

Illuminant Chromaticity from Image Sequences

Véronique Prinet Dani Lischinski Michael Werman
The Hebrew University of Jerusalem, Israel

Abstract

We estimate illuminant chromaticity from temporal sequences, for scenes illuminated by either one or two dominant illuminants. While there are many methods for illuminant estimation from a single image, few works so far have focused on videos, and even fewer on multiple light sources. Our aim is to leverage information provided by the temporal acquisition, where either the objects or the camera or the light source are/is in motion in order to estimate illuminant color without the need for user interaction or using strong assumptions and heuristics. We introduce a simple physically-based formulation based on the assumption that the incident light chromaticity is constant over a short space-time domain. We show that a deterministic approach is not sufficient for accurate and robust estimation: however, a probabilistic formulation makes it possible to implicitly integrate away hidden factors that have been ignored by the physical model. Experimental results are reported on a dataset of natural video sequences and on the GrayBall benchmark, indicating that we compare favorably with the state-of-the-art.

1. Introduction

Although a human observer is typically able to discount the color of the incident illumination when interpreting colors of objects in the scene (a phenomenon known as *color constancy*), the same surface may appear very different in images captured under illuminants with different colors.

Estimating the colors of illuminants in a scene is thus an important task in computer vision and computational photography, making it possible to white-balance an image or a video sequence, or to apply post-exposure relighting effects. However, most existing *color constancy* and *white balance* methods assume that the illumination in the scene is dominated by a single illuminant color.

In practice, a scene is often illuminated by two different illuminants. For example, in an outdoor scene the illuminant color in the sunlit areas differs significantly from the the illuminant color in the shade, a difference that becomes more apparent towards sunset. Similarly, indoor scenes of-

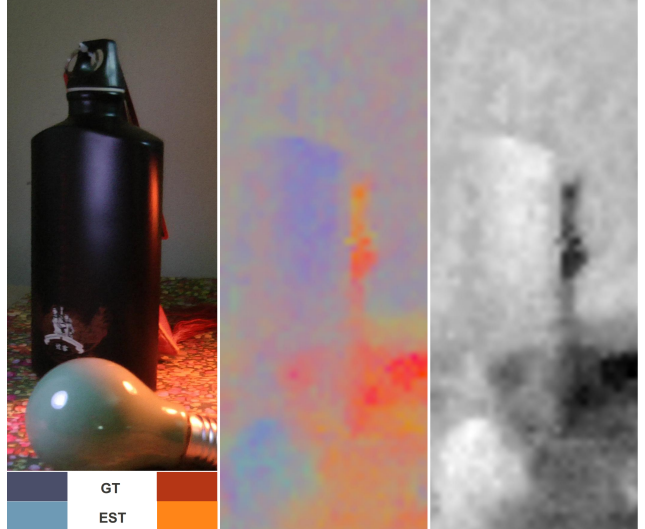


Figure 1. Left: First frame of a sequence recorded under two light sources and corresponding illuminant colors (ESTimated and Ground Truth). Middle: Locally estimated incident light color $\{\mathbf{l}^s\}$. Right: Light mixture coefficients $\{\alpha^s\}$ (see Section 4).

ten feature a mixture of artificial and natural light. Hsu *et al.* [7] propose a method for recovering the linear mixture coefficients at each pixel of an image, but rely on the user to provide their method with the colors of the two illuminants.

By using multiple images, we can formulate the problem of illuminant estimation in a well-constrained form, thus avoiding the need of any prior or additional information provided by a user, as most previous work do.

The main contribution of this work is two-fold:

1. We introduce a new physically-based approach to estimate illuminant chromaticity from a temporal sequence; we show experimentally that the distribution of the incident light at edge-points, where specularities may be often encountered, can be modeled by a Laplace distribution; this enables one to estimate the global illuminant color robustly and accurately using the MAP estimation framework.
2. We show that our approach can be extended to scenes

lit by a spatially varying mixture of two different illuminants. Hence, we propose an efficient framework for estimating the chromaticity vectors of both illuminants, as well as recovering their relative mixture across the scene.

Our framework can be applied to natural images sequences, indoor or outdoor, as long as specularities are present in the scene. We validate our illuminant estimation approach on existing as well as new datasets and demonstrate the ability to white-balance and relight such images.

The rest of the paper is organized as follows: after a short review of the state-of-the-art in color constancy, we describe a new method for estimating the illuminant color from a natural video sequence, as long as some surfaces in the scene have a specular reflectance component (Section 3). We then extend this method for the completely automatic estimation of two illuminant colors from a sequence, along with the corresponding mixture coefficients, without requiring any additional input (Section 4). We present results in Section 5, including a comparison with the state-of-the-art that shows our approach is competitive.

2. Related work

Color constancy has been extensively studied and we refer the reader to a recent survey [5] for a relatively complete review of the state-of-the-art in this area. In this section we briefly review the most relevant methods to our work, namely those based on the dichromatic model [15, 17, 20], and methods concerned with multi-illuminant estimation [7, 6]. Note that *none* of these approaches focus on video or temporal sequences: to our knowledge, the only work dealing with illuminant estimation in videos is based on averaging results from existing frame-wise methods [14, 19].

