

Segmentation and Estimation of Spatially Varying Illumination

Lin Gu, *Student Member, IEEE*, Cong Phuoc Huynh, *Member, IEEE*, and Antonio Robles-Kelly, *Senior Member, IEEE*

Abstract

In this paper, we present an unsupervised method for segmenting the illuminant regions and estimating the illumination power spectrum from a single image of a scene lit by multiple light sources. Here, illuminant region segmentation is cast as a probabilistic clustering problem in the image spectral radiance space. We formulate the problem in an optimisation setting which aims to maximise the likelihood of the image radiance with respect to a mixture model while enforcing a spatial smoothness constraint on the illuminant spectrum. We initialise the sample pixel set under each illuminant via a projection of the image radiance spectra onto a low-dimensional subspace spanned by a randomly chosen subset of spectra. Subsequently, we optimise the objective function in a coordinate-ascent manner by updating the weights of the mixture components, the sample pixel set under each illuminant and the illuminant posterior probabilities. We then estimate the illuminant power spectrum per pixel making use of these posterior probabilities. We compare our method with a number of alternatives for the tasks of illumination region segmentation, illumination colour estimation and colour correction. Our experiments show the effectiveness of our method as applied to one hyperspectral and three trichromatic image datasets.

Index Terms

illuminant segmentation, region segmentation, illumination estimation, spatially varying illumination, multiple light sources

Lin Gu is with the Bioinformatics Institute (BII), Agency for Science and Technology Research (A*STAR), Singapore. E-mail: gulin@bii.a-star.edu.sg.

Cong Phuoc Huynh and Antonio Robles-Kelly are with NICTA (National ICT Australia), and the Research School of Engineering at the Australian National University, Australia. E-mail: {cong.huynh, antonio.robles-kelly}@nicta.com.au.

I. INTRODUCTION

Detecting and estimating illuminant colours are important tasks with applications in recognition and classification based on photometric invariants [1], white balancing [2], colour correction, digital media production and graphics [3]. Despite its importance, the recovery and identification of illuminant colours in a scene has proven to be a difficult task in uncontrolled real world scenes. This is mainly due to the fact that the recovery of region-wise illumination from a single image is an under-constrained problem [4]. As a result, existing methods often assume a uniform illumination power spectrum throughout the scene [5], [6]. Other works have made statistical assumptions on the image colours to remove the dependency on material colours, such as the “grey world” assumption [7], which implies that the statistical mean of the scene colours is achromatic. The “white patch” hypothesis [6] relies on the existence of a white pixel in the scene for illuminant colour inference. Both of the methods above are special instances of the “shades of grey” method [8] and are akin to edge-based colour constancy methods [9], which assume that the statistical mean of the spatial derivative across an image is only dependent on the colour of the illuminant, rather than the scene materials. Further, for colour balancing and correction, Hsu *et al.* [2] recovered a set of dominant material colours to estimate the local mixture coefficients of the lights. Recently, several authors proposed methods for white balance under mixed lighting conditions using user-assisted inputs and sparse interpolation [10], [11].

The global illumination and the statistical assumptions above are often too restrictive for real world scenes since shadows, skylight and direct sunlight can cause significant illuminant variations across the image [12], [13]. To handle varying illumination, Finlayson *et al.* [13] mapped image chromaticities to those of standard colour patches rendered under canonical illuminants. The Retinex theory [6], [14] has also been used for the removal of local illumination by considering the spatial changes in the image intensity. Based on the grey world hypothesis, Ebner [15] employed the local space average colour to perform colour correction. Wang and Samarasinghe [16] detected and estimated multiple illuminants from a single image of an object with known geometry and Lambertian reflectance using a region-based recursive least squares method. More recently, Finlayson *et al.* [17] detected the illumination colour in images using a chromagenic camera.

Unfortunately, the algorithms above do not contain a robust mechanism to distinguish illumination changes from reflectance variations. Further, the uneven nature of the shading over the surface often

adversely affects the stability of the results. To overcome these shortcomings, Barnard *et al.* [18] addressed the colour constancy problem by combining constraints based on the variations over the surface reflectance and illumination. To do this, they employed prior knowledge about plausible illuminants to characterise the image chromaticity changes caused by illumination variation. More recently, Bleier *et al.* [4] fused the outputs of different colour constancy algorithms after performing them on super-pixels in an image. However, this approach requires a substantial training process with complicated learning algorithms. Meanwhile, Riess *et al.* [19] recovered the illuminant colour per super-pixel from a single image by grouping local estimates into regions with similar illuminant colours. In a related approach, to recover the illumination colour in non-uniformly lit scenes, Gijssenij *et al.* [20] estimated the average illuminant colour in local image regions and combined these estimates in a subsequent pass.

In this paper, we present an unsupervised method for the segmentation of illumination regions and the estimation of the illuminant power spectrum in a scene lit by multiple light sources. This method is one of the few that are able to handle uncontrolled illumination in real-world environments using only input from a single image. Unlike supervised methods, such as that in [4], it does not involve complicated setups for learning prior knowledge from training data. Moreover, the method is applicable to a wide variety of natural scenes.

To solve the problem, we make the following assumptions

- We assume the radiance spectra at pixels illuminated by the same illuminant follow a log-concave distribution. Such a condition does not overly restrict the applicability of our method to real-world scenes since log-concave probability distributions include a wide variety of densities such as the Normal, Beta, Gamma, Subbotin, Logistic and Laplace distributions. Moreover, our approach is not limited to Lambertian reflectance or Munsell spectra, being applicable to general real-world scenes.
- We assume that the illumination power spectrum is piece-wise constant or slowly varying across image regions. This local smoothness assumption permits the enforcement of the spatial consistency of illumination in natural images and has been introduced to cope with cases where the radiance spectra within small isolated image patches with purely saturated hues do not belong to the radiance distribution of the surrounding pixels. Note that, without the spatial smoothness assumption, this kind of patches would be labelled as being illuminated by a separate illuminant. Hence, this assumption prevents the presence of isolated illumination regions resulting from these patches. Moreover, this

spatial assumption does not imply the spectral smoothness of the illuminant spectrum. Rather, the resulting illuminant segments are formed due to the differences between the sets of radiance spectra under each illuminant.

Here, we view the problem of segmenting illumination regions as a clustering one with respect to a mixture model. The clustering process occurs in an iterative manner where we initialise the algorithm with an effective separation of the illuminant regions based on the projection onto random subspaces presented in [21]. To this end, we employ a kernel density estimator [22] to approximate the likelihood of the image irradiance occurring under a particular light source. In addition, the local smoothness constraint on the illuminant allows the correction of spatial discontinuities across the resulting illumination regions. Subsequently, the average illuminant for each region is computed by applying illuminant colour estimation methods used elsewhere for single light sources, such as Grey-World [7], Grey-Edge [9] and White-Patch [6]. Finally, a per-pixel estimation of the illuminant spectrum can be obtained as the weighted average of the illuminant spectrum per region, where the weights are the posterior probabilities of the illuminants.

Therefore, the primary contribution of this work is to provide a means for segmenting a scene into illumination regions. Our method is designed in such a way that any single illuminant colour estimation method can be employed to recover an estimate of the local illumination in each region. This confers several advantages over previous works. Firstly, the number of illuminants and their colours do not need to be pre-determined. This contrasts with other approaches elsewhere such as the work reported in [2]. In fact, one can initialise our algorithm with a sufficiently large number of illuminants and the method will eventually converge to a minimal number of distinctive light sources. Secondly, our method does not require user-intervention. Being completely unsupervised sets it apart from methods such as those in [10], [11]. In addition, unlike the work in [20], the illumination colour can be segmented without sampling image patches. As a result, the illuminant colour is independent of the sampling strategy and the patch size. Furthermore, by estimating the illumination boundaries, we avoid the risk of over-smoothing the illuminant spectra, which is an issue encountered by Ebner’s local-averaging method [15].

The paper is organised as follows. In Section II, we provide the background on the spectral radiance space used throughout the paper. Subsequently, we formulate the illuminant segmentation and estimation problem as an statistical inference one and derive a coordinate-ascent optimisation strategy in Section III. We show results for our method in comparison to a number of alternatives in Section IV. Finally, we

conclude the work presented here in Section V.

II. SPECTRAL RADIANCE SPACE

In this section, we examine the structure of the radiance spectra in a scene illuminated by multiple light sources. The idea is to view the illuminant segmentation problem as a clustering one in the spectral radiance space. To commence, we depart from the image formation process for a single illumination setting. Let us consider a location u in the scene that is illuminated by a spatially uniform power spectrum $L(\lambda)$, where λ is the wavelength variable. The spectral radiance $I(u, \lambda)$ reflected from that location can be expressed as

$$I(u, \lambda) = L(\lambda)S(u, \lambda), \quad (1)$$

where $S(u, \lambda)$ is the reflectance function the at location u and the wavelength λ .

