Detecting Multiple Illuminants

Clément Fredembach and Graham Finlayson
School of Computing Sciences
University of East Anglia
Norwich, NR4 7TJ, UK
{cf,graham}@cmp.uea.ac.uk

Abstract— **The presence of multiple illuminants in an image is an obstacle for many computer vision algorithms as well as most illuminant estimation algorithms. Multiply lit scenes do, however, occur in many situations, such as the presence of shadows or daytime indoor environments.**

Common ways to minimize the problems created by multiple lights are to perform a given task only in regions containing a single light or to remove the non-prevailing illumination, thereby rendering the entire image under a single illuminant.

Before these methods can be implemented though, it is necessary to first detect the parts of the image that are differently illuminated.

In this paper, we propose a novel method to detect multiple illuminants using the chromagenic theory. We start by taking two pictures of each scene, a normal one and one where a colored filter is placed in front of the camera, and then estimate the illuminant incident to each pixel of the image.

We constrain the illuminant estimation problem so that we look for a set of linear transforms where each region in the image is indexed by one of the transforms of the set. By pre-segmenting the image, we obtain illumination masks that detect multiple illuminants very accurately including occlusion shadows, which have shown to be a problem for most current shadow detection methods.

Experiments on a variety of real images show excellent results and validate our approach.

I. Introduction

Multiply lit images occur when two or more spectrally illuminants are present in a scene. Given that illuminants can strongly vary in direction, intensity and color (often all at the same time), images that admit multiple illuminants can be, and often are, problematic for many computer vision tasks such as tracking [13], scene analysis [15], object recognition [24][28] and face recognition [31]..

Additionally, in digital photography and computational colour constancy, a correct estimation of the scene illuminant is critical. These estimation algorithms, however, work under the assumption that a single illuminant is present [16], [10], [9], [2] and, as a result, make wrong estimates when two or more illuminants exist in a scene.

Unfortunately, multiply-lit scenes frequently occur in the real world, such as indoor environments (with a mixture of natural light through the window and flood light from the ceiling) and outdoor shadows.

Shadows will be our primary focus in this paper, for a couple of reasons: their prevalence in natural images and the often sharp transitions between non-shadow and shadow regions is all the more problematic. For these reasons, shadow detection is certainly the most studied occurrence of multiple illumination.

In a typical outdoor scene, the non-shadow parts of the image are illuminated by a mixture of direct sunlight and skylight. In contrast, shadow regions are lit mostly by skylight. These two illumination sources differ significantly both in brightness and colour -see Fig. 1 for an illustration- and, as a result, so do the image pixel values corresponding to shadow and non-shadow regions.

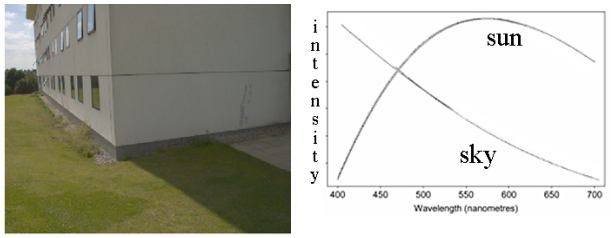

Fig. 1. *An outdoor image containing a shadow. And the Spectral Power Distribution of both illuminants: sun+sky light and sky-light only. Note the difference across the visible spectrum.*

In photography, shadows are often accidental and/or unwanted artifacts that in some conditions (e.g., cityscapes, flash, point light source) cannot be avoided. Another important attribute of shadows is that, when working with images that have a large bit depth, the presence of a shadow can indicate a High Dynamic Range (HDR) image. HDR images cannot be displayed on typical CRT monitors, however, if one can remove or attenuate the shadow, the dynamic range can then be compressed and the image properly displayed.

The rest of this paper is organized as follows: Section 2 reviews the current state of the art in shadow detection, section 3 introduces the chromagenic illuminant estimation algorithm, first introduced in [7]. Section 4 puts the chromagenic algorithm in perspective with the illuminant detection problem. Section 4 details the case where two illuminants are present in a scene and present the results of shadow detection. Section 5 looks at the robustness of the algorithm when more or less than two illuminants are present. Finally, section 6 concludes the paper.

II. Background

Detecting shadows is a difficult problem since shadows are created in diverse ways and can greatly vary in intensity, colour, shape and sharpness. As a result, additional information is often needed for an accurate detection.

This additional information can be provided in several ways. The most commonly used approach is to use a se-

quence of images instead of a single image.

A. Video-based methods

Weiss in [32] observed that given an outdoor video sequence over a long period of time: cast shadows (due to objects occluding the sun) move. It follows that the edges which are constant throughout the frames are related to the scene structure and not to the shadows. Weiss showed that a shadow free background could be obtained by taking the median edges of the sequence, effectively removing the shadows without extracting them.

Weiss approach has been extended by Matsushita et al. in [22] and [23] using several light sources and multiple cameras to recover, using multiview stereo algorithms, a view-dependant model of a reflectance only background image. They were also able, with a thresholding operator, to recover shadow masks from the illumination images.

