

# Detecting Illumination in Images

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

Mark S. Drew  
School of Computing Science  
Simon Fraser University  
Vancouver, V5A 1S6, Canada  
mark@cs.sfu.ca

## Abstract

*In this paper we present a surprisingly simple yet powerful method for detecting illumination—determining which pixels are lit by different lights—in images. Our method is based on the chromagenic camera, which takes two pictures of each scene: one is captured as normal and the other through a coloured filter. Previous research has shown that the relationship between the colours, the RGBs, in the filtered and unfiltered images depends strongly on the colour of the light and this can be used to estimate the colour of the illuminant. While chromagenic illuminant estimation often works well it can and does fail and so is not itself a direct solution to the illuminant detection problem. In this paper we dispense with the goal of illumination estimation and seek only to use the chromagenic effect to find out which parts of a scene are illuminated by the same lights.*

*The simplest implementation of our idea involves a combinatorial search. We precompute a dictionary of possible illuminant relations—that might map RGBs to filtered counterparts—from which we select a small number  $m$  corresponding to the number of distinct lights we think might be present. Each pixel, or region, is assigned the relation from this  $m$ -set that best maps filtered to unfiltered RGB. All  $m$ -sets are tried in turn and the one that has the minimum prediction error over all is found. At the end of this search process each pixel or region is assigned an integer between 1 and  $m$  indicating which of the  $m$  lights are thought to have illuminated the region.*

*Our simple search algorithm is possible when  $m = 2$  (and  $m = 3$ ) and for this case we present experiments that show our method does a remarkable job in detecting illumination in images: if the 2 lights are shadow and non-shadow, we find the shadows almost effortlessly. Compared to ground truth data, our method delivers close to optimal performance.*

## 1. Introduction

Much of computer vision, image processing and imaging in general is predicated on the assumption that there is a single prevailing illuminant lighting a scene. However, often there are multiple lights. Common examples include outdoor scenes with cast and attached shadows, indoor office environments which are typically lit by skylight and artificial illumination and the spot-lighting used in commercial premises and galleries. Relative to these mixed lighting conditions, many imaging algorithms (based on the single light assumption) can fail. Multiply-lit scenes cause problems for tracking [14], scene analysis [15], object recognition and illuminant estimation. The latter problem is almost always solved under a single light assumption[1] yet ignoring multiple lights can often lead to very poor image reproduction[13].

While we are interested in multiply-lit scenes in general, in this paper shadows will be our primary focus. Indeed, not only are cast shadows the most commonly encountered multiple illumination situation they are arguably the most difficult to deal with: shadow edges are often quite strong and it is this sudden transition between coloured lights that causes problems in applications such as tracking. An example of a cast shadow is shown in Fig. 1. Alongside we show the Planckian illuminants corresponding to the color temperatures of the light we measured inside (bluish) and outside the shadow (yellowish).

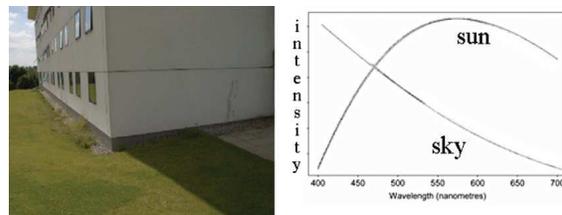


Figure 1. An outdoor image containing a shadow. And the Spectral Power Distribution of both illuminants: sun+sky light and sky-light only. Note the difference across the visible spectrum.

In this paper, we propose a novel method to detect multiple illuminants present in a scene — using a chromagenic camera and invoking a simple combinatorial detection strategy. A chromagenic camera captures two images for every scene: the first is a conventional image and the second is taken through a coloured filter [7]. Chromagenic theory has shown that the relationship between filtered and unfiltered responses depends strongly on, but varies with, illumination and that this relationship can be exploited in illuminant estimation. Interestingly, while good estimates are possible they are not guaranteed and chromagenic theory does not fully solve the illuminant estimation problem. Here, we seek to exploit the relationship between filtered and unfiltered RGBs not to determine the colour of the light but only to discriminate between different lights.

