A Comparison Framework for the Evaluation of Illumination Compensation Algorithms

Vassilios Vonikakis
Advanced Digital Sciences Center (ADSC)
University of Illinois at Urbana-Champaign
1 Fusionopolis Way, #08-10 Connexis North Tower, Singapore
Email: bbonik@adsc.com.sg

Rigas Kouskouridas and Antonios Gasteratos
Department of Production & Management Engineering
Democritus University of Thrace
Vassilisis Sofias 12, GR-671 00 Xanthi, Greece
Email: {rkouskou, agaster}@pme.duth.gr

***Abstract*—This paper presents a new comparison framework, with the view to help researchers in selecting the most appropriate illumination compensation algorithm to serve as a preprocessing step in computer vision applications. The main objective of this framework is to reveal the positive and negative characteristics of the algorithms, rather than providing a single metric to rank their overall performance. The comparison tests, that comprise the proposed framework, aim to quantitatively evaluate the efficiency of algorithms in diminishing the effects of illumination in images. The proposed framework utilizes synthetic images, with artificial illumination degradations, which are enhanced by the tested algorithms. It represents a useful tool for the selection of illumination compensation algorithms as preprocessing in other applications, due to a) its quantitative nature, b) its easy implementation and c) its useful estimations regarding many algorithm characteristics.**

I. INTRODUCTION

Illumination is among the most important factors affecting the performance of the majority of computer vision algorithms. Underexposed or overexposed image regions caused by non-uniform illumination may pose a significant challenge to computer vision algorithms, since any image characteristic, such as edges, colors or local features, become a lot harder to detect. To overcome the above illumination problems, a typical approach is the use of a preprocessing illumination compensation technique, which can minimize the effects of under/overexposed image regions on the captured images.

Several algorithms which can compensate for the effects of illumination have been presented in the literature, coming from many different disciplines. Although they may have totally different objectives, such as the enhancement of images, the estimation of the appearance of scenes (biological vision) or the decomposition of images into illumination and reflectance (intrinsic images), their illumination compensation characteristics, have established them as part of many imaging or computer vision applications, such as, image retrieval [1], stereo [2], shadow removal [3], real-time video enhancement [4], the correspondence problem [5] and eye detection [6]. This diversity, has given rise to an interesting problem; how to choose the most appropriate algorithm, among many different types (e.g. image enhancement, illumination decomposition, appearance estimation etc.) to be used as a preprocessing step in a particular imaging or computer vision application.

Image enhancement algorithms, are usually evaluated by three major approaches: visual inspection, psychophysical experiments and image quality metrics (IQM). Visual inspection is, by its nature, subjective and difficult to draw accurate conclusions about the individual characteristics of the algorithms. Psychophysical experiments are studies in which human observers are asked to quantitatively evaluate specific characteristics of the enhanced images, such as contrast, brightness, naturalness, colorfulness etc., or to rank the results of a number of enhancement algorithms [7]. Although it delivers quantitative results, this approach still remains subjective to a certain degree and difficult to reproduce. IQM are measures used for the evaluation of perceived characteristics of imaging systems or of image processing techniques [8]. Generally, IQMs usually calculate particular image characteristics (e.g. brightness, contrast, colorfulness etc.) or predict the distortion introduced into the image by processing algorithms. Intrinsic image algorithms attempt to decompose the image into reflectance, illumination and specularities [9]. Benchmark datasets, providing ground truth for many different objects [10] have been used for their evaluation. Finally, scene appearance algorithms attempt to compute the appearance of a scene, as it would have been perceived by a human observer [11]. Their evaluation is more complex and usually requires psychophysical experiments [12].

All the above evaluation approaches are designed for a specific type of algorithm. Yet, they give little insight regarding the algorithms' potential as preprocessing, since the attributes which are important for this task may be different from the original objective of the algorithm. For example, an image enhancement technique may not exhibit good quality of results, in terms of naturalness or contrast, but at the same time, it may still be good for diminishing the effects of shadows and thus be appropriate for preprocessing in another system.