In [17], Leone et al. looked at frame differences from video surveillance data. Their goal was to distinguish textured objects from shadows using a Matching Pursuit algorithm [20]. The idea of using texture classifiers has also been used in [11] and [27] based on the assumption that, while shadows alter both brightness and colour, the texture of a shadow region is essentially unchanged from its non-shadow equivalent and so, texture information can therefore be used in finding out whether a moving element in a sequence is a shadow cast by an object or the object proper.

In an approach based on computer graphics natural image matting methods, Chuang et al. [3] have been able to extract both shadow and lit images from a video. Their method however assumes a that primary point light source illuminates the scene and that fairly strong shadows are cast.

B. Single Image Methods

When having only a single image at disposal, more information is needed, either by using a trained classifier, making assumptions about the scene or through user supplied hints.

Jiang et al. in [12] detected shadows by segmenting the input image and classifying its regions as either shadow or non-shadow. To do so, they proposed that the darkest image regions are possible shadows. These candidates were then evaluated based on their geometry, assuming that the shape of shadow regions differed from the one of actual objects. This approach suffers from its simplicity and is only accurate in scenes with simple object shapes and non-textured backgrounds.

Following a gradient-based approach, Tappen et al. [30] classified image derivatives, depending on their direction and amplitude, as either shadow induced, material induced or undetermined. This three-way labelling is obtained with a classifier trained on a variety of reflectance and shadow transitions. The undetermined derivatives are then, in a second step, assigned a shadow or material label using a belief propagation algorithm that propagates information from reliably classified pixels.

In [18], Levine et al. also classified image edges as either shadow or material changes. They used a support vector machine, trained on colour and luminance ratios, for classification. In their method, it is assumed that a shadow transition occurs when there is an important change in luminance coupled with a weaker change in colour. Based on their detection results, they also proceeded to remove shadow regions from the image by assigning them the average brightness and colour values of their neighboring regions.

Both the Levine and Tappen approaches often work well. However, there are significant failures which manifest themselves in images: in part, these methods fail because some pixel derivatives can indicate *both* a shadow and a material edge, a common occurrence in the case of occlusion shadows. Our method, however, is able to detect complete occlusion shadows (where all the surrounding edges are coincidental luminance/material edges).

Finlayson et al. in [8] have proposed a method, based on minimum projection entropy that produce illumination-free images, that can in turn be used to create illumination masks. their method however assumes that the lights are Planckian and that the camera sensors are narrow-band; both assumptions are not necessarily verified in the real-world.

More recently, Wu and Tang [33] used a Bayesian approach to extract shadows where they opted for user supplied hints to disambiguate regions that could be wrongly estimated by an automatic method. Their method delivers good results but cannot be ported to an automatic framework.

In contrast, the approach we propose here is based on the premise that we have two images: a normal one and one where a colored filter is placed in front of the camera. Other two-images algorithms exist, such as Yoon et al. alternate point light source method [34], or the use of flash/no-flash image pairs, where the combination of these images can be used to either estimate the illuminant [5], [29] or to remove shadows [19] (REF CVPR06).

While these methods can provide remarkable results, they all require a very controlled environment. Yoon et al. method demands multiple point light sources that can be switched on and off at will, while all the flash methods are constrained by the physics of the flash itself, i.e., the objects must not be too dark, too bright, too close or too far.

The first advantage of our proposed method is that it can be used with any camera on any scene, no matter the ambient light or the scene content. Moreover, our method is not limited to shadow detection but is also applicable to most other instances of multiple illumination.

III. The Chromagenic Algorithm

The chromagenic illuminant estimation algorithm proceeds as follows: Let $S(\lambda)$ be the descriptor of surface reflectances, $E(\lambda)$ the scene illuminant SPD, $Q_k(\lambda)$ the camera sensitivities (we consider here trichromatic cameras, so $k = \{R, G, B\}$) and $F(\lambda)$ be the transmittance of the colour

filter placed in front of the camera.

The sensor responses of the unfiltered, $\underline{\rho}$, and filtered, $\underline{\rho}^F$, image can be written as:

$$\rho_k = \int_\omega E(\lambda)S(\lambda)Q_k(\lambda)d\lambda \quad (1)$$

$$\rho_k^F = \int_\omega E(\lambda)S(\lambda)F(\lambda)Q_k(\lambda)d\lambda \quad (2)$$

thus, for each scene we recover six responses per pixel that form the input to the illuminant estimation problem. For the purposes of this work, we set out to recover $\underline{\rho}_E$: the RGB of a white surface under the scene illuminant E.

Let us first consider the equations of filtered and unfiltered image formation (1) and (2). We can approximate the filtered image by posing a second illuminant, $E^F(\lambda)$ so that it is equivalent to putting the filter $F(\lambda)$ in front of the light source $E(\lambda)$, i.e., $E^F(\lambda) = F(\lambda)E(\lambda)$. We can therefore think of $\underline{\rho}$ and $\underline{\rho}^F$ as sensor responses of a single surface under two different illuminants. It has been shown in [21] and [6] that when the same surfaces are viewed under two lights, the corresponding RGBs can, to a good approximation, be related by a linear transform and so we use a 3×3 matrix to relate the RGBs captured with and without the coloured filter. We can then write:

$$\underline{\rho}^F = T_E^F \underline{\rho} \quad (3)$$

where T_E^F is a 3×3 linear transform that depends on both the chromagenic filter and the scene illuminant. Equation (3) implies that, given the chromagenic filter and sensor responses under a known illuminant, we can predict the filtered responses. In the problem of illuminant estimation, however, we know both the filtered and unfiltered responses but not the illuminant. Moreover, the task of finding the illuminant corresponds to finding T_E^F. If we know all possible illuminants *a priori* we can, for a given filter, determine the transforms T_E^F for every illuminant. We can then estimate which of these pre-computed transforms best fits the pair of filtered-unfiltered responses and thus, determine the illuminant.

