

Intrinsic Image Decomposition Using a Sparse Representation of Reflectance

Li Shen, *Member, IEEE*, Chuohao Yeo, *Member, IEEE*, and Binh-Son Hua

Abstract—Intrinsic image decomposition is an important problem that targets the recovery of shading and reflectance components from a single image. While this is an ill-posed problem on its own, we propose a novel approach for intrinsic image decomposition using reflectance sparsity priors that we have developed. Our sparse representation of reflectance is based on a simple observation: neighboring pixels with similar chromaticities usually have the same reflectance. We formalize and apply this sparsity constraint on local reflectance to construct a data-driven second-generation wavelet representation. We show that the reflectance component of natural images is *sparse* in this representation. We further propose and formulate a global sparse constraint on reflectance colors using the assumption that each natural image uses a small set of material colors. Using this sparse reflectance representation and the global constraint on a sparse set of reflectance colors, we formulate a constrained ℓ_1 -norm minimization problem for intrinsic image decomposition that can be solved efficiently. Our algorithm can successfully extract intrinsic images from a single image, without using color models or any user interaction. Experimental results on a variety of images demonstrate the effectiveness of the proposed technique.

Index Terms—Intrinsic image decomposition, Sparse reconstruction, Multi-resolution analysis

1 INTRODUCTION

INTRINSIC image decomposition addresses the problem of separating an image into its reflectance and shading components. This decomposition of intrinsic images is of importance in both computer graphics and computer vision applications. First, the intrinsic decomposition facilitates advanced image editing in graphics applications such as re-texturing, re-colorization and re-lighting. Second, the extracted intrinsic images benefit many computer vision algorithms. Shading images are preferred inputs to algorithms such as shape from shading while reflectance images can be used for tasks such as segmentation and image white balance. Furthermore, most vision algorithms from low-level image analysis to high-level object recognition implicitly assume that its input image is a reflectance image.

Typically, in intrinsic image decomposition, an input image \mathcal{I} is modeled as a per-color-channel product of a reflectance component \mathcal{R} and a shading (or illumination) component \mathcal{L} , and the aim is to decompose \mathcal{I} into \mathcal{R} and \mathcal{L} . In this paper, we process these components in the log domain. Denote by I , R and L the logarithms of \mathcal{I} , \mathcal{R} and \mathcal{L} , respectively. Thus, we are given:

$$I = R + L$$

and wish to recover R and L . Therefore, recovering the two intrinsic components from a single input image remains a challenging problem because of its severely

ill-posed nature: given an input image that is composed from its reflectance and shading components, the number of unknowns is twice the number of equations. To solve this problem, further constraints are needed.

In this paper, we propose two novel priors on reflectance for single image intrinsic image decomposition. Our approach is based on the following two simple observations:

- Two neighboring pixels that share similar chromaticities are likely to have similar reflectances.
- Natural images are usually dominated by a small set of material colors.

The first observation describes a local sparseness on reflectance; similar local sparseness constraints have been used in previous methods such as [1], [2]. From this observation on local reflectance, we apply multi-resolution analysis (MRA) to construct a new data-driven *second-generation wavelet* representation [3] of reflectance, so as to convert what appears to be a local constraint into a global constraint. We show that the reflectance component of natural images is *sparse* in such a representation, which leads to our first new prior, i.e., a global sparse representation of reflectance. Using this wavelet representation of reflectance, we formulate a constrained ℓ_1 -norm minimization problem for intrinsic image decomposition to solve for the reflectance component. The decomposition produced by our method is therefore the global optimum of a convex optimization problem.

The second observation, that the set of reflectance spectra in a natural image is sparse, draws from the work of Omer and Werman [4]. It leads to our second new prior which is formulated as a global sparsity constraint on the set of reflectance colors that can be integrated into

- L. Shen and C. Yeo are with the Institute for Infocomm Research, Singapore. E-mail: {lshen, chyeyo}@i2r.a-star.edu.sg.
- B. Hua is with Department of Computer Science, National University of Singapore University. E-mail: huabinhs@comp.nus.edu.sg.

the earlier constrained ℓ_1 -norm minimization problem. We show that using this prior can improve the recovery of the global structures of shading and reflectance, which in turn leads to further improvements in our intrinsic image decomposition.

The rest of the paper is organized as follows. Section 2 discusses related work. The new sparse priors are introduced in Section 3, while the optimization framework for performing intrinsic image decomposition using the proposed priors is described in Section 4. Section 5 presents experimental results on various test images. Finally, concluding remarks are presented in Section 6.

2 RELATED WORK

The problem of intrinsic image decomposition into reflectance and shading components was first introduced by Barrow and Tenenbaum [5]. The reflectance component describes the intrinsic albedo of a surface, which is illumination-invariant. The shading component corresponds to the amount of reflected light from the surface, which depends on surface geometry, reflection function and illumination condition.

Some previously proposed methods use additional information from multiple images to resolve the inherent ambiguities. For example, user registered images captured under different illumination conditions can be used [6], [7], [8]. The approach by Troccoli and Allen [9] used a laser scan of the scene and multiple lighting and viewing conditions to perform relighting and to estimate reflectance.

To overcome the severely ill-posed nature of the problem, previous methods for intrinsic image decomposition from a single image used either a strong prior or assumption. Using the Retinex strategy, local derivatives can be analyzed in order to distinguish between shading induced and reflectance induced image variations [1], [2], [10], [11], [12]. Training-based approaches have also been proposed to classify image derivatives into reflectance changes or shading changes [13], [14], [15]. With trained classifiers, Tappen et al. obtained good decomposition results from a single image by solving a global optimization problem with belief propagation [14]. A major drawback of these previous methods is that the decomposition is analyzed locally within a small window. One exception is the work of Shen et al. [16] which proposed a global optimization algorithm incorporating both the Retinex constraint and non-local texture constraint to obtain global consistency of image structures.

More recently, a user-assisted method has been proposed by Bousseau et al. [17]. Focusing on diffuse objects, they used the assumption that local reflectance colors lie on a plane and derived a closed-form least squares system which can be solved together with additional user-supplied constraints. Their method obtained impressive results on the presented test images. However, the method requires precise user strokes and their “color plane” assumption on local reflectance values is

incompatible with many practical cases such as multi-color surfaces and gray-scale input images.

In contrast, our priors are independent of color models on local surfaces. Furthermore, by using the two new global sparse priors on reflectance, the proposed method in this paper can automatically recover the intrinsic images from a single image without additional information.

Our method is partially inspired by the work of Fattal et al. [18] on the construction of data-dependent second-generation wavelets for edge-preserving image processing. Different from first-generation wavelets consisting of translates and dilates of a single pair of scaling and wavelet functions, second-generation wavelets allow them to change according to spatial particularities of the data. The *Lifting scheme* first introduced by Sweldens [3] is an efficient implementation of the fast wavelet transform for constructing bi-orthogonal wavelets through space. Fattal et al. [18] proposed the edge-avoiding wavelets (EAW) constructed using a data-prediction lifting scheme based on the edge content of the input image. In this paper, we utilize the lifting scheme [3] to construct a new data-dependent MRA based on the local reflectance sparseness using the chromaticity information.

3 SPARSE PRIORS ON REFLECTANCE

In this section, we show how to derive the proposed sparse representation of the reflectance component of natural images from a simple local constraint on reflectance and formulate the global sparsity constraint of reflectance based on the representation. We also present the sparse prior on reflectance spectra and show how to use that prior by introducing a total-variations-like cost term.

3.1 Sparse Reflectance Representation

3.1.1 Local Sparseness of Reflectance

Our method is based on an observation of a local sparseness of reflectance, where neighboring pixels of similar chromaticity have similar reflectance. We can exploit this observation to build a local sparse representation of reflectance by minimizing the following cost function:

$$J(R) = \sum_i \left\| R(i) - \sum_{j \in N_i} \widehat{w}_{ij} R(j) \right\|_1, \quad (1)$$

where $R(i)$ is the RGB vector that represents the reflectance of pixel i and N_i is the set of neighboring pixels of i . \widehat{w}_{ij} is a set of normalized non-negative weights which sum to one. This weight should be large when two neighboring chromaticities are similar, and small when they are different.

