Sensor Transforms for Invariant Image Enhancement

Mark S. Drew[1], *Chao Chen*[1], *Steven D. Hordley*[2], *and Graham D. Finlayson*[2]
[1]*School of Computing Science, Simon Fraser University Vancouver, British Columbia V5A 1S6 Canada*
{*mark,chaochen*}*@cs.sfu.ca*
[2]*School of Information Systems, The University of East Anglia, Norwich NR4 7TJ United Kingdom*
{*steve,graham*}*@sys.uea.ac.uk*

Abstract

The invariant image [1, 2] formed from an RGB image taken under light that can be approximated as Planckian solves the colour constancy problem at a single pixel. The invariant is a very useful tool for possible use in a large number of computer vision problems, such as removal of shadows from images [3]. This image is formed by projecting log-log chromaticity coordinates into a 1D direction determined by a calibration of the imaging camera. The invariant can be formed whether or not gamma-correction is applied to images and thus can work for ordinary webcam images, for example, once a self-calibration is carried out [3]. As such, the invariant image is an important new mechanism for image understanding. Since the resulting greyscale image is approximately independent of illumination, it is impervious to lighting change and hence to the presence of shadows. However, in forming the invariant image, it can sometimes happen that shadows are not completely removed. Here, we consider the problem of simple matrixing of sensor values so that the resulting invariant image is improved. To do so, we consider the calibration images and apply an optimization routine for establishing a 3×3 matrix to apply to the sensors, prior to forming the invariant, with an eye to improving lighting invariance. We find that an optimization does indeed improve the invariant. The resulting image generally has smaller entropy value because the invariant value is smoothed out across former shadow boundaries; thus the new invariant more smoothly captures the underlying intrinsic reflectance properties in the scene.

1. Introduction

The *invariant image* is formed from a colour RGB image in a very simple fashion, First, one forms ratios R/G and B/G (or a variant which uses division by the geometric mean of R,G,B [4]). If lighting is approximately Planckian, then in Wien's approximation the simple exponential form of the illuminant SPD leads to the conclusion that as temperature T changes, characterizing illuminant colour, a log–log plot of the 2-dimensional {log(R/G), log(B/G)} values for any single surface forms a straight line. Thus lighting change reduces to a linear transformation along an almost straight line, even for real data with only approximately Planckian lighting. For many patches, mean-subtracted log-log plots all cluster around a single line through the origin that characterizes lighting change. The invariant image is the greyscale image that results from projecting log-log pixel values onto the direction orthogonal to lighting change. Since shadows are approximately derived from lighting change, within and outside the umbra, the invariant image greatly attenuates the shadowing.

As well, the second assumption used in justifying such a line is the assumption that sensors are quite narrowband. However, even if sensors are indeed not narrowband then they can be "sharpened" using a 3×3 sensor sharpening matrix [5].

However, here we would wish to carry out a similar simple matrix transform on camera sensors in order not to explicitly sharpen sensors, but instead to best improve the underlying assumption behind the development of the invariant image. I.e., here we explicitly seek to enhance the linearity of such log-log plots by means of a simple matrix transform of RGB values.

Of course, such a transform amounts to a sensor transform for the camera itself. Here we apply a number of techniques learned from finding the best sharpening transform characteristics for digital cameras [6]. We find that the most straightforward and fast optimization derives from straightening the log-log curves themselves, rather than using an optimization objective based on 1D projected values of calibration image patches. The minimization is fast, and results in lighting-change curves clustering much closer to a single lighting-change direction for all patches.

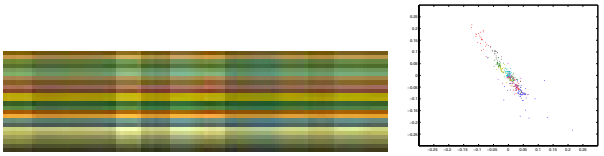
Once a calibration is carried out, for a single target of patches such as the Macbeth ColorChecker, then the *same* lighting-invariant direction can be applied to any new image, for the camera that was calibrated as above.

