Names and Shades of Color for Intrinsic Image Estimation

Marc Serra, Olivier Penacchio, Robert Benavente, and Maria Vanrell
Computer Vision Center / Computer Science Dept.
Universitat Autònoma de Barcelona
Campus UAB, Building O, 08193 Bellaterra (Barcelona), Spain
{ mserra , penacchio , robert , maria }@cvc.uab.cat

Abstract

In the last years, intrinsic image decomposition has gained attention. Most of the state-of-the-art methods are based on the assumption that reflectance changes come along with strong image edges. Recently, user intervention in the recovery problem has proved to be a remarkable source of improvement.

In this paper, we propose a novel approach that aims to overcome the shortcomings of pure edge-based methods by introducing strong surface descriptors, such as the color-name descriptor which introduces high-level considerations resembling top-down intervention. We also use a second surface descriptor, termed color-shade, which allows us to include physical considerations derived from the image formation model capturing gradual color surface variations. Both color cues are combined by means of a Markov Random Field. The method is quantitatively tested on the MIT ground truth dataset using different error metrics, achieving state-of-the-art performance.

1. Introduction

The decomposition of a scene into a set of intrinsic images containing one single physical characteristic was first proposed by Barrow and Tenenbaum in [2]. Scene intrinsic characteristics such as reflectance, shading, and depth provide useful information for improving computer vision tasks such as segmentation, tracking, and recognition.

In this paper, we focus on the decomposition of an image into its reflectance and shading components. Reflectance describes how the light is reflected at each point of a specific object and shading represents the perceived reflection due to the position of the light source and the object shape. Recovering these two characteristics from a single image I amounts to estimate a reflectance image and a shading image such that

$$I(x,y) = I_{\text{shading}}(x,y) \cdot I_{\text{reflectance}}(x,y). \quad (1)$$

This problem is clearly ill-posed because the number of unknowns is higher than the number of equations. Consequently, approaches for estimating image intrinsic components rely on making simplifying assumptions or adding constraints to the problem.

Most of the previous approaches are based on the assumption that large image derivatives are due to reflectance changes while low derivatives are caused by smooth illumination variations. Then, classifying image derivatives as being caused by either reflectance or shading variations, one can estimate the reflectance and shading images by integrating each type of derivatives. This assumption was the basis of the Retinex algorithm [14] and has been exploited latter in different ways such as directly thresholding gradients on color images [10], comparing derivatives of the original image with the derivatives of an illumination invariant image [9], using classifiers trained on a set of representative examples showing reflectance and shading changes [21, 20], or combining image gradients with local texture cues [18].

Some approaches have avoided the use of edges' information by combining steerable filters decompositions with texture and color cues [13], or by imposing local and global sparsity constraints on the reflectance representation [19]. Recently, a physical model of image formation has been used to directly derive intrinsic components of an image [3]

Finally, another way to simplify the problem is to include additional information by considering image sequences [23, 15] or using interaction with the user [7, 17].

From the analysis of the previous approaches we can draw some conclusions:

- The exclusive use of edges to recover surface reflectance is not sufficient since a small missclassified edge can provoke an error over a wide area. Surface attributes such as color and texture are essential cues to improve edge-based proposals.

- Results from methods that include user interaction suggest that top-down intervention yields a clear advantage for dealing with the ill-posed nature of re-

flectance recovery. Hence, we argue for the need of high-level attributes to describe image content.

- Few efforts have been done to exploit the information derived from the assumption that image formation obeys a specific physical model. Intrinsic image algorithms can benefit from these models since they account for changes in image appearance due to geometry and illumination of the scene.

At the light of the previous conclusions, we propose the introduction of color surface attributes based on color names instead of an edge-based approach. These attributes provide high-level information resembling top-down intervention in the reflectance recovery. Afterwards, we add a second descriptor, termed color-shade, that allows us to take into account physical considerations on color surface variations due to the geometry and lighting of a scene. This descriptor, which assumes Shafer's dichromatic reflection model for image formation [16], is introduced to address the lack of stability of the color-name descriptor in the presence of strong variations in the geometry or illumination of a scene. Color-name and color-shade descriptions are finally combined by means of a Markov Random Field.