Our illuminant detection algorithm is, in spirit, simplicity itself: we assume that a set of  $m$  relations suffices to map the unfiltered RGBs to filtered counterparts. We assign to each image region the best relation. We repeat this process for all  $m$ -sets of relations and choose the  $m$  relations that in competition with all possible other sets best account for the data. At the end of this process each pixel or region is assigned a number between 1 and  $m$ . A pixel or region that is assigned the same number is deemed to have been illuminated by the same light.

Experiments on a variety of real images show excellent results and serve to validate our approach. Illumination detection is shown for both shadow and non shadow images. While these results are, we believe, compelling in their own right (see Figure 6) we were anxious to corroborate our method using ground truth data. Thus, we carried out a simple psychophysical experiment where observers had to identify which parts of images were illuminated by the same light. We could then compare these ‘illuminant maps’ with the detections delivered by our algorithm. We found that we are in agreement with the ground truth over 90% of the time. Moreover, given the variability in the detections made by the observers this is close to the best that might reasonably be achieved.

The rest of this paper is organized as follows: Section 2 reviews the current state of the art in shadow detection and section 3 introduces the chromagenic theory [7] which is used in detecting illuminants. Section 4 introduces our new illuminant detection algorithm. Experiments validate our method in section 5. Finally, section 6 concludes the paper.

## 2. Background

The prior art in the area of illuminant detection mainly concerns finding shadows in images. Shadows are found by loosely exploiting image context: context might be temporal (video), edge-based (local variation), two lights (before and after a flash) or the context might be supplied by a user

who can indicate some candidate shadow regions[27].

### 2.1. Video-based methods

Weiss in [26] observed that given a fixed camera position and 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. Effectively, Weiss finds non-shadow edges by calculating the median edge over a sequence of images. Reintegrating the median edge field results in a shadow free image (from which the position of shadows is easily calculated). Other video methods use multiple cameras[19, 20], assumptions about the illuminant[2] or task specific information [12], [22] and [16].

### 2.2. Single Image Methods

Tappen et al. [24] propose a gradient based approach where image derivatives are classified as illumination or reflectance edges depending on their direction and magnitude and, in cases of ambiguity, on other edges in the neighbourhood. While this simple method can work well, there is no a priori reason why exactly the same edge cannot be due to light or material changes. Levine et al. [17], present a similar method where edges are classified as either shadow or material changes. insight is incorporated into their classifier: shadows take place where there is an important change in luminance

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 edges can indicate *both* a shadow and a material edge, a common occurrence in the case of occlusion shadows.

Finlayson et al. in [9] observed that shadow illumination is a different colour from non-shadow light and the variation in light colour in a scene is 2 dimensional (and can be well characterised by two variables: intensity and colour temperature). Rather surprisingly, they showed that because of this two dimensionality they could, with an appropriate photometric calibration step, make a greyscale intrinsic image from RGB which depended only on reflectance. Shadow edges could then be found by comparing edges in the intrinsic and RGB images (those that appear in the latter but not the former must be shadows). More recently Finlayson et al. [8] showed that the greyscale image could be found by a statistical analysis of the colours in an image and have proposed a method to find the intrinsic image (and hence shadows) without any calibration. The Finlayson et al. approach, which can work well, still suffers from ambiguity problems: it may wrongly identify the presence (or absence of a shadow) and like the Levine and Tappen approaches does not account for occluding shadows.

### 2.3. Two-Light Methods

Arguably, neither long sequences of video or a single RGB image will render the shadow detection problem tractable. It is reasonable therefore to seek an in-between solution. Yoon et al. propose that shadows can be found using an alternate point light source method [28]. One can also make use of flash/no-flash image pairs, where the combination of these images can be used to either estimate the illuminant [4, 23] or to remove shadows [5].

These two-light approaches both make the shadow detection problem easier to solve and also they deliver more accurate detection. However, probing a scene with sufficiently bright flashes to find shadows is not practical (there is no flash that lightens shadows outdoors).

### 3. Chromagenic Theory

Consider the simple Lambertian model of image formation, for a chromagenic camera. 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:

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

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

integrating over visible wavelengths  $\omega$ . Thus for each scene we recover six responses per pixel, and these will form the input to our illuminant detection algorithm.

It has been shown in [18] 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  $3 \times 3$  matrix to relate the RGBs captured with and without the coloured filter. We thus write:

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

where  $T_E^F$  is a  $3 \times 3$  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.

