INTRINSIC DECOMPOSITION FROM A SINGLE RGB-D IMAGE WITH SPARSE AND NON-LOCAL PRIORS

Yujie Wang [1], *Kun Li* [1*], *Jingyu Yang* [1], *Xinchen Ye* [2]

[1] Tianjin University, Tianjin 300350, China
[2]Dalian University of Technology, Dalian 116600, China

ABSTRACT

This paper proposes a new intrinsic image decomposition method that decomposes a single RGB-D image into reflectance and shading components. We observe and verify that, a shading image mainly contains smooth regions separated by curves, and its gradient distribution is sparse. We therefore use ℓ_1-norm to model the direct irradiance component—the main sub-component extracted from shading component. Moreover, a non-local prior weighted by a bilateral kernel on a larger neighborhood is designed to fully exploit structural correlation in the reflectance component to improve the decomposition performance. The model is solved by the alternating direction method under the augmented Lagrangian multiplier (ADM-ALM) framework. Experimental results on both synthetic and real datasets demonstrate that the proposed method yields better results and enjoys lower complexity compared with two state-of-the-art methods.

Index Terms— Intrinsic decomposition, RGB-D image, sparse representation, non-local correlation.

1. INTRODUCTION

Intrinsic image decomposition is an essential task for many applications in computer vision and graphics[1, 2, 3]. It aims at decomposing an image into several specific components that encode material and lighting characteristics of the scene described in the image. The most common decomposition is to separate the image into reflectance and shading components [4]. The reflectance image represents the reflectance property of object materials under invariant light, while the shading image describes all effects introduced by light. Thus a successful decomposition would be beneficial to many applications, such as image relighting, image editing, and shape from shading.

Image formation is a complex process involved with many factors and the process is irreversible. The same image might be obtained from different configurations of scenes and lighting conditions. Therefore, extracting the reflectance and shading components from an image is an ill-posed problem. To overcome this, many priors and assumptions, including color Ritinex [5], texture cues [6] and color sparsity [7], are proposed. Despite consistent effort on this problem, decomposition results are still unsatisfactory for generic images due to severe its ill-poseness. With the commoditization of depth cameras, several methods use depth information for better intrinsic decompose of RGB images [3, 8, 9]. Most works impose a smoothness prior, usually by the total energy (equivalently ℓ_2 norm) of finite differences to make the problem well-posed [8, 9]. However, such ℓ_2-norm based priors tend to penalize large differences, and are easily affected by noise and outliers. The reflectance image of a natural scene is approximately piecewise constant and the shading image varies smoothly except for boundaries of different surfaces.

In this paper, we propose a new intrinsic image decomposition method for a single RGB-D image with sparse and non-local priors. Inspired by [8], we decompose the RGB-D image into reflectance and shading components, where the shading part is further separated into three sub-components, *i.e.*, direct irradiance, other irradiance and illumination color. We use ℓ_1 norm to model the reflectance component and the direct irradiance component (the dominant sub-component extracted from shading component) based on our observation that finite differences of the reflectance and shading images are sparse. We also design a non-local prior weighted by a bilateral kernel on a large neighborhood to fully exploit structural correlation in the reflectance image, and the weight between two neighboring pixels is computed based on the patches similarity centering on the pair-wise pixels from the chromaticity information. We adopt the alternating direction method under the augmented Lagrangian multiplier (ADM-ALM) framework [10] to solve the model. The proposed method is evaluated on synthetic dataset and real-world datasets. Results demonstrate that the proposed method yields better intrinsic decomposition and enjoys low complexity compared with the state-of-the-art methods. Our code will be publicly available on the project website.

The main contributions of this work are summarized as:

- A sparsity constraint is imposed on the reflectance component and the direct irradiance sub-component in the

*Corresponding author: lik@tju.edu.cn. This work was supported in part by the National Natural Science Foundation of China (Grant 61571322 and 61372084).

978-1-5090-6067-2/17/$31.00

shading component. Based on our observation and analysis, the local finite differences of reflectance and shading images present Laplacian distributions, and can be well-modeled by using ℓ_1-norm as a regularizer.