Note that the model in Equation 1 has been used widely in the literature for illumination recovery [23]. Further, we refer to existing research on linear models of real-world reflectance spectra. In an early experiment, Cohen [24] found that the spectral reflectance of 433 chips in the Munsell Book of Color lie in a subspace spanning three basis components. Subsequently, Maloney [25] analysed the fitting of Munsell color samples and naturally occurring reflectance spectra to a linear combination of two to six basis elements. Building on these experimental results, Marimont and Wandell developed a low-dimensional linear model of real-world spectral reflectance [26] where the surface reflectance $S(u, \lambda)$ is expressed as a weighted sum of wavelength-dependent basis functions $S_i(\lambda)$. This model is expressed as

$$S(u, \lambda) = \sum_i w_i(u)S_i(\lambda), \quad (2)$$

where the basis functions $S_i(\lambda)$ can be obtained via Principal Component Analysis (PCA) [27] or Independent Component Analysis (ICA) [28].

Combining Equation 1 and 2, the scene radiance can be written as a linear combination of the basis radiance spectra as follows

$$I(u, \lambda) = \sum_i w_i(u)B_i(\lambda), \quad (3)$$

where $B_i(\lambda) = L(\lambda)S_i(\lambda)$ is the spectral radiance of the i^{th} basis component.

Equation 3 implies that the radiance of all real-world materials under a fixed illuminant spectrum form a linear subspace spanned by the basis vectors $B_i(\cdot)$. This observation serves as the foundation for gamut

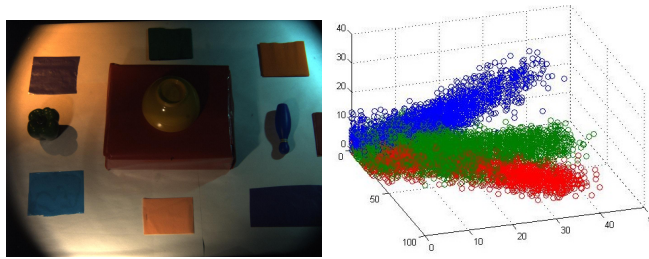


Fig. 1. The colour distribution of a scene illuminated by three different light sources whose hues are blue, orange and yellow. The left panel shows the trichromatic rendering of a sample hyperspectral image as yielded by the standard colour matching functions [30]. The right panel shows three clusters of trichromatic colour vectors, each for an illumination region. These have been plotted in a three dimensional coordinate system where the axes represent the R, G and B values in the range between 0 and 255.

mapping methods under uniform illumination [29]. Furthermore, here we note that each basis vector $B_i(\cdot)$ is dependent on the illuminant spectrum $L(\cdot)$. As a consequence, the subspace spanned by the vectors $B_i(\cdot)$ is skewed by the illuminant spectrum and so is its principal orientation in the spectral radiance space.

The observation above is illustrated in Fig 1. In the left panel, we show a sample hyperspectral image rendered in colour as yielded by the standard colour matching functions proposed by Stiles and Burch [30]. The hyperspectral image has been acquired in the visible wavelength range between 400 and 700 nm at every 10 nm in the spectral domain. To acquire the image, we have used an hyperspectral camera equipped with Liquid Crystal Tunable Filters (LCTFs). We illuminated the input scene using three illuminants with different spectral power distributions corresponding to blue, orange and yellow hues in a spatially overlapping fashion in the scene.

Note that the irradiance in the hyperspectral image is approximately linear with respect to the scene radiance. Further, the simulated colour values are linear combinations of the spectral bands, with weights being the spectral sensitivity of the standard colour matching functions. As a result, the colour image shown in the left panel is linearly related to the scene radiance spectrum, being akin to RAW imagery captured by digital colour cameras without further processing. For the sake of visualisation, we have quantised the image to an 8-bit dynamic range. To enhance the visibility of dark regions, we have used intensity bracketing, therefore inevitably creating saturated pixels.

The right-hand panel shows the clusters of colour vectors at all the image pixels in the RGB colour space, which are labelled in different colours according to the local illuminant. Note that these colour

vectors indeed form clusters in the spectral radiance space with orientations skewed by the respective illuminant. This is the basis observation for our clustering approach, allowing for the separation of pixels illuminated by each light source.

III. ILLUMINANT SEGMENTATION AND ESTIMATION

We now consider the problem of segmenting and estimating the spatially varying illuminant spectra in the scene. We would like to stress that, although the following derivation takes the number of scene illuminants as input, it is, in general, not a requirement for our method. In fact, the number of illuminants can be estimated by a pre-processing step akin to that in [31]. Although the method in [31] specifically deals with a specific form of multi-modal probability distributions, i.e. the mixture of Gaussian distributions, it can be generalised to any multi-modal distribution in a straightforward manner. In relation to our problem setting, this processing step also applies to the non-parametric model of radiance spectrum distribution we describe in Section III-A. In our implementation, the algorithm commences with a sufficiently large number of illuminants. Illuminants found to be similar in terms of their angular difference are then merged and their contributions to each pixel are consolidated at the end of the algorithm.

As before, we assume that the illuminant power spectrum varies smoothly in the spatial domain. This assumption permits the computation of individual illuminant colours by standard algorithms. Such an assumption also gives rise to a sampling approach which yields pixel subsets illuminated by each illuminant.

A. Objective Function

We formulate the target function of the clustering problem as follows. Firstly, we denote the spectral radiance vector at pixel u as $\mathbf{I}(u) \triangleq (I(u, \lambda_1), I(u, \lambda_2), \dots, I(u, \lambda_K))$, where λ_k 's, $k = 1, \dots, K$, are the sampled wavelengths of the image. Here, we consider a scene illuminated by M illuminants L_1, L_2, \dots, L_M and denote α_m as the mixing coefficient of radiance spectrum distribution under the m^{th} illuminant. The mixing coefficients $\alpha_m, m = 1, \dots, M$ indicate the prior probability of the m^{th} illuminant and satisfy the conditions $\sum_{m=1}^M \alpha_m = 1$ and $\alpha_m \geq 0, \forall m = 1, \dots, M$. In addition, we denote the cluster of pixels illuminated by the m^{th} illuminant as Ω_m , where $\Omega_i \cap \Omega_j = \emptyset \forall i \neq j$, and the union of these as $\Omega \triangleq \cup_{m=1}^M \Omega_m$.

We cast our problem as one that finds the clusters $\Omega_1, \Omega_2, \dots, \Omega_M$ that maximise the likelihood of the input image radiance. Here, we formulate the likelihood of the spectral radiance vector $\mathbf{I}(u)$ given the pixel clusters as

$$p(\mathbf{I}(u)|\Omega) = \sum_{m=1}^M \alpha_m p_m(u), \quad (4)$$

where $p_m(u) \triangleq p(\mathbf{I}(u)|\Omega_m)$ is the probability density function (PDF) of the radiance vector $\mathbf{I}(u)$ given the set of pixels Ω_m illuminated by the m^{th} illuminant.

Here, we approximate this PDF using kernel density estimation [22]. To this end, we rely on a kernel function $K_h(\cdot)$ with bandwidth h . This function aims to express the probability density of a radiance spectrum $\mathbf{I}(u)$ in terms of its distances to the radiance spectra of the sample pixels in Ω_m . Formally, the kernel density estimate of PDF $p_m(u)$ is given by

$$p_m(u) \approx \frac{1}{n_m h} \sum_{v \in \Omega_m} K_h(\|\mathbf{I}(u) - \mathbf{I}(v)\|), \quad (5)$$

where $n_m = |\Omega_m|$ is the number of sample pixels illuminated by the m^{th} illuminant and $\|\cdot\|$ is the L^2 -norm of the vector argument.

Combining Equations 4 and 5, we can rewrite the likelihood of the spectral radiance vector $p(\mathbf{I}(u))$ as

$$p(\mathbf{I}(u)|\Omega) \approx \sum_{m=1}^M \frac{\alpha_m}{n_m h} \sum_{v \in \Omega_m} K_h(\|\mathbf{I}(u) - \mathbf{I}(v)\|), \quad (6)$$

where the kernel function penalises large arguments, *i.e.* it attains a peak when the argument value is near zero and vanishes when the latter exceeds the kernel width. Examples of kernel functions include the Gaussian kernel function, which is expressed as

$$K_h^G(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2h^2}\right) \quad (7)$$

and the Epanechnikov kernel [32]

$$K_h^E(z) = \begin{cases} \frac{3}{4} \left(1 - \frac{z^2}{h^2}\right) & \text{if } |z| \leq h \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Further, we enforce an additional constraint that imposes the smooth spatial transition of the illumination in the scene. This is because it is often the case that the illumination changes smoothly rather than arbitrarily across the spatial domain of naturally occurring scenes. This implies that the contribution of the constituting illuminants at each pixel varies smoothly. As a result, we introduce a smoothness constraint

on the posterior probability $p(L_m|\mathbf{I}(u))$ over the spatial domain \mathcal{I} . Such a constraint can be formulated as minimising the following gradient norm of the posterior probability

$$\mathcal{R} = \sum_{u \in \mathcal{I}} \sum_{m=1}^M \left(\frac{\partial p(L_m|u)}{\partial x} \right)^2 + \left(\frac{\partial p(L_m|u)}{\partial y} \right)^2, \quad (9)$$

where x and y are the coordinates of the pixel u .

