Entropy Minimization for Shadow Removal

Graham D. Finlayson
University of East Anglia, Norwich, UK
graham@cmp.uea.ac.uk *

Mark S. Drew
Simon Fraser University, Vancouver, Canada
mark@cs.sfu.ca †

Cheng Lu
Micron Technology, Inc., San Jose, CA, U.S.A
chenglu@micron.com

Nov. 20, 2007

Abstract.

Recently, a method for removing shadows from colour images was developed [Finlayson, Hordley, Lu, and Drew, PAMI2006] that relies upon finding a special direction in a 2D chromaticity feature space. This “invariant direction” is that for which particular colour features, when projected into 1D, produce a greyscale image which is approximately invariant to intensity and colour of scene illumination. Thus shadows, which are in essence a particular type of lighting, are greatly attenuated. The main approach to finding this special angle is a camera calibration: a colour target is imaged under many different lights, and the direction that best makes colour patch images equal across illuminants is the invariant direction. Here, we take a different approach. In this work, instead of a camera calibration we aim at finding the invariant direction from evidence in the colour image itself. Specifically, we recognize that producing a 1D projection in the correct invariant direction will result in a 1D distribution of pixel values that have smaller entropy than projecting in the wrong direction. The reason is that the correct projection results in a probability distribution spike, for pixels all the same except differing by the lighting that produced their observed RGB values and therefore lying along a line with orientation equal to the invariant direction. Hence we seek that projection which produces a type of intrinsic, independent of lighting reflectance-information only image by minimizing entropy, and from there go on to remove shadows as previously. To be able to develop an effective description of the entropy-minimization task, we go over to the quadratic entropy, rather than Shannon’s definition. Replacing the observed pixels with a kernel density probability distribution, the quadratic entropy can be written as a very simple formulation, and can be evaluated using the efficient Fast Gauss Transform. The entropy, written in this embodiment, has the advantage that it is more insensitive to quantization than is the usual definition. The resulting algorithm is quite reliable, and the shadow removal step produces good shadow-free colour image results whenever strong shadow edges are present in the image. In most cases studied, entropy has a strong minimum for the invariant direction, revealing a new property of image formation.

Keywords: Illumination, reflectance, intrinsic images, illumination invariants, color, shadows, entropy, quadratic entropy

1. Introduction

Illumination conditions confound many computer vision algorithms. In particular, shadows in an image can cause segmentation, tracking, or recognition algorithms to fail. An illumination-invariant image is therefore of great utility in a wide range of problems in both computer vision and computer graphics.

An interesting feature of this problem is that *shadows* are approximately but accurately described as a *change of lighting* (Finlayson et al., 2002). Hence, it is possible to cast the problem of removing shadows from images into an equivalent statement about removing (and possibly later restoring) the effects of lighting in imagery.

* This work was supported by the Leverhulme Trust.
† This work was supported by the Natural Sciences and Engineering Research Council of Canada

Removal of outdoor cast shadows has been addressed before in the literature but typically not based on a photometric approach. For example, (Cho et al., 2005) uses background subtraction in YCbCr colour space and gradient thresholding to extract moving blobs in colour to detect cast shadows in traffic surveillance video. The objective is to find a moving blob that is not in a shadow region. In (Liu et al., 2006), another method based on time-varying data and a Gaussian mixture model uses the early video-based model in (Stauder et al., 1999) to partially remove effects of lighting from sequences (and cf. (Weiss, 2001)) but the proposed insensitivity to illumination depends on a slowly changing penumbra and does not work for strong shadows. In fact, a substantial amount of work on shadow detection has been concerned with moving shadows (Prati et al., 2003; Nadimi and Bhanu, 2004; Martel-Brisson and Zaccarin, 2007), whereas here we concentrate on single still images. In this paper we focus on a physically-based rather than image-processing approach in order to gain understanding of the underlying image formation process.

A method was recently devised (Finlayson et al., 2002; Finlayson and Hordley, 2001; Finlayson and Drew, 2001; Drew et al., 2003; Finlayson et al., 2006) for the recovery of an invariant image from a 3-band colour image. The invariant image, originally 1D greyscale but subsequently derived as a 2D chromaticity, is independent of lighting, and also has shading removed: it forms a type of *intrinsic image*, independent of illumination conditions, that may be used as a guide in recovering colour images that are independent of illumination conditions. While the essential definition of an intrinsic image is one that captures full reflectance information (Barrow and Tenenbaum, 1978), including albedo information, here we claim only to capture only chromaticity information, not full reflectance. Nevertheless, invariance to illuminant colour and intensity means that such images are free of shadows as well, to a good degree (Finlayson et al., 2006). Although shadow removal is not always perfect, the effect of shadows is so greatly attenuated that many algorithms can easily benefit from the new method; e.g., a shadow-free active contour based tracking method shows that the snake can without difficulty follow an object and not its shadow, using the new approach to illumination colour invariance (Jiang and Drew, 2003; Jiang and Drew, 2007). In place of standard luminance images used in vision, if in an application the effects of lighting would usefully be removed then arguably the greyscale version of the invariant image should be used instead.

The method works in a very simple way: Suppose we form chromaticity band-ratios, e.g., G/R, B/G for a colour 3-band RGB image, and suppose we further take logarithms. An interesting feature to note is that, under simplifying assumptions set out below, the scatterplot values for pixels from the same surface, but under different lighting fall on a straight line; and every such line, for different surfaces, has the same slope. This remarkable fact still hold true approximately even when the guiding, simplifying assumptions are broken. Since shadowing is a result of a difference in lighting, we can use this physics-based insight to devise a shadow-removal scheme. This paper uses evidence internal to any particular image, based on an entropy measure, to find the slope of such lines. Projection orthogonal to this special direction results in a 1D greyscale image that has shadows approximately removed. We also derive a 2D colour version of such an invariant image.

The method devised finds an intrinsic reflectivity image motivated by the assumptions of Lambertian reflectance, approximately Planckian lighting, and fairly narrowband camera sensors. Nevertheless, the method still works well when these assumptions do not hold. A crucial piece of information is the angle for an "invariant direction" in a log-chromaticity space. Originally, this information was gleaned via a preliminary calibration routine, using the camera involved to capture images of a colour target under different lights. Subsequently, it was shown in principle (Finlayson et al., 2004) that we can in fact dispense with the calibration step, by recognizing a simple but important fact: the correct projection is *that which minimizes entropy* in the resulting invariant greyscale image. In this paper, the entropy based method is examined in detail, and in order to carry out an efficient search over smooth values that are not subject to quantization problems, we replace the Shannon's entropy measure, used previously, by a Quadratic Entropy measure such that a Gaussian mixture model of the probability density function (pdf) produces an analytic formula. We show that such quadratic entropy values are much smoother and usually produce only a single minimum, making this approach the most efficient. The quadratic entropy can be evaluated in linear time using a Fast Gauss Transform, leading to a simple method for finding the invariant direction. Shadow removal in full colour, by means of comparing edges in the original and in the invariant image and then subsequent re-integration, follows.

The paper is organized as follows. In §2, we briefly recapitulate the motivation for a projection-based definition of an illuminant invariant, and set out the relevant equations. Section 3 looks at how the entropy minimization scheme plays out for a set of synthetic colour patches, on the one hand, and then for a set of actual paint patches in a calibration chart. Section 4 considers the issue of how an effective entropy-minimization algorithm should proceed, and argues that an efficient approach is possible, based on replacing the definition of entropy by the quadratic form of Renyi's entropy. Finally, we apply the method devised to unsourced images, from unknown cameras under unknown lighting, with unknown processing having been applied. Results are again strikingly good, leading us to conclude, in §8, that the method indeed holds great promise for developing a stand-alone approach to removing shadows from (and therefore conceivably re-lighting) any image, e.g. images from consumer cameras.

2. The Invariant Image

2.1. Log-Chromaticity Projection and Entropy Minimization

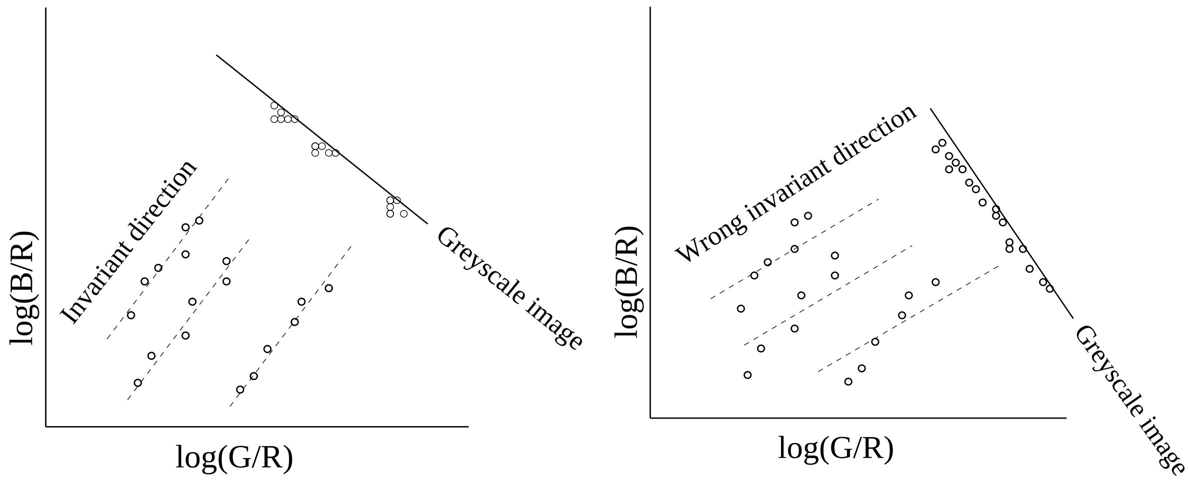
In order to motivate the study, we first briefly set out the strategy for developing an illumination invariant image, and the rationale determining entropy minimization as the key insight for finding such an image.

Consider a calibration scheme, for a particular colour camera, wherein a target composed of coloured patches (or just images of a rather colourful scene) are imaged under different illuminants — the more illuminants the better. Then knowledge that these are registered images of the same scene, under differing lighting, is put to use by plotting the capture RGB values, for each of the patches used, as the lighting changes. If pixels are first transformed from 3D RGB triples into a 2D band-ratio chromaticity colour space, $\{G/R, B/R\}$ say, and then logarithms are taken, the values across different illuminants tend to fall on *straight lines* in a 2D scatter plot. And in fact all such lines are *parallel*, for a given camera (Finlayson and Hordley, 2001), as illustrated in Fig. 1(a).

So change of illumination simply amounts to movement along such a line. Thus it is straightforward to devise a 1D, greyscale, illumination-invariant image by projecting the 2D chromaticity points into a direction perpendicular to all such lines. The result is hence a greyscale image that is independent of lighting, and is, therefore, a type of intrinsic image (Barrow and Tenenbaum, 1978) that portrays only the inherent reflectance properties in the scene. Since shadows are mostly due to change in the illuminant intensity and colour — i.e., differing lighting — such an image also has shadows removed.

Below, we discuss the restrictions on this straight-line model, but it may be useful to look at shadows and lighting colour in an example. Fig. 2(a) shows a typical consumer-grade camera TIFF image, with a strong shadow present. Here, the image processing software applied is typically aimed at a "preferred" (i.e., pleasing) rendition, rather than photometric accuracy, and the number of processing steps in the camera software can be substantial (Ramanath et al., 2005).

The standard definition of *chromaticity*, i.e., colour contents without intensity, is defined in an L_1 norm: $\boldsymbol{r} = \{r, g, b\} \equiv \{R, G, B\}/(R + G + B)$. Fig. 2(b) shows this colour content for the image.



(a) (b)

Figure 1. Intuition for finding best direction via minimizing the entropy. (a); Log-ratio feature space values for paint patches fall along parallel lines, as lighting is changed. Each patch corresponds to a single probability peak when projected in the direction orthogonal to the direction of lighting change. (b): Projecting in the wrong direction leads to a 1D pdf which is less peaked, and hence of larger entropy.

Notice that the colour of the shadow is basically a deep blue; since this is an outdoor shot on a clear day, this is not surprising in that the light for shadowed pixels is mostly from the sky dome, whereas light for non-shadowed pixels is comprised of both sky-light as well as direct sunlight. Thus shadowing is seen to be an effect due to change of lighting colour as well as intensity.

The invariant greyscale is shown in Fig. 2(e) where we see the shadow is no longer present. In (Drew et al., 2003), a 2D-colour chromaticity version of the invariant image, as in Fig. 2(f), is recovered by projecting orthogonal to the lighting direction and keeping the 2D colour location information, and also putting back an appropriate amount of lighting along the lighting direction. While 2(f) looks flat and the colours somewhat false, intrinsic images created this way are useful in computer vision: e.g. see (Jiang and Drew, 2003; Jiang and Drew, 2007).

