# Perceptually-based restoration of backlit images

***Blind submission***

## Abstract

*Scenes with back-light illumination are problematic when captured with a typical LDR camera, as they result in dark regions where details are not perceivable. In this paper, we present a method that, given an LDR backlit image, outputs an image where the information that was not visible in the dark regions is recovered without losing information in the already well-exposed parts of the image. Our method has three main steps: first, a variational model is minimized using gradient descent, and the iterates of the minimization are used to obtain a set of weight maps. Second, we consider the tone-mapping framework [3] that depends on four parameters. Two different sets of parameters are learned by dividing the image in the darker and lighter parts. Then, we interpolate the two sets of parameter values in as many sets as weighting maps, and tone-map the original image with each set of parameters. Finally, we merge the new tone-mapped images depending on the weighting maps. Results show that our method outperforms current backlit image enhancement approaches both quantitatively and qualitatively.*

## Introduction

Conventional cameras fail in capturing all the details and contrast that we perceive with the naked eye. One typical such scenario is a backlit scene where the details in the bright regions are captured properly when a short time exposure is used but at the expense of a poor rendering in the dark regions. And when a long exposure time is used, the dark regions are captured properly but the bright intensity values are clipped. No single exposure value will allow to capture the entire scene properly.

With the advancement of High Dynamic Range (HDR) imaging techniques over the past few years it is now possible to potentially capture the full range of light information of a scene [18]. And tone-mapping allows to render them on a display. Tone-mapping algorithms can also be adapted also to perform image enhancement and modify the style of single LDR images [14].

In this paper, we deal with the improvement of single LDR backlit images to recover those details that are not visible in the image. Our approach first obtains a set of weight maps that join together regions with similar luminance range. This step is performed by minimizing an energy functional related to Retinex and perceptual color processing [1, 2]. Then, a set of different tone curves are learned based on the framework given by [3], that hinges on findings in psychophysical experiments [12]. Finally, we merge the weight maps together with the images obtained by tone-mapping the original image with the different tone-curves.

This paper is divided as follows. Next section details the related work. Later we explain our approach. We then move towards the results section where we prove the adequacy of our method both quantitatively and qualitatively. The paper ends with the concluding remarks.

## Related Work

Different approaches have been defined to solve the problem of backlit images. We can divide them into local image enhancement methods, fusion-based methods, and segmentation-based methods.

Local image enhancement methods are the simpler solutions for the problem. Examples of this type of approaches are local histogram equalization methods like the CLAHE approach [28], the original Retinex model [13] and its multi-scale Retinex framework [17] or some other versions that aim at preserving naturalness [10, 23]. The former approach is prone to halo artifacts and can be alleviated with edge-aware processings, such as the ones proposed by Farbman *et al.* [5], and Paris *et al* [16]. Yuan and Sun [25] proposed an automatic exposure correction algorithm that is also useful for backlit image restoration. Finally, a recent variational model based on the log-transform [6], or an image enhancement method inspired by the Dark Channel prior [4] can also be used for this problem.

Regarding fusion-based methods, Fu *et al.* [7] first separate the image into reflectance and illumination, from which three new possible illuminations are derived. These illuminations are weighted and blended together to obtain a new illumination that is combined with the original reflectance to obtain the final result. Ying *et al.* [24] proposed a multi-exposure fusion framework that makes use of a new camera model that allows to synthesize multi-exposure images.

The first segmentation-based method for this problem was proposed by Tsai and Yeh [21]. They threshold the luminance of the image and linearly stretch the backlit detected regions. Recently, Li and Wu [15] proposed a learning-based method to classify superpixels of the image into backlit or frontlit regions and then a different tone-mapping is applied to each of these segments.

## Proposed approach

Our approach is inspired by the observation that segmentation-based methods tend to fail when an object of dark color is present in the properly illuminated region, as it is commonly selected as belonging to the dim region. To alleviate this problem we propose a variational-based region split where both the original input values and the local contrast of the object are considered to create a set of weight maps. The local contrast will have a high value for a dark object in the proper illumination region, while it will have a low value for any object of the same luminance found in the dimmer region; therefore allowing us to distinguish between the two cases. Later, once the weight maps are defined, we compute as many tone-curve as weight maps, we apply the tone-curves to the original image, and finally merge the tone-mapped images. We now explain each of the different stages in detail.

