L_0 CO-INTRINSIC IMAGES DECOMPOSITION

Haipeng Dai[1], *Wei Feng*[1,3,*], *Liang Wan*[2,3], *Xuecheng Nie*[1,2]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] School of Computer Software, Tianjin University, Tianjin, China
[3] Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China, Tianjin, China

ABSTRACT

In this paper, we focus on *co-intrinsic decomposition*, a new problem that performs intrinsic decomposition on a pair of images simultaneously, which share the same foreground with arbitrarily different illuminations and backgrounds. We specifically demand the common foreground across different images to share same reflectance values. For the purpose of efficiency and feasibility, we perform the co-intrinsic decomposition at superpixel-level and propose a uniform approach to automatically derive non-local reflectance relationships via a unsupervised L_0 sparsity between superpixels from intra- and inter-images. We present a unicolor-light-based intrinsic model, from which we construct a non-local L_0 sparse co-Retinex model that imposes feasible constraints on shading, reflectance and environment light, respectively. The co-intrinsic decomposition is finally modeled as a quadratic minimization problem that leads to a fast closed form solution. Extensive experiments show plausible results of our approach in extracting common reflectance components from multiple images. We also validate the benefits of our results in boosting the accuracy of image co-saliency detection.

***Index Terms*—** Co-intrinsic images decomposition, L_0 sparsity, superpixels, unicolor-light, quadratic minimization

1. INTRODUCTION

Intrinsic image decomposition is the problem of separating an image into a reflectance layer and a shading layer [1], which can benefit many computer vision and graphic applications like material alteration, relighting, segmentation and object tracking. Mathematically, the intrinsic model represents an input image I as the pixelwise product of the reflectance R and the shading S, i.e. $I = S \cdot R$, which has unknowns of twice the number of the equations. To resolve the severe ill-posedness, state-of-the-art methods have proposed various types of local constraints [2, 3, 4] and global sparsity constraints [5, 6], or rely on user interactions [7]. On the other hand, some works incorporate additional information from multiple registered images of the same scene under different illumination conditions, including [8] that assumes a fixed viewpoint and [9] which considers different viewpoints.

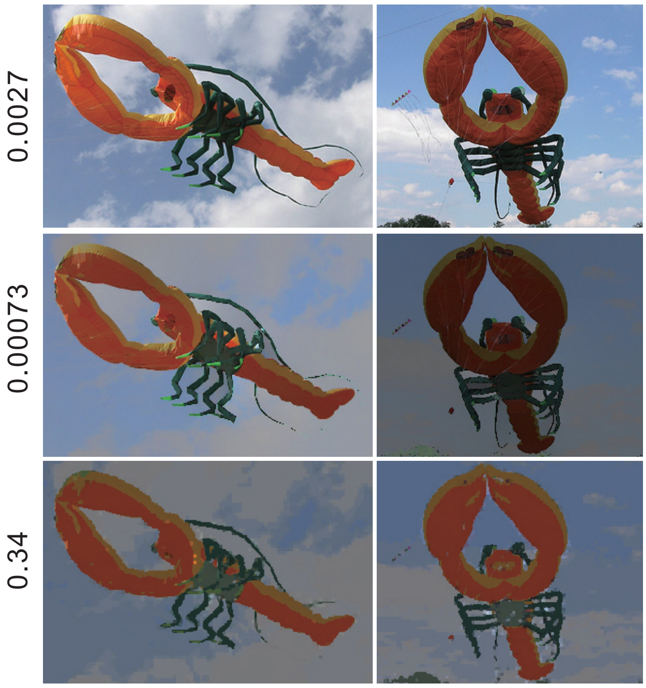

Fig. 1. The first row is the input image pair “Kite”. The second row is two reflectance images obtained by state-of-the-art single image intrinsic decomposition method[6]. The last row is the coherent reflectance images obtained using the proposed co-intrinsic decomposition. The left column show the histogram cosine similarity of foreground objects.