The rest of the paper is organized as follows. Section 2 introduces the color cues that are used to recover reflectance and shading images as explained in sections 3 and 4. Section 5 presents results and concluding remarks are given in section 6.

2. Our approach

To explain our approach for intrinsic image recovery, we first introduce the color-name and color-shade descriptors. Next, we outline the conditional inference approach we adopt to combine these color cues.

2.1. Color-name descriptor

The color-name descriptor associates the linguistic terms humans use for describing objects to image colors. Basic color names were first defined by Berlin and Kay [6]. They were deduced from a large anthropological study based on speakers of 20 different languages and specific documentation from 78 other languages. The authors concluded that the universal basic color terms defined in most evolved languages are 11. Subsequent psychophysical experiments have generated data that allow these basic names to be accurately specified [4] and computationally implemented [5]. Accordingly, color-naming models provide perceptually-based quantizations of the RGB color space, which present a higher discrimination power with respect to usual chromaticities, as proven for classification tasks [11].

We use the color-naming model proposed by Benavente *et al*. [5], where a color name category is modeled as a

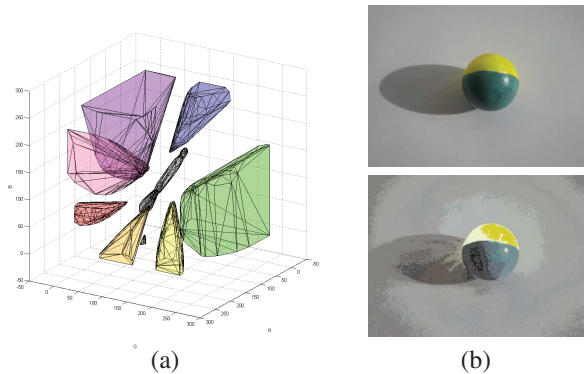(a)   (b)

Figure 1: Color names in the RGB space. (a) Plotted volumes represent those RGB's with probability equal to 1. (b) Image labeled with the color-name descriptor.

fuzzy set with a membership function that, given a color sample, assigns a value between 0 and 1 to represent the color-name membership. The model uses the set of names proposed by Berlin and Kay [6], namely $\mathcal{N} = \{$ *black, white, red, green, yellow, blue, brown, purple, orange, pink, grey*$\}$. Since the model forces the sum of all memberships to be 1 for any pixel, the membership values can be considered as probabilities. The color-name descriptor of a pixel r_i, denoted $ND(r_i)$, is a 11-dimensional real-valued vector whose components are the probabilities of labeling the given pixel with each one of the color names in $\mathcal{N}$. More explicitly,

$$ND(r_i)_j = p(N_j|r_i), \forall j = 1, ..., 11, \qquad (2)$$

where N_j is the j-th color name in the set of basic color names $\mathcal{N}$. Figure 1(a) shows the volumes of the RGB space where each of the 11 color names have probability 1.

The color-name descriptor has two interesting properties. First, it is relatively invariant to small photometric changes since wide areas of a single reflectance surface assume the same label, and small changes in shading only cause gradual changes in the descriptor. Next, it provides a sparse representation of color since very few coordinates of the 11-dimensional vectors are non-zero (usually up to three). Since this descriptor yields the labels of a conditional inference labeling problem (see Section 3), we only allow three coordinates to be non-zero and we further discretize the probability vector by quantizing the coordinates to $\{0, 0.25, 0.5, 0.75, 1\}$ while keeping the constraint that they sum to 1. This means that a color can be described with a maximum of three names (for instance, greyish blue-green), which is a perceptually consistent constraint since in the model very few colors are in the boundary of four color names [5]. Although with such restrictions 671 labels are theoretically possible, only 250 of them are actually used

since the others correspond to unfeasible combinations of colors (such as bluish yellow-purple). The set of labels is denoted by $\mathcal{L}$.

