

Illumination-invariant representation for natural colour images through SIFT matching

Frédéric Sur

▶ **To cite this version:**

Frédéric Sur. Illumination-invariant representation for natural colour images through SIFT matching. ICASSP - 38th International Conference on Acoustics, Speech, and Signal Processing, May 2013, Vancouver, Canada. pp.1962 - 1966, 10.1109/ICASSP.2013.6637996 . hal-00804736

HAL Id: hal-00804736

https://inria.hal.science/hal-00804736

Submitted on 26 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ILLUMINATION-INVARIANT REPRESENTATION FOR NATURAL COLOUR IMAGES THROUGH SIFT MATCHING

Frédéric SUR

LORIA - projet Magrit
Université de Lorraine, CNRS, INRIA, France

ABSTRACT

Illumination conditions may severely impair many computer vision applications. Defining an illumination-invariant representation from a colour image is therefore of great interest. Assuming that two views of a scene are available, we show that it is easy to derive such a representation from correspondences between interest points, within a Lambertian reflectance / Planckian lighting model. An application to shadow removal or lightening is also discussed.

Index Terms— Illumination-invariant representation, Lambertian reflectance, Planckian lighting, shadows, SIFT.

1. INTRODUCTION

This article explains how to derive illumination-invariant representations from two colour pictures of a 3D scene. Such a representation should not depend on the lighting intensity, colour, or direction. The article is based on a theory developed by Finlayson et al. [1], briefly recalled in this section.

In the Lambertian reflectance model and assuming that the camera sensitivity is a Dirac function in each channel and that the lighting is Planckian, the camera response in each channel (typically $k = R, G, B$) writes:

$$\rho_k(x,y) = q_k \sigma(x,y) E_{\lambda_k}(x,y) S_{\lambda_k}(x,y) \tag{1}$$

where λ_k is the wavelength to which channel k is sensitive, q_k is a constant across the image, $\sigma(x,y)$ is the Lambertian shading at pixel (x,y) (i.e. the dot product between the normal of the surface imaged at (x,y) and the illumination direction), $E_{\lambda_k}(x,y)$ is the spectral power distribution of the light at wavelength λ_k, and $S_{\lambda_k}(x,y)$ is the surface reflectance. Among these terms, only σ depends on the light direction.

Within Wien's approximation of Planck's law, E_λ is parametrized by the colour temperature T of the illuminant via: $E_\lambda(x,y) = c_1 \lambda^{-5} e^{\frac{-c_2}{T\lambda}} I(x,y)$ where c_1 and c_2 are constants and I is the overall light intensity.

It is possible to derive at each pixel (x,y) from an RGB image the so-called log-chromaticity values χ'_1 and χ'_2:

$$\chi'_1(x,y) := \log\left(\frac{\rho_R(x,y)}{\rho_G(x,y)}\right) = \log\left(\frac{s_R(x,y)}{s_G(x,y)}\right) + \frac{1}{T}(e_R - e_G) \tag{2}$$

$$\chi'_2(x,y) := \log\left(\frac{\rho_B(x,y)}{\rho_G(x,y)}\right) = \log\left(\frac{s_B(x,y)}{s_G(x,y)}\right) + \frac{1}{T}(e_B - e_G) \tag{3}$$

where $s_k(x,y) := q_k \lambda_k^{-5} S_{\lambda_k}(x,y)$ and $e_k := -c_2/\lambda_k$.

Consequently, the two-vector $\boldsymbol{\chi}'$ is parametrized by colour temperature T via $\boldsymbol{\chi}'(x,y) = \boldsymbol{s}(x,y) + \frac{1}{T}\boldsymbol{e}$ where $\boldsymbol{s}$ depends on the imaged surface and on the camera, and $\boldsymbol{e}$ only depends on the camera. Note that the direction of the light (in σ) and the light intensity I have been factored out

Finlayson et al's illumination-invariant representation is the 1D image $\mathcal{I}'$ obtained by factoring out the information along the line spanned by $\boldsymbol{e}$: $\mathcal{I}'(x,y) = \boldsymbol{\chi}'(x,y)^T \cdot \boldsymbol{e}^\perp$. A 2D invariant representation is also defined as: $\widetilde{\boldsymbol{\chi}'} = P_{\boldsymbol{e}^\perp}\boldsymbol{\chi}'$ where $P_{\boldsymbol{e}^\perp}$ is a projection in the 2D log-chromaticity space onto a line orthogonal to $\boldsymbol{e}$ in 2D.