However, there exist many computer vision tasks that demand the suppression of the variances caused by different lighting and imaging conditions. The typical applications are image co-segmentation [10, 11] and co-saliency detection [12], in which the images share the same foreground yet with arbitrarily different illuminations and backgrounds. In this paper, we address this new problem, which we call *co-*

* is the corresponding author. Email: wfeng@tju.edu.cn. This work is supported by the Program for New Century Excellent Talents in University (NCET-11-0365), NSFC (61100121, 61100122), Tianjin science foundation for youth (12JCQNJC00100), and the National Science and Technology Support Project (2013BAK01B01).

intrinsic images decomposition, of performing joint intrinsic decomposition on a pair of images simultaneously. We require the reflectance values of the common foreground to be consistent across different images, regardless of the varying illumination conditions, different viewpoints and backgrounds.

Note that an image pair here may not come from the same scene as in [8, 9]. Intuitively, one naive solution to this problem is to conduct intrinsic decomposition separately for the two images. However, clearly this naive solution does not consider the same foreground constraint into, and hence cannot guarantee consistent reflectance values for the common objects across different images. One example is demonstrated in Fig. 1(b), which independently exploits the state-of-the-art single image intrinsic decomposition method [6] on each image. It is noted that the reflectance values of the foreground object deviate from each other in the image pair.

To solve the co-intrinsic images decomposition problem, we propose a simple yet effective approach using unsupervised non-local L_0 reflectance sparsity at superpixel level. To the best of our knowledge, our work is the first try for co-intrinsic decomposition of multiple images with the same foreground and different backgrounds and illuminations. In our approach, we first present a unicolor-light-based intrinsic model, which accounts for the global change of color tones in two images. Based on it, we construct a *co-Retinex* model that integrates four criteria together to constrain the co-intrinsic model, which contains feasible constraints on shading, reflectance and environment light. For the purpose of efficiency and to obey the gradient sparse prior of reflectance, we represent the reflectance component of each image as a set of superpixels, and develop a uniform approach to automatically derive non-local reflectance relationships within each single image and across different images. The decomposition result is further obtained by solving a superpixel (reflectance) and pixel (shading) hybrid quadratic minimization problem, which has a fast closed form solution.

We demonstrate our approach on several real-world image pairs, with visibly apparent illumination changes and different backgrounds. The plausible results are obtained in extracting common reflectance components from an image pair. In addition, we validate the effectiveness of our approach in the context of image co-saliency detection, yielding promising improvement to the accuracy.

2. RELATED WORK

Intrinsic image decomposition was first proposed by Barrow and Tenenbaum [1] specifying the task of separating an image into shading and reflectance components. For single image intrinsic decomposition, the Retinex [2] is one of the most successful model that assumes large local gradients attribute to the reflectance change and small local gradients attribute to the shading change. The reflectance can then be obtained by integrating the thresholded gradient field. Color retinex model extends the retinex algorithm by considering chromaticity gradients along with brightness gradients in color images. As reported by [13], color retinex is one of the best intrinsic decomposition methods that use local priors.

Instead of using binary threshold in the Retinex, Tappen et al. [14] trained an SVM classifier to discriminate the reflectance gradients from shading gradients. According to [13], it empirically tended to overfit the training data. Besides automatic decomposition, Bousseau et al. [7] imposed a local planar constraint of the reflectance and used overlapping local windows to propagate user guidance.

In addition to local priors, color sparsity prior [15] that assumes natural scenes are dominated by a small number of material colors has been adopted in many recent works [16, 6]. For instance, Shen and Yeo [5] applied a total-variation cost on the set of reflectance values within the image. Zhao et al. [6] constructed a global sparsity constraint using texture analysis, requiring local texture patches belong to several clusters. In addition, the intrinsic image decomposition problem is less constrained when using additive information through multiple images of the same scene with different illuminations [8]. Laffont et al. [9] used an image collection of different viewpoints and illuminations of the same scene to construct 3D points and match pixels in different images.

In all the state-of-the-art intrinsic decomposition methods, a fundamental problem is to eliminate the ill-posedness and well-define the problem. Thus, generally speaking, one key problem in intrinsic decomposition is to seek proper priors, either data-driven or heuristic, about the desired reflectance and shading components.

3. OUR APPROACH

The proposed co-intrinsic decomposition aims at simultaneously processing a pair of images. Different from the registered images used in [9], our input image pair shares the same foreground yet each with arbitrary background, different viewpoints and varying illuminations. To handle these characteristics, in the following, we first present a unicolor lighting based intrinsic model which accounts for the varying tones, then we propose a co-Retinex model that takes local smooth constraint, model constraint and non-local L_0 sparsity constraints on the image pair into consideration.