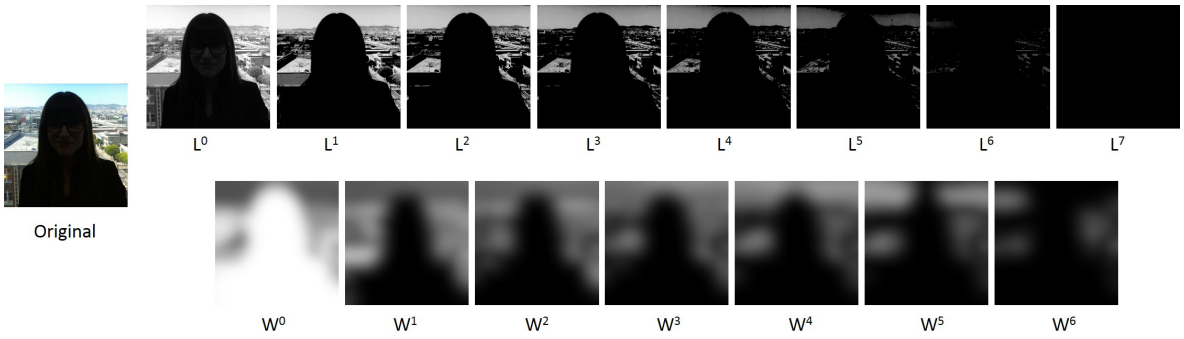


***Figure 1.*** *Original image (left). the iterates from the gradien descent (top) and the weight maps associated to these iterates (bottom).*

### *Variational region split*

Given an original RGB color image $I_{srgb}$, we perform its linearization to obtain $I$ (i.e. $I = I_{srgb}^{2.2}$). We then compute its luminance $L_0$ and minimize the following image energy functional

$$E(L) = \frac{\beta}{2}\int_{\Omega}(L(x)-L_0(x))^2dx + \frac{\alpha}{2}\int_{\Omega}(L(x)-\mu_L)^2dx + \tau\int_{\Omega}L(x)dx - \eta\int\int_{\Omega^2}\omega(x,y)|L(x)-L(y)|dxdy \quad (1)$$

where $\Omega$ is the image domain, $\alpha,\beta,\eta$, and $\tau$ are positive scalars, $L_0$ is the luminance of the original image, $\mu_L$ is the mean of the luminance of the original image, and $\omega(x,y)$ is a Gaussian kernel. This functional has its roots in [1, 2] where it was shown that when $\tau = 0$ it follows perceptual principles and has ties with the Retinex theory of color vision. Also, it has been proven useful for color gamut mapping [26, 27] and image dehazing [8, 9].

The minimization of Eq.1 will decrease the original values of the image (third term of the functional) while at the same time it will maximize the local contrast of the image (last term of the functional), and it will not move far away from the original image and the original mean (first two terms of the functional).

We compute the steady-state of Eq.1 through gradient descent

$$L^{k+1}(x) = \frac{L^k(x) + \Delta t\left(\alpha\mu_L + \beta L_0(x) + \tau + \frac{\eta}{2}R_{L^k}(x)\right)}{1+\Delta t(\alpha+\beta)}, \quad (2)$$

where the initial condition is $L^{k=0}(x) = L_0(x)$, $\Delta t$ is the time step, and $R_{L^k}(x)$ indicates the contrast function

$$R_{L^k}(x) = \frac{\sum_y w(x,y)s\left(L^k(x)-L^k(y)\right)}{\sum_y w(x,y)}. \quad (3)$$

In the case where the value of any pixel $x$ at any iteration is negative we clip it to 0. It is clear that, due to the influence of the $\tau$ parameter, this gradient descent will tend towards a black image. Also, it will do so by firstly dropping to zero those areas of the image that were dimmer. As stated at the beginning of the section, the reason behind this second fact is the behaviour of the local contrast term: it will be high for a dark object present in the properly illuminated part; in contrary, it will be low for any object of the same luminance found in the dim part of the image.

