Sharpening from Shadows: Sensor Transforms for Removing Shadows using a Single Image

Mark S. Drew and Hamid Reza Vaezi Joze
School of Computing Science, Simon Fraser University, Vancouver, British Columbia, Canada V5A 1S6
{***mark,hrv1***}***@cs.sfu.ca***

Abstract

Illumination conditions in images, such as shadows, can cause problems for both humans and computers. As well as shadows obscuring some features in images for human observers, many computer vision algorithms such as tracking, segmentation, recognition, and categorization are challenged by varying illumination. Previously, shadow removal algorithms were proposed that require recording a sequence of calibration images of a fixed scene over different illumination conditions, say over a day. As another alternative, calibration is replaced by using information in the single image itself, seeking a projection that minimizes entropy and allows one to generate a grayscale image that has shadows effectively eliminated. In this paper we wish to improve the entropy-based method by carrying out a sensor sharpening matrix transform first. In preceding work such a sensor transform for shadow removal was sought by utilizing many calibration images. Here, instead, we replace the calibration information by user interaction: we ask the user to identify two (or more) regions in a single image that correspond to the same surface(s) in shadow and not in shadow. Then using image data from these regions only, we generate a sensor sharpening transform via an optimization aimed at minimizing the difference between in-shadow and out-of-shadow pixel values once they are projected to grayscale. Again, entropy minimization is the driving force leading to a correct sensor matrix transform. Results show that, compared to using the camera sensors as-is, the sensor sharpening is beneficial for better shadow removal.

1. Introduction

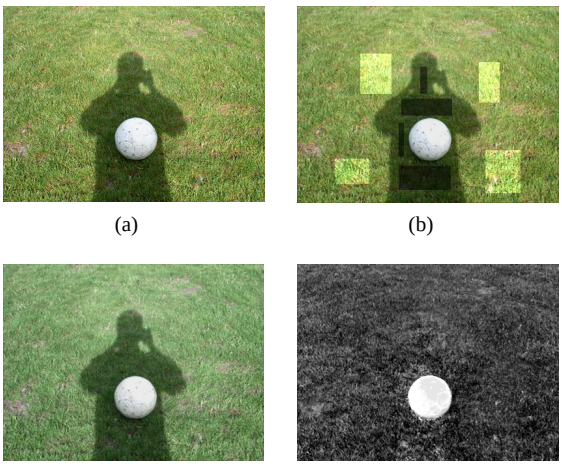
An illumination-invariant image [1] can be formed from a color RGB image in a straightforward fashion, provided enough data is gathered. First, in one embodiment one forms ratios R/G and B/G (or a variant which uses division by the geometric mean of R,G,B [2]). If lighting is approximately Planckian, then in Wien's approximation the simple exponential form of the illuminant SPD leads to the conclusion that as temperature T changes, characterizing illuminant color, a log-log plot of the 2-dimensional band-ratio chromaticity $\{log(R/G), log(B/G)\}$ values for any single surface forms a straight line. Thus lighting change reduces to a linear transformation along an approximately straight line, even for real data with non-Planckian, real lighting. Using many image patches, e.g. images of a Macbeth ColorChecker target, if we collect image information across many color temperatures, say using a light-box or simply forming images outdoors over different phases of daylight, then mean subtracted log-log 2D-chromaticity plots all cluster around a single line through the origin that characterizes lighting change. The invariant image is the grayscale 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.

However, the trick is to find the correct direction in which to project. Calibration data supplies that information. In another variant, it is noted that, if indeed one knew the correct projection direction in chromaticity color space, then pixels from patches of the same material but under different lighting would project to similar values. Hence for correct projection, the probability distribution of projected values would consist more of peaks at equal projected values, rather than a broader distribution. Thus the entropy would be minimized for the best projection direction.

The second assumption used in justifying the notion that changing color-temperature generates a straight line plot is the assumption that sensors are quite narrow-band: this implies that a logarithm of the inverse color-temperature in the Planckian law simply converts to a linear term in $1/T$ for a straight line. So narrow-band sensors more faithfully follow the model used in motivating this log-log line formulation, and sharpened sensors remove shadows better. Moreover, even if sensors are indeed not narrow-band then they can be sharpened using a 3×3 sensor sharpening matrix [3].