2. Invariant Image Formation

Consider image formation for a Lambertian surface. Suppose there are $i = 1..L$ lights, each with SPD $E^i(\lambda)$ well-described by Wien's approximation of a Planckian source [7], with

$$E^i(\lambda) = I^i c_1 \lambda^{-5} e^{-c_2/(\lambda T)} \quad (1)$$

in directions $\boldsymbol{a}^i$ with intensity I^i. (The Planckian assumption is in fact not crucial [2, 1].) If the surface projecting to retinal point x has spectral surface reflectance $S(\lambda)$ then, for a 3-channel delta-function narrowband sensor camera with spike sensor sensitivities $q_k = q_k(\lambda_k)$, $k = 1..3$, the



(a) (b)

Figure 1: *(a): Macbeth ColorChecker, HP912 camera, 14 lights. (b): Mean-subtracted log-ratios, HP912 camera.*

RGB-vector response would be

$$\begin{aligned}\rho_k &= \sum_{i=1}^{L} c_1 \boldsymbol{a}^i \cdot \boldsymbol{n}\, S(\lambda_k) I^i \lambda^{-5} e^{-c_2/(\lambda_k T)}\, q_k \\ &= \left[\sum_{i=1}^{L} (c_1 I^i \boldsymbol{a}^i)\right] \cdot \boldsymbol{n}\, S(\lambda_k) \lambda^{-5} e^{-c_2/(\lambda_k T)}\, q_k \\ &\equiv \widetilde{\boldsymbol{a}} \cdot \boldsymbol{n}\, S(\lambda_k) \lambda_k^{-5} e^{-c_2/(\lambda_k T)}\, q_k \quad . \end{aligned} \tag{2}$$

Now consider band-ratio chromaticities r_k defined as the ratio of the $k = 1..2 \equiv \{R, B\}$ sensor values ρ_1, ρ_3 divided by the green sensor value ρ_2. Thus our chromaticities are 2-vectors:

$$r_k = \frac{\rho_k}{\rho_G} \quad , \quad k = 1..2 \quad . \tag{3}$$

Since we divide, the shading term $(\widetilde{\boldsymbol{a}} \cdot \boldsymbol{n})$ is removed. Taking logarithms, we can concisely state the result by defining several terms:

$$\begin{aligned} e_\mu &= -c_2/\lambda_\mu \;,\; s_\mu = S(\lambda_\mu) \;, \\ u_\mu &= \lambda_\mu^{-5}\, q_\mu \;,\; w_\mu = \log(u_\mu/u_p) \qquad \mu = 1..3 \quad . \end{aligned} \tag{4}$$

Then for $k = 1..2$ we have

$$\log r_k = \log(s_k/s_2) + w_k + (e_k - e_2)\frac{1}{T} \;. \tag{5}$$

The meaning of the above expression is that, for every surface patch, the 2-vector is formed as a constant vector plus a vector $(e_k - e_2)$ times the inverse temperature: as lighting changes (i.e., as temperature T changes), pixel values are constrained to a straight line in 2D log colour space. For broadband sensors, similar equations hold [4]).

The main issue in finding an invariant to temperature T, then, becomes finding 2D vector $(e_k - e_2)$ in eq. (5). This is a calibration task: for a given camera (and gamma correction setting) we measure a colour target under many different lights — it would be sensible to restrict considerations to lights that fall close to Planckian, at least in terms of lying close to the Planckian locus on the chromaticity diagram. Here we use outdoor lighting.

Consider the image in Fig. 1(a). Here we have imaged a Macbeth ColorChecker in 14 different outdoor lighting conditions: each of the 24 patches is shown as a row, with each patch having 14 columns corresponding to the lights. The camera used was an experimental Hewlett-Packard HP912 digital still camera modified to produce raw output. The figure shows median values over each patch, under each light: since not all 14 images could be taken in one session, images of the ColorChecker were not registered. To find vector $(e_k - e_2)$ we first form ratios and then logs. For each patch, the mean is subtracted from all 14 data points, so that data is clustered around a single line through the origin, as in Fig. 1(b).