In the next sections we prove that this set of labels is a reliable sparse representation of color to recover reflectance. In Figure 1(b) we show an example of the color-name labels assigned to an image.

2.2. Color-shade descriptor

The color-shade descriptor is based on the method proposed by Vazquez *et al.* in [22]. In this work, the authors propose to describe scene reflectances by a Ridge Analysis of the color Distributions (RAD method). A ridge is a list of points connecting the local maxima of a color distribution in the RGB histogram space. In Figure 2(b), we show a 3D representation of the color distribution of the image in 2(a). The corresponding ridges (connected maxima) detected by the RAD method on the 4D color histogram distribution are shown in Figure 2(c), where we can see four ridges corresponding to the blue, red, orange and white parts of the image. By looking at the ridges, we can see how all the shades of each color are represented. For example, the white ridge spans colors from the lightest white to the darker gray present in the shadowed part of the object. In the ideal case, the RAD method provides a single ridge for each reflectance surface of the image.

The physical model underlying the RAD method is the dichromatic model described by Shafer in [16]. In this model, all the color variations of a surface, including shading effects and highlights, span a 2D plane in the RGB space which is defined by two vectors: one in the direction of the surface's albedo, and the other in the direction of the illuminant. Hence, the dichromatic model, and, therefore, the RAD method, provide a compact representation of all the variations of a single-color surface can present due to illumination changes and the geometry of the scene.

Given an image, the RAD method returns a set of ridges $\mathcal{R} = \{R_1, ..., R_n\}$. From this set of ridges, we define the color-shade descriptor of an image pixel r_i as

$$SD(r_i) = \arg \min_{R_j \in \mathcal{R}} dist(r_i, R_j), \qquad (3)$$

where $dist(r_i, R_j)$ represents the (Euclidian) distance between the RGB value r_i and the points of the ridge R_j.

To sum up, the ridges provide useful information for enhancing the color-name representation and allows us to deal with the variations of color names in the presence of strong illumination effects, i.e. shading and highlights. For example, two pixels belonging to the same reflectance object but with very different RGB values, e.g. one in a shadowed part of the object, the other in a brighter part, are connected by their nearest ridge. Following this approach, one can consistently name pixels within a single reflectance area allowing
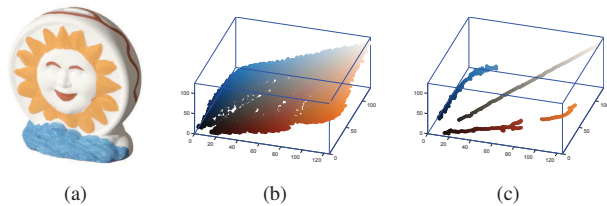

(a)   (b)   (c)

Figure 2: RAD method (a) Original image. (b) Color image distribution. (c) Detected ridges.

for shading changes.

2.3. Method outline

Our algorithm is based on the main assumption that a single material reflectance can be described by a unique color name provided by the descriptor introduced in section 2.1. Spatial coherence for this descriptor is then achieved by propagating evidence through a homogeneous MRF. The color name descriptor introduces two interesting properties:

- An accurate color edge localization, where relevant color edges are perfectly located by changes in color names.

- A meaningful surface interpretation based on standard prior knowledge compiled from psychophysical data.

However, when strong shading variations occur, irrelevant edges can appear within a homogeneous surface provoking a non-desirable over-segmentation. To deal with this problem, we use the color-shade descriptor introduced in section 2.2 to incorporate physical information to the previously defined MRF (by breaking its homogeneity). This second step stems from the assumption that changes in shading within an area of uniform reflectance yield to connected distributions of points in the RGB histogram. This step provides:

- An efficient combination criterion for the excessive segmentation of color names in shaded and near highlight areas.

- A relation between physical and name changes that could be exploited to estimate the color of the light source (beyond the scope of this paper).