This paper attempts to resolve this issue. It proposes a new comparison framework which estimates certain characteristics of illumination compensation algorithms, regardless of the field from which they are coming, important for their success as preprocessing in other vision systems. It should be clarified that the aesthetic evaluation of enhancement results or psychophysics is out of the scope of this paper. For this reason, it does not make use of observers or psychophysically derived IQMs. The primary objective is the estimation of the preprocessing potential of illumination compensation algorithms to other applications, in an easily reproducible way. More specifically, the proposed framework does not attempt to rank algorithms according to a single statistical measure, but it

978-1-4673-5791-3/13/$31.00 ©2013 IEEE

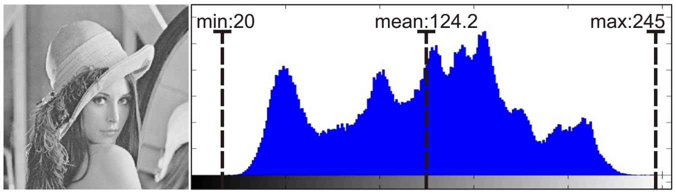

Fig. 1. The *Lena* image and its exposure characteristics.

rather aims to reveal their positive and negative characteristics, in various performance categories. As such, it attempts to answer specific questions like "*What is the maximum shadow strength that algorithm X can successfully handle?*", "*Which algorithm between X and Y is better in correcting overexposed image regions?*", or "*How good is algorithm X in preserving the visual information in the correctly-exposed regions?*". This type of information is essential for facilitating the selection of an illumination compensation algorithm as preprocessing in a particular imaging or computer vision application. The proposed framework utilizes synthetic images with various degrees of artificial illumination degradations; the synthetic images are enhanced by the evaluated algorithms and their results provide strong indications about their overall performance.

The paper is organized as follows: Section 2 presents a detailed description of the proposed comparison framework. Section 3 gives an extensive example for the evaluation of four illumination compensation algorithms, using the proposed framework. Finally, concluding remarks and discussion are presented in section 4.

II. Comparison Framework

The proposed framework excludes psychophysics from the evaluation of the enhanced images, and employs computer generated test images, which can be easily used as benchmarks by other researchers. A set of 40 specialized test images is generated, comprising the most common illumination combinations that can be encountered in non-controlled environments, i.e. uniform/non-uniform illumination and underexposed/overexposed image regions. Finally, a statistical analysis is performed, concerning the enhancement performance of tested algorithms on the set of these 40 test images.

A. Constructing the test images

The 40 test images are generated by applying artificial degradations on an image, used as Ground Truth (GT). The GT image can be any correctly exposed image of a scene, under mostly uniform illumination, without underexposed or overexposed regions. In our implementation of the framework, the *Lena* image is selected as GT, since it is one of the most widespread images in the field and includes a variety of surface types, such as textured, curved or flat. More importantly though, as Fig. 1 indicates, it has a very well balanced histogram, with a mean value of 124.2, which is very close to the middle of the scale (127.5). Additionally, there is no clipping of visual information in the lower part (min=20) or the upper part of the scale (max=245). All these are indications of overall good exposure. *Any* image with similar characteristics could, thus, be used as GT. For existing vision systems, well-exposed images captured with the system's camera could be used as GT, incorporating the camera's characteristics in the GT image. Fig. 2(a) depicts the flow chart of the proposed comparison framework.

The artificial degradations applied to the GT image, are generated by approximating two kinds of illumination (uniform and non-uniform), along with two kinds of degradations (underexposed and overexposed regions). In the case of uniform illumination, the degradations are applied to the whole GT image, whereas, for non-uniform illumination, the degradations are applied only to a specific part of it, leaving the other part intact. Underexposure degradations are simulated using the following equation, which is based on the multiplicative relation between reflectance and illumination.

$$UE_{ij} = (1 - IL_{ij}) \times GT_{ij} \tag{1}$$

where UE is the underexposed-degraded test image, IL is a function that determines type of illumination at each spatial position, and (i, j) are the pixel coordinates.

Similarly to the underexposed regions, overexposure degradations are simulated using the following equation.

$$OE_{ij} = B - [(1 - IL_{ij}) \times (B - GT_{ij})] \tag{2}$$

where OE is the overexposed-degraded test image, B is the maximum possible value of the GT image (usually 255), IL is a function that determines the type of illumination and (i, j) are the pixel coordinates.