Then in [2] it is suggested that we find the dominant direction by simply forming the Singular Value Decomposition (SVD) of the data points shown— the first eigenvector is the main direction. However, instead of a least-squares based regression we make use of a robust Least Median of Squares regression, since such a robust method will find the best line without regard to outlier values that may arise during the optimization.

3. Optimizations on Log-Log Plots versus Optimizations on Invariant Values

In [6] we considered a number of optimizations for spectral sharpening with positivity, and here we carry over some of the same techniques.

We wish to find a simple linear transform of camera sensors such that the invariant image derived from patches in Fig. 1(a) is essentially constant across each row (i.e., across lights). The invariant image that corresponds to the original collection of patch images can be most easily visualized as in Fig.2, where invariant values are shown as a height map. We wish each line across the image, corresponding to a single patch, to be as constant as possible.

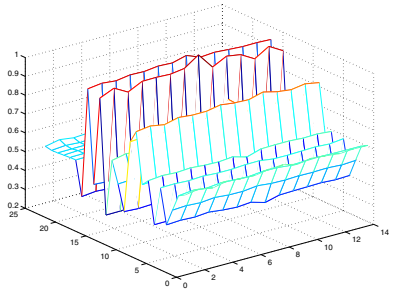


Figure 2: *Invariant image values, HP912 camera.*

We have found that another visualization scheme works very well: we map values to a colour map and display those. Fig. 3 shows such an image for the invariant patches of the ColorChecker, with the invariant derived as in [2]. We can see from Fig. 3 that the invariant is not truly constant across illuminants.

Now we wish to determine a simple 3×3 matrix $\boldsymbol{M}$ such that, after matrixing, the new sensors produce a more constant invariant. I.e., phrasing the problem in terms of sensors for the moment, rather than RGB values, if we measure sensor sensitivities using 31 samples from 400nm to 700nm at 10 nm intervals, then we arrive at a 3×31 matrix $\boldsymbol{Q}$. A matrixing scheme would be given in terms of 3×3 matrix $\boldsymbol{M}$ as forming new sensors $\widetilde{\boldsymbol{Q}}$ as combinations of the old sensors:

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}\,\boldsymbol{M} \;. \tag{6}$$

This can immediately be translated into how RGB values change by transformation to the new space. Suppose we have a set of $N \times 3$ colours $\boldsymbol{\rho}$. Then in the new colour space, these values are transformed to values

$$\widetilde{\boldsymbol{\rho}} = \boldsymbol{\rho}\,\boldsymbol{M} \;. \tag{7}$$

To form the invariant image for the new colours, we again take ratios and logs, find the direction of lighting change, and project orthogonal to that direction.

The main issue in forming an optimization scheme to improve the invariant is in determining what to optimize, and where to initialize. We can start by trying to optimize

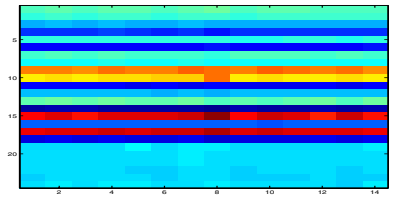


Figure 3: *HP912 Invariant image as values in a color map.*

the invariant image itself; i.e., we could try to force Fig. 2 to have straighter row plotlines, using an error measure such as

$$invt\_image\_error \;=\; \sum_{i=1}^{N}\sum_{j=1}^{L}(x_{ij}-\bar{x}_i)^2/\bar{x}_i \qquad (8)$$

where, for invariant values x_{ij}, $\bar{x}_i$ is the mean over all illuminants $j = 1..L$ for patch i, over $i = 1..N$ patches (here, $N = 24$ and $L = 14$). Here we wish to minimize $invt\_image\_error$ over a range of possible matrices $\boldsymbol{M}$.