In a third stage, once information from names and shades has been propagated, we modify the reflectance description provided by the MRF to match the intensities of the recovered reflectance to those of the original image. A scheme of the method is given in Figure 3.
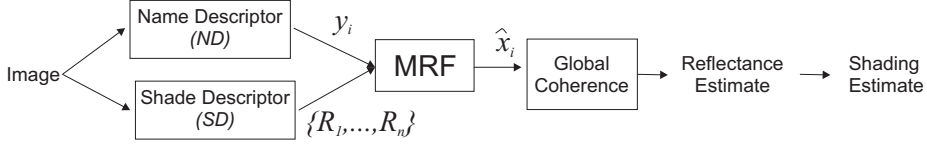
Figure 3: Block diagram of our method for intrinsic image estimation.

3. Reflectance recovery by MRF inference

In this section we present the MRF which combines the two color cues our method uses to recover reflectance.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the graph that represents the input image, where the set of vertices $\mathcal{V}$ correspond to random variables x_i associated to the set of pixels of the image (one node for each pixel), and $\mathcal{E}$ is the set of undirected edges representing relationships between pairs of adjacent pixels (using a 4-neighborhood system). The set of maximal cliques $\mathcal{Cl}$ is formed by the edges of the graph $\{x_i, x_j\}$, where i and j are adjacent pixels, and the cliques of the form $\{x_i, y_i\}$, for each pixel $i \in \mathcal{V}$, where y_i stands for the observation at pixel i.

Both random variables x_i and observations y_i are reflectance values as expressed by the color-name descriptor outlined in the previous section. Accordingly, the set of labels $\mathcal{L}$ is a set of 11-dimensional vectors.

The energy function of our MRF has the following form:

$$E(\mathbf{x}) = \mu \sum_{i \in \mathcal{V}} D(x_i, y_i) + (1 - \mu) \sum_{\{i,j\} \in \mathcal{E}} V(x_i, x_j), \quad (4)$$

where $D(x_i, y_i)$ is the singleton potential defined on each node x_i and $V(x_i, x_j)$ is the pairwise potential defined on a pair of neighboring pixels. The contribution of both terms in the global energy is weighted using a parameter $\mu \in [0, 1]$.

The labeling $\hat{\mathbf{x}}$ that minimizes equation 4 was found with the α-expansion graph cut algorithm [1] presented in [8].

In the following sections we detail how we define the potentials to integrate the information from the two color cues.

3.1. Singleton potential: color name

The singleton potential $D(x_i, y_i)$ measures to which extent the labeling $\mathbf{x}$ fits the observed data $\{y_i\}_{i \in \mathcal{V}}$. In practice, this potential can be interpreted as the cost of assigning x_i a label different from the label of observation y_i.

For computing the singleton potential, we chose the L_1 distance in the Euclidian space of 11-dimensional probabilities vectors:

$$D(x_i, y_i) = \|x_i - y_i\|_1, \forall i \in \mathcal{V}. \quad (5)$$

3.2. Pairwise potential: color shade

In classic MRFs, the set of pairwise potentials $V(x_i, x_j)$ measure the non-smoothness of the labeling $\mathbf{x}$ and can be interpreted as the cost of assigning different labels to neighboring pixels. These potentials are first defined using the Euclidian distance as $V(x_i, x_j) = \|x_i - x_j\|_1, \forall (i, j) \in \mathcal{E}$. However, in the pairwise potential of our MRF, we also include information from the color-shade descriptor by weighting the value of the distance between each pair of neighboring pixels. The main idea underlying this formulation is that pairs of pixels belonging to the same ridge should belong to the same surface and therefore should share similar labels: the cost of holding different names should be higher for neighboring pairs of pixels whose observed RGB values belong to the same ridge. Following this idea, we define the pairwise potential as

$$V(x_i, x_j) = \omega_{ij} \|x_i - x_j\|_1, \quad (6)$$

where (x_i, x_j) is a pair of neighboring pixels and ω_{ij} weights the classical smoothness term according to the relative position of the RGB values r_i and r_j of pixels i, j and the ridges of the color-shade descriptor as explained below.

