Intrinsic Image Decomposition Using Optimization and User Scribbles

Jianbing Shen, *Member, IEEE*, Xiaoshan Yang, Xuelong Li, *Fellow, IEEE*, and Yunde Jia

***Abstract*—In this paper, we present a novel high-quality intrinsic image recovery approach using optimization and user scribbles. Our approach is based on the assumption of color characteristics in a local window in natural images. Our method adopts a premise that neighboring pixels in a local window having similar intensity values should have similar reflectance values. Thus, the intrinsic image decomposition is formulated by minimizing an energy function with the addition of a weighting constraint to the local image properties. In order to improve the intrinsic image decomposition results, we further specify local constraint cues by integrating the user strokes in our energy formulation, including constant-reflectance, constant-illumination, and fixed-illumination brushes. Our experimental results demonstrate that the proposed approach achieves a better recovery result of intrinsic reflectance and illumination components than the previous approaches.**

***Index Terms*—Energy optimization, illumination, intrinsic images, reflectance, user scribbles.**

I. Introduction

INTRINSIC IMAGES are usually referred to the separation of illumination (shading) and reflectance components from an input photograph. It was originally proposed by Barrow and Tenenbaum [2] for presenting the scene properties, including the illumination of the scene and the reflectance of the surfaces of the scene. Intrinsic images can be recovered either from a single image [25], [40] or from image sequences [14], [18]. We focus on the particular case of separating a single image into illumination and reflectance components in this paper. It is well known that automatic intrinsic image algorithms are error prone due to the fundamental ill posedness of the intrinsic image recovery problem. As will be reviewed in the related work section, the unsuitable assumptions and restrictions in intrinsic image algorithms make them insufficient to produce high-quality illumination and reflectance components.

Color is usually employed as a key cue for identifying shading gradients, such as the approaches in [6] and [25]. Tappen *et al.* [25] proposed an algorithm for recovering intrinsic images from a single photograph. Their approach was based on a trained classifier, which classified the image derivatives as being caused by either illumination or reflectance changes. Then, the illumination and reflectance images were calculated according to the classified derivatives. By using pieces of paper colored with a marker as training images, Tappen *et al.* [26] created a set of ground-truth intrinsic images where they used color to measure shading separately from reflectance. Their approach then learned a weighting function using the nonlinear regression and estimated the intrinsic shading and reflectance images. Recently, Shen *et al.* [34] assumed that similar textures correspond to similar reflectance and improved the retinex-based algorithm for intrinsic image decomposition. More recently, Bousseau *et al.* [38] presented an approach for obtaining the intrinsic images with user interaction from a single image. Their approach calculated the illumination component and reflectance component by introducing a propagation energy.

It is still an open challenge on how to recover the high-quality intrinsic images when only a single photograph is available. Due to its inherent ill posedness, the automatic decomposition of intrinsic images on a single image cannot be solved correctly without additional prior knowledge on reflectance or illumination. In order to resolve its ambiguities, Bousseau *et al.* [38] presented a user-assisted approach for calculating the intrinsic images from a single image. Their approach required some user interactions to indicate the regions of constant reflectance or illumination and achieved intrinsic image decomposition by a propagation energy function. However, their approach was based on the assumption of reflectance variations lying locally in a plane in color space, and their method would introduce the incorrect decompositions when it handled the cases such as a black-and-white checkerboard texture.

In the context of the previous work on intrinsic image recovery, we can see that our paper makes the following research contributions.

1) We present a novel automatic intrinsic image recovery algorithm using optimization, which focuses on the local continuity assumption in reflectance values to define the new energy function.
2) A new weight function within patch properties in a local window is proposed for making the pixels which have large different shading values share the similar reflectance

Manuscript received October 29, 2011; revised April 16, 2012; accepted July 5, 2012. Date of publication August 14, 2012; date of current version April 16, 2013. This work was supported in part by the Key Program of NSFC-Guangdong Union Foundation under Grant U1035004, by the National Natural Science Foundation of China under Grants 60903068, 61125106, 91120302, and 61072093, by the National Basic Research Program of China (973 Program) under Grant 2012CB316400, and by the Program for New Century Excellent Talents in University (NCET-11-0789). This paper was recommended by Associate Editor D. Goldgof.

J. Shen, X. Yang, and Y. Jia are with the Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: shenjianbing@bit.edu.cn; 20907118@bit.edu.cn; jiayunde@bit.edu.cn).

X. Li is with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2012.2208744

values. Our new weight function can help to improve the accuracy of the recovered intrinsic images.

3) An efficient intrinsic image decomposition approach is proposed by adding three types of user scribbles to the energy function, which improves the performance of the results by the automatic algorithm.

The remainder of this paper is organized as follows. In Section II, we first present an overview of previous intrinsic image recovery approaches. Next, we describe our algorithm including the energy optimization issues and implementation details in Section III. Then, in Section IV, we perform a variety of experimental comparisons on natural images to validate the efficiency of the proposed method. Finally, the conclusions and discussions are shown in Section V.

II. Related Work

The intrinsic image recovery aims at decomposing an image into illumination and reflectance components. Various intrinsic image decomposition methods have been proposed in recent years. Several heuristic cues are exploited to estimate the intrinsic components with extra constraints or user intervention. We describe these approaches in detail in this section and further address the main advantages and disadvantages which led us to develop our novel method in this paper.

The problem of decomposing a single image into intrinsic images remains very difficult [1], [2], [10], [17], [20], [24], [25], [38], [40]. Since it is an ill-posed problem, various kinds of models and assumptions are proposed to obtain intrinsic images by utilizing additional information in the computer vision and image processing literatures. It has been proven to be a great success to estimate intrinsic images from image sequences [14], [18], [19]. Weiss [14] focused on the intrinsic image decomposition using an image sequence where the reflectance is constant and the illumination changes. Weiss treated the intrinsic image decomposition as a maximum-likelihood estimation problem and used natural image statistics to estimate a single reflectance image and multiple illumination images by a prior. Matsushita *et al.* [18] derived a time-dependent intrinsic image decomposition method by incorporating an illumination eigenspace into Weiss' framework. Thus, *a priori* information was provided to derive the intrinsic images from a single video frame by the generated eigenspace. Another important work by Matsushita *et al.* [19] is the approach to estimate intrinsic images with biased illumination from image sequences. They explicitly modeled spatial and temporal constraints over the image sequence and minimized a regularization function by taking advantage of the biased image derivatives. Although these approaches have demonstrated to produce high-quality intrinsic image decomposition results, the required multiple images have limited their general applications.

The decomposition approaches from a single image are more favorable for general practicality. The previous representative methods [15], [25], [26], [34], [38], [40] mainly focused on analyzing local derivatives for determining image feature variations due to shading or reflectance, such as gradient variations and texture cues. One of the earliest approaches to separate an image into its reflectance and shading components is the retinex algorithm [1]. Their approach assumed that large derivatives were attributed to reflectance changes, while smaller derivatives were due to shading changes in a scene. Then, the intrinsic images were decomposed by integrating their respective derivatives across the input image. Based on the assumption that shading variations do not alter chromaticity, Funt *et al.* [6] further extend the retinex approach to be applicable for color images by associating reflectance derivatives to significant chromaticity changes. By integrating the previous retinex approaches [1], [8], [9], Kimmel *et al.* [16] proposed a variational retinex model for intrinsic image recovery. Their method formulated the illumination estimation as a quadratic optimization, and they exploited the spatial correlation in the reflectance and shading components.

The retinex methods [1], [6], [16] are intuitively simple and efficient; however, the real scenes do not always hold the aforementioned assumption of scene conditions. Thus, another important class of intrinsic image recovery methods is proposed, which is based on the local gradients. These approaches used heuristic rules [7], [12], [17] or trained classifiers [15], [25], [26], [29] to classify each image derivative or edge junction as being caused by shading or a change in the surface's reflectance. Freeman and Viola [12] presented an approach by combining the psychophysical and the Bayesian computational model. The wavelet subband prior probabilities were assigned to different reliefs for an image by their computational model, and the intrinsic images were calculated for the most probable interpretation in a Bayesian framework. Finlayson *et al.* [21] derived an illumination-invariant image by entropy minimization for the task of shadow removal from a single color image without resorting to any calibration. Their method also derived an intrinsic image decomposition method based on assumptions of the Lambertian reflectance and Planckian lighting. Tappen *et al.* [25], [26] utilized color information and binary classifiers to recognize the derivatives caused by reflectance or shading. They used the generalized belief propagation algorithm to disambiguate the areas of the image where the classification was not clear or incorrect with the local analysis. However, it was difficult to comprehensively define the trained classifiers, which were suitable for all the possible range of shading and reflectance configurations. Moreover, it was not always possible to correctly classify the scene pixels as reflectance or shading changes only by local appearance evidence.