Equation 9 states the general smoothness constraint on the continuous spatial domain. To express this constraint over the image lattice, we consider a neighbourhood $\mathcal{N}(u)$ around the pixel u . We then discretise the above equation in terms of the difference of the posterior probability $p(L_m|\cdot)$ between the pixel u and its neighbour v , which yields

$$\mathcal{R}^* = \sum_{m=1}^M \sum_{u \in \mathcal{I}} \sum_{v \in \mathcal{N}(u)} [p(L_m|u) - p(L_m|v)]^2, \quad (10)$$

Therefore, the illuminant recovery and segmentation problem is cast as a maximisation of the log likelihood of the image radiance spectra subject to the minimisation of the spatial variation of illuminant contributions. Thus, we formulate the objective function as a combination of the log likelihood, *i.e.* the data error term, and the spatial variation of the posterior probability of illuminants, *i.e.* the regularisation term. Hence, our problem amounts to finding the mixing coefficient α_m and the (maximal) sample set of pixels Ω_m for the m^{th} illuminant so as to maximise the function

$$F = \sum_{u \in \mathcal{I}} \log p(\mathbf{I}(u)|\Omega) - \beta \mathcal{R}^*, \quad (11)$$

where $\beta \geq 0$ is the Lagrangian multiplier of the regulariser \mathcal{R}^* .

In the following section, we present an iterative approach aimed at maximising the cost function in Equation 11. At each iteration, we re-estimate the weight α_m 's and expand the subset of pixels Ω_m so as to maximise the cost function above. We update these two variables by considering the similarity between the pixel radiance spectra in the scene. Meanwhile, the spatial smoothness constraint propagates illuminant labels to neighbouring pixels. Herein, we describe the algorithm in its natural flow, commencing with the initialisation of the pixel sets $\Omega_1, \Omega_2, \dots, \Omega_M$. We then turn our attention to the optimisation process.

B. Initialisation

As mentioned earlier, we optimise the cost function in Equation 11 with respect to the illuminant cluster parameter, using a coordinate ascent approach [33]. Here, we model the distribution of pixel

radiance spectra illuminated by multiple lights as a mixture of several components. In this section, we aim to select the initial sample pixels for the cluster corresponding to each scene illuminant.

The initialisation procedure is derived from the component separation method described by Kannan *et al.* [21]. As before, we make a weak assumption that the radiance spectra at pixels illuminated by the same light follow a log-concave distribution. In fact, log-concave densities cover a wide range of commonly occurring distributions such as Gaussians and the exponential family. Therefore, this assumption does not overly restrict the applicability of our method to real-world scenes.

The algorithm proposed by Kannan *et al.* [21] is based on a theorem that for any mixture of log-concave distributions, the Singular Value Decomposition (SVD) subspace of sample points is close to the means of the samples from the mixture components, where “close” is defined in terms of the sample variances. Further, the method hinges on the intuition that the distance between a point and the mean of its component distribution can only shrink upon projection, thus magnifying the ratio of inter-component to intra-component distances [21].

Following this intuition, the method in [21] employs a random projection of the input points onto a lower dimensional subspace of the original space. At each point (e.g. projected radiance spectrum) in this subspace, we compute the principal variance over its neighbourhood, which is the largest one along all the directions in W . The point with the maximal principal variance is deemed to belong to the sparsest part of its distribution. Using this point as a reference, one can select an initial set of pixels belonging to the same radiance spectrum distribution as that point based on the distance between their radiance spectra and the radiance spectrum of the reference point. Specifically, this criterion involves applying a threshold to the ratio of the radiance spectrum distance to the standard deviation of the neighbourhood in the subspace. In [21], the threshold is related to the number of components and the input sample size.

In Algorithm 1, we summarise the initialisation of our method. Throughout the pseudocode in Algorithm 1 we use the symbol $|\cdot|$ to denote the cardinality of a set, $\|\cdot\|$ is the L^2 -norm, $\Pi_X(\cdot)$ as the projection operator onto a subspace X , $\text{Var}(\cdot)$ is the variance operator and ϵ_1, ϵ_2 and ϵ_3 are constants.

The initialisation takes place over M iterations, in each of which we select a subset of sample pixels Ω_m that are representative of the component distribution corresponding to the m^{th} light. At the beginning of each iteration, we randomly sample a subset of pixels \mathcal{K}_m from the unassigned pixels \mathcal{J}_m in the image, *i.e.* those pixels that have not been sampled at the previous iterations, as shown in Line 3. Subsequently,

Algorithm 1 Initialise the pixel clusters $\Omega_1, \Omega_2, \dots, \Omega_M$.

Require: $\mathbf{I}(u)$: the radiance spectra at each image pixel u .

M : The number of mixture components.

- 1: {Initialise the current set of pixels} $\mathcal{J}_1 \leftarrow \mathcal{I}$
 - 2: **for** $m = 1, \dots, M$ **do**
 - 3: Randomly select $\mathcal{K}_m \subset \mathcal{J}_m$.
 - 4: $W_m \leftarrow \text{span}(\{\mathbf{I}(u) | u \in \mathcal{K}_m\})$.
 - 5: $\forall u \in \mathcal{J}_m \setminus \mathcal{K}_m$, let $\mathbf{Q}(u) \leftarrow \Pi_{W_m}(\mathbf{I}(u))$.
 - 6: **for** $u \in \mathcal{J}_m \setminus \mathcal{K}_m$ **do**
 - 7: Define $\mathcal{T}(u) \leftarrow \{v | \|\mathbf{Q}(u) - \mathbf{Q}(v)\| \leq \epsilon_1\}$ { $\mathcal{T}(u)$ is the neighbourhood of u when projected onto W_m .}
 - 8: $\bar{\mathbf{P}} \leftarrow \frac{1}{|\mathcal{T}(u)|} \sum_{v \in \mathcal{T}(u)} \mathbf{Q}(v)$.
 - 9: $\mathbf{A} \leftarrow [\mathbf{Q}(v_1) - \bar{\mathbf{P}}, \dots, \mathbf{Q}(v_{|\mathcal{T}(u)|}) - \bar{\mathbf{P}}]$ where $v_i \in \mathcal{T}(u)$.
 - 10: $\sigma(u) \leftarrow$ the largest singular value of \mathbf{A} .
 - 11: **end for**
 - 12: Find $u_0 \in \mathcal{J}_m \setminus \mathcal{K}_m$ such that $\sigma(u_0)$ is maximal.
 - 13: Find a direction $V_m \subset W_m$ which maximises $\text{Var}(\Pi_{V_m}(\{\mathbf{Q}(u) | u \in \mathcal{K}_m\}))$.
 - 14: Let U_m be the subspace of W_m orthogonal to V_m .
 - 15: Find the maximal pixel subset $\Omega_m^{(0)} \subset \mathcal{J}_m \setminus \mathcal{K}_m$ such that $\|\mathbf{I}(u) - \mathbf{I}(u_0)\| \leq \epsilon_2$ and $\|\Pi_{U_m}(\bar{\mathbf{I}}(u)) - \Pi_{U_m}(\mathbf{I}(u_0))\| \leq \epsilon_3 \forall u \in \Omega_m^{(0)}$, where $\bar{\mathbf{I}}(u) = \frac{\sum_{u \in \Omega_m} \mathbf{I}(u)}{|\Omega_m|}$.
 - 16: $n_m^{(0)} \leftarrow |\Omega_m^{(0)}|$.
 - 17: $\alpha_m^{(0)} \leftarrow \frac{n_m^{(0)}}{\sum_{m=1}^M n_m^{(0)}}$.
 - 18: Update the remaining set of pixels $\mathcal{J}_{m+1} \leftarrow \mathcal{J}_m \setminus \Omega_m^{(0)}$.
 - 19: **end for**
 - 20: **return** $\Omega_m^{(0)}, \forall m = 1, \dots, M$.
-

we obtain the projection $\mathbf{Q}(u)$ onto the M -dimensional subspace W_m spanned by the radiance spectra at the pixels in \mathcal{K}_m as shown in Line 5. At this stage, the random subset \mathcal{K}_m used to obtain the low-dimensional subspace is discarded from \mathcal{J}_m to ensure that the samples selected for each component are independent from W_m . Later, as shown between Lines 7 and 10, we find the principal variance $\sigma(u)$ of

the projected neighbours of $\mathbf{Q}(u)$ in the subspace W_m . This is achieved by first constructing a matrix \mathbf{A} whose columns are the mean-centred projection of the neighbours of $\mathbf{Q}(u)$ in W_m . The principal variance $\sigma(u)$ is found to be the largest singular value of \mathbf{A} .

In Line 12, the algorithm aims to find a pixel $u_0 \in \mathcal{J}_m \setminus \mathcal{K}_m$ with the maximal principal variance $\sigma(u_0)$. Thereafter, in Line 13, we find a direction $V_m \subset W_m$ onto which the projection of \mathcal{K}_m has maximal variance. In Line 14, we find a subspace $U_m \subset W_m$ that is orthogonal to the direction V_m . At the end of each iteration, we choose the initial sample pixels $\Omega_m^{(0)}$ under the m^{th} illuminant as those close to pixel u_0 in both the original radiance space and the projection subspace U_m , as shown in Line 15.