Before outlining the actual algorithm, it is worth pointing out two cases where, depending on the filter or the sensor sensitivities, chromagenic colour constancy is not possible: if the filter has a neutral density or if the camera sensors behave like Dirac delta functions.

If the chosen filter has a neutral density, i.e., its transmittance does not vary across the spectrum, the relationship between filtered and unfiltered RGBs will be a constant scaling (the same for all lights). If we write:

$$F(\lambda) = \alpha, \ \forall\lambda \quad (4)$$

where α is a constant value, then

$$\underline{\rho}^F = \alpha\underline{\rho} \ , \ \forall S, E \quad (5)$$

It follows that the 6D responses will in fact span only three dimensions and thus solving for colour constancy is impossible.

If we suppose Dirac-type sensors where the non-null response of the k^{th} sensor is at the wavelength λ_k, we can rewrite equations (1) and (2) as:

$$\rho_k = E(\lambda_k)S(\lambda_k)Q_k(\lambda_k) \quad (6)$$

$$\rho_k^F = E(\lambda_k)S(\lambda_k)F(\lambda_k)Q_k(\lambda_k) \quad (7)$$

It follows that $\rho_k^F = F(\lambda_k)\rho_k$ and that the responses are, again, three dimensional and their relation depends neither on the reflectances nor on the scene illuminant. Additionally, while not as limiting as the neutral density case, using a rank-deficient filter -e.g., a pure red filter- will deliver poor constancy since significant information is lost -in this case the relationship between the blue pixels.

Barring the cases outlined above, the transforms can be pre-computed by choosing a set of n typical scene illuminants: $E_i(\lambda), \ i = 1, \ldots, n$ and a set of m surface reflectances: $S_j(\lambda), \ j = 1, \ldots, m$ representative of the real world. For each illuminant i, we create a $3 \times m$ matrix $\mathcal{Q}_i$ whose j^{th} column contains the sensor response of the j^{th} surface illuminated by the i^{th} illuminant. We also create $\mathcal{Q}_i^F$, which contains the equivalent filtered responses. For each illuminant, we can then define the transform matrix as:

$$\mathcal{T}_i = \mathcal{Q}_i^F \mathcal{Q}_i^+ \quad (8)$$

where $+$ denotes the Moore-Penrose pseudo-inverse operator, $\mathcal{Q}^+ = (\mathcal{Q}^{\text{T}}\mathcal{Q})^{-1}\mathcal{Q}^{\text{T}}$. $\mathcal{T}_i$ can then be described as the transform that best maps, in a least square sense, unfiltered to filtered responses under illuminant i. Because it is a least squares fit, $\mathcal{T}_i$ will not, in practice, map the responses without errors. What matters, however, is that the error committed when mapping responses under illuminant i is the smallest when the corresponding transform $\mathcal{T}_i$ is used.

Once the n transforms have been pre-computed, the illuminant estimation proceeds as follows: let $\mathcal{Q}$ and $\mathcal{Q}^F$ denote the $3 \times m$ matrices of unfiltered and filtered RGBs of arbitrary reflectances under an unknown light. For each plausible illuminant we calculate the fitting error, e_i, as:

$$e_i = \|\mathcal{T}_i\mathcal{Q} - \mathcal{Q}^F\|, \ \ i = 1, \ldots, n \quad (9)$$

under the assumption that $E_i(\lambda)$ is the actual scene illuminant. We then choose the transform that minimizes the error and surmise that it corresponds to the scene illuminant. Our estimated illuminant is $E_{\text{est}}(\lambda)$ where

$$\text{est} = \arg\min_{\text{i}}(\text{e}_\text{i}) \ \ \text{i} = 1, \cdots, \text{n} \quad (10)$$

It was shown in [7] that if both reflectances and illuminants can be *exactly* described by three basis functions each, i.e., they are three dimensional, then the chromagenic algorithm delivers perfect illuminant estimation. In natural scenes, however, these dimensions are generally higher [26], [14] and so there are estimation errors.

IV. Chromagenic Illuminant Detection

In this section, we propose a new method to detect shadows and other types of multiple illumination. Our approach, based on the chromagenic method for illuminant

estimation outputs very precise binary illumination maps that can even accurately detect occlusion shadows in cases where all the edges surrounding a shadow region are coincidental material/illuminant edges, a non-feasible task for most shadow detection methods.

In Chapter 3, we reviewed the chromagenic algorithm for illuminant estimation. This algorithm can accurately estimate an image illuminant and, importantly, is pixel-based. One can therefore assume that the same methodology can be applied to illuminant detection. To do so, given a pair of filtered and unfiltered images as well as a set of pre-computed linear transforms $\mathcal{T}_i$, one simply applies the chromagenic illuminant estimation algorithm and, for each pixel in the image, records which transform best maps the unfiltered RGB values to their filtered counterparts. The best transform for each pixel indexes the incident illuminant for that pixel.