This model proves robust enough for giving shadow-free representations which can be subsequently used in applications such as shadow removal [2, 3] or shadow-resistant tracking [4]. Indeed, contrary to non-shadow regions, shadows are not directly lightened by the source but rather by a diffuse light with a different resulting colour temperature. Hence, shadows are no more present in the $\mathcal{I}'$ or $\widetilde{\boldsymbol{\chi}'}$ representations.

A difficulty is to estimate the direction $\boldsymbol{e}$. In Sec. 2 and 3 we show that it is possible to estimate $\boldsymbol{e}$ from two views of a scene. Proof-of-concept experiments about shadow lightening are discussed in Sec. 4. Some related works are in Sec. 5.

2. DERIVING AN ILLUMINATION-INVARIANT IMAGE FROM TWO VIEWS OF A SCENE

2.1. Displacement in the log-chromaticity space

Let us note $\boldsymbol{\chi}'_A$ (resp. $\boldsymbol{\chi}'_B$) the log-chromaticities of two images (A and B respectively), as defined in Eq. (2-3). Let (x,y) be an interest point in A, and (x',y') be the corresponding interest point in B, in the sense that both points are the projection of the same physical 3D point. Then the log-chromaticity displacement is defined as $\boldsymbol{\chi}'_B(x',y') - \boldsymbol{\chi}'_A(x,y) = \boldsymbol{s}_B(x,y) - \boldsymbol{s}_A(x',y') + \frac{1}{T_B}\boldsymbol{e}_B - \frac{1}{T_A}\boldsymbol{e}_A$. Since the surface reflectance at two corresponding points is the same, $S_\lambda(x,y) = S_\lambda(x',y')$, hence $\boldsymbol{s}_B(x,y) = \boldsymbol{s}_A(x',y')$. Since the same camera is used for A and B: $\boldsymbol{e}_B = \boldsymbol{e}_A$. Con-

sequently, the log-chromaticity displacement simplifies into:

$$\boldsymbol{\Delta} := \boldsymbol{\chi}'_B(x', y') - \boldsymbol{\chi}'_A(x, y) = \left(\frac{1}{T_B} - \frac{1}{T_A}\right)\boldsymbol{e}. \quad (4)$$

which is collinear with $\boldsymbol{e}$ if $T_A \neq T_B$.

2.2. Estimating $\boldsymbol{e}$

The 2D-invariant $\widetilde{\boldsymbol{\chi}'}$ is the projection of the log-chromaticity $\boldsymbol{\chi}'$ onto a line spanned by $\boldsymbol{e}^{\perp}$. Hence, an estimation of any vector collinear with $\boldsymbol{e}$ is sufficient. The proposed process is to:
1. Compute the log-chromaticities $\boldsymbol{\chi}'_A$ and $\boldsymbol{\chi}'_B$;
2. Match interest points between both images;
3. Estimate $\boldsymbol{\Delta} = \boldsymbol{\chi}'_B(x', y') - \boldsymbol{\chi}'_A(x, y)$ for each correspondence $(x, y) - (x', y')$ (a bilinear interpolation of the RGB channels is used to compute the log-chromaticities);
4. Estimate the orientation of $\boldsymbol{e}$ in a robust way (to get rid of spurious measurements): keep only correspondences such that the norm $|\boldsymbol{\Delta}|$ is above half of the median value of all $|\boldsymbol{\Delta}|$'s, then estimate the orientation of $\boldsymbol{e}$ as the mean value of the 30% of the samples around the maximum in the histogram of the orientations of the $\boldsymbol{\Delta}$'s.

In step 2, correspondences are obtained either by SIFT [5] or by ASIFT [6] in case of large viewpoint changes (both yield subpixel accuracy), and RANSAC enforces either a homography constraint or an epipolar constraint [7].

2.3. Invariant chromaticity image

Once the orientation of $\boldsymbol{e}$ is known, the 2D illumination-invariant representation $\widetilde{\boldsymbol{\chi}'}$ can be computed. This easily yields a 3D illumination-invariant image which is better suited for visual purposes, the so-called invariant chromaticity image. It is obtained by the method detailed in [2, 3].