3.1. The unicolor-based intrinsic model

As we know, the shading is the joint effect of the object's geometry and the environment lighting. Despite the lighting direction, the environment lighting in the two images of a pair can be quite different in color tones. To suppress the variation of color tones, we assume that the shading can be formulated as the product of a global environment lighting l_e, which is constant for all the pixels in one image, and a local shading component M, which varies pixel by pixel. Let I be an input image, and p denotes one pixel. The shading S_p for pixel p is mathematically defined as:

$$S_p = l_e M_p, \tag{1}$$

where the environment lighting l_e is a RGB color vector, and M_p is a non-negative value representing the magnitude of perceived illumination.

Based on the unicolor lighting, the intrinsic image model is given by $I_p = l_e M_p R_p$, where R_p denotes the reflectance of pixel p. Taking the logarithm at both sides yields:

$$I_p = l_e + M_p + R_p, \tag{2}$$

where for simplicity, we reuse the symbols to represent their log values. Recalling the conventional intrinsic model $I_p = S_p + R_p$, ours includes one more variable l_e to reflect the global color tone.

3.2. Co-retinex model

Considering two images in a pair, the shading or reflectance of "neighboring" pixels within one image and across two images are subject to various constraints. It is noticed that the "neighboring" pixels do not only refer to the neighbors in a local window, they also refer to the correlated distant pixels within single image and across images. We formulate the co-intrinsic images decomposition problem as the minimization of the following objective function:

$$E(l_e, M, R) = E_c(l_e, M, R) + \lambda_m E_m(M) + \lambda_r E_r(R) + \lambda_e E_e(M). \tag{3}$$

Here, E_c is a term to incorporate the intrinsic model constraint. E_m is a local term to formulate the shading smoothness constraint. E_r represents the non-local L_0 sparsity constraint on the reflectance values of superpixels from intra- and inter-images. We further add E_e to constrain the brightness scale. λ_m, λ_r and λ_e are positive weights, and we set $\lambda_m = 10$, $\lambda_r = 100$, and $\lambda_e = 1000$ in our implementation. Also note that for simplicity, the notation of these objective functions uses one symbol, e.g. l_e, to represent the corresponding variables from two images, e.g. l_e^1 and l_e^2.

Before describing the definitions of these functions, we have to point out the number of unknowns is twice the number of equations for single image intrinsic decomposition. To make the computation practical, we propose to represent the reflectance R at superpixel level, due to the retinex assumption that reflectance differences between adjacent pixels are small. The superpixels are constructed for each image via the SLIC algorithm [17].

Model constraint. In our model, we minimize the three components l_e, M and R simultaneously. Hence we force a model constraint where the combination of the three components is close to the original pixel value. It is defined as:

$$E_c(l_e, M, R) = \sum_k \sum_{p \in I^k} (I_p^k - l_e^k - M_p^k - R_{u(p)}^k)^2, \tag{4}$$

where $u(p)$ represents the superpixel that pixel p belongs to.

Shading smoothing constraint. According to the retinex assumption, the shading value varies smoothly within a local image window. Hence, we define $E_m(M)$ as:

$$E_m(M) = \sum_k \sum_{p \sim q} (M_p^k - M_q^k)^2, \tag{5}$$

where $p \sim q$ encodes the 4-connected relationship of pixels in one image.

Non-local L_0 sparse reflectance constraint. We formulate the non-local reflectance constraint based on the assumption that superpixels with similar chromaticity values have similar reflectance. Similar assumptions have been employed in several previous works [5, 16, 6]. The difference lies in that we construct the non-local correlation from both intra-images and inter-images. Specifically, we formulate the non-local sparse representation as a L_0 minimization problem, which is defined on the set of superpixels from two images of a pair. The definition of $E_r(R)$ is presented in Section 3.3.

Brightness scale constraint. To disambiguate the scale problem where $I = S + R$ and $I = (S - d) + (R + d)$ are both the solutions, we demand the brightest pixels to have unit shading magnitude 1. Since $\log 1 = 0$, the brightness scale constraint is defined as:

$$E_e(M) = \sum_k \sum_{p \in \Gamma^k} (M_p^k)^2, \tag{6}$$

where Γ_i^k is the set containing the brightest pixels in image I^k. Here, We choose the brightest pixels by comparing intensity values from both images.

