated as inter-image constraints and shadows are properly handled with approximate estimation of shadow field. Then, we relax the over-smooth constraints on shading in conventional Retinex theory [10] to fit for natural situations in which shadows and occluding boundaries are commonly presented. Finally, the surface continuity constraint is added to account for special cases. To solve this optimization problem efficiently, we propose an iterative algorithm by interleaving the minimization with respect to different variables. By minimizing this novel energy function, we are able to produce accurate recovery of surface normals and reflectance. Experimental results

T. Luo and J. Shen are with the Beijing Key Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: shenjianbing@bit.edu.cn).

X. Li is with the State Key Laboratory of Transient Optics and Photonics, Center for Optical Imagery Analysis and Learning, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China.

show that our method yields better normal estimations to the state-of-the-art uncalibrated photometric stereo algorithms and generates better albedo maps in the meantime. As for intrinsic image decomposition, results on the real and synthetic scenes show that the proposed approach outperforms previous methods using multiple images as input. Our source code will be publicly available online.[1]

In summary, we make the following major contributions in this paper.

1) A novel energy optimization framework is proposed to accurately estimate normal and reflectance from an input image sequence. Our energy function integrates photometric and Retinex constraints, and we relax the over-smooth assumption in conventional Retinex methods according to shadow masks and normal variations.

2) The energy minimization problem is solved by a two-step iterative optimization algorithm. The accuracy of normal and reflectance values are refined jointly during the iterations.

3) We detect shadows in the shading map of the input images to eliminate the influence by reflectance variation. Shadow information is then used to assist the recovery of normal and reflectance.

The rest of this paper is organized as follows. Section II reviews the related work on normal and reflectance recovery. Section III provides the basic formulation for intrinsic image decomposition and photometric stereo. Section IV describes our energy optimization framework in detail. Section V outlines our iterative optimization algorithm. Section VI gives the experimental results, and Section VII concludes this paper.

II. RELATED WORK

Our approach is built upon intrinsic image and photometric stereo techniques. These two problems are long studied topics in the computer vision community. Though lots of theories and algorithms have been developed during the last decades, how to recover precise normal and reflectance in real images still remains an open challenge.

A. Intrinsic Images

Since intrinsic images decomposition is apparently ill-posed, assumptions about the imaging environment should be made to imply additional constraints. Barrow and Tenenbaum [1] showed that edges and regions provided useful heuristics. A lot of early work focused on classifying edges as caused either by reflectance or shading variations. Retinex theory [10] states that strong edges are caused by reflectance variation in a Mondrian world where reflectance is piecewise constant and shading varies smoothly. Horn [12] applied this theory to recover albedo. Funt *et al.* [13] classified edges by thresholding gradients in the chromaticity map of a color image. Sinha and Adelson [14] discriminated edges by junction analysis in a world of painted polyhedra. Finlayson *et al.* [15] used edges in an invariant

chromaticity image by entropy minimization, and then removed shadows by edge analysis in that chromaticity image [16]. Tappen *et al.* [17] relied on a classifier trained on image derivatives to distinguish shading and reflectance gradients. Later work explored local [2], [18]–[21] and global [5], [22], [23] assumptions on reflectance, and color was usually treated as a significant cue. Shen *et al.* [22] assumed that pixels with similar textures should have the same reflectance. Bousseau *et al.* [2] assumed that local reflectance variations lied on a plain of the RGB color space. Jiang *et al.* [20] interpreted luminance variations as illumination changes by correlation (CORR) analysis on local mean luminance and its amplitude. Shen *et al.* [18], [19] adopted a premise that reflectance of a pixel could be represented by the weighted summation of its neighboring pixels. Shen and Yeo [23] introduced the global sparsity on reflectance, which is based on the findings in [24] that natural image contains only a small set of dominant colors. Garces *et al.* [5] clustered image regions of the same reflectance in the CIELab color space. Serra *et al.* [21] presented two kinds of surface descriptors: the color-name descriptor as a fuzzy set representation and the color-shade descriptor by a ridge analysis of the color distributions.

For quantitatively evaluation and comparison, Grosse *et al.* [11] provided a ground truth data set for intrinsic images, which is referred as the Massachusetts Institute of Technology (MIT), intrinsic images data set. With the help of this data set, statistical approaches [25]–[27] were introduced. Gehler *et al.* [25] formulated the global sparsity prior on reflectance into a latent variable random field model by performing a MAP estimation. Barron and Malik [26] integrated multiple priors on albedo and shape into a mixed model and found the most likely explanation for single grayscale images [26] and later took natural illumination into consideration [27]. For image editing tasks using intrinsic images, Bousseau *et al.* [2] provided three kinds of interactive brushes for users to indicate the constant-reflectance, constant-illumination, and fixed-illumination regions. Similar user interaction strategy was integrated in the optimization framework of [18]. While user intervention provides precise constraints for intrinsic image decomposition, local regions often require careful and tedious manipulations.

Multiple input images of the same scene provide additional information for intrinsic images decomposition. Weiss [28] formulated a maximum-likelihood (ML) estimation problem from image sequences under a fixed viewpoint and varying lighting directions. Similar inputs with biased illumination were exploited in [29], from which they built a minimization problem with spatial and temporal constraints. Hauagge *et al.* [53] have solved the intrinsic images problem using the registered photos of the same scene as an image stack. Lee *et al.* [30] acquired image and depth videos using a Kinect camera, and Laffont *et al.* [6] used online photo collections in which viewpoints and illuminations vary greatly. Both methods constructed a coarse 3-D geometry of the scene in a preprocessing step. We believe that image sequences can be further explored for accurate normal and reflectance recovery.
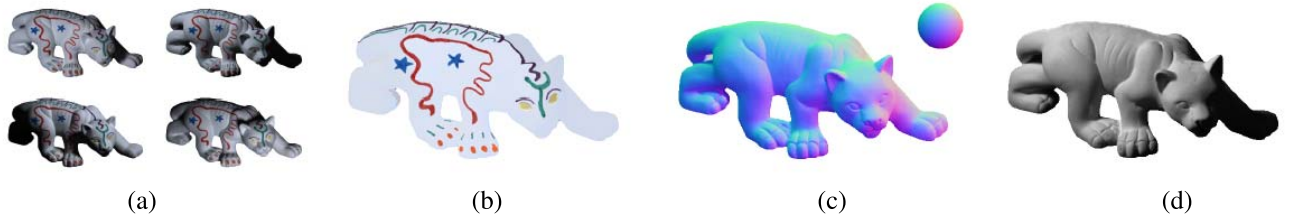
Fig. 1.   Proposed algorithm takes multiple images of the same object under varying lighting conditions as input, and produces accurate normal and reflectance maps. (a) Four input images of the panther object from the MIT intrinsic images data set [11]. (b) Recovered reflectance. (c) Recovered normal. (d) Shading map corresponding to the top-right image of (a).

B. Photometric Stereo

Our work is more related to the uncalibrated photometric stereo problem when incident lighting directions are unknown. Hayakawa [31] recovered surface normals of a Lambertian scene up to 3×3 linear ambiguity using singular value decomposition (SVD). By enforcing surface integrability, Yuille and Snow [32] further reduced the 3×3 linear ambiguity to the generalized bas-relief (GBR) ambiguity [33] with only three indeterminate parameters.

Recent work on uncalibrated photometric stereo mainly focused on resolving the GBR ambiguity. The property of surface bidirectional reflectance function (BRDF) embeds information to solve the GBR ambiguity. Non-Lambertian phenomena (interreflections [34] and specularities [35], [36]) provide useful cues for disambiguation. More complicated reflectance models were exploited such as the Torrance and Sparrow model [37]. Alldrin et al. [38] proved that GBR transformations increased the entropy of albedo distribution. Tan et al. [39] exploited the geometry of reflectance symmetries. They relied on isotropy and reciprocity constraints generated by highlight areas of input images for disambiguation. Wu and Tan [40] proposed an algorithm to automatically calibrate photometric stereo by restoring a special low-rank structure of estimated BRDF slices. Shi et al. [41] proposed a self-calibrating photometric stereo method using color and intensity profiles. Favaro and Papadhimitri [42] showed that locations of maximal diffuse brightness on a curved surface encoded sufficient information for GBR disambiguation. At those maxima locations of Lambertian diffuse reflectance (LDR), the surface normals are in the reverse direction of the incident lighting. A closed-from solution can be formulated by maximizing the corresponding inner products of surface normals and lighting directions. Compared with conventional uncalibrated photometric stereo, these BRDF-based autocalibration algorithms worked well for real-world objects.