Let $\pi(r_i)$ be the (orthogonal) projection of the pixel value r_i onto its associated ridge $SD(r_i)$ and let θ_{ij} be the angle formed by the lines $(r_i r_j)$ and $(\pi(r_i)\pi(r_j))$. We distinguish three cases of relative position between a pair of pixels' RGB values (r_i, r_j) and the set of ridges of an image (see Figure 4).

Case A. $\omega_{ij} = \alpha$ if the two pixels lie on two different ridges, *i.e.* $SD(r_i) \neq SD(r_j)$;

Case B. $\omega_{ij} = \beta$ if the two pixels lie on the same ridges, *i.e.* $SD(r_i) = SD(r_j)$, but the direction they determine is not parallel to the ridge, *i.e.* $\theta_{ij} > thr$, where thr is a parameter fixed once for all;

Case C. $\omega_{ij} = \gamma$ if the two pixels lie on the same ridges, *i.e.* $SD(r_i) = SD(r_j)$, and the direction they determine is parallel to the ridge, *i.e.* $\theta_{ij} \leq thr$.

The main idea underlying the choice of the parameters α, β, and γ is that the cost of holding different names should be higher for pairs of pixels whose observed RGB values belong to the same ridge. In particular, this cost should dramatically increase if, in addition, they form a segment
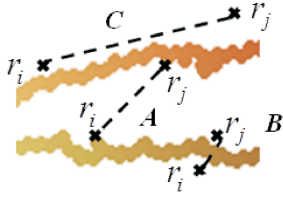
Figure 4: Schema of cases A, B, and C for the pairwise potential computation on ridges.

whose direction is collinear to the ridge's direction (case (C) in Figure 4) because this corresponds to the paradigmatic case of two pixels belonging to the same reflectance object but with different shadings. Accordingly, (α, β, γ) should verify the inequalities $\alpha \leq \beta \leq \gamma$ and $\alpha \ll \gamma$.

3.3. MRF output

The output of the MRF consists of an array of probability vectors. However, what we expect to recover are reflectance values (*i.e.* RGB triplets). Accordingly, we need a way to set a link between RGB and probability values.

Since we first discretize the probability vectors, many RGB values are mapped to a single vector by the color-name descriptor. This provides a partition $\coprod_{v \in \mathcal{L}} S_v$ of the RGB cube, where S_v is the set of RGB values associated to label v. These sets turn out to be convex. The map to RGB values is defined by associating each probability vector (i.e. label $v \in \mathcal{L}$) with the center of mass of the RGB convex region it defines.

4. Adding global scene coherence

Up to this point, we have only used local and semi-local color information for constructing reflectance estimates. The MRF yields a representative RGB value for each area of uniform reflectance. However, these semi-local representatives are likely to lack global consistency. In particular, the intensity of the estimated reflectance may not reflect that of the original image, which can cause undesired shading patterns. We addressed this problem by adding global coherence to the intensity of the reflectance values provided by the belief propagation algorithm. To this end, we modified the intensity of the RGB descriptor of each uniform reflectance area according to that of the original image.

Let $I = \bigcup_{i \in \mathcal{U}} U_i$ be the partition of an image into its areas of uniform reflectance provided by the MRF. Let L_i and L_i^{orig} be the intensities of the RGB triplet of the area U_i and of the same area in the original image, respectively. Ideally, to reflect the real shading, the ratio of intensities should verify, for each pair of areas in contact U_i and U_j,

$$L_i/L_j = L_i^{\text{orig}}/L_j^{\text{orig}}. \tag{7}$$

However, the connectivity between uniform reflectance areas is complex and no modification can make all the intensity ratios similar to that of the original image in general. In practice, we minimize the differences using mean square error (MSE). In this minimization problem, we wanted regions sharing a long frontier to have a higher weight. The length of the frontier between two regions (denoted l_{ij} for regions i and j) is defined to be the amount of pixels in both regions which have a neighboring pixel (assuming 4-neighborhood) belonging to the other region.