However, here we would wish to carry out a similar simple matrix transform on camera sensors to implicitly sharpen sensors, but not as the driving goal. Instead, we want to explicitly matrix the sensors such that shadow removal is benefited. I.e., we mean to best improve the underlying assumption behind the development of the invariant image. Therefore, 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. (We do not require knowledge of the camera sensors, since we utilize only RGB values, but the effect is the same.) Note that a generic, data-independent, sharpening matrix is derivable [4], but here we wish to do better by using the image data itself.

In previous work, a similar sensor sharpening transform had been accomplished [5, 6]. However, in that method extensive camera calibration data was used, indeed imaging a ColorChecker under many different lights. Here we mean to accomplish the same end, sensor sharpening to enhance shadow removal, but replacing such calibration data by instead using only a single im-



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

Figure 1. *(a): Original image; (b): User-supplied shadow/non-shadow regions; (c): Original image after matrixing; (d): Invariant image from color-space transformed image.*

age, and asking a user to identify in-shadow and out-of-shadow regions. No need of labeling of regions as corresponding need be carried out by the user, and no identification as in- or out-of shadows is required: instead, all in- and out-of-shadow region pixel values are simply concatenated, and again an entropy-minimization routine is employed. However, here entropy minimization is adopted in a new objective function in an optimization. The result is a single 3×3 matrix transform, for the single image under examination, that leads to enhanced shadow removal. Using only one image instead of an image sequence of a fixed scene over different illumination conditions provides a much easier and convenient mechanism for shadow removal from color camera still images, albeit it does require some simple user intervention.

Once we have obtained the optimum 3×3 matrix transform, $\boldsymbol{M}$, we apply that transform to the image (equivalent to having applied $\boldsymbol{M}$ to the sensors) and find the illumination invariant image using the algorithm proposed in [1, 7]. The result provides better attenuation of shadows. The output obtained is a 1-dimensional, grayscale image, illumination invariant and hence approximately shadow-free. To generate a full-color (3-dimensional color) shadow-free image, we could then go on to apply the method in [8], comparing edges in the original image and edges in 1-dimensional illumination invariant image. However, here we focus on the task of generating a better grayscale invariant image in the first place.

To see how this workflow proceeds, consider image Fig. 1(a), showing a strong cast shadow. This image was taken with no gamma correction, and in fact all processing (cf. [9]) except demosaicing turned off. Images are shown displayed in the sRGB color space [10]. The idea is that the user supplies in- and out-of shadow information by identifying (at least two) regions in the image, as in Fig. 1(b). The optimization proposed (see §3) then implies a color space transform, leading to a new color image as in Fig. 1(c): in the new color space, invariant image generation by entropy minimization is guaranteed to better reduce shadows. The projected grayscale output, for this image, is shown in Fig. 1(d).

2. Invariant Image Formation

To motivate sensor sharpening as an important mechanism for improving shadow removal, here we briefly recapitulate illumination invariant image formation.

Consider the RGB color formed at a pixel from 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 the RGB color $\boldsymbol{\rho}$ at any pixel results from an integral over the visible wavelengths:

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

If camera sensitivity $Q_k(\lambda)$ is exactly a Dirac delta function $Q_k(\lambda) = q_k\delta(\lambda - \lambda_k)$ (that is to say, a perfectly narrow-band sensor), with q_k the strength of the sensor $q_k = Q_k(\lambda_k)$, then the equation reduces to the simpler form

$$\rho_k = E(\lambda_k)S(\lambda_k)q_k \tag{2}$$

Now suppose lighting can be approximated by Planck's law:

$$E(\lambda, T) \approx I\, c_1\lambda^{-5}\left(e^{-\frac{c_2}{\lambda T}} - 1\right) \tag{3}$$

with light strength I. (Constants c1 and c2 equal $3.74183 \times 10^{-16}Wm^2$ and $1.4388 \times 10^{-2}m^\circ K$, respectively.) For illumination in the temperature range $2500K$ to $10000K$ the term $e^{-\frac{c_2}{\lambda T}} \gg 1$ and Wien's approximation [11] can be used:

$$E(\lambda, T) \approx I\, c_1\lambda^{-5}e^{-\frac{c_2}{\lambda T}} \tag{4}$$

Then returning to the narrow-band sensor response in eq. (2), RGB color ρ_k, $k = 1..3$, is simply given by

$$\rho_k = Ic_1\lambda_k^{-5}e^{-\frac{c_2}{\lambda_k T}}S(\lambda_k)q_k \tag{5}$$