Shadows cause important issues for accurate normal estimation since they are not considered in most of the reflectance models. Previous work either used shadowless images as input or treated shadows as outliers. Several work detected shadows by thresholding in spatial [31], [43] or temporal domain [44]–[46]. Attached shadows were often eliminated by an iterative process with updated lighting directions [32], [47], but they could be used as a useful heuristic for normal estimation [48]. Chandraker et al. [34] combined shading and shadow constraints into a fast graph cuts framework. Wu et al. [49]

eliminated shadows and specularities by recovering a low-rank intensity matrix.

III. BASIC FORMULATIONS

A. Intrinsic Images

Intrinsic image decomposition aims at separating an image into shading and reflectance components. The intensity of a pixel is the product of its shading and reflectance. The decomposition is formulated as

$$I_p = s_p R_p \qquad (1)$$

where I_p denotes the intensity value of pixel p and s_p and R_p represent its shading and reflectance values, respectively. R_p is a 3-D vector and s_p is a scalar if the color of the incident lighting is white. Directly solve for s_p and R_p using (1) is apparently underconstrained, because for an RGB pixel four unknowns need to be solved while only three constraints are available. Even using multiple images as input, constraints from different color channels do not provide too much new information. Thus, a least-square solution cannot generate satisfying decompositions.

B. Photometric Stereo

Lambertian image formation model is the most commonly used reflectance model for photometric stereo. The appearance i of an ideally diffusing surface point observed under a directional lighting is described by the Lambertian law

$$i = \max\left(\rho \mathbf{n} \cdot \mathbf{l}, 0\right) \qquad (2)$$

where ρ is the diffuse albedo, $\mathbf{l} \in \mathbb{R}^3$ denotes the incident lighting, and $\mathbf{n} \in \mathbb{R}^3$ represents the surface normal. For notation convenience, $\mathbf{n}$ is out-pointing and $\mathbf{l}$ is in the reverse direction of the incident lighting.

Using multiple images of the same object under varying lighting conditions, a set of linear equations are generated by (2). The photometric stereo problem is then formulated into the matrix form

$$\mathbf{I} = \mathbf{NL} \qquad (3)$$

where $\mathbf{I} \in \mathbb{R}^{P \times F}$ is the set of input images, in which $\mathbf{I}_{ij}$ stores the ith pixel under the jth lighting condition. $\mathbf{N} \in \mathbb{R}^{P \times 3}$ encodes normal (direction of the row vectors) and albedo (length of the row vectors) at each pixel. $\mathbf{L} \in \mathbb{R}^{3 \times F}$ is the lighting matrix, and each column of $\mathbf{L}$ denotes an incident lighting direction. If $\mathbf{L}$ is known a priori, a least square

solution of $\mathbf{N}$ is good enough for ideal Lambertian scene. Otherwise, the recovered $\mathbf{N}$ and $\mathbf{L}$ might suffer from a 3×3 linear ambiguity [31]. This ambiguity arises because different combinations of shapes, albedo distributions and lighting directions can produce identical appearance.

IV. Our Optimization Framework

In this section, we present an optimization model which incorporates uncalibrated photometric stereo constraints and assumptions used by recent intrinsic image techniques. Similar to most of the related work, we make the following assumptions.

1) The relative position of the camera and the object is fixed so that surface points are perfectly aligned across images.
2) For each image, we assume the object is illuminated by a point light source at infinity and the color of light is white. That is to say, the incident lighting direction is constant for the whole scene.

To start with, we make some basic notations used throughout this paper. Suppose that we are given F input images $I_1, I_2, \ldots, I_F$ of the same object under different incident lightings $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_F$. P denotes the number of surface points, or the number of corresponding image pixels. We use the terms albedo and reflectance interchangeably and its value is denoted by ρ. Surface normal is represented as $\mathbf{n} \in \mathbb{R}^3$. We use subscript letter p for pixel index and f for image index, e.g., $I_{p,f}$ denotes intensity value of the pth pixel in the fth image.

A. Problem Formulation

Inter-image constraints used by photometric stereo and intra-image constraints used by intrinsic images can be integrated into the following energy optimization framework:

$$\underset{\mathbf{n}, \rho}{\arg \min} E = E^M + E^R + E^S \qquad (4)$$

where E^M is defined by the image formation model, E^R encodes our extended Retinex assumptions, and E^S corresponds to shape constraints that encourage an integrable surface. In the following sections, we present the definitions and details of these energy terms.

B. Model Constraints

Shadows in real images can be generally classified into two categories: attached and cast shadows. Attached shadows refer to dark surface areas facing away from the light source. For those surface points, the Lambertian model treats their intensities as 0. Shadows can also occur when the incident lightings are blocked before they reach a certain surface, even though the surface normals may form a sharp angle with the lighting direction. Those dark areas are referred to as cast shadows where direct lightings from the light source cannot reach due to the obstruction by other objects or other parts of a nonconvex object. Dealing with cast shadows is rather difficult because they are nonlocal to surface geometry.

Shadows can be viewed as a scaling factor multiplied to the intensity of the unshadowed point, the scaling factors comprise shadow field. We use C to denote shadow field of an image. If point p is not in shadow C_p is equal to 1, and if p is in the umbra area C_p is equal to a constant x. Otherwise, the value of C_p is between x and 1. With the definition of shadow field, a shadowed image I can be represented as the product of the shadow field and a shadowless image I'

$$I = I' * C \qquad (5)$$

where $*$ denotes per-pixel multiplication and C is the shadow field of the shadowed image I.

Shadow detection in a single-image is difficult and often relies on user interaction. Now that we have multiple images of the same object under different lighting conditions, the shadows vary across images while surface albedo remains constant. This special property on inputs provides a hint for shadow removal. Weiss [28] proved that median operation in the image derivative domain can eliminate varying illumination effects. We then use the result by ML estimation to identify shadow regions of the input image. We first use Weiss' method to recover a pseudoreflectance image $\hat{R}$. Note $\hat{R}$ is free of shadow but still contains lots of shading effect when the illumination is biased [29].

We use $C_{p,f}$ to represent the shadow field of pth pixel in the fth input image, and $m_{p,f}$ denotes the corresponding shadow mask. According to (5), $C_{p,f} = I_{p,f}/\hat{R}_p$. Although the shadow field is not very accurate, we could still easily detect shadow areas from it. Attached shadows and cast shadows can be identified in the same way. A binary shadow mask is then estimated as

$$m_{p,f} = \begin{cases} 1 & \text{if } C_{p,f} < \theta \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

where θ is a threshold value classifying shadow and nonshadow area. Because the shadow field is free of reflectance variations, θ is automatically determined as the first valley point in the histogram of the shadow field. Simple thresholding in the original image has been used to detect shadows in [31], but it is unreliable in the presence of albedo variations. When the surface is painted with multiple colors, the darkness of a pixel might be caused by either shadows or the brightness of surface color. Our shadow detection approach is robust to this situation, while naive thresholding leads to lots of misclassifications. As shown in Fig. 2, the hoofs of the deer are not in shadow region, but they are misclassified as shadow effect by simple thresholding because of their intrinsic black color (reflectance).

In natural images, pixels in shadow regions may have an intensity value greater than 0 because of interreflection, which is also ignored by the original Lambertian model. Here, we add an ambient term a to each surface point. The ambient lighting intensity is assumed to be constant in the whole scene. In practice, we assign the minimum intensity value of an input image as the ambient lighting intensity.