The normalized weight, \widehat{w}_{ij} , is derived from a weighting function, w_{ij} . We define w_{ij} based on the difference

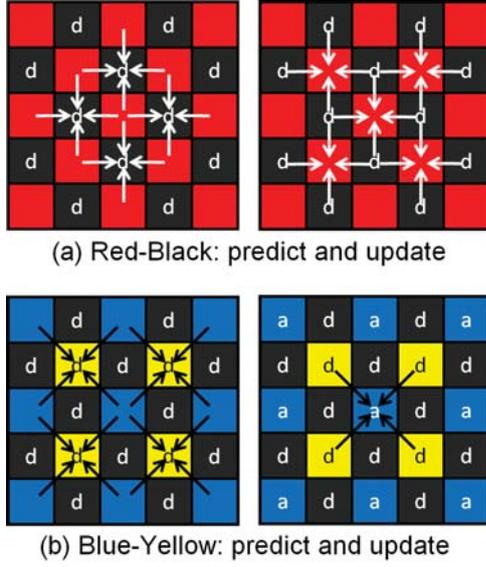


Fig. 1. RBW construction; the “a” and “d” labels in the boxes shows the locations of the approximation and detail coefficients respectively. (a) illustrates the horizontal/vertical lifting in the Red-Black stage. (b) illustrates the diagonal lifting in the Blue-Yellow stage.

between two neighboring chromaticities normalized by a local chromaticity variance:

$$w_{ij} = \begin{cases} 0 & \text{if } |C(i) - C(j)| > t_c, \\ \exp\left(-\frac{\|C(i) - C(j)\|^2}{0.3\sigma_i^2}\right) & \text{otherwise.} \end{cases} \quad (2)$$

where $C(i) = R(i)/\|R(i)\|$ is the chromaticity of the reflectance at i , and σ_i^2 is the average chromaticity variance across color channels in a local window of 5×5 pixels and clipped such that it has a minimum value of 10^{-4} . t_c is a threshold of chromaticity difference. We set t_c to a small value ($t_c = 0.02$ in our experiments) so that the weight only takes effect when chromaticities are very similar, otherwise there is no dependence between the pixels. For natural images, we can assume that there always are neighboring pixels around a pixel that have similar reflectance, i.e., $\sum_{j \in N_i} w_{ij} > 0$. Later, we will discuss the special case that all the neighboring pixels have very different reflectance in Section 3.1.2.

Similar weighting functions based on intensity values are used widely in image segmentation (e.g., [20], [21]) and colorization (e.g., [22], [23]) algorithms, where they are usually referred to as *affinity functions*. In a twist from previous methods, we use this formulation on chromaticity values to enforce the local sparsity of reflectance.

3.1.2 Global Sparseness of Reflectance using Multi-Resolution Analysis

To enforce the local reflectance sparseness constraint introduced in Section 3.1.1 at a global level, we use a multi-resolution analysis approach. We construct the MRA using a data-prediction lifting scheme based on the

-
- 1: **Input:** Input image
 - 2: **Output:** $a^K, \{d^k\}_{k=1}^K$
 - 3: **initialize:** $a^0 \leftarrow$ Input Image
 - 4: **for** $k \leftarrow 0$ **to** $K - 1$ **do**
 - 5: **Do Red-Black Stage begin**
 - 6: **Split:** Decompose a^k into the coarse data, $a_{C_{rb}}^k$, at locations C_{rb}^k (red) and fine data, $a_{F_{rb}}^k$, at locations F_{rb}^k (black)
 - 7: **Predict:** Use $a_{C_{rb}}^k$ to predict $a_{F_{rb}}^k$
 - 8: **for each** $i \in F_{rb}^k$ **do**
 - 9:
$$d^{k+1}(i) \leftarrow a_{F_{rb}}^k(i) - \sum_{j \in N_i} \widehat{w}_{ij}^k a_{C_{rb}}^k(j) \quad (3)$$
 - 10: **end for**
 - 11: **Update:** Use d^{k+1} to update $a_{C_{rb}}^k$
 - 12: **for each** $i \in C_{rb}^k$ **do**
 - 13:
$$a_{C_{rb}}^k(i) \leftarrow a_{C_{rb}}^k(i) + \frac{1}{2} \sum_{j \in N_i} \widehat{w}_{ij}^k d^{k+1}(j) \quad (4)$$
 - 14: **end for**
 - 15: **end**
 - 16: **Do Blue-Yellow Stage begin**
 - 17: **Split:** Decompose $a_{C_{rb}}^k$ into coarse data, $a_{C_{by}}^k$, at locations C_{by}^k (blue), and fine data, $a_{F_{by}}^k$, at locations F_{by}^k (yellow)
 - 18: **Predict:** Use $a_{C_{by}}^k$ to predict $a_{F_{by}}^k$
 - 19: **for each** $i \in F_{by}^k$ **do**
 - 20:
$$d^{k+1}(i) \leftarrow a_{F_{by}}^k(i) - \sum_{j \in N_i} \widehat{w}_{ij}^k a_{C_{by}}^k(j)$$
 - 21: **end for**
 - 22: **Update:** Use d^{k+1} to update $a_{C_{by}}^k$
 - 23: **for each** $i \in C_{by}^k$ **do**
 - 24:
$$a_{C_{by}}^k(i) \leftarrow a_{C_{by}}^k(i) + \frac{1}{2} \sum_{j \in N_i} \widehat{w}_{ij}^k d^{k+1}(j)$$
 - 25: **end for**
 - 26: **end**
 - 27: **end for**
-

Fig. 2. Lifting scheme of forward weighted red-black wavelet transform

chromaticity configurations of the input image. Following [18], we utilize the *red-black wavelets* (RBW) which is a lifting-based second-generation wavelet on rectangular grids introduced by Uytterhoeven et al. [19].

RBW is a two-step lifting construction for 2D signals that uses the quincunx lattices illustrated in Fig. 1. The pixels are first split into the red and black subsets as in Fig. 1. Each black pixel is predicted using the four nearest red pixels, and the computed detail pixels, d^{k+1} , are stored at the black pixels. Then, the red pixels are

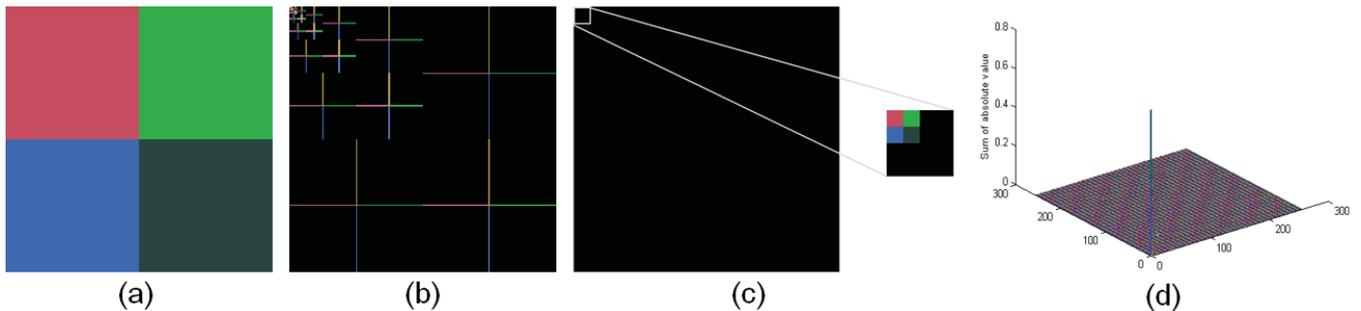


Fig. 3. (a) A synthetic image with 4 homogeneous regions. (b) RBW transform with equal weighting. (c) Proposed WRBW transform; note that all the coefficients are 0 except for the top left corner as shown in the zoomed-in box. (d) A 3D plot of the WRBW coefficients (across RGB).

updated using the computed detail coefficients stored at the black pixel locations. The updated red pixels are decomposed further into the blue and yellow subsets as shown in Fig. 1. The yellow pixels are predicted using their four diagonally-located neighbors at the blue pixel locations, and the computed detail pixels, d^{k+1} , are stored at the yellow pixels. Finally, the blue pixels are updated using the computed detail coefficients stored at the yellow pixel locations, and the computed approximation coefficients, a^{k+1} , are stored at the blue pixels. Fig. 2 shows the lifting scheme of the forward transform of the weighted red-black wavelets (WRBW). By inverting each of these lifting steps, an image can be reconstructed from the wavelet coefficients. We perform $K = \lfloor \log_2(\min(w, h)) \rfloor$ levels of decomposition, where w and h are respectively the width and height of the image.

