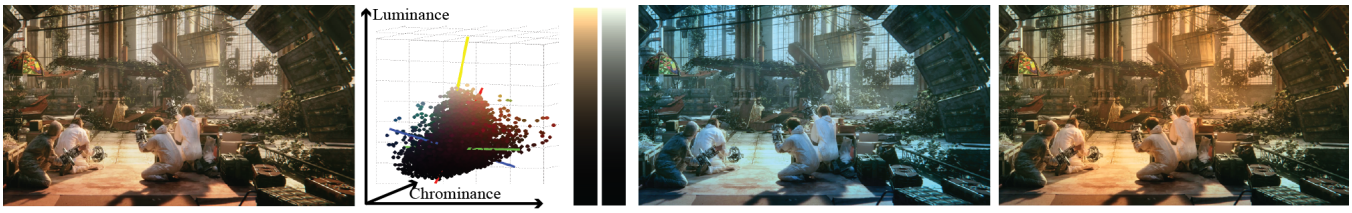
Mixed Illumination Analysis in Single Image for Interactive Color Grading

Sylvain Duchêne
b<>com & Technicolor
France
sylvain.duchene@b-com.com

Carlos Aliaga
Universidad de Zaragoza & Technicolor
Spain/France
caliaga@unizar.es

Tania Pouli
Technicolor
France
tania.pouli@technicolor.com

Patrick Perez
Technicolor
France
patrick.perez@technicolor.com



(a) Input Image (b) Shading Hue Extraction (c) Our Grading per Illuminant (d) Our Grading per Illuminant

Figure 1: Our method allows for complex edits of a single image: Given a single image input (a), we analyze within it a suitable collection of local color variations to extract distinct illumination hues, two in this case (b); The input image can now be modified simply by changing each illuminant hue, thus obtaining different color gradings with minimal user effort (c,d). (Images from *Tears of Steel* open source movie - (CC) Blender Foundation | mango.blender.org)

ABSTRACT

Colorists often use keying or rotoscoping tools to access and edit particular colors or parts of the scene. Although necessary, this is a time-consuming and potentially imprecise process, as it is not possible to fully separate the influence of light sources in the scene from the colors of objects and actors within it. To simplify this process, we present a new solution for automatically estimating the color and influence of multiple illuminants, based on image variation analysis. Using this information, we present a new color grading tool for simply and interactively editing the colors of detected illuminants, which fits naturally in color grading workflows. We demonstrate the use of our solution in several scenes, evaluating the quality of our results by means of a psychophysical study.

CCS CONCEPTS

• **Computing methodologies** → **Image manipulation**; *Computational photography*;

KEYWORDS

Mixed Illumination, Color Grading, Illuminant detection

ACM Reference format:
Sylvain Duchêne, Carlos Aliaga, Tania Pouli, and Patrick Perez. 2017. Mixed Illumination Analysis in Single Image for Interactive Color Grading. In *Proceedings of Expressive Symposium, Los Angeles, California, USA, July 2017 (EXPRESSIVE 2017)*, 10 pages.
https://doi.org/10.475/123_4

EXPRESSIVE 2017, July 2017, Los Angeles, California, USA

ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/10.475/123_4

1 INTRODUCTION

A crucial phase of movie post-production is that of color grading, where creative experts alter the color, look, and feel of content so that it visually expresses the intent of the director. The process of color grading involves two key tasks. On the one hand, content captured under different light source temperatures needs to be adjusted to attain a consistent appearance. More crucially, colorists can modify the image colors globally or locally to guide the focus of the viewer, create a particular atmosphere and translate the vision of the director into an image.

A commonly used technique to apply local color edits is *color keying*, where a mask over a specific range of colors is created and subsequently used for applying color edits. For example, to diminish the red reflections on an actor's face caused by a nearby red light, a mask selecting reddish pixels might be created first. Then the hue and saturation for these corresponding pixels might be subtly adjusted. Although this method can provide pixel-perfect, temporally coherent masks, it cannot separate the influence of

illumination from underlying object colors. In the above example, the red hue of the illumination will be modified in conjunction with the actor's skin tone, potentially leading to unnatural transitions. Furthermore, color keying implies that all other red objects in the scene will be selected and modified as well.

An alternative approach is *rotoscoping*, where a precise mask is manually created by the professional to isolate a particular object or area of the image. This process is very labor-intense and time-consuming, and recent advances towards automation have reduced the need of manual intervention [27]. However, the challenge of separating and editing illumination in the scene remains open.

To address this task algorithmically, the influence of the illumination needs to be estimated in the image, which represents an extremely ill-posed problem. It becomes particularly challenging when working with a single image as input, since each pixel is the integration of lighting, geometry and reflectances of the scene, all unknown. In the context of intrinsic decomposition of single images, existing methods assume grey illuminant only to separate the shading from the reflectance in the image, but cannot distinguish the influence of different light sources [4, 20]. On the other hand, several automatic methods exist for white balancing images — that is to say removing the influence of colored light sources — which could serve as a first step for extracting the influence of lighting colors present in the scene, but they are either appropriate only for scenes with a single illuminant [17], or require user input to separate different lighting colors [9, 24]. All in all, none existing method allows to edit light source influence independently without altering the image quality, making this approach impracticable in the context of movie production.

Our aim is to allow users to modify the color of multiple illuminants with a tool that fits the color grading production workflow. The requirements to add such tool to the colorists' toolset are to be intuitive, interactive and to preserve the image quality. Ultimately, our goal is to provide a solution that allows users to modify the color of each illuminant separately after capture, by detecting and separating the influence of different light sources in mixed illumination environments, using only a single image as input.

To achieve this goal, we observe that under mild assumptions, similar local chrominance variations in the image are more likely explained by the influence of a given illuminant than by reflectance. Based on this observation, we carefully select and analyze local color variations in the image. A suitable clustering of these variations in the CIELab color space, where chrominance and luminance are decorrelated, allows us to infer the hue of different dominant illuminants, as well as their respective zones of influence, with no need to attempt intrinsic decomposition. As a result, the colors of these illuminants can then be modified in the scene without affecting contrast and details. Our solution allows for real-time edits in an intuitive manner that fits naturally within existing color grading tools and workflows.

In summary, our main contribution is two-fold:

- A method to estimate the hues of potential illuminants within a single image by analyzing a selection of local chrominance variations, with no need of user intervention or semantic knowledge of the scene of the image.
- A new interactive tool for modifying the hue of each detected illuminant and efficiently propagating such edits in real-time, easing the process for practitioners specially for complex edits.

One of the key objectives in this work is to meet the requirements of movie post-production. We achieve high quality, natural results through an intuitive workflow that fits color grading practices, while offering real time performance.