An example of such processing is shown in Fig. 2 where a reflectance image from the Nascimento set [25] is illuminated with two distinct lights from E_{87} (the 87 lights from [1]). We see from (2 right) however that the detected illuminants -each pixel has for value the index of its best transform- do not correspond to the input, even though the images are perfectly registered. The problem here is that the chromagenic approach is efficient when multiple surfaces are present in the scene but is also fragile, when a single surface is present in the scene.

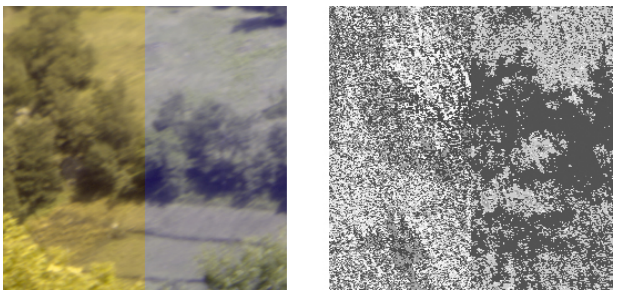

Fig. 2. *Left: One of the reflectance image from Nascimento et. al dataset; the left and right halves of the image are illuminated by two different lights. Right: The result of illuminant detection using the standard chromagenic algorithm. Each pixel of the image has for value the index of the transform that best maps it to its filtered counterpart.*

To address this stability issue, we transform the illuminant estimation algorithm in one of illuminant discrimination -or detection. Importantly, we do not aim to recover the actual scene illuminants but, instead, we look for the transforms that best discriminate the multiple illuminants in the scene irrespectively of the estimation accuracy. The starting point of our approach is to suppose there are m lights present in the image. In practice $m \leq 2$ will be appropriate for most images, $m = 2$ is particulary important for it represents the shadow detection case.

Let E_N be a set N lights for which we carry out the chromagenic preprocessing step and solve for the N relations, the 3×3 linear transforms $\mathcal{T}_i$, that best map RGBs to filtered counterparts.

Suppose we now select m elements in N, denoting the corresponding subset E_m, with $E_m \subset E_N$. Taking each pixel in turn, we determine which of the m relations best maps its RGB values to the filtered ones. Once each pixel is assigned a single of the m relations, we can assess how well the subset E_m accounts for the data by calculating its error:

$$\text{error}_{\text{E}_\text{m}} = \sum_{\text{i=1}}^{\text{m}} \|\mathcal{T}_\text{i}\ \text{I}_\text{i} - \text{I}_\text{i}^\text{F}\| \tag{11}$$

Where I_i and I_i^F are the unfiltered and filtered RGB pairs of the image I that are best mapped by $\mathcal{T}_i$.

Equation (11) measures how well *one* subset of E_N models the transition from the image to its filtered equivalent. It does not, however, guarantee that this particular subset is optimal -it does not even tell how good this subset is. To minimize detection errors, we therefore have to evaluate equation (11) for all possible m-elements subsets of E_N. Let $\mathcal{E}(m)$ be the set of all m-elements subsets of E_N. The $E_{\text{opt}} \in \mathcal{E}(m)$ that best describes the relation between the image and its filtered counterpart is:

$$E_{\text{opt}} = \arg \min_{E_m \in \mathcal{E}(m)} (\text{error}_{\text{E}_\text{m}}) \tag{12}$$

Once we have found subset E_{opt} that minimizes the mapping error among all m-subsets, we create our illuminant map M as:

$$M(x) = \arg \min_{i \in m} \|\mathcal{T}_i I(x) - I^F(x)\| \tag{13}$$

i.e., the x^{th} pixel of M takes the value of the index of the transform that best maps the x^{th} pixel of I to its filtered response.

Implemented naively, obtaining M can be computationally laborious. The computational cost is proportional to the cardinality of the set $\mathcal{E}(m)$. If we chose m lights among N, then:

$$\sharp(\mathcal{E}(m)) = \frac{N!}{(N-m)!m!} \tag{14}$$

Considering our set of 87 lights, the number of different m-sets is $(3,741)$, $(105,995)$ and $(2.2\ 10^6)$, for $m = 2, 3, 4$ respectively. A brute force search is only really possible for small m, i.e., $m = 2$ or $m = 3$.

V. The Two Illuminants Problem

The case where $m = 2$ is the commonest instance of the multiple illumination problem. Indeed, in every day circumstances such as the presence of shadows or the combination of natural and artificial light sources in indoor environments, the number of different illuminants is rarely greater than two.

Limiting ourselves to the case where two illuminants are present also makes the problem more tractable since there are only $\frac{N^2 \pm N}{2}$ relations to test for if we have a set of N lights (the $\pm N$ comes from the decision to include or not pairs of identical illuminants). In our experiments, since we have either 87 -for the synthetic case- or 86 -for the real images- illuminants: the number of possibilities is smaller than $4,000$.

