A Markov Random Field Framework for Finding Shadows in a Single Colour Image

Cheng Lu and Mark S. Drew
School of Computing Science, Simon Fraser University, Vancouver (CANADA)
Corresponding author: Mark S. Drew (mark@cs.sfu.ca)

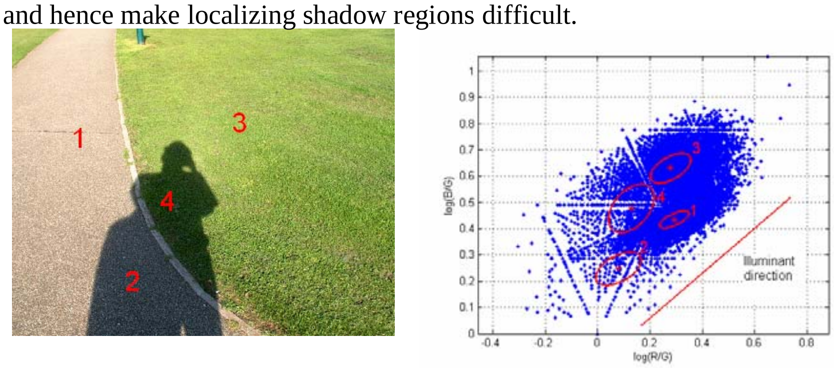
ABSTRACT

Many computer vision algorithms, such as segmentation, tracking, and stereo registration, are confounded by shadows in images. Hence finding shadows in colour images is an important research issue. As opposed to the majority of techniques, which either need a sequence of images or require geometric information on images, this paper proposes an illuminant discontinuity measure by which shadow edges can be locally identified. We model the problem of finding shadows by a Markov random field which uses the new measure. The Markov random field provides a computational framework whereby local and area constraints can be optimized such that shadows can be segmented in a single colour image. Results are presented for real images and show accurate shadow extraction.

1. INTRODUCTION

Estimating and removing illuminant effects in colour images has been a major research subject in computer vision. This problem is difficult due to the confounding interaction between intrinsic object reflectance and illumination in the formation of colours. Shadows, as one illumination effect, indicate a local change in both the colour and the intensity of illumination. Several authors have addressed the problem of finding shadows. Generally, they either extract shadows from a sequence of images,[1] or find a shadow matte based on geometrical knowledge.[2] In order to remove shadows, an illumination invariant image has been introduced,[3] by which a greyscale image for an input colour image can be found which is invariant to both colour and intensity of the scene illuminant.

For the purpose of producing an invariant image, we start by representing a daylight illumination E by Wien's approximation to Planck's formula, which can be parameterized by its colour temperature T alone. An invariant image is derived based on the fact that under Planckian lights, and for a given Lambertian surface imaged by a camera whose sensors are fairly narrowband (as for an ideal delta function sensor camera), the 2-d scatter plot of logarithms of ratios of colour components, *R/G* versus *B/G*, say, produce an approximately straight line as the illuminant colour temperature T changes. Let us recapitulate how this works. Let ρ be the response of a 3-sensor colour camera; we have $\rho_k = \sigma \int E(\lambda) S(\lambda) Q_k(\lambda) d\lambda$ where $Q_k(\lambda)$ denotes the spectral sensitivity of the kth camera sensor ($k=R,G,B$), σ is a Lambertian shading term, and $S(\lambda)$ is the surface reflectance. Now let us form the log-band-ratio 2-vector chromaticity $\chi : \chi = \log \rho_k / \rho_p$, $k=R,B$, $p=G$. Substituting the expressions for ρ into this log-chromaticity formula, summarizing in a vector form we have: $\chi = s + (1/T)e$, where s is a 2-vector which depends on surface and camera but is independent of the illuminant, and e is a 2-vector which is independent of surface (but still depends on the camera). Thus we can see that as illumination changes (T varies) the log-chromaticity for a given surface moves along a straight line. It follows that if a target consisting of different colour surfaces is imaged by this camera, as illumination changes log-chromaticity will follow a set of straight lines. Each line corresponds to a colour surface; each point on a line corresponds to a particular illuminant. Since the direction of the lines depends only on the properties of the camera, all such lines are parallel. We call this direction the *illuminant direction*. The invariant image is formed by projecting 2-vectors onto the direction orthogonal to this direction. In Fig. 1(a), the photographer's cast shadow lies on the path and the grass. Fig. 1(b) plots the 2-d log-ratio chromaticity points and the illuminant direction, which is parallel to the red line. There are four main regions in this image: lit path, shadow path, lit grass and shadow grass, marked 1-4 in Fig. 1(a). We also draw four ellipses in Fig. 1(b) corresponding to the main concentrations of points for the four regions. When projecting chromaticity points to the direction orthogonal to the red line, the four clusters fall into two portions: one for lit grass and shadow grass; and the other for lit path and shadow path. Thus, points in shadow have the same value as points in corresponding nonshadow regions. In a 1-d image for these points, the shadow will be mostly eliminated. Note that the projection operation eliminates not only shadow effects but also all information along the illuminant direction. For real images, many factors affect the distribution of the chromaticity along the illuminant direction, such as noise and surface texture. As well, real cameras have non narrow-band sensors, which could make the distribution complex. A straightforward observation is that in Fig. 1(b), the four regions have many points overlapped, although the majority are separated. It is expected that when points are projected into 1-d quantities, much information which is irrelevant to the illuminant will coalesce. This will prove troublesome for detecting the edges in the invariant image, which in turn may impact the accuracy of shadow edge detection,[4] and hence make localizing shadow regions difficult.