- A non-local prior that considers non-local similarity weighted by a bilateral kernel is designed to fully exploit structural correlation in the reflectance component, remedying the short-sighted local correlation in former methods.

- The proposed method is quite fast. For 1024×436 images, the proposed method takes 140 seconds on average.

2. RELATED WORK

Decomposing an image into a reflectance image and a shading image is first introduced by Barrow and Tenenbaum [4]. Then, there are many scientists putting effort into developing models to produce better decomposition results. The Ritinex theory [5] is an early and successful model assuming that reflectance changes bring in large gradient variations while the small gradient variations are caused by shading changes. Although the Ritinex theory works well in a Mondrian world, it is not suitable for all the real-world images. Later models and theories utilize additional information, such as multiple images [11], chromaticity gradient [1] and user-interaction [12]. Dai *et al.* propose a co-intrinsic method based on ℓ_0-norm which simultaneously decomposes a pair of images with the same foreground [2]. Besides, some algorithms aim at performing intrinsic decomposition on video sequences[13, 14].

Recently, with the development and commoditization of depth sensors, such as Kinect, it is convenient to simultaneously capture an RGB image and a depth image. Therefore, some methods use depth cues to construct their models[3, 8, 9]. Shi *et al.* [9] propose an intrinsic decomposition method for RGB-D videos. Barron and Malik [3] propose a joint estimation for shape, illumination and reflectance, but this method is time-consuming (about 3 hours for a 1024×436 image). Chen and Koltun [8] propose a simple model for intrinsic decomposition of RGB-D images. Jeon *et al.* [15] improve the decomposition quality by handling textures in the intrinsic image decomposition. However, most above works use ℓ_2-norm regularization that is sensitive to noise and outliers.

In this paper, we propose a new intrinsic image decomposition model for a single RGB-D image with sparse and non-local priors, based on the observation that the reflectance and shading images are sparse on the finite-difference domain. The model is efficiently solved by the alternating direction method under the augmented Lagrangian multiplier framework.

3. MOTIVATION

Sparse Prior on Reflectance and Shading: Many methods use quadratic smoothness prior assuming the Gaussian distribution of pairwise differences for the reflectance and shading components. However, for most natural scenes, reflectance and shading are piecewise smooth. Therefore, the differences between pixels are sparse, and should be modeled by a heavy-tailed distribution, rather than being dense and modeled by a rapidly vanishing Gaussian distribution. This is verified in Fig.1. The normalized histograms of differences in the neighborhood are presented, together with fitted Laplacian distributions and Gaussian distributions. For the reflectance component, we connect every pixel with 4 random pixels in a 9×9 window. For the shading component, we choose 12 nearest points in a six-dimensional space for each point. As shown in the figure, the Laplacian distribution fits the histogram significantly better than the Gaussian distribution, suggesting the use of sparsity-promoting ℓ_1-norm in the reflectance and shading components.

Non-local Prior on Reflectance: Usually, pixels with similar color in the RGB image tend to belong to the same material and thus have the same reflectance value. Traditional methods [8, 15] use local correlation and pixel chromatic difference to judge the pair-wise pixel similarity. This short-sighted local judgement cannot provide enough information to preserve global structures in reflectance component, which leads to the information cross-leakage between different components. Exploiting non-local correlation has received tremendous success in various fields such as image denoising [16] and depth recovery [17]. This motivates us to use a patch-based non-local prior weighted by a bilateral kernel to better regularize the reflectance component.

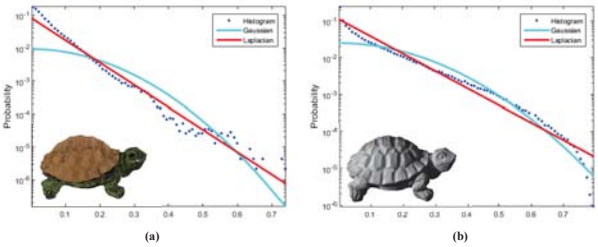

Fig. 1. Normalized histograms and the associated fitted Laplacian and Gaussian distributions for pairwise errors of (a) reflectance and (b) shading for MIT-Berkeley Intrinsic Images dataset [3].