To test our framework for illuminant detection, we start by considering the reflectance images of Nascimento et al [25] and E_{87}: the SFU set of 87 measured lights. We generate images by illuminating each half of the image with a different light. Then, we use our algorithm to classify pixels as '0' or '1' depending on which of the two illuminant best maps them to the filtered values, thus creating a binary mask of illumination. This procedure is illustrated in Fig. 3 where the two images are illuminated by skylight/sunlight and floodlight/sunlight respectively, thereby reproducing two of the most frequent naturally occurring two-illuminant scenes.

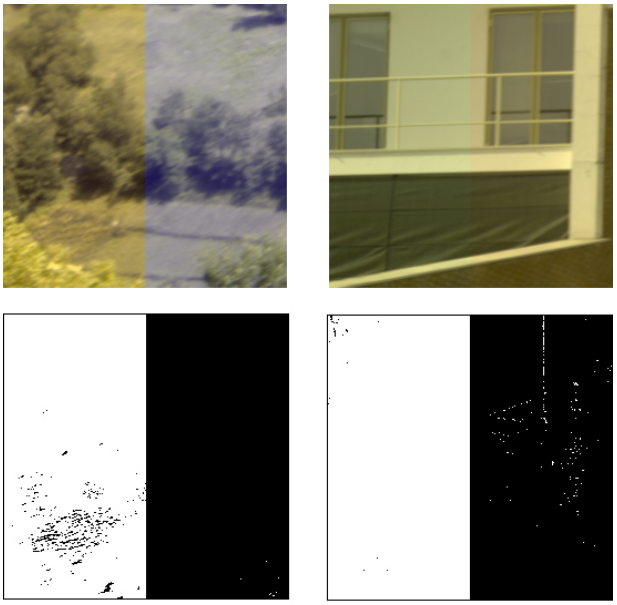

Fig. 3. *Top row: 2 images from the Nascimento dataset. The left image is illuminated by both skylight and sunlight, The right image by neon light and sunlight. Bottom row: The pixel wise classification of illuminants for the two images.*

The results illustrate that, while generally accurate, the classification is noisy even though the images are perfectly registered. The presence of noise is expected since the classification is still done at pixel-level only; the confusion can be explained by a misclassification due to having a single reflectance for each computation. To improve the detection accuracy, we propose to perform the classification not at pixel-level but at region-level.

The main insight of a region-based approach is that, over an area, the predominance of a class of pixels is correlated with the prevailing illuminant. It follows that we can modify the pixel-level classification to take into account neighboring information.

To do so, let $\mathbf{R}$ be a partition of I into K distinct regions $R_j,\ j = \{1, \ldots, K\}$. We formulate our region-based approach by rewriting equation (13), defining the illuminant map for a region R_j, M_{R_j}, to be the result of a function over the filtered and unfiltered corresponding region of the image.

$$M_{R_j} = \mathcal{F}_{R_j}(I_{R_j}, I^F_{R_j}) \tag{15}$$

where I_{R_j} represents the pixels of I that belong to region R_j; M_{R_j} will take a single value for the entire region. For each region of the image, the region is labelled using the index of the transform that best maps it to its filtered response. We use here the function $\mathcal{F}_{R_j}$ to define what that best mapping is. The function can take any form and we illustrate here two: majority voting and minimization of pixel-based error over the region.

In the majority voting case, the image is first processed according to equation (11). Then, within each region, the number of pixels that belong to each class is counted. The entire region is assigned the label of the majority of pixels:

$$\mathcal{F}_{R_j}(I_{R_j}, I^F_{R_j}) = \arg\max_i(\sharp(\mathcal{T}_i I_{R_j} - I^F_{R_j})) \tag{16}$$

That is, the function takes for value the index i which maps the largest number of pixels in R_j with the least error (compared to all other transforms).

The region-based labelling can also be done using a min error metric. In that case, the whole region is evaluated using both illuminants in turn. The region is then labelled according to which one of the two transforms has minimal error, i.e.,

$$\mathcal{F}_{R_j}(I_{R_j}, I^F_{R_j}) = \arg\min_i \|\mathcal{T}_i I_{R_j} - I^F_{R_j}\| \tag{17}$$

The outcome of both methods is illustrated in Fig. 4 where the image is partitioned in 8×8 blocks (the outcome of both region-based labelling is identical, so a single map is shown). On most of our experiments, we have not found any significative difference between either method, so in the following we keep our framework coherent and use, for region classification, the method of error minimization.

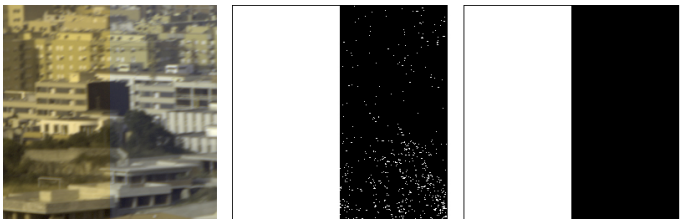

Fig. 4. *Left: The original image with two illuminants. Middle: the result of pixel-wise classification. Right: the illuminant mask processed on 8×8 regions with the methods of either majority voting and error minimization. The results of both region-methods is identical; in general no significant difference is observed between the two methods.*

Using a region-based labelling has an additional advantage when illuminant detection is performed on real images: image registration (or lack thereof). Consider the pair of images shown in Fig. 5; the images appear to be registered but, at a pixel-level, it is actually not the case. Consequentially, the detection will be noisier than in the synthetic case (Fig. 5: right). A region-based approach is therefore more adapted for a cleaner classification.