Closed form optimization Clearly, our energy function $E(l_e, M, R)$ is a well-defined quadratic function, including shading magnitude M_p for each pixel p, superpixel level reflectance R_u for each superpixel u, and environment lighting l_e from two images. Following the similar idea in [6], the quadratic function has a unique global minimum that can be solved by the conjugate gradient algorithm.

3.3. Non-local L_0 sparse reflectance constraint

Let $\mathcal{U}$ denote the set of superpixels combined from both images. For each superpixel u, we construct the sparse representation by solving the following L_0 minimization problem:

$$\min_{\alpha_u} \|f_u - D_u \cdot \alpha_u\|, \quad s.t. \ \|\alpha_u\|_0 \leq Z. \tag{7}$$

Here, f_u is the feature of superpixel u, elaborated later. The sparse dictionary D_u is define as:

$$D_u = [f_1, \ldots, f_{u-1}, f_{u+1}, \ldots, f_{|\mathcal{U}|}], \tag{8}$$

where $|\mathcal{U}|$ denotes the set size of $\mathcal{U}$. The parameter $\alpha_u \in \mathbb{R}^{|\mathcal{U}|-1}$ is the coefficient vector for sparse representation of f_u over D_u. The L_0 norm $|\alpha_u|_0$ computes the number of

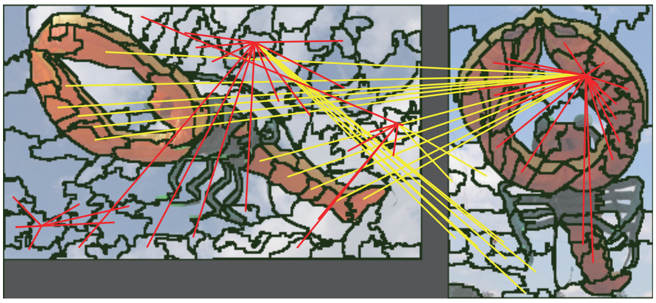

Fig. 2. Significant reflectance correlations. Red and yellow lines indicate reflectance constraints within an image and across images.

non-zero coefficients. Z controls the representation sparsity, which is set as $Z = 50$ in our experiments.

After solving the minimization problem using OMP algorithm [18], the non-zero components of α_u can be taken as the correlation weights between superpixel u and its correspondent superpixels within two images. Fig. 2 demonstrates the top 5 percents of correlations for a pair of images. Accordingly, we define $E_r(R)$ as the following:

$$E_r(R) = \sum_u \sum_{v,\alpha_u(v)\neq 0} \alpha_u(v)^2 (R_u - R_v)^2, \tag{9}$$

where $\alpha_u(v)$ is one element of α_u, representing the correlation between superpixels u and v.

An important issue in the above L_0 minimization is the representation of the feature vector f_u of superpixel u. In our paper, we adopt a histogram-based feature representation. Specifically, we first perform K-means clustering on the chromaticity values of all pixels from the two images to find J clusters. For each superpixel, we construct a J-dimensional histogram h_u. An element $h_u(j)$ represents how many pixels within a specific superpixel belong to the j-th cluster. h_u is normalized by dividing the pixel number of the superpixel. We then employ cosine-similarity to measure the distance between two feature vectors.

4. EXPERIMENTAL RESULTS

To evaluate the proposed method, we collected testing image pairs from MFC dataset [10, 9, 19]. The two images in a given image pair have visually apparent illumination changes. We compare our method with two state-of-the-art single image intrinsic decomposition methods, i.e. the close-formed solution [6] and optimization-based solution [4]. The two methods are performed separately on the two images of a pair.