4. THE PROPOSED METHOD

Let I be the input RGB image. Our task is to decompose I into a reflectance image R and a shading image S. Each color channel obey the following multiplicative observation model: $I_p = R_p S_p$ for each pixel p. The shading image S is influenced by several physical factors, including geome-

try, shadows and inter-reflections. Therefore, referred to [8], we further decompose *S* into three components. Specifically, the RGB image *I* is decomposed into four components: a reflectance image *R*, a direct irradiance image *D*, an other irradiance image *O* and an illumination color image *C*. The reflectance image *R* describes how the materials of the object reflect the incoming light. The direct irradiance image *D* represents the irradiance that is modeled by local shading algorithms without considering shadows or inter-reflections. The other irradiance image *O* encodes the effects that *D* does not take into account. The illumination color image *C* represents the color of environment illumination. Finally, the value of each pixel p in *S* is computed according to $S_p = D_p O_p C_p$. For each pixel p, it satisfies $I_p = R_p D_p O_p C_p$, and we transform it into logarithmic domain by taking logarithms on both sides: $i_p = r_p + d_p + o_p + c_p$, which is an additive model for easier mathematical manipulation.

The intrinsic decomposition model is formulated as the minimization of the following energy function:

$$\underset{\mathbf{X}=(\mathbf{R},\mathbf{D},\mathbf{O},\mathbf{C})}{E(\mathbf{X})} = E_{data}(\mathbf{X}) + E_{smooth}(\mathbf{X}), \tag{1}$$

where $E_{data}(\mathbf{X})$ and $E_{smooth}(\mathbf{X})$ are the data term and the smoothness term, respectively. The data term measures the total error of four components compared with the original RGB image, while the smoothness term consists of four regularizers that make the original ill-posed problem well-posed.

The data term is defined as follows:

$$E_{data} = \sum_p \left\| \sqrt{lum(I_p)}(i_p - r_p - \mathbf{1}d_p - \mathbf{1}o_p - c_p) \right\|_2^2, \tag{2}$$

where d_p and o_p are scalar while i_p, r_p, c_p are three-dimensional vectors, $\mathbf{1} = [111]^\top$ is a vector with all elements being one. $lum(I_p)$ is the illumination of I_p and we use $\sqrt{lum(I_p) + \varepsilon}$ to avoid the case in which $lum(I_p)$ equals zero ($\varepsilon = 0.001$).

For the compact representation in algorithm derivation, we define the following matrix/vector form of the variables to reformulate the data term (2):

$$\begin{aligned} \mathbf{W} &= \text{diag}\left(\sqrt{lum(i_1)}, \cdots, \sqrt{lum(i_n)}\right), \\ \mathbf{K} &= [1\ 1\ 1]. \end{aligned} \tag{3}$$

Then, the data term is rewritten as:

$$E_{data} = \left\| \mathbf{W}(\mathbf{I} - \mathbf{R} - \mathbf{D}\mathbf{K} - \mathbf{O}\mathbf{K} - \mathbf{C}) \right\|_2^2. \tag{4}$$

The smoothness term is defined as

$$E_{smooth} = \lambda_R E_R + \lambda_D E_D + \lambda_O E_O + \lambda_{O'} E_{O'} + \lambda_C E_C, \tag{5}$$

where λ_R, λ_D, λ_O, $\lambda_{O'}$ and λ_C are the weights for the smoothness terms: E_R, E_D, E_O, $E_{O'}$ and E_C, respectively. These terms are described below.

Reflectance: The smoothness term E_R of reflectance component is defined as:

$$E_R = \sum_{\{p,q\} \in \mathcal{N}_\mathbf{R}} \sqrt{\alpha_{p,q}} \left\| r_p - r_q \right\|_1, \tag{6}$$

where $\mathcal{N}_R$ represents the set of pairwise correlations defined on random neighboring pixels, and $\alpha_{p,q}$ is pairwise weighting coefficient between pixel p and q.