Since real images have noisier masks, the regions shape and size noticeably influence the final results. This dependency is illustrated in Fig. 6 where we process the pixel-level mask using both a 8×8 window and regions obtained with a segmentation algorithm (in this case, the meanshift

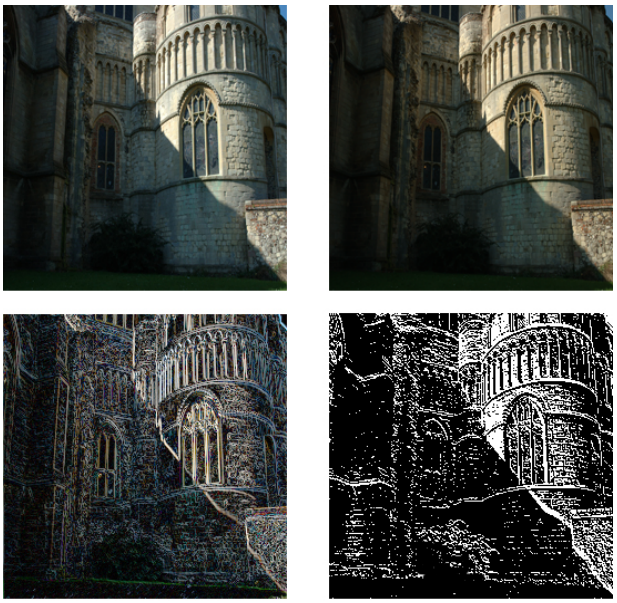

Fig. 5. *Top Row left and right: the original and filtered image, while they appear to be registered, this is not the case at pixel level. Bottom row: Registration differences (left) and the result of pixel-based classification (right). The detection is mostly accurate but noisy.*

algorithm [4]). While in both case we significantly oversegment the image, the meanshift segmentation is better for our purposes as it preserves the image edges and results in more accurate masks. We point out that any segmentation method that preserves most of the edge structure of the image would be suited for our type of processing.

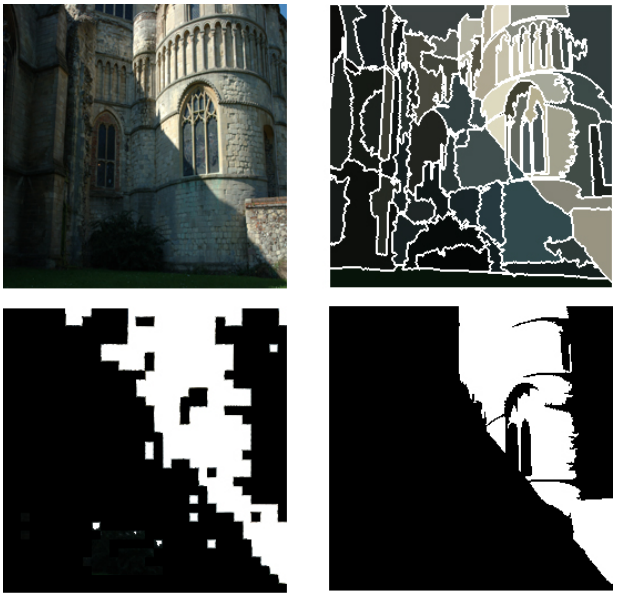

Fig. 6. *Top row: The original image and the meanshift segmentation. Bottom row: the results from partitioning the image with* 8×8 *blocks and with the meanshift algorithm. Since it preserves the edges, the segmentation results yields more accurate masks.*

To visualize the results, the image can be decomposed according to the binary mask. Doing so allows us to actually see what parts of the image are detected as being differently illuminated. The results, Fig. 7, show that the image is effectively segmented in shadow and non-shadow regions.

Fig. 7. *The original image and its segmentation according to the binary mask obtained with the chromagenic illuminant detection. Both the shadow and lit parts of the image are accurately segmented.*

Results for various situations of indoor and outdoor lighting are shown in Fig. 8. For every image we used meanshift to segment the images using its standard parameters and a minimum region size of 0.5% the size of the image. Despite some minor accuracies, the illuminations are well separated. One of the main strengths of this method is its ability to detect occlusion shadows even when all of the region edges are coincidental material and shadow edges, a major improvement over gradient or region comparison methods that generally assume the reflectances on both sides of a shadow edge are identical.

VI. $\mathbf{m \neq 2}$

When two illuminants are present in an image and the detection algorithm is constrained to discriminate them, we saw that this discrimination was accurate. It is however difficult to infer the number of illuminants present in an image *a priori*, so we want to analyze the algorithm behavior when a single illuminant is present but we try to find two.

In essence, we run the exact same test as in the previous section, but on images that contain only a single illuminant. The results (Fig. 9) show that, while the algorithm is looking for the best pair of illuminants, the returned map is almost unitary. From these results, we infer that we can assume the maximum number of lights in the image to be higher than in reality and still obtain accurate results.

For completeness, we now address the case where more than two illuminants are present. While theoretically possible, the number of different mappings for $m = 3$ and 4 is $105,995$ and $2.2\ 10^6$ respectively (for our set of 87 transforms). Also, in natural images, there are very few occurrences of more than three lights. The case of three illuminants can usually be put down to the presence of two distinct lights plus the mixture of those lights (such as the penumbra in shadows where the transition is not immediate).