For a surface point p under different lighting conditions, the albedo remains constant and the intensity differences are
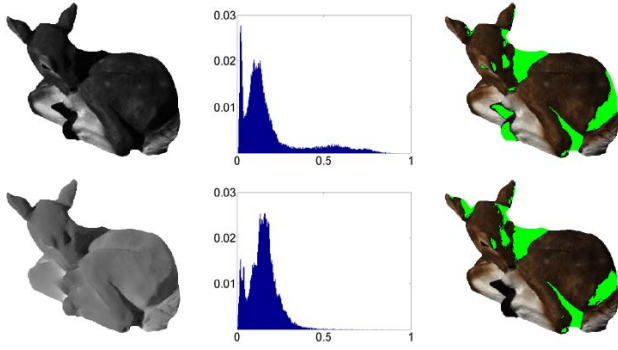
Fig. 2. Comparison on shadow detection. Top row from left to right: grayscale intensity map of the input image, histogram of the intensity map, and detected shadow mask using simple thresholding [31] directly in the intensity domain. Bottom row from left to right: shadow field of the input image, histogram of the shadow field, and detected shadow mask by our approach.
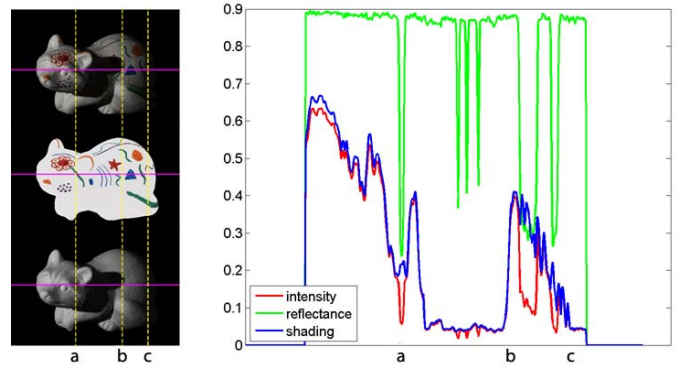

Fig. 3. Boundary analysis. Left: from top to bottom shows original image of the raccoon object [11], the ground truth reflectance, and the ground truth shading. Right: plots of intensity, reflectance, and shading along the scan row of the object (marked as purple lines in the left image). (a) Reflectance edge. (b) Cast shadow boundary. (c) On an attached shadow boundary.

caused by incident light variation according to the image formation model. These constraints belong to inter-image constraints. Based on the Lambertian law, the estimated shadow field, and the ambient lighting assumption, we define the following energy term for the pth pixel across images:

$$E_p^M = \sum_f C_{p,f} \| I_{p,f} - \rho_p \mathbf{n}_p \cdot \mathbf{l}_f - a_f \|^2 \quad (7)$$

where C is the shadow field and a denotes the ambient lighting. This energy term corresponds to a weighted least square solution. It is reasonable to use the shadow field as weighting factors; since shadows violate the Lambertian law and only intensity values in the lit area are trustable. Higher weights are given to pixels in the lit area, and lower weights are given to those in shadow.

Each surface point has a set of inter-image constraints. By summing over all surface points, we get the total image formation constraints

$$E^M = \sum_p E_p^M. \quad (8)$$

C. Extended Retinex Constraints

The Retinex theory [10] has been widely used for intrinsic image decomposition. Assuming adjacent pixels have similar shading and reflectance values in most cases, the closed-form solution to Retinex in the log domain has been proposed in [30] and [50]. If there are large chromaticity differences between the adjacent pixels that indicate the albedo changes, the smoothness constraints on reflectance can be relaxed. A weighting factor was then defined by thresholding the chromaticity differences of neighboring pixels. The intrinsic image decomposition is solved by minimizing the following quadratic Retinex energy function:

$$\arg\min f = \sum_{(p,q)\in\mathcal{N}} \left[(s_p - s_q)^2 + w_{pq}(\rho_p - \rho_q)^2 \right] \quad (9)$$

$$w_{pq} = \begin{cases} 0 & \text{if } \|\hat{c}_p - \hat{c}_q\|_2 > t \\ 100 & \text{otherwise} \end{cases} \quad (10)$$

where s_p is the log shading value of pixel p, ρ_p is the log reflectance value, $\mathcal{N}$ is the set of all adjacent pixel pairs, $\hat{c}_p$ denotes the normalized color value as the chromaticity,

t is a given threshold, and w_{pq} represents a weight controlling the reflectance similarity of neighboring pixels. Substituting $i_p = \rho_p + s_p$ into (9), we get a quadratic form on ρ_p or s_p.

The Retinex theory assumes smooth shading and piece-wise constant reflectance. The parameter w_{pq} defined in (10) accounts for abrupt changes of reflectance, but shading still remains smooth over the scene. Obviously, object boundaries and cast shadows violate the smoothness assumption of shading. This can lead to significant errors in the estimated shading image at surface and shadow boundaries. It also indicates that shadow and normal information provides a useful cue to reduce the over-smoothness of shading.

Fig. 3 shows the boundary situations in a real image. At surface point (a), the quick change of intensity is caused by reflectance variation. However, at shadow boundaries (b) and (c), reflectance is approximately constant so that step changes of intensity are caused by shading variations. A significant change of shading at both cast and attached shadow boundaries will lead to significant change of intensity in the observed image. This kind of derivative is misclassified as caused by reflectance variation by typical Retinex algorithms. Now that we have detected shadow areas by the ML estimator, shadow boundaries can be directly identified. The smoothness of shading is discouraged at these boundary positions by a weighting factor defined by the shadow mask.

We describe the extended Retinex assumptions as intra-image constraints because they are related to appearances of neighboring pixels in the same image. Use extra information obtained by our framework, we extend typical Retinex constraints to fit for more general imageries. For the fth input image, the energy term is defined as

$$E_f^R = \sum_p \sum_{q \in N_p} \left[w_{pq}^r (\rho_p - \rho_q)^2 + w_{f,pq}^s \left(\frac{I_{p,f}}{\rho_p} - \frac{I_{q,f}}{\rho_q} \right)^2 \right] \quad (11)$$

$$w_{pq}^r = \begin{cases} 0 & \text{if } \|\hat{c}_p - \hat{c}_q\|_2 > t^r \\ 10 & \text{otherwise} \end{cases}$$

$$w_{f,pq}^s = \begin{cases} 0 & \text{if } m_{p,f} + m_{q,f} = 1 \\ 1 & \text{otherwise} \end{cases} \quad (12)$$

where N_p is the 3×3 neighbors of pixel p and w^r and w^s are the weighting factors allowing step changes of reflectance and shading, respectively. $\hat{c}_p$ is the chromaticity of pixel p, t^r is a given threshold for reflectance edge classification, and m is the shadow mask estimated in the previous section. Using multiple images at a fixed viewpoint as input, we calculate the chromaticity map by normalizing RGB values of the same pixel across all images.

If normal information is known *a priori*, the smoothness of shading can be further reduced. When two pixels in the lit area have similar normals, their shading values should be similar according to the Lambertian law. For a smooth curved surface in the absence of shadows, shading varies smoothly. Shading discontinuity occurs at surface boundaries where normal directions change quickly. Therefore, we can discourage smoothness of shading at those positions by modifying the definition of w^s as

$$w^s_{f,pq} = 0, \quad \text{if } (1 - \mathbf{n}_p \cdot \mathbf{n}_q) > t^s \tag{13}$$

where t^s is a threshold for shading changes. Note that (12) and (13) are complementary for the definition of w^s. Attached shadow boundaries may be identified by normal variation, while cast shadow boundaries often lie on a smooth surface where normal variations are smooth. In this situation, only (12) gives a reasonable w^s. However, when two adjacent surfaces with distinct normal directions both lie in the lit area, only (13) takes effect.

Each input image provides a set of intra-image constraints. By summing over all F input images, we get the total extended Retinex constraints

$$E^R = \sum_f E^R_f. \tag{14}$$

D. Object Shape Constraints

Some surface points might stay in the shadow area for all the lighting directions. Their normals cannot be precisely updated by the model constraints term E^M. However, if surface integrability is enforced, we can interpolate normal values of the neighboring surface points. Rather than involving complex partial derivative operations as in [32] and [51], we add a local smooth term on the estimated normals

$$E^S = \sum_{p \in D} \sum_{q \in N_p} \| \mathbf{n}_p - \mathbf{n}_q \|^2 \tag{15}$$

where D denotes the set of pixels always in shadow. This local smooth constraint on normals works when a pixel always stays shadowed and thus compensates other energy terms.

V. Optimization

Now, we come back to discuss how to minimize the total objective function E in (4). Since this optimization problem contains multiple variables with large scale constraints, direct solution by gradient descent is time-consuming. Also, the extended Retinex term E^S is not jointly convex in ρ because of the division operation. Typical optimization techniques might easily fall into local minima. To overcome these difficulties,

we propose a two-step iterative refinement algorithm for the optimization of E, and use the LDR maxima method of [42] to generate proper initial lighting directions. The original objective function E is divided into two convex subfunctions by interleaving the minimization with respect to $\mathbf{n}$ and ρ.

Our algorithm consists of the following steps.

