Intrinsic Image Decomposition from Pair-wise Shading Ordering

Yuanliu Liu, Zejian Yuan, Nanning Zheng

Institute of Artificial Intelligence and Robotics
Xi'an Jiaotong University

Abstract. An image is composed by several intrinsic images including the reflectance and the shading. In this paper, we propose a novel approach to infer the shading image from shading orders between pairs of pixels. The pairwise shading orders are measured by two types of methods: the brightness order and the low-order fittings of local shading field. The brightness order is a non-local measure, which does not rely on local gradients, and can be applied to any pair of pixels. In contrast, the low-order fittings are effective for pixel pairs within local regions of smooth shading. These methods are complementary, and they together can capture both the local smoothness and non-local order structure of shading. Further, we evaluate the reliability of these methods by their robustness to perturbations, including the errors in reflectance clustering, the variations of reflectance and shading, and the spatial distances. We adopt a strategy of local competition and global Angular Embedding to integrate pairwise orders into a globally consistent order, taking their reliability into account. Experiments on the MIT Intrinsic Image dataset and the UIUC Shadow dataset show that our model can effectively recover the shading image including those deeply shadowed areas.

1 Introduction

An image is produced by several factors jointly, including the reflectance of the material, the shape of the surface, the positions and the colors of the illuminants and the parameters of the camera. Barrow and Tenenbaum [1] proposed to decompose a single image into intrinsic images, each of which captures a distinct aspect of the scene. The shading image captures the incident illumination at each pixel, while the reflectance image reflects the albedo of the surfaces. However it is essentially an underconstrained problem to recover the shading and reflectance from a single image. To solve this problem, additional constraints expressing the properties of the scene and the objects are needed. Most widely used properties include the local smoothness of shading [2][3][4][5][6], the local smoothness of reflectance [3][7][8][6] and the global sparsity of reflectance [4][7][8][9]. However, the smoothness of shading are not applicable to pixels separated by a shadow edge, while the smoothness of reflectance will be broken by the albedo change. The sparsity of global reflectance are not valid for complex scenes containing too many colors. How to add constraints to proper variables remains an open problem.

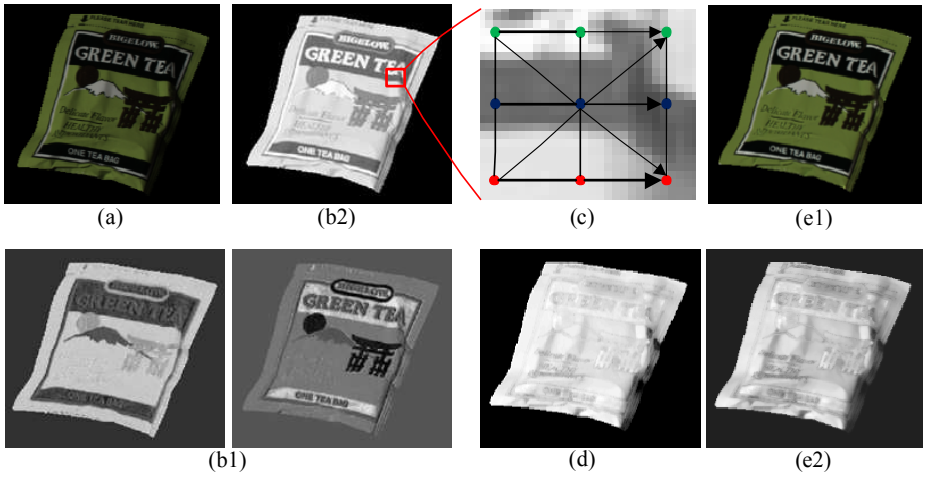

Fig. 1. The flow chart of decomposing a single image into shading and reflectance. (a) The input image; (b1) The projection of the log image onto the 2D shadow-free plane, and pixels with similar reflectance have similar values in this plane; (b2) The brightness map got from projecting the log image along the norm of the shadow-free plane; (c) The pairwise order of the shading intensity. Dots with the same color represent pixels with similar reflectance. The arrows point from the higher shading intensities to the lower ones. The lines without any arrow indicate nearly equal shading intensities. The width of the lines indicate the reliability of the estimations, while thicker ones stand for higher reliability; (d) The shading map in log space, which is the output of Angular Embedding; (e1) and (e2) are the recovered reflectance and shading image, respectively.

An initial work is the Retinex theory proposed by Land and McCann [2]. It assumes that small gradients in images are caused by shading variations, while large gradients are resulted from reflectance changes. By ignoring the edges corresponding to shading changes, a reflectance image can be recovered by integrating over the left gradient field. The edge types is classified by a threshold of the magnitude of gradients. This kind of classification is more-or-less inaccurate. Some shadow edges are quite strong under certain circumstances [10], while the reflectance edges between similar colors are relatively weak. Another way to classify edges is to utilize shadow-free color spaces [11][12][13][14][9]. Basically, if an edge appears in the raw image but not the shadow-free image, it should be a shading edge. Other methods for classifying edges include using classifiers trained on representative patches with shading or reflectance changes [15] [16], and utilizing texture information [3]. All these edge-based methods suffer from a problem that a single misclassified edge will provoke errors to a wide area of the reflectance image during integration [17].