Aside from the intrinsic image decomposition approaches studied in the previous work, several image editing and video processing tools employing the intrinsic images are developed in recent years, such as video surveillance [19], shadow removal [28], intrinsic colorization [32], image retexturing [38], and image segmentation [37]. Liu *et al.* [32] presented an example-based intrinsic colorization technique based on illumination differences between grayscale target and color reference images. Their approach first recovered the reflectance image of the target scene from multiple color references obtained by Web search, and then, they transferred color from the color reflectance image to the grayscale reflectance image through relighting with the illumination component of the target image. Yacoob and Davis [37] first derived three intrinsic images,

including reflectance, smooth-surface shading, and mesostructure roughness shading images. Then, they proposed an image segmentation approach based on the appearance of the mesostructure roughness shading image.

Several other approaches on energy optimization have been presented. User scribble-based approaches for image editing by energy optimization are very prevalent, such as image colorization [22], image photomontage [23], interactive tone mapping [27], and image edit propagation [35]. Agarwala *et al.* [23] developed an interactive digital photomontage system that combined parts of a set of photographs into a composite image by energy optimization and user scribbles. Levin *et al.* [22] used scribbles to define constraints between adjacent pixels and propagated them to the entire image colorization by solving an optimization problem. Lischinski *et al.* [27] proposed an optimization method to propagate local tonal adjustment, which required the user to draw a set of sparse scribbles indicating the desired changes in tone and color. Then, their method automatically propagates these initial adjustments to the rest of the image. An and Pellacini [35] presented an edit propagation algorithm which enforced the scribble-based constraints over all pairs of points in the image.

III. Our Approach

Assuming that the surfaces are the Lambertian objects and a single light color in the previous work, we express the intrinsic image decomposition process as

$$I = sR \tag{1}$$

where I represents the input image, s denotes the illumination component, and R is the reflectance component. The task of intrinsic image decomposition is to solve the two unknowns s and R on the right-hand side of (1). For each pixel $i \in I$, we denote it as $I_i = (I_{ir}, I_{ig}, I_{ib})$, $R_i = (R_{ir}, R_{ig}, R_{ib})$, and $I_i = s_i R_i$. It is obvious that the aforementioned equation is ill posed with two unknowns and one known, and our goal is to recover s and R from I. Based on our observation of local color characteristics in natural images, we design a new decomposition approach using optimization and some user brush constraints.

Our approach is based on the assumption of local color characteristics in natural images: In a local window of an image, the changes of pixel values are usually caused by the changes of the reflectance, i.e., the pixels with the similar intensity values share the similar reflectance values [41]. Our aforementioned assumption is inspired by both the property of a change in color between pixels often indicating a reflectance change [3] and the behavioral evidence that humans usually distinguish variation in shading from variation in reflectance [31], [33]. Thus, the reflectance value of one pixel can be represented by the weighted summation of its neighborhood pixel values as follows:

$$R_i = \sum_{j \in N(i)} \omega_{ij} R_j, \quad \omega_{ij} = e^{-(Y_i - Y_j)^2 / \left(2\sigma_i^2\right)} \tag{2}$$

where ω_{ij} measures the similarity of the reflectance value between pixel i and pixel j. ω_{ij} is determined by the intensity values of the Y channel in the input image. In our implementation, we convert the original image from RGB color space into YUV color space to get the Y channel. Y_i and Y_j represent the intensity values of pixel i and pixel j, respectively. σ_i denotes the variance of the intensities in a local window (e.g., 5×5 pixels) around i. In image segmentation [11], [13], [30], [39] and image colorization algorithm [22], similar weight functions are used extensively and referred to as affinity functions.

A. Energy Based on Local Windows

In order to make the pixels that have large different shading values share the similar reflectance values, we reconsider the assumption that the changes of shading values will lead to the proportional changes of its R, G, B color channel values. Therefore, we improve the aforementioned weight function (2) as follows:

$$\begin{aligned} \omega_{ij} &= e^{-\left[\langle \tilde{I}_i, \tilde{I}_j \rangle^2 / \sigma_{iT}^2 + (Y_i - Y_j)^2 / \sigma_{iY}^2\right]} \\ \langle \tilde{I}_i, \tilde{I}_j \rangle &= \arccos(\tilde{I}_i * \tilde{I}_j) \\ &= \arccos(\tilde{I}_{ir}\tilde{I}_{jr} + \tilde{I}_{ig}\tilde{I}_{jg} + \tilde{I}_{ib}\tilde{I}_{jb}) \end{aligned} \tag{3}$$

where $\langle \tilde{I}_i, \tilde{I}_j \rangle$ denotes the angle between the vector $\tilde{I}_i$ and $\tilde{I}_j$. Similar to [25], we normalize the RGB triplet I_i and I_j as a vector to obtain $\tilde{I}_i$ and $\tilde{I}_j$. σ_{iT} represents the variance of the angle between pixel i and the pixels in a local window around i, and σ_{iY} denotes the variance of the intensities of the pixels in a local window around i.

Fig. 1 gives the error comparisons of the intrinsic image using different weight functions. We use the image patches in a local window to calculate the error between the recovered intrinsic images and the ground-truth intrinsic images from the MIT data set [36]. It is obvious that the error by our new weight function (3) is smaller than the error by the rude weight function (2). Based on the previous analysis, we now can define a new energy function to obtain the intrinsic images from a single input photograph

$$E(R, \tilde{s}) = \sum_{i \in P} \left(R_i - \sum_{j \in N(i)} w_{ij} R_j \right)^2 + \sum_{i \in P} (I_i \tilde{s}_i - R_i)^2 \tag{4}$$

where $N(i)$ represents the local neighborhood window (e.g., 5×5 pixels) of pixel i, the first energy term denotes the constraint for R in a local window, and the second energy term constrains R and s in (1); here, $\tilde{s} = 1/s$. We then optimize the above equation to obtain the intrinsic images

$$\begin{aligned} &\arg\min_{R, \tilde{s}} E(R, \tilde{s}) \qquad \forall i \in P, \\ &\qquad 0 \le R_{ir} \le 1,\ 0 \le R_{ig} \le 1, 0 \le R_{ib} \le 1. \end{aligned} \tag{5}$$

B. Energy Based on User Scribbles

Previous automatic approaches [16], [25], [26] usually work well for the images of the simple scenes to obtain the desired