Assuming  $n$  training illuminants we can precompute  $n$  transforms:  $T_i = \mathcal{P}_i^F \mathcal{P}_i^+$ . Here  $\mathcal{P}$  denotes the  $3 \times M$  RGB responses for a training set of reflectances; the subscript  $i$  denotes dependence on illumination and the superscript  $F$  on the filter. The superscript  $+$  denotes the Moore-Penrose (least-squares) inverse of the matrix. Illumination estimation proceeds as follows: let  $Q$  and  $Q^F$  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:

$$e_i = \|T_i Q - Q^F\|, \quad i = 1, \dots, n \quad (4)$$

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. The estimated illuminant is  $E_{\text{est}}(\lambda)$  where

$$\text{est} = \arg \min_i (e_i) \quad i = 1, \dots, n \quad (5)$$

Though simple, the chromagenic illuminant estimation algorithm was found to deliver significantly [21] better performance than all other algorithms tested. Unfortunately, while *on average* the illuminant is estimated with good accuracy, the worst case performance was found to be no better than the other algorithms. Indeed, neither the chromagenic algorithm nor any of its competitors are sufficiently accurate that they themselves can be used as a vehicle for illuminant detection.

To illustrate this point we took a multispectral image from Foster and Nascimento[25] and made a synthetic image where the left half of the image was illuminated by a yellowish light and the right hand side by a bluish light. We now, at each pixel, run the chromagenic algorithm and try to determine which light is present. Remembering that the chromagenic illuminant estimation algorithm returns an integer index to an illumination in the set (here we assume a set of 87 standard[1] typical, and representative, lights) we can code this illumination number in greyscale. The results of the experiment are shown in Figure 2: the original 2-illuminant image is on the left and the recovered per pixel illumination map on the right. Ideally, the left and right sides of the output image should be a single gray value.

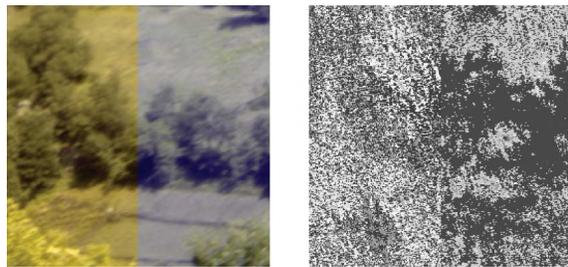


Figure 2. Left: One of the reflectance images from Nascimento et al. dataset; the left and right halves of the image are illuminated by two different lights. Right: The result of illuminant detection using the standard chromagenic algorithm. Each pixel of the image has for value the index of the transform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

This experiment is far from being a complete failure: we do see an edge in the middle of the image. However, the

edge is not strong and the left and right hand sides of the image are far from being classified as two distinct classes. Moreover, this example is perhaps the simplest case we might encounter: two lights illuminating two distinct and continuous regions. The pattern of illuminants in images in general will be much more complex.

#### 4. Chromagenic Illuminant Detection

Of course, it is not surprising that we found it difficult to estimate illumination colour at a pixel. Illuminant Estimation is known to be a hard problem. Moreover the antecedent illuminant detection methods (see section 2) do not try to estimate light colour but rather to detect when the illumination changes (almost exclusively they set out to identify shadow edges). Here we similarly set out to solve the detection problem but in contradistinction to all antecedent methods attempt to do this *at each pixel*. As we will see in the experimental section, the ability to operate pixelwise allows for (what we believe to be the very first time) the detection of occluding shadows in images.

Let us begin by carrying out a simple synthetic two-light experiment. First, we randomly select 10 surfaces from the 1995 standard reflectances compiled by Barnard[1]. We divide this set into two disjoint sets A and B, each of which has at least 3 reflectances. We then render these sets under two different randomly chosen coloured lights, giving two sets of RGBs (denoted  $I$ ) + filtered RGBs ( $I^F$ ). For A and B we have corresponding pairs:  $[I_A I_A^F], [I_B I_B^F]$ . For each response set we calculate the best  $3 \times 3$  transforms that maps filtered to unfiltered responses:  $T_A = I_A[I_A^F]^+$  and  $T_B = I_B[I_B^F]^+$ . Because the maps  $T_A$  and  $T_B$  are not exact, it is possible for an RGB response  $\underline{p}_A$  that  $\|T_A \underline{p}_A - \underline{p}_A^F\| > \|T_B \underline{p}_A - \underline{p}_A^F\|$ : the transform that minimises error overall might be suboptimal for individual surfaces. The likelihood or otherwise of this circumstance goes to the heart of our method. As we shall see if we use the wrong transform often then we will do a poor job of detecting illumination. Conversely, a small error rate would augur well for our algorithm.