The comparison results are demonstrated in Fig. 1, Fig. 3 and Fig. 4. Obviously, single image decomposition methods can not get consistent reflectance values for the common foreground in the image pairs. For example, when using [6], the doll looks more bright red in the first image, yet dark red in the other image (Fig. 3). Also see the red regions in Fig. 3, where the tower's reflectance is more brightish than that of the symmetric one. Even within a single image, the two methods under comparison may still be insufficient to generate consistent reflectance for distant correlated objects. As for [4], since it does not directly constrain distant pixels, the shading and shadow effects can be remained on the reflectance image (see the regions pointed by the yellow arrows in Fig. 4). Although Zhao et al. [6] constructed non-local reflectance constraint between distant pixels, they only considered textured pixels. Hence, it may not work very well for objects with uniform colors (see Fig. 1 where the reflectance of the kite's head suffers from the lighting effect). Instead, we use chromaticity feature to construct non-local reflectance correlation directly. As our results show, we can eliminate the shadow and make distant pixels share much consistent reflectance values, within and across a pair of images.

Next, we quantitatively evaluate the results by computing the similarity of the same foreground objects in the reflectance images for each method. We construct a histogram to represent the foreground reflectance for one image, and measure the cosine-similarity between two images. Here, we quantize each RGB channel to M bins, and the histogram is a M^3-dimension vector after normalization. Table 1 lists the similarities of different methods. We can see that our method achieves the highest similarities for all testing images.

We validate the benefit of our co-intrinsic decomposition in image co-saliency detection task. Fig. 5 shows the co-saliency results by the method of [12] on original image pairs "Doll", and our reflectance image pair. From the result, we can clearly see that the jointly extracted reflectance images enable more complete and consistent co-saliency detection result. The right chart is the average precision, recall and F1-measure value on all image pairs.

Table 1. The similarity of the same foreground reflectance values in an image pair.

	Methods in [6]	Methods in [4]	Our Method
Kite	7.348e-04	9.657e-04	0.3043
Doll	1.685e-04	0.0019	0.0518
St Basile	0.0135	0.0058	0.0334
Bucky	0.0112	0.0152	0.0337

5. CONCLUSIONS

In this paper, we address co-intrinsic images decomposition, a new problem that performs intrinsic decomposition on a pair of images simultaneously. Different from registered images decomposition, the images in a pair share the same foreground yet with arbitrarily different illuminations, viewpoints, and backgrounds. As the first attempt to this prob-