The predict and update steps of the Red-Black stage are defined by Equation (3) and Equation (4) respectively; the predict and update steps of the Blue-Yellow stage are similar and only differ in the neighborhood used. The multi-scale weights, \widehat{w}_{ij}^k , are normalized from the weights computed using Equation (2) with the chromaticity information at every scale. At coarser scales, the neighboring chromaticities around a pixel might be significantly different; for such pixels, the normalized weights, \widehat{w}_{ij}^k , are set to zero. It is interesting to note that proposed set of wavelet weights actually contains information about the chromaticity configurations of the image at every scale.

By using the weighted scheme, the proposed wavelets are designed with a support that is biased towards neighboring pixels with similar chromaticity values. At the predict step, the prediction operation is the same as the term being summed in Equation (1). With the weights used, the prediction of each reflectance value would be weighted more towards neighboring reflectance with similar chromaticity values leading to most of the detail coefficients, $\{d^k\}_{k=1}^K$, being zero or close to zero except at pixels where $\sum_{j \in N_i} w_{ij}^k = 0$. At the update step, instead of merely preserving the approximation average, the update of each reflectance value is again weighted more towards neighboring re-

flectance with similar chromaticity values. This can thus be regarded as a chromaticity distribution preserving down-sampling that attempts to keep local reflectance values as close to each other as possible at each scale. Overall, the WRBW transform is expected to lead to sparse reflectance components due to the combination of chromaticity distribution preserving down-sampling and the chromaticity-based weighted prediction.

Fig. 3 illustrates the sparse nature of the proposed WRBW representation for an image satisfying the local sparseness constraint, where we use a synthetic image with 4 homogeneous color regions with different chromaticities. The detail coefficients are zero where the input image is flat, resembling the transform results by the original RBW. Near the edges, since the proposed wavelets are designed with a support that avoids containing both the edge and the pixels with different chromaticities, the wavelets response to such edges diminishes. For this synthetic image, the coefficients obtained by the proposed WRBW transform are all zero except the four approximation coefficients at the coarsest level a^K . Compared to the RBW coefficients, our WRBW coefficients show a stronger sparsity.

3.2 Sparse prior on reflectance component

We formulate a global sparse constraint on the reflectance component of natural images by using the multi-scale representation described in Section 3.1.

We denote the WRBW forward transform operator by \mathcal{B}_w , and the backward transform operator by \mathcal{B}_w^{-1} . Then, the reflectance component of a natural image can be represented in the wavelet domain as:

$$\widehat{R} = \mathcal{B}_w R$$

where \widehat{R} are the wavelet coefficients of the reflectance.

Recall from Equation (2) that when the chromaticities of the neighboring pixels around a pixel are significantly different, $w_{ij} = 0$ for all neighbors; therefore, from Equation (3), $\widehat{R}(i)$ stores the actual reflectance value at that location and scale. When carrying out the initial wavelet decomposition, we keep track of this set of locations, Γ , where the chromaticities of neighboring pixels are

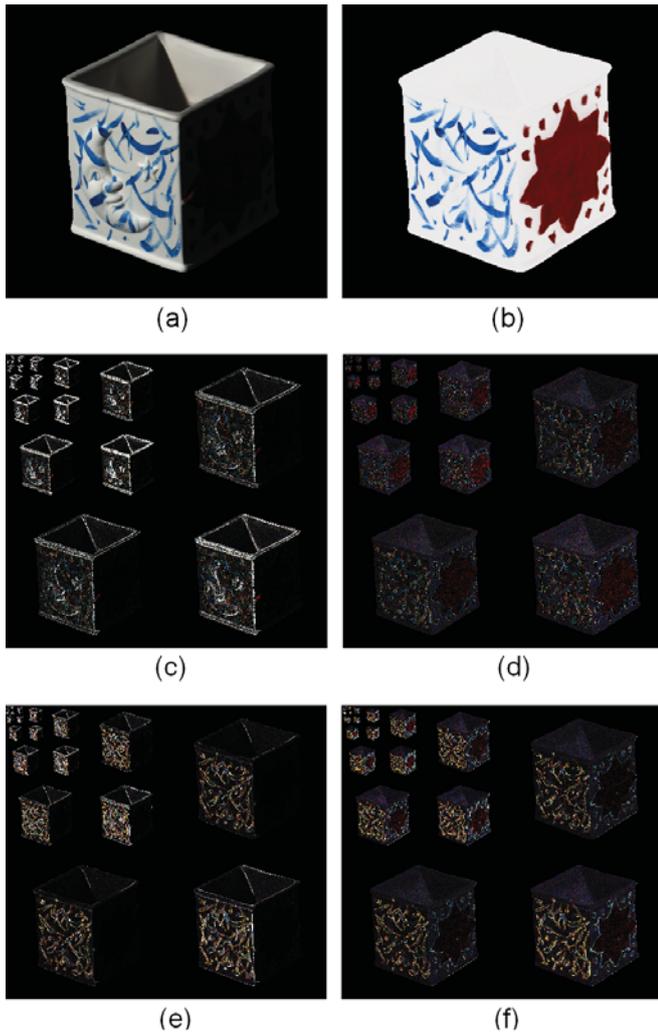


Fig. 4. Illustrate the physical meaning of the proposed sparse prior on reflectance $\Lambda\hat{R}$ for a natural image. (a) Original image. (b) Reflectance image. (c) and (d) $\Lambda\hat{I}$ and $\Lambda\hat{R}$ obtained by the proposed WRBW, respectively. (e) and (f) $\Lambda\hat{I}$ and $\Lambda\hat{R}$ obtained by the EAW transform, respectively

significantly different. For convenience, we will denote the complement of Γ as $\bar{\Gamma}$. For coefficients not in Γ , i.e., in $\bar{\Gamma}$, the neighboring pixels have similar chromaticities, and $\hat{R}(i)$ stores the local reflectance difference.

Assuming that the observation of local reflectance sparsity holds, most of the coefficients of $\hat{R}(i)$ in $\bar{\Gamma}$ should be zero or close to zero. Hence, we formulate the sparse constraint on reflectance component by minimizing the following cost term:

$$E_{sr} = \left\| \Lambda \hat{R} \right\|_1, \quad (5)$$

where Λ is a diagonal weighting matrix for $\hat{R}(i)$ with diagonal entries given by:

$$\Lambda_{i,i} = \begin{cases} 1 & i \in \bar{\Gamma}, \\ \varepsilon & \text{otherwise,} \end{cases}$$

where ε is a small value we set as 10^{-5} in our implementation.