3. EXPERIMENTS AND DISCUSSION

The dataset is made of pairs of images taken with a consumer Canon EOS 350D DSLR camera. Images are initially in Raw format, and rendered within Canon DPP software using linear processing (hence the final RGB channels are proportional to the ρ_k in eq. (1)) and resized to 1024×683 pixels. An exponential γ-correction simply amounts to multiplying $\boldsymbol{\chi}'$ by γ, which does not interfere with the estimation of the direction of $\boldsymbol{\Delta}$. However, we find out that the non-linear processing of DPP (usually used to get images pleasing to the eye) is not a simple γ-correction, and actually interferes with $\boldsymbol{\Delta}$.

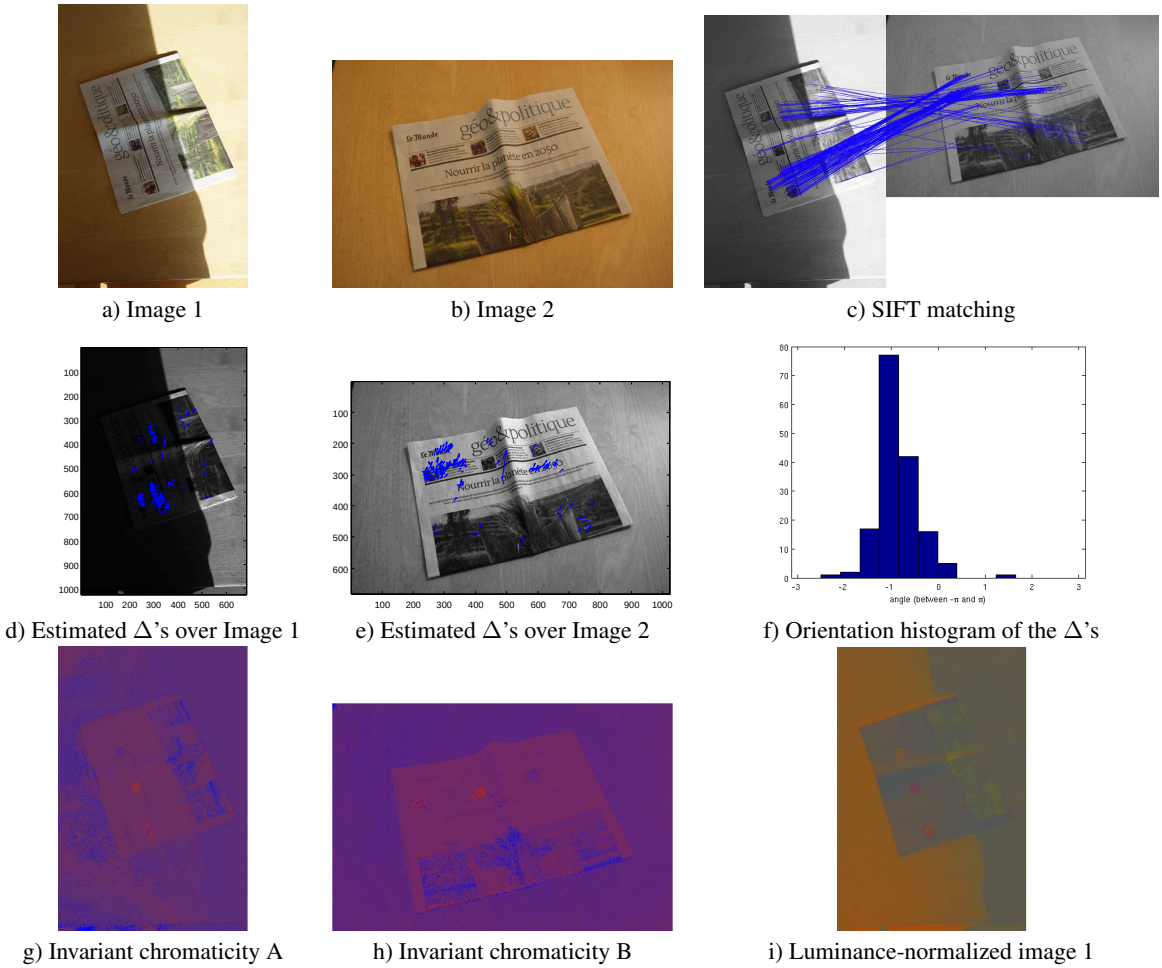
In Fig. 1, a) and b) depict the two images (non-linearly rendered for illustration purposes.) Image A is taken under direct sun illumination; a cast shadow can be seen. Image B is taken under artificial light. c) shows the 161 ASIFT correspondences. The $\boldsymbol{\Delta}$ vector is estimated for every correspondence, and superimposed on the luminance, originating from the interest points, in d) and e). The variations of this vector are actually limited, and a dominant orientation appears, as expected from Sec. 1. f) is the histogram of the orientations of the $\boldsymbol{\Delta}$'s, showing an extremum around -50^o. g) and h) are the invariant chromaticity images (Sec. 2.3.) Shadows are barely visible. The direction of the light and the illumination intensity have actually been factored out. For comparison, i) is the luminance-normalized image A (i.e. each channel is divided by $R + G + B$). As noted by Finlayson et al., it shows that such a representation, invariant only to illumination intensity changes, still contains shadow information.

