

Correlation-Based Intrinsic Image Extraction from a Single Image

Xiaoyue Jiang¹, Andrew J. Schofield¹, and Jeremy L. Wyatt²

¹ School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK

² School of Computer Science, University of Birmingham, B15 2TT, UK
{x.y.jiang,a.j.schofield}@bham.ac.uk, jlw@cs.bham.ac.uk

Abstract. Intrinsic images represent the underlying properties of a scene such as illumination (shading) and surface reflectance. Extracting intrinsic images is a challenging, ill-posed problem. Human performance on tasks such as shadow detection and shape-from-shading is improved by adding colour and texture to surfaces. In particular, when a surface is painted with a textured pattern, correlations between local mean luminance and local luminance amplitude promote the interpretation of luminance variations as illumination changes. Based on this finding, we propose a novel feature, local luminance amplitude, to separate illumination and reflectance, and a framework to integrate this cue with hue and texture to extract intrinsic images. The algorithm uses steerable filters to separate images into frequency and orientation components and constructs shading and reflectance images from weighted combinations of these components. Weights are determined by correlations between corresponding variations in local luminance, local amplitude, colour and texture. The intrinsic images are further refined by ensuring the consistency of local texture elements. We test this method on surfaces photographed under different lighting conditions. The effectiveness of the algorithm is demonstrated by the correlation between our intrinsic images and ground truth shading and reflectance data. Luminance amplitude was found to be a useful cue. Results are also presented for natural images.

1 Introduction

In standard imagery, pixel intensities depend on both the reflectance properties of objects in the scene and its illumination conditions. Images representing these underlying properties are called intrinsic images [1]. The extraction of such images can improve many computer vision methods such as: object recognition, light source estimation and shape-from-shading.

The extraction of intrinsic images is an ill-posed problem. A variety of cues have been proposed to constrain this problem. Early approaches were based on the Retinex theory of lightness constancy in humans [2]. This theory rests on the assumption that lighting changes are smooth whereas reflectance changes are abrupt; this difference can be used to distinguish illumination from reflectance. However, abutting flat surfaces at different orientations in a 3D world produce

abrupt changes in illumination leading Shina and Adelson [3] to propose a 2-stage process wherein luminance junctions are classified as illumination or reflectance using local heuristics and then reclassified if necessary by a global analysis that reconstructs 3D shapes. The method works well in stylised stimuli where edges are easily defined. With supervised learning, Bell and Freeman [4] reconstructed shading and reflectance from classified steerable filter coefficients.

Another common approach is to use colour as a key for identifying illumination gradients, based on the assumption that hue is illumination invariant (see for example, [5], [6] and [7]). An illumination map can then be derived by reintegrating only those gradients that arise from illumination. However, hue is not entirely illumination-invariant: outdoor shadows are tinted blue [8] and hue is poorly specified in dark shadows. In addition hue based methods can be confused by small image features, although Tappen et al. [7] provides a reasonable solution to this problem by training a classifier to distinguish shadow and reflectance edges. Finlayson et al.'s colour based method [9] defines an illumination-invariant colour space to discriminate shadows from reflectance variations, but this requires a calibrated camera. Further, since humans can distinguish shadows from reflectance changes in monochrome images [10] colour cannot be the only cue that enables such a separation.

Like hue, certain texture properties are also invariant to illumination and therefore a potential cue for deriving intrinsic images. Shen et al. [11] applied texture consistency as a constraint for decomposing images into shading and reflectance. This algorithm identifies groups of pixels sharing illumination invariant texture features and adjusts those aspects that are not illumination invariant until the groups are more consistent. The result is to discount illumination to produce a reflectance map. However, this method uses a computationally expensive optimization procedure. Finally, intrinsic images can be extracted from image sequences using multiple images to constrain the problem [12,13,14]. Such algorithms produce good results but are limited by the need for multiple images.

We introduce a new algorithm to extract intrinsic images from single images. The method combines colour and texture with a new metric (luminance amplitude, see Section 2) which is used by humans to differentiate shading and reflectance [15]. We combine these metrics with a steerable filter decomposition and use inter-cue correlations to identify frequency/orientation components that belong to the shading and reflectance maps respectively. We test this approach on images for which we have ground truth data and compare our novel luminance amplitude cue with the more established texture and colour cues.

2 Basic Cues

Luminance amplitude: Assuming Lambertian reflectance, the intensity value $I(x, y)$ of every pixel in an image is the product of the incident lighting $L(x, y)$ and the reflectance $R(x, y)$ at that point: $I(x, y) = R(x, y) \times L(x, y)$. For surfaces with a painted texture we can measure the mean and variance (luminance amplitude) of pixel intensities in local regions. If a change in local mean intensity

is caused by a change in illumination then luminance amplitude should vary in the same direction. This occurs because illumination multiplies light and dark reflectances in the texture by a common factor. Thus correlated changes in mean luminance and luminance amplitude indicate illumination changes.