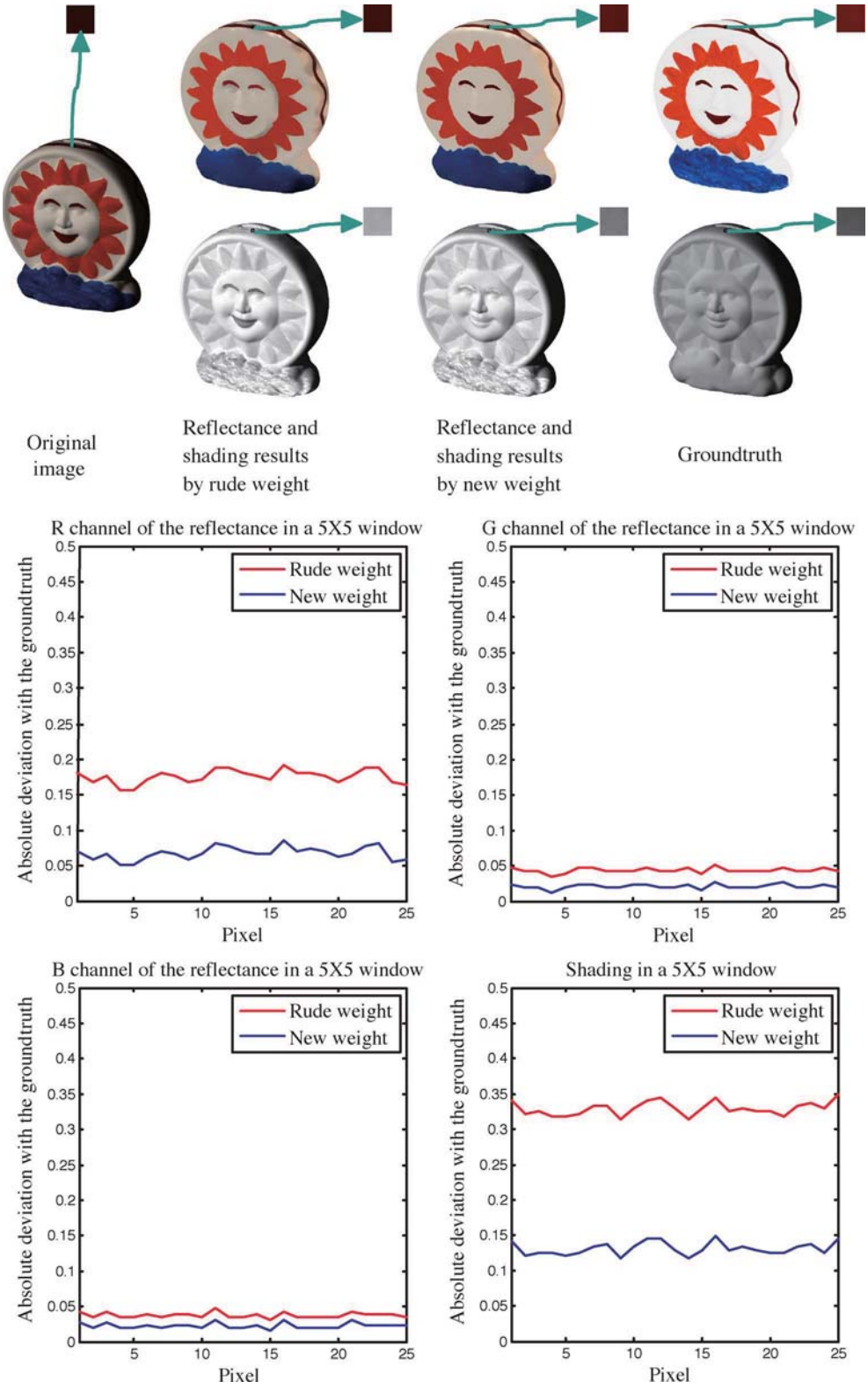

Fig. 1. Error comparisons of intrinsic image recovery using the rude weight function (2) and our new weight function (3). Compared with the ground-truth data, the recovered intrinsic images by our new weight function are more accurate.

intrinsic images. Unfortunately, such class of automatic approaches is not appropriate when an image contains complicated illumination and reflectance information. We propose to specify local constraint cues by integrating the user scribbles in our energy formulation.

Unlike Tappen's method [25], they classified each derivative as being caused by either an illumination change or a reflectance change. We employ the user scribbles to indicate each intensity change as being caused by either a shading change or a reflectance change. Thus, we adopt three types of user scribbles

Fig. 2. Illustration of our optimization approach. Note that (b) and (c) are the results by our automatic approach (5), while (g) and (h) are the results by our full approach with user scribbles (7). (a) Input. (b) Reflectance (auto). (c) Illumination (auto). (d) User scribbles. (e) Reflectance by the approach of Bousseau *et al.* [38]. (f) Illumination by the approach of Bousseau *et al.* [38]. (g) Reflectance (ours). (h) Illumination (ours).

as Bousseau *et al.* [38] did, including constant-reflectance brush, constant-illumination brush, and fixed-illumination brush

$$E^R(R) = \sum_{i \in P} z\left(B_i^R\right) \sum_{k \in B_i^R} (R_i - R_k)^2$$

$$E^s(\tilde{s}) = \sum_{i \in P} z\left(B_i^s\right) \sum_{k \in B_i^s} (\tilde{s_i} - \tilde{s_k})^2$$

$$E^{fix}(\tilde{s}) = \sum_{i \in B_i^{fix}} (\tilde{s_i} - \tilde{s}_{fix})^2.$$

The constant-reflectance brush or the constant-illumination brush constrains the pixels which share the same reflectance or illumination, respectively. While the fixed-illumination brush specifies the absolute illumination values by the user constraints. Now, we can define the new energy optimization function by adding the user scribbles as follows:

$$E_C(R, \tilde{s}) = E(R, \tilde{s}) + \lambda_R E^R(R) + \lambda_s E^s(\tilde{s}) + \lambda_{fix} E^{fix}(\tilde{s}) \tag{6}$$

where λ_R, λ_s, and λ_{fix} measure the importance to our image decomposition model with the user's constant-reflectance cue, constant-illumination cue, and fixed-illumination cue, respectively. We set $\lambda_R = \lambda_s = \lambda_{fix} = 1.0$ to give equal importance of the scribbles to the energy optimization function, and the satisfying results are produced with these default parameter settings in our implementation.

Thus, we recover the intrinsic images by optimizing the following energy equation:

$$\arg\min_{R,\tilde{s}} E_C(R, \tilde{s}) \qquad \forall i \in P,$$
$$0 \le R_{ir} \le 1,\ 0 \le R_{ig} \le 1,\ 0 \le R_{ib} \le 1,\ \tilde{s} \ge 1. \tag{7}$$

In summary, the energy (5) and (7) can be viewed as a quadratic equation, which is based on the variables R_{ir}, R_{ig}, R_{ib}, and $\tilde{s}_i$. Then, we compute the first derivatives with these equations and make them equal to zero. Finally, we use the Gaussian–Seidel method iteratively to optimize these equations and get the optimal solutions in our implementation. For more implementation details, please see the Appendix. Since our energy function is quadratic and this optimization problem can be solved using a number of standard methods, we adopt the Gaussian–Seidel method [4] to solve our energy optimization for its simplicity, and we believe that a multigrid solver [5] will improve the computational efficiency of our energy optimization.

Fig. 2 gives an illustration example of the intrinsic image recovery by our optimization approach. The results as shown in Fig. 2(b) and (c) are obtained by optimizing (5), which is an automatic process. We then improve the performance by adding the user scribbles through optimizing (7), which are shown in Fig. 2(g) and (h). In particular, the image example shown in Fig. 2 consists of a black-and-white region of clothing texture, which belongs to the cases that cannot be handled by the previous user-assisted approach of Bousseau *et al.* [38].

Compared with the results by the approach of Bousseau *et al.* [38] [see Fig. 2(e) and (f)], our optimization approach produces better results and exhibits natural edges and structures.

Now, we will discuss how to solve the multiplicative ambiguity between shading and reflectance in both our automatic method and user scribble-based approach. In order to obtain the high-quality intrinsic images, it is important to reduce the multiplicative ambiguity and recognize the color information caused by the reflectance or shading changes. Tappen *et al.* [25] used the derivatives and learning method to classify the scene pixels as the reflectance or shading changes in [25] and [26]. In our automatic approach, we use the color angle $\langle \tilde{I}_i, \tilde{I}_j \rangle$ in a local window (e.g., 5×5 pixels) to reduce the multiplicative ambiguity between shading and reflectance information. In order to further eliminate such multiplicative ambiguity, we employ the user scribbles to indicate the scene pixels as reflectance or shading changes. For instance, we use the constant-reflectance brush to constrain the pixels that share the same reflectance, which eliminates the ambiguity and recognizes the color information caused by the shading changes.

IV. Experimental Result

In our experiments, most of the input photographs are taken from the previous work in [25] and [38] for comparison. The ground-truth images of the MIT data set are taken from the Web sites.[1] For the experimental results by the user-assisted intrinsic image recovery method [38], we directly use the original pseudocode implementation[2] provided by the respective authors in our comparison.

In order to demonstrate the effectiveness of the proposed approach, we first compare our method with the approach in [25] and [38]. As shown in Fig. 3, our algorithm achieves better reflectance and illumination decomposition results, such as the clear edges and texture information of the baby's red sweater. The main limitation of Tappen's algorithm [25] is that a binary labeling strategy cannot handle areas where both the reflectance and illumination variations occur, so the result by their approach exhibits some visual artifacts in the highly textured areas and under the mixed lighting conditions.

We next show a more challenging comparison example of a synthetic image for intrinsic image recovery in Fig. 4. The result by Bousseau's approach [38] interprets the black pixels of the eyes as shadow [see Fig. 4(b)], which is incorrect for the illumination image. In contrast, our approach is more consistent with the ground-truth data, particularly interpreting both the black pixels of the eyes and the pink pixels of the nose as reflectance [see Fig. 4(c)].

