

Illumination-Invariant Image from 4-Channel Images: The Effect of Near-Infrared Data in Shadow Removal

Sorour Mohajerani¹, Mark S. Drew², Parvaneh Saeedi¹, {¹School of Engineering Science, ²School of Computing Science}, Simon Fraser University, Vancouver, BC, Canada {smohajer/mark/psaeedi}@sfu.ca,

Abstract

Removing the effect of illumination variation in images has been proved to be beneficial in many computer vision applications such as object recognition and semantic segmentation. Although generating illumination-invariant images has been studied in the literature before, it has not been investigated on real 4-channel (4D) data. In this study, we examine the quality of illumination-invariant images generated from red, green, blue, and near-infrared (RGBN) data. Our experiments show that the near-infrared channel substantively contributes toward removing illumination. As shown in our numerical and visual results, the illumination-invariant image obtained by RGBN data is superior compared to that obtained by RGB alone.

Introduction

Images captured by cameras/sensors are the product of an interaction between illumination sources, objects' surfaces, and the image acquisition device. In regular images, light from sun/sky/an artificial illumination source impinges upon an object and the object's surface reflects the light, with a camera recording the reflected energy. The captured image may include shadows, which indeed represent a change in the illumination of the scene (at least one object under two different illumination conditions).

In one research stream, removing illumination effects from regular RGB color images has been done via log-chromaticity with camera calibration [1] and entropy minimization [2]. In addition, recently, deep learning-based methods have been used to remove shadows [3].

However, removing illumination has not been investigated in RGBN images. These images have an extra channel of near-infrared (N) compared to regular images, which are in the visible spectrum (RGB) only. The N channel usually covers the spectrum in the range of 800 – 1000nm. RGBN images have been widely used in many fields such as remote sensing [4], medical imaging[5], and computer vision [6].

In this work, we investigate the effectiveness of the N channel in the process of shadow removal. Our experiments show that having the N channel allows generating a better illumination-invariant image than that obtained by visible channels alone. As a result of this benefit, the performance of shadow removal process improves.

Previous Works

Several studies have been conducted to address the problem of shadow detection and removal in RGB images [7, 8, 9, 10]. However, the number of research works which address those problems by considering near-infrared data is limited. For instance, Rüfenacht et al. [11] have used the information in near-infrared channel to detect shadows. The authors have noticed that many objects which are dark in the visible image are brighter in the N image. They have used this observation to distinguish dark objects from shadows, which are dark in both visible and

near-infrared channels. The authors in [12] have utilized a near-infrared channel to extract a shadow probability map to find the location of penumbra and umbra shadows. They have, then, re-lit those areas in RGB images to match the lightness (L channel in CIELab color space) of non-shadow areas.

Methodology

The energy reflected from an object is captured by a camera to form an image. If the object's surface is modeled as a Lambertian surface, for each pixel of this image, the following equation obtains:

$$R_k = \sigma \int E(\lambda, T) S(\lambda) Q_k(\lambda) d\lambda, \quad k \in \{1, 2, 3, 4\} \quad (1)$$

where R_k and σ denote the intensity of a pixel for channel k of the image and the shading coefficient, respectively. $E(\lambda)$, $S(\lambda)$, and $Q_k(\lambda)$ represent the spectral power distribution of the illumination at wavelength λ , the surface spectral reflectance of the object's surface, and the sensitivity response of the camera, respectively. T is the temperature or color of the light.

Since the sensitivity of narrow-band cameras, Q_k , are usually approximated with a delta function with a magnitude of q_k at the central wavelength of each channel (λ_k), Eq. (1) is simplified to:

$$R_k = \sigma \int E(\lambda, T) S(\lambda) q_k \delta(\lambda - \lambda_k) d\lambda = \sigma E(\lambda_k, T) S(\lambda_k) q_k \quad (2)$$

Another approximation to have a simple model for image formation is that the illumination is restricted to the Planckian locus [13]. Based on Wien's approximation for temperatures between 2500K to 10000K:

$$E(\lambda, T) \simeq I a_1 \lambda^{-5} e^{-\frac{a_2}{\lambda T}} \quad (3)$$

where a_1 , a_2 , and I denote the Planck constant, the Boltzmann constant, and a multiplicative coefficient I to represent the intensity of the light. By substituting Eq. (3) in Eq. (2), the intensity at each pixel is as follows:

$$R_k = \sigma I a_1 \lambda_k^{-5} e^{-\frac{a_2}{\lambda_k T}} S(\lambda_k) q_k \quad (4)$$