C. Optimisation

We summarise all the steps of our optimisation approach in Algorithm 2. Note that our optimisation outputs the weight α_m , the maximal pixel subsets Ω_m for each of the illuminants in the scene and the contribution of each light L_m at each pixel u , *i.e.* $p(L_m|u)$. The iterations of our algorithm terminate when the estimated mixing coefficients and the subsets of sample pixels $\Omega_m, m = 1, \dots, M$ converge, *i.e.* no further updates are made.

To recover the mixed illumination at each pixel, we first compute the average illuminant power spectrum L_m over the subset of pixels Ω_m . We obtain this estimate by applying an illuminant colour estimation method, such as the Grey-World [7], Grey-Edge [9] and White-Patch algorithm [6] to this subset of pixels. With the average illuminant power spectrum for each region, we can then re-estimate the per-pixel illumination using the probabilities $p(L_m|u)$, as follows

$$L(u) = \sum_{m=1}^M p(L_m|u)L_m. \quad (12)$$

For our optimisation, we depart from the initial set of pixels and maximise the cost function in Equation 11. Using the approximation in Equation 6, we can rewrite this cost function as

$$F \approx \sum_{u \in \mathcal{I}} \log \left(\sum_{m=1}^M \frac{\alpha_m}{n_m} \sum_{v \in \Omega_m} K_h(\|\mathbf{I}(u) - \mathbf{I}(v)\|) \right) - \beta \mathcal{R}^* - |\mathcal{I}| \log h, \quad (13)$$

where $|\mathcal{I}|$ is the total number of image pixels.

Since the last term on the right hand side of Equation 13 is a constant, we can simply remove it from further consideration. We also note that the current setting assumes an implicit relationship between the

Algorithm 2 Compute the mixture coefficient α , the posterior probability $p(L_m|u)$ and the pixel subset Ω_m .

Require: $\mathbf{I}(u)$: the radiance spectra at each image pixel u .

M : The number of mixture components.

- 1: Initialise $\alpha_m^{(0)}, \Omega_m^{(0)}, m = 1, \dots, M$ as in Algorithm 1.
 - 2: **repeat**
 - 3: Update $p^{(t+1)}(L_m|u)$ by solving Equation 16.
 - 4: $\alpha_m^{(t+1)} = \frac{\sum_{u \in \mathcal{I}} p^{(t+1)}(L_m|u)}{|\mathcal{I}|}$ (Equation 18).
 - 5: $\forall u \in \mathcal{I}$, define $l(u) = \operatorname{argmax}_{1 \leq m \leq M} p^{(t+1)}(L_m|u)$.
 - 6: **for** $m = 1, \dots, M$ **do**
 - 7: Define $\Upsilon_m = \{u \in \mathcal{I} | l(u) = m \text{ and } p^{(t+1)}(L_m|u) > t_l\}$.
 - 8: $\forall u \in \Upsilon_m$, let $\mathcal{S}^r(u)$ be the set of r pixels with their radiance spectra nearest to $\mathbf{I}(u)$.
 - 9: $\Omega_m^{(t+1)} \leftarrow \Omega_m^{(t)} \cup (\cup_{u \in \Upsilon_m} \mathcal{S}^r(u))$.
 - 10: **end for**
 - 11: $t \leftarrow t + 1$.
 - 12: **until** $\alpha_m^{(t)}, \Omega_m^{(t)}, p^{(t)}(L_m|u)$ are close to $\alpha_m^{(t-1)}, \Omega_m^{(t-1)}, p^{(t-1)}(L_m|u)$ within a tolerance.
 - 13: **return** $\alpha_m^{(t)}, \Omega_m^{(t)}, p^{(t)}(L_m|u), \forall m = 1, \dots, M$.
-

posterior probability $p(L_m|u)$ and the likelihood of the image radiance $p_m(u)$ by Bayes rule as follows

$$p(L_m|u) = \frac{\alpha_m p_m(u)}{\sum_{q=1}^M \alpha_q p_q(u)}. \quad (14)$$

As a result, we can constrain the optimisation using Equation 14. To this end, we adopt a coordinate-ascent optimisation approach by designing an iterative algorithm consisting of two interleaving steps per iteration. The first step re-estimates the posterior probability distribution $p(L_m|\mathbf{I}(u))$, whereas the second one estimates the mixing coefficient α_m and expands the set Ω_m for each illuminant once the posterior probability distribution is at hand. In the following subsections, we index these variables to the iteration number t and describe the optimisation steps in detail.

1) *Updating the posterior probabilities $p(L_m|u)$:* With the initial mixture coefficients and sample pixels under each illuminant, we proceed to estimate the posterior probability. Let us use the shorthand $p^{(t)}(L_m|u)$ to denote the estimate of $p(L_m|u)$ given the values of $\alpha^{(t)}$ and the sample pixels set $\Omega^{(t)}$ at the t^{th} iteration. Also, let $\bar{p}^{(t)}(L_m|u)$ denote the Bayesian estimate of $p(L_m|u)$ given by Equation 14 with respect to $\alpha^{(t)}$

and $p_m^{(t)}(u) \triangleq p(\mathbf{I}(u)|\Omega_m^{(t)})$.

Since only the second term, *i.e.* the spatial smoothness term, of the cost function in Equation 13 is dependent on $p(L_m|u)$, this optimisation step can be viewed as that aiming to find a posterior probability which minimises

$$F_1 = \beta \sum_{m=1}^M \sum_{u \in \mathcal{I}} \sum_{w \in \mathcal{N}(u)} [p(L_m|u) - p(L_m|w)]^2 + \gamma \sum_{m=1}^M \sum_{u \in \mathcal{I}} (p(L_m|u) - \bar{p}^{(t)}(L_m|u))^2, \quad (15)$$

where $\gamma \geq 0$ is the weight of the Bayesian constraint.

Setting the derivative of F_1 with respect to $p(L_m|u)$ to zero, we obtain

$$(\beta|N(u)| + \gamma)p(L_m|u) - \beta \sum_{w \in \mathcal{N}(u)} p(L_m|w) = \gamma \bar{p}^{(t)}(L_m|u) \quad (16)$$

Note that, following Equation 16, there exists a linear system of equations with respect to the variables $\{p(L_m|u)|u \in \mathcal{I}\}$ for each illuminant m . To solve such a system under the constraints that $0 \leq p(L_m|u) \leq 1$ and $\sum_{m=1}^M p(L_m|u) = 1$, we use the standard linear solver technique in [34].

2) *Updating the mixing coefficients α_m* : Having estimated the posterior probability, we now estimate the mixture coefficients α_m . We note that only the first term of the function in Equation 13, which we denote τ_1 , is dependent on this subset of variables. For the current estimate of the posterior distribution $p^{(t+1)}(L_m|u)$, the expected value of this term is

$$\mathbb{E}(\tau_1) = \sum_{u \in \mathcal{I}} \sum_{m=1}^M \log \alpha_m p^{(t+1)}(L_m|u) + \sum_{u \in \mathcal{I}} \sum_{m=1}^M \log \left(\frac{\sum_{v \in \Omega_m} K_h(\|\mathbf{I}(u) - \mathbf{I}(v)\|)}{n_m} \right) p^{(t+1)}(L_m|u). \quad (17)$$

Thus, we maximise the function above with respect to α under the constraint $\sum_{m=1}^M \alpha_m = 1$. By setting the derivative of the equation above with respect to α to zero, we obtain

$$\alpha_m^{(t+1)} = \frac{\sum_{u \in \mathcal{I}} p^{(t+1)}(L_m|u)}{|\mathcal{I}|} \quad (18)$$

3) *Expanding the subsets of pixels Ω_m* : With the posterior probabilities $p(L_m|u)$ and the mixture coefficients α_m in hand, we turn our attention to expanding the subset of sample pixels Ω_m for each illuminant L_m . This implies that, for each m , we find a subset Ω_m that maximises the weighted sum $\sum_{u \in \mathcal{I}} \log \left(\frac{\sum_{v \in \Omega_m} K_h(\|\mathbf{I}(u) - \mathbf{I}(v)\|)}{n_m} \right) p^{(t+1)}(L_m|u)$, where the weights $p^{(t+1)}(L_m|u)$ are obtained in the previous step.

Before proceeding further, we define the following notation. For each pixel u in the image, we define its dominant light $l(u)$ to be the one with the largest posterior probability, *i.e.* such that $l(u) = \operatorname{argmax}_{1 \leq m \leq M} p(L_m|u)$. Also, we denote Υ_m as the subset of pixels having the m^{th} illuminant as their

dominant illuminant, *i.e.* $\Upsilon_m = \{u \in \mathcal{I} | l(u) = m\}$ such that $p(L_m | u) > t_l$, where t_l is a preset threshold close to 1.