Colour and Texture: Unlike illumination changes, reflectance variations are characterized by complex variations in pixel values based on a number of potential cues. If the pattern of a texture changes (e.g. a change in granularity or dominant orientation), then any associated change in mean intensity, over a large enough patch, might reasonably be regarded as a change in surface reflectance $R(x, y)$. Colour is another diagnostic feature for reflectance changes. If hue and intensity vary together this is likely to signal a reflectance change. Thus positive correlations between colour and luminance or texture and luminance indicate reflectance changes.

3 Steerable Filter Based Feature Extraction

Our algorithm is based on the relationships between intensity, luminance amplitude, texture and colour, as described above. The overall framework for the algorithm is shown in Fig. 1. We use steerable filters to decompose the image into its constituent orientation/frequency bands. These filters provide a general framework that can decompose images and completely reconstruct the originals from the resulting components [16] or, as here, construct partial images from selected components. We apply the steerable filter bank S_L to the raw luminance values (luminance modulations LM) extracting a full set of luminance components ($LM_{ij}, (i = 1, \dots, N; j = 1, \dots, M)$, where N is the number of orientations, and M is the number of frequency bands in S_L). We also apply the filter bank to estimate variations in local amplitude (AM), texture (TM), and hue (HM). We then calculate the correlation between LM and AM, TM and HM in each orientation/frequency band. If a component of LM is positively correlated to AM but not TM or HM, it is deemed to convey shading information. The illumination

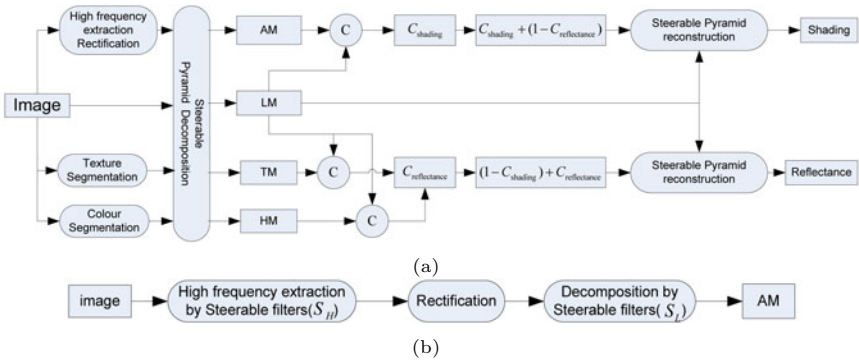


Fig. 1. (a) Flowchart of the overall algorithm, (b) Flowchart of AM extraction

(or shading) image is then constructed from only these components. Similarly, components of LM that are positively correlated to TM and HM but not AM are used to construct the reflectance image.

3.1 Extraction of AM

The equation for image intensity ($I = R \times L$) shows that, for fixed reflectance, intensity increases with increased illumination (L). Assuming that lighting is constant in a local region but that reflectance varies due to texture, convolving both sides of the lighting equation with a high-frequency filter $F_h(\cdot)$ produces the equation $F_h(I) = F_h(R) \times L$. Large scale variations in L will modulate the magnitude of the filter responses such that the envelope of $F_h(I)$ will be determined by L . Rectifying the output of $F_h(I)$ will demodulate this envelope which can be processed by further filtering with $F_l(\cdot)$,

$$F_l(r(F_h(I))) = F_l(r(F_h(R) \times L)) \quad (1)$$

where $r(x) = \text{abs}(x - \mu_x)$ is the rectification of signal x . The low-frequency filter $F_l(\cdot)$ detects low-frequency information in $r(F_h(I))$, hence Eq.1 can be written as $F_l(r(F_h(I))) \approx F_l(r(L))$. The response of $F_l(r(F_h(I)))$ is a measure of local amplitude which, if the texture is uniform, will be correlated with illumination.

In practice, the high-frequency part of the input image I_H is extracted by the steerable filters: reconstructing I_H from only the high-frequency responses. We then rectify I_H about its mean value, i.e. $r(I_H) = \text{abs}(I_H - \mu_{I_H})$, and apply steerable filters (S_L), as used for LM, to decompose $r(I_H)$ yielding components of AM that match those extracted for LM. The flowchart for extracting AM and an example AM component are shown in Fig. 1(b) and Fig. 2(c), respectively.

3.2 Extraction of TM

Texture modulation (TM) should represent transitions between different texture patterns. The extraction of TM relies on texture segmentation; a difficult problem in itself which we do not attempt to solve in full here. However, for our purposes segmentation based on Gabor-features works well. We calculate the Gabor responses to the images in different orientation and frequency bands, and then use the principal component analysis (PCA) to extract common texture features. Next we use fuzzy clustering to classify the responses into non-continuous regions of similar texture. We then create a simple texture map by block filling texture regions with their own mean intensity, see Fig. 2(f), before applying steerable filters (S_L) to extract components matching the LM signals, see Fig. 2(d).