Thus, we calculated the probability that the wrong transform best maps filtered to unfiltered RGB (i.e., that  $T_A$  best maps  $\underline{p}_B$  and  $T_B$  best maps  $\underline{p}_A$ ). This was true less than 2% of the time. We can distinguish correctly which light illuminates a given surface more than 49 times out of 50. Moreover, we repeated the experiment and added Gaussian noise at the 2% level and found that this result held (classification was slightly worse but we were still correct about 98% of the time).

##### 4.1. The Algorithm

Let us begin by assuming that for  $N$  lights we carry out the chromagenic preprocessing step and solve for the  $N$  re-

lations that best map RGBs to filtered counterparts. Individual relations are denoted  $f^i$  and the set of relations as:  $\mathfrak{R} = \{f^1, f^2, \dots, f^N\}$ . Note that here we do not necessarily assume that the relation is a  $3 \times 3$  matrix transform but rather, for generality, assume an arbitrary function  $f : \mathfrak{S}^3 \rightarrow \mathfrak{S}^3$ , where  $\mathfrak{S}$  is the set of possible integers in a colour image (for example, for 16-bit colour channels,  $\mathfrak{S}$  is the set  $[0..65535]$ ). Possible non-linear relations include polynomial mapping transforms[11] and Look Up Tables[10].

Suppose we now select an  $m$ -element subset  $\mathcal{R} \subset \mathfrak{R}$ . Taking each pixel, or region, in turn we determine which of the  $m$  relations best maps the RGB(s) to the filtered counterpart(s). Once each pixel, or region, is assigned a single relation it is a simple matter to calculate how well the set of  $m$  relations  $\mathcal{R}$  accounts for our data. Of course there are many possible  $m$ -element subsets  $\mathcal{R}$  in  $\mathfrak{R}$ . Mathematically, the set of all  $m$  element subsets of  $\mathfrak{R}$  is denoted  $\mathfrak{R}(m)$ . The  $\mathcal{R} \in \mathfrak{R}(m)$  which best describes the relation between image and filtered counterpart overall is then found through an optimisation procedure (which, in this paper, is essentially a brute-force searching algorithm). This effectively finds the  $m$  best mappings, and thus an  $m$ -level labelling of pixels. E.g., in the case  $m = 2$  this amounts to a binary labelling of pixels. Ideally, the labelling should identify which parts of the image are illuminated by different lights. If for example an image has a cast shadow this binary map should (say) have a 0 everywhere in the image that is in shadow and 1 elsewhere.

In general we determine which areas are illuminated by different lights by solving:

##### General statement of optimisation:

$$\arg \min_{\mathcal{R}, i_k} \sum_k \|f^{i_k}(I_k) - I_k^F\|, \quad (6)$$

with  $\mathcal{R} \in \mathfrak{R}(m)$

$$i_k \in \{1, 2, \dots, m\}$$

where  $\|\cdot\|$  is an appropriate measure of fit.

If  $I_k$  is a single pixel then  $\|\cdot\|$  is some simple scalar function, e.g. the sum of absolute values of vector components, or the square root of the sum of absolute values squared. If  $I_k$  is a region there is scope to make  $\|\cdot\|$  a more robust measure, e.g. the median deviation. Assuming some colour calibration is carried out then a perceptual measure of colour error (such as CIE Lab) might be used.