Physically-based modeling. Shafer [15] introduced the physically-based dichromatic reflection model, which decomposes the observed radiance into diffuse and specular components. This model has been used by a number of researchers to estimate illuminant color [10, 9, 3]. More recently, Tan *et al.* [17] defined an inverse intensity-chromaticity space, exploiting the different characteristics of diffuse and specular pixels in this space to estimate the illuminant’s RGB components. Yang *et al.* [20] operate on a pair of images, simultaneously estimating illuminant chromaticity, correspondences between pairs of point, and specularities. While all of these approaches are based on the physics of reflection, most of them encountered limited success outside laboratory experiments (*i.e.*, on complex images in uncontrolled environment and lighting).

The closest work to our single illuminant estimation method (described in Section 3) is that of Yang *et al.* [20],

which is based on a heuristic that votes for discretized values of the illuminant chromaticity Γ . In contrast to [20], starting from the same equations, we formulate the problem of illuminant estimation in a probabilistic manner to implicitly integrate hidden factors that have been ignored by the underlying physical model. This results in a different, simple yet robust approach, making it possible to reliably estimate the global illuminant chromaticity from natural image sequences acquired under uncontrolled settings.

Multi-illuminant estimation. The need for multi-illuminant estimation arises when different regions of a scene captured by a camera are illuminated by different light sources [16, 7, 2, 8, 6]. Among these, the most closely related to our work is the one of Gisenji *et al.* [6], who propose estimating the incident light chromaticity locally in patches around points of interest, before estimating the two global color illuminants. This method is mainly practical but not theoretically justified. Conversely, our approach, which is also based on a local-to-global framework, is mathematically justified and based on inverting the compositing equation of the illuminant colors. Also related is the work of Hsu *et al.* [7], who address a different but related problem: performing white balance in scenes lit by two differently colored illuminants. However, this work operates on a single image assuming that the two global illuminant chromaticities are known and focuses on recovering their mixture across the image. In contrast, the method we describe in Section 4 operates on a temporal sequence of images, automatically recovering both the illuminant chromaticities and the illuminant mixtures.

3. Single illuminant chromaticity estimation

3.1. The dichromatic reflection model

The dichromatic model for dielectric materials (such as plastic, acrylics, hair, etc.) expresses the light reflected from an object as a linear combination of diffuse (body) and specular (interface) components [15]. The diffuse component has the same radiance when viewed from any angle, following Lambert’s law, while the specular component captures the directional reflection of the incident light hitting the object’s surface. After tristimulus integration, the color $\mathbf{I}$ at each pixel $\mathbf{p}$ may be expressed as:

$$\mathbf{I}(\mathbf{p}) = \mathbf{D}(\mathbf{p}) + m(\mathbf{p}) \mathbf{L}, \quad (1)$$

where $\mathbf{D} = (D_r, D_g, D_b)$ is the diffusely reflected component and $\mathbf{L} = (L_r, L_g, L_b)$ denotes the global illuminant color vector, multiplied by a scalar weight function $m(\mathbf{p})$, which depends on the spatial position and local geometry of the scene point visible at pixel $\mathbf{p}$. The specular component $m(\mathbf{p}) \mathbf{L}$ has the same spectral distribution as the incident light [15, 10, 3]. In this section we assume that the spectral

distribution of the illuminant is identical everywhere in the scene, making $\mathbf{L}$ independent of the spatial location $\mathbf{p}$.

3.2. Illuminant chromaticity from a sequence

Extending the model in eq. (1) to a temporal sequence of images, assuming that the illuminant color $\mathbf{L}$ does not change with time, gives:

$$\mathbf{I}(\mathbf{p}, t) = \mathbf{D}(\mathbf{p}, t) + m(\mathbf{p}, t) \mathbf{L}. \quad (2)$$

Consider a 3D point $\mathcal{P}$ projected to $\mathbf{p}$ at time t , and to $\mathbf{p} + \Delta\mathbf{p}$ at time $t + \Delta t$. If the incident illumination at $\mathcal{P}$ has not changed, the diffuse component reflected at that point also remains the same:

$$\mathbf{D}(\mathbf{p}, t) = \mathbf{D}(\mathbf{p} + \Delta\mathbf{p}, t + \Delta t). \quad (3)$$