For both underexposed and overexposed-degraded test images, we have identified three types of illumination, as indicated by the following equations.

$$IL_{ij}^{const} = IL_{max} \tag{3}$$

$$IL_{ij}^{step} = u(j - \frac{imx}{2}) \cdot IL_{max} \tag{4}$$

$$IL_{ij}^{grad} = \frac{j}{imx} \cdot IL_{max} \tag{5}$$

where IL^{const} corresponds to a uniform (constant) illumination, IL^{step} corresponds to a sharp illumination transition (step) and IL^{grad} is an illumination gradient. IL_{max} is the strength of illumination, with $IL_{max} \in [0, 1]$, $u(\cdot)$ is the unitary step function and imx is the width of the GT image. When $IL_{max} = 0$, underexposed or overexposed image regions disappear and the degraded test image is equal to the GT image. When $IL_{max} = 1$, the degradation strength is maximum, resulting to the complete loss of any visual information. For all the intermediate values of IL_{max}, the strength of the degradation varies linearly between these two extremes.

In our implementation we focused on the uniform illumination of equation (3) and the non-uniform (step) illumination of equation (4). We preferred the step illumination to the gradient of equation (5) since the former poses a grater challenge for illumination compensation algorithms, by triggering the appearance of possible halo artifacts in the region of the sharp illumination transition. Thus, using the illumination of equation (4) may expose these kind of limitations, which are important to know. In order to test the results of algorithms in various degrees of illumination, test images with 10 different under/overexposure strengths are generated, for both uniform and

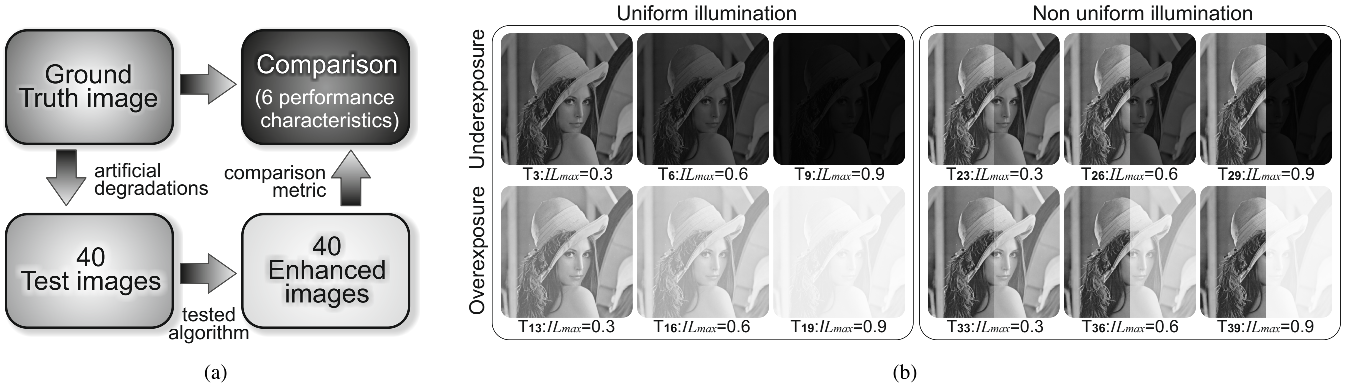

Fig. 2. (a) The block diagram of the proposed framework; (b) A subset of the 40 proposed test images of the framework.

non-uniform illumination. This essentially means that parameter $IL_{max} \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95\}$. Eventually, a set of 40 different test images ($T_1 - T_{40}$) is generated, a subset of which is depicted in Fig. 2.

- $T_1 - T_{10}$: Uniform underexposure (IL^{const}).
- $T_{11} - T_{20}$: Uniform overexposure (IL^{const}).
- $T_{21} - T_{30}$: Non-uniform underexposure (IL^{step}).
- $T_{31} - T_{40}$: Non-uniform overexposure (IL^{step}).