Each image pixel, with four values corresponding to the four channels of RGBN, satisfies Eq. (4). We can eliminate the effect of shading and light intensity in the image by dividing pixel values by one of the R, G, B, N values. For instance by dividing by the G channel:

$$\frac{R_k}{R_G} = \left(\frac{\lambda_k}{\lambda_G}\right)^{-5} e^{-\frac{a_2}{T} \left(\frac{1}{\lambda_k} - \frac{1}{\lambda_G}\right)} \frac{S(\lambda_k) q_k}{S(\lambda_G) q_G} \quad (5)$$

By taking logarithms on both sides of Eq. (5), we form the band-ratio chromaticities [9]:

$$\Psi_k \equiv \log\left(\frac{R_k}{R_G}\right) = \frac{-a_2}{T} \left(\frac{1}{\lambda_k} - \frac{1}{\lambda_G}\right) + \log\left(\frac{S(\lambda_k) q_k \lambda_k^{-5}}{S(\lambda_G) q_G \lambda_G^{-5}}\right) \quad (6)$$

We can form another version of chromaticity to have even a simpler and more convenient equation than Eq. (6). Instead of

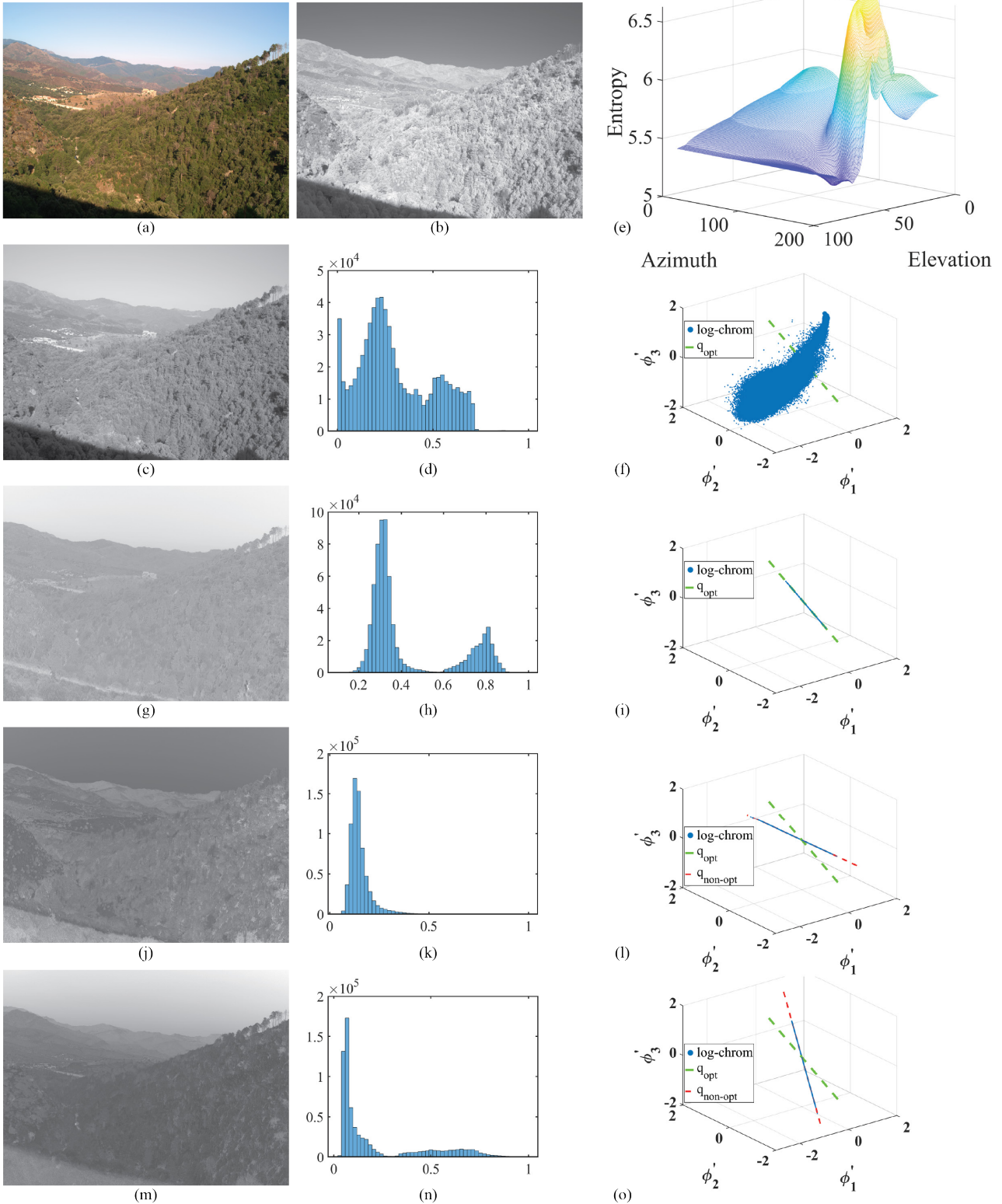


Figure 1: Illumination-invariant image formation, (a) RGB image, (b) Near-infrared channel, (c) luminance image obtained by $(R + G + B + N)/4$, (d) histogram of (c), (e) entropy variation w.r.t. *azimuth* and *elevation* angles, which shows the entropy of the projected data onto all of the possible q vectors, (f) mapped log-chromaticities in 3D space, the vector in dashed green line depicts the optimal projection vector, (g) grayscale illumination-invariant image, (h) histogram of (g), (i) projected log-chromaticities onto the optimal projection vector, (l) and (o) projection of the data onto two non-optimal examples of q (dashed red vectors), (j) and (m) the corresponding invariant images as the result of non-optimal projections, (k) and (n) histograms of (j) and (m). The entropy obtained by q vectors from (i), (l), and (o) is 5.16, 6.58, and 5.64, respectively.

dividing pixel values by one of the channels, we divide them by the geometric mean of the pixels. This approach not only avoids bias towards one of the channels but also leads to a format in which log-chromaticities can be further simplified.