Fig. 2 shows that the invariant direction $\boldsymbol{e}$ can actually be estimated in various situations. In the Ride-On Truck experiment, image A is taken under sun and image B under artificial light (32 corresp., orientation $\simeq -60^o$); in the Pine Cone experiment, image A is taken under sun and image B under a cloudy sky (28 corresp., orientation $\simeq -62^o$ modulo 180^o.) In the Pine Cone 2 experiment, both images were taken a few seconds apart, under the same light. The histogram of the $\boldsymbol{\delta}$ orientations shows two well separated modes, differing by an angle of 180^o. Hence the invariant direction can still be estimated ($\simeq -55^o$ from 515 corresp.) One can wonder why an invariant direction still exists, since here the temperature colour does not vary. It is mainly due to the limited accuracy of SIFT keypoint localization. The RGB values of corresponding points are actually not exactly the same in both images since the interest points correspond to slightly different physical 3D points. These 3D points are not lit up exactly in the same way (probably because of diffuse lighting.) However, they have the same reflectance. Thus, Eq. 4 holds and the $\boldsymbol{\Delta}$ vectors are still oriented towards a common direction. The two leftmost images of Fig. 3 shows that the $\boldsymbol{\Delta}$ vectors have actually approximately the same direction. Comparing the distribution of $\boldsymbol{\chi}'_A$ and of $\boldsymbol{\chi}'_B$ shows a displacement along a dominant direction (but in both ways), even if it is less prominent than in the Newspaper experiment, where a frank colour temperature change occurs. Note that the marginal distribution of the log-chromaticities over the orthogonal direction $\boldsymbol{e}^{\perp}$ is likely sharper than in any other direction. It is exactly the idea behind [3] where the "flatness" of a probability distribution is measured by Shannon's entropy, and the direction is obtained by entropy minimization. In [3] the chromaticities are computed for every pixel (not just interest points.) Our procedure, relying on two images, is much simpler. A thorough comparison with [3] has yet to be done.

4. APPLICATION TO SHADOW LIGHTENING

Shadows are not present in the invariant chromaticity representation. Hence, their edges can be extracted by keeping the edges in the original image that are not also edges in the illumination-invariant representation. Mimicking [2, 3], images are first segmented with Mean-Shift [8] to make edge retrieval easier (we use the Matlab wrapper[1] for EDISON[2].)

[1] http://www.wisdom.weizmann.ac.il/~bagon/
[2] http://coewww.rutgers.edu/riul/research/code/EDISON/



a) Image 1 b) Image 2 c) SIFT matching

d) Estimated Δ's over Image 1 e) Estimated Δ's over Image 2 f) Orientation histogram of the Δ's

g) Invariant chromaticity A h) Invariant chromaticity B i) Luminance-normalized image 1

Fig. 1. Newspaper experiment. The reader is kindly asked to zoom in the pdf file.

This yields a shadow-edge map $M(x,y)$ such that $M(x,y) = 0$ if there is a shadow edge at (x,y) and 1 otherwise. Then the following Poisson equation is solved for each $\log(\widetilde{\rho_k})$:

$$\Delta \log(\widetilde{\rho_k})(x,y) = \nabla \cdot \big(M(x,y)\, \nabla \log(\rho_k)(x,y)\big) \qquad (5)$$

with Neumann conditions (derivatives at the boundary set to zero). We solve this equation by the simple method in the Fourier domain from [9]. This method gives each $\log(\widetilde{\rho_k})$, up to an additive constant. In the same way as [2], we normalize each channel such that the 1% brightest pixels (which are thus not in shadow) of $\log(\widetilde{\rho_k})$ have the same intensity as the 1% brightest pixels of $\log(\rho_k)$.