To expand Ω_m we adopt a greedy approach. We note that the kernel value $K_h(\|\mathbf{I}(u) - \mathbf{I}(v)\|)$ is large when the vectors $\mathbf{I}(u)$ and $\mathbf{I}(v)$ are close. For each pixel $u \in \Upsilon_m$, to maximise $K_h(\|\mathbf{I}(u) - \mathbf{I}(v)\|)$, we choose a subset $\mathcal{S}^r(u)$ of r -nearest neighbours of u in the spectral radiance space, where r is a preset constant. The union of the subsets $\mathcal{S}^r(u)$, where $u \in \Upsilon_m$, are then augmented to the current subset of pixels under the m^{th} illuminant. In other words, the sample pixels under the m^{th} illuminant is updated using the rule $\Omega_m^{(t+1)} \leftarrow \Omega_m^{(t)} \cup (\cup_{u \in \Upsilon_m} \mathcal{S}^r(u))$.

IV. EXPERIMENTS

In this section, we illustrate the utility of our method for the segmentation and estimation of non-uniform illumination from a single image. In our experiments, we compare our method to a number of alternatives. These are the method of Gijsenij *et al.* [20], the Retinex algorithm [35] and the approach proposed by Ebner [15]. For our method and Gijsenij *et al.*'s, we apply a variety of existing methods, such as the Grey-World [7], White-Patch [6] and Grey-Edge [9] to recover the illuminant colour in each illuminant segment. The comparison presented here involves several variants of the above methods, including the general Grey-World (general GW) [20] with L^8 -norm, the first-order Grey-Edge (1^{st} GE) and 2^{nd} -order Grey-Edge (2^{nd} GE). Regarding Gijsenij *et al.*'s method, we employ its highest performing variant using a grid-based sampling strategy with a patch size of 10×10 , as mentioned in [20].

For purposes of comparing the illumination segmentation results, we have employed k -means to segment the illumination images produced by Ebner's [15], Retinex [35] and Gijsenij *et al.*'s method [20]. Note that the Gijsenij *et al.*'s method [20] assumes the correct number of illuminants in the scene. This contrasts with our method, which does not require the number of illuminants to be supplied a priori. Rather, the initialisation step in Algorithm 1 determines the number of illuminants automatically. In addition, for our method, we have employed a Gaussian kernel to estimate the probability density function of the radiance vector $\mathbf{I}(u)$ according to Equations 4 and 5. Here we set the kernel width h to be the standard deviation of all the pair-wise distances between the radiance spectra in the input scene.

It is noted that several operations throughout the digital signal processing of a camera pipeline, such as non-linear enhancement of chroma and hue, optical glare and spatial variation of camera responses, may interfere with the linearity of data. However, for photometric evaluation tasks, the image intensity is

Our method	77.09 \pm 19.10				
Ebner's [15]	43.02 \pm 9.76				
Retinex [35]	44.76 \pm 12.78				
	Grey World	White Patch	General GW	1 st GE	2 nd GE
Gijssenij <i>et al's</i> [20]	75.19 \pm 22.05	76.81 \pm 19.74	74.35 \pm 21.02	73.78 \pm 19.73	74.00 \pm 20.20

TABLE I

THE SEGMENTATION ACCURACY ϵ (IN %) FOR THE HYPERSPECTRAL IMAGE DATASET *Foster-spectral*.

Our method	65.77 \pm 16.12				
Ebner's [15]	47.07 \pm 7.00				
Retinex [35]	48.21 \pm 6.73				
	Grey World	White Patch	General GW	1 st GE	2 nd GE
Gijssenij <i>et al.'s</i> [20]	70.97 \pm 20.97	71.20 \pm 19.81	70.89 \pm 20.34	67.95 \pm 19.75	69.24 \pm 19.91

TABLE II

THE SEGMENTATION ACCURACY ϵ (IN %) FOR THE PSEUDOCOLOUR IMAGE DATASET *Foster-colour*, WHICH WE RENDER USING THE DATASET COLLECTED BY FOSTER *et al.* [36].

required to be directly proportional to the scene radiance. This implies that linearity between the image irradiance and the scene radiance should be preserved throughout the image acquisition process [37]. This is important since the effectiveness of illuminant colour estimation algorithms cannot be quantified reliably on non-linear imagery due to the effect of dynamic range compression [38]. As a result, we have opted for the following databases in which the acquired image irradiance is approximately linear with respect to the scene radiance. To assure the linearity property, we have relied on the fact that these datasets are either generated from spectral data including reflectance [36] and illumination spectra [39] or captured in the RAW data format [20], [4].

Our method	77.69 \pm 19.41				
Ebner's [15]	61.24 \pm 18.00				
Retinex [35]	72.16 \pm 13.33				
	Grey World	White Patch	General GW	1 st GE	2 nd GE
Gijssenij <i>et al's</i> [20]	71.13 \pm 25.79	71.71 \pm 21.09	69.47 \pm 24.93	61.08 \pm 21.52	63.99 \pm 23.75

TABLE III

THE SEGMENTATION ACCURACY ϵ (IN %) FOR THE NATURAL IMAGES CAPTURED BY GIJSENIJ *et al.* [20].

The first of these was synthetically generated from the reflectance images of eight rural and urban scenes captured by Foster *et al.* [36]. These images were captured with a Peltier-cooled digital camera and a fast VariSpec tunable liquid-crystal filter producing a spatial resolution of 1344×1024 pixels and a spectral resolution of 10 nm over the wavelength range of 400 - 720 nm. The intensity response was recorded with 12-bit precision and was linear to the scene radiance. The maximum pixel intensity was kept within 86 - 90% of saturation. To compute image reflectance, the illumination spectrum was obtained from a grey Munsell surface placed in the scene.

To simulate the spatial variation of illumination, we divided each image in the above database into quadrants and apply the illumination spectrum of a different light to each region. These illumination power spectra were randomly selected from a set of 81 illuminant spectra collected by Barnard *et al.* [39]. This procedure aims to simulate two or three different illuminant spectra across the scene in a similar fashion to the method reported in [20]. We also generate pseudocolour imagery from these images making use of the colour matching functions in [40]. Note that the application of the colour matching functions is a linear operation whereby the resulting colour image is equivalent to a RAW image captured by digital colour cameras without further processing. We denote Foster *et al.*'s hyperspectral dataset *Foster-spectral* and the corresponding pseudocolour version *Foster-colour*.

In addition, we employ two other datasets of colour images illuminated by multiple illuminants. One of them consists of the natural outdoor scenes captured by Gijsenij *et al.* [20]. We denote this dataset herein *Gijsenij*. The ground truth colour of the light sources in this dataset were obtained from several grey balls placed in the scene, with locations manually annotated. Using a Sigma SD10 camera with a Foveon X3 sensor, the authors acquired the images at a spatial resolution of roughly 384×256 . The camera white balancing mode was kept fixed at the overcast setting. The images are captured in the RAW format and converted to linear sRGB for experimentation purposes [20].

The last dataset under consideration is that acquired by Bleier *et al.* [4]¹. Similarly, we have denoted this dataset as *Bleier*. The dataset contains 408 images in the RAW format of four scenes acquired under two or three illuminants. It also includes ground truth illumination colour images acquired using grey spray-painted scenes. The data was captured with a Canon EOS 550D camera and a Sigma 17-70mm lens. The image intensity value is linear as the RAW data from the camera was converted using ddraw

¹The dataset is accessible at <http://www5.informatik.uni-erlangen.de/research/data/multi-illuminant-dataset/>

with gamma set to 1.0 and without any white balancing. To avoid under-exposure and over-exposure artifacts, different fixed shutter speeds were used across the 17 different illumination conditions.

We note that even with RAW images converted by dcrw, the quantisation of digital data introduces errors that may render similar hues indistinguishable. Therefore, the ability to distinguish objects with very similar hues using digital image data is limited by the level of quantisation induced when digitising cameras intensity response to the scene irradiance.

In the qualitative results presented herein, we show imagery that has been gamma corrected only to enhance the visibility of the dark regions. Other than that, the input imagery used for the quantitative evaluation of our method and the alternatives has not been modified from the original, with special care taken to preserve the linearity of the data.

A. Illumination Region Segmentation

We first examine the accuracy of the segmented illuminant regions resulting from the above methods. To provide a quantitative evaluation of the segmentation results, we adopt an accuracy measure based on the overlap between the ground-truth and the estimated region for each illuminant. This measure has been employed elsewhere to quantify the segmentation accuracy of the PASCAL Visual Object Classes (VOC) Challenge [41]. Let A_m^{gt} and A_m^{est} be respectively the ground-truth and the estimated region of the m^{th} illuminant. The segmentation error is then formulated as

$$\epsilon = \frac{1}{M} \sum_{m=1}^M \frac{A_m^{gt} \cap A_m^{est}}{A_m^{gt} \cup A_m^{est}}, \quad (19)$$

where $A_m^{gt} \cap A_m^{est}$ denotes the intersection of the estimated and ground truth regions of the m^{th} light and $A_m^{gt} \cup A_m^{est}$ their union.