2 RELATED WORK

A common task in image manipulation is the modification of image colors. This can be in order to correct an undesirable tint in the image, as would be the case when white balancing photographs, or to achieve a particular creative look or style, as is the case when color grading a film. In either scenario, one can modify image pixels directly without considering the image content, or can attempt to understand the underlying image structure to modify scene properties, such as lighting or materials, in a more meaningful sense. Our work draws inspiration from both. As such key techniques from both classes will be discussed here, focusing on creative image edits.

2.1 Pixel-oriented Editing

Arguably the most common class of methods aimed at modifying image colors and style is collectively known as *color transfer*. The goal of such methods is to transfer the color characteristics of a reference image that exemplifies a desired look onto another image or video [19]. This can be achieved by globally modifying the color distribution of the image [28, 29] or by considering higher level, localized statistics in the image to guide the assignment of colors [34].

Recently, the idea of color transfer has been employed in the context of film color grading, offering an efficient way of applying a particular style to a film. Rather than using a single reference image, Xue et al. [35] analyzed clips from a variety of feature films, extracting distinctive styles that could then be applied at will to new content. A video reference was used by Bonneel et al. [6] to define the target style for a sequence. Their approach requires a user-driven separation of foreground and background such that luminance histograms in "shadows", "midtones" and "highlights" bands can be modified akin to the manner a colorist would perform such a task in grading software. A temporal filtering is performed, smoothing points of higher temporal variation to ensure temporal coherence in the final result.

Another alternative for modifying colors in an artistic context is to extract a representative color palette from the image and modify the palette colors directly. The main challenge in this case is the extraction of a representative color palette, which clearly separates different parts of the image in a meaningful manner. Different statistical tools have been proposed to this end and for the subsequent recoloring step, for example using optimized clustering approaches based on k-means [11, 36]. Recently, the concept of color palettes has been extended decomposing the image into coherent layers, not only based on the color content but also considering the structure [2, 33]. Although both palette and layer approaches offer simple means for editing images that fit well within photo

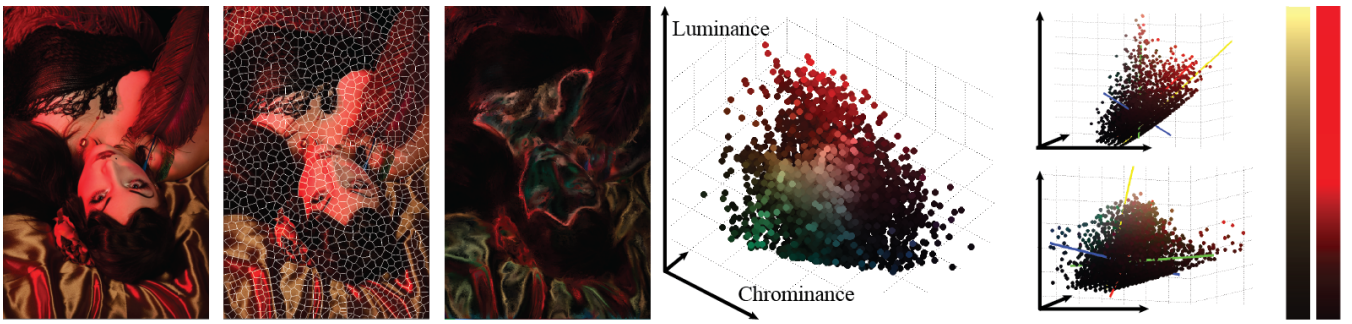


Figure 2: Proposed analysis pipeline: The input image (a) is over-segmented into compact super-pixels (b); Local radiance variations in the image are sampled within super-pixels (c) and gathered into the orthogonal, opponent color space CIELab (d); Spectral clustering using chrominance-based similarities and per-cluster orientation analysis is performed (e); The distinct hues inferred from these clusters (f) explain main observed shading variations and can be used to perform per-illuminant color grading of the scene. Note that the radiance variations in (c) have been spatially interpolated by gradient-aware filtering for visualization. (Photograph copyright: Rémi Cozot.)

editing workflows, they do not explicitly separate illumination from surface colors.

Although their goal is not primarily creative, *white balancing* or *color constancy* techniques are also concerned with modifying image colors, specifically for correcting the bias of illumination colors in the scene [17, 22]. Traditionally white balancing methods assume a single illuminant, but more recently, methods addressing mixed illumination conditions have appeared. Considering multiple illumination colors, Ebner et al. [18] compute, through filtering, local average colors that are used as location-dependent white points. Alternatively, the image can be divided into patches, and local, per-patch white points may be computed and clustered to determine the most representative illuminant colors [3, 5, 23]. Although these methods can work well in cases where the illumination varies smoothly across a surface, they cannot handle sharp discontinuities in the image, leading to potential halo artefacts. This has been partially addressed by detecting different illuminants but offering a global correction according to user preferences [12], however such a solution would severely limits the editing freedom in a creative context.

Within the context of mixed lighting, Hsu et al. [24] formulate white balancing as a matting problem, whereby the contributions of known light colors must be untangled. In addition to white balancing, this method allows modifying the chrominance of each identified light source for subsequent editing. However, this approach is limited to two sources and critically relies on the capacity of the user to estimate the color of the two illuminants. It also requires that a given reflectance is observed under a number of different proportions of the two illuminants. The user-assisted method of Boyadzhiev [9] tackles these issues in the context of mixed lighting, considering scenes with more than two sources. To achieve this however, the user needs to indicate regions with similar reflectance, as well as neutral regions that should be white, using scribbles on the image.

2.2 Inverse Rendering based Approaches

The approaches discussed so far consider pixel information but do not attempt to understand the underlying image structure. Although pixel-based approaches remain simpler algorithmically, knowledge of the geometry, materials and illumination interacting in the original scene would permit more complex and precise edits to be performed.

To retrieve the different contributions comprising the measured radiance at each pixel, the inverse rendering problem can be considered. In each of the RGB channels, the radiance $I_p \in \mathbb{R}$ measured at pixel p in image I results from the sum of the radiances emitted and reflected by corresponding 3D point $\mathbf{x}_p$ in the direction $\boldsymbol{\omega}_p$ of the line of sight:

$$I_p = L_e(\mathbf{x}_p, \boldsymbol{\omega}_p) + L_r(\mathbf{x}_p, \boldsymbol{\omega}_p). \quad (1)$$

The reflected radiance reads:

$$L_r(\mathbf{x}_p, \boldsymbol{\omega}_p) = \int_{\Omega_p} R(\mathbf{x}_p, \boldsymbol{\omega}_i, \boldsymbol{\omega}_p) L_i(\mathbf{x}_p, \boldsymbol{\omega}_i) \mathbf{n}_p \cdot \boldsymbol{\omega}_i d\omega_i, \quad (2)$$

where $R(\mathbf{x}_p, ., .)$, $L_i(\mathbf{x}_p, .)$, $\mathbf{n}_p$ and Ω_p are respectively the bi-directional reflectance, the incident light, the surface normal and the viewing hemisphere at point $\mathbf{x}_p$.