The issue of initialization is important since generally such an optimization will find only a local minimum. The first point to note is that since eq. (5) depends on the narrowband sensor assumption for accuracy, the idea of spectral sharpening is pertinent [5]. An important consideration is that any RGB values created in the new colour space should be nonnegative [6] — we can carry out a constrained optimization either in a sensor space, if we know the sensors, or using a data-driven spectral sharpening approach [5].

We can expect that sharpening should help our error invariant and indeed this may be the case. On the other hand, this is not guaranteed. To examine sharpening, consider what happens when we simply raise the colour-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ [7], to some power. Certainly, for positive powers we see considerably narrower sensors, as in Fig. 4(a) (each sensor is normalized in the L1 norm). Fig. 4(b) shows the error values for these curves. We note that for negative powers, when the curves become flatter, the error in the invariant image also drops. This should not really be surprising, since in the process of finding the invariant we form colour ratios. Certainly, if we allow unconstrained mixing, the "best" invariant would be formed by sensors that all have the same mix of the original RGB sensors. That way, every invariant value would be zero.

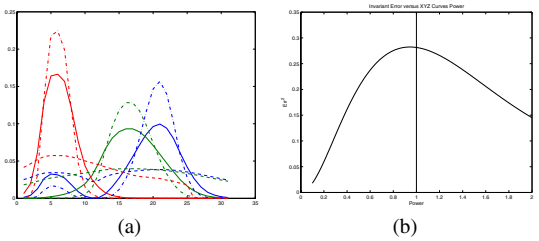
Thus we see that we must constrain an optimization from deriving sensors that are too rank-reduced. The matrix rank (integer) of the new set of RGB values $\tilde{\boldsymbol{\rho}}$ is a product of the rank of the original RGBs $\boldsymbol{\rho}$ and the rank of the transform matrix $\boldsymbol{M}$. Therefore our restriction amounts to encouraging the rank of $\boldsymbol{M}$ to remain near the rank of an initializing full-rank matrix $\boldsymbol{M}_0$.

Hence an optimization can be stated as

$$\min_{M} \;\; \sum_{i=1}^{N}\sum_{j=1}^{L}(x_{ij}-\bar{x}_i)^2/\bar{x}_i \;-\; \alpha \cdot \mathrm{rank}(\boldsymbol{M}\,)$$

with constraints

$$\begin{cases} \sum_{j=1}^{3} M_{ij} \;=\; 1, & i = 1..3,\ \mathrm{L}_1 \text{ normalization} \\ \boldsymbol{\rho}\,\boldsymbol{M} \;\geq\; \mathbf{0}\,, & \text{non-negative colours} \end{cases} \qquad (9)$$



(a) (b)

Figure 4: *Powers of colour-matching functions. (a): Curves (sum==1); (b): Errors for powers of curves from power=0.1 to 2.0: vertical line is power=1.0 .*

The objective of the first constraint is to limit the linear combinations of sensors allowed: only sensors that form a convex sum are allowed. The second constraint states the reasonable property that transformed colours are positive. The second term in the objective function is meant to encourage a non-rank-reducing matrix $\boldsymbol{M}$. To define a non-integer rank function, we take

$$rank \;=\; \lambda_3/\lambda_1 \quad , \qquad (10)$$

where λ_i, $i = 1..3$ are the singular values in an SVD decomposition of matrix $\boldsymbol{M}$, and use a constant α to control the optimization (here we set α=0.1).

An alternative optimization is to recognize that the original problem, that of reducing as much as possible the lines in Fig. 1(b) to a single straight line, amounts to the same objective. That is, we could instead use an optimizer

$$\log r_k \;\rightarrow\; \text{mean-subtracted } \log r_k \quad ;$$
$$\min_{M} \;\; R^2(\log r_k) \;-\; \alpha \cdot rank(\boldsymbol{M}\,)$$

with constraints

$$\begin{cases} \sum_{j=1}^{3} M_{ij} \;=\; 1, & i = 1..3,\ \mathrm{L}_1 \text{ normalization} \\ \boldsymbol{\rho}\,\boldsymbol{M} \;\geq\; \mathbf{0}\,, & \text{non-negative colours} \end{cases} \qquad (11)$$

where R^2 is the coefficient of determination of the second dimension in the log-log plot versus the first dimension — how well is the data explained by a single line through the origin: $R^2 = sum\_of\_squares$ of the data minus that of the residuals.

