Intrinsic Images Using Optimization

Jianbing Shen, Xiaoshan Yang, Yunde Jia
Beijing Laboratory of Intelligent Information Technology, School of Computer Science
Beijing Institute of Technology, Beijing 100081, P. R. China
{shenjianbing, 20907118, jiayunde}@bit.edu.cn

Xuelong Li
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, Shaanxi, P. R. China
xuelong_li@opt.ac.cn

Abstract

In this paper, we present a novel intrinsic image recovery approach using optimization. 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 of a single image having similar intensity values should have similar reflectance values. Thus the intrinsic image decomposition is formulated by optimizing an energy function with adding a weighting constraint to the local image properties. In order to improve the intrinsic image extraction results, we specify local constrain cues by integrating the user strokes in our energy formulation, including 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 by previous approaches.

1. 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 [1] 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 [16, 7] or from image sequences [18, 11, 10]. 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 images recovery problem. Assumptions and restrictions in intrinsic images algorithms make them insufficient to produce high quality illumination and reflectance components from a single image.

Colour is usually employed as a key cue for identifying shading gradients, such as the approaches in [2, 4, 5, 16]. Tappen *et al.* [16] 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.By using pieces of paper colored with a marker as test images, Tappen *et al.* [15] created a set of ground truth intrinsic images where they used color to measure shading separately from reflectance. Their approach then learned a weight function using the non-linear regression and estimated the intrinsic shading and albedo images.

Recently, Shen *et al.* [13] assumed that similar textures correspond to similar reflectance, and then they improved Retinex based algorithm for intrinsic images decomposition. More recently, Bousseau *et al.* [3] presented an approach for obtaining the intrinsic images with user scribbles from a single image. Their approach calculated the illumination component and reflectance component by introducing a propagation energy. Besides the intrinsic images decomposition approaches are studied in the previous work, several image editing tools employing the intrinsic images are also developed in recent years, such as intrinsic colorization [9] and image retexturing [3]. Liu *et al.* [9] presented an example-based intrinsic colorization algorithm based on illumination differences between grayscale target and color reference images. Their approach first recovered reflectance image of the target scene from multiple color references by web search, and then they transferred color to the grayscale

reflectance image through relighting with the illumination component of the target image.

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 can not be solved correctly without additional prior knowledge on reflectance and illumination. In order to resolve its ambiguities, Bousseau *et al.* [3] presented a user-assisted approach for calculating the intrinsic images from a single image. Their approach required some user interactions for indicating the regions of constant reflectance or illumination, and achieved intrinsic images decomposition by a propagation energy. However, their approach was based on the assumptions of reflectance variations lying locally in a plane in color space, and it would introduce the incorrect decompositions when it handled the cases such as a black-and-white checkerboard texture. In this paper, we present a novel automatic intrinsic images recovery algorithm using optimization, which is based on the newly defined energy function using a local continuity assumption in reflectance values. Moreover, a novel user scribble based energy and a new weight function are proposed to improve the accuracy of the recovered intrinsic images.

2. 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 images extraction is to solve the two unknowns s and R on the right hand side of Equation (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 above 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 improve the performance by some user scribble constrains.

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, that is to say, the pixels with the similar intensity values share the similar reflectance values. Our aforementioned assumption is inspired by the property of a change in color between pixels often indicating a reflectance change [12]. Thus the reflectance value of one pixel can be represented by the weighted summation of its