1) *Preprocessing:* As described in Section IV, for each input image I_f, we first estimate its shadow field C_f using the ML estimator and detect its shadow mask m_f. Some pixels always stay in the lit area across all images, so we mark these pixels as robust pixels. Robust pixels are used to refine lighting directions during iteration.

2) *Initialization:* We locate the maxima positions of LDR, and recover the initial lighting direction for each image by the LDR method presented in [42]. In the meantime, initial normal $\mathbf{n}_0$ and reflectance ρ_0 values are also calculated.

3) *Update Reflectance:* We transform the extended Retinex term E^R into the log domain to construct a convex quadratic energy function. Division in the intensity domain corresponds to subtraction in the log domain

$$E^R_f = \sum_p \sum_{q \in N_p} \left\{ w^r_{pq} (\rho_p - \rho_q)^2 + w^s_{f,pq} \right.$$
$$\left. \times [(I_{p,f} - \rho_p) - (I_{q,f} - \rho_q)]^2 \right\}. \tag{16}$$

Weights are still calculated by (12). Using the result of previous iteration ρ as the initial value, we minimize E^R to refine the estimated reflectance. Normals by previous iteration can be used to update w^s by (13).

4) *Update Normals and Lighting Directions:* Substituting $\mathbf{l}$ and ρ into (7) yields a weighted least square problem with a closed-form solution. Then, we calculate new lighting directions using updated normals of robust pixels detected in step 1.

5) *Iteration:* Steps 3 and 4 are iterated until the estimated normal and reflectance cannot be further improved.

The Appendix presents our detailed derivations for steps 3 and 4. The two subproblems both have closed-form solutions. In our experiments, we find that several iterations are generally enough to generate stable results.

VI. Experimental Results

In this section, we test the performance of the proposed method using synthetic images and real-world photographs. Since our approach jointly recovers surface normals and reflectance, we compare with the state-of-the-art algorithms for both uncalibrated photometric stereo and intrinsic image decomposition.

A. Comparison With Photometric Stereo Approaches

Using the LDR initial lighting directions, our algorithm directly recovers surface normals and albedo by iterative energy minimization. Unlike our method, the LDR photometric stereo method is based on matrix decomposition. Starting from the SVD normal and lighting matrix, LDR searches for the optimal GBR transformation matrix. Another difference
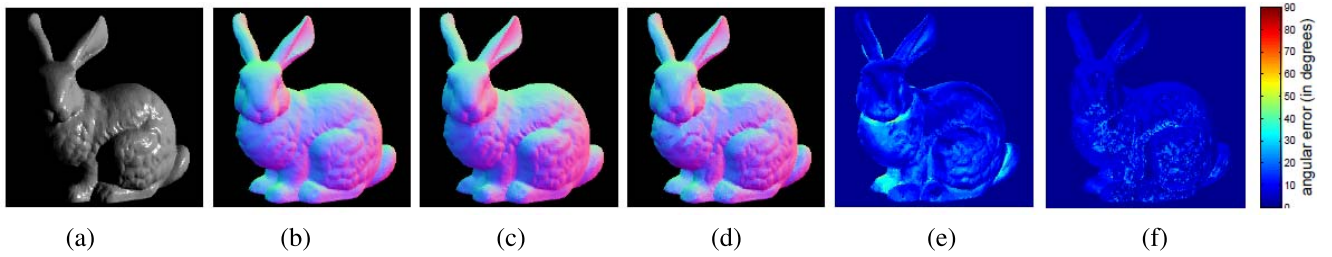
Fig. 4.   Normal estimations from 40 images of the synthetic bunny data set. (a) One of the input images. (b) Ground truth normal map. (c) and (d) Recovered normal maps by LDR + RPCA and our method. (e) and (f) Angular error maps corresponding to (c) and (d).

lies in how to handle shadows. Favaro and Papadhimitri [42] performed a preprocessing step based on the RPCA method proposed in [49] to remove shadows, while we use shading information to approximate shadow effect in the energy formulation. We do not explicitly remove shadows from input images. We first compare our approach with the LDR + RPCA method on a set of synthetic Bunny images. These images are generated in [52]. Lighting directions are randomly selected. Specularities are synthesized using the Cook-Torrance reflectance model and cast/attached shadows are generated under each lighting condition. This test set consists of 40 input images. Fig. 4 shows normal estimations of our approach and LDR + RPCA, along with the ground truth normal map. For quantitative evaluation, we calculate the angular error between the normal estimations and the ground truth. Angular error maps [Fig. 4(e) and (f)] clearly show that our normal is superior to that of LDR + RPCA, especially in cast shadow areas. The mean angular error (MAE) of LDR + RPCA is 11.75° and our estimation only has MAE of 7.39°.

As discussed in Section II, Allrin *et al.* [38] resolved the GBR ambiguity by minimizing entropy (ME) of the albedo distribution. Favaro and Papadhimitri [42] used LDR to get rid of this ambiguity. These approaches showed the state-of-the-art results, so we compare our experimental results with their recovered normals and reflectance. In this experiment, we use RPCA as a preprocessing step for the calibrated method, ME and LDR, but not for our method. As shown in Fig. 5, we first compare ME + RPCA, LDR + RPCA, and our approach on three real object data sets: red fish by Alldrin[2]; cat and owl by Goldman and Seitz.[3] These data sets are commonly used to test uncalibrated photometric stereo algorithms, and we mainly select objects with spatial varying reflectance since our algorithm relies on color information. Fig. 5 shows normal maps recovered by calibrated photometric stereo, ME + RPCA, LDR + RPCA, and our approach. We can see that our approach yields better results to other methods in the absence of shadows. Cast shadows cause problems for the other three methods even when RPCA is used, while our weighted least square energy term handles shadows well. For the cat object, previous methods treat cast shadow effects as the results caused by normal variation. Our approach successively reveals the coherence at these shadow regions (Fig. 5, last column).

[2]http://neilalldrin.com/research
[3]http://www.cs.washington.edu/education/courses/csep576/05wi/projects/project3/project3.htm
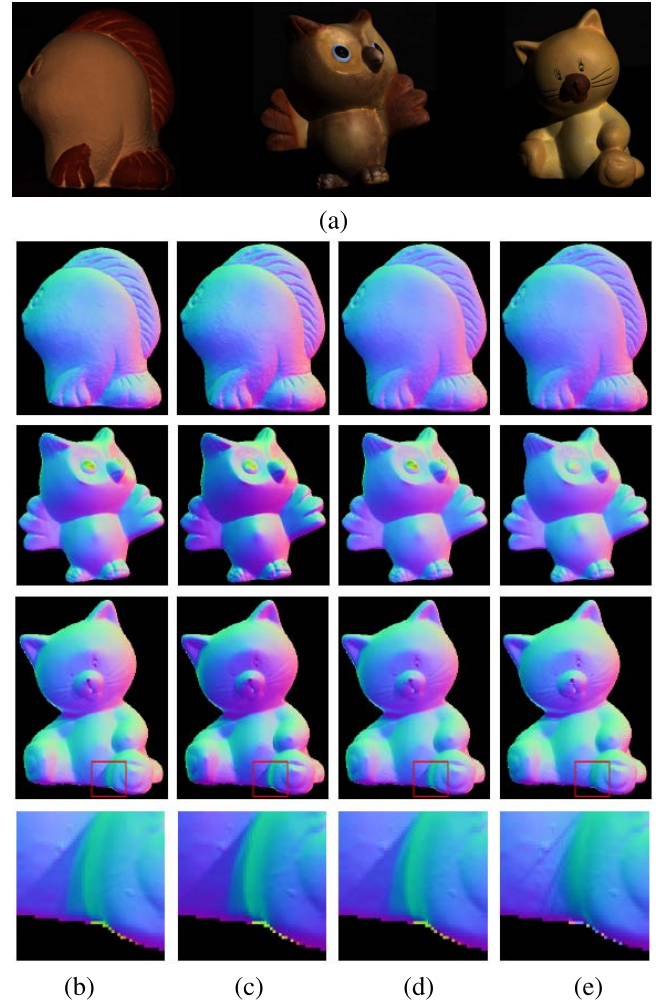

Fig. 5.   Comparison results of normals. (a) Test objects: red fish, owl, and cat. (b) Result by calibrated photometric stereo. (c) Result by ME + RPCA [38]. (d) Result by LDR + RPCA [42]. (e) Our approach. The last row shows close-ups of the cat object.