In practice, we found that the second optimizer, eq. (11), converged much more rapidly than did the first, eq. (9), and hence that is the optimizer used here. Optimization was by gradient-descent.

For colour patches created synthetically, for XYZ sensors imaging the Macbeth chart under illuminants A, C, and the Judd daylights, the optimization yields sensors as in Fig. 5. The effect seems rather minor, but the straightening of curves in log-log space is dramatic: the coefficient of determination R^2 (square of the correlation coefficient) without a sensor transform is only 0.4449 because lighting changes log-log values quite nonlinearly, whereas after a transform it is increased to 0.9661 . This large change is apparent in the log-log plots in Fig. 6. [1] Generally, we found that the best optimization, for both the XYZ curves

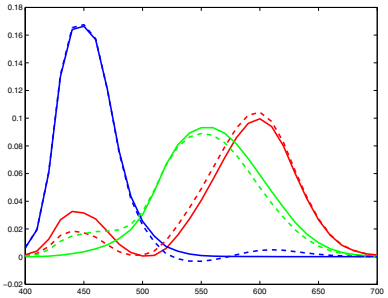
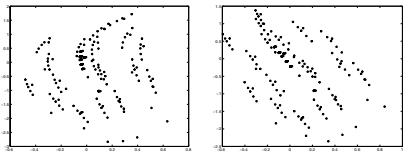


Figure 5: *Transformed XYZ curves.*

[1] Note that we do not need to know the sensors to do the optimization. We show them here because the curves help us see the balancing of sharpening versus equal-mixing.

and sensors for the Kodak DCS420 digital camera that we also tested, was given by the optimization (11) with initialization at spectrally sharpened sensors.



(a) (b)

Figure 6: *Log-log plots for XYZ sensors: (a): No $\boldsymbol{M}$ transform. (b): With transform.*

We also carried out data-driven sharpening on RGB values for the Kodak DCS420 digital camera, and then optimization. The camera sensors change as in Fig. 7, and straightening is as in Fig. 8. For this camera, R^2 goes from 0.4343 to 0.9368.

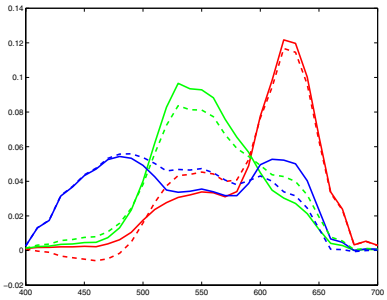
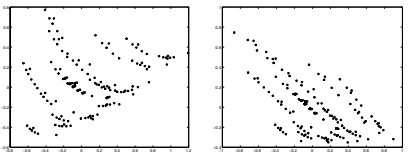


Figure 7: *Transformed Kodak DCS420 sensors.*



(a) (b)

Figure 8: *Log-log plots for Kodak DCS420 camera: (a): No $\boldsymbol{M}$ transform. (b): With transform.*

4. Results

4.1. Transformed Macbeth Patches

Getting back to our experimental data for the HP912 camera, the optimization using eq. (11) leads to a matrix $\boldsymbol{M}$ that substantially straightens the collection of lines in log-log space. Fig. 9(a) shows the set of colour patches transformed by matrix $\boldsymbol{M}$. For the HP912 camera, we found that an initial matrix given by data-based sharpening was did not give as tight a set of log-log lines as did an initializer using $diag(1, 1, 1)$, so we simply used the unit matrix to begin the search. Interestingly, an SVD decomposition of the original set of RGB patches yields singular values in the ratio 1.0, 0.182, 0.071, so the patches themselves are quite