In Figure 2, we provide a qualitative demonstration of the illuminant segmentation results. The figure includes sample illuminant segmentation maps for four scenes, one from each dataset. The first column shows the input image while the remaining ones show the illuminant regions recovered by the different methods in separate columns. In the figure, each colour in the segmentation maps corresponds to a distinct illuminant label. In the last column, we show the results produced by the variant of Gijsenij *et al.*'s method [20] that employs the Grey-World assumption for local illumination estimation. This is the variant implemented by Gijsenij *et al.* [20] that generally yields the highest performance. As observed from this figure, our method delivers more accurate segmentation maps than the others. While Ebner's and

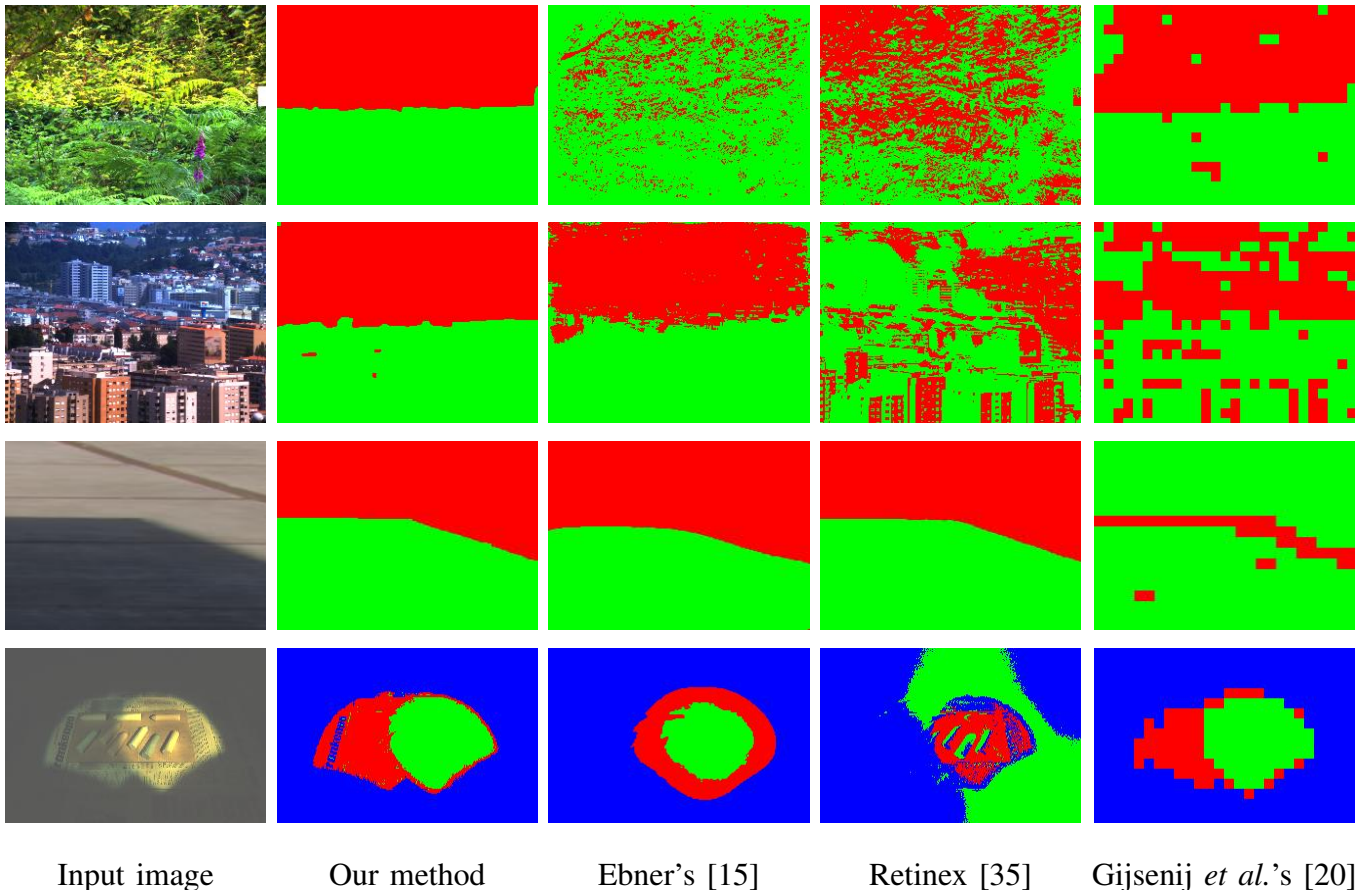


Fig. 2. Illuminant regions segmented by our method and a number of alternatives. In each row (from top to bottom), we show results for a sample image from each of our datasets, *i.e.* *Foster-spectral*, *Foster-colour*, *Gijssenij* and *Bleier*. First column: Input image; Second to last columns: Illuminant segments yielded by our method, Ebner's [15], Retinex method [35] and Gijssenij *et al.*'s [20], respectively.

the Retinex algorithms often result in small discontinuous fragments, Gijssenij *et al.*'s method [20] yields patchy and isolated regions due to the nature of its grid-based sampling strategy. These observations are consistent across our experiments, as shown in the following quantitative evaluation.

In Table I, we show the segmentation accuracy as defined in Equation 19 for the spectral images in the dataset *Foster-spectral*. Similarly, Table II shows the segmentation results for the pseudocolour version of these images, *i.e.* the *Foster-colour* dataset. For the former dataset, our method slightly outperforms the variants of Gijssenij *et al.*'s method [20] and greatly outperforms the others. In addition, both of these methods outperform the remaining alternatives for the above two datasets. In addition, we show the illumination segmentation results for the natural images captured by Gijssenij *et al.* [20] in Table III. For this dataset, our approach outperforms all the alternatives by at least 5.5%. The numerical results in the above tables demonstrate the effectiveness of our method for the task of illumination segmentation as compared to the alternatives.

	Grey-World	White-Patch	General GW	1 st GE	2 nd GE
Our method	22.88 ± 11.86(25.03)	19.60 ± 7.05(19.24)	20.64 ± 9.18(20.16)	20.68 ± 9.30(20.59)	19.84 ± 9.09(20.35)
Gijssenij <i>et al.</i> 's [20]	23.18 ± 11.67(25.87)	19.57 ± 9.87(21.31)	22.05 ± 11.22(23.99)	20.78 ± 8.94(21.81)	20.23 ± 8.79(21.23)
Ebner's [15]	35.96 ± 12.04(38.04)				
Retinex [35]	27.11 ± 9.96(28.45)				

TABLE IV

ANGULAR ERRORS (IN DEGREES) FOR THE ILLUMINANT SPECTRUM ESTIMATED FROM THE HYPERSPECTRAL IMAGE DATASET *Foster-spectral* [36]. HERE, WE REPORT THE MEANS AND THE STANDARD DEVIATIONS OF THE ANGULAR ERROR WITH THE MEDIAN ERROR IN PARENTHESES.

	Grey-World	White-Patch	General GW	1 st GE	2 nd GE
Our method	15.39 ± 8.66(14.76)	14.05 ± 4.67(13.60)	15.47 ± 5.48(15.28)	16.94 ± 6.31(16.26)	16.53 ± 6.02(15.78)
Gijssenij <i>et al.</i> 's [20]	14.49 ± 8.87(12.69)	14.48 ± 7.25(12.59)	14.81 ± 8.18(13.07)	17.07 ± 6.07(16.58)	16.57 ± 5.91(15.77)
Ebner's [15]	24.42 ± 13.02(21.46)				
Retinex [35]	19.58 ± 9.15(18.44)				

TABLE V

ANGULAR ERRORS (IN DEGREES) FOR THE ILLUMINANT COLOUR IN THE PSEUDOCOLOUR IMAGE DATASET *Foster-colour*, WHICH WAS SYNTHESIZED FROM THE DATASET COLLECTED BY FOSTER *et al.* [36]. HERE, WE REPORT THE MEANS AND THE STANDARD DEVIATIONS OF THE ANGULAR ERROR WITH THE MEDIAN ERROR IN PARENTHESES.

Note that, despite the spatial smoothness assumption on the illumination, our algorithm can still handle images with sharp illumination edge or shadow edges. In such situations, we can set a low value of β for the regulariser \mathcal{R}^* in Equation 11 so that the separation of illuminant segments is governed mainly by the illuminant likelihood term, i.e. the second term on the left-hand side of the equation. In fact, we have set $\beta = 5$ for all the datasets under study. This approach has been shown to be effective on images with large variations across edges, especially those in the *Gijssenij* and the *Bleier* datasets. The relevant segmentation results can be observed in Figure 2 and in Table III.