Now suppose we first form band-ratio chromaticities from color values given by this equation, dividing ρ_R and ρ_B by ρ_G, and then take the logarithm:

$$\begin{aligned} r_k &\equiv log(\rho_k/\rho_G) \\ &= log(s_k/s_G) + (e_k - e_G)/T\ ,\ k = R, B \end{aligned} \tag{6}$$

where we define $s_k = c_1\lambda_k^{-5}S(\lambda_k)q_k$ and $e_k = -c2/\lambda_k$. As temperature T changes, 2-vectors r_k, $k = R, B$, will follow a straight line in 2-d chromaticity space. Moreover, this will be true for pixels that correspond to each surface spectral reflectance $S(\lambda)$, so that for all materials the lines will be parallel, with slope $(e_k - e_G)$.

Calibration then amounts to determining the 2-vector direction $(e_k - e_G)$ in the color space of logs of ratios. Note that as in all chromaticity operations, we effectively remove intensity information I. As well, in eq. (1), we omitted any multiplicative shading term, e.g., Lambertian shading, but this will also disappear in r_k. Note, however, that we did not include specular highlights, so these may in actual trials form a difficulty.

The invariant image, then, is formed by projecting 2-d colors into the direction $\boldsymbol{e}^{\perp}$ orthogonal to the 2-vector $(e_k - e_G)$. The result of this projection is a single scalar which we then code as a grayscale value. Since we have projected orthogonal to the direction of lighting change, shadows disappear.

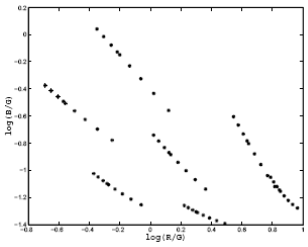


Figure 2. *Chromaticities for 7 different colors, imaged under a set of different Planckian illuminants*

Experiments have shown [8] that images of the same scene containing objects of different colors illuminated by any complex lighting field (including lights of different colors and intensities) will map to the same invariant image. Most importantly for this paper, shadows which occur when there is a change in light but not surface will disappear in the invariant image.

However, we draw the reader's attention to a possible problem. Specifically, the invariant is designed to work for Planckian lights. Additive combinations of Planckians (which might result indoors when there is mixed light from a Tungsten source, say, and outdoor illumination through a window) is non-Planckian. However, because the Planckian locus is a very shallow crescent shape, additive combinations of light tend to fall close to the locus. Experimentally, the invariant image factors out the effect of lighting even for additive combinations of Planckian illuminants [12].

Nevertheless, we still have a problem remaining: we have assumed narrow-band sensors, whereas typical sensor sensitivity curves for cameras can be quite broad [13]. The next section addresses this problem.

3. Finding the Sensor Transform

In the formation of an illumination invariant image, we need log chromaticity change with illumination to be linear with inverse color temperature $1/T$, and in turn this falls out from the assumption of Planckian lighting and sensors that are fairly narrow-band. However sensors are not typically narrow-band in reality, and hence the linear shift of chromaticities is only approximate. Fig. 2 demonstrates this property by displaying chromaticities for surfaces of 7 different colors (two are grays), imaged under different illuminants. Clearly, chromaticities for each surface do not fall exactly on a line, but nevertheless the curve might indeed be approximated as a line. We wish to make these curves as linear as possible in order to get a best possible result when applying the linearity assumption to form the illumination invariant image. We posit that this can be accomplished by linearly transforming the RGB values, effectively transforming the camera sensors.

The strategy proposed in this paper is to add additional information provided by the user, in the form of delineating (possibly a collection of) regions within shadows and outside shadows. These need not be labeled as belonging together in shadow/non-shadow pairs, but are simply concatenated. The idea is to include in a subset of pixels values from sun and shadow for the *same* material, e.g., grass in sun and grass in shadow. Since such pixels are from the same surface, but with different illumination conditions, they fall on a single line in a log-log plot. Each different material will fall on a different line, but all lines should be approximately parallel.