The geometric mean of an image at each pixel is defined as:

$$R_m = \left(\prod_{k=1}^{N_k} R_k \right)^{\frac{1}{N_k}} \quad (7)$$

where $N_k = 4$ for RGBN images. Forming a log-chromaticity space by dividing pixel values by the geometric mean and taking logs, we will have the following equation:

$$\Phi_k \equiv \log\left(\frac{R_k}{R_m}\right) = \frac{-a_2}{T} \left(\frac{1}{\lambda_k} - \frac{1}{\lambda_m} \right) + \log\left(\frac{S(\lambda_k)q_k\lambda_k^{-5}}{S(\lambda_m)q_m\lambda_m^{-5}}\right) \quad (8)$$

where λ_m and q_m denote the geometric mean of central wavelengths and geometric mean of camera sensitivity magnitudes, respectively. This equation can be rewritten as follows:

$$\Phi_k = \frac{1}{T}(e_k - e_m) + \log\left(\frac{s_k}{s_m}\right), \quad (9)$$

$$e_k = \frac{-a_2}{\lambda_k}, \quad e_m = \frac{-a_2}{\lambda_m}, \quad s_k = S(\lambda_k)q_k\lambda_k^{-5},$$

$$s_m = S(\lambda_m)q_m\lambda_m^{-5}$$

This formulation shows these important points: (1) log-chromaticity values of an image at a certain type of surface indeed live on a line (in a 4D space) parametrized by T , with slope $e_k - e_m$; (2) all pixels belonging to one surface, under multiple illumination colors (multiple T), are located on that line; (3) the offset of the line ($\log(s_k/s_m)$) is independent of the illumination and represents only surface characteristics.

Having used geometric means for getting chromaticities, the log-chromaticities satisfy the following criteria at each pixel in 4D space: $\Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 = 0$. This, in fact, is the equation of a subspace in 4D space orthogonal to the normal vector $u = [1, 1, 1, 1]^T / \sqrt{4}$. To get the equivalent location of log-chromaticity values in the 3D space, we project them onto the subspace orthogonal to u . This leads to a dimension reduction from 4D to 3D, ($\Phi_k \in \mathbb{R}^4 \rightarrow \Phi'_i \in \mathbb{R}^3$).