The method of Debevec et al. [14] recovers the different terms in (1)-(2) with the help of a 3D scanned geometry and the acquisition of both lighting and reflectance, which allows subsequent editing of both lighting and reflectance. With the emergence of photogrammetry, the method of Laffont et al. [26] allows lighting transfer between multiple captures of the same scene under different lighting conditions, R and L_i being estimated under the assumption of Lambertian reflectance. The method of Duchêne et al. [16] allows the editing of R and L_i for outdoor scenes only captured under similar lighting conditions, by inferring iteratively an image driven completion of L_i. For all the methods described above, the sun, the sky as well as the influence of indirect illumination are considered separately to estimate the incoming radiance received per point. However, all these approaches can edit either R and/or L_i.

To the best of our knowledge, there exists no automatic single-image method that allows the manipulation of the incoming radiance. Existing single-image methods focus instead on retrieving the reflectance R to edit it. The intrinsic decomposition model simplifies the reflected radiances (2) in image I as the product of a Lambertian reflectance R and a shading layer S, which is monochromatic in the methods of [4, 8, 20].

Our method draws inspiration from all the above. Similar to automatic white balancing approaches, we detect different illumination colors and their influence, without input from the user. However, we rely on some of the assumptions that underlie intrinsic images. We target our approach towards color grading applications but do not require the selection of reference content or the computation of an intrinsic decomposition. Instead, we provide a simple editing interface that fits color grading workflows because the main quality for a production tool are to be interactive, to preserve quality and to leave control to the artists. We consider that an artist is not going to use a tool if a global driven solution updates his local constraints, in the same way user assisted rotoscopy methods [27] are preferred to automatic rotoscopy algorithms in the production industry [10].

3 THEORY AND OVERVIEW

In absence of emitted radiance and assuming Lambertian reflectance, the rendering Equations 1-2 simplify to:

$$I_p = R(\mathbf{x}_p) \underbrace{\int_{\Omega_p} L_i(\mathbf{x}_p, \boldsymbol{\omega}_i)\mathbf{n}_p \cdot \boldsymbol{\omega}_i d\omega_i}_{:=S(\mathbf{x}_p)}, \tag{3}$$

where $S(\mathbf{x}_p)$ is the shading at pixel p, which depends on both the surface geometry at point $\mathbf{x}_p$ and the light source model at this point. This is the classic decomposition into reflectance and shading that intrinsic imaging aims to untangle.

Now let us consider two nearby points $\mathbf{x}_p$ and $\mathbf{x}_q$ that belong to the same surface of the 3D scene and project onto neighboring pixels p and q in the image plane. Assuming that they share the same reflectance, the local radiance variation between the two points reads:

$$|I_p - I_q| = c_{pq}|S(\mathbf{x}_p) - S(\mathbf{x}_q)|, \tag{4}$$

where $R(\mathbf{x}_p) = R(\mathbf{x}_q) = c_{pq}$ is independent of the lighting and of the geometry, hence of the shading variation $|S(\mathbf{x}_p) - S(\mathbf{x}_q)|$. Let's assume further that the lighting of the two points is primarily due a same illuminant. The intensity of their respective illuminations by this common source still depends on the surface orientation and light attenuation. The intrinsic decomposition problem from a single radiance variation remains ill-posed, even under our assumptions. However, if we consider multiple point pairs within the image that approximately follow these assumptions, we observe that the associated collection of radiance variations (4) in RGB space exhibit several elongated arms. The intuition is that these few trends that emerge from local radiance variations in the image are due to the dominant illuminants in the scene.

It is difficult to analyze this behaviour in RGB space due to its correlation properties [30]. However, we note that a light source emits with a constant hue and saturation within all directions. Hue and saturation of illumination are thus the same for the two points $\mathbf{x}_p$ and $\mathbf{x}_q$. Saturation being the ratio between chroma and luminance, the chroma variation due to illumination with constant hue and saturation will be correlated to the luminance variation. Based on these observations, we can formulate our key insight: if a subset of the local image radiance variations exhibit luminance and chroma correlations that amount to the same hue, the corresponding pairs of 3D points (and associated pixels) are likely to be lit by the same illuminant of said hue.

To facilitate the discovery of such groups, we opt to work in the CIELab opponent color space, which separates luminance (L) from chrominance (a-b) information. Further, a constant hue in CIELab translates to a constant ratio between the a and b components. To illustrate this, we simulated radiance variations under constant hue (with random perturbations) and correlated luminance/chroma (Figure 3). In an orthogonal, opponent color space such as CIELab, these variations are organized along 3D lines, which should make their extraction possible from real data.

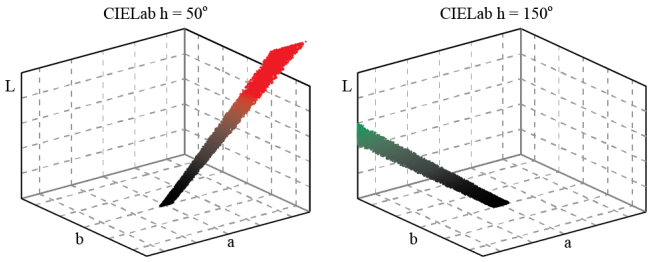


Figure 3: Local radiance variations due to a given illumination are likely to have similar hue and saturation, while luminance may vary. In CIELab space, variations with such characteristics are organized along a line with a large span of luminances L and constant ratio of a and b values. We show two such structures with simulated radiance variations for two distinct hues.

Considering a complete scene under mixed illumination, local radiance variations due to the shading are thus likely to concentrate around a few constant hue directions (with varying chroma and lightness) that correspond to distinct illuminants. We aim at detecting these meaningful lines within the collection of sampled variations in CIELab space. Not only these structures will give access to illumination colors but also to associated pixel locations.

Once obtained, these lines and associated locations provide a powerful means to edit illumination in the scene. To grade the image, we can independently manipulate them to adjust illumination colors. In contrast to previous work, this allows us to apply complex edits with no need to estimate per-pixel mixtures of different illumination colors or to perform intrinsic decomposition.

In practice, a challenge is to sample pixel pairs such that they are likely to satisfy our two key assumptions: constant reflectance and shared main illuminant across the pair. The spatial sparsity of reflectance changes (while reflectances themselves can take a multitude of values) and the limited number of illuminants in most scenes suggest that suitable pairs should be found within image fragments output by a fine-grained image segmentation. We opt for an off-the-shelf super-pixel extractor and compute a small number of variation samples within each super-pixel.

As summarized below and explained in detail in Sections 4 and 5, our approach has three main steps: (i) Over-segmentation of the image and sampling of radiance variations; (ii) Extraction of most likely illuminant hues from the collected variations; (iii) Real time interactive grading based on the identified illuminant hues. Our only input is the single image to be edited, captured under mixed illumination conditions. Intermediate results for several steps of our algorithm are shown in Figure 2. Section 5.2 describes several important optimizations to speed-up preprocessing steps (i-ii).