3.3 Extraction of HM

Colour (more specifically hue) is an important feature for estimating reflectance as it is, more-or-less, illumination invariant. We derive a 4-dimensional intensity-free colour vector,

$$F_{colour}(x, y) = (r_{xy}/\|I_{xy}\|, g_{xy}/\|I_{xy}\|, b_{xy}/\|I_{xy}\|, h_{xy}) \quad (2)$$

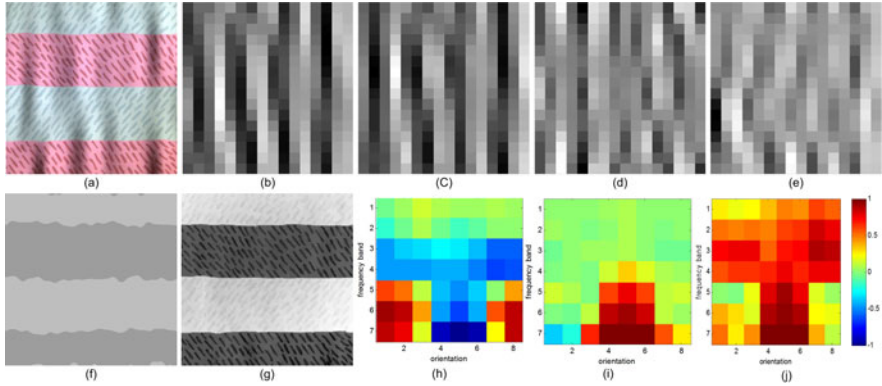


Fig. 2. Intermediate processing: (a) the input image, (b) one LM component of (a), (c-e) the matching AM, TM and HM components respectively, (f-g) texture and colour maps for (a), (h-j) pictorial representation of correlations between LM and AM, TM and HM respectively; x-axis shows component orientation, y frequency, cell colour indicates correlation coefficients

where (r_{xy}, g_{xy}, b_{xy}) represent RGB values, $\|I_{xy}\| = \sqrt{r_{xy}^2 + g_{xy}^2 + b_{xy}^2}$ is the norm of the RGB triple, and h_{xy} is hue. We apply the mean shift algorithm [17] to this vector so as to achieve colour segmentation, block filling each region in the hue map according to its mean luminance, see Fig. 2(g). Once again steerable filters (S_L) are used to extract HM components from the hue segmentation results, see Fig. 2(e).

4 Reconstruction of Shading and Reflectance

Extracting intrinsic images makes the implicit, but not always appropriate, assumption that luminance variations are either due to reflectance or illumination but not both. Our algorithm is based on this assumption but not bound by it. By estimating the correlation between each LM component and reflectance and shading respectively we can assign individual components to shading and reflectance in some proportion. Correlation coefficients C_{ij}^{la} , between the corresponding components LM_{ij} and AM_{ij} are used to measure the relationship between luminance and luminance amplitude, see Fig. 2(h). If LM_{ij} and AM_{ij} are positively related ($C_{ij}^{la} > 0$) we presume that LM_{ij} results from illumination and retain some proportion of it when reconstructing the illumination map. If $C_{ij}^{la} < 0$, LM_{ij} is used in the reflectance map. Hence we rename C_{ij}^{la} as C_{ij}^{shd} because when positive this measure places the component into the shading map.

The correlation coefficient C_{ij}^{lt} between component LM_{ij} and TM_{ij} describes the relationship between luminance and texture. If $C_{ij}^{lt} > 0$, LM_{ij} should be treated as a reflectance change and used in the reflectance map. Because TM is extracted from texture blocks, see Fig. 2(f), its low-frequency components are more reliable than its high frequency components. We use the frequency index j

to weight coefficients C_{ij}^{lt} to reflect their reliability. The updated coefficient \tilde{C}_{ij}^{lt} between LM_{ij} and TM_{ij} is,

$$\tilde{C}_{ij}^{lt} = C_{ij}^{lt} \times \frac{j}{M+1} \quad (3)$$

where $j = M$ indicates the lowest frequency band. Similarly, the correlation coefficient C_{ij}^{lh} between LM_{ij} and HM_{ij} measures the relationship between luminance and hue changes. If $C_{ij}^{lh} > 0$, LM_{ij} should be included in the reflectance map. As texture and colour are both positively related to reflectance changes we can combine their relationship with LM into a single measure (C_{ij}^{ref}) as follows,

$$C_{ij}^{ref} = \frac{\tilde{C}_{ij}^{lt} + C_{ij}^{lh}}{2} \quad (4)$$

In order to correctly divide LM components between shading and reflectance we need to consider their relationship with both properties. If we treat local amplitude, texture and colour as equally reliable, then the correlation coefficients for reconstructed shading ($C_{ij}^{rec-shd}$) and reflectance ($C_{ij}^{rec-ref}$) are given by Eq.5 and Eq.6, respectively.