To illustrate the physical meaning of the proposed sparse prior on reflectance for a natural image, we used the “box” example from the MIT intrinsic image database [24]. This is shown in Fig. 4. We perform the WRBW on both the original image and the reflectance image using the same set of the weights that was computed using the *chromaticity* information of the original image. The WRBW coefficients obtained from the original and reflectance images, $\Lambda\hat{I}$ and $\Lambda\hat{R}$, respectively, are shown in Fig. 4(c) and (d). For comparison, we show the $\Lambda\hat{I}$ and $\Lambda\hat{R}$ obtained by using the EAW transform proposed by Fattal [18]. The EAW transform is similar to the proposed WRBW transform except that their weights are computed from the color intensities of the original image while our weights are computed using the chromaticity information based on the local reflectance sparseness. We can see that $\Lambda\hat{I}$ obtained by our WRBW transform actually contains the shadow information of the original image. However, the EAW coefficients $\Lambda\hat{I}$ have not such a nature.

3.3 Sparse prior on reflectance spectra

The second prior comes from an additional constraint that the total number of reflectance values (or colors) is small within each image. Omer and Werman [4] have shown that scenes are dominated by a small number of material colors. In other words, the set of reflectance spectra is sparse.

We formulate the constraint on having a sparse set of reflectance spectra by applying a total variations-like cost on the set of reflectance coefficients within the image in Γ (see Section 3.2). We denote the cardinality of Γ by $M = |\Gamma|$. Let \mathcal{T} be the operator that computes all $\frac{M \times (M-1)}{2}$ differences between the reflectance values found in the locations stored in Γ . In other words, \mathcal{T} is a sparse matrix, with one row for each possible combination of indices in Γ that corresponds to computing the difference between the reflectance values at the indices. For example, if the k th combination is between index i and j , then $\mathcal{T}_{ki} = 1$ and $\mathcal{T}_{kj} = -1$. We would use the prior of $\mathcal{T}\hat{R}$ being sparse by minimizing the following cost term:

$$E_{sc} = \left\| \mathcal{T} \hat{R} \right\|_1. \quad (6)$$

To reduce the number of operations when computing the term $\mathcal{T}\hat{R}$, we trim the number of entries by removing the constraint for pairs of coefficient positions in Γ that are likely to have different reflectance values. We do so by first computing the forward weighted RBW on the original image. Then, a constraint between 2 locations i and j in Γ is added only if the square difference between the normalized coefficient values (across color channels) is smaller than a threshold, t_w . In our experiments, we use $t_w = 10^{-4}$.

4 INTRINSIC IMAGE DECOMPOSITION

In this section, we will show how we can use the two sparse priors introduced in Section 3 for intrinsic image decomposition by formulating an appropriate optimization problem. We also present a method for further refinement of the decomposition using a matting-based approach.

4.1 Optimization

Assuming that the illumination component changes smoothly over the scene, we can apply a smoothness constraint on L by adding a Laplacian-based cost at all locations:

$$E_{\text{smooth}} = \sum_i \|\Delta L(i)\|_2$$

where Δ denote the Laplacian operator. This smoothness regularization of L both ensures that every pixel has an equation that constrains it and controls the smoothness of the illumination component.

We substitute L by $I - R$ and express this smoothness constraint on the illumination with the following cost term:

$$\begin{aligned} E_{\text{smooth}} &= \|\Delta L\|_2^2 \\ &= \|\Delta I - \Delta \mathcal{B}_w^{-1} \widehat{R}\|_2^2. \end{aligned}$$

The smoothness constraint can be considered to be a set of measurements on the reflectance coefficients, i.e.,

$$y \approx A\widehat{R}$$

where

$$A = \Delta \mathcal{B}_w^{-1} \text{ and } y = \Delta I \quad (7)$$

If surfaces in the scene are diffuse or near-diffuse, we can assume that the input image chromaticity is the same as the reflectance chromaticity. The weights of the WRBW transform are thus computed according to Equation (2) using the chromaticity of the input image.

To recover \widehat{R} , we would solve the following constrained ℓ_1 -norm minimization problem by using the sparse reflectance representation prior from Equation (5) together with the smoothness constraint on illumination:

$$\min_{\widehat{R}} \|\Lambda \widehat{R}\|_1 \text{ s.t. } A\widehat{R} = y$$

This optimization problem can be solved using an ℓ_1 -regularized least-squares solver, e.g., [25], [26] by rewriting the optimization problem as:

$$\min_{\widehat{R}} \left\| A\widehat{R} - y \right\|_2^2 + \lambda \|\Lambda \widehat{R}\|_1 \quad (8)$$

where λ is a regularization parameter.

Further including the sparse prior on reflectance spectra from Equation (6), we obtain the following optimization:

$$\min_{\widehat{R}} \left\| A\widehat{R} - y \right\|_2^2 + \lambda \|\Lambda \widehat{R}\|_1 + \mu \|\mathcal{T}\widehat{R}\|_1, \quad (9)$$

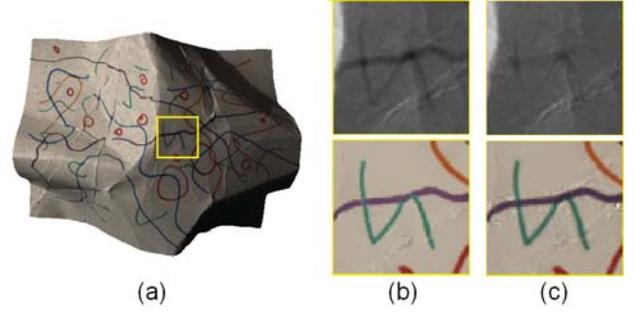


Fig. 5. Separation results illustrating soft matting refinement for “paper2” image. (a) Original image. (b) Before refinement (zoom-in of yellow box). (c) After refinement (zoom-in of yellow box).

where μ is an additional regularization parameter.

We note here that we can take advantage of the fact that the A and A^T operators can be implemented efficiently without the need to perform full matrix multiplication. The inverse WRBW transform, \mathcal{B}_w^{-1} , can be computed using wavelet lifting, while the inverse dual WRBW transform, \mathcal{B}_w^{-1T} , can also be computed using wavelet lifting by switching the order of the predict and update steps and manipulating the weights used. Moreover, the Laplacian operator, Δ , can be implemented as an image filter.

4.2 Soft matting

Small changes in the reflectance component that are co-located with those in the chromaticity component, which could be caused by phenomena such as color bleeding, could be wrongly assigned to the shading component since the local color sparsity constraint described in Section 3.1.1 is no longer valid. Here, we apply a refinement process to solve this problem.

We first express each intrinsic component as the product between a scalar intensity, $r = \|R\|$ or $l = \|L\|$, and a chromaticity, $R_c = R/r$ or $L_c = L/l$:

$$I = rR_c + lL_c.$$

Denoting $\alpha = r/(r+l)$, we express the intensity value at each pixel as a mixture of two values weighted by α :

$$I = \alpha \overline{R_c} + (1 - \alpha) \overline{L_c}$$

where $\overline{R_c} = (r+l)R_c$ and $\overline{L_c} = (r+l)L_c$. Therefore, we can apply a closed-form framework of matting [27] to refine the separation. To do so, we first perform an initial decomposition by solving one of the two optimization problems presented earlier in Eqns. (8) and (9). We then compute an initial value of α , denoted by $\tilde{\alpha}$, at the pixels from edges in the decomposed image, and propagate α on those edges using the matting Laplacian algorithm of Levin et al. [27]. Rewriting $\alpha(x)$ and $\tilde{\alpha}(x)$ in their vector forms, we minimize the following cost function:

$$J(\alpha) = \alpha^T \Sigma \alpha + (\alpha - \tilde{\alpha})^T G (\alpha - \tilde{\alpha})$$

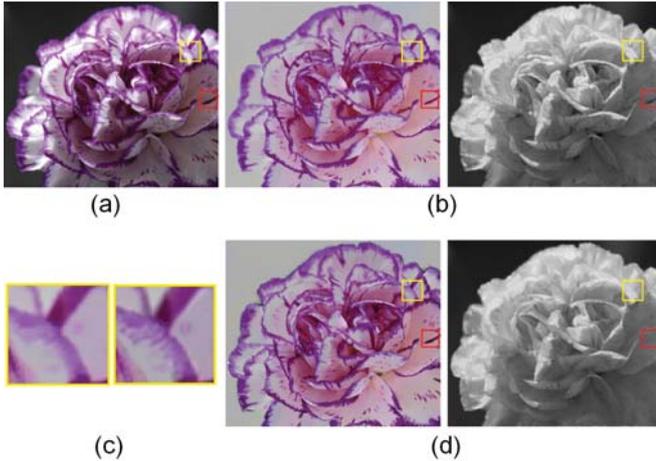


Fig. 6. Separation results illustrating soft matting refinement for a flower image. (a) Input image. (b) Separation results before matting refinement. (d) Separation results after matting refinement. (c) Zoomed-in reflectance results within the yellow box (Left: before refinement; Right: after our matting refinement).

where G is a diagonal matrix of weights. We set $G_{ii} = 0$ when pixel i is at an edge, and $G_{ii} = 100$ otherwise. The matrix Σ is the matting Laplacian matrix [27]. The optimal α can be obtained by solving the following sparse linear system: $(\Sigma + G)\alpha = G\tilde{\alpha}$.