B. Performance metric

The 40 test images are tested one by one against the GT image, in the way the flow chart of Fig. 1a depicts. A theoretically perfect illumination compensation algorithm, would output an image identical to the GT image. The departure of the algorithm output from the GT image is measured using normalized Root Mean Square Error, which is the performance metric upon which, all the evaluations of the proposed framework are based. The formula of this metric is given in the following equation:

$$m_k^s = 1 - \frac{1}{B}\sqrt{\frac{\sum\limits_{i\in S}\sum\limits_{j\in S}\left(GT_{ij} - E\left(T_k\right)_{ij}\right)^2}{N_S}} \qquad (6)$$

$\forall(i,j) \in S = \{S_D, S_{ND}, S_W\}$. m_k^S is the performance metric applied to the S region of test image T_k, taking values in the interval [0, 1]. N_S is the number of pixels belonging to region S, B is the maximum possible value of the GT image (usually 255), $E(T_k)$ is the test image T_k, enhanced by the evaluated algorithm and (i, j) the pixel coordinates. Set S comprises three different image regions, on which the performance metric can be applied: S_D, which is the degraded region of the image, S_{ND}, which is the non-degraded region of the image and S_W, which is the whole size of the image. According to equation (6), the 'theoretically perfect' enhancement algorithm would exhibit a value of $m_k^S = 1$.

The final result of the proposed framework is a set of 6 different graphs that describe 6 corresponding performance characteristics, as they change across different degrees of illumination strength.

1) Improvement of uniformly underexposed regions (T_1-T_{10}, S_W).
2) Improvement of uniformly overexposed regions (T_{11}-T_{20}, S_W).
3) Improvement of non-uniformly underexposed regions (T_{21}-T_{30}, S_D).
4) Improvement of non-uniformly overexposed regions (T_{31}-T_{40}, S_D).
5) Preservation of intact regions for underexposure (T_{21}-T_{30}, S_{ND}).
6) Preservation of intact regions for overexposure (T_{31}-T_{40}, S_{ND}).

Cases 5 and 6 are very important since a particular algorithm may perform very well in enhancing, for example, the underexposed regions of an image, yet, it may also affect negatively the correctly exposed ones. The latter is an unwanted side-effect, as it lowers the quality of output images and passes undetected when the metric is applied globally, to the whole image size (S_W). Algorithms which can enhance under/overexposed areas, without affecting the correct ones, will exhibit better performance as preprocessing for other applications. To our knowledge, no other comparison approach has taken into consideration this attribute.

III. Applying the Framework

In order to demonstrate the proposed comparison framework, four different illumination compensation algorithms were evaluated. Vassilios Vonikakis (VV) algorithm [13] is a center-surround image enhancement algorithm, based on the Human Visual System. The Multi-Scale Retinex with Color Restoration (MSRCR) [14] and Saponara Retinex (SR) [15] are two different versions of the Retinex algorithm. Finally, the Fused Logarithmic Transform (FLOG) is an improvement of the classic logarithmic mapping, employing also a multi-scale pyramid [16]. Fig. 3 depicts the results of the proposed framework. These six graphs can give important insights regarding an algorithms performance profile, highlighting its strengths and weaknesses and thus, helping researchers to assess its preprocessing potential for other applications.

The most prominent characteristic of the VV algorithm is that it targets its enhancement specifically to the degraded image regions, affecting minimally the correctly exposed ones. This is evident in Fig. 3e and Fig. 3f, where VV clearly