$$C_{ij}^{rec-shd} = C_{ij}^{shd} + (1 - C_{ij}^{ref}) \quad (5)$$

$$C_{ij}^{rec-ref} = (1 - C_{ij}^{shd}) + C_{ij}^{ref} \quad (6)$$

However, because we cannot decompose the original image into infinitely narrow orientation and frequency bands, and because, in real lighting situations, some variations in intensity are caused by reflectance and shading together, some components LM_{ij} will strongly correlate with both reflectance and shading. In this situation, we need to assign a weight to each correlation coefficient according to the reliability of the texture and colour segmentation results. That is,

$$C_{ij}^{rec-shd} = wC_{ij}^{shd} + (1 - w)(1 - C_{ij}^{ref}) \quad (7)$$

$$C_{ij}^{rec-ref} = w(1 - C_{ij}^{shd}) + (1 - w)C_{ij}^{ref} \quad (8)$$

$$w = \frac{T_S}{T_S + T_R/k} \quad (9)$$

where T_R ($T_R \in [0, 1]$) estimates the reliability of texture and colour segmentations and $T_S = 1 - T_R$ the reliability of the amplitude modulations. Although texture and colour segmentation are based (in principle) on illumination-invariant features, illumination changes can still influence the segmentation results. Therefore we use an illumination parameter k to adjust the reliability of these cues. k is the image's *key* value which is given by the global contrast of an image [18],

$$k = \frac{L_{max} - L_{av}}{L_{max} - L_{min}} \quad (10)$$

where L_{av} , L_{min} and L_{max} are the logarithmic average, minimum and maximum of the luminance respectively. More extreme (harsh, high contrast) lighting conditions produce bigger key values down-weighting texture and colour. Fig. 3(b) shows the key values of images under different lighting conditions.

After deciding the correlation coefficients for every LM component, estimates of shading and reflectance can be reconstructed as follows:

$$I_{shd} = S_L \otimes \begin{cases} C_{ij}^{rec_shd} \times LM_{ij} & \text{if } C_{ij}^{rec_shd} > 0 \\ 0 & \text{if } C_{ij}^{rec_shd} \leq 0 \end{cases} \quad (11)$$

$$I_{ref} = S_L \otimes \begin{cases} C_{ij}^{rec_ref} \times LM_{ij} & \text{if } C_{ij}^{rec_ref} > 0 \\ 0 & \text{if } C_{ij}^{rec_ref} \leq 0 \end{cases} \quad (12)$$

where \otimes is the reconstruction of steerable filters S_L with weighted LM_{ij} . Due to the self-inverting characteristics of steerable filters, the same filters can be used for decomposition and reconstruction [16]. Correlation coefficients $C_{ij}^{rec_shd}$ and $C_{ij}^{rec_ref}$ determine how much each component will contribute to the relevant intrinsic image. More positive correlations produce stronger weights, but negative correlations produce zero weights.

5 Post-processing of the Reconstructed Images

As will be shown in Section 6, the reconstruction process described above is reasonably effective. However it is not perfect and we now outline some post processing steps that improve the final results.

5.1 DC Component of Shading Image

During the reconstruction process some LM components will be set to zero. Therefore the resulting images may lose their DC value. Although this DC value will not influence the overall appearance of the reconstructed images, it may affect subsequent processing. An alternative estimate for the reflectance image can be derived from the shading image as follows:

$$I_{dRef} = I_{org}/I_{shd} \quad (13)$$

where I_{org} is the original image. Therefore the problem of calculating the DC component can be transferred to an assessment of the reflectance image I_{dRef} . If the initial estimate of shading is accurate enough, then I_{dRef} should convey uniform intensity distributions within each texture class. Thus an evaluation of the texture consistency within I_{dRef} can be used as a cost function to optimize I_{shd} . The optimization problem is

$$E_{shd}(V_{DC}) = \arg \min_{T_i} \sum_i F_{txt}\left(\frac{I_{org}}{I_{shd} + V_{DC}}, T_i\right) \quad (14)$$

where V_{DC} is the DC value to be optimized. The function $F_{txt}(I, T_i)$ represents texture consistency evaluated for image I against the texture segmentation results T_i ($i = 1, \dots, p$), where p is the number of textures in the image I . We

model the texture distribution as a normal distribution $N(\mu_i, \sigma_i)$, hence texture consistency is defined as

$$F_{txt}(\hat{I}_{dRef}, T_i) = -\frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(\hat{I}_{dRef}(T_i) - \mu_i)^2}{2\sigma_i^2}\right) \quad (15)$$

where $\hat{I}_{dRef} = I_{org}/(I_{shd} + V_{DC})$; $\hat{I}_{dRef}(T_i)$ is the region that belongs to texture T_i in image \hat{I}_{dRef} . Texture distributions are sampled from the most reliable regions of \hat{I}_{dRef} where the corresponding shading value I_{shd} is high. The resulting shading image $\hat{I}_{shd} = I_{shd} + V_{DC}$ is a better estimate of true shading than I_{shd} .

5.2 Compensation of Reflectance in Regions of Deep Shadow

When the frequency of an AM component is greater than that of any shading in the image, it will represent small variations in reflectance not shading. Thus LM tends not to correlate with AM in high-frequency bands, but LM will still correlate well with TM and HM in these bands. Therefore high frequency components tend to be allocated to the reflectance image rather than the shading image. However, shadows tend to suppress the luminance range of textures. Looking ahead, this is seen in Fig. 5(d) where textured areas of the reconstructed reflectance image are slightly erased in regions that were in shadow (cf Fig. 2(a)).