Many photometric stereo algorithms only work for isolated objects on a black background with simplified lighting and surface conditions. Here, we present the experimental results on a complex real-world model. Laffont *et al.* [6] built a diffuse model of the St. Basil cathedral which contained complex geometric details and spatially varying albedo. Fixed-viewpoint images are rendered under realistic lighting conditions. We choose 14 daylight images under autumn or
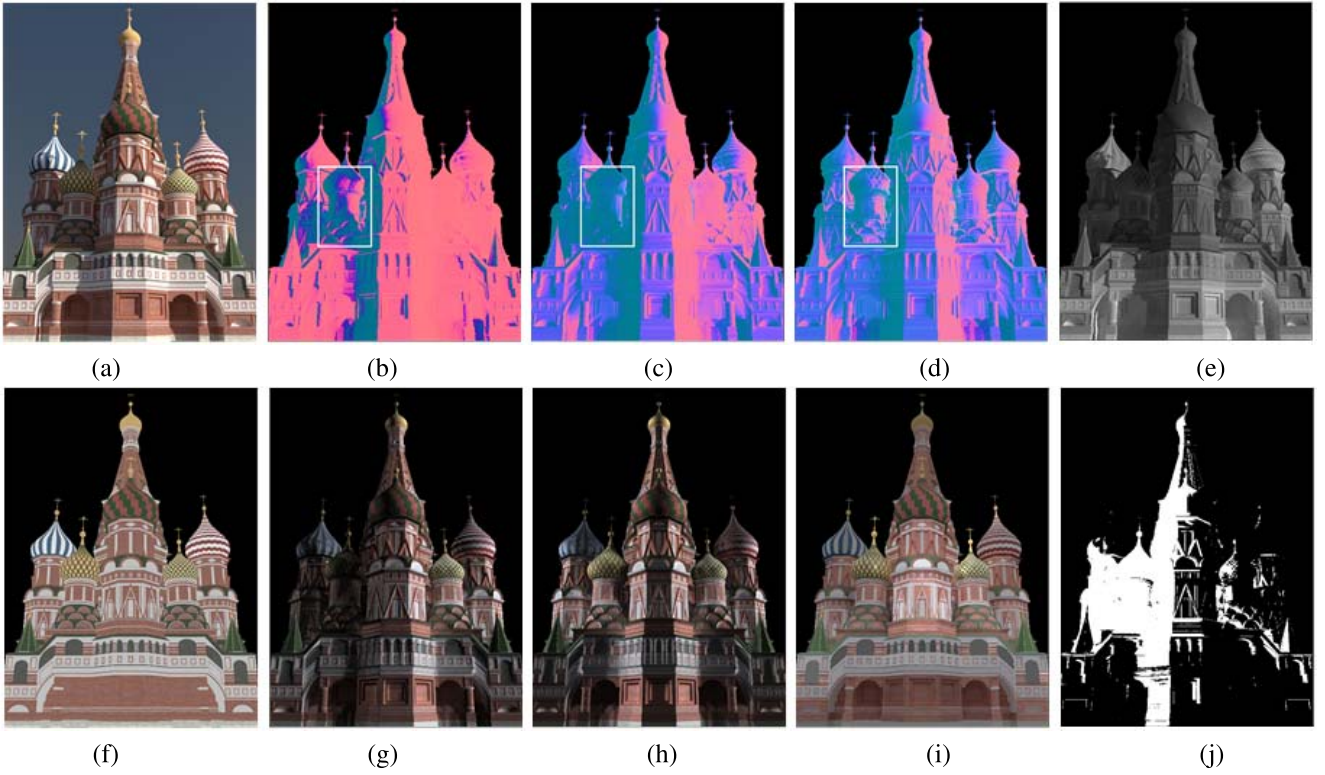
Fig. 6. Results for the St. Basil cathedral scene. (a) Input. (b) ME normal. (c) LDR normal. (d) Our normal. (e) Our shading. (f) GT albedo. (g) ME albedo. (h) LDR albedo. (i) Our albedo. (j) Our shadow. Note in (b)–(d), regions marked by the white box clearly show that our normal estimation is more realistic than those of ME and LDR.

summer illuminations to test the proposed approach. The RPCA shadow removal method produces a poor performance on this set, so we directly apply ME and LDR on the original input images. Fig. 6 shows normal and albedo maps estimated by ME, LDR, and our approach. We also present our estimated shading image and the shadow mask in Fig. 6. From Fig. 6(b)–(d), we can observe that our normal reveals more geometric details than those of ME and LDR. Our approach successfully handles most of the shadows, while ME and LDR could not produce realistic normal estimation in these occluded regions. Cast shadows are also solved in our method and our normal map is more continuous in local regions where shading and reflectance both change quickly. As shown in Fig. 6 (e)–(g), heavy shading effects are retained in the albedo maps of ME and LDR, while our approach successfully recovers a coherent albedo map.

Resolving the GBR ambiguity by entropy minimization or via diffusion maxima works well for a wide variety of natural objects, but they suffer from a common degenerative case in which the object has plane surfaces. Under this situation, the distribution of albedo is a sum of several delta functions after all kinds of GBR transformations, and normals on a plane are in the same direction so that there exists no maxima points of LDR. We provide results for this kind of degenerative case in Fig. 7. The box object from the MIT data set [11] has several plane surfaces with small raised areas. Apparently, the two front surfaces facing left and right should have distinct normal directions, and parallel planes should have the same
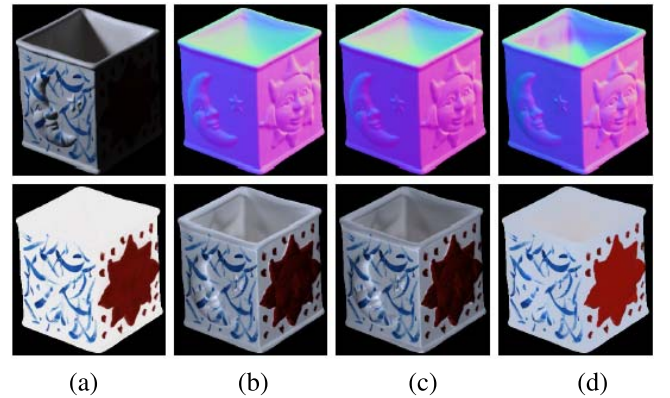

Fig. 7. Illustration of a degenerative case for ME [38] and LDR [42] methods. (a) Top: one of the input images and bottom: ground truth reflectance. (b) Normal and reflectance maps by ME [38]. (c) Results by LDR [42]. (d) Our results. Note that parallel planes should have the same normal directions, and adjacent surface with different normals should be clearly distinguished. Our algorithm successively reveals these properties in the resulting normal map.

surface normals. From Fig. 7, we can see that our approach yields consistent normals in parallel surfaces, and adjacent surfaces with different normals are clearly distinguished in our result [Fig. 7(d)]. While the other two methods [38], [42] show a bad performance on this object. Surface boundaries are successively eliminated, and shading variations are almost totally removed in our recovered reflectance map [Fig. 7(d)]. For the objects with a few discrete planes, Zhou and Tan [54] proposed the ring-light photometric stereo approach.
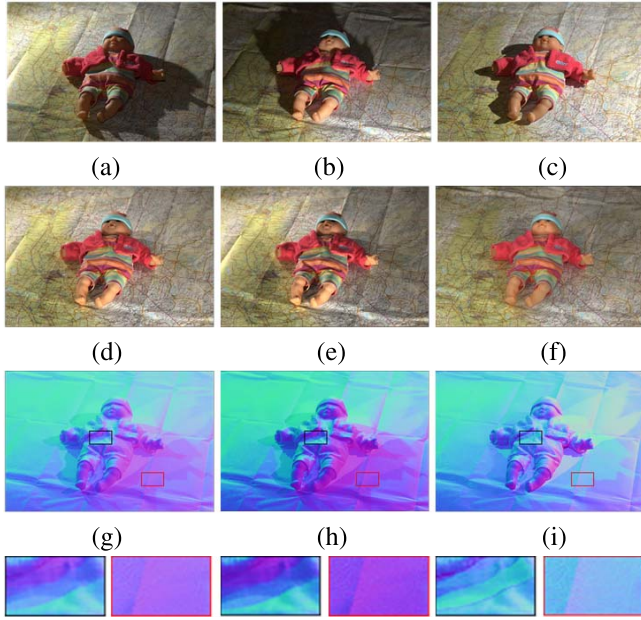
Fig. 8.   Comparison results for the captured real scene (ME [38], LDR [42], and our method). (a)–(c) Three input images (out of eight). (d)–(f) ME, LDR, and our albedo. The last row is the close-ups for above normal results [black and red rectangles in (g)–(i)]. When using colors to represent normal directions, the background should be green regardless of cast shadows.