(a) (b)

Figure 1 (a): Input colour image; 1-4 label four main regions. (b): Log chromaticity for the input image: four ellipses corresponding to the main regions.

Fig. 1(b) shows that the four clusters constitute two pairs, each of which aligns along the illuminant direction: a shadow-nonshadow grass pair, and a shadow-nonshadow path pair. This leads to a general observation: rather than projecting all 2-d chromaticities into 1-d quantities, we can simply compare two pixels on both sides of a region edge: if their chromaticities belong to one of the pairs; the edge is probably a shadow edge. We call two pixels with a neighbour relation across a shadow-nonshadow edge an *illuminant discontinuity pair*. In this paper we propose a continuous angle between vectors as a criterion for measuring the illuminant discontinuity. We then define a Markov random field to model our binary shadow segmentation problem. The MRF approach combines spatial context and local chromaticity features to interpret shadow-nonshadow edges. The segmented shadows are found as a result of energy minimization of this random field.

2. ILLUMINANT DISCONTINUITY MEASUREMENT

We have shown that, under certain assumptions, log-ratio chromaticity for a single surface will follow a straight line as illumination changes. Reformulating our formulation $\chi = s + (1/T)e$ in vector form, we have $\chi = [\chi_1; \chi_2] = [s_1 + (1/T)e_1; s_2 + (1/T)e_2]$, where χ_1 denotes log*R/G*, and χ_2 denotes log*B/G*. Now if the same surface is imaged by the same camera under a different light, characterized by T', we have $\chi' = [\chi_1'; \chi_2'] = [s_1 + (1/T')e_1; s_2 + (1/T')e_2]$. If we perform vector subtraction on χ, χ', we have a new vector describing the change of illuminant: $[(1/T - 1/T')e_1; (1/T - 1/T')e_2]$. This is independent of surface reflectance, and thus captures the illumination direction.

There are two primary concerns: First, this conclusion is based on Wien's approximation of Planck lights — does this approximation work for real lights? Second, how to determine if there is an illuminant discontinuity (shadow-nonshadow edge) between two neighbouring pixels, i.e. whether the line linking them lies in the illuminant direction.

It has been shown[5] that the Planckian constraint on the form of the illuminant can be relaxed to a much more general form, with CIE standard lights still represented very closely. In one variant this is found by performing a principal component analysis on the log of a set of CIE daylights. The great majority of the variance is accounted for by the first eigenvector, so the log of CIE daylight can be represented by $L = L_m + pb$, where L_m is the mean of the log of illuminants, p is the first eigenvector, and b is a weighting factor. L_m and p are functions of λ, whereas varying b produces different lights. The variation of b with λ is small compared with the variation with T, and so b may be considered a function of T. This leads to a new form for representing the log of illuminants: $L=a(\lambda)+u(\lambda)f(T)$. Clearly, Wien's approximation of Planckian light is a special case of this form. Thus, we can rewrite the vector subtraction as: $[(b-b')e_1;(b-b')e_2]$, where e_1 and e_2 are functions of λ and independent of T. This form verifies that under natural lights, the log chromaticities indeed fall on a line as illuminants vary.

Now, we are aiming at specifying a criterion by which illuminant discontinuity pairs can be determined, i.e. if the two pixels align along the illuminant direction. For a particular camera, the illuminant direction can be computed either by a pre-calibration scheme[3] or an automatic entropy minimization method.[4] We denote the vector for the illuminant direction by θ_0. For the purpose of measuring the difference of the two directions, we propose factoring the difference directly into the similarity measure calculation by using the normalized inner product of two vectors: $Q_{ij} = <\theta_{ij}, \theta_0> / (\|\theta_{ij}\| \|\theta_0\|)$, where θ_{ij} is the vector linking the two chromaticity points for neighbouring pixels i,j, so that Q_{ij} is cos of the angle between θ_0 and θ_{ij}.

3. MARKOV RANDOM FIELD

The MRF-based segmentation model is defined by the contextual relationships within a local neighbourhood structure. Since our goal is the assertion of local discontinuity constraints, we will consider only first order random fields, both simplifying the model and limiting the computational complexity. The formulation of our Gibbs model will be similar to others used for segmentation,[6] except for a number of variations due to the characteristics of our illuminant discontinuity measure.