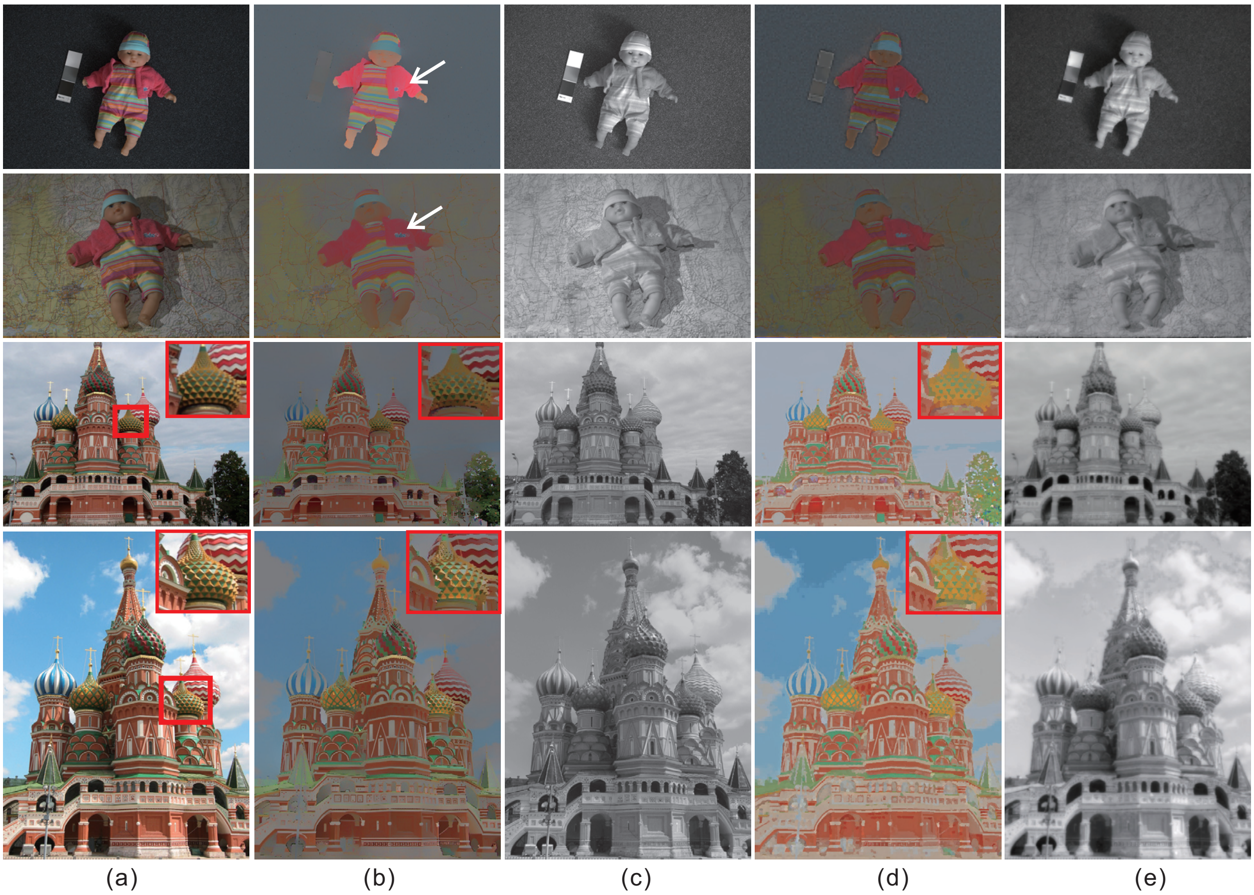

Fig. 3. (a) Input image pairs ”Doll” and ”St Basile”. (b)-(c) Reflectance and shading images by the method in [6]. (d)-(e) Reflectance and shading images by our method

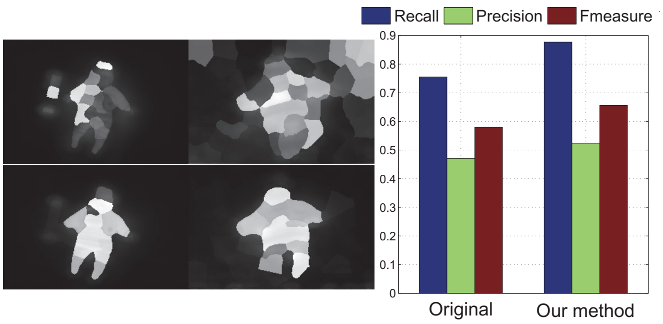

Fig. 5. Co-saliency detection result on the original image pair ”Doll” and our reflectance image pair.

lem, we propose an effective approach using unsupervised non-local L_0 reflectance sparsity at super-pixel level. To account for the global change of color tones in two images, we extend intrinsic model to a unicolor lighting-based intrinsic model. We then propose a co-Retinex model that constrains local and non-local pixels/superpixels from both intra- and inter-images. By combining the set of superpixels from two images, we can derive non-local reflectance correlation within and across images in a uniform manner. The effectiveness of the proposed approach is validated through experiments on co-intrinsic images decomposition and further co-saliency or co-foreground detection.

6. REFERENCES

[1] H.G. Barrow and J.M. Tenenbaum, “Recovering intrinsic scene characteristics from images,” in *Computer vision systems*, 1978.

[2] E. H. Land and J. J. McCann, “Lightness and retinex theory,” *Journal of the Optical society of America*, vol. 61, no. 1, pp. 1–11, 1971.

[3] M. F. Tappen, E. H. Adelson, and W. T. Freeman, “Estimating intrinsic component images using non-linear regression,” in *CVPR*, 2006.

[4] J. Shen, X. Yang, Y. Jia, and X. Li, “Intrinsic images using optimization,” in *CVPR*, 2011.