As pointed out by Bousseau *et al.* [38], the results by their approach can be sensitive to the quality of input images when dealing with JPEG compression images. Fig. 5 compares the intrinsic images recovered by Bousseau *et al.* [38] with the results by our automatic approach. Clearly, the JPEG artifacts and noise by Bousseau *et al.*'s approach become visible [see top row of Fig. 5(c)]. In contrast, our automatic approach obtains

[1] http://people.csail.mit.edu/rgrosse/intrinsic/.
[2] http://artis.imag.fr/Publications/2009/BPD09.

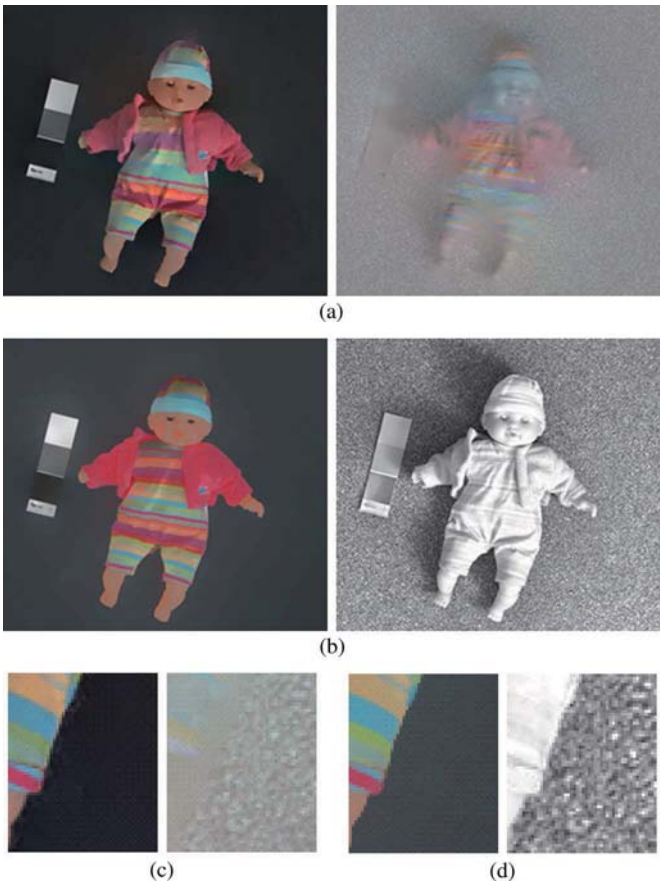

Fig. 3. Comparison results with the automatic approach of Tappen *et al.* [25]. (a) Results by the approach of Tappen *et al.* [25]. (b) Our results. (c) Close-ups of (a). (d) Close-ups of (b).

Fig. 4. Comparison results with the ground-truth data from a synthetic image [38]. Compared with the user-assisted approach of Bousseau *et al.* [38], our optimization method creates better results, particularly interpreting the black pixels of the eyes as reflectance. First row: Reflectance. Second row: Shading. Third row: Close-ups. (a) Input. (b) Bousseau's method [38]. (c) Our method.

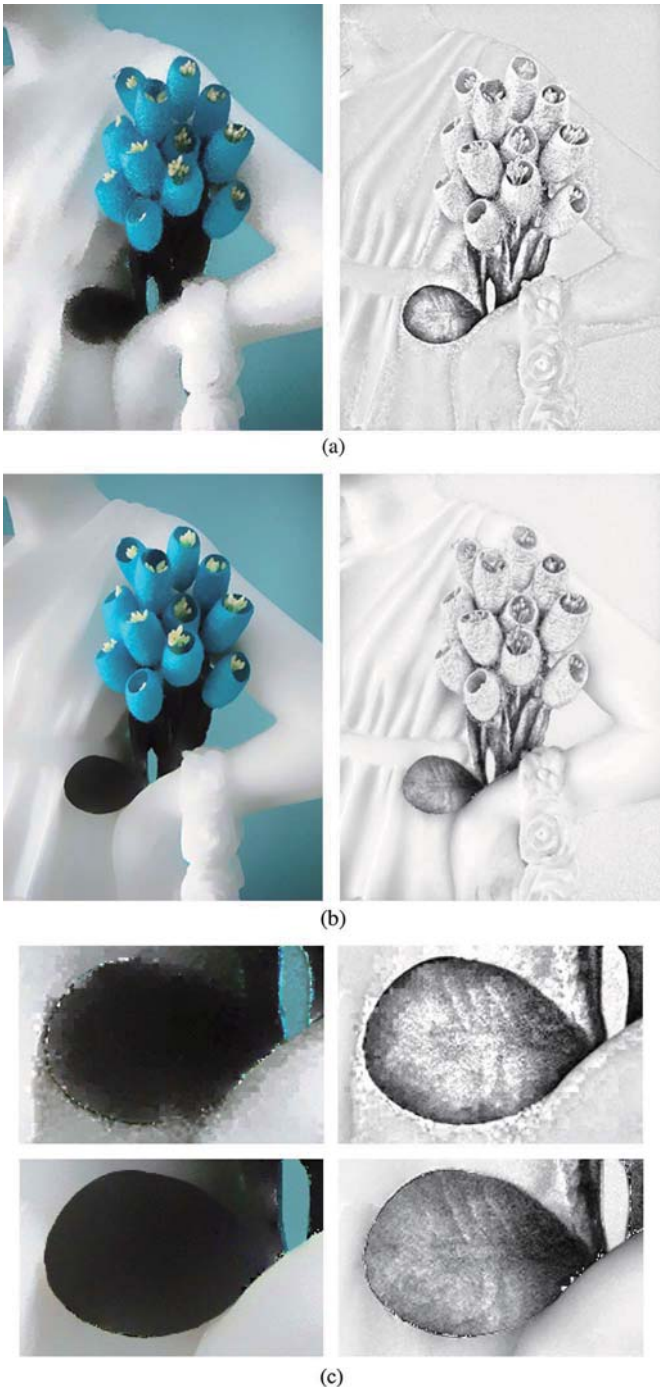

Fig. 5. Comparison between the intrinsic image recovery method in [38] and our automatic approach for dealing with the JPEG compression image. (a) Results by the approach of Bousseau *et al.* [38] (automatic). (b) Results by our automatic approach. (c) Enlargement of (a) and (b).

better visual quality with less JPEG compression artifacts [see Fig. 5(b)] and preserves more shadows and highlight components in both the reflectance and the illumination image [see bottom row of Fig. 5(c)].

In order to fix the global disambiguation between dark objects in light and bright objects in light, we allow the user to improve the intrinsic image results by a few user scribbles. Again, for the example shown in Fig. 6, our initial automatic intrinsic images can be refined by adding fixed-illumination, constant-reflectance, or constant-illumination scribbles. It is

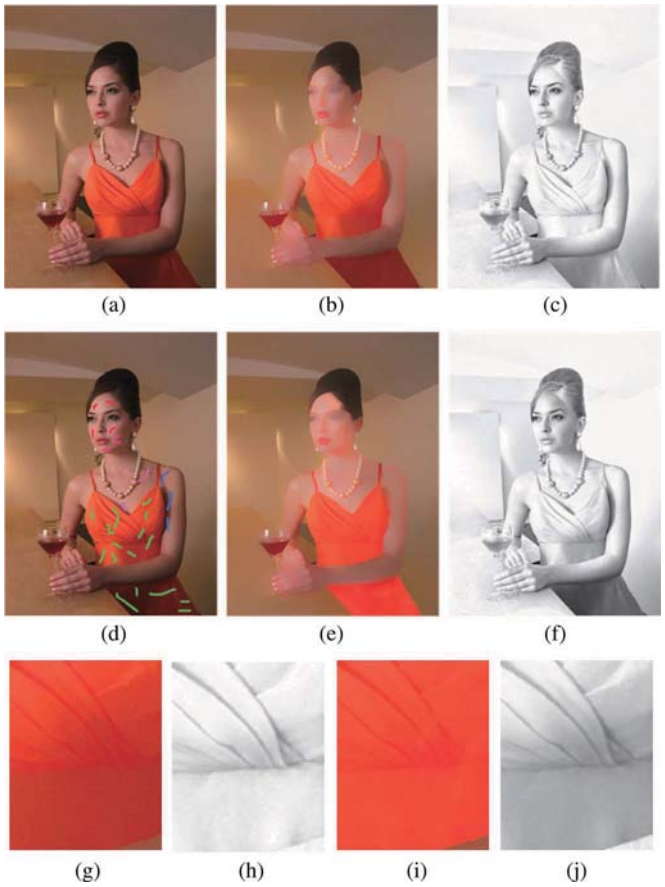

Fig. 6. Illustration of our approach with the user scribbles. (a) Input. (b) Reflectance (auto). (c) Shading (auto). (d) User scribbles. (e) Reflectance (our full approach). (f) Shading (our full approach). (g) and (h) Enlargement of (b), (c), (e), and (f). Note that the initial automatic intrinsic images are refined by adding the user scribbles, particularly the reflectance and illumination images of the orange color pixels of the one-piece dress are more accurate.

clear that the orange color pixels of the one-piece dress are correctly separated into the reflectance image [see Fig. 6(e)]. The shading pixels of the face skin are also correct after adding the fixed-illumination scribbles [see Fig. 6(f)]. This intrinsic image decomposition with user scribbles is more accurate because the different orange color regions of the one-piece dress do share a similar shading or reflectance [see Fig. 6(e) and (f)].