This method suffers from inherent drawbacks: some objects that do not correspond to shadows cannot be seen in the invariant chromaticity images though (such as the greyish body of the ride-on truck over the grey tiles), edge detection is a difficult problem, and shadows have not necessarily sharp edges, even with segmentation. Here, the images were not deliberately chosen to ease shadow removal. In spite of it, we succeed in removing or at least lightening shadows (Fig. 4.)

5. RELATION TO PRIOR WORK

Shadow removal is an ill-posed problem addressed by a large literature. Despite its limitations, Finlayson's et al. approach does not necessitate learning as in [10] or [11]. The present contribution is to derive the illumination-invariant representation defined in [1] from SIFT correspondences of two natural images. The seemingly restrictive assumptions of the model proves to be realistic enough, although more robust invariant representations exist (see e.g. [12].) The proposed method is designed to work if the colour temperature of the light changes between the two views, but we have noted that it still works without any colour temperature change. The proposed approach does not need a colour checker chart (unlike [2]), a complex optimization stage (unlike [3] or [12]), or a static camera (unlike [13] for flash/no-flash pairs.) Let us also mention [14] where multi-view patch matching is used for the colour constancy problem.

Acknowledgements. Thanks are due to Margot and Colin for the loan of the ride-on truck and of the pine cone.