To make our approach slightly more general we allow the goodness of fit operation to be carried out pixelwise but will assign lighting labels on a *region by region* basis. Suppose

we compute an assignment of  $n$  regions indexed by  $k$ ,  $k = 1, 2, \dots, n$  in an image. Now, let  $I_{k,j}$  denote the  $j$ th pixel in the  $k$ th region. We now, initially, assign the relation labels  $i_{k,j}$  by minimising:

**Region-driven statement of optimisation:**

$$\begin{aligned} \arg \min_{\mathcal{R}, i_{k,j}} \sum_{j,k} \|f^{i_{k,j}}(I_{k,j}) - I_{k,j}^F\|, \\ \text{with } \mathcal{R} \in \mathfrak{R}(m) \\ i_{k,j} \in (1, 2, \dots, m) \end{aligned} \tag{7}$$

We assign labels to regions based on the fits to the underlying pixels:

$$i_k = \text{bestlabel}(\{i_{k,j} : I_{k,j} \in I_k\}) \tag{8}$$

Here, function  $\text{bestlabel}()$  must choose which label to assign to region  $k$ , of all the up to  $m$  labels assigned to pixels  $I_{k,j}$  in region  $k$ . An obvious candidate for function  $\text{bestlabel}()$ , and the one we use in this paper, is the mode function. E.g., if  $I_k$  has 100 pixels and, of those 100, and 90 have a relation label  $i$ , then the mode is also  $i$  and the overall label for the region should be  $i$ .

We remark that, implemented naively, minimising (6) or (7) is computationally laborious. The computational cost is proportional to the cardinality of the set  $\mathfrak{R}(m)$ . If, say, there are 50 relations in  $\mathfrak{R}$  then the cardinality of the  $m$ -set  $\mathfrak{R}(m)$  is  $\frac{50!}{m!(50-m)!}$  which for  $m = 2, 3, 4, 5$  is equal to 1225, 19600, 230300 and 2118110. A brute force search is only really possible for small  $m$ , i.e.  $m = 2$  or  $m = 3$ .

Thus far we have assumed that we have a precomputed set of relations. If we do not happen to have available any pre-computed mappings from unfiltered to filtered responses, then a robust statistical procedure can be used to find the best mapping from one image to the other provided that at least half the image (plus 1 pixel) is approximately associated with that mapping. Pixels not associated correctly are outliers and belong to the second label.

**5. Experiments**

Arguably, the  $m = 2$  case is the most interesting and most common case. Indeed, it is hard to think of normal circumstances when  $m$  is much larger than 2. Thus a brute force search is tractable (indeed it is fast) and we can use either of the optimisations we presented in the last section.

In our first experiment we wish to improve upon the results shown in Figure 2. The poor illuminant detection results shown on the right hand side of this figure is due to us allowing all lights to be candidates for the illumination at each pixel. Here, let us assume that there are just two lights illuminating the scene: then it follows we should only

consider a single pair of transforms when attempting to determine which parts of the scene are illuminated by which light. Thus for each pair of transforms, at a pixel, we calculate the RMS error and assign the transform that delivers the minimum error. We carry out this process for all possible pairs of transforms: the best pair overall is the one which when applied pixel-wise across the scene (one transform per pixel) results in the minimum prediction error. In effect we implement (6) as a simple combinatorial search. The resulting illuminant detection result is shown in Figure 3 below. Notice that we have (almost perfectly) correctly discovered which of the two lights is illuminating which parts of the image.

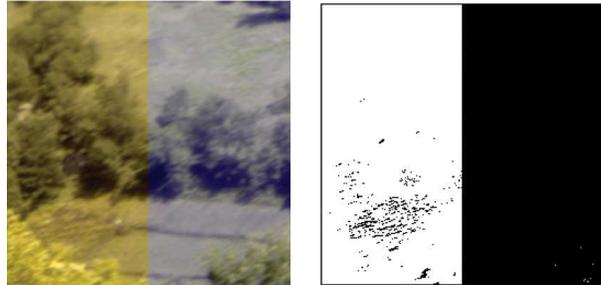


Figure 3. Left: One of the reflectance images from Nascimento *et al.* dataset; the left and right halves of the image are illuminated by two different lights. Right: The result of illuminant detection using the chromagenic algorithm set out here. Each pixel of the image has for value, 1 or 0, the index of the transform that best maps it to its filtered counterpart.

Now let us implement the algorithm shown in (6) and (7) and test it on natural images. We begin by creating the set of all possible illuminant maps  $\mathfrak{R}$ , which in this case consists of fifty  $3 \times 3$  matrix transforms. These transforms were calculated by imaging a standard colour reference chart (the Macbeth ColorChecker ) under 50 lights one at a time, with and without a coloured filter, using a Nikon D70 camera (which outputs linear raw unprocessed images). The 50 lights were chosen to be representative of typical lights that are encountered every day and included: bluesky only, bluesky + sun, overcast sky, fluorescent light and incandescent illumination. The Macbeth ColorChecker has 24 different coloured patches and so we solved for each  $3 \times 3$  transform by regressing the 24 unfiltered RGBs onto the filtered counterparts.