Quantitative Comparison: In order to quantitatively compare the existing algorithms, several objective performance measures of intrinsic image decomposition have been proposed recently, such as mean square error (MSE) scores, local MSE (LMSE) scores [36], and absolute LMSE (aLMSE) scores [40]. The MSE, LMSE, and aLMSE metrics are defined as follows [36], [40]:

$$\mathrm{MSE}(I_{GT},\hat{I}) = \arg\min_{a} \|I_{GT} - a\hat{I}\|$$

$$\mathrm{LMSE}(I_{GT},\hat{I}) = \sum_{w} \arg\min_{a} \left\| I_{GT}^{w} - a\hat{I}^{w} \right\|$$

$$\mathrm{aLMSE}(I_{GT},\hat{I}) = \sum_{w} \arg\min_{a} \left\| (I_{GT}^{w} - \mu_{GT}^{w}) - a\left(\hat{I}^{w} - \mu_{I}^{w}\right) \right\|$$

where $\hat{I}$ is the estimated intrinsic image and I_{GT} is the ground-truth image in a local window w. μ_{GT}^{w} and μ_{I}^{w} are the mean

TABLE I
MSE, LMSE, AND aLMSE MEASURE STATISTICS WITH [25] AND [38] USING THE MIT DATA SET [36]

		Results by [25]			Results by [38]			Our results		
ID	Image	LMSE	MSE	aLSME	LMSE	MSE	aLSME	LMSE	MSE	aLSME
1	apple	0.0364	0.3564	0.2390	0.0436	0.4032	0.3803	0.0102	0.0336	0.1111
2	box	0.0891	0.5042	0.5063	0.0274	0.1332	0.1070	0.0115	0.0138	0.0565
3	cup1	0.0599	0.4492	0.4246	0.0067	0.0116	0.0437	0.0055	0.0080	0.0356
4	cup2	0.0606	0.4061	0.4684	0.0164	0.1733	0.0778	0.0073	0.0114	0.0396
5	deer	0.0773	0.3902	0.3665	0.0950	0.9166	0.3077	0.0319	0.0658	0.1744
6	dinosaur	0.0984	0.5577	0.4845	0.0610	0.5871	0.1393	0.0212	0.0368	0.0935
7	frog1	0.1287	0.4877	0.4765	0.0377	0.0826	0.1726	0.0287	0.0755	0.1404
8	frog2	0.1747	0.5662	0.6419	0.0771	0.3004	0.2772	0.0238	0.0711	0.1427
9	panther	0.0699	0.4097	0.5036	0.0136	0.5196	0.0686	0.0049	0.0073	0.0354
10	paper1	0.0310	0.3167	0.4431	0.0085	0.0568	0.0613	0.0125	0.0157	0.1107
11	paper2	0.0307	0.3098	0.4482	0.0134	0.1080	0.0911	0.0161	0.0195	0.1314
12	pear	0.0534	0.3384	0.3101	0.0190	0.2179	0.1301	0.0102	0.0332	0.1026
13	phone	0.0811	0.3681	0.3493	0.0143	0.5164	0.0482	0.0112	0.0398	0.0477
14	potato	0.1016	0.3585	0.4966	0.0219	0.0490	0.1479	0.0140	0.0313	0.1129
15	raccoon	0.0843	0.4785	0.4710	0.0097	0.0229	0.0596	0.0077	0.0139	0.0529
16	squirrel	0.1123	0.4539	0.4562	0.0982	0.6005	0.3382	0.0374	0.0624	0.1921
17	sun	0.0307	0.2591	0.3763	0.0105	0.0340	0.0815	0.0070	0.0144	0.0566
18	teabag1	0.0666	0.3391	0.4052	0.4131	0.8853	0.5613	0.0631	0.0743	0.1421
19	teabag2	0.0644	0.3211	0.2607	0.2742	0.6842	0.3696	0.0307	0.0553	0.1236
20	turtle	0.1251	0.4309	0.5905	0.0293	0.0998	0.1942	0.0247	0.0592	0.1819
	Mean	0.0788	0.4051	0.4359	0.0645	0.3201	0.1828	0.0190	0.0371	0.1042

*Note that all evaluations are calculated by the mean values of summing the shading and reflectance images.

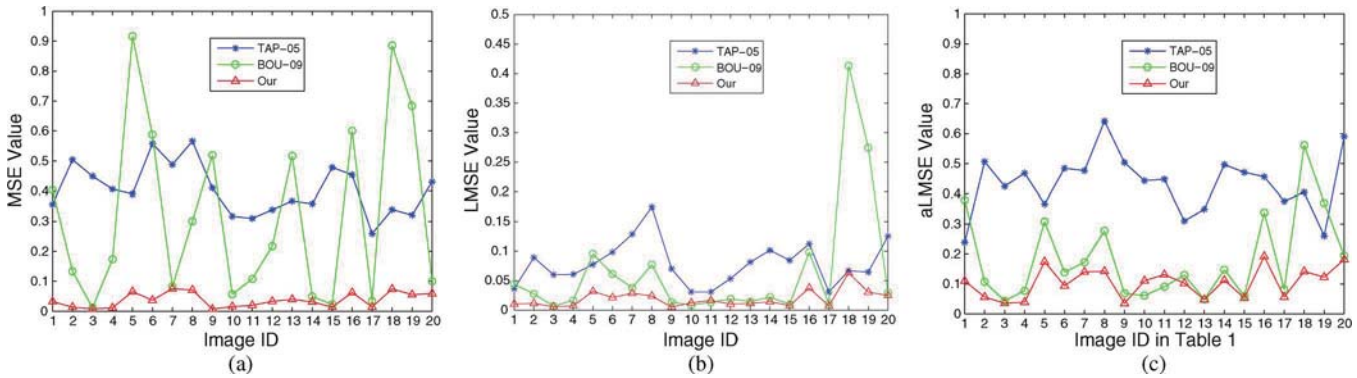

Fig. 7. Quantitative evaluation: Plots of Table I. (a) MSE comparison. (b) LMSE comparison. (c) aLMSE comparison.

values for the ground-truth image I^w_{GT} and the estimated intrinsic image $\hat{I}^w$ in a local window w.

The LMSE measure associates incorrect classification with edge information and defines a meaningful error metric, while the aLMSE measure is insensitive for variations in mean intensity through subtracting out the mean values of the two intrinsic images. In order to quantitatively compare our algorithm with other different methods, we compute three statistics, including MSE, LMSE, and aLMSE scores, respectively, which are shown in Table I. Fig. 7 gives a more intuitive figure illustration of Table I. As we would expect, our optimization approach performs better than both Tappen's approach [25] and Bousseau's method [38] and achieves less error decomposition. Fig. 8 gives the performance comparisons on the MIT data set [36]. The first column in Fig. 8(b) and (c) denotes the average performance. While the second and third columns are the performance on the reflectance and shading estimates, respectively. Note that the value of MSE and aLMSE is referred to left y-axis, while the value of LMSE is referred to the right y-axis.

Execution Time Comparison: Table II shows the execution time between Bousseau *et al.*'s decomposition method [38] and our intrinsic image optimization algorithm. It is worth noting that our optimization approach takes slightly more computation time than Bousseau *et al.*'s method while achieving better intrinsic image recovery quality with more accurate results. Note that we utilize the Gaussian–Seidel algorithm for optimizing the energy equation with 200 iterations.