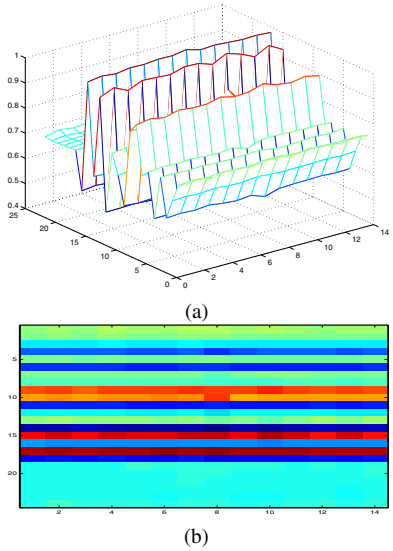
(a) (b)

Figure 9: *(a): Macbeth ColorChecker under transformed HP912 camera. (b): Transformed log-log plot for HP912 camera.*

rank-reduced. The matrix $\boldsymbol{M}$ has singular values in the ratios 1.0, 0.807, 0.579, and the resulting set of RGBs gives ratios 1.0, 0.155 0.043 . Thus some rank-reduction is indeed taking place.

Nevertheless we do achieve the goal of a substantial increase in the constancy of invariant values for each patch: Fig. 9(b) shows that indeed the lines in a log-log diagram do become tighter around the new illumination-change direction. (The new direction also has a slope 7° greater than the original direction.) The coefficient of determination without a sensor transform is 0.8647, whereas after a transform it is increased to 0.9254 . As well, Fig. 10 shows that in either visualization, the invariant patch values (1D, greyscale) are more constant compared to Figs. 2, 3.

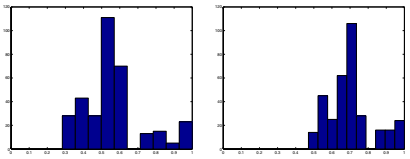
Since the rows in Fig. 10 are smoothed out, there are fewer values in the invariant image. Hence the histogram for the invariant image is more concentrated. Fig. 11 shows the histograms for the original and the improved invariant image: the entropy in the original is 5.856 whereas this value is decreased to 5.590 in the new invariant. This reflects the fact that the latter is smoother. Generally, the entropy measure for the invariant image provides a simple indicator of how well we have done in smoothing out changes over illumination change boundaries.



(a)

(b)

Figure 10: *HP912 camera: (a): Invariant colour patch values for transformed sensors. (b): Invariant from colour patches values in a color map.*



(a) (b)

Figure 11: *HP912 camera: (a): Histogram of original invariant values from colour patches. (b): Histogram of invariant values for transformed sensors.*

4.2. Transformed Images

Consider the image in Fig. 13(a). The second figure, (b), shows the invariant image, found using the method in [2]. The image is histogram-equalized for display. Fig. 13(c) shows the invariant derived from the image transformed by matrix $\boldsymbol{M}$. Note that once a set of patch images has been calibrated, as in §4.1 above, we need not re-compute $\boldsymbol{M}$ again.

To compare the two invariant images, we can plot the mean over columns of a horizontal scan from the sunny part of the grass in the image to the shady side. The standard deviation (dividing by the mean of the subimage) goes from 0.0208 with no transform to 0.0121, so the new invariant does about twice as well. We can understand this result by considering the edge maps produced by a Canny edge-finder (with high threshold .08 and standard deviation of the Gaussian filter 1.0, applied to both old and new invariant images), in Figs. 13(d,e). The edge map is smoother across the shadow edge, for the new invariant: the change in texture now occurs not at the shadow boundary but instead at places where the grass in shade changes because it is more trampled down (see Fig. 13(f)).