The derivation of the matting Laplacian matrix in [27] is based on a color line assumption, i.e, within a small window, foreground (background) colors lie on a straight line in color space. Since this assumption still holds for natural shading/reflectance images, it is also valid to use the matting Laplacian matrix in intrinsic images. Fig. 5 shows the results of matting refinement for the “paper2” example. We can see that the “ghost” markings in the shading component are reduced. Fig. 6 shows the refined results for a flower image. As we can see, the “ghost” markings in the shading component are reduced, such as that within the red rectangle. Fig. 6(c) shows the zoomed-in reflectance component within the yellow rectangle where the block artifacts in the reflectance are suppressed after applying the matting refinement.

5 EXPERIMENTAL RESULTS

In this section, we provide various experimental validation of the proposed method. We first evaluate the performance of our method on a benchmark dataset with known ground truth [24]. Then, we compare our method with the user-assisted approach of [17].

In the experiments¹, we test two variations of the proposed intrinsic image decomposition algorithm. First, we only use the sparsity constraint on the reflectance representation in the algorithm which solves (8); this

1. In the paper, all the separation results are shown in *AI γ amma* with gamma correction = 1, and A is a scale

TABLE 1
LMSE for CR, SR and SRC over the MIT Intrinsic Images dataset

Example	LMSE		
	CR	SR	SRC
box	0.013	0.0036	0.0018
cup1	0.007	0.0043	0.0030
cup2	0.011	0.0052	0.0045
deer	0.041	0.0413	0.0419
dinosaur	0.035	0.0317	0.0216
frog1	0.066	0.0558	0.0483
frog2	0.071	0.0587	0.0472
panther	0.011	0.0075	0.0078
paper1	0.004	0.0019	0.0014
paper2	0.004	0.0027	0.0021
raccoon	0.015	0.0052	0.0048
sun	0.003	0.0024	0.0023
squirrel	0.072	0.0856	0.0794
teabag1	0.032	0.0280	0.0280
teabag2	0.023	0.0151	0.0141
turtle	0.069	0.0349	0.0174
Average	0.030	0.0240	0.0204

will be referred to as SR. Then, we use both the global sparsity constraints on the reflectance representation and reflectance colors which solves (9); this will be referred to as SRC. In our implementation, we use the fast NESTA method [26] for both SR and SRC to solve the constrained ℓ_1 -norm minimization problem.

5.1 Benchmarking Results on MIT Intrinsic Images Dataset

A benchmark dataset with ground-truth (GT) was presented in [24] for performance evaluation of intrinsic image algorithms. We test our methods, SR and SRC, with this dataset. Following [24], we use local mean squared error (LMSE) from the ground truth to measure decomposition quality. We compare with the conventional color Retinex algorithm (CR) [12], which performed best among single image based methods in the study of [24]. All the separation results of our methods here are before the refinement process. The LMSE values² used for comparisons are computed using the color retinex algorithm made available by the MIT Intrinsic Images dataset [24].

This dataset contains three categories: artificially painted surfaces, printed objects, and toy animals. We display one example from each category in Table 2. With conventional Retinex constraints, pixels that contain significant reflectance derivatives should be smooth in shading. Using the local constraint, the CR method correctly identifies most of the markings as reflectance changes. However, it leaves some “ghost” markings in the shading and some residues of the cast shadows in the reflectance images because the sharp edges contain a mixture of large and small image radiances. In contrast, SR eliminate many of these ghost because by using multi-resolution analysis, our method enforces the

2. Some of the computed values could be slightly different from that presented in [24] because of the convergence algorithm.

TABLE 2

Decomposition results by Color Retinex and our proposed methods on three images from the MIT intrinsic dataset

Example	Raccoon	Cup1	Turtle
Input			
GT			
CR	 <i>LMSE = 0.015</i>	 <i>LMSE = 0.007</i>	 <i>LMSE = 0.069</i>
SR	 <i>LMSE = 0.0052</i>	 <i>LMSE = 0.0043</i>	 <i>LMSE = 0.035</i>
SRC	 <i>LMSE = 0.0048</i>	 <i>LMSE = 0.0030</i>	 <i>LMSE = 0.0174</i>

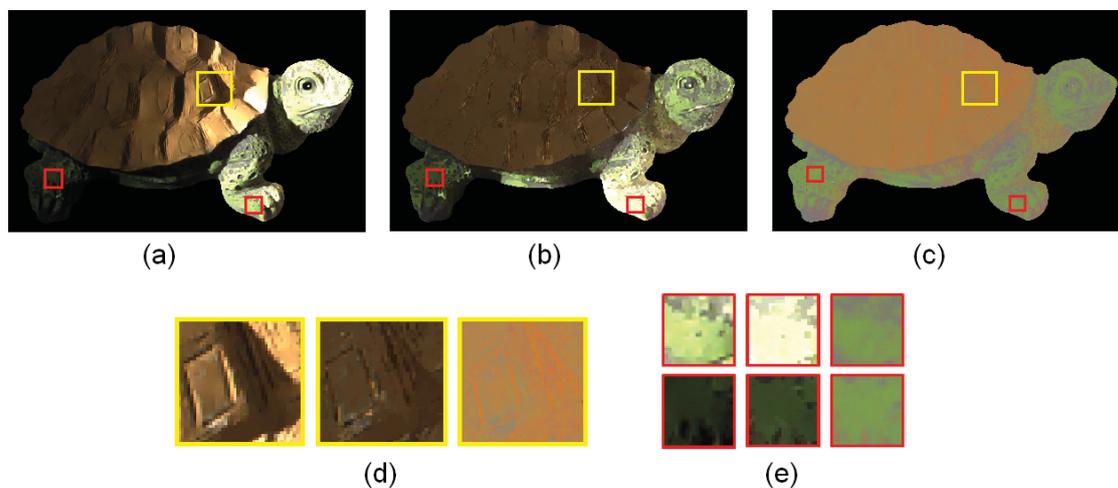


Fig. 7. Reflectance recovered using CR, SR and SRC on the "turtle" image. (a) CR. (b) SR. (c) SRC. (d) Zoom-in of yellow patch for CR, SR and SRC (left to right). (e) Zoom-in of red patches for CR, SR and SRC (left to right).

