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 ©2017 IEEE

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.

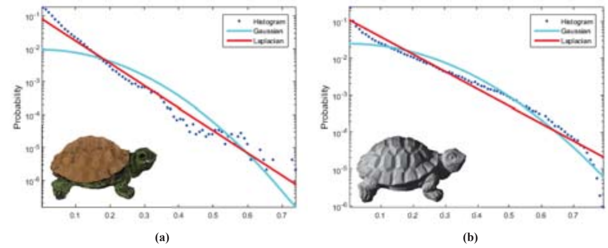

Fig. 1. Normalized histograms and the associated fitted Laplacian and Gaussian distributions for pairwise errors of (a) reflectance and (b) shading for MIT-Berkeley Intrinsic Images dataset [3].

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-