Another stream of research combines different types of constraints softly by additive energies [3][7][4][5][8][6][18]. The energy of smoothness constraints are often modeled as the negative log of normal distributions over the gradients of shading and reflectance. The global sparsity of reflectance can be captured by

minimizing either the Rényi Entropy of reflectance [8], the cluster-wise variation in a shadow-free space [4][19] or the number of different reflectance values [7]. An energy minimization process ensures that the recovered reflectance and shading image best satisfy the synthesized constraints. Although these methods avoid hard classification of gradients, they bring in a new problem that different kinds of constraints are hinged together and relaxed for compromission. As a result, the element constraints may not be satisfied accurately. More specifically, the normal distribution of shading gradients tend to smooth the intensive shadow edges, while the smoothness of reflectance will blur the texture.

We propose to infer the shading image from shading orders, which capture not only the shading smoothness between nearby pixels but also the difference between distant pixels or those separated by shadow edges. The flow chart of our method is shown in Fig. 1. To estimate the shading orders, we introduce a brightness measure derived from the Bi-illumination Dichromatic Reflection Model (BIDR) [20]. The brightness has a linear relationship to the shading intensity, while the reflectance determines the bias. Combining brightness with different properties of the scene results in different estimation methods. The sparsity of reflectance ensures that the pixels can be clustered to a limited number of categories. For pixels with the same reflectance, the shading orders can be estimated from the brightness orders directly, since their biases will be canceled out. Unlike the edge-based methods, the brightness order does not rely on gradients that are sensitive to noise and image blur. It does not add any prior distribution to the shading either, so it can preserve different kinds of shading changes including those sharp shadow edges. For pixels with different reflectance, we first estimate the difference of biases between different categories. The smoothness of shading implies that within a small patch the shading is nearly constant, so the bias difference equals to the brightness order. For pixel pairs within local areas under smooth shading, their shading orders can also be estimated by low-order (constant or linear) models of the local shading image. Thanks to the linear relation between shading and brightness, the coefficients can be easily got by fitting the local brightness image.

The estimation methods above are complementary, each of which aims at specific image regions that the underlying properties are valid. Unlike the weighted summation used by the MRF models, we select the most reliable estimation for each pair of pixels, so unreliable estimations will not interfere the results. The reliability of the estimations are determined by the validity of the underlying properties, which is evaluated based on multiple cues, not limited to the magnitude of gradients. The densely sampled local shading orders together with their reliability are fed to Angular Embedding (AE) [21], resulting in a globally consistent shading image. AE uses complex matrix to encode pairwise orders and their confidences simultaneously. It uses spectral decomposition to get a near-global optimal solution. Moreover, it adopts a cosine error function, which is proved to be more robust to outliers than the traditional L1 or L2 errors [21].

The paper is organized as follows. In Sec. 2, we begin with projecting the raw image into a shadow-free plane and a brightness dimension. Then we estimate

the shading order either from the brightness order or a low-order fitting of the local shading map. The confidences of these estimations are evaluated in Sec. 3, and the process of inferring the global shading orders is described in Sec. 4.

2 Pairwise Order of Shading from Shadow Free Projection

In the log space, a surface with only diffuse reflection can be represented by

$$\log I^i = \log F^i C_b^i L_a^i + \log\left(\frac{\gamma}{M^i} + 1\right) \tag{1}$$

where the superscript $i \in \{r, g, b\}$ indexes the RGB channels. F is the response of the camera sensor. C_b is the body reflection. L_a is the ambient illumination, and $M = L_a/L_d$ is the ratio between the ambient illumination and the direct illumination. BIDR [20] assumes that both of the direct and ambient illuminants are constant across the scene, so the illuminant ratio M is constant. When there are multiple direct illuminants with the same color, their effect can be summed up. $\gamma \in [0, 1]$ is the shading intensity.

The BIDR model [20] delivers a shadow-free plane whose normal direction is:

$$w^i = \frac{\log(\frac{1}{M^i} + 1)}{\sqrt{\sum_i (\log(\frac{1}{M^i} + 1))^2}} \tag{2}$$

It is the direction of the parallel lines pointing from the dark pixels (whose $\gamma = 0$) to the bright ones (whose $\gamma = 1$) with the same body reflectance C_b. We call it the brightening direction. See Fig. 2 for an example. The pixels are projected into the space spanned by the dimension along w and the plane UV perpendicular to it:

$$\begin{aligned} I^w &= \log I \cdot w \\ I^u &= \log I \cdot u = \log F C_b L_a \cdot u \\ I^v &= \log I \cdot v = \log F C_b L_a \cdot v \end{aligned} \tag{3}$$

where u and v are unit base vectors of the UV plane. $\cdot$ is the inner product. We can see that the projections I^u and I^v are invariant to the shading intensity, and pixels with similar reflectance C_b are staying closely together. The brightening direction is assumed to be exactly the direction that the distribution of the pixels in its null space gets the minimum entropy [22][20]. Fast Gaussian Transform is adopted to calculate the Rényi Entropy efficiently [22]. Histogram-based techniques can further accelerate the process [23].