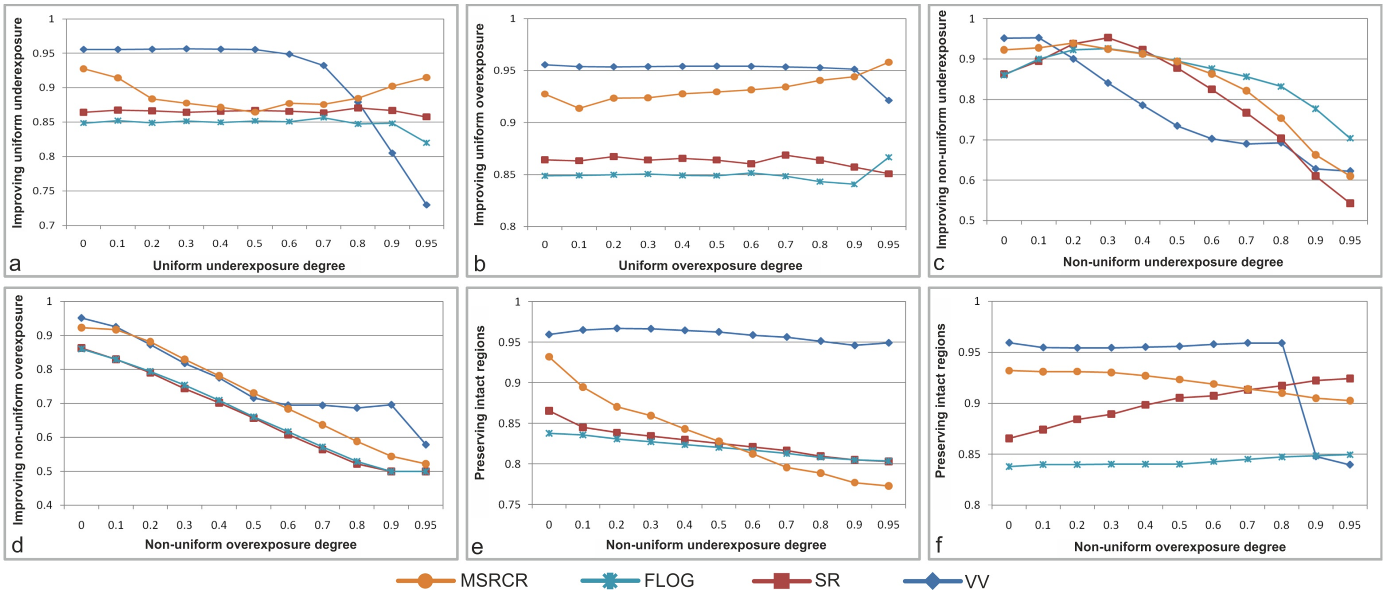

Fig. 3. Results of the proposed framework.

outperforms all the other algorithms, exhibiting a m_k^S score of more than 0.95, approximately. This characteristic might be important for both aesthetic reasons, as well as in many computer vision applications, because it ensures the preservation of visual information in the non-degraded regions. However, such a good characteristic comes at a cost; the enhancement of non-uniform underexposed regions is worst in most of the cases, compared to the other algorithms (Fig. 3c). Apart from that, VV is good in enhancing uniformly underexposed or overexposed images with moderate degradation strength. This is evident in Fig. 3a and Fig. 3b, where VV outperforms all the other algorithms for underexposure strengths up to 0.8. After that, its performance degrades rapidly, ending last for very strong uniform underexposures. As a conclusion, the VV algorithm exhibits moderate illumination compensation characteristics, with its strongest point being the overall appearance of the image, since it does not affect the correctly exposed regions. When it comes to computer vision applications, VV should be given preference only for moderate illumination degradations.

MSRCR is not ranked first in any evaluation category, although it exhibits the second best performance in most of the cases. It is generally very good in enhancing any kind of underexposure, both uniform and non-uniform (Fig. 3a, Fig. 3c). Concerning overexposure, MSRCR is a good candidate for any illumination condition, since it also exhibits competitive results in most of the cases (Fig. 3b, Fig. 3d). An important feature is that the preservation of the correctly exposed image regions is satisfactory for overexposure, but degrades rapidly for underexposure cases (Fig. 3f and Fig. 3e respectively). As a conclusion, MSRCR is an algorithm which can be used reliably for illumination compensation, both for aesthetic correction of images, as well as preprocessing for other algorithms.

Similarly to MSRCR, SR is not ranked first in any evaluation category exhibiting, in general, a rather moderate performance. In the case of uniform illumination, SR could be a good choice of enhancement algorithm, due to the high predictability of its results; it exhibits an almost constant output of approximately 0.86 for any degree of uniform under/overexposure (Fig. 3a, Fig. 3b). However, when illumination is non-uniform, its performance drops steadily (Fig. 3c, Fig. 3d). In such illumination conditions, it exhibits good enhancement of moderately underexposed regions, very low enhancement of overexposed ones, while the preservation of the correctly exposed image regions is not its strongest point. Similarly to MSRCR, SR exhibits a potential for both aesthetic and preprocessing use.

FLOG is an algorithm which exhibits a superior performance in enhancing strong underexposure under non-uniform illumination conditions (Fig. 3c). However, this attribute comes at a cost, since, at the same time it also affects the correctly exposed regions; it exhibits the worst performance in most of the cases in Fig. 3e and Fig. 3f. In this sense, FLOG exhibits an opposite performance compared to VV, which preserves almost perfectly the correctly exposed regions, at the expense of underexposure enhancement. Additionally, FLOG is not to be recommended for enhancing overexposed regions under non-uniform illumination conditions (Fig. 3d). Similarly, its performance is the worst, in most of the cases of uniform illumination (Fig. 3a, Fig. 3b). For these reasons, FLOG should be given preference mostly in preprocessing for severe shadow cases, rather than in the aesthetic enhancement of images.