V. CONCLUSION

We have presented a novel energy optimization approach for separating high-quality intrinsic images from a single input photograph. Our approach is based on a premise that neighboring pixels in a local window having similar intensity values should have similar reflectance values. Thus, the intrinsic image decomposition is formulated by an energy function with the addition of a novel weighting constraint to the local image characteristics. We further improve the accuracy of the intrinsic images by introducing the local constraints of user scribbles in our energy function, such as constant-reflectance, constant-illumination, and fixed-illumination brushes. Our experimental results demonstrate that our approach achieves a better recovery of intrinsic reflectance and illumination components than the previous approaches.

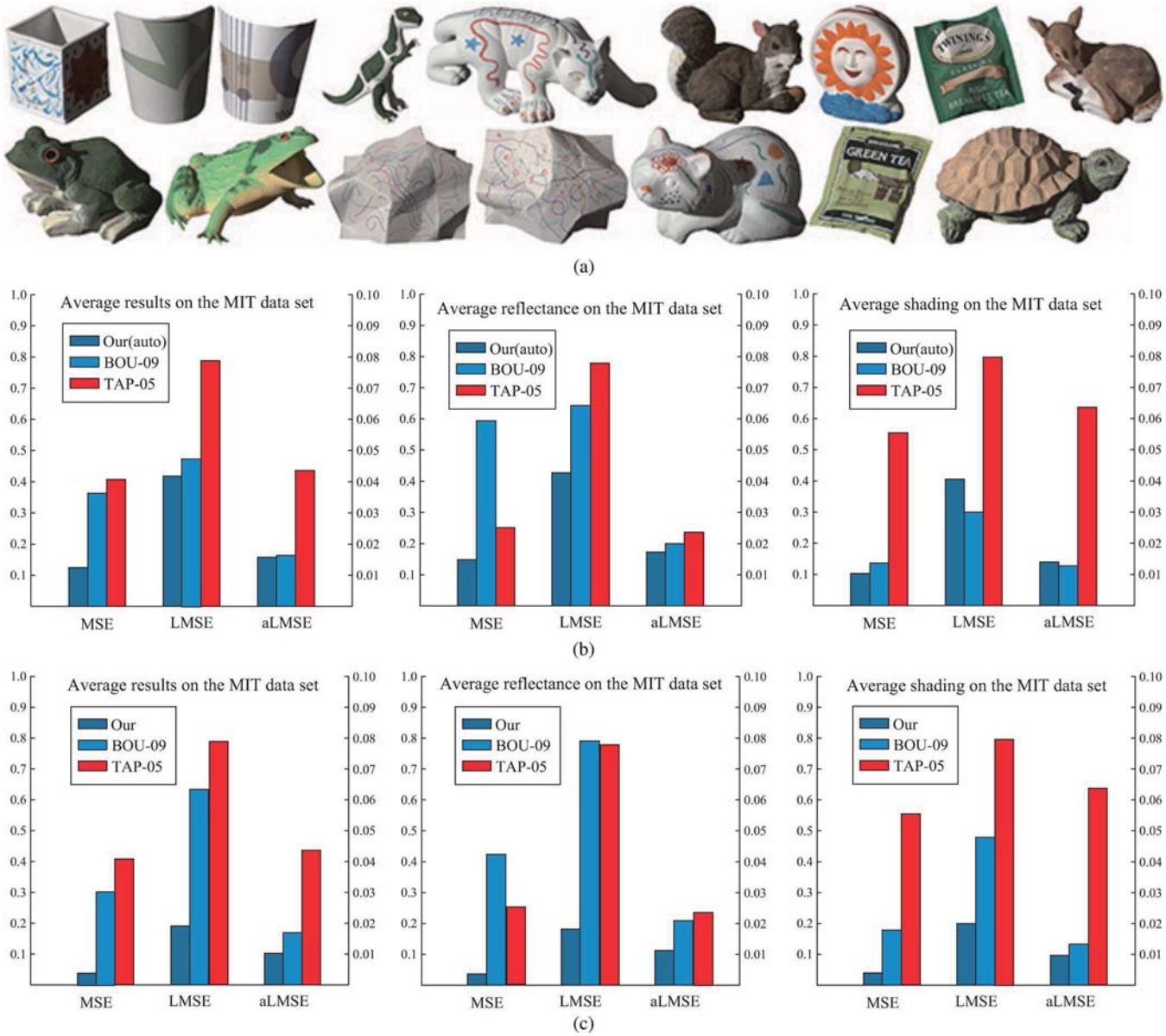

Fig. 8. Performance comparisons on MIT data set with the approaches by Tappen *et al.* [25] and Bousseau *et al.* [38]. Note that we abbreviate Tappen *et al.*'s [25] approach (TAP-05), and abbreviate Bousseau *et al.*'s [38] algorithm (BOU-09). (a) MIT data set [36]. (b) Comparing our automatic approach (5) with the methods by Tappen *et al.* [25] and Bousseau *et al.* [38]. (c) Comparing our full approach (7) with the methods by Tappen *et al.* [25] and Bousseau *et al.* [38].

TABLE II
EXECUTION TIME COMPARISON WITH THE METHOD BY BOUSSEAU *et al.* [38]

	image size	Bousseau *et al.*'s approach [38]	our method	user brushes
Figure 2	476×640	27.68	55.27	34
Figure 3	500×450	20.11	40.54	46
Figure 4	750×603	45.03	84.54	52
Figure 5	600×800	42.08	86.65	none
Figure 6	512×683	32.43	62.81	29

*Time is measured in seconds on a laptop with 2.4GHz i5 CPU and 2GB RAM.

Additionally, the main limitation of the proposed approach lies in its high computational cost. Our current unoptimized implementation takes approximately 42 s to process one 600×800 image on a laptop. Incorporating the graphic processing unit (GPU)-based technique into our intrinsic image recovery framework may help to provide an interactive feedback for the users. To address this issue, we are planning to implement our approach as a real-time application where the users can benefit from the useful scribbles for rapidly previewing the final results.

APPENDIX
IMPLEMENTATION DETAILS

In this section, we describe our exact formulations with more detailed explanations for completeness. We will derive the optimization algorithm with (5) for the interested reader, and we first revisit (5) as follows:

$$\arg\min_{R,\tilde{s}} E(R,\tilde{s}), \forall i \in P, 0 \le R_{ir} \le 1, 0 \le R_{ig} \le 1, 0 \le R_{ib} \le 1$$

$$E(R,\tilde{s}) = \sum_{i \in P} \left(R_i - \sum_{j \in N(i)} w_{ij} R_j \right)^2 + \sum_{i \in P} (I_i \tilde{s}_i - R_i)^2.$$

The above energy optimization can be viewed as a quadratic equation which is based on the variables R_{ir}, R_{ig}, R_{ib}, and $\tilde{s}_i$. After the first derivative with the above equation, we get

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ir}} = 2\left(R_{ir} - \sum_{j\in N(i)} w_{ij}R_{jr}\right) - 2(\tilde{s}_i \cdot I_{ir} - R_{ir}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jr} - \sum_{k\in N(j)} w_{jk}R_{kr}\right) \quad (8)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ig}} = 2\left(R_{ig} - \sum_{j\in N(i)} w_{ij}R_{jg}\right) - 2(\tilde{s}_i \cdot I_{ig} - R_{ig}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jg} - \sum_{k\in N(j)} w_{jk}R_{kg}\right) \quad (9)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ib}} = 2\left(R_{ib} - \sum_{j\in N(i)} w_{ij}R_{jb}\right) - 2(\tilde{s}_i \cdot I_{ib} - R_{ib}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jb} - \sum_{k\in N(j)} w_{jk}R_{kb}\right) \quad (10)$$

$$\frac{\partial E(R,\tilde{s})}{\partial \tilde{s}_i} = 2I_{ir}(\tilde{s}_i \cdot I_{ir} - R_{ir}) + 2I_{ig}(\tilde{s}_i \cdot I_{ig} - R_{ig}) + 2I_{ib}(\tilde{s}_i \cdot I_{ib} - R_{ib}). \quad (11)$$