Thus, our purpose is to find a set of scalars $\{\lambda_i | i \in \mathcal{U}\}$ which modify our estimated reflectance intensities in order to enhance the global coherence of our recovered shading scene (explicitly, L_i is substituted by $\lambda_i L_i$). Mathematically, we can define a function W depending on such scalars as

$$W(\{\lambda_i\}_{i \in \mathcal{U}}) = \sum_{(i,j) \in \mathcal{R}^2, i<j} l_{i,j} \|\lambda_i L_i L_j^{\text{orig}} - \lambda_j L_j L_i^{\text{orig}}\|_2 \tag{8}$$

and find the set of values that minimize it:

$$\{\lambda_{\mathbf{i}}\}^* = \arg \min_{\{\lambda_{\mathbf{i}}\}} W(\{\lambda_{\mathbf{i}}\}). \tag{9}$$

This can be done by applying MSE and imposing a lower bound to the solution (otherwise we could obtain the trivial solution $\lambda_i = 0, \forall i \in \mathcal{R}$). Once we have such a set of parameters λ_i, we can multiply all the modified reflectance values by the same factor to obtain new values within the set $[0, 255]$. Figure 5 shows an example of how global coherence can improve the final recovered images.
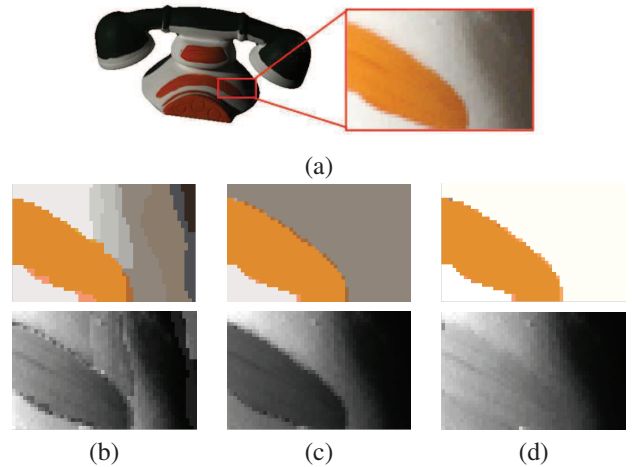

(a)

(b) (c) (d)

Figure 5: (a) Original Image. Different reflectance and shading estimates are shown: (b) Using only local color coherence. (c) Adding semi-local ridge observations. (d) Considering global scene coherence

5. Experiments

In this section we evaluate the performance of our approach. First, we discuss the existing error metrics for intrinsic image evaluation that have been proposed in previous works. Afterwards, we test our method on the MIT dataset [12], which has become the standard set to test intrinsic image algorithms. We quantitatively and qualitatively compare our results to the ones obtained by several previous approaches.

5.1. Error Metrics

Several error metrics have been proposed in previous works to evaluate intrinsic image algorithms.

One of the most used metrics in previous works is the local MSE (LMSE), which was proposed by Grosse *et al*. [12] as an appropriate measure for edge-based methods. They claimed that, for such methods, MSE is too restrictive because images with just a small misclassified edge can have a large MSE. However, Jiang *et al*. [13] argued that LMSE sometimes has a low value in images that are not qualitatively good. To overcome this problem they defined a new metric based on LMSE, the absolute LMSE (aLMSE), and also proposed to evaluate intrinsic images using the correlation measure, which computes the similarity, *i.e.* statistical dependency, between two images independently of their mean values.

Whereas correlation and MSE are global error measures, the others are variations of the global measures and are computed as an average of local error on small image windows.

We consider that LMSE is biased towards edge-based methods. Hence, in the next section, although we evaluate our method with four error metrics, we will focus our analysis on the results for global measures, *i.e.* MSE and correlation.

5.2. Results

We have estimated the intrinsic images of the full MIT dataset composed of 20 images whose reflectance and shading ground truth are available. For each image, the error has been obtained by averaging the results on the reflectance and shading estimates.