We illustrate the three lights case on both reflectances images (where the lights are distinct) and on real images (where the distinction is blurred). For real images, we look at the difference between assuming the presence of either two or three illuminants. The results (Fig. 10) show that the detection, in the synthetic case, is as accurate as when

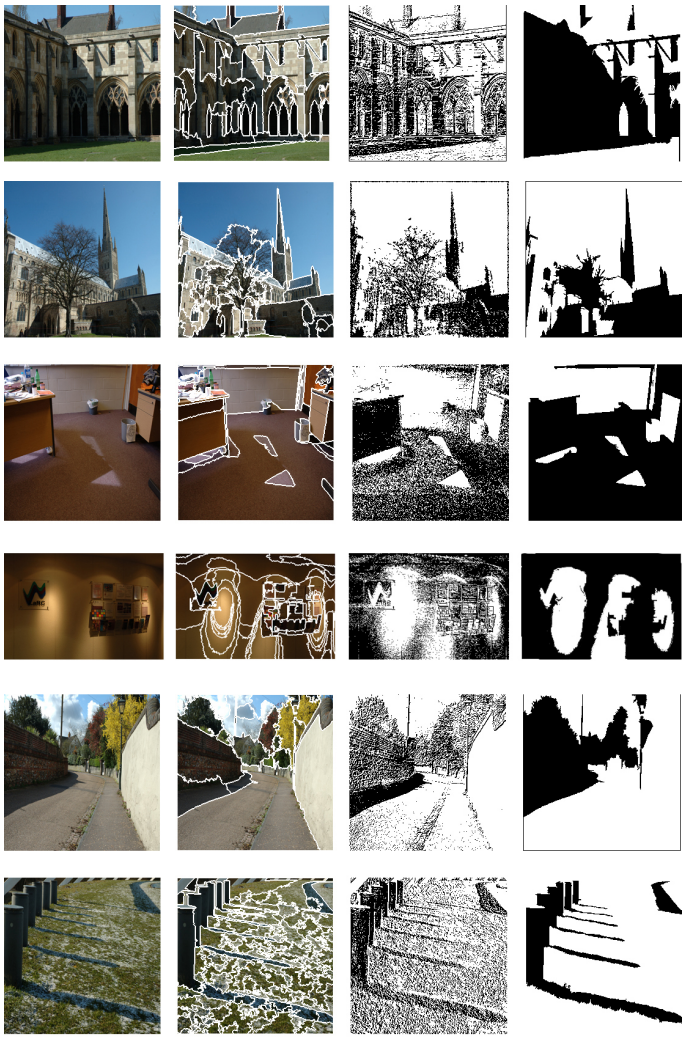

Fig. 8. *Columns from left to right: the original images, the mean-shift segmentation; the region-based classification and the pixel-based classification.*

Fig. 9. *columns from left to right, 1) the original images containing a single illuminant; 2) the pixel-based pair classification: the map is almost unitary but for some minor noise; 3) the results using region processing: again, almost all regions are labelled using a single illuminant.*

two illuminants are sought. On real images, we see that the transitory regions are picked up as different illuminants but the improvement does not necessarily justify the increase in complexity.

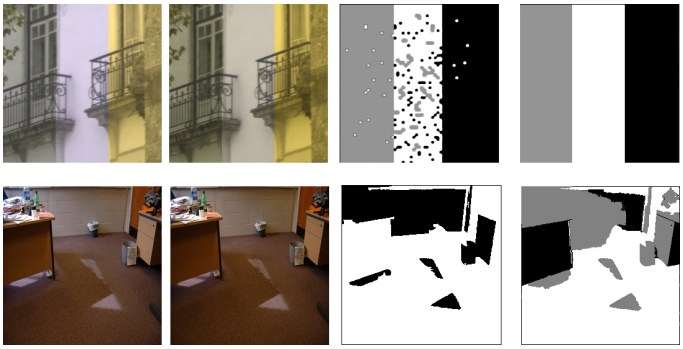

Fig. 10. *Detection assuming three illuminants. Top row: synthetic images and results from both pixel and region classification. Bottom row: indoor scene with two main illuminants. Assuming two illuminants sperate them (3rd image). Assuming three illuminants enables to refine the classification depending on the relative proportion of the two illuminants in the scene.*

We point out that, for illumination detection, the number of transforms $\mathcal{T}_i$ can be reduced. Indeed, since we are interested in finding out illumination difference irrespectively of the accuracy of illuminant estimation, we can limit the number of transforms used so that they still cover the possible gamut of lights but with a coarser sampling.

This can be achieved, for example, by looking at the images in the training set and manually selecting a single transform per illuminant class (i.e., keeping only a single sunlit image to represent the whole class of sunlight illuminants). Alternatively, we obtained similar results to the ones presented throughout this section by plotting the mapped 2D chromaticities of a white patch with all the transforms and selecting a smaller number (15 out of 87) using the k-means algorithm to form clusters.