They used constraints in lighting configurations, while our method mostly relied on reflectance properties. For this kind of case, our approach mainly benefits from the iterative refinement of lighting directions. The input images of our algorithm can be captured under randomly selected lighting directions such as the synthetic Bunny experiment in Fig. 4, while the ring-light photometric stereo method requires lights to lie on a view centered cone.

In addition to the above isolated objects and synthetic scenes, we further compare these approaches on a captured real scene. These photographs are generously provided in [6]. Input images contain complicated textures (the map in the background), and fine-scale geometries (shape of the doll), as well as the heavy cast shadows. Compared with ME and LDR methods, both our recovery reflectance map and our estimated normal map are more precise especially in the cast shadow areas, as shown in Fig. 8. The background almost lies on the same plane, and thus normals of the background should be similar in Fig. 8. When using colors to represent normal directions, the background should be green [Fig. 8(i)].

B. Comparison With Intrinsic Image Algorithms

Intrinsic image algorithms extract reflectance and shading maps from an input image or image sequences, while our approach directly produces surface normals and reflectance as output. We could render a shading map from scratch using the recovered lighting direction, object normals, and shadow field. However, our shadow field is just an approximation so that shadows cannot be truly revealed. Furthermore, intrinsic image decomposition does not assume the Lambertian image formation model. Therefore, the shading maps are still

calculated by pixel-wise division using (1). Existing intrinsic image algorithms [2], [23], [28], [50] also obtain the shading maps in this way.

We test the proposed approach on the MIT intrinsic images data set [11], which has become the benchmark for evaluating intrinsic image algorithms. We compare with the conventional CR algorithm [13], Weiss's multiimage algorithm (W) [28], and a combination algorithm integrating both methods Weiss's and color Retinex (WCR) algorithm, [11]. These three algorithms adopted the same input as ours, and they were reported to achieve good performances on the MIT data set. The corresponding implementations are generously provided online in [11]. Although single-image-based algorithms have made much progress in recent years, we mainly focus on comparison with multiple-image approaches since they generally yield better results than single-image ones. We use the 10 fixed-viewpoint images for each object as input for all these algorithms, and perform quantitative evaluation on the original image.

To perform quantitative comparisons, several error metrics have been proposed in [11] and [20]. We use mean squared error (mse) and CORR to measure global errors. Local errors are evaluated by local mse (Lmse) and absolute Lmse (aLmse). For mse, Lmse, and aLmse, a smaller score means less error, while greater CORR coefficient indicates better results for the CORR metric. To ensure a consistent measurement, we slightly modified the CORR metric to non-CORR (NCORR) coefficient. These metrics are defined as

$$\mathrm{MSE}(I_{\mathrm{gt}}, I) = \arg\min_{\alpha} \| I_{\mathrm{gt}} - \alpha I \|$$

$$\mathrm{NCORR}(I_{\mathrm{gt}}, I) = 1 - \frac{E[(I_{\mathrm{gt}} - \mu_{\mathrm{gt}})(I - \mu_I)]}{\sigma_{\mathrm{gt}}\sigma_I}$$

$$\mathrm{LMSE}(I_{\mathrm{gt}}, I) = \sum_{w} \arg\min_{\alpha} \| I_{\mathrm{gt}}^w - \alpha I^w \|$$

$$\mathrm{aLMSE}(I_{\mathrm{gt}, I}) = \sum_{w} \arg\min_{\alpha} \| (I_{\mathrm{gt}}^w - \mu_{\mathrm{gt}}^w) - \alpha(I^w - \mu_I^w) \|$$

where I_{gt} means the ground truth vector, I represents the estimated vector, μ is the mean value, σ denotes the variance, and the superscript w stands for a sliding window. Note that α allows the estimation to be any scalar multiple of the ground truth.

Fig. 9 shows average evaluation results on the MIT data set. For a fair comparison, the two thresholds t_r and t_s of our algorithm are determined exactly as in [11] by cross-validation. We separately compute errors for reflectance and shading because local metrics on shading and reflectance are not linearly dependent. Final errors are calculated by taking the mean scores for each object of the data set. As shown in Fig. 9, both reflectance and shading estimations produced by our approach are better than the rest three algorithms both locally and globally, since both inter- and intra-image constraints are incorporated into our energy function. Because lighting condition varies across images, inter-image constraints contribute more to the estimation of shading component. Our method tries to balance the relation between the reflectance estimation and the shading recovery. We focus more on the coherence of shading. All these metrics allow a scalar α
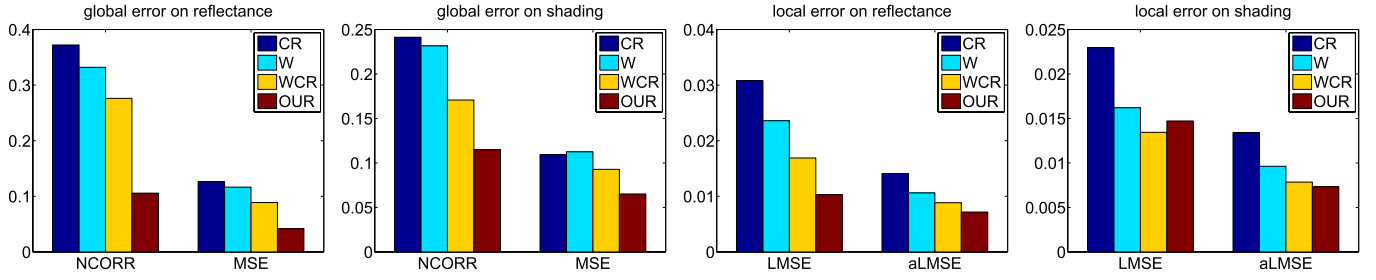
Fig. 9.   Quantitative evaluation on the MIT data set. For each metric, errors are calculated by averaging scores on all test objects. CR denotes results by the conventional color Retinex (CR) algorithm [13], W stands for results by Weiss's method [28], and WCR indicates results by the combinational approach in [11].
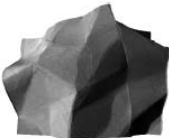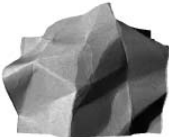
Fig. 10.   Results on three representative objects from the MIT data set. GT indicates ground truth, W stands for results by Weiss's method [28], and WCR indicates results by a combinational approach [11]. The NCORR coefficient below each result is obtained by averaging shading and reflectance errors.

multiple to the ground truth. For local metrics, α differs among local windows, thus spatial coherence is ignored to some extent. Jiang *et al.* [20] and Serra *et al.* [21] also argue that Lmse is not qualitatively good and sometimes biased, and the global metrics are more consistent with human vision. When considering reflectance and shading together, our method outperforms others on the quantitative test.

Reflectance recovered by our method is closely similar to the ground truth, and shading effects including shadows are totally removed in the reflectance maps (last row in Fig. 10).

Only a few reflectance residuals can be observed in our shading images, and geometry details are properly retained. In Fig. 10, we show results on three representative objects from the MIT data set [11]. Each object represents one category of the data set: raccoon for painted objects with smooth shape and regionally constant reflectance, paper1 for print-material objects with locally varying reflectance, and turtle for animal models with highly mixed reflectance and shading variations. We also present NCORR coefficients for every result in Fig. 10. Weiss's algorithm removes shadows in the
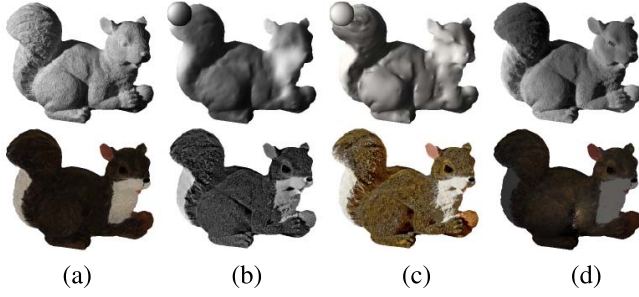
Fig. 11. Comparison results for the squirrel object. (a) Ground truth shading and reflectance maps. (b) Results by SAIFS [26]. (c) Results by SIRFS [27]. (d) Our results. Our algorithm preserves fine-scale properties in the shading image.

reflectance maps, but shading effect is obviously retained, especially for raccoon and paper1 (third row in Fig. 10). The shading map produced by Weiss's method is somewhat too dark but is totally free of reflectance residuals. The combined method improves reflectance by Weiss's algorithm to some extent, but also introduces residuals to the shading image (fourth row in Fig. 10). Our approach yields the best results for all the three objects. Reflectance recovered by our method is closely similar to the ground truth, and shading effects including shadows are totally removed in the reflectance maps (last row in Fig. 10). Only a few reflectance residuals can be observed in our shading images, and geometry details are properly retained.