Thus, the illuminant color $\mathbf{L} = (L_r, L_g, L_b)$ can be derived from equations (2) and (3). For each component $c \in \{r, g, b\}$, we have:

$$\begin{aligned} I_c(\mathbf{p} + \Delta\mathbf{p}, t + \Delta t) - I_c(\mathbf{p}, t) \\ = (m(\mathbf{p} + \Delta\mathbf{p}, t + \Delta t) - m(\mathbf{p}, t)) L_c, \end{aligned} \quad (4)$$

since the diffuse component in the right-hand side cancels out due to property (3). Denoting the left hand side of eq. (4) by $\Delta I_c(\mathbf{p}, t)$, and normalizing both sides of the equation, we obtain (whenever $\|\Delta\mathbf{I}(\mathbf{p}, t)\| \neq 0$):

$$\frac{\Delta I_c(\mathbf{p}, t)}{\|\Delta\mathbf{I}(\mathbf{p}, t)\|_1} = \frac{L_c}{\|\mathbf{L}\|_1} = \Gamma_c, \quad (5)$$

where $\|\mathbf{Y}\|_1 = \sum_{c \in \{r, g, b\}} Y_c$.¹

Hence $\mathbf{\Gamma} = (\Gamma_r, \Gamma_g, \Gamma_b)$ is the global incident light chromaticity vector, simply obtained by differentiating (and normalizing) the RGB irradiance components of any point $\mathbf{p}$ with a specular component, tracked between two consecutive frames t and $t + \Delta t$. Note that this formulation assumes that the displacement $\Delta\mathbf{p}$ is known: $\Delta\mathbf{I}(\mathbf{p}, t)$ is the difference between image $\mathbf{I}$ at $t + \Delta t$ and the wrapped image at t .

So far we *implicitly* assumed that: (1) a change in the specular reflection occurs at $\mathbf{p}$ from time t to time $t + \Delta t$; and (2) the displacement $\Delta\mathbf{p}$ is estimated accurately. These two factors suggest that reliable sets of points to evaluate eq. (5) accurately are edge-points extracted from each frame. The rationale behind this is that flow/displacement estimation is usually robust at edges (because edges are discontinuities that are preserved/invariant over time, unless occlusion or shadows appear). More importantly, edge points delineate local discontinuities or object's surface boundaries (with large local curvature) and specularities are likely to be observed at these points. The counter argument

¹By abuse of notation we use $\|\cdot\|$, even though it can take negative values, and thus is not a norm, strictly speaking.

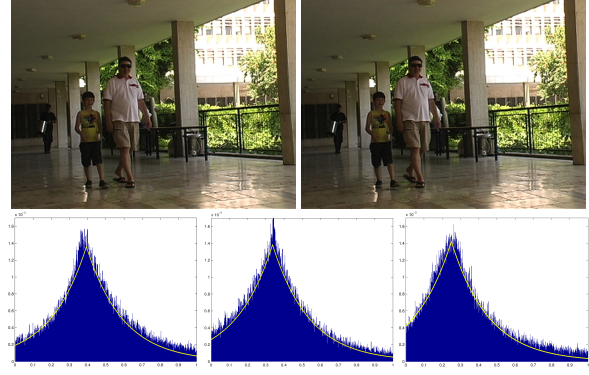


Figure 2. Top: two successive frames of a video sequence; Bottom: empirical distributions $P(x_c)$ for the Red, Green and Blue channels and approximations by a Laplace distribution (yellow curve).

to this choice might be that pixel values at edges often contain a mixture of light coming from two different objects; experimentally, we found that this is not a limiting factor. In Section 5, we experimentally compare a number of different point choice strategies and their impact on the results.

3.3. Robust probabilistic estimation

Yang *et al.* [20] already proposed estimating illuminant chromaticity from a pair of images using eq. (5), demonstrating their method on certain, suitably constrained image pairs. In this section, we propose an alternative probabilistic estimation approach, which is simpler, yet robust enough to reliably estimate $\mathbf{\Gamma}$ from natural image sequences.

Equation (5) is based on Shafer's physically-based dichromatic model [15, 10]. It does not, however, take into account several factors which also might affect the observed scene irradiance in a noticeable way: (i) the effect of the incident light *direction* is neglected; (ii) *local inter-reflections* are not taken into account; they can, however, account for a significant amount of light incident to a given object [13]; as a result, the assumption of a single and uniform global illuminant might not be completely valid everywhere; (iii) the statistical nature of the image capture process (*e.g.*, camera noise) is ignored.

We therefore cast the problem of illuminant chromaticity recovery in a probabilistic framework, where $\mathbf{\Gamma}$ is obtained using Maximum-a-Posteriori estimation:

$$\hat{\mathbf{\Gamma}} = \arg \max_{\mathbf{\Gamma}} P(\mathbf{\Gamma}|\mathbf{x}) \quad (6)$$

where $\mathbf{x} = \{(x_r(\mathbf{p}), x_g(\mathbf{p}), x_b(\mathbf{p}))\}$ is an observation vector consisting of all the pixels $\mathbf{p}$ of a temporal image sequence. Applying Bayes' rule, we express:

$$P(\mathbf{\Gamma}|\mathbf{x}) \propto P(\mathbf{x}|\mathbf{\Gamma})P(\mathbf{\Gamma}), \quad (7)$$

and reasonably assuming that all illuminants are equiprobable ($P(\Gamma) = \text{const}$), we rewrite the right-hand side:

$$P(\Gamma|\mathbf{x}) \propto P(\mathbf{x}|\Gamma) \quad (8)$$

$$\propto \prod_{c \in \{r, g, b\}} P(x_c|\Gamma) = \prod_{c \in \{r, g, b\}} P(x_c|\Gamma_c).$$

Above we made the additional assumption that the observed channels x_c are mutually independent, and depend only on the corresponding illuminant channel Γ_c .