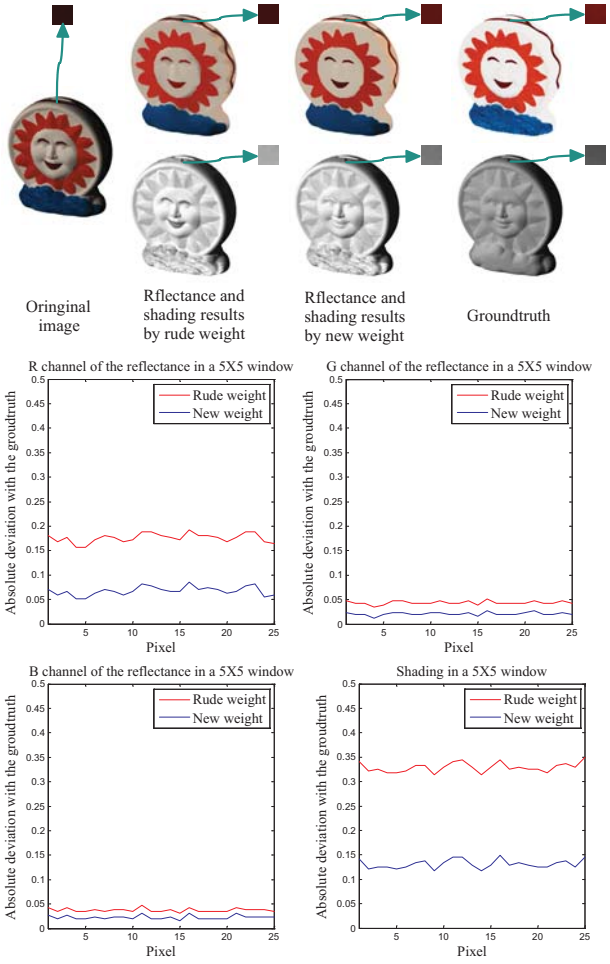


Figure 1. Comparisons of intrinsic images recovery using the rude weight function (Equation (2)) and our new weight function (Equation (3)). Compared to the ground truth data, the intrinsic images by our new weight function are more accurate.

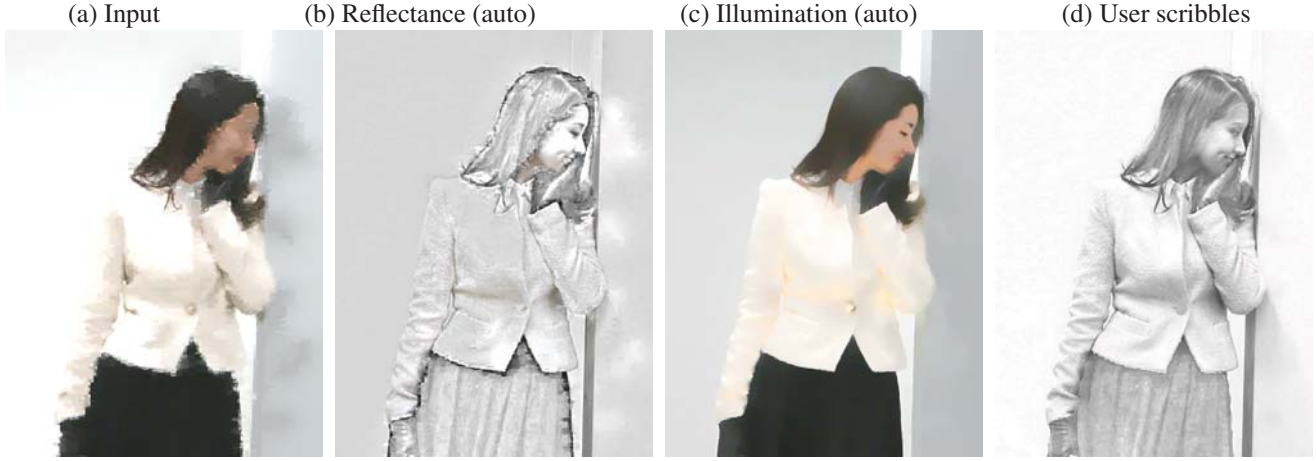
neighborhood pixel values as follows:

$$R_i = \sum_{j \in N(i)} \omega_{ij} R_j$$

$$\omega_{ij} = e^{-(Y_i - Y_j)^2 / 2\sigma_i^2} \tag{2}$$

where ω_{ij} measures the similarity of the reflectance value between pixel i and pixel j. ω_{ij} is determined by the intensity value 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 around i. In image segmentation [14, 17] and image colorization [8] algorithms, similar weight functions are used extensively and referred to as affinity functions.