We wish to find a simple 3×3 matrix transform $\boldsymbol{M}$ of camera sensors such that these pixels are as linear as possible in the log-log plot.

Suppose that matrix $\boldsymbol{P}$ consists of the set of $n \times 3$, RGB values for n pixels selected by the user — both shadow and non-shadow pixels. $\boldsymbol{M}$ is the 3×3 matrix such that after matrixing, the new sensors produce a more constant invariant. Thus an optimization can be stated as follows:

$$\begin{aligned} \min_{\boldsymbol{M}} \quad & f(\boldsymbol{M}, \boldsymbol{P}) \; - \; \alpha \; \times \; rank(\boldsymbol{M}) \\ \text{subject to} \quad & \textstyle\sum_{j=1}^{3} M_{ij} = 1 \qquad i = 1..3 \\ & |M_{ii} - 1| \leq \beta \qquad i = 1..3 \\ & |M_{ij}| \leq \beta, \qquad i \neq j \end{aligned} \tag{7}$$

The objective of the first constraint is to limit the linear combinations of sensors allowed so only sensors that form a convex sum are allowed. The objective of the second and third constraints is to limit the changes in the sensor so that, e.g., red should be red after applying the transform, and similarly for blue and green: i.e., diagonal elements of the matrix transform should be limited to a small excursion from unity. The β parameter applies this limitation (here we set $\beta = 0.3$).

The first term in the objective function consists of a function $f(\boldsymbol{M}, \boldsymbol{P})$ to evaluate whether the pixels in $\boldsymbol{P}$ lie along a line in the log-log plane after application of matrix $\boldsymbol{M}$. The measure employed in this paper is to evaluate the entropy, for the minimum-entropy projection to grayscale for user-selected pixels, for any candidate matrix $\boldsymbol{M}$. That is, for any candidate $\boldsymbol{M}$, we project pixels $\boldsymbol{P}$ over all angles $\theta = 0..180^\circ$ (just a few angles are required in practice) and take as our objective function value f the least entropy η over θ. We repeat this calculation for all candidate matrices $\boldsymbol{M}$.

Suppose that function $\boldsymbol{\chi}$ returns 2-vector log chromaticities for an RGB triple. Then objective function f can be defined as follows:

$$f(\boldsymbol{M}, \boldsymbol{P}) = \min_{0<\theta<\pi} \eta(\boldsymbol{\chi}(\boldsymbol{P} \; \boldsymbol{M}) \begin{bmatrix} cos(\theta) \\ sin(\theta) \end{bmatrix}) \tag{8}$$

The second term in the objective function (7) is meant to encourage a non-rank-reducing matrix $\boldsymbol{M}$. To define a non-integer effective-rank function, we take

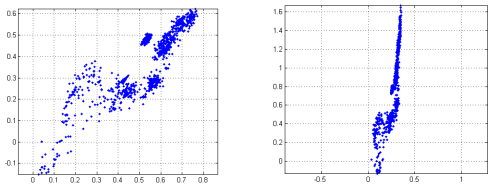
$$rank \; = \; \lambda_3/\lambda_1 \tag{9}$$

where λ_i, $i = 1..3$, are the singular values in an SVD decomposition of matrix $\boldsymbol{M}$, in decreasing order. We simply use a constant α to control the rank constraint in the optimization (we set it to 0.05).

Fig. 3 demonstrates the difference between images upon applying such a matrix transform: in effect, the color space is changed. To illustrate the effect of optimizing for shadow removal, Fig. 4 shows a plot of the user-selected pixels of the image of Fig. 3 in the log-log chromaticity plane, before and after applying the matrix transform. We note that the chromaticities are clearly much more linear in the new color space, after applying the matrix transform.

(a) (b)

Figure 3. *(a) Original image; (b) Image after matrixing with optimization result.*



(a) (b)

Figure 4. *(a) User-selected pixels including shadow and non-shadow values, in 2D-color log-chromaticity plot; (b) Selected pixels after matrixing.*

With knowledge of the user-selected shadow and non-shadow regions, both corresponding to the same materials, we are guaranteed by the optimization to arrive at a more linear set of lines in log-chromaticity space. Since such lines indeed correspond to both shadow and sun regions, projection will effectively remove shadows, or at least greatly mitigate them.

In the next section, we consider the shadow-removal capability of the proposed method.