Sampling the variations (Section 4.1). We first segment the input image using the super-pixel method of Duan and Lafarge [15] (Figure 2(b)). The geometry of these super-pixels is particularly adapted to the sampling of relevant radiance variations, *i.e.*, which are most likely due to shading variations within the image. A visualization of such variations is given in Figure 2(c).

Mining the variations (Section 4.2). We identify in the CIELab opponent space the most likely illuminant hues based on the sampled color variations (Figure 2(d-f)). Chrominance-based spectral clustering of the collected data in this space and robust directional analysis of each cluster, provide a set of illuminants, each with a distinctive hue and associated super-pixels in the image.

Grading content (Section 5). The hue of identified illuminants can be modified independently, with corresponding changes being automatically applied to the appropriate image locations with the help of a domain transform filter [21]. In order to preserve texture details, these changes are applied to chrominance only, as done by colorists.

We apply our method to a variety of challenging scenes, keeping in mind the performance and quality requirements imposed by production workflows. Typically, all the image pre-processing is performed in less than 10 seconds, while subsequent editing can be applied and visualized in real-time. Our approach is the first to propose multi-illuminant editing of a single image, based on fast, automatic image analysis. It offers a tool that complements the usual toolset – rotoscoping, matting and keying – used by a colorist.

4 EXTRACTING ILLUMINANTS

As briefly discussed in the previous section, we first sample local radiance variations within small, homogeneous segments of the image. To identify potential coloured illuminants, we then look for chromatic similarities among these variations in an opponent color representation. We provide next the details of these different steps.

4.1 Sampling Shading Variations

Selecting pairs of nearby pixels such that they have similar reflectance and that their shading difference in (4) provides useful information on the incoming light is challenging. Choosing them in the same object fragment, as the use of super-pixels will strongly favour, is not sufficient. In order for their shadings to differ at all, their surface normals should be different, *e.g.*, they should not sit on a nearly flat surface. Ideally, we would like to place these pixel pairs near object visual boundaries, where the local changes of normal direction are often more pronounced than in other regions. In absence of geometric information on the scene, we will instead look for pixel pairs whose positioning in a super-pixel make them more likely to be in this ideal configuration.

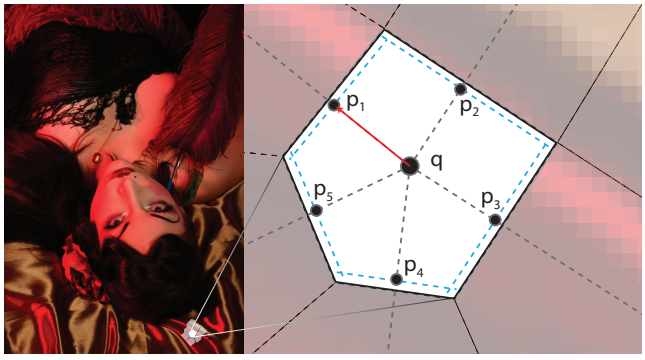


Figure 4: To sample color variations within each super-pixel, all directions toward neighbouring super-pixels (gray dashed lines) are sampled by positioning q at the centroid and the p's near the boundaries. (Photograph copyright: Rémi Cozot.)

Over-segmenting the image into small homogeneous regions, a.k.a super-pixels, has already proved very useful to compute single image intrinsic decomposition [4, 20], where they help finding areas of similar reflectance. For our purpose, we want in addition to measure tangible and meaningful shading variations within super-pixels. We must keep these considerations in mind when selecting the segmentation technique.

K-means clustering [25], mean shift segmentation [13] and SLIC super-pixels [1] are among the most popular over-segmentation methods. However, we find the recent approach of Duan and Lafarge [15] more suited to our specific requirement of getting not too uniform segments while, at the same time, respecting as much as possible the structures in the image (the occlusion boundaries in particular). The super-pixels in this approach are Voronoi cells, *i.e.*, they form convex polygonal regions defined by their centroids, with borders aligned to strong contours when present. This unique geometric feature provides natural sampling directions inside a segment: Taking q as the centroid of a super-pixel, and moving as far as possible in the cell along its edges in the associated Delaunay tessellation (*i.e.*, along the lines to the centroids of the adjacent cells) is an efficient way to sample shading variations. See example in Fig. 4 where five pairs (q, p_i), $i = 1 \cdots 5$, are thus obtained.

Note that the segmentation method of [15] is parametrized by the minimum distance between two centroids, a parameter that controls in turn the final number of super-pixels. We set it to 20.0 pixels in all experiments, obtaining between 1000 and 1500 super-pixels in average for 1000-pixel wide images. With this setting, the size of the super-pixels suits our purpose (respecting well the image structures while providing sufficient segment-wise shading variations) and their limited number allows efficiency, as discussed in Section 5.2

For each super-pixel, the sampling procedure provides as many samples as it has neighbours. This results in an overall collection of N_s RGB variations, $|\mathbf{I}_{q_n} - \mathbf{I}_{p_n}| \in \mathbb{R}^3_+$, $n = 1 \cdots N_s$, with absolute value taken component-wise and N_s amounting approximately

to 5 times the number of super-pixels. An example of sampled variations is shown in Fig. 2(c) where, for visualization purpose, they are extrapolated through gradient-aware filtering.

4.2 Mining Variation Similarities

From the set of N_s samples, we now aim at inferring a small number of dominant illumination hues that would best explain the shading variation embedded in each sample. As discussed in Section 3, no useful information can be directly extracted by conducting this analysis in RGB due to the cross-channel correlation. Using instead an opponent color space that separates chrominance from lightness allows us to focus on chromatic variations, and later on, to apply edits on chrominance, hence without affecting lightness within the image. To this end, we project the collection of RGB variations $\{|\mathbf{I}_{q_n} - \mathbf{I}_{p_n}|\}_{n=1}^{N_s}$ into the CIELab color space, yielding a new collection of points $\{\mathbf{y}_n\}_{n=1}^{N_s}$, where $\mathbf{y}_n = [y_n^{\mathsf{a}}, y_n^{\mathsf{b}}, y_n^{\mathsf{L}}]^\top$ is the CIELab counterpart of $|\mathbf{I}_{q_n} - \mathbf{I}_{p_n}|$ (Fig. 2(c)). We shall denote $\mathbf{z}_n = [y_n^{\mathsf{a}}, y_n^{\mathsf{b}}]^\top$ the chrominance part only.