The entropy in the original invariant image is 5.295, and is reduced to 4.939 in the new invariant; this is clear from a histogram of pixels values, in Fig. 12. Thus the new

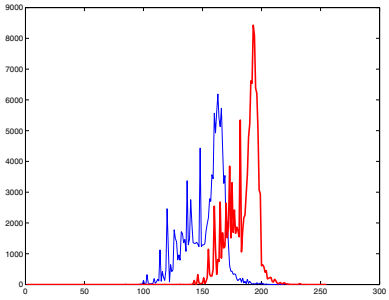


Figure 12: *HP912 camera: Histograms for real image; original (blue) and new (red– the rightmost plot) invariant images.*

invariant is essentially a smoother version of the underlying intrinsic reflectance properties in the scene.

The properties in this typical image are carried over to the other images shown below.

5. Linearized Images

Since users will generally not have available linear data, we considered linearizing output JPEG images from the HP 912 camera, and then forming an invariant. To do so, we created two 24×14 images of the ColorChecker, one linear and one typical output JPEG. We then "linearized" the set of patch RGBs via a function fit of the form

$$F(x) = -C_1 \cdot \log\left(\frac{C_2}{f^{C_3}(x) + C_4} - 1\right),$$

$$x = R \text{ or } G \text{ or } B,\ f(x) = C_5 * R + C_6 * G + C_7 * B$$

and applied the same linearization to JPEG camera images.

Finding a matrix $\boldsymbol{M}'$ for the linearized ColorChecker, we applied it to real image RGBs. Fig. 15(a) shows linearized versus correct linear Blue channel pixel values, for the image Fig. 13(a). Fig. 15(b) shows the output image using a matrix $\boldsymbol{M}$ derived from the original JPEG ColorChecker patches, and Fig. 15(c) shows the invariant for a linearized image with application of a matrix $\boldsymbol{M}'$ derived from the linearized ColorChecker. Clearly, linearization of JPEG images results in a better invariant.

6. Conclusions

In this paper we have outlined an optimization method that promotes creation of a more shadow-invariant image. In Fig. 14 we show a few more input and output images (for purposes of display, the raw input camera RGB values have been converted to the standard sRGB[8] colour space). For these images again, the entropy is decreased, and shadow boundaries are attenuated. In future, we wish to explore more fully the space of possible optimizations, with a view to developing techniques that can be applied to unsourced imagery from the web, rather than controlled imagery from a single camera.

Acknowledgements

Mark Drew gratefully acknowledges support of the Natural Sciences and Engineering research Council of Canada, and Graham Finlayson and Steven Hordley acknowledge support of the Hewlett-Packard corporation.

References

[1] G.D. Finlayson, S.D. Hordley, and M.H. Brill. Illuminant invariance at a single pixel. In *8th Color Imaging Conference: Color, Science, Systems and Applications.*, pages 85–90. Society for Imaging Science & Technology (IS&T)/Society for Information Display (SID) joint conference, 2000.

[2] G.D. Finlayson and S.D. Hordley. Color constancy at a pixel. *J. Opt. Soc. Am. A*, 18(2):253–264, Feb. 2001. Also, UK Patent application no. 0000682.5. Under review, British Patent Office.

[3] G.D. Finlayson, S.D. Hordley, and M.S. Drew. Removing shadows from images. In *ECCV 2002: European Conference on Computer Vision*, pages 4:823–836, 2002. Lecture Notes in Computer Science Vol. 2353, http://www.cs.sfu.ca/∼mark/ftp/Eccv02/shadowless.pdf.

[4] G.D. Finlayson and M.S. Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *ICCV'01: International Conference on Computer Vision*, pages II: 473–480. IEEE, 2001.

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

[6] M.S. Drew and G.D. Finlayson. Spectral sharpening with positivity. *J. Opt. Soc. Am. A*, 17:1361–1370, 2000.

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

[8] http://www.srgb.com.