Suppose we are given a colour image pixel set $X=\{x_i\}$, on a first order neighbourhood system represented by a set N of all unordered pairs, and each variable x_i takes a label l_i in L={shadow, nonshadow}. We first calculate the adjacent-pixel illuminant discontinuity criterion Q_{ij} for neighbouring pixels i,j. As the intensity is also an important cue for determining shadow-nonshadow edges, constraints on intensity difference between neighbouring pixels are necessary. Given the intensity difference R_{ij} between i,j, we rewrite the discontinuity criterion in the form of a weighted sum: $D_{ij}=wQ_{ij}+(1-w)R_{ij}$. Then the energy E can be formulated as follows:

$$E(l) = \sum_i \sum_{j \in N_i} D_{ij}\delta'(l_i, l_j) + \beta(1-\delta(l_i, l_j)),$$

$\delta(l_i, l_j) = 1$ if ($l_i = l_j$), and $\delta(l_i, l_j) = 0$ if $l_i \neq l_j$;

$\delta'(l_i, l_j) = 1$ if ($l_i = l_j$) or ($l_i \neq l_j$ and ($I_i > I_j$ and l_i is shadow) or ($I_i < I_j$ and l_i is nonshadow)), and

$\delta'(l_i, l_j) = 0$ if ($l_i \neq l_j$ and ($I_i > I_j$ and l_i is nonshadow) or ($I_i < I_j$ and l_i is shadow)),

where I is the pixel's intensity. β controls the threshold of the discontinuity. If $D_{ij} < \beta$ (not a candidate discontinuity pair), then it is cheaper to pay the price D_{ij} and set no discontinuity, i.e. both $\delta(l_i, l_j)$ and $\delta'(l_i, l_j)$ are 1. $\delta(l_i, l_j)$ and $\delta'(l_i, l_j)$ are indicator functions for labelling discontinuity. If there is a discontinuity between i and j, and the shadow-nonshadow labels for i, j are consistent with their intensity, then both $\delta(l_i, l_j)$ and $\delta'(l_i, l_j)$ will be 0, so the cost is β. If the two labels are not consistent with their intensity, i.e. the lower intensity pixel is labelled nonshadow, then $\delta'(l_i, l_j)$ is 1 and $\delta(l_i, l_j)$ is 0, and so the penalty is $D_{ij}+\beta$.

The model is no more complicated than a standard Potts model and so is well-understood and easily implemented. The primary drawback is that texture and noise may confuse the discontinuity measure. To remove these factors, the Mean Shift method is used to filter the image first, and the edges are then detected on the filtered image. Our illuminant discontinuity criterion is calculated on only these edges, yielding a fast region-labelling scheme. A second undesired effect is that shadow edges usually are not sharp but diffuse, so that the chromaticity of neighbouring pixels may be too close to each other, leading to an untrustworthy angle. To overcome this effect, for two pixels across an edge the illuminant discontinuity criterion is calculated using the means of two blocks of pixels on the two sides of the edge.

4. RESULTS

The Gibbs Sampler[7] will be used to optimize the model. Optimization is iterated until equilibrium is reached, producing a complete image labelling. The free parameter w controls the significance of the illuminant discontinuity criterion in relation to the intensity. Since shadow-nonshadow borders prefer illuminant changes, we simply set $w=0.7$. Results on real images are shown in Fig. 2. Shadows are clearly present in each colour image, and the binary mask image results show that shadows are very successfully found.

Figure 2: Input colour images and extracted shadows.

5. CONCLUSION

We have presented a Markov Random Field-based model for finding shadows. This model use a vector-angle difference measure between two log ratio chromaticities and includes weights to take intensity difference into account. In our model, β and w are free parameters set empirically. Future work will be aimed at modifying the model to allow estimating these parameters and segmenting shadows simultaneously.

References

1. Y. Matsushita and K. Nishino, "Illumination normalization with time-dependent intrinsic images for video surveillance," IEEE Trans. Pattern Anal. Mach. Intell., 26, pp. 1336-1348 (2004).
2. Y. Y. Chung, D. B Goldman, B. Curless, D. Salesin and R. Szeliski, "Shadow matting and compositing," in Proceedings of ACM SIGGRAPH 2003, Vol. 22, pp. 494-500.
3. G.D. Finlayson and S.D. Hordley, "Colour constancy at a pixel," J. Opt. Soc. Am. A, 18, 253-264 (2001).
4. G.D. Finlayson, M.S. Drew, and C. Lu, "Intrinsic images by entropy minimization," in ECCV: European Conf. on Computer Vision (Springer-Verlag, Berlin 2004) pp. 582-595.
5. J.A. Marchant and C.M. Onyango, "Color invariant for daylight changes: relaxing the constraints on illuminants," J. Opt. Soc. Am. A, 18, 2704-2706 (2001).
6. P. Fieguth and S. Wesolkowski, "Highlight and shading invariant color image segmentation using simulated annealing," Energy Minimization Methods in Computer Vision and Pattern Recognition III (Sophia Antipolis, France, 2001), pp.314-327.
7. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 6, No.6 (1984).