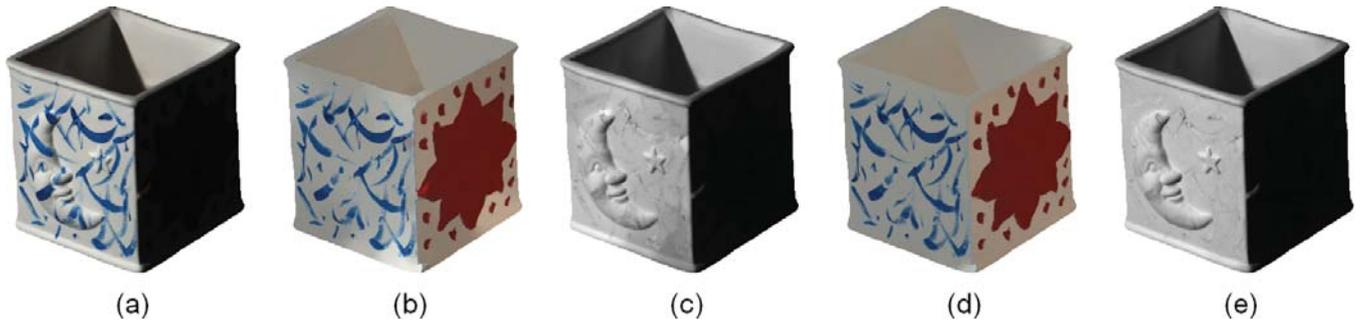


Fig. 8. Intrinsic image decomposition results for the “box” image. (a) Input image. (b-c) Separation results using SR (LMSE = 0.003606). (d-e) Separation results using SRC (LMSE = 0.001835)

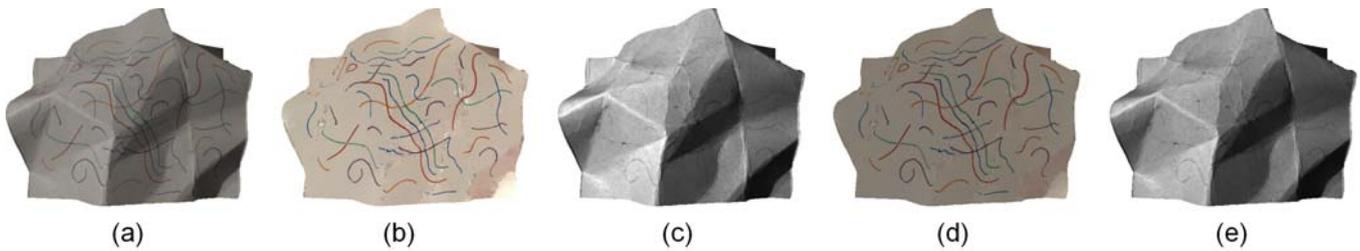


Fig. 9. Intrinsic image decomposition results for the “paper1” image. (a) Input image. (b-c) Separation results using SR (LMSE = 0.001871). (d-e) Separation results using SRC (LMSE = 0.001395)

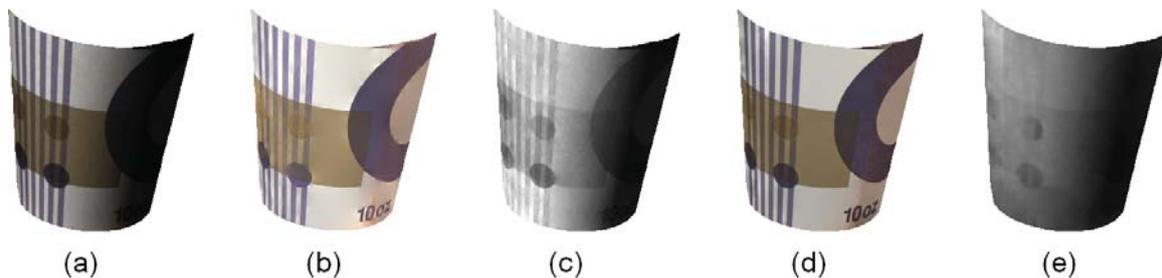


Fig. 10. Example of failure of the local sparseness of reflectance assumption in the “cup2” image. Note that in this case, there exists intensity change with constant hue which corresponds to a change in reflectance and not shading. (a) Input image. (b-c) Separation results of CR method (LMSE = 0.011). (d-e) Separation results of our SRC method (LMSE = 0.0045)

sparse constraint on neighboring reflectance at every scale. The constraint on the multi-resolution representation broadens the influence of local cues to help resolve the ambiguous local inferences.

The “turtle” image in Table 2 is challenging for the CR method. The shell of the turtle exhibits big variations in shading and shadows that arise from the 3D weave pattern. With only local cues, CR misses much of the global and local shading structure in the recovered shading image because the algorithm misinterprets many image gradients as purely reflectance changes due to the large color differences. In contrast, SR can better handle the gradual shading change across the image as well as the local shading variations, and accurately recovers the shape of the shell surface. This difference can be more clearly seen in the closeup of a small shell region in Fig. 7(d).

To illustrate the benefit of the global sparsity constraint on reflectance color, we also compare the results obtained with the SRC method in Fig. 7. As shown in Fig. 7, the forequarter and hindquarter of the turtle are two distinct regions. In the decomposition with SR and CR, shading and reflectance in each of these regions are computed separately, which results in recovered reflectances that are inconsistent, as seen in Fig. 7(a) and (b). With the non-local sparse constraint on reflectance colors, the recovered reflectance with SRC has a smaller set of reflectance values, which leads to a more consistent decomposition as shown in Fig. 7(c). This can be seen more clearly for a closeup of the small regions on the two feet Fig. 7(e). With this global constraint on reflectance colors, the SRC method can correctly recover the global shading and reflectance structure that cannot easily be inferred using local cues alone.

Fig. 8 and 9 show two other examples where SRC effectively eliminates cast shadows on the surfaces from the reflectance. Chromaticity values might change in dark regions caused by cast shadows. Since the proposed WRBW is designed with a support that avoids pixels with different chromaticities, the wavelets response to such edges would be diminished. In the decomposition with SR, shading and reflectance in these shadow regions are thus computed separately from the neighboring regions, which results in the inconsistent reflectances as shown in Fig. 8(b) and Fig. 9(b). With the sparse constraint on reflectance colors, the reflectance values recovered by SRC in these regions are more similar to the ones which are not in shadows. The proposed SRC method can better deal with this problem, and more accurately removes the cast shadow found inside the box in Fig. 8(d) and the shadow at upper right in Fig. 9(d).

Quantitative comparisons on all the dataset images are provided in Table 1 where we compared the LMSE of CR and our proposed methods, SR and SRC. The SR method outperforms CR for most of the objects and the SRC method generally has the best performance. However, CR outperforms our proposed methods on a few examples, “deer”, and “squirrel”. This is a result of our assumption on the local sparseness of reflectance being invalid. Fig. 10 exemplifies the problem with the proposed methods on the “cup2” image. There are some places on the cup surface where neighboring pixels with similar chromaticities have different reflectance, and that is where our methods fail to properly separate reflectance and shading. For the cup2 example, even though the local sparsity prior is invalid for these places, the separation results of our method are still better than the ones of the color retinex method.

5.2 Comparison with user-assisted approaches

Here, we compare our method with that of Bousseau et al. [17], which uses the following global constraints provided by a user: sets of pixels with similar reflectance, sets of pixels with similar illumination, and locations and shading values of pixels with known illumination.

Accurate decomposition results can be achieved by using the global constraints of shading and reflectance provided in the form of user scribbles. However, users may not always provide useful scribbles. Fig. 11(a) shows the decomposition results with ground truth scribbles, which has a LMSE of 0.00055. To simulate the effect of having inaccurate scribbles, we used scribbles with the same fixed illumination values as before but with positions that are randomly perturbed by up to 15 pixels; the result is shown in Fig. 11(b), with a LMSE of 0.0011. The result of our proposed method without any user interaction is shown in Fig. 11(c), which has a LMSE of 0.0015.

Fig. 12 shows further comparisons with the method proposed by Bousseau et al. In Fig. 13, we show the zoomed-in separation results for the cloth example. We

can see that Bousseau’s method leaves some “ghost” markings in the shading (c) and some residues of the cast shadows in the reflectance component (d). Fig. 14 shows the comparison with Tappen et al.’s work [15] and Bousseau et al.’s. We also compare our method with Tappen et al.’s method [14] for a gray-scale image example in Fig. 15. For gray-scale images, we compute the WRBW using pixel intensity. It is evident that our technique can generate visually comparable results from a single image without any additional information.