(a) Input (b) Reflectance (auto) (c) Illumination (auto) (d) User scribbles



(e) Reflectance by [3] (f) Illumination by [3] (g) Reflectance (ours) (h) Illumination (ours)

Figure 2. Illustration of our optimization approach. Note that (b) and (c) are the results by our automatic approach (Equation (5)), while (g) and (h) are the results by our full approach with user scribbles (Equation (7)).

2.1. Energy based on local windows

In order to make the pixels that have large different shading values share the similar reflectance weighting values, we re-consider 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 (Equation (2)) as follows:

$$\omega_{ij} = e^{-[<\tilde{I}_i,\tilde{I}_j>^2/\sigma_{iT}+(Y_i-Y_j)^2/\sigma_{iY}]},$$

$$< \tilde{I}_i, \tilde{I}_j >= \arccos(\tilde{I}_i * \tilde{I}_j) = \arccos(\tilde{I}_{ir}\tilde{I}_{jr}+\tilde{I}_{ig}\tilde{I}_{jg}+\tilde{I}_{ib}\tilde{I}_{jb}) \quad (3)$$

where $< \tilde{I}_i, \tilde{I}_j >$ denotes the angle between the vector $\tilde{I}_i$ and $\tilde{I}_j$. Similar to [16], 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, σ_{iY} denotes the variance of the intensities of the pixels in a local window around i.

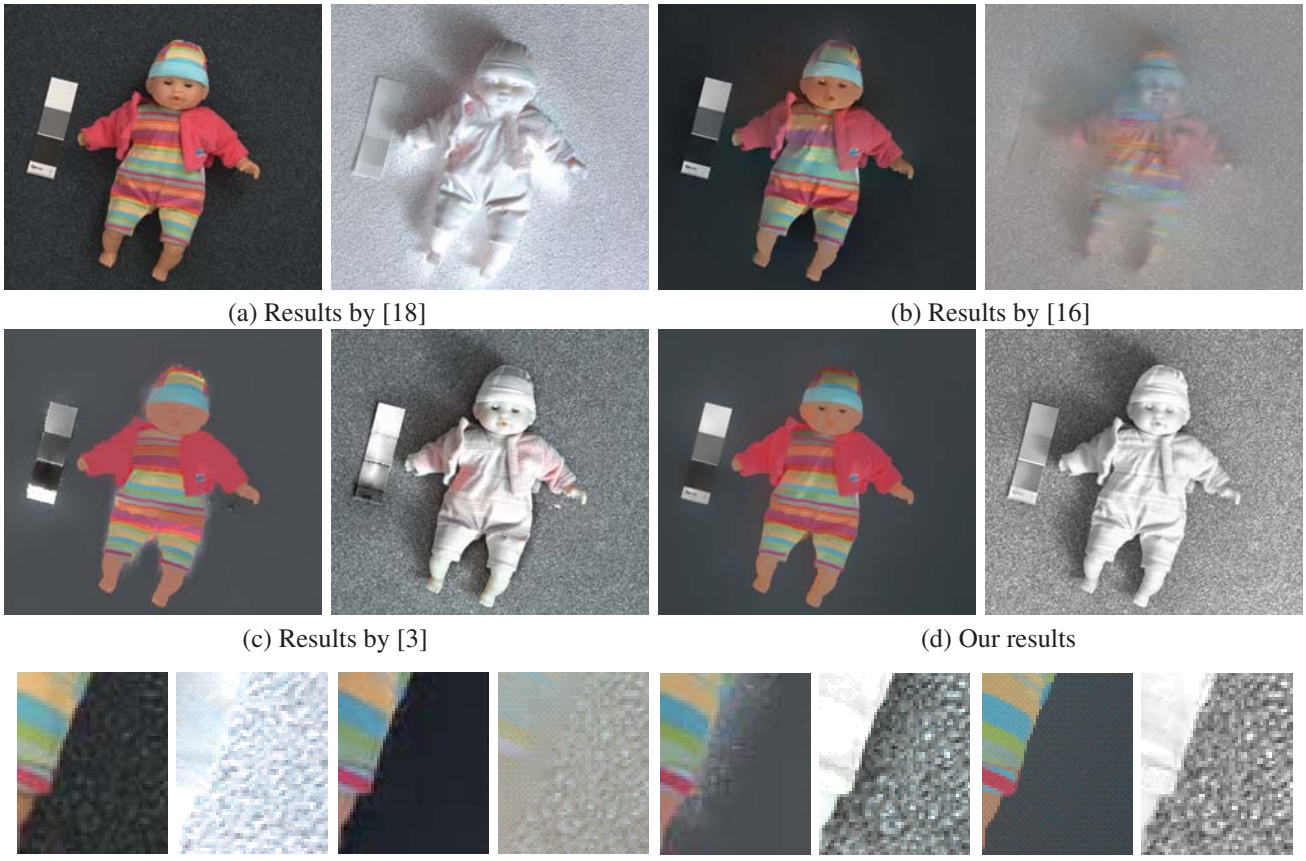
Figure 1 gives the illustration of comparing the accuracy of the recovered intrinsic images using different weight functions. We use the image patches in a local window for calculating the per-pixel error between the recovered intrinsic images and the ground truth intrinsic images from MIT data set [6]. It is obvious that the error by our new weight function (Equation (3)) is smaller than the error by rude weight function (Equation (2)). Based on the previous analysis, now we can define a new energy function to obtain the intrinsic images from a single input photograph:

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

where N(i) represents the local neighborhood window of pixel i, the first energy term defines the constraint for R in a local window, and the second energy term constrains R and s in Equation (1), here $\tilde{s} = 1/s$. We optimize the above equation to obtain the intrinsic images:

$$\arg\min_{R,\tilde{s}} E(R, \tilde{s}),$$

$$\forall i \in P, 0 \leq R_{ir} \leq 1, 0 \leq R_{ig} \leq 1, 0 \leq R_{ib} \leq 1 \quad (5)$$



(a) Results by [18] (b) Results by [16]

(c) Results by [3] (d) Our results

(e) (f) (g) (h)

Figure 3. Comparison results with the automatic approach of Tappen *et al.* [16] and the user-assisted approach of Bousseau *et al.* [3]. (e)-(h) Close-ups of (a)-(d).

2.2. Energy based on user brushes

Previous automatic approaches usually work well for the images of simple scene to obtain the desired 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 constrain cues by integrating the user strokes in our energy formulation.

Unlike Tappen's method [16], they propose an efficient method to classify each derivative as being caused by either illumination or a reflectance change. We employ the user scribbles to indicate each intensity as being caused by either a shading change or a reflectance change. Thus, we adopt three types of user scribbles as [3] did, including constant-reflectance brush, constant-illumination brush and fixed-illumination brush.

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

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

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

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

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

where λ_R, λ_s and λ_{fix} measure the importance to our image decomposition model with the user's constant-reflectance cues, constant-illumination cues and fixed-illumination cues, respectively.

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

$$\arg\min_{R, \tilde{s}} E_C(R, \tilde{s}),$$

$$\forall i \in P, 0 \leq R_{ir} \leq 1, 0 \leq R_{ig} \leq 1, 0 \leq R_{ib} \leq 1, \tilde{s} \geq 1 \quad (7)$$

Figure 2 gives an illustration example of intrinsic images recovery by our optimization approach. The results shown in Figures 2(b) and (c) are obtained by our automatic energy optimization (Equation (5)). As shown in Figure 2(g) and (h), we then add some user scribbles to improve the performance by optimizing Equation (7). In particular, the input image shown in Figure 2(a) consists of a black-and-white clothing texture region, which belongs to the case that cannot be handled by the previous user-assisted approach [3]. Compared to the results by Bousseau *et al.* [3] (Figure 2(e) and (f)), our optimization approach produces better results of the recovered intrinsic images, and exhibits natural edges and structures.