Recent advances on image modeling show that image processing using non-local correlation achieves significant improvements over previous schemes using local correlation alone [16, 17]. Therefore, we design the coefficient $\alpha_{p,q}$ on a non-local neighborhood with a bilateral kernel. Let $\mathcal{P}_p$ be a patch centered at pixel p. For every pixel p, we first generate K random neighbors $\mathcal{N}(p) = \{\mathcal{P}_{q_i}\}_{i=1}^K$ over the entire image. Then, the coefficient α_{p,q_i} is defined as follows:

$$\alpha_{p,q_i} = \exp\left(\frac{||\mathbf{B}_p \circ (\mathcal{P}_p - \mathcal{P}_{q_i})||_2^2}{\sigma_1^2} \right), \tag{7}$$

where σ_1 controls the decay rate of the exponential function, $\circ$ represents the element-wise multiplication. We use a bilateral filter kernel $\mathbf{B}_p$ to weight the distance of neighboring patches, which has a strong response for pixels of similar intensities to the center pixel p, and hence carries the shape information of local image structures. The kernel $\mathbf{B}_p$ is defined in the extracted patch $\mathcal{P}_p$ as follows:

$$\mathbf{B}_p(p,t) = \exp\left(-\frac{||p-t||_2^2}{\sigma_2^2} \right) \exp\left(-\frac{||I_p - I_t||_2^2}{\sigma_3^2} \right), \tag{8}$$

where t is a neighboring pixel of pixel p in the patch $\mathcal{P}_p$, and σ_2 and σ_3 are parameters of the bilateral kernel to adjust the importance of the spatial distance and intensity difference, respectively.

We define a matrix $\mathbf{Q}_R \in \{\sqrt{\alpha_{p,q}}, -\sqrt{\alpha_{p,q}}\}^{|\mathcal{N}_\mathbf{R}| \times |\mathbf{I}|}$ for concise presentation. Concretely, each row of $\mathbf{Q}_R$ corresponds to a pair in $\mathcal{N}_R$ and each column corresponds to a pixel in $\mathbf{I}$. Each row in $\mathbf{Q}_R$ has only two nonzero entries. For example, assuming the r^{th} row in $\mathbf{Q}_R$ associates with the pair $\{p,q\}$, then the entry linking to the reference pixel p is set at $\sqrt{\alpha_{p,q}}$, while the one linking to the neighboring pixel q is set at $-\sqrt{\alpha_{pq}}$. So, the smoothness term of the reflectance is formulated in the matrix form as:

$$E_R = \left\| \mathbf{Q}_R \mathbf{R} \right\|_1. \tag{9}$$

Direct irradiance: The direct irradiance component represents the "virtual" illumination in which every points would have received from the light source directly without other objects' contributions. Hence, two points in the scene have similar direct irradiance when they have similar positions and similar normals. According to our piecewise smooth assumption, the smoothness of the direct irradiance is formulated as:

$$E_D = \sum_{\{p,q\} \in \mathcal{N}_D} \left\| d_p - d_q \right\|_1, \tag{10}$$

where $\mathcal{N}_D$ denotes the set of the pairwise connections. We connect every point with k nearest points in the six-dimensional space which consists of the 3D position and the surface normal of the point. The surface normal of p is computed from the depth values of p and nearby points.

Similar to the smoothness term of reflectance, we define a differential matrix $\mathbf{Q}_D \in \{1,-1\}^{|\mathcal{N}_D|\times|\mathbf{I}|}$. Every row of $\mathbf{Q}_D$ corresponds to a pair in $\mathcal{N}_D$ and each column corresponds to a point in $\mathbf{I}$. There are only two nonzero values in each row of $\mathbf{Q}_D$. If the r^{th} row in $\mathbf{Q}_D$ associates with the pair $\{p,q\}$, the p^{th} column in this row will be set at 1 and the q^{th} column will be set at -1. Then, we reformulate the smoothness term of the direct irradiance as follows:

$$E_D = \left\|\mathbf{Q}_D\mathbf{D}\right\|_1. \tag{11}$$

Other irradiance: Similar to [8], we use the other irradiance component to describe the objects' contributions including shadow, occlusion and inter-reflectance. Hence, the smoothness prior imposed on this irradiance is that two points have similar values when they have similar positions in the object space. The prior is formulated as follows:

$$E_O = \sum_{\{p,q\}\in\mathcal{N}_O} \left\|o_p - o_q\right\|_2^2. \tag{12}$$

We construct $\mathcal{N}_O$, the set of the pairwise connections in the other irradiance component, by connecting every point with K nearest other points in the three-dimensional object space. For the compact representation, we define a differential matrix $\mathbf{Q}_O \in \{1,-1\}^{|\mathcal{N}_O|\times|\mathbf{I}|}$ like $\mathbf{Q}_D$. Then, the smoothness term of the other irradiance component is rewritten as

$$E_O = \left\|\mathbf{Q}_O\mathbf{O}\right\|_2^2. \tag{13}$$

Besides, a simple constraint on the magnitude of other irradiance is defined as

$$E_{O'} = \sum_p \left\|o_p\right\|_2^2. \tag{14}$$

We introduce an identity matrix $\mathbf{Q}_{O'} \in \mathbb{R}^{|\mathbf{I}|\times|\mathbf{I}|}$ for the compact representation and Eq. (14) is rewritten as

$$E_{O'} = \left\|\mathbf{Q}_{O'}\mathbf{O}\right\|_2^2. \tag{15}$$

Illumination color: Similar to [8], the illumination color component $\mathbf{C}$ is defined as

$$E_C = \sum_{\{p,q\}\in\mathcal{N}_C} \sqrt{\gamma_{p,q}}\left\|c_p - c_q\right\|_2^2, \tag{16}$$

where $\sqrt{\gamma_{p,q}}$ adjusts the influence of the Euclidean distance between 3D positions of p and q, and $\gamma_{p,q}$ is computed as

$$\gamma_{pq} = 1 - \frac{\left\|\overline{p}-\overline{q}\right\|_2}{\max\limits_{\{p,q\}\in\mathcal{N}_\mathbf{c}}\left\|\overline{p}-\overline{q}\right\|_2}, \tag{17}$$

where $\overline{p}$ and $\overline{q}$ represent the 3D positions of p and q, respectively.

We define a matrix $\mathbf{Q}_C \in \{\sqrt{\gamma_{p,q}}, -\sqrt{\gamma_{p,q}}\}^{|\mathcal{N}_C|\times|\mathbf{I}|}$ for concise presentation. The definition of $\mathbf{Q}_C$ and $\mathcal{N}_C$ is similar to the definition of $\mathbf{Q}_R$ and $\mathcal{N}_R$. Then, we rewrite Eq. (16) as

$$E_\mathbf{c} = \left\|\mathbf{Q}_C\mathbf{C}\right\|_2^2. \tag{18}$$

The final energy function has the following compact form with matrix-vector notations:

$$\begin{aligned}\min_{\mathbf{X}=(\mathbf{R},\mathbf{D},\mathbf{O},\mathbf{C})} &\left\|\mathbf{W}(\mathbf{I}-\mathbf{R}-\mathbf{D}\mathbf{K}-\mathbf{O}\mathbf{K}-\mathbf{C})\right\|_2^2 + \lambda_R\left\|\mathbf{Q}_R\mathbf{R}\right\|_1 \\ &+ \lambda_D\left\|\mathbf{Q}_D\mathbf{D}\right\|_1 + \lambda_O\left\|\mathbf{Q}_O\mathbf{O}\right\|_2^2 + \lambda_{O'}\left\|\mathbf{Q}_{O'}\mathbf{O}\right\|_2^2 \\ &+ \lambda_C\left\|\mathbf{Q}_C\mathbf{C}\right\|_2^2.\end{aligned} \tag{19}$$