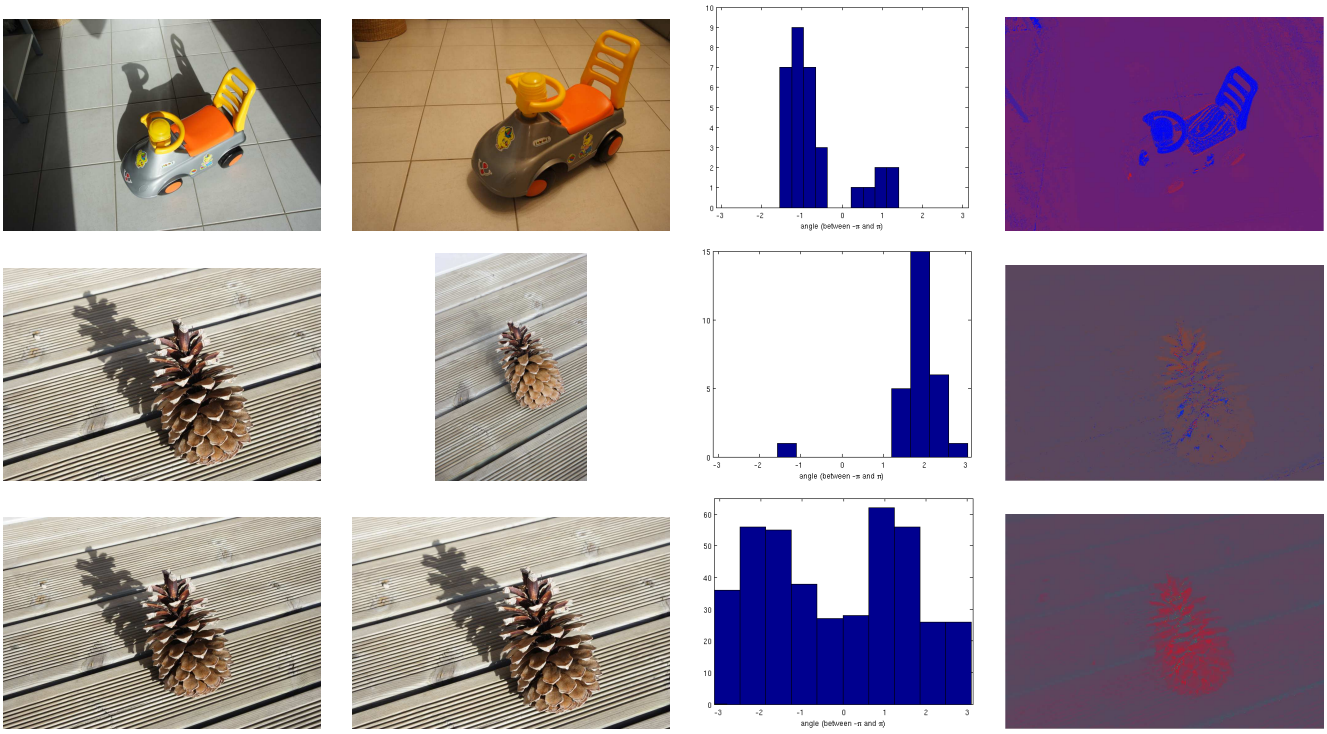


Fig. 2. From top to bottom: Ride-on Truck, Pine Cone, and Pine Cone 2 experiments: from left to right: image A, image B, histogram of Δ orientation among SIFT correspondences, and invariant chromaticity image A (information about illumination direction, intensity, and colour has actually been removed, and hence also shadows.)

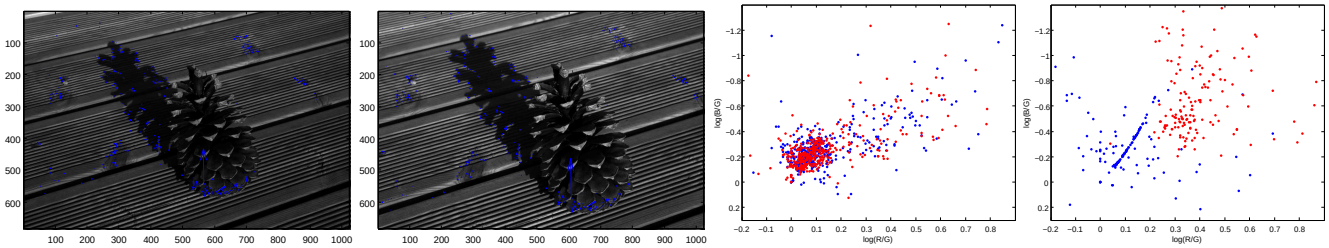


Fig. 3. Pine Cone 2 experiment. From left to right: displacement $\boldsymbol{\Delta}$ over images 1 and 2, distribution of $\boldsymbol{\chi}'_A$ (in blue) and $\boldsymbol{\chi}'_B$ (in red), and comparison with the Newspaper experiment (far right.)



Fig. 4. Application to shadow lightening for image A of Newspaper, Ride-on Truck, and Pine Cone experiments.

6. REFERENCES

[1] G.D. Finlayson and S.D. Hordley, "Color constancy at a pixel," *Journal of the Optical Society of America A: Optics, Image Science & Vision*, vol. 18, no. 2, pp. 253–264, 2001.

[2] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 59–68, 2006.

[3] G. D. Finlayson, M. S. Drew, and C. Lu, "Entropy minimization for shadow removal," *International Journal of Computer Vision*, vol. 85, no. 1, pp. 35–57, 2009.

[4] H. Jiang and M. S. Drew, "Shadow resistant tracking using inertia constraints," *Pattern Recognition*, vol. 40, no. 7, pp. 1929–1945, 2007.

[5] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[6] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[8] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[9] J.-M. Morel, A. B. Petro, and C. Sbert, "A PDE formalization of retinex theory," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2825–2837, 2010.

[10] R. Guo, Q. Dai, and D. Hoeim, "Single-image shadow detection and removal using paired regions," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, 2011, pp. 2033–2040.

[11] J.-F. Lalonde, A.A. Efros, and S.G. Narasimhan, "Detecting ground shadows in outdoor consumer photographs," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Heraklion, Crete, Greece, 2010, vol. 2, pp. 322–335.

[12] S. Ratnasingam and T.M. McGinnity, "Chromaticity space for illuminant invariant recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3612–3623, 2012.

[13] C. Lu and M.S. Drew, "Practical scene illuminant estimation via flash/no-flash pairs," in *Proceedings of the IS&T/SID Color Imaging Conference*, Scottsdale, AZ, USA, 2006, pp. 84–89.

[14] T. Owens, K. Saenko, A. Chakrabarti, Ying Xiong, T. Zickler, and T. Darrell, "Learning object color models from multi-view constraints," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, 2011, pp. 169–176.