We need to partition this collection of chromatic variations $\mathbf{z}_n$s into an unknown number of groups. To this end, we resort to normalized cut [32], a variant of spectral clustering with automatic determination of the number of clusters. We first construct the symmetric order-N_s weight matrix $W = [w_{mn}]$ defined by:

$$w_{mn} = \exp -\frac{\|\mathbf{z}_m - \mathbf{z}_n\|}{2\sigma^2}, \; m, n = 1 \cdots N_s, \tag{5}$$

where $\sigma = 4.0$ in all experiments. The similarity weight between two chrominance variations is a decreasing function of their chromatic distance. The normalized Laplacian matrix is deduced from W:

$$L = \mathbb{I} - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}, \tag{6}$$

where $\mathbb{I}$ is the order-N_s identity matrix and $D = \mathrm{diag}(d_1 \cdots d_{N_s})$ is the diagonal matrix of degrees $d_n = \sum_m w_{mn}$. This matrix is positive semi-definite with eigenvalues $\lambda_0 = 0 \leq \lambda_2 \leq \cdots \leq \lambda_{N_s-1}$ and associated eigenvectors $\mathbf{u}_0 \cdots \mathbf{u}_{N_s-1}$. Given K, the target number of clusters, this eigen-decomposition is used to embed all input vectors as follows: the n-th input vector $\mathbf{z}_n$ is now mapped to the n-th row of the $N_s \times K$ matrix $U = [\mathbf{u}_1 \cdots \mathbf{u}_K]$. These K-dim representations are clustered with K-means. The number of clusters is automatically set as $K = \min\{k : \frac{\lambda_{k+1}}{\lambda_k} \leq 0.97\}$.

As a result of previous step, CIELab variation samples $\{\mathbf{y}_n\}_{n=1}^{N_s}$, and associated super-pixels as in Fig. 6(b-e), are partitioned into K groups, according to chrominance similarities. Based on our assumptions, each of these clusters should be explained by a single dominant illuminant with a distinct hue. Denoting $\mathcal{J}_k \subset (1, N_s)$ the index set of the k-th cluster, the set $\{\mathbf{y}_n\}_{n \in \mathcal{J}_k}$ should accordingly exhibit a one-dimensional structure that corresponds to a constant hue with varying chroma and lightness. The estimation of this linear structure is delicate: small variations, which form the majority of the cluster, are noisier and much less informative than the few ones with large magnitudes. In order to make robust the analysis, we proceed iteratively. A first estimate of the 3D line is defined as the one passing through the center of mass of the set and following the most vertical among its three principal directions computed through PCA (shown as red, green and blue axes in Fig. 2(e)), that is the one capturing the largest span of lightness variations within the samples. The set is split in two parts by the plane normal to this selected direction and passing through the center of mass. The bottom subset is discarded and the procedure is iterated, until 15% only of the original cluster is left. The final estimate is the line joining the center of mass of the final sample set to the origin of CIELab, as shown in yellow in Fig. 2(e). The intersection of this line with the maximum lightness plane ($\mathsf{L} = 100$) is used to compute the hue of the illuminant for the cluster.

(a) Input Image (b) Propagation Layer (c) Our Grading

Figure 5: In the input image (a), three illuminant hues where identified –see last row of Fig. 6–, and the yellow one is modified with a green shift. This hue shift is applied to the sampling points inside the super-pixels attached to this illuminant, and propagated to all pixels (b) to obtain the grading (c). For visualization purposes, the propagated chrominance shift is shown in RGB here. (Image from *Tears of Steel* open source movie - (CC) Blender Foundation | mango.blender.org)

5 COLOR GRADING

Once the K dominant illumination hues and their respective zone of influence (at the super-pixel level) are extracted, we can use them to modify the illumination colors in the image. As we target color grading workflows, two main requirements have to be met. First, the edits should look natural and consistent with the image structure. Second, the proposed tool should allow users to modify parameters and control the resulting image look interactively. We discuss next our design decisions in the light of these requirements.

5.1 Quality Grading

One of the critical constraints in a grading tool is the preservation of details within images. For this reason, our approach is directly inspired from the chrominance shifting method, used extensively by colorists to grade content. This approach is also used in color transfer methods for its ability to preserve the histogram shape [9].

For the editing part of our system, we remain in the CIELab color space, to benefit from its separation of luminance and chrominance information, and we assume input images are encoded in RGB with a 2.2 gamma. Given a user-specified chrominance shift $\mathbf{\Delta} = [\Delta^{\mathsf{a}}, \Delta^{\mathsf{b}}]^\top$ to the k-th illuminant, this shift is first applied to all sampling pixels p_n, $n \in \mathcal{J}_k$, on the inner border of super-pixels that this illuminant dominates. Using these localized chrominance shifts as soft constraints, a pixel-level field of shifts is computed through structure-aware diffusion with the "domain transform filter" [21], using the implementation proposed by [31]. We visualize these chrominance shifts as an RGB propagation layer in Fig. 5(b). Their application to the chrominances in the original input image produces the grading. Choosing the domain transform filter has three advantages:

- It incorporates the localized target shifts as soft constraints and propagates them in a soft manner;
- Yet, it is aware of the luminance discontinuities of the input image, thus respecting its structure;
- Its implementation is based on a recursive filter, allowing great performance on both CPU and GPU.

To further improve the coherency of the propagated chromaticity shifts, the super-pixel labelling is pre-processed such as to eliminate isolated labels: if the illuminant label k of a super-pixel is absent from its direct neighbours, this super-pixel (and associated samples) is relabelled according to the dominant label among its neighbours. For super-pixels at the border of the image, which have fewer direct neighbours, second-order neighbours are also taken into account for assessing isolation.

The chrominance of all pixels is modified in the final image. If only one illuminant is edited, as in Fig. 5, this modification is stronger for pixels whose shading is dominated by the edited illuminant and more subtle, even indiscernible in some locations, for other pixels. Also, note that since all modifications are applied to the a and b channels in CIELab, the luminance of the image remains unchanged, which is key to preserve the original detail of the input image.

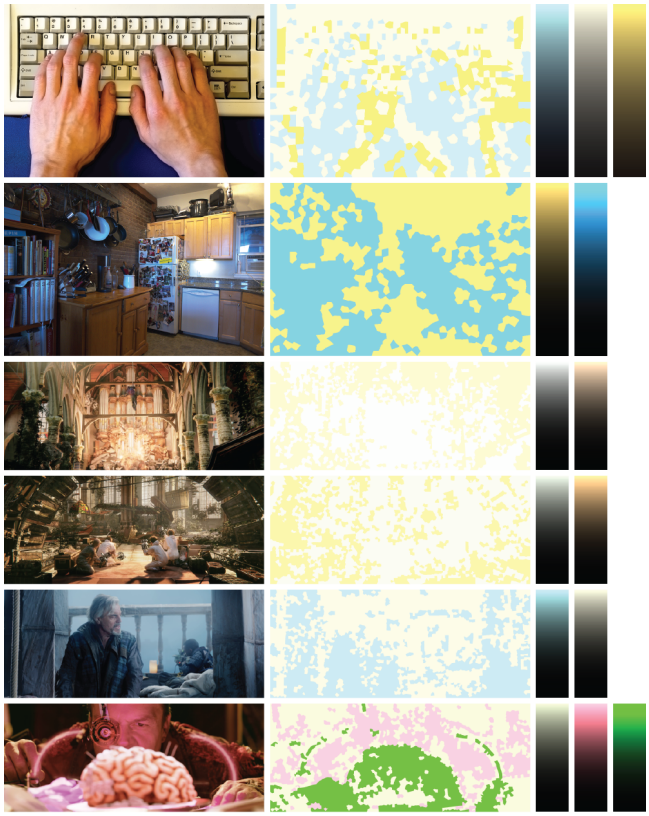
It should be noted that, although the edits of the image are computed and applied in the $\mathsf{a}-\mathsf{b}$ chrominance channels, we have the flexibility to express the illuminant modification in any color space and to translate it in terms of chrominance shift. As can be seen in the accompanying video, we provide editing tools that control a and b channels directly as well as tools operating in the more user friendly HSV space. We show in Figure 8 a collection of grading results that demonstrate the possibilities of our method on a variety of scenes.