To solve the problem, we first transform the minimization (19) into the following form with an auxiliary variable $\mathbf{A}$ and $\mathbf{B}$:

$$\begin{aligned}\min_{\mathbf{R},\mathbf{D},\mathbf{O},\mathbf{C},\mathbf{A},\mathbf{B}} &\left\|\mathbf{W}(\mathbf{I}-\mathbf{R}-\mathbf{D}\mathbf{K}-\mathbf{O}\mathbf{K}-\mathbf{C})\right\|_2^2 + \lambda_R\left\|\mathbf{A}\right\|_1 \\ &+ \lambda_D\left\|\mathbf{B}\right\|_1 + \lambda_O\left\|\mathbf{Q}_O\mathbf{O}\right\|_2^2 + \lambda_{O'}\left\|\mathbf{Q}_{O'}\mathbf{O}\right\|_2^2 \\ &+ \lambda_C\left\|\mathbf{Q}_C\mathbf{C}\right\|_2^2, \\ \text{s.t.}\quad &\mathbf{A} = \mathbf{Q}_R\mathbf{R}, \\ &\mathbf{B} = \mathbf{Q}_D\mathbf{D}.\end{aligned} \tag{20}$$

Then, we solve the constrained minimization (20) using the augmented Lagrangian method (ALM) by converting it to iterative minimization of its augmented Lagrangian function:

$$\begin{aligned}&L(\mathbf{A},\mathbf{B},\mathbf{R},\mathbf{D},\mathbf{O},\mathbf{C},\mathbf{Y}_1\mathbf{Y}_2,\mu_1,\mu_2) \\ &\quad= ||\mathbf{W}(\mathbf{I}-\mathbf{A}-\mathbf{D}\mathbf{K}-\mathbf{O}\mathbf{K}-\mathbf{C})||_2^2 \\ &\quad+ \lambda_R||\mathbf{A}||_1 + \lambda_D||\mathbf{B}||_1 + \lambda_N||\mathbf{Q}_O\mathbf{O}||_2^2 \\ &\quad+ \lambda_{O'}||\mathbf{Q}_{O'}\mathbf{O}||_2^2 + \lambda_C||\mathbf{Q}_C\mathbf{C}||_2^2 \\ &\quad+ \langle\mathbf{Y}_1, \mathbf{A}-\mathbf{Q}_R\mathbf{R}\rangle + \frac{\mu_1}{2}||\mathbf{A}-\mathbf{Q}_R\mathbf{R}||_F^2 \\ &\quad+ \langle\mathbf{Y}_2, \mathbf{B}-\mathbf{Q}_D\mathbf{D}\rangle + \frac{\mu_2}{2}||\mathbf{B}-\mathbf{Q}_D\mathbf{D}||_F^2,\end{aligned} \tag{21}$$

where (μ_1, μ_2) are positive constants, ($\mathbf{Y}_1$,$\mathbf{Y}_2$) are Lagrangian multipliers, and $\langle\cdot,\cdot\rangle$ denotes the inner product of two matrices considered as long vectors. We resort to the alternate direction method (ADM) [10] to optimize $\mathbf{A}$, $\mathbf{B}$, $\mathbf{R}$, $\mathbf{D}$, $\mathbf{O}$ and $\mathbf{C}$ separately at each iteration

5. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method on the NYU dataset [3] (Section 5.1) and the MPI-Sintel dataset [18] (Section 5.2), compared with two state-of-the-art methods [8, 15]. To test the robustness of the proposed

Fig. 3. Results on one image from MPI-Sintel dataset: (a) input RGB image and depth map, (b) results of [8], (c) results of [15], and (d) our results.

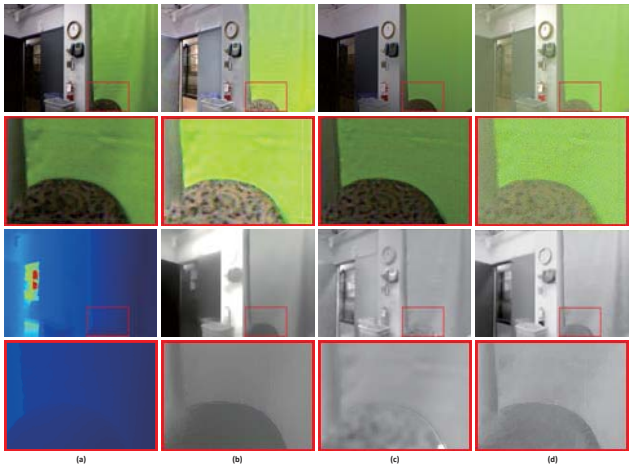

Fig. 2. Results on one image from NYU dataset: (a) input RGB image and depth map, (b) results of [8], (c) results of [15], and (d) our results.