Barron and Malik [26] built a statistical model for intrinsic images in the MIT data set. By performing a ML optimization, they recovered shape, albedo, and illumination from a single grayscale image (SAIFS). Later, they integrated color constancy [27] to their model and then estimated shape, illumination, and reflectance from a single color image (SIRFS). Their approach was built upon techniques of shape from shading, shape from contour, and intrinsic images. Therefore, they could only estimate coarse shading and reflectance maps. In contrast, our framework produces more accurate and fine-scale results of normal, reflectance, and shading maps with multiple images as input. Fig. 11 shows the comparison results with their approaches. Results of the other two methods are directly copied from the supplementary material of [27] with unknown laboratory illumination. It is clear that single-image-based methods cannot handle shadows well. Cast shadows of the squirrel are obviously misclassified into their reflectance maps by SAIFS [26] and SIRFS [27]. In addition, only coarse shape is inferred from their over-smoothed shading map. The squirrel object is an extremely challenging case because its surface consists of highly mixed shading and reflectance effects, and this kind of fine-scale property is hard to preserve by typical intrinsic image algorithms. Despite difficulty, the high variance of local shading is successively revealed in our results [Fig. 11(d)].

VII. Conclusion

In this paper, we have proposed a novel energy optimization framework to accurately recover normal and reflectance of an object from multiple images taken from a fixed viewpoint under different lighting conditions. To build the energy function, we add partial shadow information and ambient lighting to the Lambertian model, and relax the over-smooth Retinex assumption on shading constraint. Experimental results have demonstrated that our approach yields better normal and reflectance maps than previous uncalibrated photometric stereo algorithms. Furthermore, it is free of typical degenerative situations such as objects with plane surfaces. Both qualitative and quantitative results on the MIT data set suggest that our method produces more accurate shading and reflectance components than the state-of-the-art intrinsic image algorithms.

Appendix
Implementation Details

In this section, we present our exact formulations with more detailed explanations for completeness, and these equations have already been described and discussed in Sections IV and V.

Update Normals

With partial shadow information, we formulated the model constraints as a weighted least square energy in (7). Here, we derive the closed-form solution to this minimization problem. Let us first review the original definition of the model constraints at pixel p

$$E_p^M = \sum_f C_{p,f} \|I_{p,f} - \rho_p \mathbf{n}_p \cdot \mathbf{l}_f - a_f\|^2$$
$$= \rho_p \sum_f \left\| C_{p,f}^{1/2} \left(\frac{I_{p,f} - a_f}{\rho_p} - \mathbf{n}_p \cdot \mathbf{l}_f \right) \right\|^2.$$

During each iteration, $\mathbf{l}$ and ρ are treated as known parameters when updating normals. At every surface point, we modify the minimization problem to the following matrix form:

$$\arg\min_{\mathbf{n}_p} E_p^M = \arg\min_{\mathbf{n}_p} \rho_p \left\| \mathbf{C}_p^{1/2} (\hat{\mathbf{I}}_p - \mathbf{L}\mathbf{n}_p) \right\|^2$$

where $\mathbf{C}_p$ is a $F \times F$ diagonal matrix with $C_{p,f}$ as the diagonal elements, $\mathbf{L}$ is the $F \times 3$ lighting matrix, $\mathbf{n}_p$ is the 3×1 normal vector, and $\hat{\mathbf{I}}_p$ is a $F \times 1$ vector defined as

$$\hat{\mathbf{I}}_p(f) = \frac{I_{p,f} - a_f}{\rho_p}.$$

By differentiating E_p^M with respect to $\mathbf{n}_p$ and equating to zero, we then obtain the closed-form solution to this weighted least square problem as

$$\frac{\partial E_p^M}{\partial \mathbf{n}_p} = \frac{\partial (\hat{\mathbf{I}}_p - \mathbf{L}\mathbf{n}_p)^T \mathbf{C}_p (\hat{\mathbf{I}}_p - \mathbf{L}\mathbf{n}_p)}{\partial \mathbf{n}_p}$$
$$= 2\mathbf{L}^T \mathbf{C}_p \mathbf{L}\mathbf{n}_p - 2\mathbf{L}^T \mathbf{C}_p \hat{\mathbf{I}}_p.$$

Therefore, the final closed-form solution is

$$\hat{\mathbf{n}}_p = \left(\mathbf{L}^T \mathbf{C}_p \mathbf{L} \right)^{-1} \mathbf{L}^T \mathbf{C}_p \hat{\mathbf{I}}_p$$
$$= \left((\mathbf{C}_p^{1/2} \mathbf{L})^T (\mathbf{C}_p^{1/2} \mathbf{L}) \right)^{-1} (\mathbf{C}_p^{1/2} \mathbf{L})^T (\mathbf{C}_p^{1/2} \hat{\mathbf{I}}_p)$$
$$= (\mathbf{C}_p^{1/2} \mathbf{L})^+ (\mathbf{C}_p^{1/2} \hat{\mathbf{I}}_p)$$

where $+$ denotes the Moore–Penrose pseudoinverse. Since these inter-image constraints are independent among the surface points, they are solved individually in parallel.

Update Reflectance

As described in Section V, the extended Retinex energy term is translated into the log domain, which results in a quadratic function of ρ as

$$E_f^R = \sum_p \sum_{q \in N_p} \left\{ w_{pq}^r (\rho_p - \rho_q)^2 + w_{f,pq}^s [(I_{p,f} - \rho_p) - (I_{q,f} - \rho_q)]^2 \right\}.$$

To minimize this quadratic object function of variables $\rho_1, \rho_2, \ldots, \rho_P$, we set the partial derivatives of ρ_p to 0, and then solve the resulting linear equations. For clarity, we organize E_f^R as

$$E_f^R = \sum_p \sum_{q \in N_p} \left\{ w_{pq}^r \Delta_\rho^2 + w_{f,pq}^s (\Delta_I - \Delta_\rho)^2 \right\}$$
$$= \sum_p \sum_{q \in N_p} \left\{ (w_{pq}^r + w_{f,pq}^s) \Delta_\rho^2 - 2 w_{f,pq}^s \Delta_I \Delta_\rho + c \right\}$$

where

$$\Delta_I = I_{p,f} - I_{q,f}$$
$$\Delta_\rho = \rho_p - \rho_q$$
$$c = w_{f,pq}^s \Delta_I^2.$$

Note that c is a constant irrespective to ρ and thus makes no contribution to locating the minimum point. The partial derivative with respect to ρ_p in E_f^R is computed as

$$\frac{\partial E_f^R}{\partial \rho_p} = 4 \sum_{q \in N_p} \left\{ (w_{pq}^r + w_{f,pq}^s) \Delta_\rho - w_{f,pq}^s \Delta_I \right\}.$$

We then substitute this result into (14) using the summation rule of partial derivative operation

$$\frac{\partial E_R}{\partial \rho_p} = \frac{\partial \sum_f E_f^R}{\partial \rho_p} = \sum_f \frac{\partial E_f^R}{\partial \rho_p}$$
$$= 4 \sum_f \sum_{q \in N_p} \left\{ (w_{pq}^r + w_{f,pq}^s) \Delta_\rho - w_{f,pq}^s \Delta_I \right\}.$$

After making $\partial E_R / \partial \rho_p = 0$ for $p = 1, 2, \ldots, P$, we then get a set of linear equations with the standard form

$$\mathbf{Ax} = \mathbf{b}$$

where $x_p = \rho_p$ for $p = 1, 2, \ldots, P$, and

$$\mathbf{A}_{pq} = \begin{cases} \sum_f \sum_{k \in N_p} (w_{pk}^r + w_{f,pk}^s) & \text{if } q = p \\ -\sum_f (w_{pq}^r + w_{f,pq}^s) & \text{if } q \in N_p \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$$\mathbf{b}_p = \sum_f \sum_{q \in N_p} w_{f,pq}^s (I_{p,f} - I_{q,f})$$

these linear equations can be solved by standard methods such as Gauss–Seidel iteration, lower-upper decomposition, the conjugate gradient algorithm, and so on. In our MATLAB implementation, we directly use the matrix left division operator for simplicity.