We solve the above problem by locally amplifying the responses of high-frequency components using the optimised shading (I_{shd} , Eq.14) to guide the adjustment of each LM_{ij} component separately. The compensation for LM_{ij} is,

$$\hat{LM}_{ij} = LM_{ij} \times F_{adj}(\alpha, \beta, t) = LM_{ij} \times \frac{t(\alpha + \beta)}{\alpha \hat{I}_{shd} + \beta} \quad (16)$$

where \tilde{I}_{shd} is produced by normalizing \hat{I}_{shd} to the range $[0, 1]$. The adjustment function $F_{adj}(\alpha, \beta, t)$ is in the range of $[t(\alpha/\beta + 1), t]$, such that the darkest part is re-scaled to $t(\alpha/\beta + 1)LM_{ij}$ and the lightest part to tLM_{ij} . Parameters α and β ($\alpha > 0, \beta > 0$) control the adjustments for the dark and light pixels, t controls the overall range of the adjustment. The objective function for optimizing these parameters is

$$E_{ref} = \arg \min \sum_k F_{txt}(\hat{LM}_{ij}, T_k) + \lambda E_{cst}(F_{adj}) \quad (17)$$

where $F_{txt}(\hat{LM}_{ij}, T_k)$ evaluates texture consistency as defined in Eq. 15. For each texture T_k , \hat{LM}_{ij} is modelled as a normal distribution $N(\mu_k^{ij}, \sigma_k^{ij})$ based on the more reliable regions where \hat{I}_{shd} is high (light regions). The function $E_{cst}(F_{adj})$ constrains the maximum rescaling produced by the adjustment function. If the estimated shading map is good, we only need to apply small adjustments and can constrain the ratio between α and β as Eq. 18. The interior-point algorithm [19] is used to solve this constrained optimization problem.

$$E_{cst}(F_{adj}) = \alpha/\beta \quad (18)$$

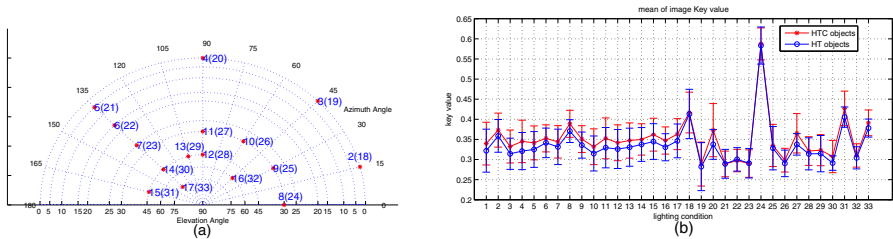


Fig. 3. (a) Asterisks show spotlight positions, numbers represent spot+diffuse conditions (those in brackets spotlight only conditions). (b) Mean key values for HT and HTC surfaces indexed by lighting condition.

6 Experimental Results

6.1 Test Set

In order to evaluate the proposed algorithm quantitatively we devised a test set containing images of 10 surfaces (5 surface shapes with two paint treatments) photographed under 33 lighting conditions. Surfaces were milled into small (57 x 64 mm) plastic blocks using a CNC Milling machine (Modella JWX-10, Roland Inc, Japan) and ArtCAM design software (Delcam plc, UK) providing multiple examples of the same surface. Surface profiles included highly coherent sinusoidal corrugations, random but oriented ripples and isotropic bumps, see Fig. 2 and 7. Two paint treatments (HTC and HT surface as described below) were applied using matte paints. Both treatments consisted of broad red and green stripes each textured with thin elements oriented differently for the two colours. For hue, texture, contrast (HTC) surfaces the darker green paint was chosen such that the green regions had lower contrast but higher mean reflectance than the red regions. For hue, texture (HT) surfaces the contrasts of the green and red regions were more similar.

Surfaces were placed into the centre of a 1m diameter integrating sphere; standing vertically and facing forward towards a pair of cameras placed either side of azimuth angle 90° , see Fig. 3(a). Here we use images from the right camera only. A bright white backlight (composed of 24 wide angle, 7 lm LEDs; NSPWR70BS, Nichia Inc, Japan) placed in the wall of the sphere behind the surfaces produced uniform diffuse illumination via reflections off the white internal surface of the sphere. A baffle placed behind the surface ensured that the backlight did not shine directly into the cameras. We placed an array of individual high brightness (29 cd), spotlight LEDs (Nichia NSPW500DS) at specific locations in the wall of the sphere, facing the surface. Only 16 of the spotlights were used in this study located as shown in Fig. 3(a). The spotlights produced a bluer light than the backlight.

We first photographed the objects under the diffuse light only (condition 1), then under the diffuse light with each of the spotlights in turn, conditions l ($l = 2, \dots, 17$). Finally we used each spotlight alone, conditions $l + 16$. We also produced a matte grey version of every surface to provide ground truth shading images. Ground truth reflectance images were obtained by dividing images taken

under the diffuse light by the shading ground truth for each surface. Images were taken from a larger database (<http://www.bold.bham.ac.uk>).