4. Results

Fig. 5 shows some results, using optimization (7). The input image is shown on the left, followed by the result of the color space transform implied by application of sharpening matrix $\boldsymbol{M}$. The third image shows the shadow-free invariant image as found after matrixing, and the fourth image shows the difference between invariants before and after applying matrix $\boldsymbol{M}$. Clearly, shadow removal has been improved by our optimization.

Fig. 6 shows several more results: we found that in every case application of a color space transform, optimized for shadow removal, improved the grayscale invariant image. While results are not perfect, they do show greatly attenuated shadows.

Other images show similar improvement. Hence we can conclude that color space optimization before invariant image generation is indeed beneficial.

In order to evaluate this improvement we need to check whether the shadow and non-shadow regions have the same appearance. Since the shadow and non-shadow regions are usually selected to contain simple, textured contents, but with different windows sizes, we compute the intensity histogram of the two regions and demonstrate the degree of similarity between the two regions using the correlation measure. Such an evaluation does not focus on edges but instead evaluates the degree of similarity of the distribution of intensity over in- and out-of- shadow regions.

Table 1 shows the correlation of the two regions in some invariant images formed without as opposed to with applying matrix $\boldsymbol{M}$. The complete set of results indicate an average improvement

Image	without $\boldsymbol{M}$	with $\boldsymbol{M}$
House	0.9473	0.9598
Path	0.9867	0.9872
Ball	0.4454	0.9648
Arch	0.6762	0.8228

Table 1. Correlation of shadow and non-shadow regions in invariant images before and after applying matrix $\boldsymbol{M}$

of 41.9% in the correlation for the idea of applying a sharpening matrix before finding an invariant image. The improvement is most apparent when the the result without application of $\boldsymbol{M}$ is not a high correlation by itself.

5. Conclusion

With the objective of removing shadows from images, in this paper we have proposed a new schema for generating illumination invariant images. The idea is that the user identifies two or more image regions, in and out of shadows. Then a color space transform is determined via an optimization aimed at enhancing shadow removal. Results show that the method effectively removes or at least greatly reduces shadows in the output grayscale invariant.

In future, we will examine the effect of noise on the proposed scheme, in particular the effect of JPEG block artifacts and the effect of various JPEG de-blocking and de-noising methods.

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, 2000.

[2] 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.

[3] 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.

[4] M.S. Drew, M. Salahuddin, and A. Fathi. A standardized workflow for illumination-invariant image extraction. In *15th Color Imaging Conference: Color, Science, Systems and Applications*, 2007.

[5] 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.

[6] M.S. Drew, C. Chen, S.D. Hordley, and G.D. Finlayson. Sensor transforms for invariant image enhancement. In *Tenth Color Imaging Conference: Color, Science, Systems and Applications.*, pages 325–329, 2002.

[7] G.D. Finlayson, M.S. Drew, and C. Lu. Intrinsic images by entropy minimization. In *ECCV 2004: European Conference on Computer Vision*, pages 582–595, 2004. Lecture Notes in Computer Science Vol. 3023.

[8] G.D. Finlayson, S.D. Hordley, C. Lu, and M.S. Drew. On the removal of shadows from images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 28:59–68, 2006.

[9] R. Ramanath, W.E. Snyder, Y.F. Yoo, and Mark S. Drew.