Now, we calculate the 2-set  $\mathfrak{R}(2)$ : the set of all subsets of  $\mathfrak{R}$  with 2 elements. Because there are 50 transforms there are ‘50 choose 2’ equals 1225 combinations. For a given relation set  $\mathcal{R}$  containing a particular pair of  $3 \times 3$  matrices, we test which matrix best maps each image pixel to the filtered counterpart. As we do so, we calculate the discrepancy, or error, between the mapped RGBs and the actual filtered responses. We repeat this process over all 1225 combinations of two lights (and hence mappings); we determine the pair

of transforms, one of which is applied at each pixel, that best maps the unfiltered to filtered image overall.

Figure 4 shows typical results of an optimisation Eq. (6) applied at the pixel level. Figure 4(a) shows the original image: since it has shadows there are clearly two lights present in the scene. Fig 4b shows the best pixel-wise illuminant detection results delivered by our algorithm. Note a pixel value of 1 or 0 is an index for the corresponding  $3 \times 3$  matrix that best maps the RGB at that pixel to its filtered counterpart. While the results are somewhat noisy we clearly have arrived at a reasonable illuminant map.

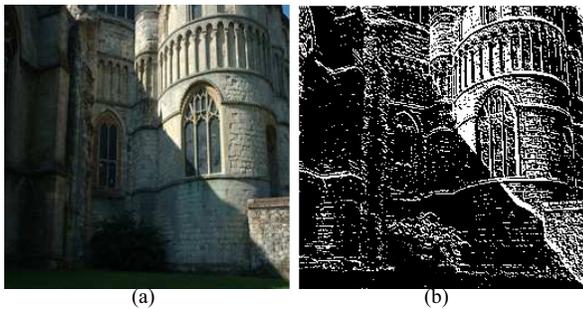


Figure 4. *Pixel-based illuminant detection: noisy.*

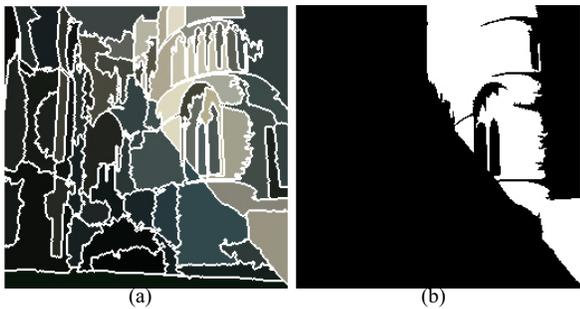


Figure 5. *Segmentation, and the result of region-based illuminant detection: clean determination of shadow areas.*

Now let us go on to apply the region-based label assignment given by optimisation Eq. (7) followed by Eq. (8). Using the Mean Shift algorithm [3], or any similarly edge-preserving segmentation algorithm, we calculate an initial segmentation of an image. Figure 5(a) shows the segmentation arrived at by the standard Mean Shift algorithm where, so as not to merge regions lit by more than light, we start with an over segmented the image. Figure 5b shows the resulting illuminant detection map. It is clear that a region based computation has returned superior detection performance.

We have found that our region based illuminant detection algorithm is remarkably adept at detecting different lights in images. Figure 5 shows a selection of results. The left im-

age is the input, the 2nd column the output of mean-shift segmentation then the 3rd column shows pixelwise classification of the illuminants. Finally the rightmost column shows the illuminant maps found by our region based algorithm.