5.2 Performance

To allow for fast pre-processing and real time editing, several choices and optimizations are necessary. For the interactive editing part, we take advantage of the recursive nature of the domain transform filter. Its highly parallel nature for both GPU and CPU allows real time manipulations.

A standard HD image typically produces about 4000-5000 super-pixels, each of them containing about 5 sampling directions (directions toward its neighbours). If all of them were sampled, the weight and Laplacian matrices (Section 4.2) would be approximately of size 25000×25000. It would not be possible to manipulate them at reasonable speed (*e.g.*, less than 1 second) on a current desktop computer, not to mention on a mobile phone where memory would be insufficient. For these reasons, we present several optimizations of the illuminant extraction method described in Section 4.

To reduce the size of the weight and Laplacian matrices, we first down-sample the input image, *e.g.*, to size 1280×720, which leaves us with about 1000 super-pixels. As explained in Section 4.2, each super-pixel is sampled in several directions from its center, to collect local chrominance variations. To reduce further the size of the manipulated matrices, we replace, for clustering purpose only, the samples from a super-pixel by a single representative one with maximum amplitude in each channel, *e.g.*, $\{|\mathbf{I}_q - \mathbf{I}_{p_i}|,\ i = 1 \cdots 5\}$ is replaced by $\max\{|\mathbf{I}_q - \mathbf{I}_{p_i}|,\ i = 1 \cdots 5\}$ in configuration from Fig. 4, with max operator being applied entry-wise. However, note that the complete set of N_s samples, around 5000 in total, is used for inferring illuminants hues.



(a) Original Image (b) Hue Label (c) Hues

Figure 6: Illumination hues detected by our algorithm for a variety of scenes and the corresponding labeling per super-pixel. (First image used with kind permission from Eugene Hsu. Second image used with kind permission from Ivaylo Boyadzev. Last four images from *Tears of Steel* open source movie - (CC) Blender Foundation | mango.blender.org)

Finally, we also enforce the sparsity of the matrix L before the eigen-decomposition by thresholding the similarity weights in Eq. 5. We note that the Maltab `eig` function performs an eigen decomposition of a 1000×1000 sparse matrix in less than 1*s* based on the Intel Math Kernel Library. For this reason, our C++ application links directly to Matlab.

6 EVALUATION

6.1 Psychophysical Evaluation

To evaluate the quality of our editing results we performed a psychophysical experiment, following the two-alternative forced choice (2AFC) methodology. In our study, the original grade and an edited result were shown side by side. Participants were asked to choose the image that they thought *was the original color graded image*. The goal of this study was to assess whether our editing

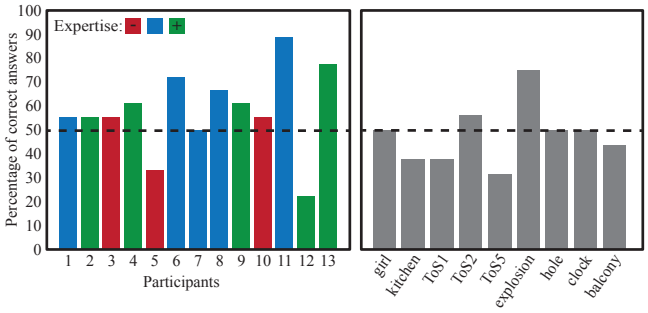


Figure 7: Summarized results per participant (left) and per scene (right) for our psychophysical experiment comparing original grades with our edited results. Correctness rates of 50% correspond to chance. Participants' level of expertise is denoted by the color of the bars in the left chart.

introduced artifacts or led to unnatural results, relative to the original editing that was applied by the colorist. As such, the main hypothesis was that if our results are natural, participants would not be able to correctly distinguish which image was the original colorist result, leading to 50% or less of correct answers.

Nine images were shown to 13 participants (10M/3F, age $\mu = 42.2, \sigma = 9.5$), with various editing styles used for the modified results, including both subtle modifications and strongly colored edits. Figure 7 shows the average responses per participant and grouped for each scene. The stimuli for this experiment can be found in the supplemental materials. We observe that for most scenes, correctness rates are at or below 50% (*i.e.*, chance), indicating that participants were not able to correctly identify the original graded image. Generally, strongly colored edits were deemed as less plausible, leading to higher rates of correct answers (*e.g.*, clock scene). Although most participants were able to correctly identify more than half of the scenes, for the majority, performances were only slightly above chance. No correlation was found between correctness of answers and participants' (self-assessed) expertise in imaging. Overall, our experimental findings suggest that the modifications achieved with our tool provide natural results that are deemed as plausible by human observers.

It should be noted that many of the images used in our study were taken from the same content (Tears of Steel open source movie). As such, it is possible that participants were additionally aided by the fact that many images exhibited a coherent style in the unmodified case, while our modifications went towards different directions in each case.

6.2 Discussion

Our approach relies on building an approximate physical model that permits plausible creative edits without recovering either the reflectance or the incoming radiance per point. Our solution can approximately estimate the hue of each illuminant affecting the scene, however one limitation to this estimation is that emitting surfaces within the image are not modelled. In the last example of Fig. 6, the hue estimation of the brain in the foreground, which is an emitting region, converged to the opposite CIELab direction since an emitting purple surface is missing green to form white. Nonetheless, our method still allowed us to manipulate this illuminant in a satisfactory fashion, as shown in Fig. 8, despite the inverted estimation.

A second observation is that our correction layer is not produced from hard constraints. Consequently, very localized shading variations might not be compensated for as a result, see in Fig. 8 (row 2) the man's fingers in the right part of the image. Using soft constraints allows us to obtain results that remain plausible visually, without determining precisely the influence of an illuminant per image region.

A key motivation for our method is that it is aimed at artistic, real-time edits, as would be necessary in a color grading session. As such, algorithmic and design decisions have been made accordingly, opting for results that are visually plausible, albeit not always physically correct. A more accurate estimation of the illuminant influence and hue might be possible with more restrictive constraints or more elaborate image analysis methods as preprocessing, however these would inevitably increase the computational complexity of our approach.