B. Illumination Spectrum Estimation

Now, we turn our attention to the efficacy of our method in estimating the illuminants in the scene. As a standard practice, we measure the error of the illuminant estimate instead of the reflectance estimate. This error measure also directly pertains to our aim of computation. The accuracy of the estimated illuminants for each image is computed as the mean of the Euclidean angular deviation of the estimated illuminant spectrum or colour vector from the respective ground-truth over the image pixels. We opt for

	Grey-World	White-Patch	General GW	1 st GE	2 nd GE
Our method	3.89 ± 1.43(3.46)	3.20 ± 1.54(2.97)	3.72 ± 1.49(2.86)	3.26 ± 0.86(3.25)	3.25 ± 1.01(3.27)
Gijssenij <i>et al.</i> 's [20]	7.06 ± 2.40(6.33)	6.26 ± 2.63(6.27)	7.03 ± 2.53(6.26)	7.17 ± 2.46(6.52)	6.81 ± 2.49(5.98)
Ebner's [15]	6.13 ± 2.21(5.82)				
Retinex [35]	6.68 ± 2.26(5.47)				

TABLE VI

ANGULAR ERRORS (IN DEGREES) FOR THE ILLUMINANT COLOUR ESTIMATED FROM THE NATURAL IMAGES ACQUIRED BY GIJSENIJ *et al.* [20]. HERE, WE REPORT THE MEANS AND THE STANDARD DEVIATIONS OF THE ANGULAR ERROR WITH THE MEDIAN ERROR IN PARENTHESES.

	Grey-World	White-Patch	General GW	1 st GE	2 nd GE
Our method	1.16 ± 0.49(1.18)	1.86 ± 1.09(1.77)	1.59 ± 0.77(1.56)	3.32 ± 0.96(3.39)	3.23 ± 1.00(3.32)
Gijssenij <i>et al.</i> 's [20]	4.93 ± 1.30(4.71)	6.04 ± 1.73(6.12)	5.15 ± 1.50(4.78)	14.52 ± 2.72(14.89)	13.82 ± 2.62(13.89)
Ebner's [15]	4.76 ± 1.33(4.57)				
Retinex [35]	5.18 ± 5.50(4.73)				

TABLE VII

ANGULAR ERRORS (IN DEGREES) FOR THE ILLUMINANT COLOUR ESTIMATED FROM THE DATASET REPORTED IN BLEIER *et al.* [4]. HERE, WE REPORT THE MEANS AND THE STANDARD DEVIATIONS OF THE ANGULAR ERROR WITH THE MEDIAN ERROR IN PARENTHESES.

this error measure as it is widely used in previous works [4], [9], [19], [20], [29], [38]. Note that the other error measure used elsewhere is the RMS chromaticity difference [18], [37]. It is worth noting, however, that the mean Euclidean angular deviation and the RMS chromaticity difference are correlated when the chromaticity vector is normalised to unit L^2 -norm.

In Tables IV and V, we report the median angular error for the illuminant spectrum estimated from the hyperspectral image dataset *Foster-spectral* and for the illuminant colour estimated from its pseudocolour version *Foster-colour*, respectively. Similarly, we report the angular error for the *Gijssenij* dataset in Table VI and those for the *Bleier* dataset in Table VII. In addition to the median error (in parentheses), we have reported the means and standard deviations in the above tables to provide more information on the error distribution.

From the tables, we can conclude that, on the *Foster-spectral* and *Foster-colour* datasets, all the variants of our method are comparable to Gijssenij *et al.*'s method and outperform the others by at least three degrees. The variants of our method and Gijssenij *et al.*'s method that employ White-Patch to estimate the illumination achieve the highest accuracy. This is due to the presence of materials with grey or white

reflectance in the scene. Moreover, the estimation accuracy of our method does not appear to vary greatly with respect to the single illuminant colour estimation method employed. However, this is not the case for Gijsenij *et al.*'s method.

On the *Gijsenij* and the *Bleier* datasets, all the variants of our method significantly outperform the remaining ones, achieving a much lower estimation error than those produced by the alternatives. This good performance is justified by the high accuracy of our method in locating illuminant regions. Gijsenij *et al.*'s method [20] does not perform as well as ours due to its assumption of constant illuminant on every sample patch. Ebner's local averaging method does not seem to perform well on these datasets since each image only consists of a small number of materials, which does not satisfy the Grey-World assumption. Lastly, the Retinex implementation in [35] relies on an initialisation based on the White-Patch assumption, which does not work for these images since they lack white materials. From the results in Tables IV–VII, we can observe a higher level of accuracy in the illumination spectra and colours estimated by our method compared to the alternatives.

It is noticeable that the angular errors in Tables IV–VII vary greatly between the datasets, from 1.16 degrees for the *Bleier* dataset to 22.88 degrees for the *Foster-spectral* dataset. This great variation can be explained by the nature of the ground truth illumination acquired for each dataset. For the dataset of natural images acquired by Foster *et al.* [36], a single illuminant spectrum was measured for the entire scene using a flat grey (Munsell N5 or N7) surface. Therefore, even after normalising the image irradiance against the measured illuminant spectrum, the effect of spatially varying illumination and indirect illumination can still be observed in the resulting image [36]. These illumination components are the confounding factors that cause inaccuracies in the illuminant estimation at each local image region even if the illuminants have been segmented correctly as shown in Figure 2. The angular error for the *Foster-colour* dataset is lower than those for the *Foster-spectral* one because the integration of spectral bands for colour rendering mitigates the error.

On the other hand, the illuminant estimates are more accurate for the *Gijsenij* and the *Bleier* datasets. Specifically, Gijsenij *et al.* [20] placed several grey balls under various illumination regions in the scene. Hence, a single illuminant colour can be estimated for each local region using the grey-World algorithm. Bleier *et al.* [4] took a more meticulous approach to measure the illuminant colour at every scene point. To do this, they painted the input scenes in grey and captured the illumination image by illuminating the

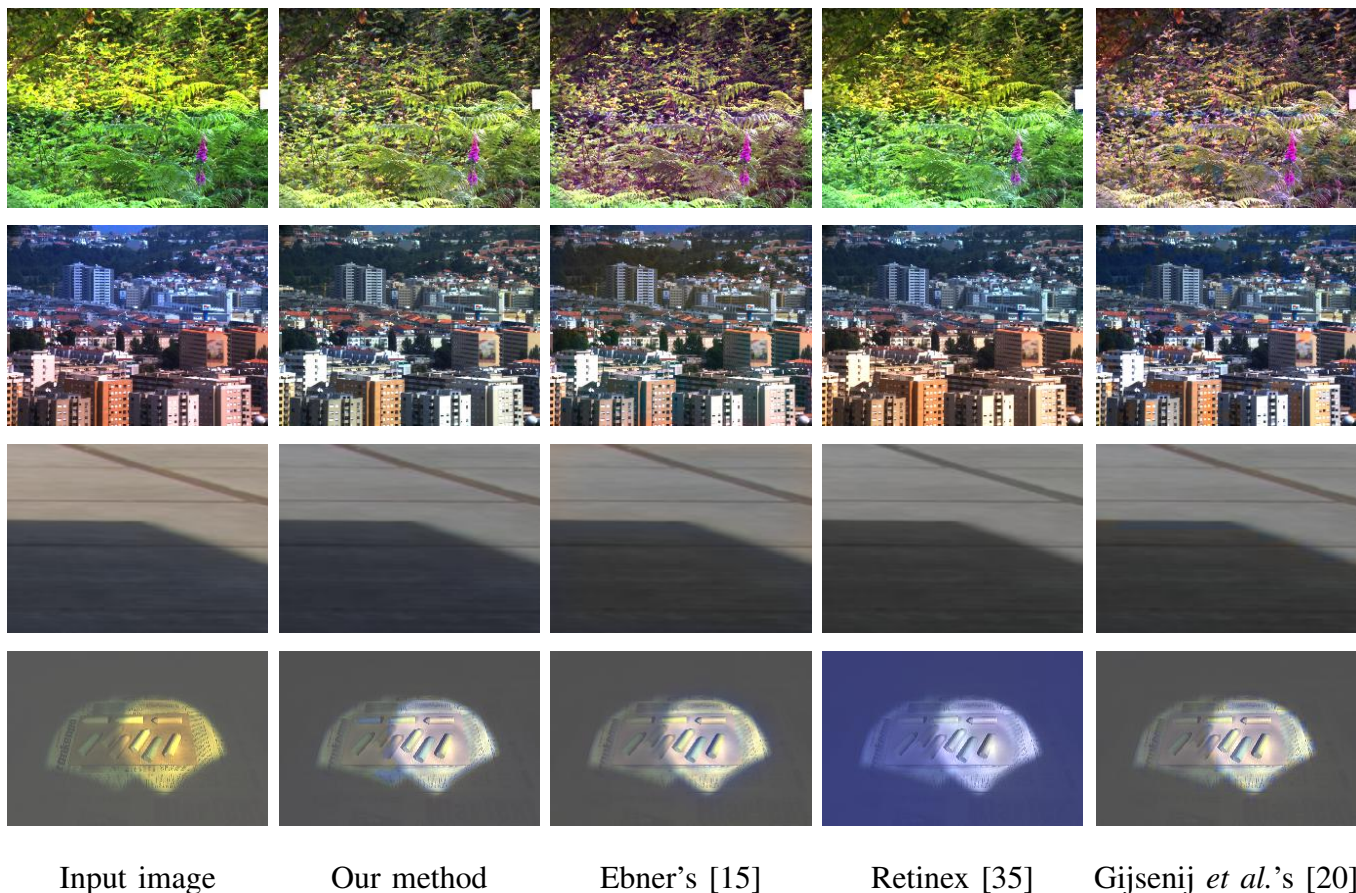


Fig. 3. Colour correction with respect to a white light source. In each row (from top to bottom), we show results for a sample image from each of the following datasets, *i.e.* *Foster-spectral*, *Foster-colour*, *Gijssenij* and *Bleier*. First column: Input images. Remaining columns, left-to-right: Colour corrected images using the illuminant recovered by the Grey-World variant of our method, Ebner's [15], Retinex method [35] and the Grey-World variant of Gijssenij *et al.*'s [20].

grey-painted scenes under the same illumination condition as the input colour image.