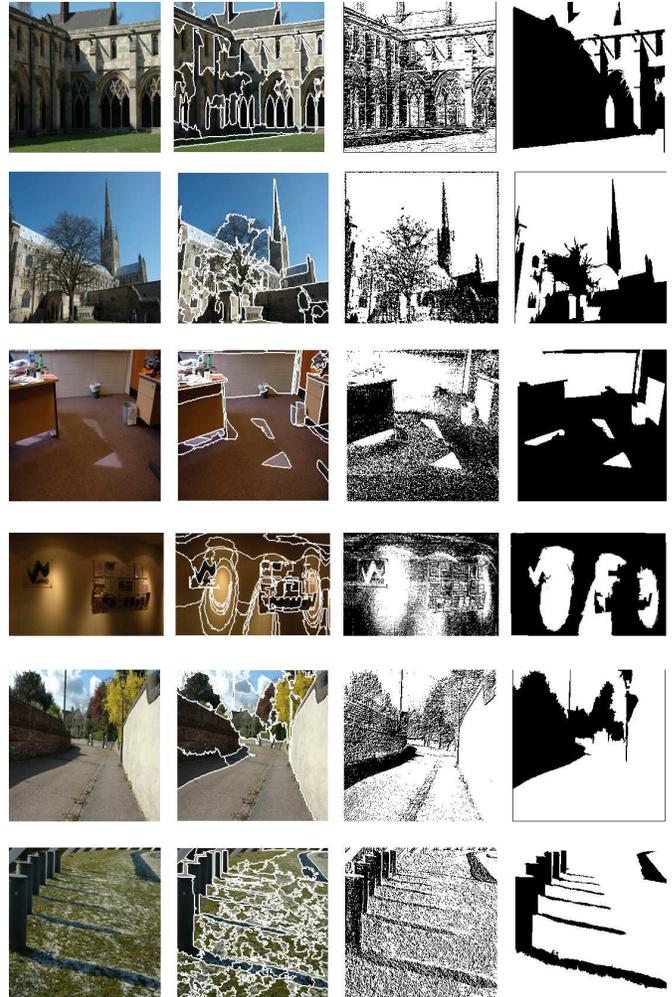


Figure 6. *From left to right: input image, segmented image, pixel-wise illuminant detection and the final illuminant map. Top to bottom: a range of multiply lit scenes*

We draw the reader’s attention to the 2nd row of Figure 5. The shadow discovered on the spire of the cathedral is an occluding shadow. That is, the surface which is in shadow is not out of shadow anywhere else in the image. As such, prior art, which assumes that the surface on either side of a shadow edge is the same, must (and does fail). Yet, the approach advanced in this paper finds occluding shadows with ease. The 4th row shows an indoor example where there are two lights. Noticed that there are no sharp edges and the pixelwise result seems to capture the uncertainty in determining which light is present where. The final bi-

nary illuminant map represents a good 2-class estimate of the lights in the scene.

The reader might be interested in what happens when there is only a single light or indeed when there are more than 2 lights present. Regarding the former case, we have found that by assuming 2 lights and solving the optimisation we generate an illuminant map that classifies almost all regions as belonging to a single light. Similarly, if we assume there are 3 lights when there are two our algorithm tends to return an almost binary map of illuminations. One exception to this was the indoor image shown in Figure 7. The result of assuming 3 lights, setting  $m = 3$ , returns a map which finds those parts of the image illuminated by outdoor (grey), indoor (black) and mixed (white).



Figure 7. The left image is an indoor scene and the image on the right shows the illuminant map recovered: outdoor (gray), indoor (black) and mixed (white).

### 5.1. Ground-Truth Data

While visually the results of illuminant detection look good, we wished to test our algorithm on ground truth data. Thus we took 5 outdoor images which had both cast and occluding shadows and segmented them using the mean shift algorithm. Each segmented image had about 50 segments. We then asked 5 non expert (but colour-normal) observers to shade in the image regions in each segmented image that they thought were in shadow (the observer could see the original unsegmented image). The unshaded regions are interpreted as being not in shadow. We say an observer judgment agrees with our algorithm if when our algorithm identifies a shadow (or non shadow) region an observer agrees.

Figure 8 shows two of the images we used. The left image is the original, the middle is the image after mean-shift segmentation. The observer is then given the edge-map from the segmentation and shades in where they think the shadows are (right image).

Agreement between the observers and our algorithm was found to be 91% over all regions. We repeated this calculation where instead of computing the percentage of regions in agreement we counted the percentage of pixels. This area weighted calculation found that our algorithm was in agreement with the observers 95.3% of the time.