Having mapped log-chromaticities in 3D, they can be used to remove the effect of illumination in the image. This is done via projecting Φ'_i onto a vector orthogonal to the direction of $v = e_k - e_m$. The result of this projection is a grayscale image in which the same surfaces under two illumination would have the same values as each other. That is why this image is called the grayscale illumination-invariant, or intrinsic, image. In other words, shadows will be removed in the invariant image (since shadow areas in an image are caused by different illumination conditions from non-shadow parts):

$$I = \Phi'_i(v^\perp)^T \quad (10)$$

where I and v^\perp denote the projected image and the vector orthogonal to illumination change, respectively.

Since the exact direction of v is unknown, following [14], we can address this problem by searching over all possible projection vectors and find out which of those leads to *minimum entropy* in the projected image:

$$q_{opt} = \operatorname{argmin}_q \left(- \sum_{i=0}^1 p_i(\Phi'_i q^T) \log(p_i(\Phi'_i q^T)) \right) \quad (11)$$

where q and q_{opt} represent projection vector and the optimal projection vector, respectively. p_i denotes the probability of having pixel values equal to i in the projected image. To attenuate noise, entropy is calculated using only 90% of the data—excluding first

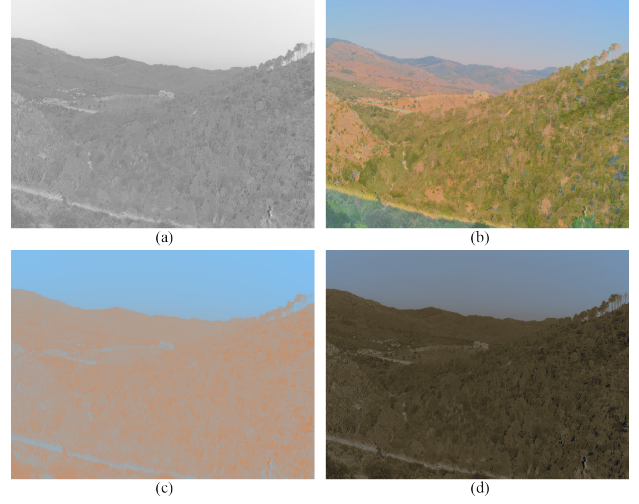


Figure 2: Invariant RGB image formation, (a) illumination-invariant image, (b) first three channels of the L1-chromaticity, (c) invariant RGB image obtained by projection (Eq. (13)), (d) invariant RGB image obtained by multiplication (Eq. (14)).

and last five percentiles of the data. In addition, the number of possible bins (discrete values of i) is limited by Scott's Rule [15].

Re-identifying q in polar coordinates rather than Cartesian limits the search space for entropy minimization. Therefore, instead of finding three unknown parameters in $q = [q_1, q_2, q_3]^T$, we search for two unknown angles of *elevation* and *azimuth* in $q = [\cos el \cos az, \cos el \sin az, \sin el]^T$ (radius is set to 1).

After finding the proper vector, the grayscale invariant image is obtained by:

$$I_{inv} = \exp(\Phi'_i q_{opt}^T) \quad (12)$$

All of the described steps are illustrated in Fig. (1). To get a shadow-free version of the image in 3D space, we need to project Φ'_i onto the found vector q_{opt} while preserving the final coordinates in 3D. This is done via a 3×3 projector ($P_{q_{opt}}$):

$$\tilde{\Phi}'_i = \Phi'_i P_{q_{opt}}, \quad \tilde{I} = \exp(\tilde{\Phi}'_i) \quad (13)$$

\tilde{I} represents a shadow-free RGB image (illustrated in Fig. 2(c)). We have found that with a simple modification we can get a visually nicer shadow-free RGB image. First, the following equation is solved to find a 3×3 transformation matrix M which maps the obtained \tilde{I} to the first three channels of L1-chromaticity of the original image ($\rho = \{R, G, B\} / (R+B+G+N)$): $\rho \approx \tilde{I}M$. Then, we calculate the approximated image and multiply each channel of it by the illumination-invariant image. This way, the values are forced to mimic/copy the patterns in invariant image, which leads to a better recovered shadow-free RGB image:

$$\tilde{I}_{aprx} = \tilde{I}M, \quad I_{rec} = I_{inv} \tilde{I}_{aprx} \quad (14)$$

where I_{rec} represents the final shadow-free RGB image (Fig. 2(d)).