6 CONCLUSION

In this paper, to address the problem of intrinsic image decomposition, we have proposed two new sparse priors on reflectance: a data-driven sparse representation of reflectance and a global sparse constraint on reflectance colors. Combining the two sparse priors, we can effectively decompose a single image into its intrinsic components.

A sparse representation is made possible by using data-dependent weighted wavelets constructed based on the local sparsity constraint on reflectance. At the same time, the constructed weighted wavelet also preserves chromaticity distribution even at coarse scales. By using a multi-resolution representation of reflectance and applying reflectance weighting to enforce the sparsity constraint at multiple scales, we can convert what appears to be a local constraint into a global constraint. We also apply a global assumption that the number of different reflectance colors in the image is small through the use of a total-variations-like cost term. The decomposition problem is formulated as a constrained ℓ_1 -norm minimization problem, and the proposed approach seeks to recover the sparse reflectance signal given smoothness constraints on the illumination component. We also discussed the color bleeding problem in the decomposition with the proposed method. Small changes in the reflectance components could be wrongly assigned to the shading component. We solve this problem by using a soft matting method based the color line assumption which holds for natural shading and reflectance images.

The optimization formulation effectively broadens the influence of local information to help resolve ambiguous local inference and our experimental results show that the decomposition significantly benefits from the global constraints.

REFERENCES

- [1] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, “A variational framework for retinex,” *International Journal of Computer Vision*, vol. 52, pp. 7–23, 2003.
- [2] B. V. Funt, M. S. Drew, and M. Brockington, “Recovering shading from color images,” in *European Conf. on Computer Vision (ECCV)*, 1992, pp. 124–132.
- [3] W. Sweldens, “The lifting scheme: A construction of second generation wavelets,” *SIAM J. Math. Anal.*, vol. 29, pp. 511–546, 1998.
- [4] I. Omer and M. Werman, “Color lines: Image specific color representation,” *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2, pp. 946–953, 2004.

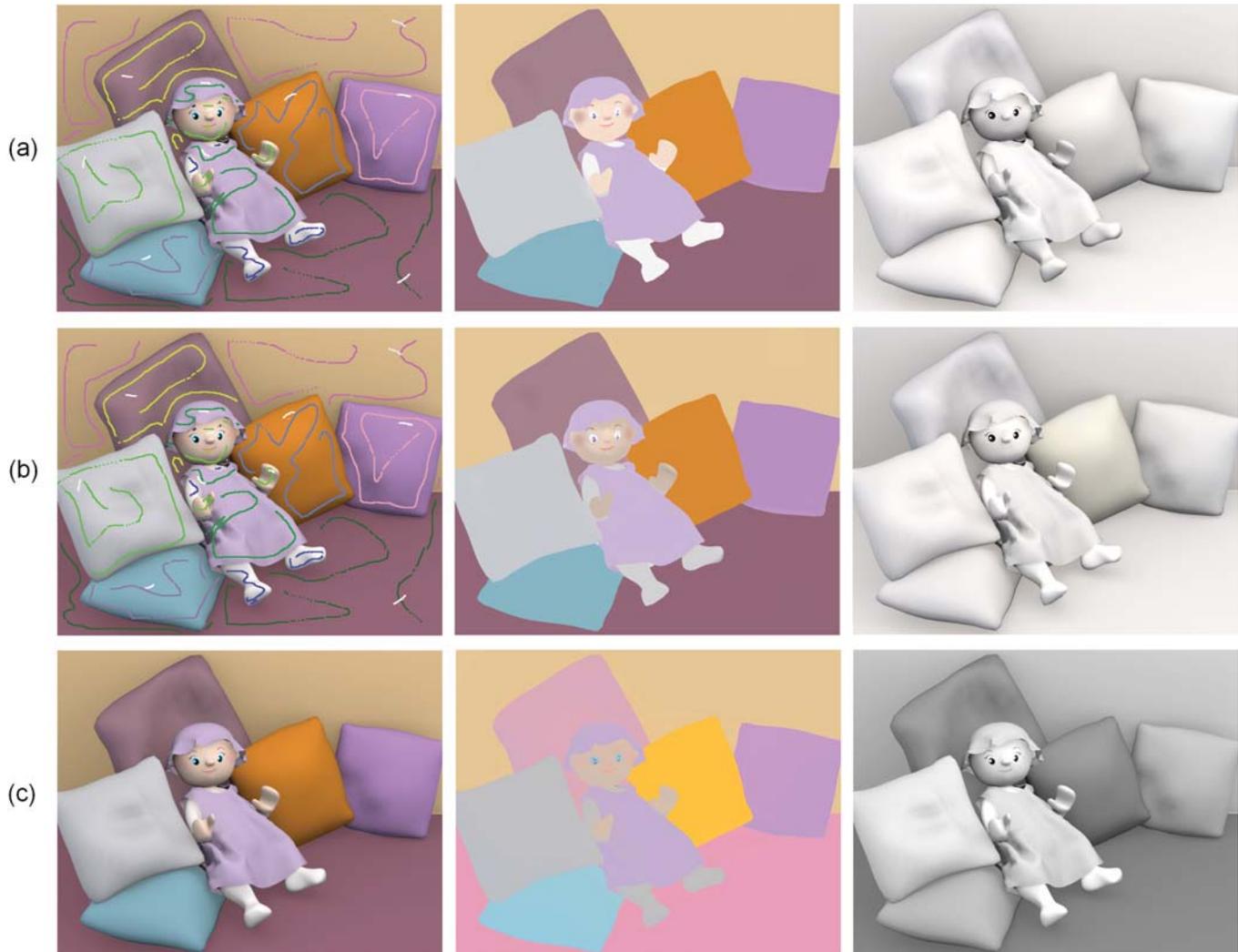
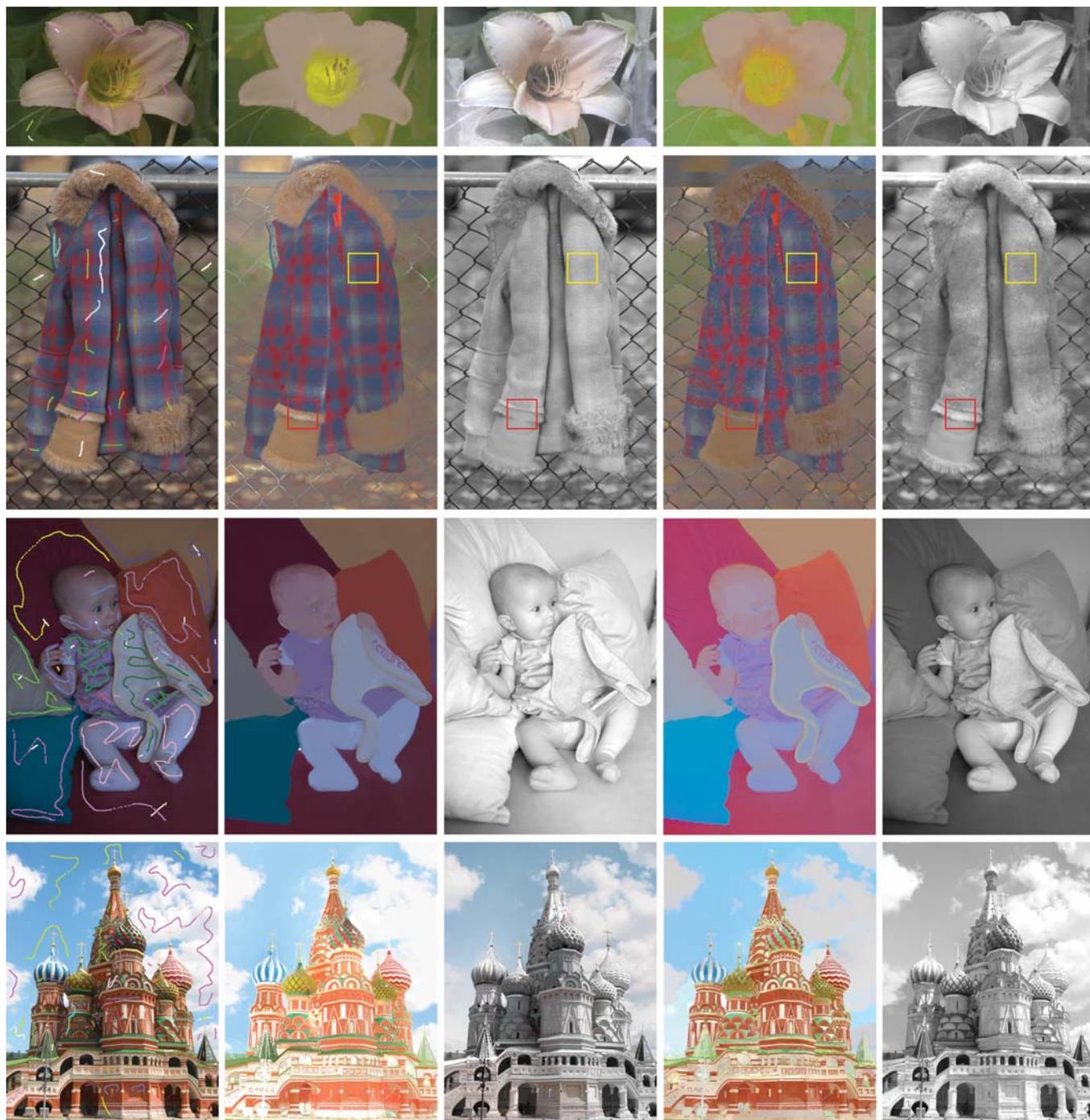