Finally, we show colour correction results based on the spatially varying illuminant colour estimated by each of the methods under consideration. To perform colour correction on the original images, we have substituted the illuminant per pixel with a single white light as the reference illuminant. In Figure 3, we show the colour-balanced output for the methods above in separate columns. Note that the colour-corrected images using the illuminant colour estimated by our method exhibit a good level of chromaticity consistency over the scene, especially within the same material regions across illuminant boundaries. Furthermore, our method yields the most neutral colour corrected imagery. In contrast, the alternatives often exhibit colour shades strongly influenced by the original illuminant and colour bleeding along illuminant boundaries.

V. CONCLUSIONS

In this paper, we have presented an algorithm to segment the illuminant regions and to estimate the illumination power spectrum in a single image of a scene lit by multiple light sources. We exploit the probability distribution of the spectral radiance spectra at pixels illuminated by the same illuminant and cast the problem as an estimation of mixture components in the spectral radiance space. Based on this approach, the problem becomes that of maximising the likelihood of the image pixels being illuminated by the scene illuminants while enforcing the smooth variation of the light contribution over the spatial domain. To this end, we apply a kernel density estimator to a subset of pixels illuminated by each illuminant. We tackle the illumination segmentation problem using a coordinate ascent scheme, which involves update operations on the mixture coefficients, the sample pixel subset under each illuminant and the posterior probability of each light per pixel. Once the pixel clusters are in hand, we estimate the illuminant spectrum per region using standard illuminant colour estimation methods for single illuminants, and re-estimate the illumination per pixel with the posterior probability of each light. We have also shown results on a number of image datasets and compared our method to alternatives elsewhere in the literature. Our experiments show the effectiveness of our method for illumination region segmentation, illumination colour estimation and colour correction.

ACKNOWLEDGMENT

NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

REFERENCES

- [1] T. Zickler, S. P. Mallick, D. J. Kriegman, and P. N. Belhumeur, "Color subspaces as photometric invariants," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 13–30, 2008.
- [2] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, "Light mixture estimation for spatially varying white balance," ser. SIGGRAPH '08, 2008, pp. 70:1–70:7.
- [3] J. Lopez-Moreno, S. Hadap, E. Reinhard, and D. Gutierrez, "Compositing images through light source detection," *Computer & Graphics*, vol. 34, no. 6, pp. 698–707, Dec. 2010.
- [4] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Tröger, and A. Kaup, "Color constancy and non-uniform illumination: Can existing algorithms work?" in *ICCV Workshops*, 2011, pp. 774–781.

- [5] C. P. Huynh and A. Robles-Kelly, "A solution of the dichromatic model for multispectral photometric invariance," *International Journal of Computer Vision*, vol. 90, no. 1, pp. 1–27, 2010.
- [6] E. Land and J. McCann, "Lightness and retinex theory," *Journal of Optical Society America*, vol. 61, no. 1, pp. 1–11, 1971.
- [7] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 337–350, 1980.
- [8] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Color Imaging Conference*, 2004, pp. 37–41.
- [9] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2207–2214, 2007.
- [10] A. Bousseau, S. Paris, and F. Durand, "User-assisted intrinsic images," in *ACM SIGGRAPH Asia*, 2009, pp. 130:1–130:10.
- [11] I. Boyadzhiev, K. Bala, S. Paris, and F. Durand, "User-guided white balance for mixed lighting conditions," *ACM Transactions on Graphics*, vol. 31, no. 6, pp. 200:1–200:10, Nov 2012.
- [12] M. Ebner, *Color constancy*, ser. Imaging Science and Technology. Wiley- IS&T, 2007.
- [13] G. D. Finlayson, B. V. Funt, and K. Barnard, "Color constancy under varying illumination," in *ICCV*, 1995, pp. 720–725.
- [14] E. H. Land, "Recent advances in retinex theory," *Vision Research*, vol. 26, no. 1, pp. 7–21, 1986.
- [15] M. Ebner, "Color constancy using local color shifts," in *European Conference on Computer Vision*, 2004, pp. 276–287.
- [16] Y. Wang and D. Samaras, "Estimation of multiple illuminants from a single image of arbitrary known geometry," in *European Conference on Computer Vision*, 2002, pp. 272–288.
- [17] G. D. Finlayson, C. Fredembach, and M. S. Drew, "Detecting illumination in images," in *IEEE International Conference on Computer Vision*, 2007.
- [18] K. Barnard, G. D. Finlayson, and B. V. Funt, "Color constancy for scenes with varying illumination," *Computer Vision and Image Understanding*, vol. 65, no. 2, pp. 311–321, 1997.
- [19] C. Riess, E. Eibenberger, and E. Angelopoulou, "Illuminant Color Estimation for Mixed-Illuminant Real-World Scenes," in *ICCV Workshops*, 2011, pp. 774–781.
- [20] A. Gijsenij, R. Lu, and T. Gevers, "Color constancy for multiple light sources," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 697–707, 2012.
- [21] R. Kannan, H. Salmasian, and S. Vempala, "The spectral method for general mixture models," in *18th Annual Conference on Learning Theory (COLT)*, 2005, pp. 444–457.
- [22] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.
- [23] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *International Journal of Computer Vision*, vol. 52, no. 1, pp. 7–23, 2003.
- [24] J. Cohen, "Dependency of the spectral reflectance curves of the munsell color chips," *Psychonomic Science*, vol. 1, no. 12, pp. 369–370, Oct 1964.
- [25] L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *Journal of Optical Society America A*, vol. 3, no. 10, pp. 1673–1683, Oct 1986.
- [26] D. H. Marimont and B. A. Wandell, "Linear models of surface and illuminant spectra," *Journal of the Optical Society of America A*, vol. 9, pp. 1905–1913, 1992.
- [27] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao, "Statistics of cone responses to natural images: implications for visual coding," *Journal of Optical Society America A*, vol. 15, no. 8, pp. 2036–2045, August 1998.

- [28] T.-W. Lee, T. Wachtler, and T. J. Sejnowski, "The spectral independent components of natural scenes," in *Proceedings of the First IEEE International Workshop on Biologically Motivated Computer Vision*, ser. BMVC '00. London, UK, UK: Springer-Verlag, 2000, pp. 527–534.
- [29] A. Gijsenij, T. Gevers, and J. van de Weijer, "Generalized Gamut Mapping using Image Derivative Structures for Color Constancy," *International Journal of Computer Vision*, vol. 86, no. 2-3, pp. 127–139, 2010.
- [30] W. S. Stiles and J. M. Burch, "N.P.L. colour-matching investigation: Final report (1958)," *Optica Acta*, vol. 6, pp. 1–26, 1959.
- [31] J. Hoffbeck and D. Landgrebe, "A method for estimating the number of components in a normal mixture density function," in *Proceedings of Geoscience and Remote Sensing Symposium*, vol. 4, 2000, pp. 1675–1677.
- [32] D. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, 1992.
- [33] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, 2000.
- [34] S. Wright, *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics, 1997.
- [35] B. V. Funt, F. Ciurea, and J. J. McCann, "Retinex in MATLAB TM," *J. Electronic Imaging*, vol. 13, no. 1, pp. 48–57, 2004.
- [36] D. H. Foster, S. M. Nascimento, and K. Amano, "Information limits on neural identification of colored surfaces in natural scenes," *Visual Neuroscience*, vol. 21, pp. 331–336, 4 2004.
- [37] F. Ciurea and B. V. Funt, "A large image database for color constancy research," in *Color Imaging Conference*, 2003, pp. 160–164.
- [38] B. Funt and L. Shi, "The rehabilitation of maxrgb," in *Color Imaging Conference*. IS&T - The Society for Imaging Science and Technology, 2010, pp. 256–259.
- [39] K. Barnard, L. Martin, B. Funt, and A. Coath, "A Data Set for Colour Research," *Color Research and Application*, vol. 27, no. 3, pp. 147–151, 2002.
- [40] W. S. Stiles and J. M. Burch, "Interim report to the Commission Internationale de l'Éclairage Zurich, 1955, on the National Physical Laboratory's investigation of colour-matching," *Optica Acta*, vol. 2, pp. 168–181, 1955.
- [41] M. Everingham, L. Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, June 2010.