method, we also compare the three methods on the noisy dataset in Section 5.3. The comparison for running time is given in Section 5.4. In our experiments, the weights λ_R, λ_D, λ_O, $\lambda_{O'}$ and λ_C are set at 0.001, 10, 0.1, 10, and 5, respectively. The empirical parameters σ_1, σ_2 and σ_3 are set at 3.05, 1000, and 0.2, respectively. We initialize the reflectance $\mathbf{R}$ by using the log-transformed input image and initialize the other three variables $\mathbf{D}, \mathbf{O}, \mathbf{C}$ with zero matrices.

5.1. Results on NYU Dataset

We test 15 images from the NYU dataset [3], compared with two other methods [8, 15]. Fig.2 gives the results for one image. For the shading component, the result of method [8] is a little over-smoothed. Although the method [15] generates more details, some textures, *e.g.*, on the back of chair, are wrongly preserved. On the contrary, our method provides more accurate geometric information in the shading component. For the reflectance component, the result of method [8] is a little bright and the result of method [15] is a little dark, while our method gives compromised result.

5.2. Results on MPI-Sintel Dataset

Because there are no ground-truth reflectance images and shading images in the NYU dataset, we quantitatively evaluate the performance of the proposed method on 12 images from the MPI-Sintel dataset [18]. As mentioned in [18], although the images of the MPI-Sintel dataset are computer-generated, they have similar statistics to natural images. With the ground truth data in this dataset, we can evaluate our method quantitatively and intuitively. Because there are some images unsuitable for evaluation due to the defects in the provided ground-truth reflectance, we chose 12 images in this experiment. Table 1 gives the quantitative results compared with two state-of-the-art methods [8, 15]. We use two error measures for evaluation: the standard mean-squared error (MSE) and the local mean-squared error (LMSE) [1]. Since the ground-truth is defined up to a scale factor, each image is adjusted by a estimated scaling factor to minimize the error[1]. As shown in the table, the reflectance images and the shading images obtained by the proposed method produce smaller error values, which means the results of our method are more similar to the ground-truth images. Qualitative evaluation for one image is shown in Fig.3.

Table 1. *Quantitative evaluation on MPI-Sintel dataset.*

Method	MSE			LMSE		
	Reflectance	Shading	Average	Reflectance	Shading	Average
[8]	0.0388	0.0277	0.0333	0.0274	0.0195	0.0235
[15]	0.0479	0.0362	0.0421	0.0229	0.0183	0.0206
Ours	**0.0270**	**0.0268**	**0.0269**	**0.0163**	**0.0176**	**0.0169**

Table 2. *Quantitative evaluation on noisy MPI-Sintel dataset.*

Method	MSE			LMSE		
	Reflectance	Shading	Average	Reflectance	Shading	Average
[8]	0.0413	0.0281	0.0347	0.0300	0.0198	0.0249
[15]	0.0497	0.0361	0.0429	0.0262	0.0184	0.0223
Ours	**0.0275**	**0.0269**	**0.0272**	**0.0168**	**0.0178**	**0.0173**

5.3. Results on Noisy MPI-Sintel Datasets

In order to evaluate the ability of the proposed method to suppress noise, we perform a quantitative comparison on the noisy images. The noisy images are obtained by adding slight additive white Gaussian noise $\mathbf{n}$ with $\sigma = 5$ to the 12 images from the MPI-Sintel dataset. The quantitative results are given in the Table 2, compared with two methods [8, 15]. It

can be seen that our method is more robust to noise by using the sparse priors on the reflectance and shading.

5.4. The Running Time

We compare the average running times of different methods on the images with different sizes in Table 3. All the experiments are run on a desktop with Intel Core i7 CPU 3770 and 32GB RAM. In a whole, our method is fast without loss of accuracy.