We can use the greyscale or the pseudo-colour invariant as a guide that allows us to determine which colours in the original, RGB, colour image are intrinsic to the scene or are simply artifacts of the shadows due to lighting. Forming the gradient of the image's colour channels, we can guide a thresholding step via the difference between edges in the original and in the invariant image (Finlayson et al., 2002; Finlayson et al., 2006). Forming a further derivative, and then integrating back, we can produce a result that is a 3-band *colour* image which contains all the original salient information in the image, except that the shadows are removed, as in Fig. 2(g). Although this method is based on the invariant image, which has shading removed, nonetheless its output is a colour image, including shading. It is worth pointing out that we have found that in implementing this process, a 2D-colour, chromaticity, illumination-invariant image is more well-behaved than the greyscale variant, and thus gives slightly better shadow removal.

Of course these applications sit on top of a well calibrated imaging system. We measure how the camera responds to light and find the invariant direction accordingly. However, often in vision tasks we do not know the provenance of the images or even if we do have a calibrated camera this calibration does change over time. Thus, the problem we consider, and solve, in this paper is the determination of the invariant image from *unsourced* imagery — images that arise from cameras that are *not calibrated*.

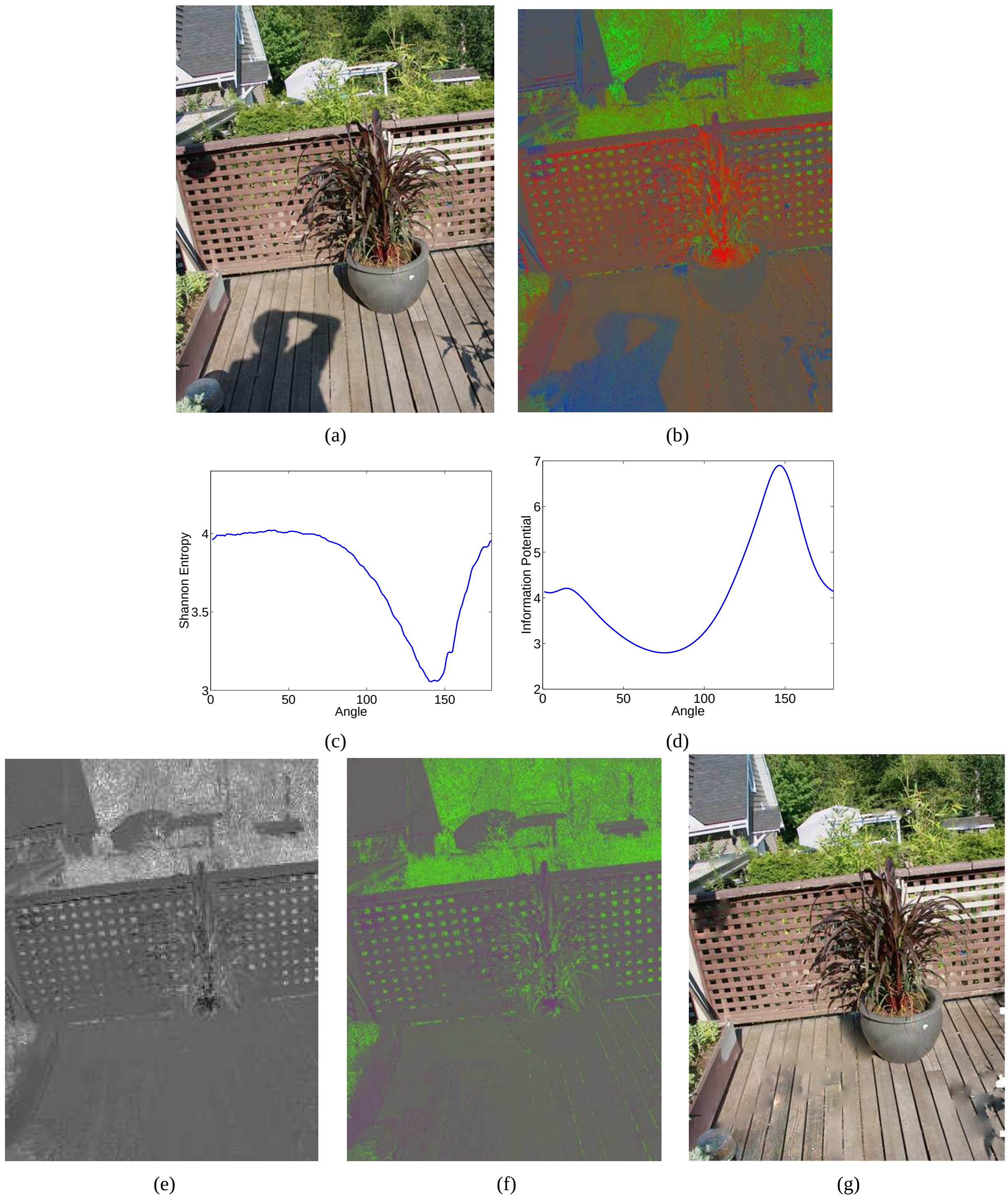


Figure 2. Colour and intensity shift in shadows: (a): Original image; (b): L_1 chromaticity image; (c): Shannon's entropy plot (we seek the minimum); (d): quadratic entropy plot (we seek the maximum of the quantity plotted); (e): greyscale 1D invariant; (f): 2D invariant L_1 chromaticity; (g): re-integrated 3D colour image.

The input is a colour image with unknown provenance, one that includes shadows, and the output is the invariant chromaticity version, with shading and shadows removed.

The fundamental idea in this paper is the observation that, without having to image a scene under more than a single illuminant, projecting in the correct direction minimizes the entropy in the resulting greyscale image. The intuition behind this statement is evident in the calibration situation, with a set of colour patches under changing lighting. Projecting onto a line perpendicular to the set of straight lines, we end up with a 1D pdf that is concentrated in peaks, as in Fig. 1(a). In a set of real images of colour patches, we indeed see a set of peaks, each well separated from the others and corresponding to a single colour patch. On the other hand, if we instead project in some other direction, as in Fig. 1(b), then instead of pixels located in sharp peaks of occurrence we expect the distribution of pixels along our 1D projection line to be spread out. In terms of histograms, in the first instance, in which we guess the correct direction and then project, we see a distribution with a set of sharp peaks, with resulting low entropy. In the second instance we instead see a broader histogram, with resulting higher entropy.

But does this idea apply in real images? We have found that, for almost every image considered that does indeed involve shadows, entropy has a strong minimum near the correct invariant direction. Changing lighting is automatically provided by the shadows in the image themselves.

Nevertheless, we have found that the method of calculating the entropy is important. Fig. 2(c) shows Shannon's entropy, for the projected, greyscale image when features are projected over angles from 1° to 180°. This entropy is calculated by choosing a bin-width, and then quantizing pixel feature values using a histogram, normalizing the histogram, and forming the standard quantity describing the entropy (see, e.g., (Li and Drew, 2004)). But this calculation can sometimes be quite sensitive to the bin-width, as shown in Fig. 3(b) (for quite a pathological case). We have found that such entropy plots can in fact have many local minima; this is discussed further in §4.1 below. Instead, a quadratic entropy plot, discussed below, is usually a good deal smoother since it is founded on a Gaussian kernel density distribution, and most often has a single strong maximum (of the quantity that must be maximized, in this case), making for a simple optimization to find the maximum. Further, quadratic entropy, which is the logarithm of an integral, is simply related to *Information Potential*. Specifically, information potential is the exponent of the negative of quadratic entropy and so minimum quadratic entropy corresponds to maximum information potential. This is an important point as the majority of the results derived below are for information potential; though, the reader should understand that information potential and quadratic entropy are simply related.

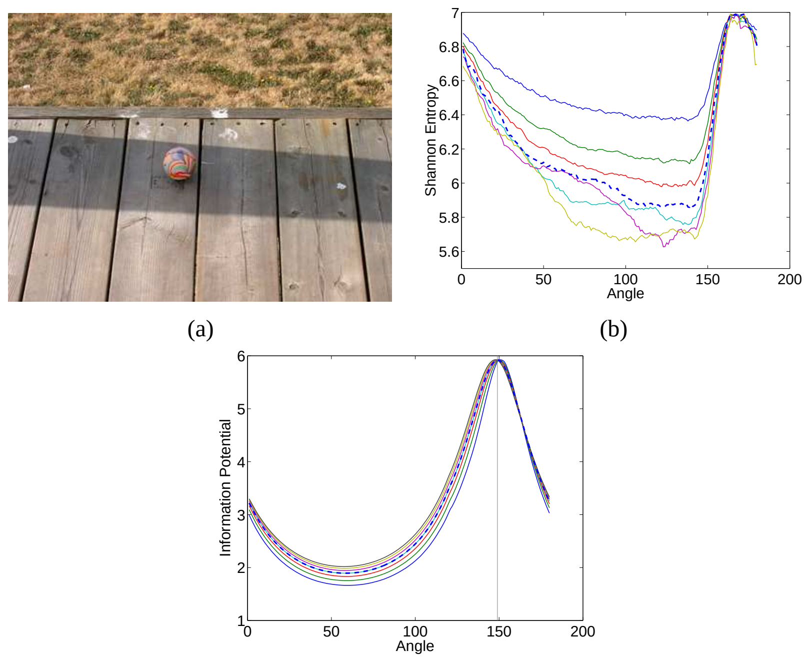
In Fig. 3(c), we show the information potential using a range of different bandwidth parameters, and notice that the maximum is quite insensitive to the bandwidth. Consequently, in this paper we go over to this definition of the entropy, as shown in Fig. 2(d) for the initial image Fig. 2(a) (here, we look for a maximum of the quantity plotted).

In §2.2, we now briefly summarize the set of theoretical assumptions regarding the problem of lighting change in imagery that lead to the straight-line hypothesis.

2.2. Theory of Invariant Image Formation

2.2.1. *Planckian Lighting, Lambertian Surfaces, Narrowband Camera*

Suppose we consider a fairly narrow-band camera, with three sensors, Red, Green, and Blue, as in Fig. 4(a) (these are sensor curves for the Sony DXC930 camera). Now if we image a set of coloured Lambertian surfaces under a particular Planckian light, e.g. in a controlled light box, then for each pixel the log of chromaticity band-ratios, say $\{\log(R/G), \log(B/G)\}$, appears as a dot in a 2D plot. Chromaticity removes shading, for Lambertian reflectances under orthography, so every pixel in each patch is approximately collapsed into the same dot (no matter if the surface is curved).



(a) (b)

(c)

Figure 3. Effect of quantization on Shannon's entropy: (a): Original image; (b): Shannon's entropy plot, with changing bin-width (we seek the minimum) – the normative bin-width value given below by eq. (15) is shown dashed, and the other curves are for multipliers of this width from 0.1 to 2.0, mapped to equal maxima; (c): quadratic entropy plot with bandwidth multiplied by factors from 0.1 to 1.9 by 0.3, with 1.0 shown dashed (note that we seek the maximum of the quantity plotted).

For example, Fig. 5(b) illustrates the log-chromaticities for the 24 surfaces of the Macbeth ColorChecker Chart shown in Fig. 5(a). The plot shows 19 distinct clusters of points — each cluster corresponds to chromaticities from a single patch (there are 19 clusters rather than 24 since the patches in the last row of the chart are all neutral in colour and so have the same chromaticity). Fig. 5(c) shows the plot of the median 2D log-chromaticities for 6 of the Macbeth surfaces under 14 different Planckians — we see a set of parallel approximately straight lines.

For narrow-band sensors (or spectrally-sharpened ones (Finlayson et al., 1994; Drew et al., 2002)), and for Planckian lights — or lights such as Daylights which behave as if they were Planckian in that their chromaticity is very close to the Planckian locus — as the correlated colour temperature T that characterizes the illuminant changes, the log-chromaticity colour 2-vector does indeed move along an approximately straight line which is independent of the magnitude and position of the lighting. (Note that the invariant direction is different for each camera.) Further, in manufacturing artificial lights, the colour rendering properties of lights are calculated using a CIE standard methodology (CIE, 1995). According to this methodology, illuminants that are far from the Planckian locus render less well than those that are close. As such most commercial lights have chromaticities close to the Planckian locus and for all commercial lights tested we discover more or less the same intrinsic reflectance image.

Let's recapitulate how linear behaviour with lighting change results from the assumptions of Planckian lighting, Lambertian surfaces, and a narrowband camera. Consider the RGB colour $\boldsymbol{R}$ formed at a pixel, for illumination with spectral power distribution $E(\lambda)$ impinging on a surface with surface spectral reflectance function $S(\lambda)$. If the three camera sensor sensitivity functions form a set $\boldsymbol{Q}\ (\lambda)$, then we have

$$R_k \ = \ \sigma \int E(\lambda)S(\lambda)Q_k(\lambda)d\lambda \ , \ k = R, G, B \ , \tag{1}$$

where σ is Lambertian shading: surface normal dotted into illumination direction.