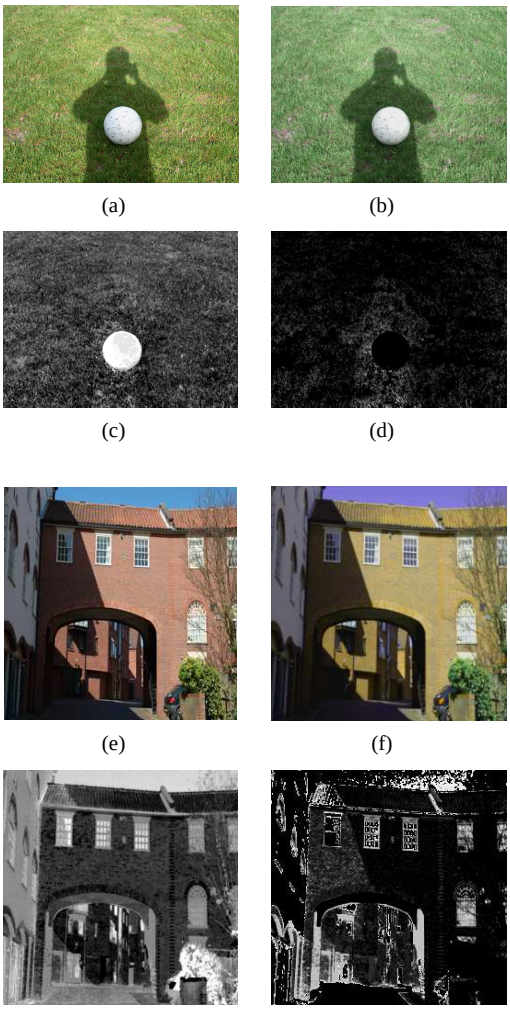


Figure 5. *(a,e) Original image; (b,f) Original image with matrix M applied; (c,g) Invariant of image with matrix M ; (d,h) Difference between invariant, without and with matrix M .*

Color image processing pipeline in digital still cameras. *IEEE Signal Processing*, 22(1):34–43, 2005.

[10] International Electrotechnical Commission. Multimedia systems and equipment – colour measurement and management – part 2-1: Colour management – default RGB colour space – sRGB. IEC 61966-2-1:1999.

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

[12] G.D. Finlayson and S.D. Hordley. Colour constancy at a pixel. *J. Opt. Soc. Am. A*, 18(2):253–264, Feb. 2001.

[13] P.M. Hubel, G.D. Finlayson, , J. Holm, and M.S. Drew. Sharp transforms for color appearance. In *5th Color Imaging Conference: Color, Science, Systems and Applications.*, 1997.

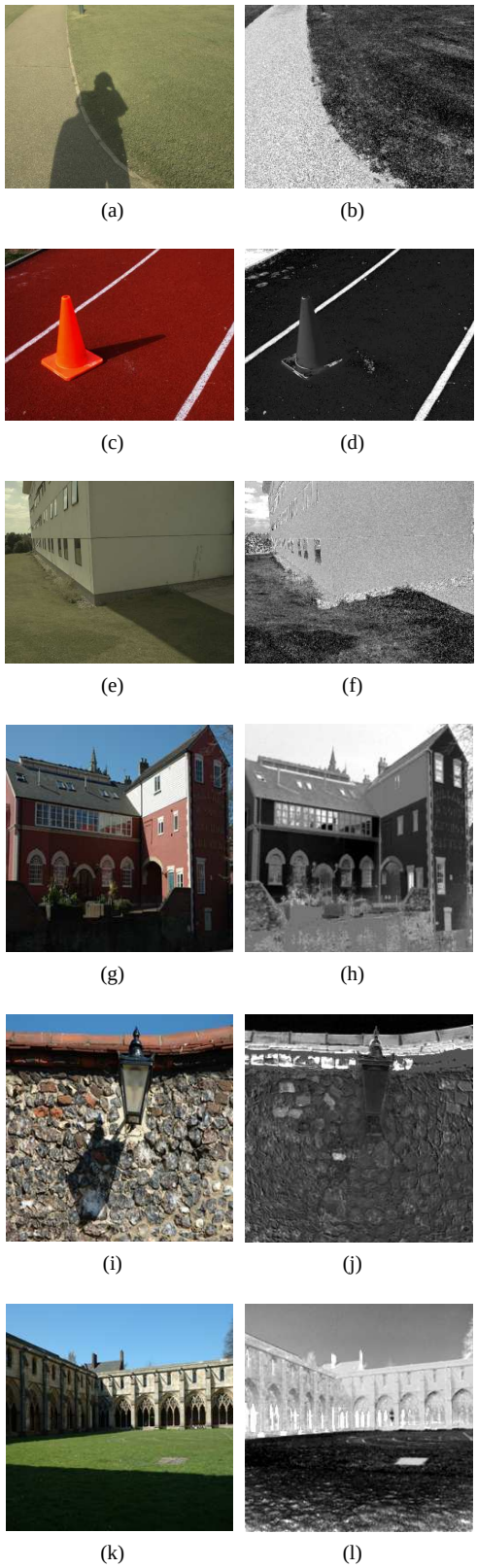


Figure 6. *Input and output images, using matrix M .*