Table 3. *Comparison of running times.*

Running time / Image Size / Method	1024×436	620×420
[8]	1160.7s	660s
[15]	121.1s	70.9s
Ours	139.8s	103.6s

6. CONCLUSIONS

In this paper, we present a new method for intrinsic decomposition from a single RGB-D image. We exploit non-local and sparse priors to make the problem well-posed. The alternating direction algorithm under the augmented Lagrangian multiplier (ADM-ALM) framework is adopted to solve the minimization of the proposed energy function. Experimental results demonstrate that our method produces better qualitative results and lower quantitative errors, and the results on the noisy images prove the robustness of the proposed method. The code of our method will be publicly available on our project website.

7. REFERENCES

[1] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *Proc. ICCV*, 2010, pp. 2335–2342.

[2] Haipeng Dai, Wei Feng, Liang Wan, and Xuecheng Nie, "L0 co-intrinsic images decomposition," in *Proc. ICME*, 2014, pp. 1–6.

[3] Jonathan Barron and Jitendra Malik, "Intrinsic scene properties from a single RGB-D image," in *Proc. CVPR*, 2013, pp. 17–24.

[4] Harry Barrow and J Tenenbaum, "Recovering intrinsic scene characteristics," *Comput. Vis. Syst., A Hanson & E. Riseman (Eds.)*, pp. 3–26, 1978.

[5] Edwin H Land and John J McCann, "Lightness and retinex theory," *JOSA*, vol. 61, no. 1, pp. 1–11, 1971.

[6] Qi Zhao, Ping Tan, Qiang Dai, Li Shen, Enhua Wu, and Stephen Lin, "A closed-form solution to retinex with nonlocal texture constraints," *IEEE Trans. PAMI*, vol. 34, no. 7, pp. 1437–44, 2012.

[7] Shen Li and Chuohao Yeo, "Intrinsic images decomposition using a local and global sparse representation of reflectance," in *Proc. CVPR*, 2011, pp. 697–704.

[8] Qifeng Chen and Vladlen Koltun, "A simple model for intrinsic image decomposition with depth cues," in *Proc. ICCV*, 2013, pp. 241–248.

[9] Jian Shi, Yue Dong, Xin Tong, and Yanyun Chen, "Efficient intrinsic image decomposition for RGBD images," in *Proc. the 21st ACM Symposium on Virtual Reality Software and Technology*. ACM, 2015, pp. 17–25.

[10] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[11] Y. Weiss, "Deriving intrinsic images from image sequences," in *Proc. ICCV*, 2001, pp. 68–75 vol.2.

[12] Adrien Bousseau, Sylvain Paris, and Frdo Durand, "User-assisted intrinsic images," *ACM Trans. Graphics*, vol. 28, no. 5, pp. 89–97, 2009.

[13] Genzhi Ye, Elena Garces, Yebin Liu, Qionghai Dai, and Diego Gutierrez, "Intrinsic video and applications," *ACM Trans. Graphics*, vol. 33, no. 4, pp. 1–11, 2014.

[14] Nicolas Bonneel, Kalyan Sunkavalli, James Tompkin, Deqing Sun, Sylvain Paris, and Hanspeter Pfister, "Interactive intrinsic video editing," *ACM Trans. Graphics*, vol. 33, no. 6, pp. 1–10, 2014.

[15] Junho Jeon, Sunghyun Cho, Xin Tong, and Seungyong Lee, "Intrinsic image decomposition using structure-texture separation and surface normals," in *Proc. ECCV*, 2014, pp. 218–233.

[16] Antoni Buades, Bartomeu Coll, and J-M Morel, "A non-local algorithm for image denoising," in *Proc. CVPR*. IEEE, 2005, vol. 2, pp. 60–65.

[17] J. Yang, X. Ye, K. Li, C. Hou, and Y. Wang, "Color-guided depth recovery from RGB-D data using an adaptive autoregressive model," *IEEE Trans. IP*, vol. 23, no. 8, pp. 3443–3458, 2014.

[18] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, and Michael J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proc. ECCV*, 2012, pp. 611–625.