If the camera sensor $Q_k(\lambda)$ is exactly a Dirac delta function $Q_k(\lambda) = q_k\delta(\lambda - \lambda_k)$, then eq. (1) becomes simply

$$R_k \ = \ \sigma \ E(\lambda_k)S(\lambda_k)q_k \ . \tag{2}$$

Now suppose lighting can be approximated by Planck's law, in Wien's approximation (Wyszecki and Stiles, 1982):

$$E(\lambda, T) \ \simeq \ I \ k_1 \lambda^{-5} e^{-\frac{k_2}{T\lambda}} \ , \tag{3}$$

with constants k_1 and k_2. Temperature T characterizes the lighting colour and I gives the overall light intensity.

In this approximation, from (2) the RGB colour $R_k, \ k = 1 \ldots 3$, is simply given by

$$R_k \ = \ \sigma \ I \ k_1 \lambda_k^{-5} e^{-\frac{k_2}{T\lambda_k}} S(\lambda_k) q_k \ . \tag{4}$$

Let us now form the band-ratio 2-vector chromaticities $\boldsymbol{c}$,

$$c_k \ = \ R_k / R_p \ , \tag{5}$$

where p is one of the channels and $k = 1, 2$ indexes over the remaining responses. For example, we could use $p = 1$ (i.e., divide by Red) and so calculate $c_1 = G/R$ and $c_2 = B/R$. We see from eq. (4) that forming this chromaticity effectively removes intensity and shading information. If we now form the log of (5), with $s_k \equiv k_1 \lambda_k^{-5} S(\lambda_k) q_k$ and $e_k \equiv -k_2/\lambda_k$ we obtain

$$\rho_k \ \equiv \ \log(c_k) \ = \ \log(s_k/s_p) + (e_k - e_p)/T \ . \tag{6}$$

Eq. (6) is a straight line parameterized by T. Notice that the 2-vector direction $(e_k - e_p)$ is *independent of the surface*, although the line for a particular surface has offset that depends on s_k. Every such line is parallel, with slope dictated by $(e_k - e_p)$.

An invariant image can be formed by projecting these 2D logs of band-ratio chromaticity $\rho_k, k = 1, 2$, into the direction $\boldsymbol{e}^{\perp}$ orthogonal to the vector $\boldsymbol{e} \ \equiv (e_k - e_p)$. The result of this projection is a single scalar which we then code as a greyscale value.

We go on in §3.2.5 to generate a 2-colour chromaticity image from the greyscale version. The images thus generated are "intrinsic" in the sense that they capture reflectance information independent of lighting. However, they are not full reflectance-only images (as specified in (Barrow and Tenenbaum, 1978)), since they bear only chromaticity information, not albedo.

Since the method stems from a Planckian illumination model, it is worth asking whether the combination of lights when not in shadow – sunlight plus skylight – breaks the model. In fact, the sum of two Planckian lights is not Planckian. However, since the Planckian locus is in fact a very shallow curve (Wyszecki and Stiles, 1982), the combination is almost Planckian. To investigate the effect of this combining of lights on the theoretical underpinnings of the method, consider the synthetic scene in Fig. 6(a). This depicts two hemispheres on a plane viewed from $(0, 0, 1)$, shaded via a full-spectrum raycaster (cf. (Bergner et al., 2009)). Here the left sphere surface material is Macbeth chart patch #4, "olive green", the right sphere is #6, "bluish green", and the plane is #2, "light skin". Lighting is a

Planckian with temperature T=10,500°K in the shade (blue lighting), and a combination of that light plus a Planckian with T=2,500°K (red lighting) outside the shadows. The camera sensors used are delta-functions at λ_k={650nm, 540nm, 450nm}. This image illustrates the different colours in shadow and non-shadow regions. Using a projection of log band-ratio chromaticities $R/G, B/G$, Fig. 6(c) shows the greyscale invariant delivered by the proposed method, and Fig. 6(d) shows the corresponding derived L_1 chromaticity, invariant image $\{R, G, B\}/(R + G + B)$. We notice that while the greyscale image is approximately independent of the lighting, it is not perfect. This is, as we supposed, due to the fact that a combination of Plankians is not itself perfectly Planckian. So in fact even in circumstances in which the theoretical model is perfectly obeyed, the method will likely not deliver a perfect result. Moreover the chromaticity invariant delivered is not precisely as we expect in that there are gradual colour changes across each sphere (although these would disappear if had only a single illuminant impinging on each of shadow and non-shadow pixels). However, indeed both attached and cast shadows are essentially removed. This example serves to show that an invariant image can be of use in establishing which edges correspond to material changes and which to lighting, at least well enough as a vehicle for shadow removal.

Before light is added back to such images (below), they are a type of intrinsic image bearing reflectivity information only. In §3.2 we recover an approximate intrinsic RGB reflectivity, akin to that in (Tappen et al., 2003; Tappen et al., 2005) but with a considerably less demanding algorithm: (Tappen et al., 2005) classified image gradients as illumination or reflectance edges depending on their direction and magnitude and, in cases of ambiguity, on other edges in the neighbourhood. Work on recovering the intrinsic reflectance and illumination of a scene flows in part from early work on Retinex (Land and McCann, 1971), and (Tappen et al., 2005) is a sophisticated development in this stream. Note that an important qualification of the domain of the present method is that whereas the method in (Tappen et al., 2005) works on either greyscale images or colour ones, the method set out here depends on colour. Allied efforts, especially in the domain of Computational Photography, have considered light mixtures (e.g., (Hsu et al., 2008)) or colour-filtered images (e.g., (Finlayson et al., 2007)).

Clearly, if we have the opportunity to calibrate our camera, then we can determine the invariant 2-vector direction $\boldsymbol{e}$. However, if we have only a single image, then we do not have the opportunity to calibrate. Nevertheless we would still like to be able to remove shadows from any image. We show in the next Section that the automatic determination of the invariant direction is indeed possible, with entropy minimization being the correct mechanism.

3. Intrinsic Images by Entropy Minimization

Here, we would like to do away with the necessity of a calibration step to gain foreknowledge of the invariant direction. We begin in §3.1 by creating a synthetic "image" that consists of a great many colour patches. Since the image is synthetic, we in fact do know the ground truth invariant direction. Examining the question of how to *recover* this direction from a single image, with no prior information, we show that minimizing the entropy provides a very strong indicator for determining the correct projection. This result provides a proof in principle for the entropy-minimizing method.

The idea being examined in this section is thus as follows: Suppose that in a single image various illuminants impinge on several paint patches. Here we use synthetic Planckian lights in order to see that the underlying theory behaves as expected. The question examined is then whether we can remove the effects of lighting from this single, synthetic image.

But how do we fare with a real camera? In §3.2 we consider a set of captured colour-patch images, taken with a known camera. Since we control the camera, and the target, we can establish the invariant direction. Then comparing to the direction recovered using entropy minimization, we find that not only is the direction of projection recovered correct (within 3 degrees), but also the minimum is global and is a very strong signal.

3.1. Entropy Minimization

If we wished to find the minimum-variance direction for lines that are formed in log-chromaticity space as the light changes, we would need to know which points fall on which lines. But what if we did not have that information?

To test the idea that entropy minimization gives an intrinsic image, suppose we start with a theoretical Dirac-delta sensor camera, as in Fig. 4(b). Now let us synthesize an "image" that consists of many measured natural surface reflectance functions interacting with many lights, in turn, and then imaged by our theoretical camera. As a test, we use the reflectance data $S(\lambda)$ for 170 natural objects, measured by Vrhel et al. (Vrhel et al., 1994). For lights, we use the 9 Planckian illuminants $E(\lambda)$ with T from 2,500° to 10,500° Kelvin with interval of 1,000°. Thus we have an image composed of 1,530 different illuminant-reflectance colour signal spectral products. This image is shown in Fig. 4(c). From left to right we have the 170 different reflectances. And, top to bottom the illuminants starting from the reddish 2500K light to the bluish 10,500K. We clearly see a colour shift from reddish to bluish. A close-up of the last 7 reflectance patches is shown in Fig. 4(d).

If we form chromaticities (actually we use geometric mean chromaticities defined in eq. (7) below instead of simple band ratios, in order to not favour one particular colour channel), then taking logarithms and plotting we have 9 points (for our 9 lights) for every colour patch. Subtracting the mean from each 9-point set, all lines go through the origin. Then it is trivial to find the best direction describing all 170 lines via applying the Singular Value Decomposition method to this data. The best direction line is found at angle $68.89°$. And in fact we know from the theoretical definition of $(e_k - e_p)$ that this angle is correct, for this camera. This verifies the straight-line equation (6), in this situation where the camera and surfaces exactly obey our assumptions. This exercise amounts to a calibration of our theoretical camera in terms of the invariant direction.

But now suppose we do not know that the best angle at which to project our theoretical data is orthogonal to about $69°$ — how can we recover this information? Clearly, in this theoretical situation, the intuition displayed in Fig. 1 can be brought into play by simply traversing all possible projection angles that produce a projection direction $\boldsymbol{e}^{\perp}$: *the direction that generates an invariant image with minimum entropy is the correct angle*.

To carry out such a comparison, we simply rotate from 0° to 180° and project the log-chromaticity image 2-vector $\boldsymbol{\rho}$ into that direction. To utilize Shannon's definition of entropy, we can form a histogram as a quantization mechanism. We must decide on a bin size, and for now we simply use 64 equally-spaced bins. And finally the entropy η is calculated: the histogram is divided by the sum of the bin counts to form probabilities p_i and, for bins that are occupied, the sum $\eta = \sum_{i=1}^{64} -p_i \log_2 p_i$ is formed.

Fig. 4(e) shows a plot of angle versus this particular entropy measure, for the synthetic image. As can be seen, the correct angle of $159 = 90 + 69°$ is accurately determined (within a degree). When we go over to a quadratic entropy, explained below in §5, we see from Fig. 4(f) that this definition of entropy also gives the correct answer (with zero error, for this case).

Fig. 4(g) shows the actual invariant greyscale "image" for these theoretical colour patches, given by exponentiating the projected log-image, with a close-up of the last 7 reflectance patches shown in Fig. 4(h). As we go from left to right across Fig. 4(f) we change reflectance. From top to bottom we have pixels calculated with respect to different lights. Recall that Fig. 4(c) shows the 170 reflectances (left to right) under the 9 Planckians (top to bottom). As opposed to the invariant image in Fig. 4(g), notice how all colours become progressively bluer top to bottom. The figure Fig. 4(g) shows the invariant image coded as greyscale, and there is zero variation from top to bottom. Yet the greyscale value does change from left to right, along our 170 surfaces. So, in summary, Fig. 4(g) tells us that the same surface has the same invariant across lights but different surfaces have different invariants (and so the intrinsic image conveys useful reflectance information).

Next, we consider an image formed from *measured* values of a colour target.

3.2. Calibration Images vs. Entropy Minimization

Now let us investigate how this theoretical method can be used for real, non-synthetic values. We acquired calibration images of a Macbeth ColorChecker over 14 phases of daylight, with results displayed in Fig. 7. (These images were taken with an experimental HP 912 digital camera with the normal nonlinear processing software disabled, but in fact the entropy minimum phenomenon persists regardless of the processing.)

3.2.1. *Geometric Mean Invariant Image*

From (4), we can remove σ and I via division by any colour channel: but which channel should we use? If we divide by red, but red happens to be everywhere small, as in a photo of greenery, say, outliers can occur. A better solution is to divide by the geometric mean (Finlayson and Drew, 2001), $\sqrt[3]{R \cdot G \cdot B}$. Then we still retain our straight line in log space, but do not favour one particular channel.

Thus we amend our definitions (5, 6) of chromaticity as follows:

$$c_k \;=\; R_k / \sqrt[3]{\Pi_{i=1}^{3} R_i}, \quad \equiv \; R_k / R_M, \tag{7}$$

and log version (Finlayson and Drew, 2001)

$$\begin{gathered} \rho_k \;=\; \log(c_k) \;=\; \log(s_k/s_M) \;+\; (e_k - e_M)/T, \; k = 1..3, \\ \text{with } \; s_k = k_1 \lambda_k^{-5} S(\lambda_k) q_k \;, s_M = \sqrt[3]{\Pi_{j=1}^{3} s_j} \;, e_k = -k_2/\lambda_k, \; e_M = -k_2/3 \sum_{j=1}^{p} \lambda_j, \end{gathered} \tag{8}$$

and for the moment we carry all three (thus nonindependent) components of chromaticity. (Broadband camera versions of eq. (8) are stated in (Finlayson and Drew, 2001).)