More specifically, we define the observed features as $x_c(\mathbf{p}) = \frac{\Delta I_c(\mathbf{p}, t)}{\|\Delta \mathbf{I}(\mathbf{p}, t)\|}$ (the left-hand side of eq. (5)), where the image points $\mathbf{p}$ are a set of edge points extracted from the image sequence. We estimate the likelihood $P(x_c|\Gamma_c)$ from its empirical distribution: we discretize x_c in n bins ranging from ϵ to 1 (*i.e.*, the set of values that the chromaticities can take), and compute the histogram of x_c . Figure 2 (bottom) illustrates the empirical distributions for the three channels x_c computed from the video frames (top).

We experimented with estimating Eq. (6) in two different ways: (i) in a purely empirical fashion, by setting $\hat{\Gamma}_c$ to the histogram mode for each channel $\mathbf{x}_c$; (ii) by observing experimentally that the histograms follow a multivariate Laplace distribution, whose maximum likelihood estimator is the median of the samples, we set $\hat{\Gamma}_c$ to the median value of $\mathbf{x}_c$, for each channel c independently. Finally, the estimated chromaticity vector is normalized so that $\sum_c \hat{\Gamma}_c = 1$. The latter approach proved to be more robust in practice.

4. Two light sources

Until now we assumed a single illuminant whose color is constant across the image. In this section we extend our approach to the common scenario where the illuminant color at each point may be modeled as a spatially varying mixture of two dominant chromaticities. Examples include: illumination by a mixture of sunlight and skylight, or a mixture of artificial illumination and natural daylight.

Our approach is partly motivated by the work of Hsu *et al.* [7] who proposed a method for recovering the mixture coefficients from a single image, when the two global illuminant chromaticities are *known*. In contrast to their approach, we use a temporal sequence (with as few as 2-3 images) but recover both the two chromaticities and their spatially varying mixture.

We assume that the mixture is constant across small space-time patches, and consequently the combined illuminant chromaticity is also constant across each patch. We further restrict ourselves to cases where the change in the view/acquisition angle between the first and the last frame is kept relatively small.

We begin by independently estimating the combined illuminant chromaticity over a set of small overlapping patches,

using the method described in the previous section separately for each patch. Since some of the patches might not contain enough edge points with specularities, making it impossible to obtain an estimate of the illuminant there, we use linear interpolation from neighboring patches to fill such holes. We then use the resulting combined illuminant chromaticity map to estimate the spatially varying mixture coefficients and the two illuminant chromaticities, as described in the remainder of this section.

4.1. Problem statement and solution

Assuming the chromaticities of the two global illuminants in the scene are given by the (unknown) normalized vectors Γ_1 and Γ_2 , we replace the incident global illumination vector $\mathbf{L}$ in eq. (2) at point $(\mathbf{p}, t)$ with a spatially varying one:

$$\mathbf{L}(\mathbf{p}, t) = k_1(\mathbf{p}, t) \Gamma_1 + k_2(\mathbf{p}, t) \Gamma_2, \quad (9)$$

where k_1 and k_2 are the non-negative intensity coefficients of Γ_1 and Γ_2 . Assuming that the incident light $\mathbf{L}(\mathbf{p}, t)$ is roughly constant across small space-time patches, we write:

$$\mathbf{L}^s = k_1^s \Gamma_1 + k_2^s \Gamma_2 \quad (10)$$

for each small space-time patch s . Normalizing both sides and making use of the fact that Γ_1 and Γ_2 are normalized, we express the local combined incident light chromaticity as a convex combination:

$$\begin{aligned} \Gamma_c^s &= \frac{L_c^s}{\|\mathbf{L}^s\|_1} = \frac{k_1^s \Gamma_{1,c} + k_2^s \Gamma_{2,c}}{\|k_1^s \Gamma_1 + k_2^s \Gamma_2\|_1} \\ &= \alpha^s \Gamma_{1,c} + (1 - \alpha^s) \Gamma_{2,c}, \end{aligned} \quad (11)$$

where $\alpha^s = \frac{k_1^s}{k_1^s + k_2^s}$, for $c \in \{r, g, b\}$. This equation resembles the compositing equation in natural image matting; a similar observation was made by Hsu *et al.* [7]. However, unlike natural image matting where the composited colors as well as α vary across the image (underconstrained problem), in our case the composited vectors Γ_1 and Γ_2 are assumed constant. This enables a more direct solution once the left-hand side (Γ^s) has been estimated.