After making $\partial E(R,\tilde{s})/\partial R_{ir} = \partial E(R,\tilde{s})/\partial R_{ig} = \partial E(R,\tilde{s})/\partial R_{ib} = \partial E(R,\tilde{s})/\partial \tilde{s}_i = 0$, we obtain the following equations:

$$2R_{ir} - \sum_{j\in N(i)} w_{ij}R_{jr} - \sum_{j\in N(i)} w_{ji}R_{jr} + \sum_{j\in N(i)} w_{ji}\sum_{k\in N(j)} w_{jk}R_{kr} - I_{ir}\tilde{s}_i = 0 \quad (12)$$

$$2R_{ig} - \sum_{j\in N(i)} w_{ij}R_{jg} - \sum_{j\in N(i)} w_{ji}R_{jg} + \sum_{j\in N(i)} w_{ji}\sum_{k\in N(j)} w_{jk}R_{kg} \quad (13)$$

$$2R_{ib} - \sum_{j\in N(i)} w_{ij}R_{jb} - \sum_{j\in N(i)} w_{ji}R_{jb} + \sum_{j\in N(i)} w_{ji}\sum_{k\in N(j)} w_{jk}R_{kb} - I_{ib}\tilde{s}_i = 0 \quad (14)$$

$$\left(I_{ir}^2 + I_{ig}^2 + I_{ib}^2\right)\tilde{s}_i - I_{ir}R_{ir} - I_{ig}R_{ig} - I_{ib}R_{ib} = 0. \quad (15)$$

Then, we use the Gaussian–Seidel method iteratively to optimize the above equations and get the optimal solutions in our implementation.

Similarly, we can optimize the energy function with user strokes (7) as follows:

$$E_C(R,\bar{s}) = E(R,\tilde{s}) + \left[\lambda_R E_R(R) + \lambda_s E_s(\bar{s}) + \lambda_{fix}E_{fix}(\tilde{s})\right].$$

After the first derivative with the above equation, we obtain

$\forall i \in B^R$:

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ir}} = 2\left(R_{ir} - \sum_{j\in N(i)} w_{ij}R_{jr}\right) - 2(\tilde{s}_i \cdot I_{ir} - R_{ir}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jr} - \sum_{k\in N(j)} w_{jk}R_{kr}\right) + 4\lambda_R z\left(B_i^R\right)\sum_{k\in B_i^R}(R_{ir} - R_{kr}) \quad (16)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ig}} = 2\left(R_{ig} - \sum_{j\in N(i)} w_{ij}R_{jg}\right) - 2(\tilde{s}_i \cdot I_{ig} - R_{ig}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jg} - \sum_{k\in N(j)} w_{jk}R_{kg}\right) + 4\lambda_R z\left(B_i^R\right)\sum_{k\in B_i^R}(R_{ig} - R_{kg}) \quad (17)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ib}} = 2\left(R_{ib} - \sum_{j\in N(i)} w_{ij}R_{jb}\right) - 2(\tilde{s}_i \cdot I_{ib} - R_{ib}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jb} - \sum_{k\in N(j)} w_{jk}R_{kb}\right) + 4\lambda_R z\left(B_i^R\right)\sum_{k\in B_i^R}(R_{ib} - R_{kb}) \quad (18)$$

$$\frac{\partial E(R,\tilde{s})}{\partial \tilde{s}_i} = 2I_{ir}(\tilde{s}_i \cdot I_{ir} - R_{ir}) + 2I_{ig}(\tilde{s}_i \cdot I_{ig} - R_{ig}) + 2I_{ib}(\tilde{s}_i \cdot I_{ib} - R_{ib}) \quad (19)$$

$\forall i \in B^s$:

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ir}} = 2\left(R_{ir} - \sum_{j\in N(i)} w_{ij}R_{jr}\right) - 2(\tilde{s}_i \cdot I_{ir} - R_{ir}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jr} - \sum_{k\in N(j)} w_{jk}R_{kr}\right) \quad (20)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ig}} = 2\left(R_{ig} - \sum_{j\in N(i)} w_{ij}R_{jg}\right) - 2(\tilde{s}_i \cdot I_{ig} - R_{ig}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jg} - \sum_{k\in N(j)} w_{jk}R_{kg}\right) \quad (21)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ir}} = 2\left(R_{ib} - \sum_{j\in N(i)} w_{ij}R_{jb}\right) - 2(\tilde{s}_i \cdot I_{ib} - R_{ib}) - 2\sum_{j\in N(i)} w_{ji}\left(R_{jb} - \sum_{k\in N(j)} w_{jk}R_{kb}\right) \quad (22)$$

$$\frac{\partial E(R,\tilde{s})}{\partial \tilde{s}_i} = 2I_{ir}(\tilde{s}_i \cdot I_{ir} - R_{ir}) + 2I_{ig}(\tilde{s}_i \cdot I_{ig} - R_{ig}) + 2I_{ib}(\tilde{s}_i \cdot I_{ib} - R_{ib}) + 4\lambda_s z\left(B_i^s\right) \sum_{k \in B_i^s} (\tilde{s}_i - \tilde{s}_k) \quad (23)$$

$\forall i \in B^{fix}$:

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ir}} = 2\left(R_{ir} - \sum_{j \in N(i)} w_{ij} R_{jr}\right) - 2(\tilde{s}_i \cdot I_{ir} - R_{ir}) - 2\sum_{j \in N(i)} w_{ji}\left(R_{jr} - \sum_{k \in N(j)} w_{jk} R_{kr}\right) \quad (24)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ig}} = 2\left(R_{ig} - \sum_{j \in N(i)} w_{ij} R_{jg}\right) - 2(\tilde{s}_i \cdot I_{ig} - R_{ig}) - 2\sum_{j \in N(i)} w_{ji}\left(R_{jg} - \sum_{k \in N(j)} w_{jk} R_{kg}\right) \quad (25)$$

$$\frac{\partial E(R,\tilde{s})}{\partial R_{ir}} = 2\left(R_{ib} - \sum_{j \in N(i)} w_{ij} R_{jb}\right) - 2(\tilde{s}_i \cdot I_{ib} - R_{ib}) - 2\sum_{j \in N(i)} w_{ji}\left(R_{jb} - \sum_{k \in N(j)} w_{jk} R_{kb}\right) \quad (26)$$

$$\frac{\partial E(R,\tilde{s})}{\partial \tilde{s}_i} = 2I_{ir}(\tilde{s}_i \cdot I_{ir} - R_{ir}) + 2I_{ig}(\tilde{s}_i \cdot I_{ig} - R_{ig}) + 2I_{ib}(\tilde{s}_i \cdot I_{ib} - R_{ib}) + 2\lambda_{fix}(\tilde{s}_i - \tilde{s}_{fix}). \quad (27)$$

After making $\partial E(R,\tilde{s})/\partial R_{ir} = \partial E(R,\tilde{s})/\partial R_{ig} = \partial E(R,\tilde{s})/\partial R_{ib} = \partial E(R,\tilde{s})/\partial \tilde{s}_i = 0$, we obtain the following equations:

$\forall i \in B^R$:

$$2(1+\lambda_R)R_{ir} - \sum_{j \in N(i)} w_{ij} R_{jr} - \sum_{j \in N(i)} w_{ji} R_{jr} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kr} - 2\lambda_R z\left(B_i^R\right) \sum_{k \in B_i^R} R_{kr} - I_{ir}\tilde{s}_i = 0 \quad (28)$$

$$2(1+\lambda_R)R_{ig} - \sum_{j \in N(i)} w_{ij} R_{jg} - \sum_{j \in N(i)} w_{ji} R_{jg} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kg} - 2\lambda_R z\left(B_i^R\right) \sum_{k \in B_i^R} R_{kg} - I_{ig}\tilde{s}_i = 0 \quad (29)$$

$$2(1+\lambda_R)R_{ib} - \sum_{j \in N(i)} w_{ij} R_{jb} - \sum_{j \in N(i)} w_{ji} R_{jb} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kb} - 2\lambda_R z\left(B_i^R\right) \sum_{k \in B_i^R} R_{kb} - I_{ib}\tilde{s}_i = 0 \quad (30)$$

$$\left(I_{ir}^2 + I_{ig}^2 + I_{ib}^2\right)\tilde{s}_i - I_{ir}R_{ir} - I_{ig}R_{ig} - I_{ib}R_{ib} = 0 \quad (31)$$