The shading intensity is fully captured by the projection I^w, and we call it the brightness. To further analyze the relation between the brightness and the shading intensity, we first approximate the log function of shading intensity in Eqn. 1 by the following linear function:

$$\log(\frac{\gamma}{M^i} + 1) \approx \frac{\gamma}{M^i} \tag{4}$$

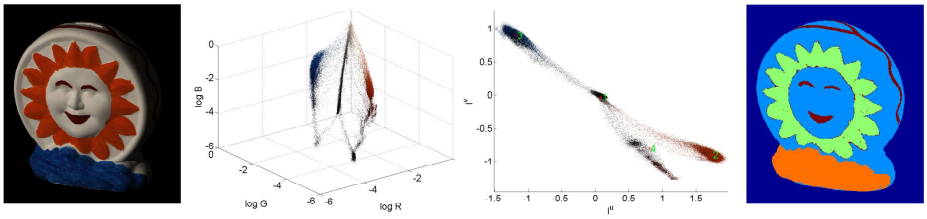

Fig. 2. The brightening direction and shadow-free plane. (a) The raw image; (b) The pixels in log RGB space. The white, orange, blue and dark red pixels form 4 nearly parallel lines; (c) Projections of pixels on the shadow-free plane. The shadow-free plane is perpendicular to the brightening direction. We can see that the pixels fall into 4 groups on the shadow-free plane, each for a distinct color. The red stars indexed by 1∼ 4 are cluster centers for white, orange, blue and dark red pixels, respectively; and (d) The cluster image. The pixels are successfully categorized by reflectance.

which is valid when $\frac{\gamma}{M^i}$ is small. Then the brightness can be rewritten to be:

$$\begin{aligned} I^w &= \log I \cdot w \\ &\approx \sum_i \log F C_b L_a \frac{\frac{1}{M^i}}{\sqrt{\sum_i \frac{1}{(M^i)^2}}} + \sum_i \frac{\frac{\gamma}{M^i}\frac{1}{M^i}}{\sqrt{\sum_i \frac{1}{(M^i)^2}}} \\ &= I^b(C_b) + \tilde{\gamma} \end{aligned} \tag{5}$$

The brightness I^w is composed of two parts: the ambient reflectance $I^b(C_b) = \sum_i \log F C_b L_a \frac{\frac{1}{M^i}}{\sqrt{\sum_i \frac{1}{(M^i)^2}}}$ that is a function of the body reflectance C_b, and the scaled shading intensity $\tilde{\gamma} = \sqrt{\sum_i \frac{1}{(M^i)^2}}\gamma$. The scale factor will not affect the shading order, so we just omit it for now. It will be recovered in Sec. 4.1.

The linear relationship between brightness and shading intensity is the basis for estimating shading orders. Consider a pair of pixels at positions p and q. According to Eqn. 5, the shading order can be estimated by:

$$O(p,q) = \tilde{\gamma}(p) - \tilde{\gamma}(q) = (I^w(p) - I^w(q)) - (I^b(p) - I^b(q)) \tag{6}$$

The ambient reflectance I^b are unknown biases varying with the body reflectance C_b. If a pair of pixels have the same body reflectance, their shading order is equal to their brightness order. Otherwise we need to know the difference of their ambient reflectance beforehand. We estimate the difference of ambient reflectance between different categories of pixels according to the sparsity of global reflectance, as described later in this section. However, the difference of ambient reflectance cannot be accurately estimated for some images, then we resort to the shading-smoothness-based methods.

In natural scenes, the shading intensities vary smoothly in most parts of the images. This property suggests that we can fit the local shading map by low-order functions.

First-order Smoothness (FS). The normal directions of flat surfaces change slowly, so the angle between the incident light and the normal direction will not change too much. According to the cosine law of the Lambertian reflection, the shading intensity will not change too much either. We assume that the first-order derivative of the shading field is almost 0 everywhere. Consequently, the adjacent pixels have identical shading intensity. This assumption is valid for nearly flat surfaces, when no shadow boundaries occur on them.

Second-order Smoothness (SS). For a local area of a smooth surface, the normal direction rotated smoothly. As a result, the shading intensity will change smoothly. We assume that the second-order derivative of the shading field is close to 0, so we can fit the local shading field centered at p by a linear function. We further assume that the adjacent pixels of p share the same body reflectance with p. Under this assumption, the slope of the linear model $\frac{\partial \tilde{\gamma}(p)}{\partial p} = \frac{\partial I^w(p)}{\partial p}$, where $\frac{\partial I^w(p)}{\partial p}$ is the first-order derivative of I^w evaluated at p.

Formally, we can estimate the order of shading intensity O between pixels p and q in the following ways:

$$O(p,q) = \begin{cases} I^w(p) - I^w(q) & \text{if } C_b(p) = C_b(q) \\ I^w(p) - I^w(q) - (I^b(p) - I^b(q)) & \text{if } C_b(p) \neq C_b(q) \\ 0 & \text{if } q \in N(p) \\ \dfrac{\partial I^w(p)}{\partial p} \cdot (p - q) & \text{if } \dfrac{\partial^2 (I^w(p))}{\partial p^2} \approx 0 \end{cases} \tag{7}$$

where $N(p)$ is the neighborhood of p, and $p - q$ is the spatial distance between p and q. In practice, we calculate the derivative and the spatial distance in horizontal and vertical directions separately. Notice that, the preconditions of the estimations are not mutual exclusive, so different methods may be applicable to the same pair of pixels. We need to choose the most reliable estimation, whose preconditions are best satisfied. See Sec. 3 for details. In the mean time, the preconditions together cover all possible situations, so we can find at least one suitable method for each pair of pixel. The redundancy and completeness of these methods result in robust estimations of pairwise shading orders.