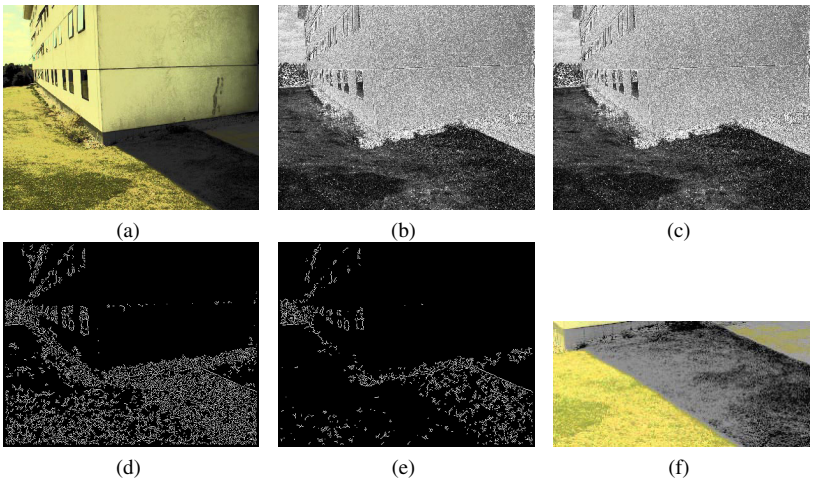


Figure 13: *(a): Original image. (b): Invariant image with no transform. (c): Invariant image using a transform. image, and red solid is the new invariant. (d): Edge map, original invariant. (e): Edge map, new invariant. (f): Detail of sun-shade boundary (histogram equalized).*

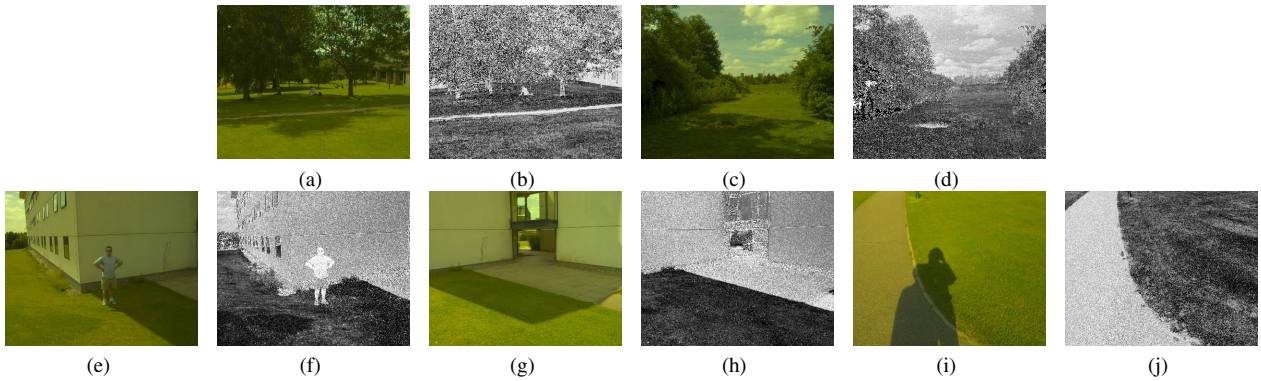


Figure 14: *Input and output images.*

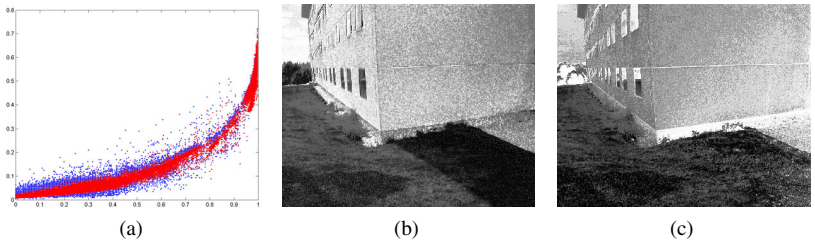


Figure 15: *LNonlinear and linearized image invariants. (a): Blue channel for Fig. 13(a): abscissa is nonlinear value, blue dots are correct linear values, and red dots are best-fit linearized values. (b): Invariant from nonlinear JPEG image. (c): Invariant from linearized image.*