Experimental Results

To quantify the contribution of the N channel to shadow removal in an image, we have compared the illumination-invariant image obtained by RGBN to that of obtained by RGB. If a shadow area is perfectly removed, regions inside that area should be similar to their adjacent non-shadow regions. In other words, two sub-regions belonging to a surface—under two different illuminations—should share similar qualities.

We have selected 6 images from the RGB-NIR public dataset [16] to conduct our experiments. The images of this dataset have been captured by Nikon D90 and Canon T1i cameras. For each scene, there are two images: (1) an RGB image captured by B+W 486 (visible) filter, (2) a grayscale near-infrared image captured by a 093 filter.

We have utilized the root mean squared error (RMSE) to measure the similarity of two adjacent sub-regions:

$$D = I_{inv}^{sh} - I_{inv}^{nonsh}, \quad RMSE = \sqrt{(1/M) \sum_{i=1}^M D_i^2} \quad (15)$$

where $RMSE$ denotes the error between two same size sub-regions in illumination-invariant image: one inside shadow (I_{inv}^{sh}) and the other one outside shadow (I_{inv}^{nonsh}). M represents the total number of pixels existing in a sub-region.

The $RMSE$ is calculated for the illumination-invariant image of RGBN and RGB. Table 1 shows the numerical results. According to this table, in all of the tested images the error between a shadow region and its neighbourhood non-shadow one is smaller in the 4D invariant image than that from a 3D image. Fig. 3 illustrates some of the visual results. In this figure, clearly, the contrast between shadow and non-shadow areas are less visible in the invariant images obtained by 4D data (bottom row) than the ones obtained by 3D data (middle row). In addition, the edge artifact (between shadow and non-shadow area) formed in an invariant image is less observable in the image obtained from 4D than the one obtained from 3D. The reason for having such an artifact is that an optimal projection vector is selected in order to minimize the entropy in the entire image. However, this very vector might not be a proper value for the regions in immediate adjacency to shadow areas. This suggests that further considerations in the optimization process are required for removing this artifact in the regions close to the boundary between shadow and non-shadow areas.

Table 1: Quantitative evaluation of shadow removal in illumination-invariant images (in %). Bold numbers indicate lower $RMSE$.

Image Name	RMSE of RGBN	RMSE of RGB
country 0010	5.078	18.055
field 0052	0.516	0.590
mountain 0026	3.755	10.497
old building 0004	0.904	1.206
street 0001	0.334	0.457
urban 0053	1.572	20.953

Summary and Conclusions

The effectiveness of using a near-infrared channel in the shadow removal process has been investigated in this paper. Illumination-invariant images obtained by the 4D RGBN inputs numerically and visually deliver better quality than that of RGB inputs. This indicates that having more spectral data (such as near-infrared) helps to remove the illumination effects in a scene.

Future Works

Since removing shadows is an important pre-processing step in many remote sensing applications, employing the proposed method on multispectral satellite images could be a future direction. Additionally, a more sophisticated optimization algorithm for finding the optimal projection vector could be utilized. This could be beneficial for removing the edge artifacts observed in Fig. 3.

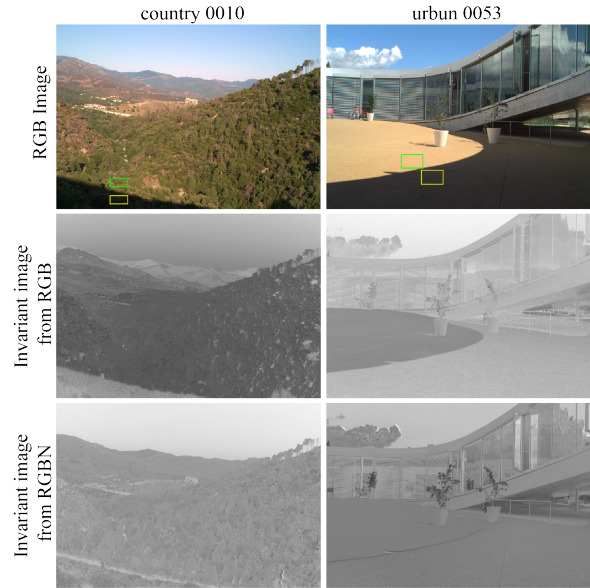


Figure 3: Some examples of illumination-invariant images. The sub-regions selected to calculate error are displayed in yellow and green boxes.