According to Eqn. 7, the differences of ambient reflectance I^b are needed for estimating shading order between pixels with different body reflectance. It is infeasible to calculate the absolute value of I^b due to several unknown factors of it (Eqn. 5). Instead, we cluster the pixels by body reflectance, and estimate the difference of ambient reflectance between different categories. We assume that pixels within a small patch have similar shading intensities. According to Eqn. 6, if the shading intensities $\tilde{\gamma}$ are the same, the difference of ambient reflectance will be equal to the difference of brightness. Fig. 3 gives an example with two categories. The image is divided into dense grids with 10 pixels in each side. We calculate the difference of ambient reflectance between the categories within each grid. Then we generate a histogram of those grid-wise measures, and take the highest peak to be the final estimation. The reliability of the estimation P is set to be the height of the highest peak accordingly.

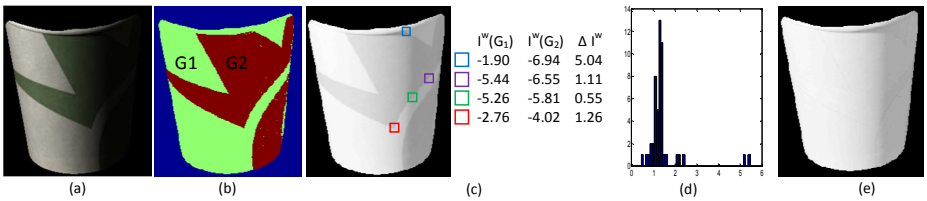

Fig. 3. Estimating the difference of ambient reflectance between categories. (a) The image. (b) The cluster label image. Green: Pixels from category $G1$; and Red: Pixels from category $G2$; (c) Left: The brightness image I^w with some representative patches indicated by the squares. Right: The median brightness of the two categories within each patch as well as their difference. The difference got from the blue square is an outlier, since there is a shadow edge inside it. So does the green one. (d) The histogram of the patch-wise differences of ambient reflectance between the two categories. The grids with only one category of pixels are ignored. The peak of the histogram is selected to be the estimated difference of ambient reflectance, which is 1.3 for this image. (e) The estimated shading intensity. It is got from adding the brightness of category G_2 by the difference of ambient reflectance 1.3 while keeping the brightness of category G_1 unchanged.

When there are multiple categories in the image, we need to estimate the difference of ambient reflectance between each pair of categories. However, some categories are not close enough in the image plane, such that none of the local patches contain pixels from both of these categories. To bridge the gap, we resort to the other categories lying between. We build an undirected graph $G = (V, E)$, where V is the set of nodes representing the categories, and E is the set of edges. The weight of the edge $E_{s,t}$ between node s and t is set to be $1/P_{s,t}$, where $P_{s,t}$ is the confidence of the estimation of their ambient reflectance order as described before. When $P_{s,t}$ is 0, it means category s and t are not adjacent and $E(s,t)$ will be cut off. We can get an estimation of the ambient reflectance difference between any two nodes by summing up the ambient reflectance differences along the path connecting them. To get the most consistent pairwise difference, we extract the Minimum Spanning Tree (MST) of the graph G. The MST ensures that there is one and only one path between any pair of nodes, so the difference of ambient reflectance between the nodes can be uniquely determined. In the mean time, the summation of confidences of the pairwise differences are maximized.

The sparsity of the reflectance spectra within a single image [24] implies that we can cluster the pixels into a small number of categories by their body reflectance. Notice that, pixels on the shadow-free plane UV are well organized by their reflectance (see Fig. 2(c) for an example). A simple k-means is used to cluster the pixels by reflectance in the shadow-free plane UV. The number of clusters is set to be the number of peaks (local maxima) in the 2D histogram of I^u and I^v. The bin size of the histogram is empirically set to be 0.03.

3 The Reliability of the Pairwise Orders

We can get several estimations of the shading order from the methods described in Eqn. 7. These methods are designed according to certain properties of the scene, such as the smoothness of shading field. However, these properties may be invalid for certain parts of the scene, so the estimations from the proposed methods are more or less different from the ground-truth. We analyze the robustness of the properties to the perturbations that may happen in the scene or to local areas. Then we evaluate the validity of these properties for each pair of pixels through calculating the joint probability of the occurrences of the perturbations they are not robust to.

The confidence of individual method is calculated by a Noisy-Or model, which is the probability of all its preconditions being satisfied:

$$C_m(p,q) = \prod_{n \in \mathcal{C}_m} 1 - P_n(p,q) \tag{8}$$

where m belongs to the set of methods in Eqn. 7, namely the Brightness Order (BO), the Brightness Order minus Bias difference (BOB), the First-order Smoothness (FS) and the Second-order Smoothness (SS) of shading. $\mathcal{C}_m$ is the set of perturbations that the estimation method m is not robust to, as listed in Tab. 1. The probability $P_n(p,q)$ measures the probability of the perturbation n occurring between pixel p and q. We first calculate a distance measure between the pair of pixels according to each feature, and translate the distance into probability by a sigmoid function in the form of $sigm(x;w) = \frac{2}{1+e^{-wx}} - 1$. Here w is a positive scalar, which is a parameter of the model.

Table 1. The robustness of the methods for estimating the pair-wise shading orders with respect to different perturbations.