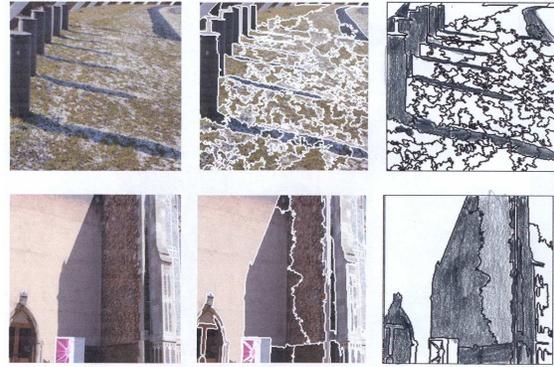


Figure 8. Left column: input images, middle column: mean shift segmentations and right column the shadows found by a human observer.

We also calculated intra-observer agreement rates and found these were 90.6% for regions and 96% when weighted for area. It is evident then that the agreement between our algorithm and the ground truth is almost as good as might be expected given the observer variability. In this sense we say that our algorithm returns close to optimal results.

## 6. 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 pair-wise without insisting on accuracy of the illuminant estimation, and by processing the results on a region rather than pixel basis, we obtained very accurate results over a variety of illuminations. Results are remarkably good, especially considering how simple and fast the method is.

## References

- [1] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms- part i: methodology and experiments with synthesized data. *IEEE Trans. on Image Processing*, 11:972–984, 2002. [1](#), [3](#), [4](#)
- [2] 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. [2](#)
- [3] 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. [6](#)
- [4] J.M. DiCarlo, F. Xiao, and B.A. Wandell. Illuminating illumination. In *Proc. of the ninth Color Imaging Conference*, pages 27–34, 2001. [3](#)

- [5] M.S. Drew, C. Lu, and G.D. Finlayson. Removing shadows using flash/noflash image edges. In *IEEE International Conference on Multimedia Expo*, pages 1–1, 2006. 3
- [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. 3
- [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. 2
- [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. 2
- [9] G.D. Finlayson, S.D. Hordley, and M.S. Drew. Removing shadows from images. In *Proc. of the 7th European Conference on Computer Vision (ECCV)*, pages 823–836, 2002. 2
- [10] G.D. Finlayson, P.M. Morovic, and S.D. Hordley. The chromagenic colour camera and illuminant estimation. In *Proc. of the 13th Color Imaging Conference*, pages 20–25, 2005. 4
- [11] Guowei Hong, Bing Han, and M. Ronnier Luo. Colorimetric characterization of low-end digital camera and its application for on-screen texture visualization. In *ICIP*, 2000. 4
- [12] 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 Workshop*, 1999. 2
- [13] P. Hubel. the perception of color at dusk and dawn. In *Proc. of the 7th Color Imaging Conference*, pages 48–51, 1999. 1
- [14] H. Jiang and M. Drew. Tracking objects with shadows. In *CME03: International Conference on Multimedia and Expo*, pages 100–105, 2003. 1
- [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. 1
- [16] A. Leone, C. Distanto, and F. Buccolieri. A shadow elimination approach in video-surveillance context. *Pattern Recognition Letters*, 27:345–355, 2006. 2
- [17] M.D. Levine and J. Bhattacharyya. Removing shadows. *Pattern Recognition Letters*, 26:251–265, 2005. 2
- [18] L.T. Maloney and B.A. Wandell. Color constancy: A method for recovering surface spectral reflectance. *Journal of The Optical Society of America part A*, 3:29–33, 1986. 3
- [19] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Shadow elimination for robust video surveillance. In *Proc. of Workshop on Motion and Video Computing*, pages 15–21, 2002. 2
- [20] Y. Matsushita, K. Nishino, K. Ikeuchi, and M. Sakauchi. Illumination normalization with time-dependent intrinsic images for video surveillance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26:1336–1347, 2004. 2
- [21] S.D.Hordley and G.D.Finlayson. Reevaluation of color constancy algorithm performance. *The Journal of the Optical Society of America A*, 23(5):1008–1020, May 2006. 3
- [22] J. Stauder, R. Melch, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Trans. on Multimedia*, 1:65–77, 1999. 2
- [23] 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. 3
- [24] 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. 2
- [25] Manchester University. Hyperspectral images. <http://personalpages.manchester.ac.uk/staff/david.foster/>, 2002. 3
- [26] Y. Weiss. Deriving intrinsic images from images sequences. In *International conference in Computer Vision (ICCV)*, pages 68–75, 2001. 2
- [27] 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. 2
- [28] 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. 3