3.2.2. *Geometric Mean 2-D Chromaticity Space*

We should use a 2D representation that is appropriate for this log chromaticity space $\boldsymbol{\rho}$. We note that, in log space, $\boldsymbol{\rho}$ is orthogonal to $\boldsymbol{u} = 1/\sqrt{3}(1,1,1)^T$. That is, $\boldsymbol{\rho}$ lives on a plane orthogonal to $\boldsymbol{u}$, as in Fig. 8 (see (Finlayson et al., 2004)): $\boldsymbol{\rho} \cdot \boldsymbol{u} = 0$.

To characterize the 2D space, we can consider the projector $\boldsymbol{P}_u^{\perp}$ onto the plane. This projector $\boldsymbol{P}_u^{\perp}$ has two non-zero eigenvalues, and its decomposition reads

$$P_u^{\perp} \;=\; I - \boldsymbol{u}\,\boldsymbol{u}^T \;=\; \boldsymbol{U}^T \boldsymbol{U}\,, \tag{9}$$

where $\boldsymbol{U}$ is a 2×3 orthogonal matrix. $\boldsymbol{U}$ rotates 3-vectors $\boldsymbol{\rho}$ into a coordinate system *in the plane*:

$$\boldsymbol{\chi} \;\equiv\; \boldsymbol{U}\,\boldsymbol{\rho}\,, \quad \boldsymbol{\chi} \;\text{ is } 2 \times 1. \tag{10}$$

Straight lines in $\boldsymbol{\rho}$ are still straight in $\boldsymbol{\chi}$. For example, we could take $v_1 = (1/\sqrt{2}; -1/\sqrt{2}; 0)^T$, $v_2 = (1/\sqrt{6}; 1/\sqrt{6}; -2/\sqrt{6})^T$), and $\boldsymbol{U} = [v_1, v_2]^T$.

In the $\{\chi_1, \chi_2\}$ plane, we are now back to a situation similar to that in Fig. 1: we must find the correct direction θ in which to project, in the plane, such that the entropy for the marginal distribution along a 1D projection line orthogonal to the lighting direction is minimized. The greyscale image $\mathcal{I}$ along this line is formed via

$$\mathcal{I} \;=\; \chi_1 \cos\theta \,+\, \chi_2 \sin\theta \tag{11}$$

and Shannon's entropy is given by

$$\eta \;=\; -\sum_i p_i(\mathcal{I})\, \log(p_i(\mathcal{I})). \tag{12}$$

We shall see below, in §5, how instead the quadratic entropy allows us to inherit the marginal pdf from the 2D pdf, as a function of θ.

3.2.3. *Test of Main Idea*

Thus the heart of this test of the entropy-minimization idea using real, measured paint-patch data, is as follows:

(a) Form a 2D log-chromaticity representation of the image.

(b) for $\theta \;=\; 1..180$

  (i) Form greyscale image $\mathcal{I}$: the projection onto 1D direction.

  (ii) Calculate entropy.

  (iii) Min-entropy direction is correct projection for shadow removal.

We would like an actual algorithm to proceed faster than this type of brute force search, of course, and that issue is addressed in §5.

3.2.4. *3-Vector Representation*

After we find θ, we can go back to a 3-vector representation of points on the projection line. We project 2D points onto a line via a 2×2 projector $\boldsymbol{P}_\theta$: if $\boldsymbol{\Theta} \;=(\cos\theta, \sin\theta)^T$, then $\boldsymbol{P}_\theta$=$\boldsymbol{\Theta}\,\boldsymbol{\Theta}^{\,T}$. We form the projected 2-vector $\boldsymbol{\chi}_{\,\theta}$ via $\boldsymbol{\chi}_{\,\theta} \;=\; \boldsymbol{P}_{\,\theta}\boldsymbol{\chi}$ and then go back to an estimate (indicated by a tilde) of 3D $\boldsymbol{\rho}$ and $\boldsymbol{c}$ via $\widetilde{\boldsymbol{\rho}} \;= \boldsymbol{U}^{\,T}\boldsymbol{\chi}_{\,\theta}$, $\;\widetilde{\boldsymbol{c}} \;= \exp(\widetilde{\boldsymbol{\rho}}\,)$. For display, we would like to move from an intrinsic image, governed by reflectivity, to one that includes illumination (cf. (Drew et al., 2003)). So before applying $\boldsymbol{U}^{\,T}$ we add back enough $\boldsymbol{e}$ so that the median of the brightest 1% of the pixels has the 2D chromaticity of the original image: $\boldsymbol{\chi}_{\,\theta} \rightarrow \boldsymbol{\chi}_{\,\theta} + \boldsymbol{\chi}_{\,extralight}$.

3.2.5. *Stable Chromaticity Image*

Once we have an estimate $\widetilde{\boldsymbol{c}}$ of the geometric-mean chromaticity (7), we can also go over to the usual L_1-based chromaticity $\{r, g, b\}$, defined as

$$\boldsymbol{r} \;\;=\; \{r, g, b\} \;=\; \{R, G, B\}/(R+G+B),\; r+g+b \equiv 1\;. \tag{13}$$

This is the most familiar representation of colour independent of magnitude (column 2 of Fig. 8 shows the L_1 chromaticity for colour images). To obtain L_1 chromaticity $\boldsymbol{r}$ from our estimate of $\boldsymbol{c}$, we simply take

$$\widetilde{\boldsymbol{r}} \;\;=\; \widetilde{\boldsymbol{c}}\,/\sum_{k=1}^{3}\widetilde{c}_k\;. \tag{14}$$

Since $\boldsymbol{r}$ is bounded $\in [0,1]$, invariant images in $\boldsymbol{r}$ are better-behaved than is $\mathcal{I}$. The greyscale image $\mathcal{I}$ for this test using images of a colour target is shown in Fig. 7(d), and the L_1 chromaticity version $\widetilde{\boldsymbol{r}}$, as per eq. (14), is shown in Fig. 7(e). We note that both greyscale and colour invariant images are stable across illuminants. The colour range for the 2D colour invariant is of course reduced compared to that of an unprojected colour target.

3.2.6. *Entropy Minimization — Strong Indicator*

From the calibration technique described in section 3.1 we in fact already know the *correct* characteristic direction in which to project to attenuate illumination effects: for the HP-912 camera, this angle turns out to be 158.5°.

We find that entropy minimization gives a close approximation of this result: 161° for both Shannon's definition of entropy and the quadratic entropy variant. First, transforming to 2D chromaticity coordinates $\boldsymbol{\chi}$, the colour patches of the target do form a scatterplot with approximately parallel lines, in Fig. 7(a). We compose an image consisting of a montage of median pixels for all 24 colour patches and 14 lights. The calculation of entropy carried out for this image gives a very strong extremum, shown in Figs. 7(b,c), and excellent greyscale $\mathcal{I}$ invariant to lighting, and chromaticity invariant, $\boldsymbol{r}$ in Figs. 7(d,e).

This completes both the theoretical and a controlled-experiment justification of the main idea — finding the invariant projection direction by entropy minimization. In the next section, we examine the issues involved when we extend this laboratory success to the realm of ordinary, non-calibration images. To dispense with a brute-force search over all angles, we also need a disciplined search mechanism, and we see that this is provided by the quadratic entropy measure, with implementation by Fast Gauss Transform.

4. Intrinsic Image Recovery Algorithm

4.1. Shannon's Entropy and Quantization

Real images are noisy and might not provide such a clean picture as in our theoretical and testing images above. As well, we must decide on a quantization procedure if we wish to utilize Shannon's definition of entropy.

4.1.1. *Quantization Problem*

Consider the colour image in Fig. 3(a): a colourful ball on a wooden deck is in a shadow cast by strong sunlight. To find the minimum entropy, we again examine projections $\mathcal{I}$ over angles 0° to 180°, for log-chromaticities $\boldsymbol{\chi}$ formed according to eqs. (7), (8), and (10). For each angle, we project the log-chromaticity, and then determine the entropy (12). However, the nature of the data, for real images, presents an inherent problem. Since we are considering ratios, we can expect noise to possibly be enhanced (although this is mitigated by the sum in eq. (13)). To begin with, therefore, we apply Gaussian smoothing to the original image colour channels. But even so, we expect that some ratios may be large. So the question remains as to what we should use as the range, and number of bins, in a histogram of a projected greyscale image $\mathcal{I}$. Using the usual, Shannon, definition of entropy, we cannot escape this quantization issue. However, the alternative Quadratic Entropy measure, used below, largely circumvents this issue by utilizing a different, kernel density driven non-parametric estimate of the pdf that automatically incorporates smoothness. We still have to choose a bandwidth parameter, but the resulting quantity is relatively independent of this choice.

We calculate Shannon's entropy by approximating the pdf with a histogram over projected 1D greyscale values. To form an appropriate bin width, we utilize Scott's Rule (Scott, 1992):

$$\text{bin\_width} \;=\; 3.5\,\text{std}(\text{projected data})\,N^{-1/3} \tag{15}$$

where N is the size of the invariant image data, for the current angle. Since there may be outlier ratios, we use the middle values only, i.e., the middle 90% of the data, to form a histogram. And the scale of the entropy for each projection is the same, since the number of bins is about the same: if we draw the samples from a Gaussian population then the first 3 standard deviations, say, from the mean plus overload at the boundaries describe all the data, and the number of bins is then just proportional to $N^{1/3}$, which is the same for every projection.

The entropy calculated is shown in Fig. 3(b); but we find from varying the bin width in Fig. 3(b) around the value in eq. (15) that this entropy may be sensitive to the bin-size. We would like to develop a smoother version of the entropy, with a clearer indication of the minimum. As well, we would like to dispense with an exhaustive search over angles and go over to a smoother curve that facilitates efficient search for the minimum. We shall see next that the Quadratic Entropy curve is smooth and also generally has a single extremum. And a Fast Gauss Transform can produce each entropy evaluation in linear time.

Fig. 3(c) shows the Information Potential, derived in the next Section from the Quadratic Entropy. We see that in this case there is a much simpler curve shape, and local quantization effects are eliminated. The result for the resulting chromaticity invariant, and reconstructed shadow-free colour image is shown in Fig. 8.

5. Quadratic Entropy and Gauss Transform

Firstly we replace our pdf over 2D chromaticity coordinates by a Gaussian kernel density mixture to ensure that entropy is calculated over smooth values. If we go over to a Quadratic Entropy measure (a special case of Renyi's entropy (Renyi, 1987)), then the entropy takes on a very simple form.

In 1D, Renyi's entropy reads

$$\eta_\alpha \;=\; \frac{1}{1-\alpha}\log\int p^\alpha(x)dx, \quad \alpha \geq 0,\; \alpha \neq 1 \tag{16}$$

where $p(x)$ is the pdf. This measure is known to approach Shannon's entropy as α goes to 1.

For the special case of $\alpha = 2$ we have

$$\eta_{quadratic} \;=\; -\log\int_{-\infty}^{\infty} p^2(x)dx \tag{17}$$

Notice that the log is outside the integral, making for a much simpler evaluation.

For the purposes of optimization, we can simply drop the log, giving the so-called *information potential*,

$$V = \int p^2(x)dx \tag{18}$$

To see how the information potential is indeed related to minimizing the entropy, consider the parallel lines in (χ_1, χ_2)-space formed for two paint patches as illumination changes, in Fig. 9(a). Suppose data points are uniformly distributed along each line, so that the projected, marginal pdf is proportional to 1 except when the line projections overlap, when the pdf is proportional to 2. Clearly, there is a singularity when the lines project to zero length along the projection axis, at the minimum-entropy angle. Fig. 9(b) shows the theoretical value of V, as the projection angle changes. The information potential for real data usually also has a similar strong, single-maximum structure: the curve in Fig. 7(c) showing the information potential for the measured patch data in Fig. 7(a) has a maximum at $160.5°$, whereas the correct angle is $158.5°$. Compared to Fig. 7(b), the quadratic entropy has a much cleaner structure that facilitates a fast search by successive evaluation of the quadratic entropy over a few angles.

The quadratic entropy is explicitly evaluated using the Parzen window technique, below.

Parzen Window

The quadratic-entropy approach approximates the pdf $p(x)$ from its N samples a_i by a Parzen window estimator (Parzen, 1962), using Gaussian kernels G with mean a_i and variance s^2:

$$p(x) = \frac{1}{N}\sum_{i=1}^{N} G(a_i, s^2) \tag{19}$$

Since a convolution of two Gaussians is a Gaussian with variance equal to the sum of variances of the constituent Gaussians and mean given by the difference of individual means, the information potential V becomes simply (Xu and Principe, 1998)

$$V \;=\; \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} G(a_i - a_j, 2s^2) \;=\; \frac{1}{N^2}\,\frac{1}{\sqrt{2\pi(2s^2)}}\sum_i\sum_j e^{-\frac{(a_i-a_j)^2}{4s^2}} \tag{20}$$

Notice that now we can evaluate the entropy directly from the data, without the need to create a pdf first.