Perturbations	BO	BOB	FS	SS
Clustering Error (CE)	Yes	No	Yes	Yes
Local Reflectance Variation (LRV)	No	No	Yes	No
Reflectance Change (RC)	No	Yes	Yes	Yes
Shadow Edges (SE)	Yes	Yes	No	No
Spatial Distance (SD)	Yes	Yes	No	Moderate

Cluster Error (CE) denotes the uncertainty of the clustering results in the shadow-free plane. When the pixels are incorrectly assigned to the categories, their ambient reflectance cannot be well represented by the estimated ambient reflectance of their categories. We model each cluster as a multivariate normal distribution, and calculate the probability of each pixel belonging to the category that the pixel is assigned to. The Cluster Error for a pair of pixels is represented as the probability that at least one of the pixels does not belong to its assigned category. In addition, the difference of ambient reflectance between color categories are inaccurately estimated sometimes, causing step edges at the

boundaries between different categories. The final Cluster Error is calculated by:

$$P_{CE}(p,q) = (1 - P_C(p)P_C(q)) \cdot sigm(e_{\hat{\gamma}}(p,q); w_1) \tag{9}$$

where P_C denotes the probability of a pixel belonging to the category that it is assigned to. The subscript $\hat{\gamma}$ stands for the brightness map I^w minus the categorical ambient reflectance. Inspired by the similarity measure in [25], we set the strength of the step edge $e_{\hat{\gamma}}(p,q)$ to be the largest magnitude of the gradients of $\hat{\gamma}$ evaluated at the pixels along the path connecting p and q.

Local Color Variance (LCV) reflects the textureness of the local regions. In highly textured areas, the blur effect over different colors will produce various mixed colors, whose ambient reflectance are unpredictable. In implementation we represent LCV as follows:

$$P_{LRV}(p,q) = sigm(\max(\sigma(I^{uv}(p)), \sigma(I^{uv}(q))); w_2) \tag{10}$$

where $\sigma(I^{uv}(p))$ is the sum of the standard variations of I^u and I^v within the 3x3 window centered at pixel p.

Reflectance Change (RC) is modeled as follows:

$$P_{RC}(p,q) = sigm(d_{uv}(p,q); w_3) \cdot sigm(e_w(p,q); w_4) \tag{11}$$

where d_{uv} is the geometric distance in shadow-free space UV. Similar to $e_{\hat{\gamma}}(p,q)$, $e_w(p,q)$ is the magnitude of the edge lying between p and q in the brightness image. It is used to distinguish regions with similar color but different reflectance intensities, especially achromatic regions like white and gray. The underlying assumption is that the edges caused by reflectance change are often stronger than those caused by shading changes.

Shadow Edges (SE) are caused by occlusions of the direct light. They are always quite intensive compared to those shading changes caused by surface normal change. We measure the probability of existing a shadow edge between pixel p and q by:

$$P_{SE}(p,q) = sigm(e_{\hat{\gamma}}(p,q); w_5) \tag{12}$$

The biased shading intensities $\hat{\gamma}$ is defined the same as that in Eqn. 9, but scaled by a different weight.

Spatial Distance (SD) affects the accuracy of fitting the local shading by low order models. The probability of a pair of pixels being far away from each other is calculated by:

$$P_{SD}(p,q) = sigm(d_s(p,q); w_6) \tag{13}$$

where d_s is the geometric distance.

4 Infer the Global Shading Map from Pairwise Shading Orders

The local estimations from different methods are combined through selecting the most confident one of them, while the confidence is calculated accordingly:

$$C(p,q) = \max_m C_m(p,q) \tag{14}$$

Now we have got a matrix O of the pairwise orders of the shading intensities together with their confidence matrix C. To get a global shading map, we need to align these pairwise measurements. Here we use the Angular Embedding method [21] to embed the shading intensities of the pixels into the angular space, such that the global orders of the embedded points keep their pairwise orders.

Let $Z(p) = e^{i\gamma(p)}$ denote the embedding of the shading intensity of pixel p on the unit circle in the complex plane. Here $i = \sqrt{-1}$. The norm of $Z(p)$ is always 1, and the angle $\Theta(p, q)$ from $Z(p)$ to $Z(q)$ is the order of shading intensity between p and q. It is expected that $\Theta(p, q)$ is consistent with the pairwise shading order $O(p, q)$, when it gets a high confidence. Angular Embedding minimizes the difference between the embedding $Z(p)$ and the estimation of it from its neighbors, $\bar{Z}(p)$ weighted by its total confidence $D(p, p)$:

$$\begin{aligned} &\min_{Z} \sum_{p} D(p,p) \cdot \|Z(p) - \bar{Z}(p)\|^2 \\ &\text{s.t. } \|Z(p)\| = 1, \forall p \end{aligned} \tag{15}$$

where D is a diagonal degree matrix:

$$D(p,p) = \frac{\sum_q (C(p,q))}{\sum_{p,q} (C(p,q))} \tag{16}$$

and

$$\bar{Z}(p) = \frac{1}{\sum_q (C(p,q))} \sum_q C(p,q) \cdot Z(q) \cdot e^{iO(p,q)} \tag{17}$$

This optimization problem is hard to solve, since it has n constraints, where n is the number of sampled pixels. In implementation, the constraints are relaxed to be $Z'DZ = 1'_n D 1_n$. To make the optimization tractable, we consider only the orders between nearby pixels. The neighborhood is set to be a square of 30 pixels in each side. The confidence of the orders between a pixel with any pixel outside its neighborhood is set to be 0.

The optimization problem in Eqn. 15 is solved by a spectral partitioning algorithm [25] with complex-valued eigenvectors. The solution is the angle of the first eigenvector $\angle Z_0$ which has the smallest eigenvalue. It should be pointed out that the angles of the points on the unit circles are within the scope of $[-\pi, \pi]$. We need to ensure that the angle between any pair of $Z(p)$ and $Z(q)$ is no larger than 2π, otherwise the points may overlap with each other. We scale the brightness image I^w and the the ambient reflectance images I^b by a positive scalar in the beginning, such that the difference of brightness is no larger than 2π. The scaling operation will not disturb the order of Z, and it will be recovered in Sec. 4.1.