Manipulating eq. (11) we derive a linear relationship between α^s and each channel of Γ^s :

$$\begin{aligned} \Gamma_c^s &= \alpha^s (\Gamma_{1,c} - \Gamma_{2,c}) + \Gamma_{2,c} \\ \alpha^s &= \frac{\Gamma_c^s - \Gamma_{2,c}}{\Gamma_{1,c} - \Gamma_{2,c}} \\ \alpha^s &= a_c \Gamma_c^s - b_c \end{aligned} \quad (12)$$

where $a_c = \frac{1}{\Gamma_{1,c} - \Gamma_{2,c}}$ and $b_c = \frac{\Gamma_{2,c}}{\Gamma_{1,c} - \Gamma_{2,c}}$ when $\Gamma_{1,c} \neq \Gamma_{2,c}$ and $\mathbf{a} = \{a_c\}$. To recover the mixture coefficients α^s we minimize the following quadratic cost function:

$$\sum_{s,c} (\alpha^s - a_c \Gamma_c^s + b_c)^2 + \epsilon \|\mathbf{a}\|^2 \quad (13)$$

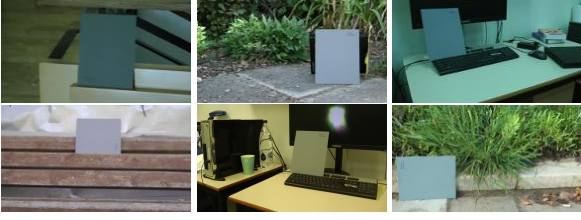


Figure 3. Video dataset recorded under normal lighting conditions using a single illuminant: the first frames of six of the sequences.

by solving for the smallest eigenvector of the associated symmetric homogeneous linear system [11]. The vector of α^s values is then obtained by shifting and scaling the resulting eigenvector’s entries to $[0, 1]$ (assuming that each of the illuminants is exclusive in at least one patch in the image).

Having obtained the mixing coefficients α^s , we recover Γ_1 and Γ_2 by solving equation (11) using least squares minimization.

5. Experimental evaluation

5.1. Implementation details

Our method is implemented in Matlab (code is available online). We used some off-the-shelf functions with the following settings:

- Illuminant chromaticity estimation is performed in linear RGB, assuming gamma of 2.2.
- Edge detection is performed using the standard Canny operator in Matlab with the default threshold of 0. For the estimation we only use edge points $\mathbf{p}$ for which $|\sum_c \Delta I_c(\mathbf{p}, t)| > T$.
- Point correspondences between frames are computed using Liu *et al.*’s SIFTFlow algorithm [12].
- Empirical distributions $P(x_c|\Gamma_c)$, are quantized to 2000 bins for single illuminant estimation, and to 500 bins for two illuminants, in the range $[0.001, 1]$. Note that the quantization imposes an upper bound on the estimation accuracy (on the order of 10^{-4} per channel). Finer quantization leads to overfitting, while coarser reduces the accuracy.
- We use 100×100 pixel tiles for two-illuminant estimation. The tiles are overlapping, with a spacing of 10 pixels. Note that this defines a sub-sampling of the original space/time domain.

5.2. Datasets and experimental settings

We evaluate the performance of single illuminant estimation on two datasets: a newly created dataset of 13 video sequences and the GrayBall database [1]. To validate the two-illuminant estimation approach, we recorded three video sequences of scenes lit by two light sources.

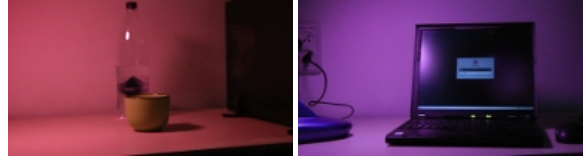


Figure 4. Video dataset simulating extreme lighting conditions: reddish $\Gamma_R = (0.54695, 0.1779, 0.27515)$, and bluish $\Gamma_B = (0.35132, 0.12528, 0.52339)$. Shown are the first frames from two sequences (out of four).

The single-illuminant dataset we created consists of video sequences captured with a high definition video camera (Panasonic HD-TM 700), at 60 fps and 1920×1080 pixels per frame. The set includes three outdoor scenes and six indoor scenes. The videos were recorded with a moving camera. A flat grey card with spectrally uniform reflectance was placed in each scene, appearing in each video sequence for a few seconds. We supplemented this dataset with two additional publicly available sequences². The remaining four sequences of this set were taken using red or blue filters (Fig. (4)), in order to simulate “extreme” lighting conditions.

The ground truth illuminant was estimated, for each sequence individually, using the grey card. We extracted pixels on the grey card over 5 consecutive frames, and computed their normalized average RGB value. For each sequence, we also computed the variance σ_c and mean angular variation β of the grey card RGB vectors to ensure that the scene complies with a constant illumination assumption ($0.1^\circ < \beta \leq 0.5^\circ$ and $1.e - 7 < \sigma^2 \leq 1.e - 5$).

As for the two-illuminant dataset, it consists of videos acquired under complex lighting conditions (Fig. 5): two incandescent lamps (blue and red), sun and skylight, incandescent lamp and natural daylight. Two grey cards were placed in the scene during acquisition, ensuring that each grey card is illuminated by only one of the illuminants. The ground truth values were computed as explained earlier.