3. Experimental results

In our experiments, most of the input images are taken from the previous work in [18, 16, 3] for comparison. In order to demonstrate the effectiveness of the proposed approach, we compared our method with the approaches in [18, 16, 3]. As shown in Figure 3, we only take one input image to run our algorithm, and achieve the best reflectance and illumination decomposition results, such as the clear edges and texture information. The main limitation of Tappen's algorithm [16] is that a binary labeling cannot handle areas where both the reflectance and illumination variations occur, so the result by their approach exhibits some visual artifacts in highly textured areas and under mixed lighting conditions. Bousseau's method [3] assumed that local reflectance variations lie in 2D subspaces of RGB color space , and some reflectance information is incorrectly separated into the illumination component in the clothing regions.

We next show a more challenging example of a synthetic image for intrinsic image recovery in Figure 4. The result by Bousseau's approach [3] interprets the black pixels of the eyes as shadow (Figure 4(b)), which is incorrect for the illumination image. In contrast, our approach is more consistent with the ground truth data, especially interprets the black pixels of the eyes as reflectance (Figure 4(c)).

In order to quantitatively compare the existing algorithms, several objective performance measures of intrinsic image decomposition have been proposed recently, such as mean squared error (MSE) scores, local mean squared error (LMSE) scores [6] and absolute LMSE (aLMSE) scores [7]. We compute three statistics, including MSE, LMSE and aLMSE scores, respectively, which are shown in Table 1. Figure 5 gives a more intuitive illustration of Table 1. As we would expect, our optimization approach performs better than both Tappen's approach [16] and Bousseau's method [3], and achieves less error decomposition. Figure 6 gives the performance comparisons on the MIT data set [6]. The first column in Figure 6(b) and (c) denotes the average performance. While the second and third columns in Figure 6(b) and (c) are the performance on the reflectance and shading estimates, respectively. Note that the values of MSE and aLMSE are referred to left y-axis, while the values of LMSE are referred to right y-axis.

(a) Input (b) Bousseau's method [3] (c) Our method

Figure 4. Comparison results with the ground truth data from a synthetic image [3]. Compared to the user-assisted approach of Bousseau *et al.* [3], our optimization method creates better results, especially interprets the black pixels of the eyes as reflectance. First row: reflectance; second row: shading; third row: close-ups.

4. Conclusions

We have presented a novel 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 optimizing an energy function with adding a novel weighting constraint to the local image characteristics. The intrinsic image results are improved by introducing the local constrains of user scribbles in our energy function. We show examples of intrinsic images where our optimization approach eliminates the visual artifacts present in previous methods and achieves a better recovery of intrinsic reflectance and illumination components.