VII. Conclusion

In this paper, we have introduced a new method to detect multiple illuminants in images based on the chromagenic theory. By forcing the lights to be examined pairwise without preoccupying ourselves about the accuracy of the illuminant estimation and processing the results on a region instead of pixels, we obtained very accurate results for a variety of illuminations. We went on to show that this method does not require prior knowledge of the number of lights in the image and were also able, when more than two illuminants were assumed, to detect shadow penumbra.

References

[1] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms- part i: methodology and experiments with synthetized data. *IEEE Trans. on Image Processing*, 11:972–984, 2002.

[2] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:1–26, 1980.

[3] Y-Y Chuang, D.B. Goldman, B. Curless, D.H. Salesin, and R. Szeliski. Shadow matting and compositing. In *ACM SIGGRAPH 2003*, pages 494–500, 2003.

[4] D. Comanicu and P. Meer. Mean shift: A robust approach towards feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 24, 2002.

[5] J.M. DiCarlo, F. Xiao, and B.A. Wandell. Illuminating illumination. In *Proc. of the ninth Color Imaging Conference*, pages 27–34, 2001.

[6] M. D'Zmura and G. Iverson. Color constancy i: Basic theory of two-stage linear recovery of spectral descriptors for lights and surfaces. *Journal of the Optical Society of America*, 10:2148–2165, 1993.

[7] G. Finlayson, S. Hordley, and P. Morovic. Colour constancy using the chromagenic constraint. In *Computer Vision and Pattern Recognition (CVPR) 2005*, pages 1079–1086, 2005.

[8] G.D. Finlayson, M.S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *Proc. of the European Conference on Computer Vision (ECCV)*, pages 582–595, 2004.

[9] G.D. Finlayson and S.D. Hordley. Color constancy at a pixel. *Journal of the Optical Society of America*, 18:253–264, 2001.

[10] D.A. Forsyth. A novel algorithm for colour constancy. *Intl Journal of Computer Vision*, 5:5–36, 1990.

[11] T. Horprasert, D. Hardwood, and L.S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *Proc. of ICCV Frame-Rate Worshop*, 1999.

[12] C. Jiang and M.O. Ward. Shadow segmentation and classification in a constrained environment. *CVGIP: Image Understanding*, 59:213–225, 1994.

[13] H. Jiang and M. Drew. Tracking objects with shadows. In *CME03: International Conference on Multimedia and Expo,*, pages 100–105, 2003.

[14] D.B. Judd, D.L. MacAdam, and G. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. *Advances in Neural Information Processing*, 54:1031–1040, 1964.

[15] G. J. Klinker, S. A. Shafer, and T. Kanade. A physical approach to color image understanding,. *International Journal of Computer Vision*, 4:7–38, 1990.

[16] E.H. Land. The retinex theory of color vision. *Scientific American*, pages 108–129, 1977.

[17] A. Leone, C. Distante, and F. Buccolieri. A shadow elimination approach in video-surveillance context. *Pattern Recognition Letters*, 27:345–355, 2006.

[18] M.D. Levine and J. Bhattacharyya. Removing shadows. *Pattern Recognition Letters*, 26:251–265, 2005.

[19] C. Lu and M.S. Drew. Practical scene illuminant estimation via flash/no-flash pairs. In *Proc. of the fourteenth Color Imaging Conference*, pages 1–1, 2006.

[20] S.G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41:3397–3415, 1993.

[21] L.T. Maloney and B.A. Wandell. Color constancy: A method for recovering surface spectral reflectance. *Journal of The Optical Scociety of America part A*, 3:29–33, 1986.

[22] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Shadow elimination for robust video surveillance. In *Proc. of Worshop on Motion and Video Computing*, pages 15–21, 2002.

[23] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakaushi. Illumanation normalization with time-dependent intrinsic images for video surveliance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26:1336–1347, 2004.

[24] Y. Moses, Y. Adini, and S. Ullman. Face recognition, the problem of compensating for changes in illumination direction. In *Europ. Conf. on Comp. Vision (ECCV)*, pages 286–296, 1994.

[25] S.M.C. Nascimento, F. Ferreira, and D.H. Foster. Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A*, 19:1484–1490, 2002.

[26] J. Parkkinen and T. Jaaskelainen. Characteristic spectra of munsell colors. *Journal of The Optical Scociety of America part A*, 6:318–322, 1989.

[27] J. Stauder, R. Melch, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Trans. on Multimedia*, 1:65–77, 1999.

[28] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7:11–32, 1991.

[29] R. Szeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 246–253, 2000.

[30] M.F. Tappen, W.T. Freeman, and E.H. Adelson. Recovering intrinsic images from a single image. In *Proc. of the Advances in Neural Information Processing Systems (NIPS)*, pages 1343–1350, 2003.

[31] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *Proc. of Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[32] Y. Weiss. Deriving intrinsic images from images sequences. In *International comference in Computer Vision (ICCV)*, pages 68–75, 2001.

[33] T.-P. Wu and C.-K. Tang. A bayesian approach for shadow extraction from a single image. In *Proc. of the tenth International Conference on Computer Vision*, pages 480–487, 2005.

[34] J.J. Yoon, C. Koch, and T.J. Ellis. Shadowflash: an approach for shadow removal in an active illumination environment. In *Proc. of the British Machine Vision Conference (BMVC)*, pages 626–645, 2002.