We also used the GrayBall database of Ciurea and Funt [1]. This dataset is composed of frames extracted from several video clips taken at different locations. The temporal ordering of the frames had been preserved, resulting in a time lapse of about 1/3 second between consecutive images [19]. The entire database contains over 11,000 images, of both indoor and outdoor scenes. A small grey sphere was mounted onto the video camera, appearing in all the images, and used to estimate a per-frame ground truth illuminant chromaticity. This ground truth is given in the camera reference system (*i.e.* RGB domain) [1]. Note that, in the GrayBall database, the illuminant varies slightly from frame to frame and therefore violates our assumption of uniform illu-

²<http://www.cs.huji.ac.il/labs/cglab/projects/tonestab/>

	<i>Edges</i>	<i>Near edges</i>	<i>Entire image</i>
Laplace	5.389	5.429	5.450
Gaussian	6.462	6.486	6.487

Table 1. Comparison of different strategies for point selection (columns) and between Laplace and Gaussian distribution modeling (rows) (see Section 3.3), $T = 10^{-1}$. The reported angular error (in degrees) is averaged over the nine video sequences recorded with normal lighting conditions.

	<i>Average</i>	<i>Best 1/3</i>	<i>Worst 1/3</i>
GE-1 [18]	6.572	2.1787	11.271
GE-2 [18]	7.150	2.958	11.723
GGM [4]	7.013	6.208	9.166
IIC [17]	8.303	3.984	12.540
Our approach	5.389	2.402	8.784

Table 2. Angular errors (in degrees) for video sequences recorded under normal lighting conditions.

mination over time. We use this dataset because it is, to our knowledge, the only publicly available temporal sequence data for which both ground truth and results of previously published methods are available.

Results are reported in terms of the angular deviation β between the ground truth Γ^g and the estimated illuminant $\hat{\Gamma}$, in camera sensor basis: $\beta = \arccos(\frac{\hat{\Gamma} \cdot \Gamma^g}{\|\hat{\Gamma}\| \|\Gamma^g\|})$.

5.3. Single illuminant estimation

We begin with an experimental validation of the claims made in Section 3.2 regarding the choice to use edge points for illuminant estimation and the use of the Laplace distribution to model $P(x_c|\Gamma_c)$. Table 1 compares between three different strategies for choosing the specular points: choosing from points detected by the Canny edge detector, choosing from points next to edges, and choosing from the entire image. Note that we do not attempt here to compare between different edge detectors, but only to validate that edges are a good source of points for our estimator. We also compare between using the Laplace model and a Gaussian model (*i.e.* using the mean of $\mathbf{x}$, instead of the median, as the estimated illuminant). As can be seen from the table, smaller errors are obtained when using edge points and the Laplace model.

Video dataset. Tables 2 and 3 report illuminant estimation accuracy for the sequences recorded under normal illumination conditions (Fig. 3) and under extreme lighting (Fig. 4). We used a temporal window of 3 frames for the former, of 5 for the latter (to account for the noise in data acquisition due to the relatively dark environment), with a time step between consecutive frames of 3ms for both (except for the two downloaded sequences² for which we set a time step of 1ms). To estimate the illuminant we exclude

	<i>Reddish</i>	<i>Bluish</i>
GE-1 [18]	8.907	13.052
GE-2 [18]	10.246	13.657
GGM [4]	15.544	25.505
IIC [17]	—	19.675
Our approach	7.708	6.236

Table 3. Average angular errors (in degrees) for video sequences recorded with red and blue filters, with $T = 10^{-1}$.

the region of the frames that contains the grey card.

We compare our approach to several state-of-the-art methods: the Grey-Edge algorithm [18], Generalized Gamut Mapping (GGM) [4], and Inverse Intensity Chromaticity method (IIC) [17]. For Grey-Edge, we use first order and second order derivatives (GE-1 and GE-2, respectively), with L_1 norm and a Gaussian filter $\sigma = 1$ [18]. For GGM we use the intersection 1-jet criteria (*i.e.* based on first order derivatives), because it was reported to give the best results on several databases [4]. The IIC method was chosen because it is a popular reference among color constancy approaches based on a physical model. We used the authors’ implementation of these algorithms. All these approaches estimate a per-frame illuminant; we average the illuminant chromaticity vector computed for each frame, and report the angular error between the mean chromaticity vector and the ground truth [19]. We report the overall mean angular error, as well as the average angular errors over the best and the worst thirds of the results of each method.

Tables 2 and 3 show that IIC performs poorly on this dataset. This can be attributed to the fact that images in uncontrolled environments contain a large amount of saturated pixels or noise, factors which are ignored by purely deterministic models. GGM does not perform well in extreme light conditions, because the very limited range of color visible in the input frames does not enable a good matching with the prior gamut used by this algorithm. On the other hand, GGM, GE-1, and GE-2, all give reasonably good results under normal lighting conditions (6.5°, 7.1°, 7.0°). Note the large variance between the best and worst thirds for the GE methods, indicating a relatively unstable behavior. Our approach outperforms all of these methods on average for both normal and extreme lighting (5.3° and 6.9°), and exhibits stable performance. Note that the advantage of our approach can be attributed to the fact that it uses the temporal sequence, while other methods reported here work on each individual frame separately.