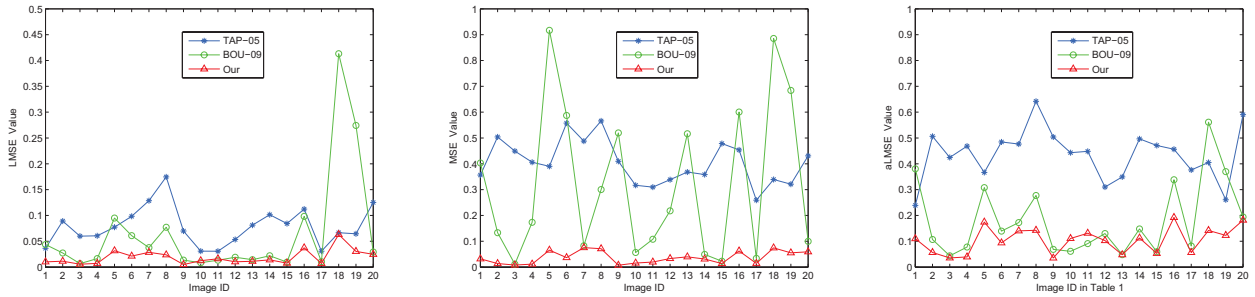
5. Acknowledgements

This work was supported by the National Natural Science Foundation of China (Nos. 60903068 and 61072093), the Key Program of NSFC-Guangdong Union Foundation (No. U1035004), and Excellent Young Teacher Research Fund of Beijing Institute of Technology (2009Y0707). The Project-sponsored by SRF for ROCS, SEM.

Table 1. MSE, LSME and aLMSE measure statistics with [16, 3] using the MIT data set [6].

		Results by [16]			Results by [3]			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.



(a) LMSE comparison (b) MSE comparison (c) aLMSE comparison

Figure 5. Quantitative evaluation: plots of Table 1.

References

[1] H. G. Barrow and J. M. Tenenbaum. Recovering intrinsic scene characteristics from images. *Computer Vision Systems*, pages 3–26, 1978.

[2] M. Bell and W. T. Freeman. Learning local evidence for shading and reflectance. *In: Proceedings of IEEE ICCV*, pages 670–677, 2001.

[3] A. Bousseau, S. Paris, and F. Durand. User-assisted intrinsic images. *ACM Transaction on Graphics*, 28(5):Article No.: 130, 2009.

[4] B. V. Funt, M. S. Drew, and M. Brockington. Recovering shading from color images. *In: Proceedings of ECCV*, pages 124–132, 1992.

[5] M. S. D. G. D. Finlayson and C. Lu. Intrinsic images by entropy minimization. *In: Proceedings of ECCV*, pages 582–595, 2004.

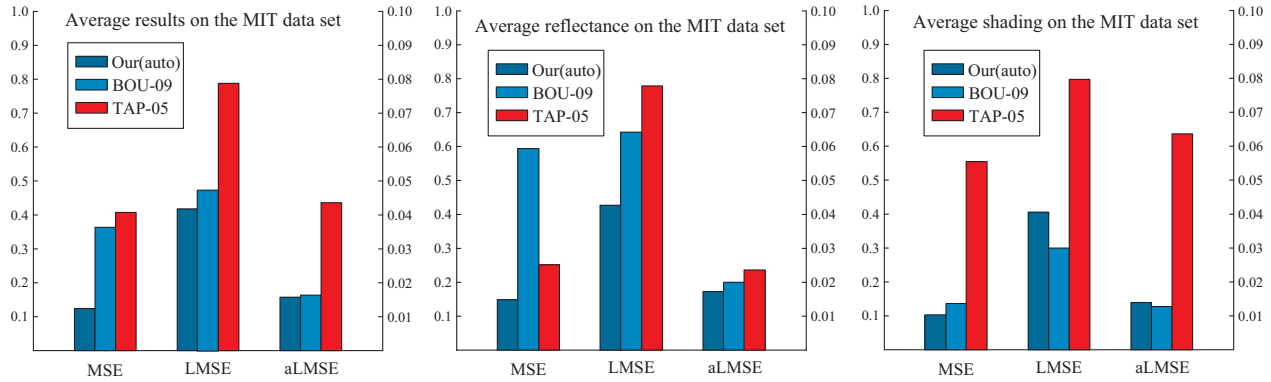
[6] R. Grosse, M. Johnson, E. H. Adelson, and W. T. Freeman. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. *In: Proceedings of IEEE ICCV*, pages 2335–2342, 2009.

[7] X. Jiang, A. J. Schofield, and J. L. Wyatt. Correlation-based intrinsic image extraction from a single image. *In: Proceedings of ECCV*, pages 58–71, 2010.