6.2 Evaluation of Extraction Method and Cue Combinations

We extracted intrinsic images for the test set while using different cues within the algorithm. We used correlations between the estimated shading or reflectance images and their respective ground truth images as a metric for assessing results. Fig. 4(a), (b) show the performance of different cues for the HTC treatment condition under all 33 lighting conditions, where H, T and A indicate that HM, TM and AM streams were 'turned on' respectively. The combination of all three cues (HTA in Fig. 4(a), (b)) is better than any cue alone or any combination of two cues. The results show that local amplitude (A in Fig. 4(a), (b)) has an important role in detecting illumination changes. When it is combined with either hue or texture (HA and TA in Fig. 4(a), (b)), it boosts performance relative to either of these cues alone.

Fig. 4(c) further summarize the results. It shows improved extraction of shading from diffuse+spot images as streams are activated. Illumination changes are much weaker in these images. Consequently, small correlation coefficients between cues make the assignment of LM components less accurate when only using one or two cues. Fig. 4(a) shows marked differences in performance for individual light sources. The poor performance for images under frontal spotlights (azimuth 90° , conditions 20, 27 & 28) is caused by the same problem.

Conversely reflectance images are very good for diffuse+spot lighting and less good for spotlights only, see Fig. 4(c). We might expect reflectance estimates to be best for images that contain relatively little shading (Fig. 4(b) conditions 1-17)

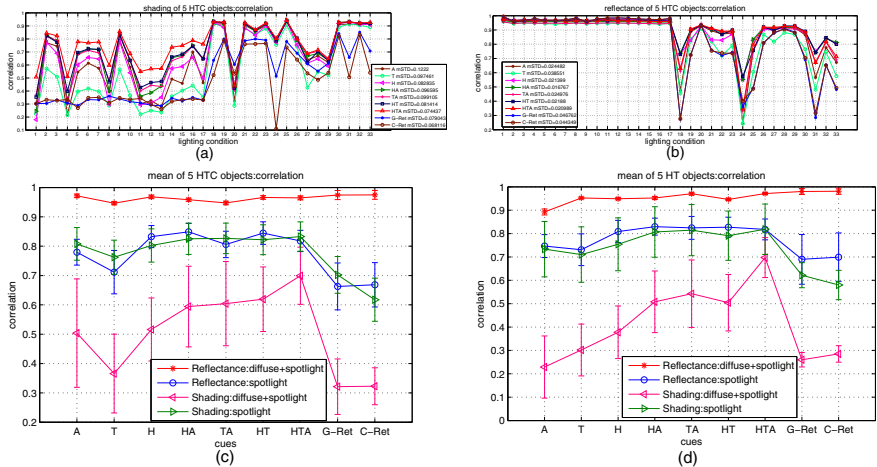


Fig. 4. Performance of different cues. Mean correlation against ground truth images for (a) shading and (b) reflectance estimates, which is the average of 5 HTC objects. (mSTD is the mean standard deviation of 5 objects across 33 lighting conditions). (c),(d) are correlation results on HTC and HT images respectively.

as only slight adjustments are needed in this case. Reflectance estimates are very poor (especially when only texture is used) for spotlights directed from the side (conditions 18, 24, & 31). These images are a special case where the light only glances on the surface much of which is in shadow, almost erasing any texture present. Fig. 4(d) summaries the results on HT surfaces. As we should expect AM alone worked better for HTC surfaces, where texture contrast and hence AM was negatively correlated with gross reflectance (LM) across texture boundaries, than HT surfaces where texture contrast did not vary. Combining cues remains helpful in this situation. Fig. 4(c),(d) also show the results of processing our stimuli with the grey-Retinex (G-Ret) and colour-Retinex (C-Ret) algorithms [20]. Our algorithm outperforms Retinex for these images.

6.3 Enhancement of Reconstructed Images

Based on the extraction of intrinsic images using all three streams (HTA) we tested the post-processing enhancements presented in Section 5. Recovering the DC component in the shading image does not alter its correlation with the ground truth stimuli but it does improve the estimate of reflectance obtained by dividing the original image by the reconstructed shading map (compare dRef and dRef-original in Fig. 5(f),(g) and examples in Fig. 5(b),(c)). Recovering the DC component of shading greatly improves it as a basis for further processing. The effects of enhancing reflectance is clearly seen by comparing Fig. 5(d) and (e). Improvements to the reconstructed reflectance images can be assessed more directly by comparing each reconstructed image (RecRef) with its compensated version (CompRef) here we see improved performance for spotlight images and the difficult side-lit cases (Fig. 5(f), (g), conditions 18, 24 & 31).