IV. Conclusion

A new comparison framework for the evaluation of illumination compensation algorithms is presented in this paper. It utilizes computer generated synthetic images, with artificial illumination degradations of various degrees. The improvements introduced by algorithms, when applied to the test images, are used to create graphs for 6 performance attributes (improvement of under/overexposure, both in uniform and non-uniform cases, along with the ability to preserve information in the intact image regions). These attributes give strong indications about the algorithms positive and negative characteristics and their preprocessing potential in other algorithms or vision systems. Consequently, the contribution of this paper is twofold; first, the proposed framework represents a benchmarking tool for evaluating illumination compensation algorithms, highlighting their important characteristics and second, it provides credence to their suitability for preprocessing in computer vision applications.

ACKNOWLEDGMENT

This study is partially supported by the research grant for ADSC's Human Sixth Sense Programme from Singapore's Agency for Science, Technology and Research (A*STAR).

REFERENCES

[1] G. Ciocca, D. Marini, A. Rizzi, R. Schettini, and S. Zuffi, "Retinex preprocessing of uncalibrated images for color-based image retrieval." *J. Electronic Imaging*, vol. 12, no. 1, pp. 161–172, 2003.

[2] W. Xiong and B. Funt, "Stereo retinex," *Image and Vision Computing*, vol. 27, no. 12, pp. 178 – 188, 2009, ¡ce:title¿Canadian Robotic Vision 2005 and 2006¡/ce:title¿.

[3] G. D. Finlayson, S. D. Hordley, and M. S. Drew, "Removing shadows from images using retinex," in *The Tenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications, CIC 2002, November 12, 2002, Scottsdale, Arizona, USA*. IST - The Society for Imaging Science and Technology, 2002, pp. 73–79.

[4] C. Iakovidou, V. Vonikakis, and I. Andreadis, "Fpga implementation of a real-time biologically inspired image enhancement algorithm." *J. Real-Time Image Processing*, vol. 3, no. 4, pp. 269–287, 2008.

[5] F. Werner, F. Maire, and J. Sitte, "Topological slam using fast vision techniques," in *Proceedings of the FIRA RoboWorld Congress 2009 on Advances in Robotics*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 187–196. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03983-6_23

[6] C. Jung, T. Sun, and L. Jiao, "Eye detection under varying illumination using the retinex theory," *Neurocomput.*, vol. 113, pp. 130–137, Aug. 2013.

[7] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Image attributes and quality for evaluation of tone mapping operators," in *Proceedings of Pacific Graphics 2006 (14th Pacific Conference on Computer Graphics and Applications)*. National Taiwan University Press, Oct. 2006, pp. 35–44.

[8] I. Avcibas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *J. Electronic Imaging*, vol. 11, no. 2, pp. 206–223, 2002.

[9] J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 3481–3487.

[10] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground-truth dataset and baseline evaluations for intrinsic image algorithms," in *International Conference on Computer Vision*, 2009, pp. 2335–2342.

[11] R. Palma-Amestoy, E. Provenzi, M. Bertalmio, and V. Caselles, "A perceptually inspired variational framework for color enhancement," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 3, pp. 458–474, 2009.

[12] C. Parraman, A. Rizzi, and J. J. McCann, "Color appearance and color rendering of hdr scenes: an experiment," pp. 72 410R–72 410R–12, 2009.

[13] V. Vonikakis, I. Andreadis, and A. Gasteratos, "Fast centre-surround contrast modification," vol. 2, no. 1, pp. 19–34, February 2008.

[14] D. Jobson, Z.-u. Rahman, and G. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *Image Processing, IEEE Transactions on*, vol. 6, no. 7, pp. 965–976, 1997.

[15] S. Saponara, L. Fanucci, S. Marsi, and G. Ramponi, "Algorithmic and architectural design for real-time and power-efficient retinex image/video processing." *J. Real-Time Image Processing*, vol. 1, no. 4, pp. 267–283, 2007.

[16] H.-S. Le and H. Li, "Fused logarithmic transform for contrast enhancement," *Electronics Letters*, vol. 44, no. 1, pp. 19–20, 2008.