GrayBall dataset. In Table 4, we compare the performance of our approach on the GrayBall database to two state-of-the-art methods [4, 19], as well as to the classical GrayWorld method for reference. Reported values for these methods are taken from the original papers. For our method, we used pairs of consecutive frames and $T = 0.1$. We did

	<i>GrayWorld</i>	<i>GGM</i> [4]	<i>GE-2</i> [19]	<i>Ours</i>
Mean	7.9	6.9	5.4	5.4
Median	7.0	5.8	4.1	4.6

Table 4. Angular errors (in degrees) for images from the GrayBall database.

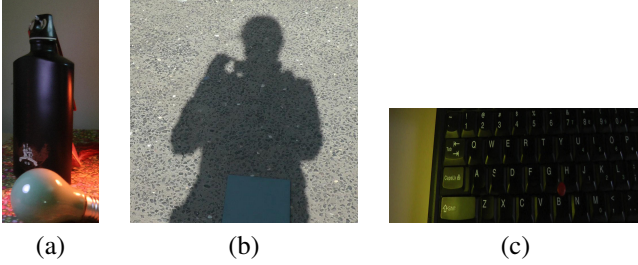


Figure 5. First frames of three sequences captured with two lights. (a) Two incandescent lamps (Γ_1 red, Γ_2 blue). (b) Outdoor scene lit by sunlight (Γ_1) and skylight (Γ_2). (c) Incandescent lamp (Γ_1 green) and natural daylight (Γ_2).

not attempt to apply IIC to this dataset, since it seems irrelevant to use this method for natural images, which contain saturated pixels and are acquired under uncontrolled lighting conditions. We refer the reader to [5] for an extended comparison of methods on this dataset, among which we include here only the best ones.

On this dataset the results obtained by Wang *et al.* [19] are equivalent to ours in term of average error (5.4°). Wang’s method uses several parameters (three to five threshold values), which have been tuned specifically on the GrayBall database (no results on other datasets are provided); we believe that the accuracy reported by the authors [19] is in part due to this parameter tuning.

5.4. Two illuminant estimation

Figure 1 shows the estimated incident light color map $\{\Gamma^s\}_{s=1}^S$, the light mixture coefficients $\{\alpha^s\}_{s=1}^S$, and the estimated light chromaticity, computed from sequence (a). During recording, the scene was illuminated by a red light from the back on the right side and a blue light from the front on the left side. The incident color map (middle) clearly captures the pattern of these two dominant light sources. The mixture coefficients map (right) indicates the relative contribution of one illuminant with respect to the other, interpolated across the image.

Table 5 reports quantitative results obtained from sequences recorded with two lights sources (Figure 5). We compare with the state-of-the-art, namely local GE-1, local GE-2, and local GrayWorld (GW) (see [6] for details). We apply local GE-1 and GE-2 using L_1 norm and Gaussian $\sigma = 2$. Results were computed using 3–5 frames from each sequence, with a time step of 2–4ms between frames.

	<i>Ours</i>		<i>Local GE</i> [6]		<i>Local GW</i>	
	Γ_1	Γ_2	Γ_1	Γ_2	Γ_1	Γ_2
Seq. (a)	9.65	5.14	31.69	4.8	12.94	10.49
Seq. (b)	5.74	4.76	9.69	9.82	5.89	8.81
Seq. (c)	7.35	6.49	17.9	5.65	7.63	5.74

Table 5. Two illuminant estimation. Angular errors (in degrees) for the estimation of Γ_1, Γ_2 on the sequences shown in Figure 5.



Figure 6. First frames of three video sequences (top) and estimated illuminant colors (bottom).

Patch/tiles size is set to 50×50 pixels in sequences (a) and (c) of our dataset, to 100×100 otherwise. From [6], we report the best result among local GE-1 and local GE-2, averaged over frames of the sequence. Overall, our method provides more accurate estimates than those obtained with the other methods. We have observed that both GE and GW tend to produce estimates biased towards gray. This makes the estimation of strongly colored illuminants (*e.g.*, the red light Γ_1 in sequence (a)) difficult.

Figure 6 shows additional results obtained on sequences for which the ground truth was not available. Motion between frames originates from camera displacement (right), object/human motion (middle), or light source motion (left). Color patches in the bottom row show the two estimated illuminant colors for each sequence. Note that the dominant tone is correctly retrieved (blue or gray mixed with yellow-orange in these three cases).

5.5. Application to white balance correction

The aim of white balance correction is to remove the color cast introduced by a non-white illuminant: *i.e.*, to generate a new image/sequence that renders the scene as if it had been captured under a white illuminant. Figure 7 demonstrates the result of applying white balance to a scene illuminated by a mixture of (warmer colored) late afternoon sunlight and (cooler colored) skylight. (Additional results, including scenes with a single illuminant, are provided in the supplementary material). Having estimated the incident light color Γ^s across the image, we simply perform the white balance separately at each pixel, instead of globally for the entire image, producing the result shown in Fig-