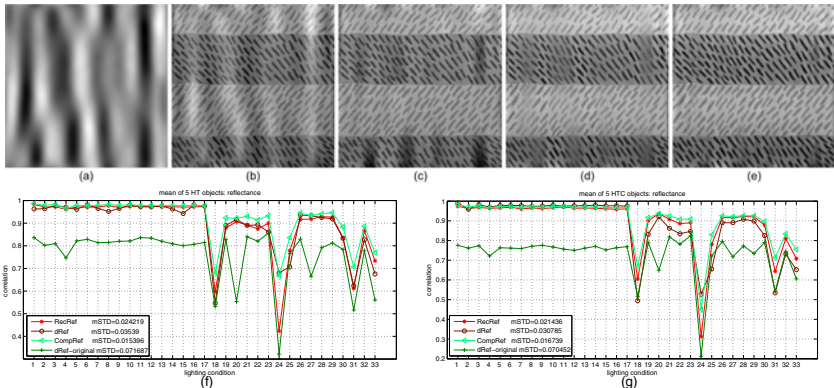


Fig. 5. Optimization results. (a) reconstructed shading for the image of Fig. 2(a), (b) reflectance derived from shading image (a). (c) reflectance derived from optimized shading. (d) reconstructed reflectance. (e) compensated reflectance. The mean correlation value between reflectance and ground truth for 5 different HT objects (f) and for 5 different HTC objects (g). mSTD is the mean stand deviation of the correlation for 5 objects across 33 lighting conditions.

6.4 Evaluation on Natural Images

Fig. 6 shows the performance of our algorithms compared to Retinex on the full MIT data set. The results for reflectance and shading have been averaged. The

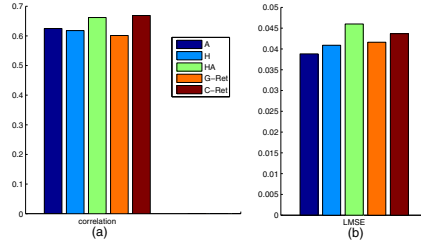


Fig. 6. Average results on the MIT data set

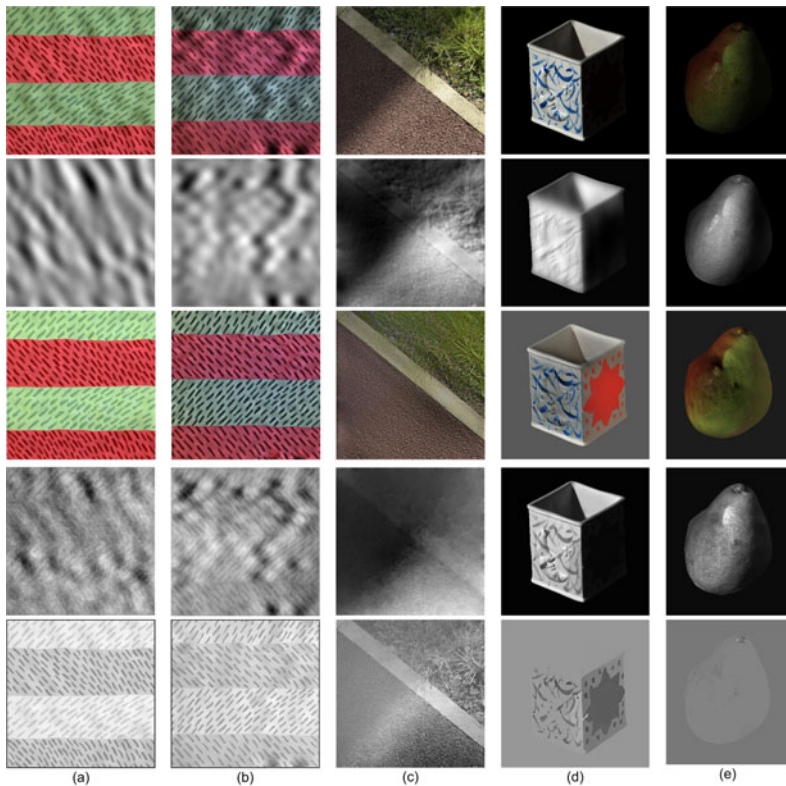


Fig. 7. Intrinsic images extracted by our algorithm and colour Retinex. Top to bottom in each panel: original image, shading and reflectance estimates from our algorithm and colour Retinex respectively. (a) isotropic ripples 1 with HTC texture, light condition $l = 14$; (b) isotropic ripples 2 with HT texture, $l = 19$; (c) a natural image. (d),(e) sample images from MIT data set.

LMSE measure [20], Fig. 6(b), favours our algorithm with only 'A' or 'H' (A low value is good.) while our correlation metric, Fig. 6(a), places our 'HA' algorithm equal to C-Retinex. (A high value is good.) We prefer our correlation metric as LMSE includes a term which can amplify the contrast of each local patch. This term is optimised on a local basis to reduce the MSE. Thus, it is possible to produce low LMSE scores from output images that do not visually match ground truth, as in the case of 'A' and 'H' only. The supplementary material provides a more detailed explanation.