7 CONCLUSION AND FUTURE WORK

We propose the first automatic method to detect multiple illuminant hues and associated regions in a single input image. We put it at work in a new color grading tool that gives intuitive control to the user at interactive rates and produces high quality results, thus meeting post-production requirements in terms of editing speed and control intuitiveness. We demonstrate the merit of our approach as a standalone tool for colorists and validate the quality of the produced results by means of a psychophysical study.

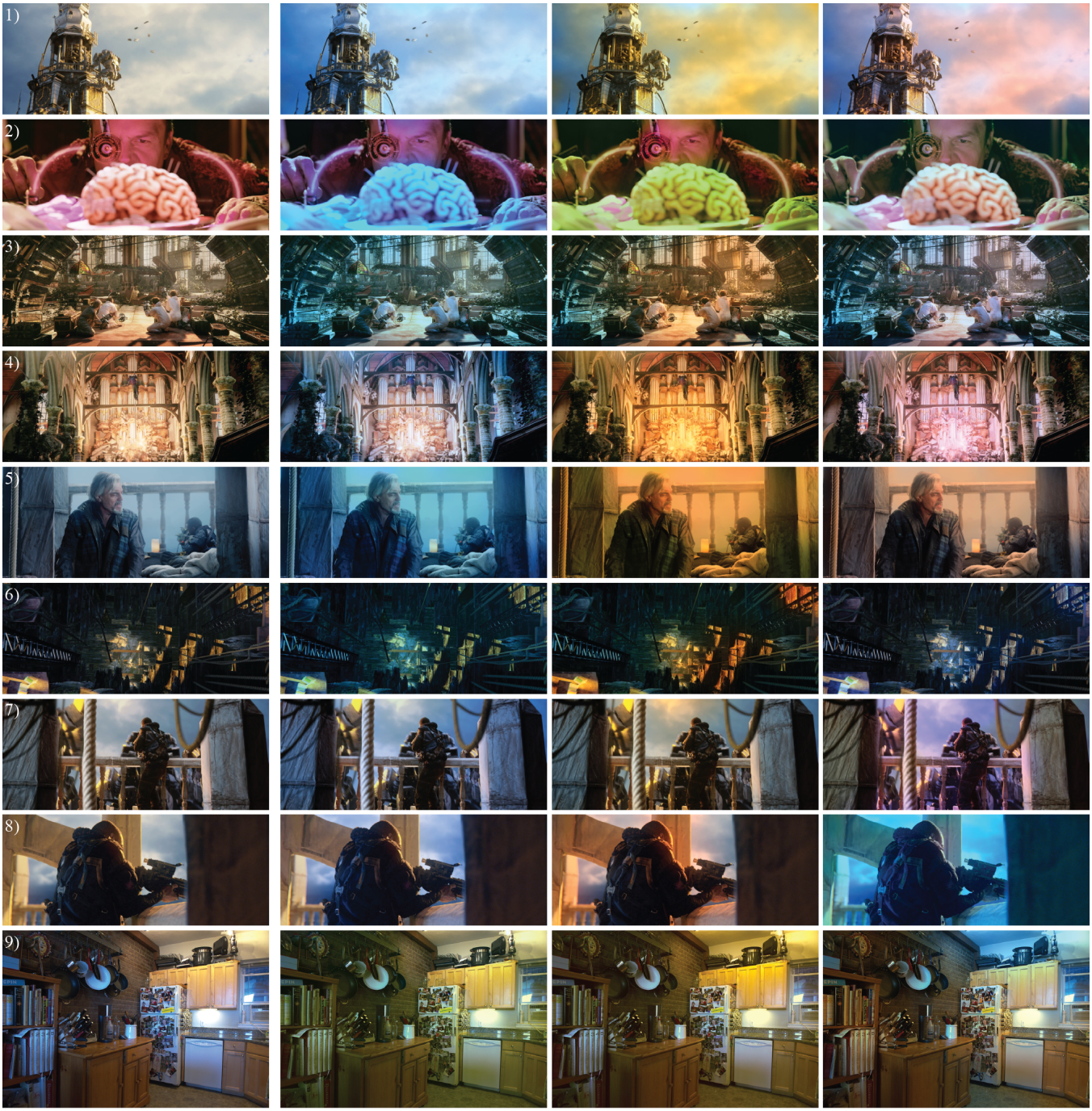
At its core, the proposed approach lies on a sparse sampling of the image radiance variations that allows us to build a physical proxy model. This approximate model makes illumination editing possible with no need to set up hard constraints or to perform intrinsic decomposition. Further, it does not require additional knowledge (assumed or measured) on the geometry of the scene and on indirect lighting.

Our method has been so far demonstrated on still images, however an extension to video content would be possible by considering an appropriate temporal smoothing approach, such as the blind temporal consistency approach of Bonneel et al. [7]. To edit video content, in accordance with typical color grading workflows, the user would work on key frames selected from each shot. Edits could be then propagated offline in a temporally consistent manner to the rest of the frames.

Our method for analyzing scene illumination opens up other possibilities. White balancing and inverse rendering in presence of multiple illuminants are among the open problems that can directly benefit from it. Progress in such directions will nonetheless require the acquisition of ground-truth datasets with realistic scenes under mixed illumination.

REFERENCES

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Machine Intell.* 34, 11 (2012), 2274–2282.



(a) Original Image (b) Our Grading Results

Figure 8: Examples of color grading obtained with our illumination editing tool. We show three distinct grading results of each of the 9 input images, demonstrating various effects, such as: (Row 1) Different edits of the contribution from the clouds; (Row 2) Two extreme hue changes of the emitting brain, followed by a shutting-down of this emitting surface; (Row 3) In the second result, the hues of the two illuminants are swapped and amplified, that is the outdoor lighting is shifted towards orange whereas the indoor lighting is given a colder tone. A similar inversion can be observed in the last edit in Row 9; (Row 8) Each of the two illuminants is first edited on its own, the sky shading on the character in the first result and the sun light received by the walls and character's face in the second one, and jointly modified in the last result. (Images 1-8 from *Tears of Steel* open source movie - (CC) Blender Foundation | mango.blender.org. Image 9 used with kind permission from Ivaylo Boyadzhiev.)

[2] Yağiz Aksoy, Tunç Ozan Aydin, Aljoša Smolić, and Marc Pollefeys. 2017. Unmixing-based soft color segmentation for image manipulation. *ACM Transactions on Graphics (TOG)* 36, 2 (2017), 19.

[3] Shida Beigpour, Christian Riess, Joost Van De Weijer, and Elli Angelopoulou. 2014. Multi-illuminant estimation with conditional random fields. *IEEE Trans. Im. Proc.* 23, 1 (2014), 83–96.