Now let us show how the 1D model above comes out of projecting 2D, (χ_1, χ_2) chromaticity-space data in a direction θ. For convenience let us write $x = \chi_1$, $y = \chi_2$. where X_i, Y_i is the 2D data and i indexes the image treating indices as a vector. If the 2D pdf is approximated as

$$p(x, y) \;=\; \frac{1}{N} \frac{1}{\sqrt{(2\pi)^2 s_1^2 s_2^2}} \sum_{i=1}^{N} \exp\left[-\frac{(x - X_i)^2}{2s_1^2}\right] \exp\left[-\frac{(y - Y_i)^2}{2s_2^2}\right] \;, \tag{21}$$

then to find the marginal probability density for this function along an axis μ, projected in the θ direction, we substitute $x = \mu \cos\theta + \nu \sin\theta$, $y = -\mu \sin\theta + \nu \cos\theta$. Also defining $M_i = X_i \cos\theta - Y_i \sin\theta$, after some algebra the projected marginal probability comes out to be

$$p_\theta(\mu) \;=\; \int_{\nu=-\infty}^{\infty} p\left(x(\mu, \nu),\, y(\mu, \nu)\right) d\nu \;=\; \frac{1}{N} \frac{1}{\sqrt{2\pi\tilde{s}^2}} \sum_i \exp\left[-\frac{(\mu - M_i)^2}{2\tilde{s}^2}\right] \tag{22}$$

with

$$\tilde{s}(\theta)^2 \;=\; s_1^2 \cos^2\theta \;+\; s_2^2 \sin^2\theta \tag{23}$$

We then use value $\tilde{s}$ that minimizes the asymptotic mean integrated squared error (AMISE) (Scott, 1992), given by

$$\tilde{s} \;=\; 1.06\, s(\theta)\, N^{-1/5} \tag{24}$$

Thus the information potential is

$$V \;=\; \int_{\mu=-\infty}^{\infty} \{f_\theta(\mu)\}^2\, d\mu \;=\; \frac{1}{N^2} \frac{1}{\sqrt{2\pi(2\tilde{s}^2)}} \sum_i \sum_j \exp\left[-\frac{(M_i - M_j)^2}{4\tilde{s}^2}\right] \tag{25}$$

Therefore, the information potential is given by a simple sum, along the projected axis.

The information potential V can be regarded as the total potential energy of the data set, with the Gaussians in the role of potential energy of data point M_i in the potential field of data point M_j. To minimize the entropy, we maximize this potential energy (Xu and Principe, 1998). In the context of data points that are free to move, the derivative of this potential is a force that drives data points into an equilibrium state such that the information potential takes on an extremum. This has been used for determining neural network parameter values that produce such optimized network outputs: the derivative of V with respect to the network parameters become derivatives of the output points, via the chain rule (Xu and Principe, 1998). In our application, the data is fixed, so we simply evaluate eq. (25).

In the Appendix, we show that the sum V can be calculated in linear time, using the Fast Gauss Transform. Quadratic entropy curves found are simple and smooth, and we have found that a maximum of V for real image data can be generated in just a few search steps. We found that on average the most critical and time-consuming step of the algorithm, namely the linear-time FGT, took about 4 microseconds per pixel (in Matlab on a single-core 3.0 Ghz P4 running Windows), or in other words some 0.5 seconds for a 300×400 image.

6. Re-Integrated Image Results

Using the re-integration method in (Finlayson et al., 2006), we can go on from our invariant image to recover a full-colour shadow-free image. The method introduced in (Finlayson et al., 2002) uses a shadow-edge map, derived from comparing the original edges to those in the greyscale invariant image.

In (Finlayson et al., 2006) we use edges from the invariant *chromaticity* image and compare to edges from a Mean-Shift (Comaniciu and Meer, 2002) processed original image. As well, rather than simply zeroing edges across the shadow edge, we use simple edge inpainting to grow edges into shadow-edge regions.

Regaining a full-colour image has two components: finding a shadow-edge mask, and then re-integrating. The first step is carried out by comparing edges in the Mean-Shift processed original image with the corresponding recovered invariant chromaticity image. We look for pixels that have edge values higher than a threshold for any channel in the original, and lower than another threshold in the invariant, shadow-free chromaticity. We identify these as shadow edges, and then thicken them using a morphological operator. For the second stage, for each log colour channel, we first grow simple gradient-based edges across the shadow-edge mask using iterative dilation of the mask and replacement of unknown derivative values by the mean of known ones. Then we form a second derivative, go to Fourier space, divide by the Laplacian operator transform, and go back to x, y space. Neumann boundary conditions leave an additive constant unknown in each recovered log colour, so we regress on the top brightness quartile of pixel values to arrive at the final resulting colour planes.

In our experiments, images show behaviour similar to that displayed in Fig. 3(c), with strong entropy minima (information potential maxima), and results quite free of shadows. Since the invariant image is basically shadow free and the re-integrated image is quite good, our intuition that minimization of entropy would lead to correct results is indeed justified.

Fig. 8 shows results from various images, from both calibrated and uncalibrated cameras, including consumer cameras. For all experiments we carried out, quadratic entropy minimization provided a strong guiding principle for removing shadows. Note that in some actual cameras, an entropy-minimization approach rather than a calibration is ideal for finding the invariant direction, since it is possible that even a change of camera settings or heating over the day in a surveillance situation can produce effectively different camera sensors. We point out to the reader that there is considerable variance in the recovered invariant angle direction over the set of images and cameras (150 degrees plus or minus 20 degrees) and so a single fixed calibration direction will not remove the effect of illumination in images.

While the results are not perfect, we believe they are pretty good. In all cases the shadows are removed or attenuated and the main look and feel of the image is retained. And, we achieved this performance without any calibration or prior learning.

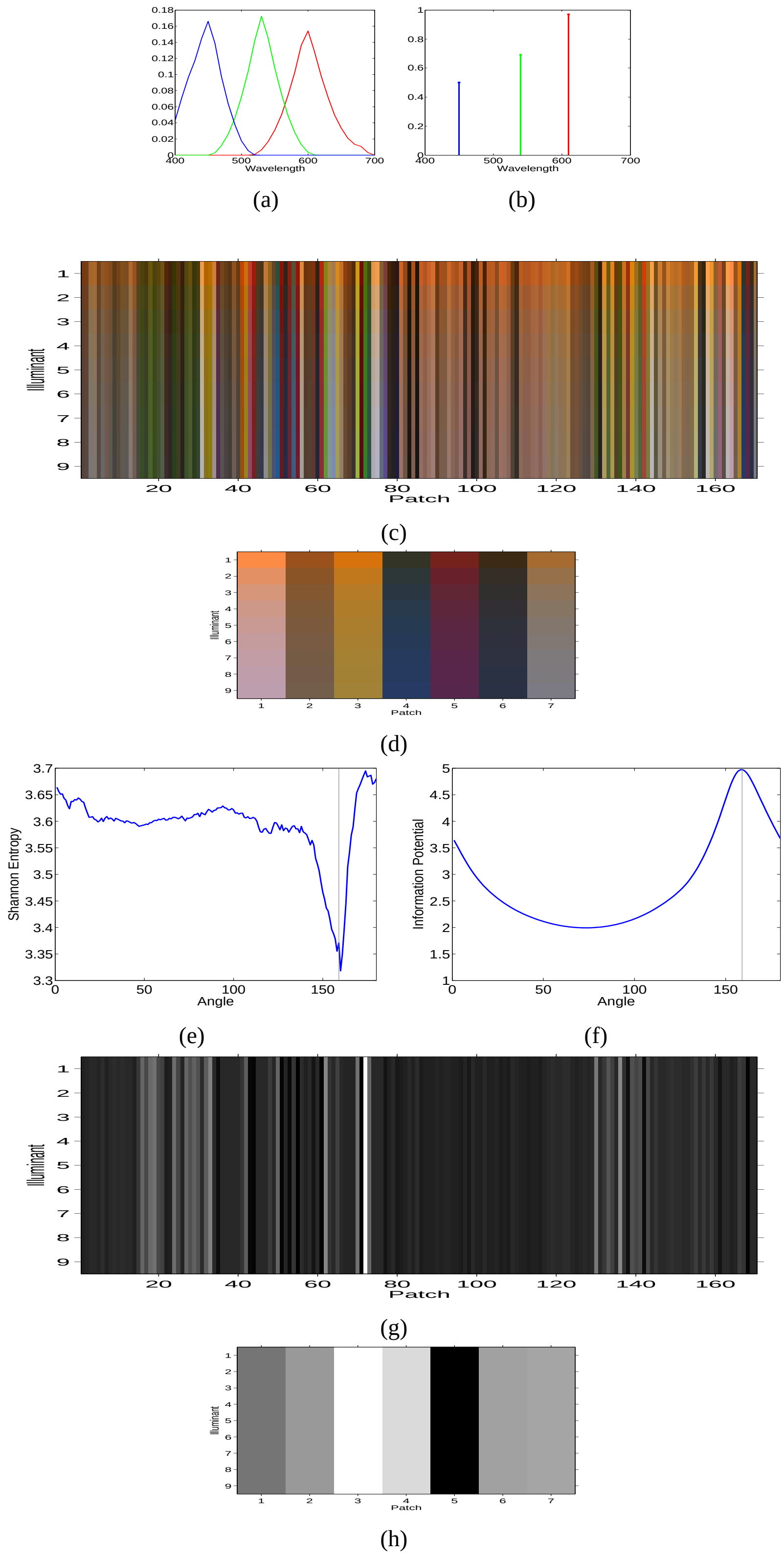
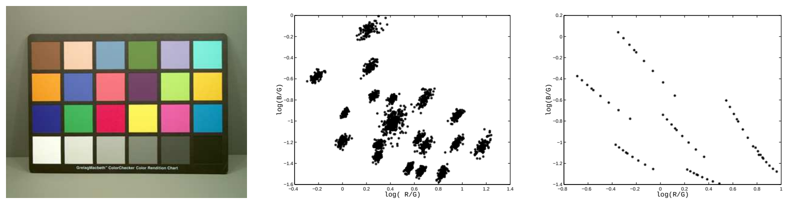
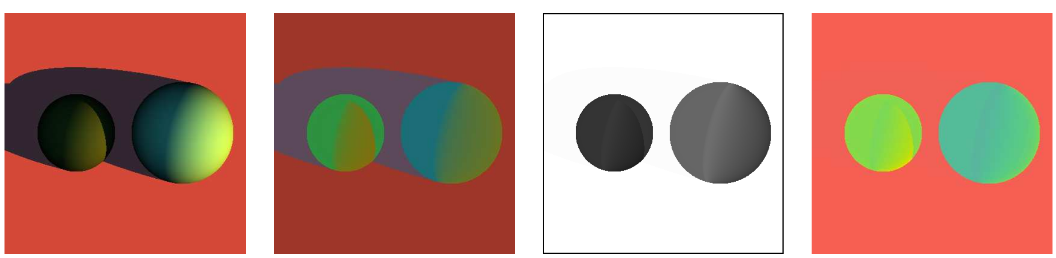


Figure 4. Synthetic values. (a): Typical RGB camera sensors — Sony DXC930 camera. (b): Theoretical narrowband RGB camera sensors. (c): An image showing all lights and surfaces. Left to right there are 170 reflectances and top to bottom the 9 Planckian lights. (d): Close-up showing the last 7 patches, under the 9 lights. (e): Minimum Shannon's entropy invariant direction gives same angle as calibration test. (f): The same angle (here we need a curve maximum) is produced by quadratic entropy. (g): Invariant image for theoretical synthetic image — same greylevels across illuminants. (h): Close-up of last 7 reflectance patches for invariant image.



(a) (b) (c)

Figure 5. (a): Macbeth ColorChecker Chart image under a Planckian light imaged with an HP912 Digital Still Camera, modified to generate linear output. (b): Log-chromaticities of the 24 patches of the imaged chart. (c): Chromaticities for 6 different patches, imaged under a set of different Planckian illuminants.



(a) (b) (c) (d)

Figure 6. (a): Two hemispheres composed of Macbeth ColorChecker Chart patches #4 and #6, on a plane composed of patch #2. Illumination is by two Planckians, and image formation is by using delta-function sensors. (b): L_1-norm chromaticity for this scene. (c): Greyscale invariant image. (d): Invariant image as chromaticity.