We compare our results to state-of-the-art approaches. Previous methods have been evaluated either on the full MIT dataset composed of 20 images or on a subset of 16 objects that we denote here as 'reduced MIT dataset'. In each case, we compare our results to those from the methods whose results, or the code to generate them, are available.

Results on the reduced MIT dataset have been compared to grey and color Retinex algorithms (obtained from [12]), the methods by Tappen *et al*. (Tap-05 [21] and Tap-06 [20]), the methods by Shen and Yao [19] (Shen-SR and Shen-SRC), and Weiss algorithm [23]. For the comparison on

the full MIT dataset we test grey and color Retinex algorithms, the methods by Jiang *et al*. [13] (Jiang-A, Jiang-H and Jiang-HA), and Weiss' algorithm.

The error metrics used to evaluate the results are the global measures MSE and correlation, and the local measures LMSE and aLMSE.

In our method, we set $\mu = 1/3$ to weight the two components of the energy function of our MRF and the dependence relations among parameters α, β, and γ defined in section 3.2 as follows: $\gamma/\alpha = 100$ and $\gamma/\beta = 2$. To initialize the network, we apply a logarithm to the input image before obtaining its color-shade descriptor.

Tables 1 and 2 show the results obtained by the evaluated methods on the reduced and full MIT datasets, respectively. As can be seen in the tables, our method obtains the best results on the global measures (*i.e.* MSE and correlation) on both the reduced and the full dataset when compared to single-image methods. In both cases, we overcome state-of-the-art results, obtaining even better results than Weiss' algorithm, which uses image sequences and, therefore, has more information than single-image based methods. Notice that for the three error metrics, the lower is the better, while for the correlation the opposite holds.

	Global measures		Local measures	
	Corr.	MSE	LMSE	aLMSE
Grey Retinex	0.6494	0.1205	0.0329	0.3373
Tap-05	—	—	0.0570	—
Tap-06	—	—	0.0390	—
Col. Retinex	0.7146	0.1108	0.0286	0.2541
Shen-SR	0.7259	0.1223	0.0242	0.2454
Shen-SRC	0.7733	0.0906	**0.0149**	**0.2147**
Ours	**0.7862**	**0.0834**	0.0340	0.2958
Weiss	0.7709	0.0900	0.0210	**0.1953**

Table 1: Results on the reduced MIT dataset (16 objects) with different error metrics. Shen-SRC results are computed on a subset of 13 objects ('deer', 'squirrel' and 'dinosaur' results were not available).

	Global measures		Local measures	
	Corr	MSE.	LMSE	aLMSE
Grey Retinex	0.6292	0.1169	0.0296	0.3789
Col. Retinex	0.7171	0.1072	**0.0257**	**0.2895**
Jiang-A	0.6262	—	0.0388	0.4036
Jiang-H	0.6179	—	0.0409	0.3655
Jiang-HA	0.6631	—	0.0460	0.3655
Ours	**0.7556**	**0.0836**	0.0305	0.3457
Weiss	**0.7619**	0.0890	**0.0191**	**0.2230**

Table 2: Results on the full MIT dataset (20 objects) with different error metrics.

As expected, local measures (LMSE and aLMSE) penalize our results and the performance of our methods considerably decreases when evaluated with these measures. However, as stated above, we consider that the evaluation of intrinsic image algorithms in terms of correlation or MSE is more accurate since these measures are more meaningful in terms of similarity to the ground truth.

In table 3, we present some qualitative results of our method on three objects of the MIT dataset. For comparative purposes, the objects shown are the ones used in [12]. These objects belong to each of the three subgroups of objects that the dataset contains, namely painted objects, printed papers, and animals. We compare our results to the ones from Color Retinex and Weiss algorithm, which are the best methods in the evaluation done in [12], and the results of the SRC method of Shen and Yao [19], which is the state-of-the-art best result.

As can be seen in the table, our method is the only one that completely avoids the cast shadow on the reflectance image of the raccoon. Although the final colors of surfaces in the reflectance are not well recovered in the turtle's reflectance image, our method forces a single reflectance value within the areas where the material color is uniform and all the shading effects due to the textured surface of the shell are correctly included in the shading image. Finally, on the tea bag most of the errors are found on the shading estimate, which includes some reflectance information. However, the reflectance image is quite well recovered.