[4] Sean Bell, Kavita Bala, and Noah Snavely. 2014. Intrinsic Images in the Wild. *ACM Trans. on Graphics (Proc. Siggraph)* 33, 4 (2014).

[5] Michael Bleier, Christian Riess, Shida Beigpour, Eva Eibenberger, Elli Angelopoulou, Tobias Tröger, and André Kaup. 2011. Color constancy and non-uniform illumination: Can existing algorithms work?. In *ICCV Workshops*.

[6] Nicolas Bonneel, Kalyan Sunkavalli, Sylvain Paris, and Hanspeter Pfister. 2013. Example-Based Video Color Grading. *ACM Trans. on Graphics (Proc. Siggraph)* 32, 4 (2013).

[7] Nicolas Bonneel, James Tompkin, Kalyan Sunkavalli, Deqing Sun, Sylvain Paris, and Hanspeter Pfister. 2015. Blind Video Temporal Consistency. *ACM Transactions on Graphics (SIGGRAPH Asia 2015)* 34, 6 (2015).

[8] Adrien Bousseau, Sylvain Paris, and Frédo Durand. 2009. User-assisted intrinsic images. *ACM Trans. on Graphics* 28, 5 (2009), 130.

[9] Ivaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Frédo Durand. 2012. User-guided white balance for mixed lighting conditions. *ACM Trans. on Graphics* 31, 6 (2012), 200.

[10] B. Bratt. 2011. *Rotoscoping: Techniques and Tools for the Aspiring Artist*. Elsevier/Focal Press. https://books.google.fr/books?id=8PsAcEhBu_0C

[11] Huiwen Chang, Ohad Fried, Yiming Liu, Stephen DiVerdi, and Adam Finkelstein. 2015. Palette-based photo recoloring. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 139.

[12] Dongliang Cheng, Abdelrahman Abdelhamed, Brian Price, Scott Cohen, and Michael S Brown. 2016. Two Illuminant Estimation and User Correction Preference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 469–477.

[13] Dorin Comaniciu, Peter Meer, and Senior Member. 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.* 24 (2002), 603–619.

[14] Paul Debevec, Chris Tchou, Andrew Gardner, Tim Hawkins, Charis Poullis, Jessi Stumpfel, Andrew Jones, Nathaniel Yun, Per Einarsson, Therese Lundgren, Marcos Fajardo, and Philippe Martinez. 2004. *Estimating Surface reflectance properties of a Complex Scene under captured natural Illumination*. Technical Report ICT TR 06 2004. http://ict.usc.edu/pubs/ICT-TR-06.2004.pdf

[15] Liuyun Duan and Florent Lafarge. 2015. Image partitioning into convex polygons. In *Proc. CVPR*.

[16] Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez-Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. 2015. Multi-View Intrinsic Images of Outdoors Scenes with an Application to Relighting. *ACM Trans. on Graphics* (2015), 16. https://hal.inria.fr/hal-01164841

[17] Marc Ebner. 2007. *Color Constancy*. John Wiley & Sons, West Sussex, England.

[18] Marc Ebner. 2009. Color constancy based on local space average color. *Mach. Vis. Appl.* 20, 5 (2009), 283–301. https://doi.org/10.1007/s00138-008-0126-2

[19] H Sheikh Faridul, T Pouli, C Chamaret, J Stauder, E Reinhard, D Kuzovkin, and A Tremeau. 2015. Colour Mapping: A Review of Recent Methods, Extensions and Applications. In *Computer Graphics Forum*. Wiley Online Library.

[20] Elena Garces, Adolfo Munoz, Jorge Lopez-Moreno, and Diego Gutierrez. 2012. Intrinsic Images by Clustering. *Computer Graphics Forum (Proc. EGSR)* 31, 4 (2012).

[21] Eduardo S. L. Gastal and Manuel M. Oliveira. 2011. Domain Transform for Edge-Aware Image and Video Processing. *ACM Trans. on Graphics (Proc. Siggraph)* 30, 4, Article 69 (2011), 69:1–69:12 pages.

[22] Theo Gevers, Arjan Gijsenij, Joost Van de Weijer, and Jan-Mark Geusebroek. 2012. *Color in computer vision: fundamentals and applications*. John Wiley & Sons.

[23] Arjan Gijsenij, Rui Lu, and Theo Gevers. 2012. Color constancy for multiple light sources. *IEEE Trans. Im. Proc.* 21, 2 (2012), 697–707.

[24] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Frédo Durand. 2008. Light mixture estimation for spatially varying white balance. In *ACM Trans. on Graphics*, Vol. 27. ACM, 70.

[25] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. 2004. A local search approximation algorithm for k-means clustering. *Comput. Geom.* 28, 2-3 (2004), 89–112. https://doi.org/10.1016/j.comgeo.2004.03.003

[26] Pierre-Yves Laffont, Adrien Bousseau, Sylvain Paris, Frédo Durand, and George Drettakis. 2012. Coherent Intrinsic Images from Photo Collections. *ACM Trans. on Graphics (Proc. Siggraph Asia)* 31, 6 (2012). https://doi.org/10.1145/2366145.2366221

[27] Wenbin Li, Fabio Viola, Jonathan Starck, Gabriel J Brostow, and Neill DF Campbell. 2016. Roto++: Accelerating professional rotoscoping using shape manifolds. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 62.

[28] Tania Pouli and Erik Reinhard. 2011. Progressive color transfer for images of arbitrary dynamic range. *Computers & Graphics* 35, 1 (2011), 67–80.

[29] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. 2001. Color transfer between images. *IEEE Computer graphics and applications* 21, 5 (2001), 34–41.

[30] Erik Reinhard and Tania Pouli. 2011. Colour spaces for colour transfer. In *Computational Color Imaging*. Springer, 1–15.

[31] Joan Sol Roo and Christian Richardt. 2014. Temporally Coherent Video De-Anaglyph. In *SIGGRAPH Talks*. https://doi.org/10.1145/2614106.2614125

[32] Jianbo Shi and Jitendra Malik. 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.* 22, 8 (2000), 888–905.

[33] Jianchao Tan, Jyh-Ming Lien, and Yotam Gingold. 2016. Decomposing Images into Layers via RGB-Space Geometry. *ACM Transactions on Graphics (TOG)* 36, 1 (2016), 7.

[34] Fuzhang Wu, Weiming Dong, Yan Kong, Xing Mei, Jean-Claude Paul, and Xiaopeng Zhang. 2013. Content-Based Colour Transfer. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 190–203.

[35] Su Xue, Aseem Agarwala, Julie Dorsey, and Holly Rushmeier. 2013. Learning and applying color styles from feature films. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 255–264.

[36] Qing Zhang, Chunxia Xiao, Hanqiu Sun, and Feng Tang. 2017. Palette-Based Image Recoloring Using Color Decomposition Optimization. *IEEE Transactions on Image Processing* 26, 4 (2017), 1952–1964.