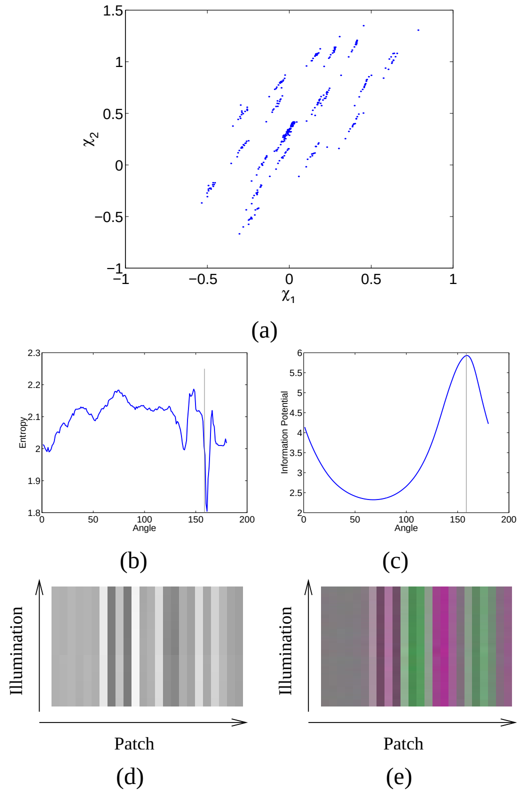


Figure 7. Measured values. (a): 2D chromaticity for measured colour patches, HP 912 camera. [Please zoom in to 300% in Acrobat see see image (a) on-screen. The image prints correctly.] (b): Minimum entropy invariant direction gives angle close to that of calibration method. (c): Same angle is found by quadratic entropy. (d): Greyscale invariant image for measured patch values — projected greylevels are same for different illuminants. (e): Recovered 2D-colour chromaticity invariant image.

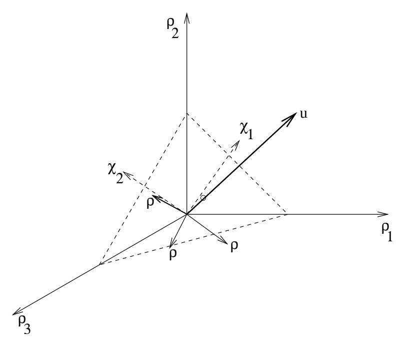
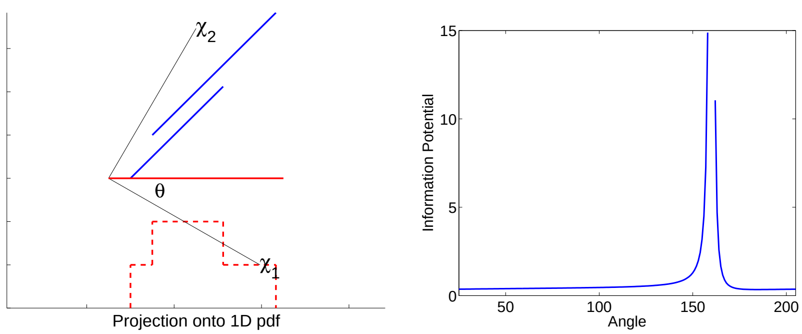


Figure 8. Geometric mean divisor implies that every $\boldsymbol{\rho}$ is orthogonal to $\boldsymbol{u}$. Basis in plane is $\{\chi_1, \chi_2\}$.



(a) (b)

Figure 9. (a): Projection of two illumination-variation lines into 1D marginal pdf by projection in θ direction. (b): Resulting (continuous) information potential for the quadratic entropy shows strong, single maximum at correct angle.

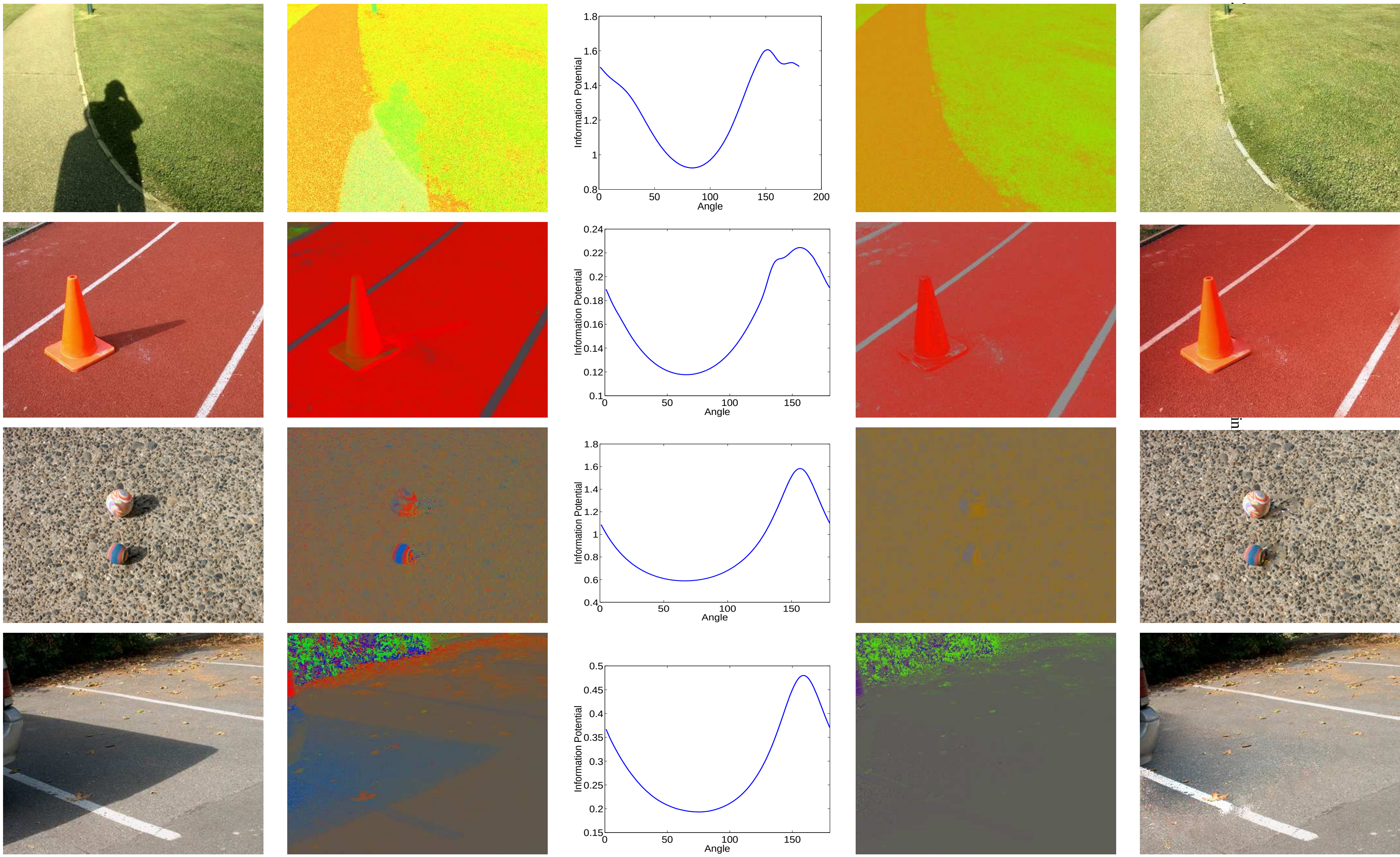

Information Potential
Angle
Information Potential
Angle
Information Potential
Angle
Information Potential
Angle

Figure 8.

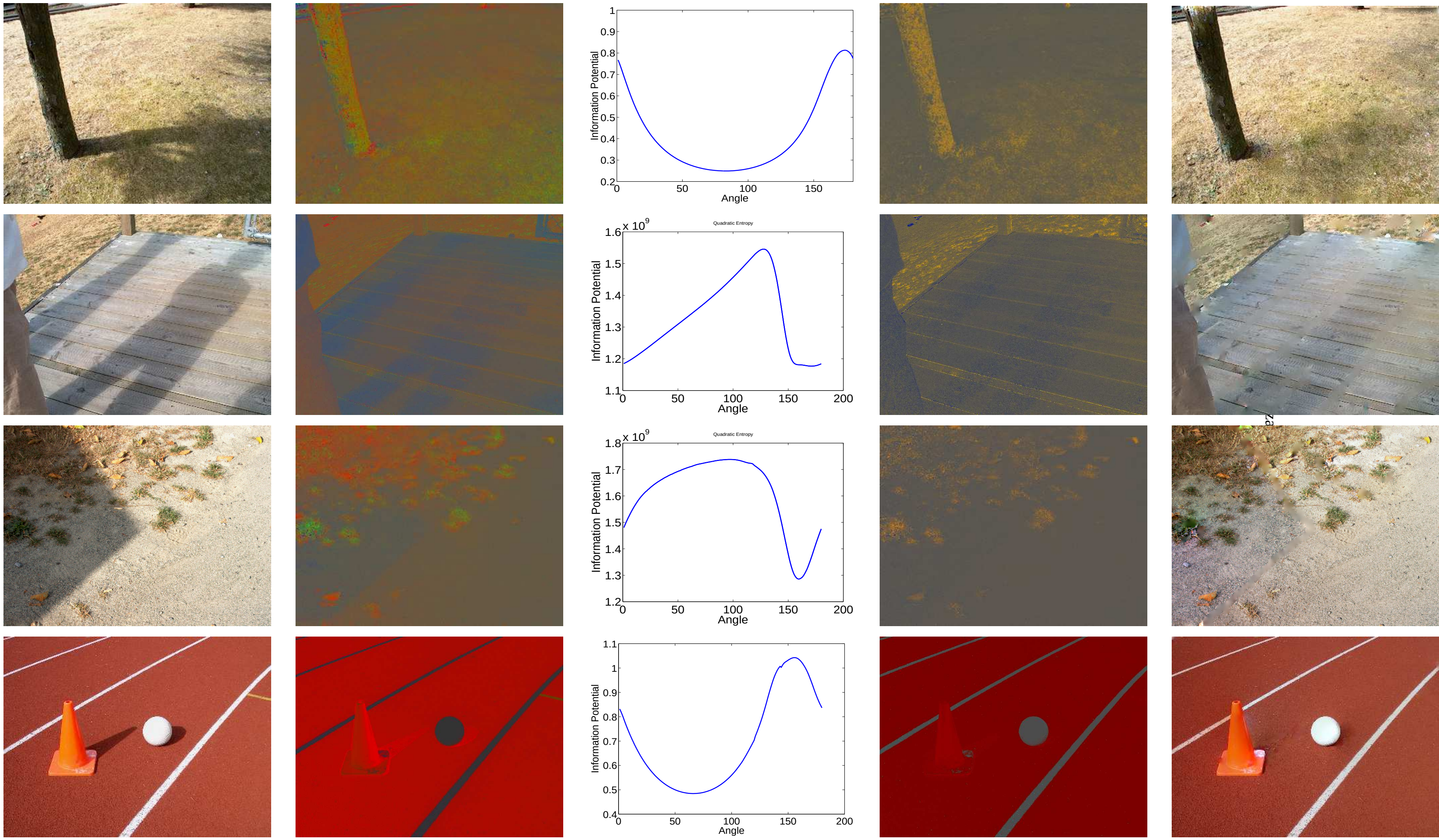


Figure 8.