6. Conclusions

In this work, we have described an approach for intrinsic image recovering based on a description of color using color names which is novel in the field. This sparse description of color was then combined with a color-shade attribute which enhances the stability of color names against strong changes in the illumination due to shading and highlights using a MRF.

Our results with the MIT dataset show that our method achieves state-of-the-art performance. This proves that color names, based on psychophysical data, provide a good basis for describing object reflectance.

In addition, the simple framework for inference we use in this work paves the way for more elaborated inference mechanisms like higher order potentials, which would allow us to include additional cues as texture, light source color, or multiple illuminants.

Acknowledgements

References

[1] S. Bagon. Matlab wrapper for graph cut, 2006.

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

[3] S. Beigpour and J. van de Weijer. Object recoloring based on intrinsic image estimation. In *IEEE International Conference on Computer Vision*, 2011.

[4] R. Benavente, M. Vanrell, and R. Baldrich. A data set for fuzzy colour naming. *Color Research and Application*, 31(1):48–56, 2006.

[5] R. Benavente, M. Vanrell, and R. Baldrich. Parametric fuzzy sets for automatic color naming. *Journal of the Optical Society of America A*, 25(10):2582–2593, 2008.

[6] B. Berlin and P. Kay. *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, 1969.

[7] A. Bousseau, S. Paris, and F. Durand. User assisted intrinsic images. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2009)*, 28(5):130:1–130:10, 2009.

[8] Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1222–1239, 2001.

[9] G. Finlayson, S. Hordley, C. Lu, and M. Drew. On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):59–68, 2006.

[10] B. Funt, M. Drew, and M. Brockington. Recovering shading from color images. In *European Conference on Computer Vision*, pages 124–132, 1992.

[11] L. D. Griffin. Optimality of the basic colour categories for classification. *Journal of the Royal Society Interface*, 3(6):71–85, 2006.

[12] 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*, pages 2335–2342, 2009.

[13] X. Jiang, A. J. Schofield, and J. L. Wyatt. Correlation-based intrinsic image extraction from a single image. In *European Conference on Computer Vision*, pages 58–71, 2010.

[14] E. Land and J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.

[15] Y. Matsushita, S. Lin, S. B. Kang, and H. Shum. Estimating intrisic images from image sequences with biased illumination. In *European Conference on Computer Vision*, pages 274–286, 2004.

[16] S. A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10(4):210–218, 1985.

[17] J. Shen, X. Yang, Y. Jia, and X. Li. Intrinsic images using optimization. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3481–3487, 2011.

[18] L. Shen, P. Tan, and S. Lin. Intrinsic image decomposition with non-local texture cues. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2008.

IMAGE			
GT			
COLOR RETINEX	Correlation=0.7533	Correlation=0.7873	Correlation=0.3325
SHEN SRC	Correlation=0.8625	Correlation=0.8454	Correlation=0.4558
OURS	Correlation=0.9687	Correlation=0.8835	Correlation=0.7020
WEISS	Correlation=0.5701	Correlation=0.9697	Correlation=0.7453

Table 3: Shading and reflectance images recovered by previous algorithms and our approach from images of the MIT database. Values below each decomposition are the corresponding correlation measures.

[19] L. Shen and C. Yeo. Intrinsic images decomposition using a local and global sparse representation of reflectance. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 697–704, 2011.

[20] M. F. Tappen, E. H. Adelson, and W. T. Freeman. Estimating intrinsic component images using non-linear regression. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1992–1999, 2006.

[21] 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):1459–

1472, 2005.

[22] E. Vazquez, R. Baldrich, J. Van de Weijer, and M. Vanrell. Describing reflectances for colour segmentation robust to shadows, highlights and textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):917–930, 2011.

[23] Y. Weiss. Deriving intrinsic images from image sequences. In *International Conference on Computer Vision*, pages 68–75, 2001.