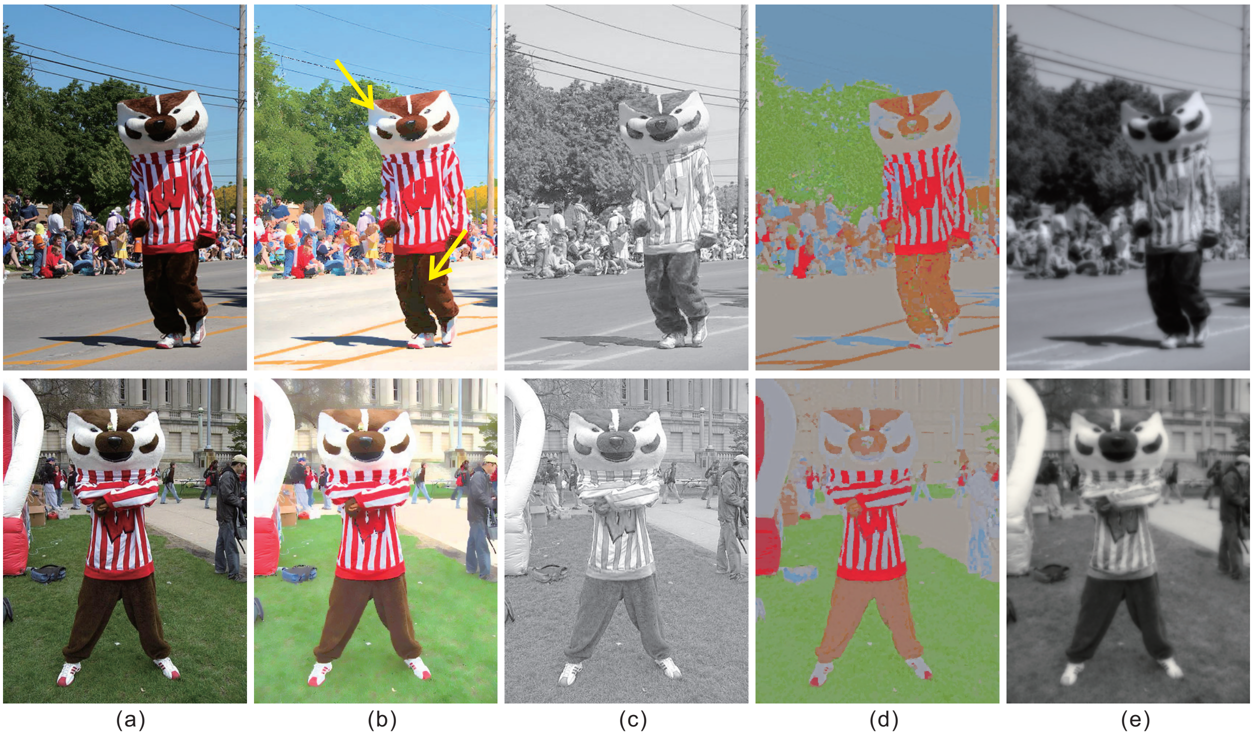

Fig. 4. (a) Input image pair "Bucky". (b)-(c) Reflectance and shading images by the method in [4]. (d)-(e) Reflectance and shading images by our method

[5] L. Shen and C. Yeo, "Intrinsic images decomposition using a local and global sparse representation of reflectance," in *CVPR*, 2011.

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

[7] A. Bousseau, S. Paris, and Frédo Durand, "User-assisted intrinsic images," *ACM Transactions on Graphics*, vol. 28, no. 5, 2009.

[8] Y. Weiss, "Deriving intrinsic images from image sequences," in *ICCV*, 2001.

[9] P. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis, "Coherent intrinsic images from photo collections," *ACM Transactions on Graphics*, vol. 31, no. 6, 2012.

[10] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "icoseg: Interactive co-segmentation with intelligent scribble guidance," in *CVPR*, 2010.

[11] W. Feng, J. Jia, and Z.Q. Liu, "Self-validated labeling of markov random fields for image segmentation," *IEEE TPAMI*, vol. 32, no. 10, pp. 1871–1887, 2010.

[12] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE TIP*, vol. 20, no. 12, pp. 3365–3375, 2011.

[13] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *ICCV*, 2009.

[14] M. F. Tappen, W. T. Freeman, and E. H. Adelson, "Recovering intrinsic images from a single image," *IEEE TPAMI*, vol. 27, no. 9, pp. 1459–1472, 2005.

[15] I. Omer and M. Werman, "Color lines: Image specific color representation," in *CVPR*, 2004.

[16] C. Rother, M. Kiefel, L. Zhang, B. Schölkopf, and P. V. Gehler, "Recovering intrinsic images with a global sparsity prior on reflectance," in *NIPS*, 2011.

[17] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[18] J Mairal, "Spams (sparse modeling software), version 2.3," 2010.

[19] L. Li, W. Feng, L. Wan, and J. Zhang, "Maximum cohesive grid of superpixels for fast object localization," in *CVPR*, 2013.