$\forall i \in B^s$:

$$2R_{ir} - \sum_{j \in N(i)} w_{ij} R_{jr} - \sum_{j \in N(i)} w_{ji}(R_{jr} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kr} - I_{ir}\tilde{s}_i = 0 \quad (32)$$

$$2R_{ig} - \sum_{j \in N(i)} w_{ij} R_{jg} - \sum_{j \in N(i)} w_{ji}(R_{jg} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kg} - I_{ig}\tilde{s}_i = 0 \quad (33)$$

$$2R_{ib} - \sum_{j \in N(i)} w_{ij} R_{jb} - \sum_{j \in N(i)} w_{ji}(R_{jb} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kb} - I_{ib}\tilde{s}_i = 0 \quad (34)$$

$$\left(I_{ir}^2 + I_{ig}^2 + I_{ib}^2 + 2\lambda_s\right)\tilde{s}_i - 2\lambda_s z\left(B_i^s\right) \sum_{k \in B_i^s} \tilde{s}_k - I_{ir}R_{ir} - I_{ig}R_{ig} - I_{ib}R_{ib} = 0 \quad (35)$$

$\forall i \in B^{fix}$:

$$2R_{ir} - \sum_{j \in N(i)} w_{ij} R_{jr} - \sum_{j \in N(i)} w_{ji} R_{jr} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kr} - I_{ir}\tilde{s}_i = 0 \quad (36)$$

$$2R_{ig} - \sum_{j \in N(i)} w_{ij} R_{jg} - \sum_{j \in N(i)} w_{ji} R_{jg} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kg} - I_{ig}\tilde{s}_i = 0 \quad (37)$$

$$2R_{ib} - \sum_{j \in N(i)} w_{ij} R_{jb} - \sum_{j \in N(i)} w_{ji} R_{jb} + \sum_{j \in N(i)} w_{ji} \sum_{k \in N(j)} w_{jk} R_{kb} - I_{ib}\tilde{s}_i = 0 \quad (38)$$

$$\left(I_{ir}^2 + I_{ig}^2 + I_{ib}^2 + 2\lambda_s\right)\tilde{s}_i - I_{ir}R_{ir} - I_{ig}R_{ig} - I_{ib}R_{ib} - \lambda_{fix}\tilde{s}_{fix}. \quad (39)$$

In our implementation, the above equations are solved iteratively by the Gaussian–Seidel method.

REFERENCES

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

[2] H. G. Barrow and J. M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in *Computer Vision Systems*, A. Hanson and E. Riseman, Eds. New York: Academic, 1978, pp. 3–26.

[3] J. M. Rubin and W. A. Richards, "Color vision and image intensities: When are changes material," *Biol. Cybern.*, vol. 45, no. 3, pp. 215–226, Oct. 1982.
[4] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 1988, pp. 305–306.
[5] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[6] B. V. Funt, M. S. Drew, and M. Brockington, "Recovering shading from color images," in *Proc. ECCV*, 1992, pp. 124–132.
[7] P. Sinha and E. Adelson, "Recovering reflectance and illumination in a world of painted polyhedra," in *Proc. IEEE ICCV*, 1993, pp. 156–163.
[8] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.
[9] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.
[10] E. H. Adelson and P. A. Pentland, "The perception of shading and reflectance," in *Perception as Bayesian Inference*, D. Knill and W. Richards, Eds. New York: Cambridge Univ. Press, 1996, pp. 409–423.
[11] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *Proc. IEEE CVPR*, 1997, pp. 731–737.
[12] W. T. Freeman and P. A. Viola, "Bayesian model of surface perception," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1998.
[13] Y. Weiss, "Segmentation using eigenvectors: A unifying view," in *Proc. IEEE ICCV*, 1999, pp. 975–982.
[14] Y. Weiss, "Deriving intrinsic images from image sequences," in *Proc. IEEE ICCV*, 2001, pp. 68–75.
[15] M. Bell and W. T. Freeman, "Learning local evidence for shading and reflectance," in *Proc. IEEE ICCV*, 2001, pp. 670–677.
[16] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *Int. J. Comput. Vis.*, vol. 52, no. 1, pp. 7–23, Apr. 2003.
[17] M. S. Keil, G. Cristobal, and H. Neumann, "Neural mechanisms for segregation and recovering of intrinsic image features," in *Proc. IEEE ICIP*, 2003, pp. 693–696.
[18] Y. Matsushita, S. Lin, S. B. Kang, and H. Y. Shum, "Estimating intrinsic images from image sequences with biased illumination," in *Proc. ECCV*, 2004, pp. 274–286.
[19] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi, "Illumination normalization with time-dependent intrinsic images for video surveillance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 10, pp. 1336–1347, Oct. 2004.
[20] A. Olmos and F. A. A. Kingdom, "A biologically inspired algorithm for the recovery of shading and reflectance images," *Perception*, vol. 33, no. 12, pp. 1463–1473, 2004.
[21] G. D. Finlayson, M. S. Drew, and C. Lu, "Intrinsic images by entropy minimization," in *Proc. ECCV*, 2004, pp. 582–595.
[22] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 689–694, Aug. 2004.
[23] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 294–302, Aug. 2004.
[24] M. W. Powell, S. Sarkar, D. B. Goldgof, and K. Ivanov, "A methodology for extracting objective color from images," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 5, pp. 1964–1978, Oct. 2004.
[25] 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.
[26] M. F. Tappen, E. H. Adelson, and W. T. Freeman, "Estimating intrinsic component images using nonlinear regression," in *Proc. IEEE CVPR*, 2006, pp. 1992–1999.
[27] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 646–653, Jul. 2006.
[28] G. D. Finlayson, S. D. Hordley, C. Lu, and M. Drew, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 59–68, Jan. 2006.
[29] M. Farenzena and A. Fusiello, "Recovering intrinsic images using an illumination invariant image," in *Proc. IEEE ICIP*, 2007, pp. 485–488.
[30] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.
[31] F. A. A. Kingdom, "Perceiving light versus material," *Vis. Res.*, vol. 48, no. 20, pp. 2090–2105, Sep. 2008.
[32] X. Liu, L. Wan, Y. Qu, T. Wong, S. Lin, C. Leung, and P. Heng, "Intrinsic colorization," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 152-1–152-9, Dec. 2008.
[33] A. J. Schofield, G. Hesse, P. B. Rock, and M. A. Georgeson, "Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces," *Vis. Res.*, vol. 46, no. 20, pp. 3462–3482, Oct. 2006.
[34] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *Proc. IEEE CVPR*, 2008, pp. 1–7.
[35] X. An and F. Pellacini, "AppProp: All-pairs appearance-space edit propagation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 40-1–40-9, Aug. 2008.
[36] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground-truth data set and baseline evaluations for intrinsic image algorithms," in *Proc. IEEE ICCV*, 2009, pp. 2335–2342.
[37] Y. Yacoob and L. S. Davis, "Segmentation using appearance of mesostructure roughness," *Int. J. Comput. Vis.*, vol. 83, no. 3, pp. 248–273, Jul. 2009.
[38] A. Bousseau, S. Paris, and F. Durand, "User-assisted intrinsic images," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 130-1–130-10, Dec. 2009.
[39] B. Wang, X. Gao, D. Tao, and X. Li, "A unified tensor level set for image segmentation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 857–867, Jun. 2010.
[40] X. Jiang, A. J. Schofield, and J. L. Wyatt, "Correlation-based intrinsic image extraction from a single image," in *Proc. ECCV*, 2010, pp. 58–71.
[41] J. B. Shen, X. S. Yang, Y. D. Jia, and X. L. Li, "Intrinsic images using optimization," in *Proc. IEEE CVPR*, 2011, pp. 3481–3487.

Jianbing Shen (M'11) received the Ph.D. degree in computer science from Zhejiang University, Hangzhou, China, in 2007.

He is currently an Associate Professor with the School of Computer Science, Beijing Institute of Technology, Beijing, China. He has published more than 30 refereed papers in journals and conference proceedings. His research interests include intrinsic images, texture synthesis, image completion, video abstraction, and high dynamic range image processing.

Xiaoshan Yang is currently working toward the M.S. degree in the School of Computer Science, Beijing Institute of Technology, Beijing, China.

His current research interests include high-quality intrinsic images and shadow removal.

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

Yunde Jia received the Ph.D. degree in mechatronics from the Beijing Institute of Technology, Beijing, China, in 2000.

He is currently a Professor of Computer Science and also a Director of the Laboratory of Media Computing and Intelligent Systems, School of Computer Science, Beijing Institute of Technology. His research interests include computer vision, media computing, human–computer interaction, and intelligent systems.