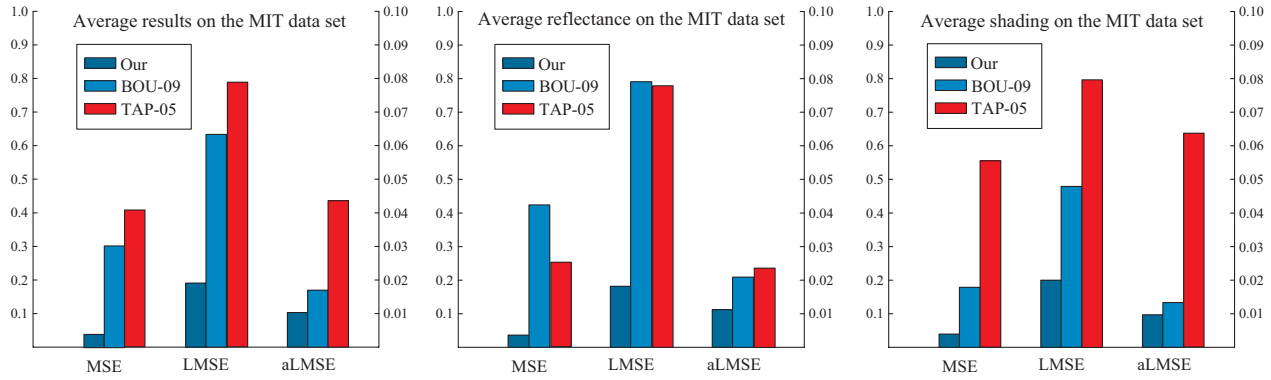
[8] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. *ACM Transaction on Graphics*, 23(3):689–694, 2004.

[9] X. Liu, L. Wan, Y. Qu, T. Wong, S. Lin, C. Leung, and

(a) MIT data set [6]



(b) Comparing our automatic approach (Equation (5)) with the methods [16, 3] on MIT data set [6]



(c) Comparing our full approach (Equation (7)) with the methods [16, 3] on MIT data set [6]

Figure 6. Performance comparisons on MIT data set [6] with the approaches [16, 3]. Note that we abbreviate Tappen *et al.*[16]'s approach (TAP-05), and abbreviate Bousseau *et al.* [3]'s algorithm (BOU-09).

P. Heng. Intrinsic colorization. *ACM Transaction on Graphics*, 27(5):Article 152, 2008.

[10] Y. Matsushita, S. Lin, S. B. Kang, and H. Y. Shum. Estimating intrinsic images from image sequences with biased illumination. *In: Proceedings of ECCV*, pages 274–286, 2004.

[11] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Illumination normalization with time-dependent intrinsic images for video surveillance. *In: Proceedings of IEEE CVPR*, pages 3–10, 2003.

[12] J. M. Rubin and W. A. Richards. Color vision and image intensities: when are changes material. *Biological Cybernetics*, 45:215–226, 1982.

[13] L. Shen, P. Tan, and S. Lin. Intrinsic image decomposition with non-local texture cues. *In: Proceedings of IEEE CVPR*, pages 1–7, 2008.

[14] J. Shi and J. Malik. Normalized cuts and image segmentation. *In Proceedings of IEEE CVPR*, pages 731–737, 1997.

[15] M. F. Tappen, E. H. Adelson, and W. T. Freeman. Estimating intrinsic component images using non-linear regression. *In: Proceedings of IEEE CVPR*, pages 1992–1999, 2006.

[16] M. F. Tappen, W. T. Freeman, and E. H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(9):777–792, 2005.

[17] Y. Weiss. Segmentation using eigenvectors: A unifying view. *In Proceedings of IEEE ICCV*, pages 975–982, 1999.

[18] Y. Weiss. Deriving intrinsic images from image sequences. *In: Proceedings of IEEE ICCV*, pages 68–75, 2001.