References

- [1] G.D. Finlayson and M.S. Drew, “4-Sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities,” in *ICCV’01: Int. Conf. on Computer Vision*, 2001, pp. II: 473–480.
- [2] G.D. Finlayson, M.S. Drew, and C. Lu, “Entropy minimization for shadow removal,” *Int. J. Computer Vision*, vol. 85, no. 1, Oct, pp. 35–57, 2009.
- [3] J. Wang, X. Li, and J. Yang, “Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal,” in *CVPR’18: Conf. on Computer Vision and Patt. Recog.*, 2018, pp. 1788–1797.
- [4] S. Mohajerani and P. Saeedi, “Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery,” in *IGARSS’19: Int. Geoscience and Remote Sensing Symp.*, 2019, pp. 1029–1032.
- [5] S. Shrestha, J. Petermann, T. Farrahi, A. Deshpande, and G. C. Giakos, “Design, calibration, and testing of an automated near-infrared liquid-crystal polarimetric imaging system for discrimination of lung cancer cells,” *IEEE Trans. on Instrum. and Measur.*, vol. 64, no. 9, pp. 2453–2467, 2015.
- [6] C. Son and X. Zhang, “Near-infrared fusion via color regularization for haze and color distortion removals,” *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 28, no. 11, pp. 3111–3126, 2018.
- [7] S. Mohajerani and P. Saeedi, “Shadow detection in single RGB images using a context preserver convolutional neural network trained by multiple adversarial examples,” *IEEE Trans. on Image Proc.*, vol. 28, no. 8, pp. 4117–4129, 2019.
- [8] S. Mohajerani and P. Saeedi, “CPNet: A context preserver convolutional neural network for detecting shadows in single RGB images,” in *MMSp’18: Int. Workshop on Multim. Signal Proc.*, 2018, pp. 1–5.
- [9] 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.*, vol. 28, pp. 59–68, 2006.
- [10] G.D. Finlayson, S.D. Hordley, and M.S. Drew, “Removing shadows from images,” in *ECCV’02: European Conf. on Computer Vision*, 2002, pp. 4:823–836, Lecture Notes in Computer Science Vol. 2353.
- [11] D. ufenacht, C. Fredembach, and S. S usstrunk, “Automatic and

- accurate shadow detection using near-infrared information,” *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 36, no. 8, pp. 1672–1678, 2014.
- [12] N. Salamati, A. Germain, and S. Siisstrunk, “Removing shadows from images using color and near-infrared,” in *ICIP’11: Int. Conf. on Image Proc.*, 2011, pp. 1713–1716.
- [13] G. Wyszecki and W.S. Stiles, *Color science: concepts and methods, quantitative data and formulae*, John Wiley & Sons, 2nd edition, 1982.
- [14] G.D. Finlayson, M.S. Drew, and C. Lu, “Intrinsic images by entropy minimization,” in *ECCV’04: European Conf. on Computer Vision*, 2004, pp. 582–595, Lecture Notes in Computer Science Vol. 3023.
- [15] D.W. Scott, *Multivariate Density Estimation: Theory, Practice and Visualization*, Wiley and Kegan Paul, 1992.
- [16] M. Brown and S. Süssstrunk, “Multispectral SIFT for scene category recognition,” in *CVPR’11: Conf. on Computer Vision and Patt. Recog.*, 2011, pp. 177–184.

Author Biography

Sorour Mohajerani received the B.Sc. and M.Sc. degrees in electrical engineering from Tehran Polytechnic and Iran University of Science and Technology. In 2017, she joined the School of Engineering Science, Simon Fraser University, Canada, as a Ph.D. student. Her current research interests include computer vision, remote sensing, and deep learning.

Mark S. Drew is a Professor in the School of Computing Science at Simon Fraser University in Vancouver. His background is in Engineering Science, Mathematics, and Physics. His interests lie in the fields of image processing, color, computer vision, computer graphics, and multimedia. He has published over 160 refereed papers.

Parvaneh Saeedi (Ph.D., P.Eng.) is an Associate Professor at the School of Engineering Science at Simon Fraser University, Burnaby, BC, Canada. Her research interests include machine learning, image processing, pattern recognition, computer vision, and artificial intelligence.