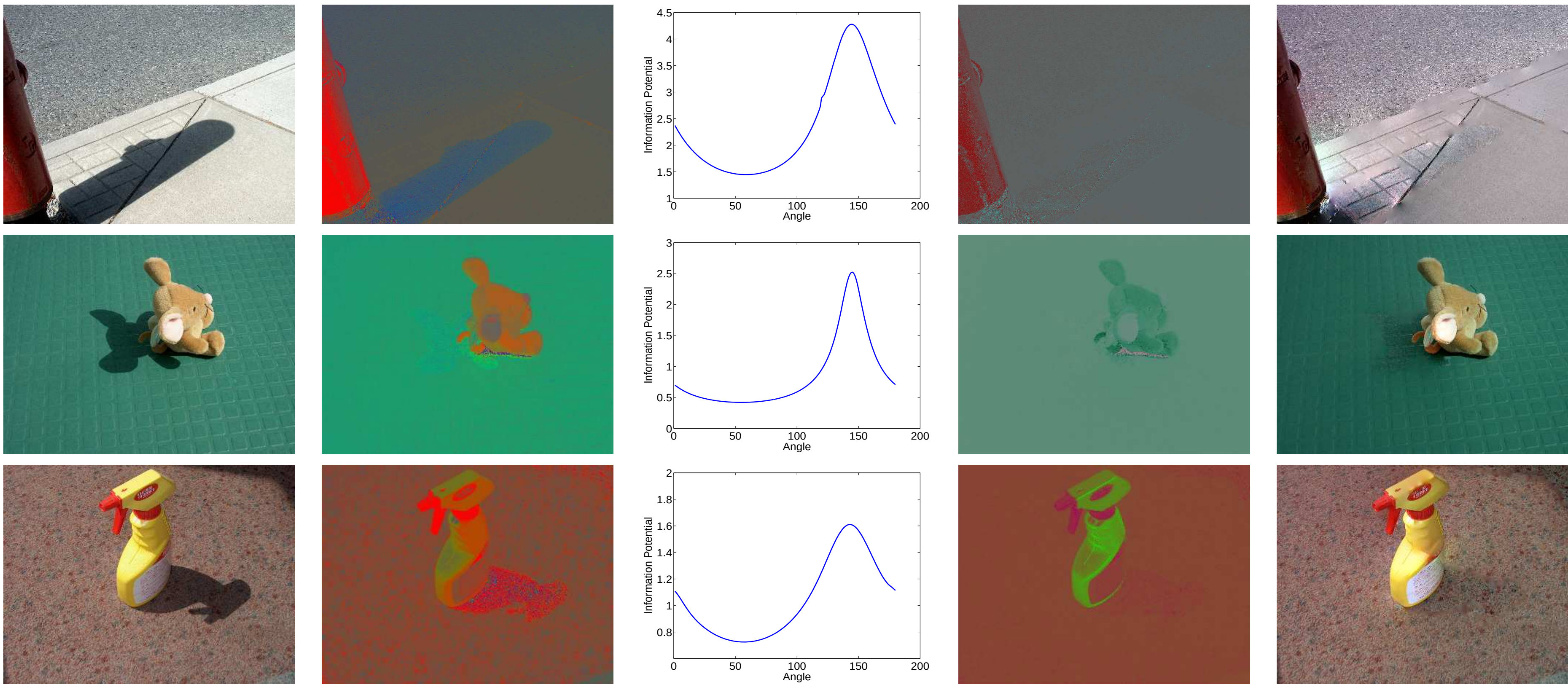


Figure 8. Additional invariant images, for minimum entropy: columns show original image, L_1 chromaticity image, information potential plot, invariant L_1 chromaticity, and re-integrated RGB colour image.

7. Are Invariant Images Intrinsic?

A question that remains is whether the invariant images produced are indeed “intrinsic” in the sense of yielding identical reflectivity results regardless of lighting conditions. To test this capability of the algorithm, we used time-lapse imaging to show shading and shadow removal stability over lighting. Fig. 9 shows a subset of several images taken outdoors over time on a variably sunny day at 20-minute intervals. For each individual image, we ran the algorithm presented here. If the invariant images produced are indeed intrinsic, then we expect to find that all invariant images are approximately equal, or at least much closer to each other than are each of the original sequence since they would consist of reflectance-only images independent of the lighting change between frames. The camera used was an inexpensive commodity camera, but with the software modified such that both preferred and raw renditions are both stored: for raw images, only demosaicing using bilinear interpolation is applied.

We find that in most of the results, the attached shadows are still somewhat apparent, although the cast shadows have been mostly removed. Nevertheless the output images are indeed closer to each other than are the originals. Since we produce chromaticity images as the invariant output, we compare closeness for the L_1 chromaticity amongst the input set across daylight conditions versus that for the output set. A simple but effective measure of the quality of image nearness is the Peak Signal to Noise ratio (PSNR) (Daly, 1992), and in fact perception-based image quality metrics have been found to offer little advantage over PSNR as a measure to evaluate the quality of image nearness. For the input set of L_1 chromaticities, if we compute the PSNR between all input images the median value is 38.70 dB. But for the output set, the median is 50.06 dB, showing a much stronger correlation: i.e., the invariant images are indeed considerably closer to generating an intrinsic representation.

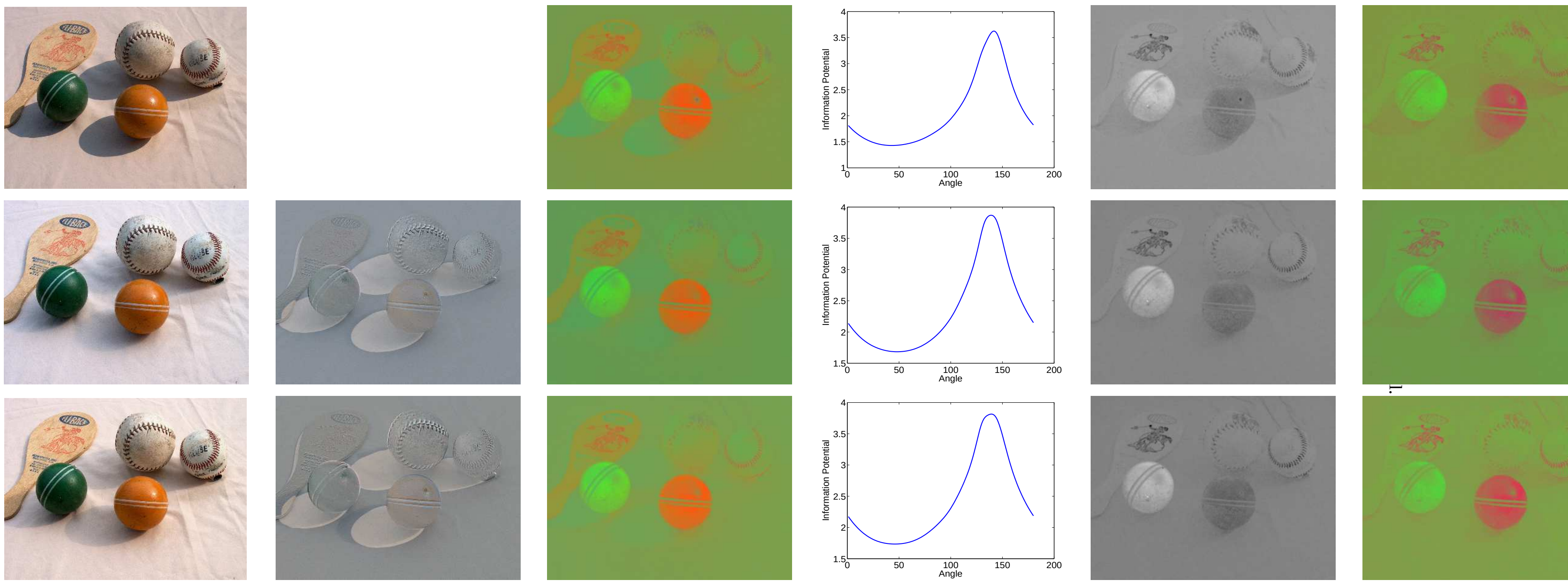


Figure 9. Invariant images using time-lapse imaging: columns show original image, difference from first image, L_1 chromaticity image (stretched into 0..1), information potential plot, invariant greyscale, and invariant L_1 chromaticity.

8. Conclusions

We have presented a method based on entropy minimization for finding the invariant direction, and thus a greyscale and thence an L_1-chromaticity intrinsic image that is independent of lighting and hence free of shadows, without any need for a calibration step or special knowledge about an image. The method appears to work quite well, and leads to good re-integrated full-colour images with shadows greatly attenuated. We found that going over to a quadratic entropy definition provides a stable and efficient vehicle for calculating the minimum-entropy lighting invariant direction.

Future work would involve a careful assessment of how onboard nonlinear processing in cameras affects results. Cameras ordinarily supply images and videos that are compressed, as well as greatly processed away from being linear images. Although the method does indeed work under such processing (see Fig. 8) it would be well to understand how JPEG artifacts impact the method. We have found that JPEG images do indeed exhibit a strong entropy minimum, just as do uncompressed images. However, the extra edges introduced due to blocking effects make re-integration more difficult.

For the re-integration step, it may be the case that consideration of a separate shadow-edge map for x and y could be useful, since in principle these are different. A variational inpainting algorithm would likely work better than our present simple morphological edge-diffusion method for traversing shadow-edges, but would be slower.

In general, the model does perform best when the underlying assumptions guiding the approach are indeed obeyed. For example, if a spectral sharpening transform (Finlayson et al., 1994) is available for a camera (or even using a generic such transform (Drew et al., 2007)) then we can expect to obtain better shadow removal from the lighting invariant. And Lambertian surfaces indeed produce the best results. A simple test of whether a surface is in fact Lambertian is that the chromaticity removes shading. The Lambertian assumption is often broken, but real images typically contain only small specular areas and these do not much affect the results. However, if we were to use scenes with large areas of non-dielectrics, this would indeed affect performance. In general, we expect the method to have limited applicability to a degree for scenes that image surfaces with BRDFs that deviate from Lambertian, such as glass, metal, etc., and likely also skin, which is complex to model (Weyrich et al., 2006).

As well, dynamic range plays an important part in consumer imaging. Under bright lighting, shadows are typically driven down to very small pixel values — say, to 2% of the maximum channel value — that may be unusable by the method presented. Also, when strong interreflections are present, in shadow regions that are very close to an object with attached shadow, the method can also not correctly remove this effect. Nonetheless, generally the method does remove, or at least diminish the presence of shadows in imagery.

Appendix: Fast Gauss Transform applied to Quadratic Entropy

In practice, computation of the information potential can be expensive. When computed naively, computation of V has complexity $O(N^2)$, where N is the number of pixels. This cost may be prohibitive when the image is large.

The Fast Gauss Transform (FGT) was introduced by Greengard and Strain (Greengard and Strain, 1991) for efficient evaluation of a weighted sum of Gaussians. It has proved to be a very efficient algorithm in a variety of applications (Yang et al., 2003; Elgammal et al., 2003; Beatson and Greengard, 1997). The discrete Gauss transform, here discussed in terms of the approximation of a 1D pdf, is to be evaluated on a grid of T points:

$$G(x_i) = \sum_{j=1}^{N} w_j \, e^{-\left(\frac{x_i - s_j}{\sigma}\right)^2} \qquad i = 1, \ldots, T. \tag{26}$$

Here, w_j are weight coefficients; $\{s_j\},\ j = 1 \ldots N$ are the data point centers of the sum of Gaussians (the *sources*); and σ is a bandwidth parameter. The sum of Gaussians is evaluated only at a set of grid points $\{x_i\},\ i = 1 \ldots T$ (the *targets*). A direct computation evaluating the sum of N source points at T targets requires $O(TN)$ exponential evaluation operations.

The FGT algorithm speeds up the computation by approximation of the Gaussian function to achieve a desired precision. The basis of the fast algorithm is the expansion of the Gaussian in terms of the Hermite functions $h_n(x)$:

$$G(x) = e^{-(x-s)^2} = \sum_{n=1}^{p} \frac{s^n}{n!} h_n(x) + \epsilon(p), \tag{27}$$

where $h_n(x)$ is defined by

$$h_n(x) = (-1)^n \frac{d^n}{dx^n} e^{-x^2},$$

and ϵ is the error introduced by truncating the Hermite series after p terms. This is a rephrasing of the Taylor series about $s = 0$.

The FGT starts by dividing the feature space (the sources) into uniform boxes with side length σ. Then the Hermite expansion is applied such that the influence of sources and targets separates. For each source s, the Gaussian can be expanded using a shifted and scaled version of Hermite functions which are located at the center s_B of the box in which the source lies.

$$\begin{aligned} e^{-\left(\frac{x-s}{\sigma}\right)^2} &= e^{-\left(\frac{x - s_B - (s - s_B)}{\sigma}\right)^2} \\ &\cong \sum_{n<p} \frac{1}{n!} \left(\frac{s - s_B}{\sigma}\right)^n h_n\left(\frac{x - s_B}{\sigma}\right) \end{aligned} \tag{28}$$

This is a so-called far-field expansion, in that it is an approximation not dependent on the distance between source and target being small.

In a similar manner, the target Gaussian field can be approximated by a Hermite expansion about the center of the target box x_B:

$$\begin{aligned} e^{-\left(\frac{x-s}{\sigma}\right)^2} &= e^{-\left(\frac{x - x_B - (s - x_B)}{\sigma}\right)^2} \\ &\cong \sum_{n<p} \frac{1}{n!} h_n\left(\frac{s - x_B}{\sigma}\right) \left(\frac{x - x_B}{\sigma}\right)^n \end{aligned} \tag{29}$$

The two expansions are identical, except that the role of sources and targets are interchanged. Eq. (29) is a so-called near-field expansion, in that it expresses a function with target x as a Taylor series about a *nearby* target box center x_B.

The FGT first calculates the expansion coefficients in eq. (28) (the monomials in s) and adds them for each source box, yielding a single expansion for each source box. These series are then shifted to the centers of target boxes using eq. (29), for the Hermite series in each target box and its nearby source box. Thus each target point has only one Taylor expansion with monomials $((x - x_B)/\sigma)^n$. In this way, a sum of Gaussians can be computed in $O(T + N)$ operations.