The angles $\angle Z_0$ keep the pairwise order of the shading intensities, but the absolute values of the angles have no determined mappings to the shading intensities. Angular Embedding allows the points to rotate as a whole around the original point. Fig. 4 gives an example. We can see that the angles of the brightest pixels in this image appear in the interval of $[-\pi, -0.5\pi]$, lower than

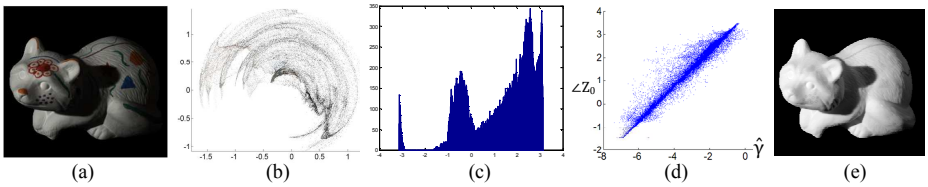

Fig. 4. An example of the Angular Embedding results. (a) The image; (b) The output embedding with the smallest eigenvalue. There is a gap between the brightest and the darkest pixels; (c) The histogram of the angles of embedding. The bins with zero counts are generated by the gap. The pixels fell into the bins to the left of the zero-count bins will be shifted to the right by 2π; (d) The relation between the output angles and the biased shading intensities $\hat{\gamma}$. Most of them stay roughly in a line; and (e) The recovered shading image in log space.

those of the dark pixels. As a result, the brightest pixels will be mistaken to be "shadowed". To solve this problem, their angles need to be increased by 2π instead. Notice that, the darkest pixels and the bright pixels are always separated by a noticeable gap on the circles in the complex plane. The gap can be easily located by the consecutive empty bins of the histogram of the angles $\angle Z_0$. All the pixels whose angles are smaller than the darkest pixels will be increased by 2π. After that, the angles will have a roughly linear relation to the biased shading intensities $\hat{\gamma}$ (See Fig. 4 (d)). That is because most of the input pairwise shading orders are calculated by the brightness orders (through the BO or BOB method), while the brightness has a global linear relationship to the shading intensities. We further normalize the angles $\angle Z_0$ to be within the interval of [0, 1], which produce the final shading intensities γ (See Fig. 4 (f)).

4.1 Recovering Shading and Reflectance

The reflectance image is regarded to be the brightened image under full direct illumination. That is, the shading intensity γ is 1 for every pixel of the image. From Sec. 4 we have already got the shading intensities, now we can recover the reflectance by raising the shading intensities of all the pixels to 1 along the brightening direction. According to Eqn. 1 and 4, the log image $logI$ is close to a linear function of γ with an unknown slope k. The recovered reflectance image is:

$$R^i = e^{logI^i + kw^i(1-\gamma)} = I^i e^{kw^i(1-\gamma)} \tag{18}$$

while the shading map can be got by:

$$S^i = e^{kw^i(\gamma-1)} \tag{19}$$

We find a reasonable slope k through a voting process. For each channel i and each pixel p, we find the scalar $\tilde{k}^i(p)$ that makes the recovered $R^i(p)$ equal to the maximum value of I^i over the image. Then we calculate the histogram of $\tilde{k}^i(p)$, and record the bin that got the most votes, denoted by $\tilde{k}^i$. Here we empirically set the bin size to be 0.05. We set the slope to be $k = \text{median}_i \tilde{k}^i$.

5 Experiments

We evaluate our method on the MIT Intrinsic Images dataset [26], which is a widely used benchmark for evaluating intrinsic decomposition methods. The images are taken in a controlled environment, where the direct illuminations are always white and the ambient illuminations are limited through painting the background to be black. We further test our method on the UIUC shadow dataset [27], where the direct illuminants and the ambient illuminants are uncontrolled.

To quantitatively evaluate the results, we use both the Mean Squared Error (MSE), the Local Mean Squared Error (LMSE) [26], the absolute LMSE (aLMSE) and the correction [28]. Among these metrics, correlation and MSE measure the error in a global way, while LMSE and aLMSE take an average of local errors on small image windows. For each image, the performance of shading image and reflectance image are calculated separately and the average of them is taken to be the result. The final result is the average of the performances over all the images. The main parameters of our model are the positive weights of the sigmoid function in Sec. 3. In our experiments we set w_1 to be $ln3/0.1$, which ensures that the sigmoid function maps a step edge of strength 0.1 to a probability of 0.5. We set $w_2 \sim w_5$ to be $ln3/0.2$, $ln3/0.08$, $ln3/0.1$ and $ln3/0.1$, respectively. Especially, we set the w_6 of the FS method to be twice as much as that of the SS method. We find the medium of the spatial distances of all the pixel pairs $\bar{d}_s$, and set w_6 to be $6ln3/\bar{d}_s$ for the FS method. These parameters are used for all the images of our experiments.