Finally Fig. 7 shows some additional results based on our test set, a natural image where shadows fall across regions of different textures and colours, and two images from the MIT test set [20]. Example output from the colour Retinex method is also shown in Fig. 7. Our algorithm gives fairly accurate estimations for shading and reflectance. Although it does not perform well on the MIT 'box' stimulus (Fig. 7(d)), it outperforms Retinex on the 'pear' stimulus (Fig. 7(e)). From these results we see that our algorithm extracts low-frequency shading very well and is suited to natural textures. However the assignment of components is not accurate for high-frequency shading, due to the limitations of the frequency bandwidth of the steerable filters (about 1 octave). Furthermore, our algorithm currently deals with each component globally. This failing differentially affects high-frequency components which tend to arise from more local features.

7 Conclusion

In this paper, we propose an algorithm to extract shading and reflectance maps from a single input image. Based on results from human vision we propose local luminance amplitude (AM) as an effective cue for separating shading and reflectance along with texture and hue. We also introduce a multi-resolution framework to decompose images into components that can then separately contribute to the shading and reflectance maps. Correlation coefficients between luminance and the different cues (AM, texture, and hue) decide the weight for each component in each map. Experiments on images of rippled surfaces under different lighting conditions showed the effectiveness of the proposed algorithm in all but a few hard cases. In the proposed algorithm, we gave each component a global weight for each reconstruction, but shading or reflectance information may exist in only specific locations, such as at shadow edges. That is, correlations between different cues may be location-dependent. Thus the algorithm could be improved by adding a local correlation measure. Our method can produce two estimates for both shading and reflectance. One estimate is derived directly from the weighted components. The other can be derived by dividing the original image by the other reconstructed intrinsic image (e.g. shading=image/reflectance). Here, we used the reconstructed shading image to improve reflectance estimates, but did not provide an equivalent enhancement for shading. In future work, considering the four initial estimates together might improve estimates of shading and reflectance.

Acknowledgement. This project is supported by EPSRC grant EP/F026269/1.

References

1. Barrow, H.G., Tanenbaum, J.M.: Recovering intrinsic scene characteristics from images. *Computer Vision systems*, 3–26 (1978)
2. Land, E.H., McCann, J.J.: Lightness and retinex theory. *Journal of the Optical Society of America A* 61, 1–11 (1971)
3. Sinha, P., Adelson, E.: Recovering reflectance and illumination in a world of painted polyhedra. In: *ICCV*, pp. 156–163 (1993)
4. Bell, M., Freeman, W.T.: Learning local evidence for shading and reflectance. In: *ICCV*, vol. 1, pp. 670–677 (2001)
5. Funt, B.V., Drew, M.S., Brockington, M.: Recovering shading from color images. In: Sandini, G. (ed.) *ECCV 1992*. LNCS, vol. 588, pp. 124–132. Springer, Heidelberg (1992)
6. Olmos, A., Kingdom, F.A.A.: A biologically inspired algorithm for the recovery of shading and reflectance images. *Perception* 33, 1463–1473 (2004)
7. Tappen, M.F., Freeman, W.T., Adelson, E.H.: Recovering intrinsic images from a single image. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1459–1472 (2005)
8. Parraga, C.A., Troscianko, T., Tolhurst, D.J.: Spatiochromatic properties of natural images and human vision. *Current Biology* 12, 483–487 (2002)
9. Finlayson, G.D., Hordley, S.D., Lu, C., Drew, M.: On the removal of shadows from images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 28, 59–68 (2006)
10. Kingdom, F.A.A.: Perceiving light versus material. *Vis. Res.* 48, 2090–2105 (2008)
11. Shen, L., Tan, P., Lin, S.: Intrinsic image decomposition with non-local texture cues. In: *IEEE Computer Vision and Pattern Recognition*, pp. 1–7 (2008)
12. Weiss, Y.: Deriving intrinsic images from image sequences. In: *ICCV*, vol. 2, pp. 68–75 (2001)
13. Agrawal, A., Raskar, R., Chellappa, R.: Edge suppression by gradient field transformation using cross projection tensors. In: *CVPR*, vol. 2, pp. 2301–2308 (2006)
14. Matsushita, Y., Lin, S., Kang, S.B., Shum, H.Y.: Estimating intrinsic images from image sequences with biased illumination. In: Pajdla, T., Matas, J.(G.) (eds.) *ECCV 2004*. LNCS, vol. 3022, pp. 274–286. Springer, Heidelberg (2004)
15. Schofield, A.J., Hesse, G., Rock, P.B., Georgeson, M.A.: Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research* 46, 3462–3482 (2006)
16. Simoncelli, E.P., Freeman, W.T.: The steerable pyramid: A flexible architecture for multi-scale derivative computation. In: *ICIP*, pp. 444–447 (1995)
17. Comanicu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 603–619 (2004)
18. Reinhard, E., Devlin, K.: Dynamic range reduction inspired by photoreceptor physiology. *IEEE Trans. Visualization and Computer Graphics* 11, 13–24 (2005)
19. Waltz, R.A., Morales, J.L., Nocedal, J., Orban, D.: An interior algorithm for non-linear optimization that combines line search and trust region steps. *Mathematical Programming* 107, 391–408 (2006)
20. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In: *ICCV* (2009)