The FGT is typically applied to Gaussian kernel evaluations where the targets are *not well-behaved* near the sources, making it necessary to use the far field Hermite expansion and the translation to a local Taylor expansion. However, here we wish to use the FGT specifically for calculating the information potential, as in eq. (25). Here, sources and targets are identical — M_i are sources, with the term for each i of the form (26) with $x_j \equiv M_j$. In this case, the Hermite expansion is equivalent to the Taylor expansion, with no need to perform the conversion from the Hermite expansion to the local Taylor series. Therefore, *a simpler* evaluation is possible for eq. (25): all points are transformed into a Hermite expansion about the centers of the boxes, and these expansions are directly evaluated at each point.

Formally, the kernel in V can be expressed as a Hermite series:

$$e^{-\frac{(M_i - M_j)^2}{4\tilde{s}^2}} \cong \sum_{n<p} \frac{1}{n!} \left(\frac{M_j - M_B}{2\tilde{s}} \right)^n h_n \left(\frac{M_i - M_B}{2\tilde{s}} \right) \tag{30}$$

where point M_j is located in a box B with center M_B and side length $\tilde{s}$. The Fast Gauss Transform for computing the information potential (25) thus consists of the following steps:

Step 1. Assign the N data points into uniform boxes with length $\tilde{s}$.

Step 2. Choose p sufficiently large to enforce a desired error precision. The error due to the truncation of the series eq. (28) after p terms satisfies the following bound in this 1D case (Greengard and Strain, 1991; Baxter and Roussos, 2002; Beatson and Greengard, 1997):

$$\left| e^{-\left(\frac{M_i - M_j}{2\tilde{s}}\right)^2} - \sum_{n<p} \frac{1}{n!} \left(\frac{M_j - M_B}{2\tilde{s}} \right)^n h_n \left(\frac{M_i - M_B}{2\tilde{s}} \right) \right| \le \left(\frac{1}{p!} \right)^{\frac{1}{2}} \left(\frac{\sqrt{2}}{4} \right)^p . \tag{31}$$

Step 3. For each box B, with center M_B, sum the Hermite polynomials, i.e. add corresponding coefficients:

$$A_n(B) = \frac{1}{n!} \sum_{M_j \in B} \left(\frac{M_j - M_B}{2\tilde{s}} \right)^n \tag{32}$$

Step 4. For each point M_i, compute the influence of all points M_j by adding the Hermite expansion for each box B.

$$\begin{aligned} \sum_j e^{-\left(\frac{M_i - M_j}{2\tilde{s}}\right)^2} &= \sum_B \sum_{M_j \in B} e^{-\left(\frac{M_i - M_j}{2\tilde{s}}\right)^2} \\ &\cong \sum_B \sum_{n \le p} A_n(B) h_n \left(\frac{M_i - M_B}{2\tilde{s}} \right) \end{aligned} \tag{33}$$

Because of the exponential decay of the Gaussian, points in a given box will have no effect (given a particular accuracy) on far-away targets. Thus it is reasonable to compute the influence of only a range of nearby boxes for each target point, where the range is determined by the desired error bound. If we take only the r closest boxes for a point in each direction (i.e., a neighbourhood of $2r + 1$ boxes centered at the point), it can be shown (Greengard and Strain, 1991) that we incur an error bounded by $e^{-r^2/4}$. Denoting the $2r+1$ nearby boxes by $IR(B)$, the summation can be approximated by

$$\sum_{B} \sum_{M_j \in B} e^{-\left(\frac{M_i - M_j}{2\tilde{s}}\right)^2} \cong \sum_{IR(B)} \sum_{n \leq p} A_n(B) h_n \left(\frac{M_i - M_B}{2\tilde{s}}\right) \tag{34}$$

Step 5. Finally, the information potential can be calculated by adding all the Gaussian approximations obtained in step 4.

$$V \cong \sum_{i} \sum_{IR(B)} \sum_{n \leq p} A_n(B) h_n \left(\frac{M_i - M_B}{2\tilde{s}}\right). \tag{35}$$

In step 3, each point contributes to exactly one expansion, so that the amount of work required to calculate the coefficients for all boxes is $O(Np)$. The amount of work required in step 4 is $O(p(2r+1))$ for each point, and $O(Np(2r+1))$ in total for all points. The desired precision ϵ dictates our choice of r and p. For calculating the information potential, the precision required is moderate, so that we can have small r and p. In this paper we use $r = 6$ and $p = 6$. Overall, the FGT algorithm achieves linear running time $O(N)$.

References

Barrow, H. and J. Tenenbaum: 1978, 'Recovering intrinsic scene characteristics from images'. In: A. Hanson and E. Riseman (eds.): *Computer Vision Systems*. Academic Press, pp. 3–26.

Baxter, B. and G. Roussos: 2002, 'A New Error Estimate of the Fast Gauss Transform'. *SIAM J. on Sci. Comput.* **24**(1), 257–259.

Beatson, R. and L. Greengard: 1997, 'A short course on fast multipole methods'. In: M. Ainsworth, J. Levesley, W. Light, and M. Marletta (eds.): *Wavelets, Multilevel Methods and Elliptic PDEs.* Oxford Univ. Press.

Bergner, S., M. Drew, and T. Möller: 2009, 'A tool to create illuminant and reflectance spectra for light-driven graphics and visualization'. *ACM Trans. on Graphics* **28**(1), 5:1–5:11.

Cho, J.-H., T.-G. Kwon, D.-G. Jang, and C.-S. Hwang: 2005, 'Moving cast shadow detection and removal for visual traffic surveillance'. In: *Australian Conference on Artificial Intelligence*. pp. 746–755.

CIE: 1995, 'Method of Measuring and Specifying Colour Rendering Properties of Light Sources, Publication 13.3, ISBN 978-3900734572'. http://www.cie.co.at/publ/abst/13-3-95.html.

Comaniciu, D. and P. Meer: 2002, 'Mean Shift: A Robust Approach Toward Feature Space Analysis'. *IEEE Trans. Patt. Anal. and Mach. Intell.* **24**, 603–619.

Daly, S.: 1992, 'The visible difference predictor: an algorithm for the assessment of image fidelity'. In: A. Rogowitz and Klein (eds.): *SPIE: Human Vision, Visual Processing, and Digital Display III*. pp. 1666:2–15.

Drew, M., C. Chen, S. Hordley, and G. Finlayson: 2002, 'Sensor Transforms for Invariant Image Enhancement'. In: *Tenth Color Imaging Conference: Color, Science, Systems and Applications.* pp. 325–329.

Drew, M., G. Finlayson, and S. Hordley: 2003, 'Recovery of Chromaticity Image Free from Shadows via Illumination Invariance'. In: *IEEE Workshop on Color and Photometric Methods in Computer Vision, ICCV'03*. pp. 32–39.

Drew, M., M. Salahuddin, and A. Fathi: 2007, 'A Standardized Workflow for Illumination-Invariant Image Extraction'. In: *15th Color Imaging Conference: Color, Science, Systems and Applications*.

Elgammal, A., R. Duraiswami, and L.S.Davis: 2003, 'Efficient Kernel Density Estimation Using the Fast Gauss Transform with Applications to Color Modeling and Tracking'. *IEEE Trans. Patt. Anal. and Mach. Intell.* **25**(11), 1499–1504.

Finlayson, G. and M. Drew: 2001, '4-Sensor Camera Calibration for Image Representation Invariant to Shading, Shadows, Lighting, and Specularities'. In: *ICCV'01: International Conference on Computer Vision*. pp. II: 473–480.

Finlayson, G., M. Drew, and B. Funt: 1994, 'Spectral sharpening: sensor transformations for improved color constancy'. *J. Opt. Soc. Am. A* **11**(5), 1553–1563.

Finlayson, G., M. Drew, and C. Lu: 2004, 'Intrinsic Images by Entropy Minimization'. In: *ECCV 2004: European Conference on Computer Vision*. pp. 582–595. Lecture Notes in Computer Science Vol. 3023.

Finlayson, G., C. Fredembach, and M. S. Drew: 2007, 'Detecting Illumination in Images'. In: *ICCV'07: International Conference on Computer Vision*.

Finlayson, G. and S. Hordley: 2001, 'Colour Constancy at a Pixel'. *J. Opt. Soc. Am. A* **18**(2), 253–264.

Finlayson, G., S. Hordley, and M. Drew: 2002, 'Removing Shadows from Images'. In: *ECCV 2002: European Conference on Computer Vision*. pp. 4:823–836. Lecture Notes in Computer Science Vol. 2353.

Finlayson, G., S. Hordley, C. Lu, and M. Drew: 2006, 'On the Removal of Shadows From Images'. *IEEE Trans. Patt. Anal. Mach. Intell.* **28**, 59–68.

Greengard, L. and J. Strain: 1991, 'The fast Gauss transform'. *SIAM J. Sci. Statist. Comput.* **12**(1), 79–94.

Hsu, E., T. Mertens, S. Paris, S. Avidan, and F. Durand: 2008, 'Light mixture estimation for spatially varying white balance'. *ACM Trans. Graph.* **27**(3), 1–7.

Jiang, H. and M. Drew: 2003, 'Shadow-Resistant Tracking in Video'. In: *ICME'03: Intl. Conf. on Multimedia and Expo*. pp. III 77–80.

Jiang, H. and M. Drew: 2007, 'Shadow Resistant Tracking using Inertia Constraints'. *Pattern Recognition* **40**, 1929–1945.

Land, E. and J. McCann: 1971, 'Lightness and retinex theory'. *J. Opt. Soc. Amer.* **61**, 1–11.

Li, Z.-N. and M. Drew: 2004, *Fundamentals of Multimedia*. Prentice-Hall.

Liu, Z., K. Huang, T. Tan, and L. Wang: 2006, 'Cast Shadow Removal with GMM for Surface Reflectance Component'. In: *ICPR06*. pp. 727–730.

Martel-Brisson, N. and A. Zaccarin: 2007, 'Learning and removing cast shadows through a multidistribution approach'. *IEEE Trans. Patt. Anal. and Mach. Intell.* **29**, 1133–1146.

Nadimi, S. and B. Bhanu: 2004, 'Physucal Models for Moving Shadow and Object Detection in Video'. *IEEE Trans. Patt. Anal. and Mach. Intell.* **26**, 1079–1087.

Parzen, E.: 1962, 'On Estimation of a Probability Density Function and Mode'. *Annals of Mathematical Statistics* **33**, 1065–1076.

Prati, A., I. Mikic, M. Trivedi, and R. Cucchiara: 2003, 'Detecting Moving Shadows: Algorithms and Evaluation'. *IEEE Trans. Patt. Anal. and Mach. Intell.* **25**, 918–923.

Ramanath, R., W. Snyder, Y. Yoo, and M. S. Drew: 2005, 'Color Image Processing Pipeline in Digital Still Cameras'. *IEEE Signal Processing* **22**(1), 34–43.

Renyi, A.: 1987, *A Diary on Information Theory*. Wiley.

Scott, D.: 1992, *Multivariate Density Estimation: Theory, Practice and Visualization.* Wiley and Kegan Paul.

Stauder, J., R. Mech, and J. Ostermann: 1999, 'Detection of Moving Cast Shadows for Object Segmentation'. *IEEE Trans. on Multimedia* **1**, 65–76.

Tappen, M., W. Freeman, and E. Adelson: 2003, 'Recovering Intrinsic Images from a Single Image'. In: *Advances in Neural Information Processing Systems 15*.

Tappen, M., W. Freeman, and E. Adelson: 2005, 'Recovering Intrinsic Images from a Single Image'. *IEEE Trans. Patt. Anal. and Mach. Intell.* **27**, 1459–1472.

Vrhel, M., R. Gershon, and L. Iwan: 1994, 'Measurement and analysis of object reflectance spectra'. *Color Research and Application* **19**, 4–9.

Weiss, Y.: 2001, 'Deriving Intrinsic Images from Image Sequences'. In: *ICCV01*. pp. II: 68–75.

Weyrich, T., W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross: 2006, 'Analysis of human faces using a measurement-based skin reflectance model'. *ACM Trans. Graph.* **25**(3), 1013–1024.

Wyszecki, G. and W. Stiles: 1982, *Color Science: Concepts and Methods, Quantitative Data and Formulas*. Wiley, New York, 2nd edition.

Xu, D. and J. Principe: 1998, 'Learning from Examples with Quadratic Mutual Information'. In: *Neural Networks for Signal Processing*. pp. 155–164.

Yang, C., R. Duraiswami, N. Gumerov, and L. Davis.: 2003, 'Improved fast Gauss transform and efficient kernel density estimation'. In: *Int. Conf. on Comp. Vision*. pp. 464–471.