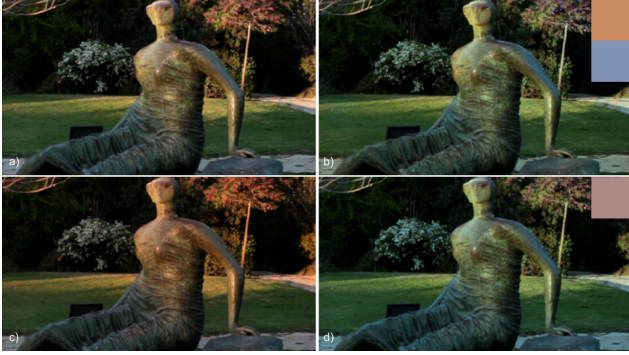


Figure 7. White balance with two illuminants. (a) Input frame of a scene illuminated by afternoon sunlight and skylight. (b) Result of spatially variant white balance correction after two illuminant estimation. (c) “Relighting” by changing the chromaticity of one of the illuminants. (d) For comparison: uniform white balance correction using a single estimated illuminant.

ure 7(b). A global white balance correction (using a single estimated illuminant) is shown in Figure 7(d) for comparison, suffering from a stronger remaining greenish color cast. The availability of the mixture coefficient map makes it possible to simulate changes in the colors of one or both illuminants. This is demonstrated in Figure 7(c), where the illuminant corresponding to the sunlight was changed to a more reddish color.

6. Conclusion

The ease with which one can acquire temporal sequences using commercial cameras and the ubiquity of videos on the web, makes natural the exploitation of temporal information for various image processing tasks. In this work, we presented an effective way to leverage temporal dependencies between frames to tackle the problem of illuminant estimation from a video sequence. By using multiple images, we can formulate the problem of illuminant(s) estimation as a well constrained problem, thus avoiding the need of prior knowledge or additional information provided by a user. Our physically-based model, embedded in a probabilistic framework (via MAP estimation), applies to natural images of indoor or outdoor scenes. Our approach is simply extended to scenes lit by two global illuminants, whenever the incident light chromaticity at each point of the scene can be modeled by a mixture of the two illuminant colors. We show that on several datasets, our results in general are comparable or improve upon the state-of-the-art both for single illuminant estimation and for two-illuminant estimation.

Acknowledgments. We thank the anonymous reviewers for their comments. This work was supported in part by the Israel Science Foundation and the Intel Collaborative Research Institute for Computational Intelligence (ICRI-CI).

References

- [1] F. Ciurea and B. Funt. A large image database for color constancy research. In *Color Imaging Conf.*, 2003. 5
- [2] M. Ebner. Color constancy based on local space average color. *Machine Vision and Applications*, 2009. 2
- [3] G. D. Finlayson and G. Shaefer. Solving for colour constancy using a constrained dichromatic reflection model. *IJCV*, 2001. 2
- [4] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalised gamut mapping using image derivative structures for color constancy. *IJCV*, 2010. 6, 7
- [5] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy: Survey and experiments. *TIP*, 2011. 2, 7
- [6] A. Gijsenij, R. Lu, and T. Gevers. Color constancy for multiple light sources. *IEEE TIP*, 2012. 2, 7
- [7] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand. Light mixture estimation for spatially varying white balance. *ACM Trans. Graph.*, 2008. 1, 2, 4
- [8] Y. Imai, Y. Kato, H. Kadoi, T. Horiuchi, and S. Tominaga. Estimation of multiple illuminants based on specular highlights detection. In *Int. Conf. on Comput. Color Imaging*, 2011. 2
- [9] G. Klinker, S. Shafer, and T. Kanade. The measurement of highlights in color images. *IJCV*, 1988. 2
- [10] H.-C. Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *J. Opt. Soc. Am. A*, 3(10):1694–1699, October 1986. 2, 3
- [11] A. Levin, D. Lischinski, and Y. Weiss. A closed form solution to natural image matting. *PAMI*, 2008. 5
- [12] C. Liu, J. Yuen, and A. Torralba. Sift flow: Dense correspondence across scenes and applications. *PAMI*, 2011. 5
- [13] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph.*, 2006. 3
- [14] J. Renno, D. Makris, T. Ellis, and G. Jones. Application and evaluation of colour constancy in visual surveillance. In *Int. Workshop on Performance Evaluation of Tracking and Surveillance*, 2005. 2
- [15] S. A. Shafer. Using color to separate reflection components. *Color Research and Applications*, 1985. 2, 3
- [16] R. Tan and K. Ikeuchi. Estimating chromaticity of multicolored illuminations. In *Workshop on Color and Photometric Methods in Computer Vision*, 2003. 2
- [17] R. Tan, K. Nishino, and K. Ikeuchi. Illumination chromaticity estimation using inverse-intensity chromaticity space. In *CVPR*, 2003. 2, 6
- [18] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE TIP*, 2007. 6
- [19] N. Wang, B. Funt, C. Lang, and D. Xu. Video-based illumination estimation. In *Int. Conf. on Comp. Color Imaging*, 2011. 2, 5, 6, 7
- [20] Q. Yang, S. Wang, N. Ahuja, and R. Yang. A uniform framework for estimating chromaticity, correspondence, and specular reflection. *IEEE Trans. Im. Proc.*, 2011. 2, 3