5.1 Results on MIT Intrinsic Image dataset

We take the image with only diffuse reflection as input, since specular reflection is out of the scope of this paper. We compare our method to the state-of-art together with some classic approaches as listed in Tab. 2. For each method, a single group of best parameters are used for all the images. Our method achieves the best performance on the correlation and LMSE metric. The SIRFS model gets the lowest MSE [8], but their method relies on priors got from training images. Weiss [29] gets the best aLMSE, but their method takes image sequences captured under different illuminants as input. Among the single-image based methods without training, our model gets the best performance over all the metrics. Specifically, our method performs much better than the method of Color Retinex [26]. One important reason is that our method explicitly take the ambient illuminant into consideration (although it is very weak in this dataset), which results in a better shadow-free plane than that used by Color Retinex. Fig. 5 gives some concrete examples. One important advantage of our method is that we can recover the reflectance of very dark areas. The reason is that our model carries out the local fusion of estimations from different methods by a maximization operation (Eqn. 14), which preserves the large shading orders between pixels separated by shadow edges. In comparison, the method of Gehler et al. [4] often smoothes the shading map through adding strong smoothness constraints, leaving residuals of shadows in the recovered reflectance image. This problem is

even more serious for the SIRFS model [8], since the smoothed surfaces of the objects in this model always generate smooth shading image. The method of Jiang-HA [28] is based on the global correlation between the mean luminance and luminance amplitude, which recovers the global shading distribution well but not the details in local regions (e.g., the symbols on the raccoon).

	Correlation	MSE	LMSE	aLMSE
Grey Retinex [2]	0.6494	0.1205	0.0329	0.3373
Tappen et al. [16]	-	-	0.0390	-
Color Retinex [26]	0.7146	0.1108	0.0286	0.2541
Jiang-A [28]	0.6184	0.1533	0.0421	0.3988
Jiang-H [28]	0.5829	0.1524	0.0483	0.3476
Jiang-HA [28]	0.6109	0.1579	0.0454	0.3631
Shen-SR [7]	0.7259	0.1223	0.0242	0.2454
Shen-SRC et al. [7]	-	-	0.0204	-
Gehler et al. [4]	0.7748	0.0985	0.0244	0.2544
Serra et al. [17]	0.7862	0.0834	0.0340	0.2958
Li et al. [18]	-	-	0.0190	-
Chang et al. [19]	-	-	0.0229	-
Ours	**0.8582**	0.0684	**0.0189**	0.2252
Weiss [29]	0.7709	0.0900	0.0210	**0.1953**
SIRFS [8]	0.8095	**0.0567**	0.0279	0.2329

Table 2. Results on the MIT Intrinsic Images dataset. Higher correlation and lower MSE, LMSE and aLMSE are better. The method SIRFS is evaluated on 8 images including cup2, deer, frog2, paper2, raccoon, sun, teabag1 and turtle, while the other images are used for training.

5.2 Evaluation under chromatic illuminations

We test our method's ability of handling both direct and ambient illuminants on the UIUC shadow dataset [27]. Fig. 6 shows several examples. We compare our method to the method of Jiang-HA [28] and the method proposed by Gehler et al. [4]. We also compare it to the region-pair-based shadow removal method proposed by Guo et al. [27]. For this method the shading map is replaced by a shadow map, in which black pixels indicate shadows and gray ones indicate penumbra. We can see that our method successfully recovers the shading map, not only for the cast shadows but also for the self shading (the first image of Fig. 6). For the image of the cup, only our method recovers the deeply shadowed area inside the cup. In the outdoor scenes (the last 3 columns of Fig. 6), the ambient illuminants are usually the blue sky, which turns the shadowed areas more blueish than the bright areas. Our model recovers their reflectance by lighting the dark pixels along the yellowish brightening direction, while the other intrinsic decomposition methods often fail to recover the color of them.

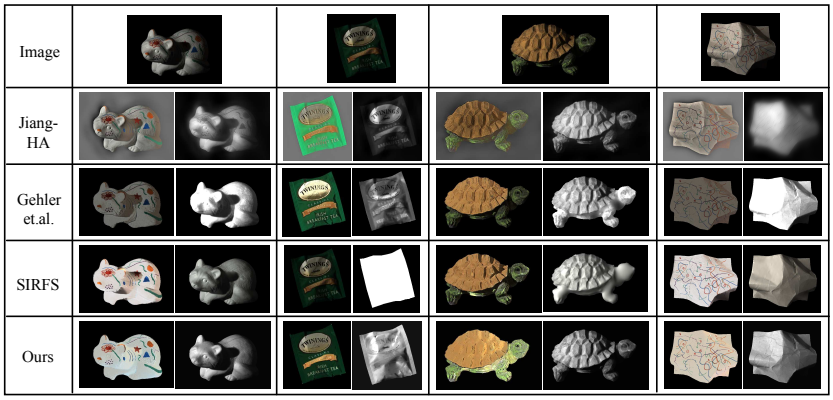

Fig. 5. Typical results on the MIT Intrinsic Images dataset.

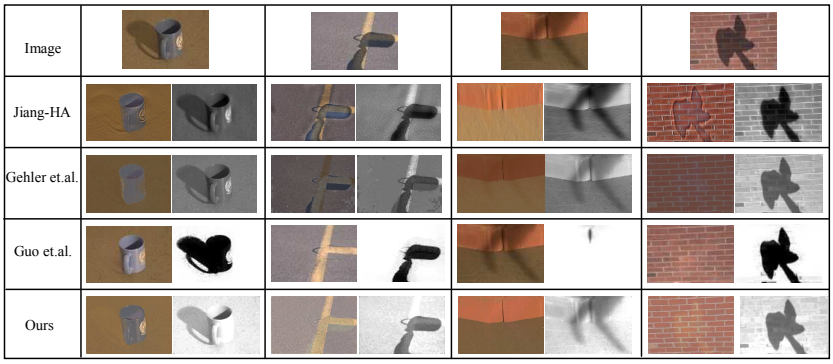

Fig. 6. Typical results on the UIUC shadow dataset.