Fig. 11. Comparison using ground truth data from a synthetic image. Bousseau et al.'s approach [17] requires fairly accurate user strokes. (a) [17]'s results when user strokes are set to ground truth values (LMSE = 0.00055). (b) [17]'s results when the positions of user strokes are randomly perturbed by up to 15 pixels (LMSE = 0.0011). (c) Proposed SRC's results without user interaction (LMSE = 0.0015).

- [5] H. Barrow and J. Tenenbaum, "Recovering intrinsic scene characteristics from images," *Computer Vision Systems*, pp. 3–26, 1978.
- [6] Y. Weiss, "Deriving intrinsic images from image sequences," in *IEEE Int'l Conf. on Computer Vision (ICCV)*, vol. 2, 2001, pp. 68–75.
- [7] Y. Matsushita, S. Lin, S. B. Kang, and H.-Y. Shum, "Estimating intrinsic images from image sequences with biased illumination," in *European Conf. on Computer Vision (ECCV)*, vol. 2, 2004, pp. 274–286.
- [8] K. Sunkavalli, W. Matusik, H. Pfister, and S. Rusinkiewicz, "Factored time-lapse video," *ACM Transactions on Graphics*, vol. 26, no. 3, p. 101, 2007.
- [9] A. Troccoli and P. Allen, "Building illumination coherent 3d models of large-scale outdoor scenes," *International Journal of Computer Vision*, vol. 78, no. 2-3, pp. 261–280, 2008.
- [10] E. Land and J. McCann, "Lightness and retinex theory," *Journal of the Optical Society of America A*, vol. 3, pp. 1684 – 1692, 1971.
- [11] B. K. P. Horn, *Robot Vision*. MIT Press, 1986.
- [12] G. D. Finlayson, S. D. Hordley, and M. Drew, "Removing shadows from images using retinex," in *Proceedings of IS&T/SID Tenth Color Imaging Conference: Color science, Systems and Applications*, 2002, pp. 73–79.
- [13] M. Bell and W. T. Freeman, "Learning local evidence for shading and reflectance," in *IEEE Int'l Conf. on Computer Vision (ICCV)*, vol. 1, 2001, pp. 670–677.
- [14] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1459–1472, 2005.
- [15] M. Tappen, E. Adelson, and W. Freeman, "Estimating intrinsic component images using non-linear regression," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. II: 1992–1999.
- [16] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–7.
- [17] A. Bousseau, S. Paris, and F. Durand, "User-assisted intrinsic images," in *SIGGRAPH Asia '09: ACM SIGGRAPH Asia 2009 papers*. ACM, 2009, pp. 1–10.
- [18] R. Fattal, "Edge-avoiding wavelets and their applications," *ACM Trans. on Graphics*, vol. 28, no. 3, pp. 1–10, Aug 2009.
- [19] G. Uytterhoeven and A. Bultheel, "The Red-Black wavelet transform," in *Signal Processing Symposium (IEEE Benelux)*, M. Moonen, Ed. IEEE Benelux Signal Processing Chapter, 1998, pp. 191–194.
- [20] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.
- [21] L. Grady, "Random walks for image segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.



(a) Bousseau's results from a single image and user scribbles

(b) Our results from a single image using SRC

Fig. 12. Comparison with the user-assisted approach of Bousseau et al. [17].



(a) Right to left: Bousseaus' and our results

(b) Right to left: Bousseaus' and our results

Fig. 13. We zoom into the separation results for details of the yellow and red patches of the cloth example.



(a) Bousseau's results from a single image and user scribbles

(b) Our results from a single image using SRC



(c) Tappen's results from a single image

Fig. 14. Comparison with the user-assisted approach of Bousseau et al. [17], and the automatic approach of Tappen et al. [14].



(a) Input image

(b) Tappen's results

(c) Our results

Fig. 15. Gray-scale image example. Comparison with Tappen et al.'s work [14].

- [22] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graphics*, vol. 23, no. 3, pp. 689–694, 2004.
- [23] X. Liu, L. Wan, Y. Qu, T.-T. Wong, S. Lin, C.-S. Leung, and P.-A. Heng, "Intrinsic colorization," *ACM Transactions on Graphics (SIGGRAPH Asia 2008 issue)*, vol. 27, no. 5, pp. 152:1–152:9, December 2008.
- [24] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground-truth dataset and baseline evaluations for intrinsic image algorithms," in *International Conference on Computer Vision*, 2009, pp. 2335–2342.
- [25] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l_1 -regularized least squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.
- [26] T. Goldstein and S. Osher, "The split bregman method for l_1 regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009.
- [27] A. Levin, D. Lischinski, and Y. Weiss, "A closed-form solution to natural image matting," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 228–242, 2008.



Li Shen received the M.Eng. degree (Panasonic Scholarship) in software science from the Osaka University, in 2002, and the Ph.D. degree (MONBUSHO Scholarship) in information systems engineering from the Osaka University, Japan 2006. From 2006 to 2008, she was a visiting researcher at Microsoft Research Asia, Beijing. Since 2009, she has been a scientist with the Computer Graphics and Interface Department at the Institute for Infocomm Research, Singapore. Her main research interests are in computer graphics & compute vision, especially in low-level vision, computational photography, and image-based rendering/modeling.



Chuohao Yeo received the S.B. degree in electrical science and engineering and the M.Eng. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 2002, and the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley in 2009.

From 2005 to 2009, he was a graduate student researcher at the Berkeley Audio Visual Signal Processing and Communication Systems Laboratory at UC Berkeley. Since 2009, he has been a Scientist with the Signal Processing Department at the Institute for Infocomm Research, Singapore, where he leads a team that is actively involved in HEVC standardization activities. His current research interests include image and video processing, coding and communications, distributed source coding, and computer vision.

Dr. Yeo was a recipient of the Singapore Government Public Service Commission Overseas Merit Scholarship from 1998 to 2002, and a recipient of Singapore's Agency for Science, Technology and Research National Science Scholarship from 2004 to 2009. He received a Best Student Paper Award at SPIE VCIP 2007 and a Best Short Paper Award at ACM MM 2008.



Binh-Son Hua is a PhD student in School of Computing, National University of Singapore. He received his B.E. from Ho Chi Minh City University of Technology, Vietnam, in January 2008. His main research focus is physically-based rendering. He is also interested in real-time rendering and computational photography.