REFERENCES

[1] H. G. Barrow and J. M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Proc. Int. Conf. Comput. Vis. Syst.*, 1978.

[2] A. Bousseau, S. Paris, and F. Durand, "User-assisted intrinsic images," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 130:1–130:10, 2009.

[3] S. Beigpour and J. van de Weijer, "Object recoloring based on intrinsic image estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 327–334.

[4] R. Carroll, R. Ramamoorthi, and M. Agrawala, "Illumination decomposition for material recoloring with consistent interreflections," *ACM Trans. Graph.*, vol. 30, no. 3, pp. 43:1–43:10, 2011.

[5] E. Garces, A. Munoz, J. Lopez-Moreno, and D. Gutierrez, "Intrinsic images by clustering," *Comput. Graph. Forum*, vol. 31, no. 4, pp. 1415–1424, 2012.

[6] P.-Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis, "Coherent intrinsic images from photo collections," *ACM Trans. Graph.*, vol. 31, no. 6, p. 202, 2012.

[7] S. Du and R. K. Ward, "Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 9, pp. 1165–1175, Sep. 2010.

[8] P.-Y. Laffont, A. Bousseau, and G. Drettakis, "Rich intrinsic image decomposition of outdoor scenes from multiple views," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 2, pp. 210–224, Feb. 2013.

[9] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Opt. Eng.*, vol. 19, no. 1, pp. 191–139, 1980.

[10] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[11] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2009, pp. 2335–2342.

[12] B. K. P. Horn, *Robot Vision*. Cambridge, MA, USA: MIT Press, 1986.

[13] B. V. Funt, M. S. Drew, and M. Brockington, "Recovering shading from color images," in *Proc. 2nd Eur. Conf. Comput. Vis. (ECCV)*, 1992, pp. 124–132.

[14] P. Sinha and E. Adelson, "Recovering reflectance and illumination in a world of painted polyhedra," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, May 1993, pp. 156–163.

[15] G. D. Finlayson, M. S. Drew, and C. Lu, "Intrinsic images by entropy minimization," in *Proc. 8th Eur. Conf. Comput. Vis. (ECCV)*, 2004, pp. 582–595.

[16] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 59–68, Jan. 2006.

[17] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 9, pp. 1459–1472, Sep. 2005.

[18] J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3481–3487.

[19] J. Shen, X. Yang, X. Li, and Y. Jia, "Intrinsic image decomposition using optimization and user scribbles," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 425–436, Apr. 2013.

[20] X. Jiang, A. J. Schofield, and J. L. Wyatt, "Correlation-based intrinsic image extraction from a single image," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 58–71.

[21] M. Serra, O. Penacchio, R. Benavente, and M. Vanrell, "Names and shades of color for intrinsic image estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 278–285.

[22] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–7.

[23] L. Shen and C. Yeo, "Intrinsic images decomposition using a local and global sparse representation of reflectance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 697–704.

[24] I. Omer and M. Werman, "Color lines: Image specific color representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jul. 2004, pp. 946–953.

[25] P. V. Gehler, C. Rother, M. Kiefel, L. Zhang, and B. Schölkopf, "Recovering intrinsic images with a global sparsity prior on reflectance," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2011, pp. 765–773.

[26] J. T. Barron and J. Malik, "Shape, albedo, and illumination from a single image of an unknown object," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 334–341.

[27] J. T. Barron and J. Malik, "Color constancy, intrinsic images, and shape estimation," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 57–70.

[28] Y. Weiss, "Deriving intrinsic images from image sequences," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Jul. 2001, pp. 68–75.

[29] Y. Matsushita, S. Lin, S. B. Kang, and H.-Y. Shum, "Estimating intrinsic images from image sequences with biased illumination," in *Proc. 8th Eur. Conf. Comput. Vis. (ECCV)*, 2004, pp. 274–286.

[30] K. J. Lee *et al.*, "Estimation of intrinsic image sequences from image+depth video," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 327–340.

[31] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *J. Opt. Soc. Amer. A, Opt., Image Sci., Vis.*, vol. 11, no. 11, pp. 3079–3089, 1994.

[32] A. Yuille and D. Snow, "Shape and albedo from multiple images using integrability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 1997, pp. 158–164.

[33] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas-relief ambiguity," *Int. J. Comput. Vis.*, vol. 35, no. 1, pp. 33–44, 1999.

[34] M. K. Chandraker, F. Kahl, and D. J. Kriegman, "Reflections on the generalized bas-relief ambiguity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2005, pp. 788–795.

[35] O. Drbohlav and R. Šára, "Specularities reduce ambiguity of uncalibrated photometric stereo," in *Proc. 7th Eur. Conf. Comput. Vis. (ECCV)*, 2002, pp. 46–60.

[36] O. Drbohlav and M. Chaniler, "Can two specular pixels calibrate photometric stereo?" in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Oct. 2005, pp. 1850–1857.

[37] A. S. Georghiades, "Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2003, pp. 816–823.

[38] N. G. Alldrin, S. P. Mallick, and D. J. Kriegman, "Resolving the generalized bas-relief ambiguity by entropy minimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–7.

[39] P. Tan, L. Quan, and T. Zickler, "The geometry of reflectance symmetries," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2506–2520, Dec. 2011.

[40] Z. Wu and P. Tan, "Calibrating photometric stereo by holistic reflectance symmetry analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1498–1505.

[41] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan, "Self-calibrating photometric stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 1118–1125.

[42] P. Favaro and T. Papadhimitri, "A closed-form solution to uncalibrated photometric stereo via diffuse maxima," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 821–828.

[43] Z. Liu, K. Huang, and T. Tan, "Cast shadow removal in a hierarchical manner using MRF," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 1, pp. 56–66, Jan. 2012.

[44] D. Xu, J. Liu, X. Li, Z. Liu, and X. Tang, "Insignificant shadow detection for video segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 1058–1064, Aug. 2005.

[45] Y. Wang, "Real-time moving vehicle detection with cast shadow removal in video based on conditional random field," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 437–441, Mar. 2009.

[46] A. Abrams, C. Hawley, and R. Pless, "Heliometric stereo: Shape from sun position," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 357–370.

[47] K. Sunkavalli, T. Zickler, and H. Pfister, "Visibility subspaces: Uncalibrated photometric stereo with shadows," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 251–264.

[48] T. Okabe, I. Sato, and Y. Sato, "Attached shadow coding: Estimating surface normals from shadows under unknown reflectance and lighting conditions," in *Proc. IEEE 12th Int. Conf. Comput. Vis. (ICCV)*, Oct. 2009, pp. 1693–1700.

[49] 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. 10th Asian Conf. Comput. Vis. (ACCV)*, 2010, pp. 703–717.

[50] Q. Zhao, P. Tan, Q. Dai, L. Shen, E. Wu, and S. Lin, "A closed-form solution to retinex with nonlocal texture constraints," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1437–1444, Jul. 2012.

[51] T.-P. Wu, J. Sun, C.-K. Tang, and H.-Y. Shum, "Interactive normal reconstruction from a single image," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 119:1–119:9, 2008.

[52] S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, "Robust photometric stereo using sparse regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 318–325.

[53] D. Hauagge, S. Wehrwein, K. Bala, and N. Snavely, "Photometric ambient occlusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 2515–2522.

[54] Z. Zhou and P. Tan, "Ring-light photometric stereo," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 265–279.
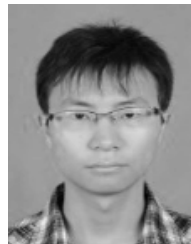
Tao Luo is working toward the M.S. degree with the School of Computer Science, Beijing Institute of Technology, Beijing, China.

His research interests include high-quality intrinsic images and photometric stereo algorithms.

Jianbing Shen (M'11–SM'12) is a Full Professor with the School of Computer Science, Beijing Institute of Technology, Beijing, China. He has authored more than 40 refereed papers in journals and conference proceedings. His research interests include computer vision and computer graphics.

Xuelong Li (M'02–SM'07–F'12) is a Full Professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China.