6 Conclusions

We proposed a model to decompose a single image into reflectance and shading by pairwise shading orders. It overcame the limitations of edge-based methods that rely solely on local gradients. The shading orders were estimated by several individual methods, each of which aimed at specific types of image regions. We experimented on different kinds of images, and achieved promising results on most of them. We adopted a new strategy of integrating local cues, that was local competition and global collaboration. The local competition prevented the cues from different sources interfering each other. Especially, it kept the sharp shadow edges from being weaken by the shading smoothness constraints.

Acknowledgement. This work was supported in part by the National Basic Research Program of China under Grant No. 2012CB316400, and the National Natural Science Foundation of China under Grant No. 91120006.

References

1. Barrow, H.G., Tenenbaum, J.M.: Recovering Intrinsic Scene Characteristics from Images. Computer Vision Systems. Academic Press (1978)
2. Land, E.H., McCann, J.J.: Lightness and retinex theory. Journal of the Optical Society of America **61** (1971) 1–11
3. Shen, L., Ping, T., Lin, S.: Intrinsic image decomposition with non-local texture cues. In: IEEE Conference on Computer Vision and Pattern Recognition. (2008) 1–7
4. Gehler, P., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. In Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F.C.N., Weinberger, K.Q., eds.: Advances in Neural Information Processing Systems. (2011) 765–773
5. Lee, K., Zhao, Q., Tong, X., Gong, M., Izadi, S., Lee, S., Tan, P., Lin, S.: Estimation of intrinsic image sequences from image+depth video. In: European Conference on Computer Vision. (2012) 327–340
6. Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: IEEE International Conference on Computer Vision. (2013)
7. Shen, L., Yeo, C.: Intrinsic images decomposition using a local and global sparse representation of reflectance. In: IEEE Conference on Computer Vision and Pattern Recognition. (2011) 697–704
8. Barron, J.T., Malik, J.: Color constancy, intrinsic images, and shape estimation. In: European Conference on Computer Vision. (2012) 57–70
9. Garces, E., Munoz, A., Lopez-Moreno, J., Gutierrez, D.: Intrinsic images by clustering. Computer Graphics Forum **31** (2012)
10. Huang, X., Hua, G., Tumblin, J., Williams, L.: What characterizes a shadow boundary under the sun and sky? In: IEEE International Conference on Computer Vision. (2011) 898–905
11. Mark, B.F., Drew, M.S., Brockington, M.: Recovering shading from color images. In: European Conference on Computer Vision, Springer-Verlag (1992) 124–132
12. Gevers, T.: Reflectance-based classification of color edges. In: IEEE International Conference on Computer Vision. (2003) 856
13. Finlayson, G., Hordley, S., Drew, M.: Removing shadows from images. In: European Conference on Computer Vision. Volume 2353. (2002) 129–132
14. Van de Weijer, J., Gevers, T., Geusebroek, J.M.: Edge and corner detection by photometric quasi-invariants. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 625–630
15. Tappen, M.F., Freeman, W.T., Adelson, E.H.: Recovering intrinsic images from a single image. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 1459–1472
16. Tappen, M.F., Adelson, E.H., Freeman, W.T.: Estimating intrinsic component images using non-linear regression. In: IEEE Conference on Computer Vision and Pattern Recognition. (2006) 1992–1999
17. Serra, M., Penacchio, O., Benavente, R., Vanrell, M.: Names and shades of color for intrinsic image estimation. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012) 278–285
18. Li, Y., Brown, M.S.: Single image layer separation using relative smoothness. In: The IEEE Conference on Computer Vision and Pattern Recognition. (2014)
19. Chang, J., Cabezas, R., Fisher III, J.W.: Bayesian nonparametric intrinsic image decomposition. In: Proceedings of the European Conference on Computer Vision. (2014)

20. Maxwell, B.A., Friedhoff, R.M., Smith, C.A.: A bi-illuminant dichromatic reflection model for understanding images. IEEE Conference on Computer Vision and Pattern Recognition (2008)
21. Yu, S.X.: Angular embedding: A robust quadratic criterion. IEEE Transactions on Pattern Analysis and Machine Intelligence **34** (2012)
22. Finlayson, G., Drew, M., Lu, C.: Entropy minimization for shadow removal. International Journal of Computer Vision **85** (2009) 35–57
23. Barron, J.T., Malik, J.: Shape, albedo, and illumination from a single image of an unknown object. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012) 334–341
24. Omer, I., Werman, M.: Color lines: image specific color representation. In: IEEE Conference on Computer Vision and Pattern Recognition. Volume 2. (2004) 946–953
25. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Conference on Computer Vision and Pattern Recognition (1997)
26. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In: IEEE International Conference on Computer Vision. (2009) 2335–2342
27. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. In: IEEE Conference on Computer Vision and Pattern Recognition. (2011) 2033–2040
28. Jiang, X., Schofield, A., Wyatt, J.: 5. In: Correlation-Based Intrinsic Image Extraction from a Single Image. Volume 6314 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2010) 58–71
29. Weiss, Y.: Deriving intrinsic images from image sequences. In: IEEE International Conference on Computer Vision. (2001) 68–75