We can use these two facts to create a set of weight maps in the following way

$$W^k(x) = \frac{L^k(x) - L^{k+1}(x)}{L^k(x)+\varepsilon} \quad (4)$$

where $\varepsilon$ is a very small value that controls division by 0 for the pixels that already reached the 0 value. We convolve the weight maps with a Gaussian Kernel to avoid problems in the edges, and we normalize them such as $\sum_k W^k(x) = 1$.

Due to the construction of Eq. 4, the first weight map will contain most of the information from the dimmer parts of the image (whenever a region gets to black all its subsequent weight maps are 0), while the latter ones will contain information about the lighter parts of the image. This is also shown in Figure 1 where we show an original image, the iterates of the variational formulation (top) and the different weight maps (bottom).

### *Tone-mapping*

In this section we explain briefly the tone mapping operator proposed by Cyriac et. al [3]. The method finds an adaptive non-linear transform that performs constrained histogram equalization which is based on the psychophysical study [12] showing subjective preference to images with flat lightness histogram. Complete histogram equalization achieved by an intensity transform based on the cumulative histogram may produce sharp changes in contrast and spurious colors. Thus, Cyriac et al. [3] modeled the transformation function as a smooth approximation of the cumulative histogram based on natural image statistics. Natural images on average has a triangular shaped luminance histogram in log-log coordinates[11, 19]. Thus, the cumulative histogram ($H$) can be modeled as a piecewise linear function with two slopes ($\gamma_L$ and $\gamma_H$) and can be approximated in linear-linear coordinates as follows:

$$H(I) = I^{\gamma(I)}, \quad (5)$$

where

$$\gamma(I) = \gamma_H + (\gamma_L - \gamma_H)\left(1 - \frac{I^M}{I^M+\mu^M}\right), \quad (6)$$

and $I$ is the normalized image and $\gamma(\cdot)$ is a slope function with $\gamma \simeq \gamma_L$ for small intensities, $\gamma \simeq \gamma_H$ for large intensities and a smooth transition at $\mu$ with a slope $M$. All these parameters are

computed from the cumulative histogram of the input intensity image. Finally, the tone mapping is given by:

$$TMO(I;M,\gamma_L,\gamma_H,\mu) = I^{\gamma(I)}. \quad (7)$$

### Merging

The most straightforward idea at this step would be to generate a set of improved images by applying different tone-curves to the original image; each of them obtained from the pixels belonging to each particular weight map. However, this solution is problematic when the weight maps represent a small number of pixels. A clear example of this are the maps $W^5$ and $W^6$ from Figure 1. Therefore, to avoid this problem we directly obtain only two tone-curves (one for the brighter parts and one for the dimmer parts), and then we interpolate the two to generate the rest of the tone-curves needed. To this end, we consider the first weight map $W^1(x)$, and depending on its values, we divide the image into two different sets as follows:

$$\begin{aligned} Set^a &= \{I(x)|W^1(x) \geq threshold\} \\ Set^b &= \{I(x)|W^1(x) < threshold\}. \end{aligned} \quad (8)$$

where $I$ is the linear version of the original image (as detailed at the beginning of the approach). Then, for each of the two sets we compute the four parameters of the previously explained tone-mapping method, therefore obtaining two different sets of parameters $Par^a = \{M^a, \gamma_L^a, \gamma_H^a, \mu^a\}$, and $Par^b = \{M^b, \gamma_L^b, \gamma_H^b, \mu^b\}$. The parameters with superindex $a$ will properly tone-map the darker part of the image, while the parameters with superindex $b$ will properly tone-map the brighter parts of the image.

Let us consider now that we have $N$ weight maps. Then, to obtain N different parameters sets we interpolate the sets $Par^a$ and $Par^b$ in the following way:

$$Par^k = Par^{k-1} + \frac{(Par^N - Par^{k-1})}{2} \quad (9)$$

where $Par^0 = Par^a$ and $Par^N = Par^b$. We use here an abuse of notation to mark that the same interpolation is done for the four different elements of the set. We select this interpolation since, apart from the first weight map that mostly contains the information about the dim region, all the others maps tend to have most of their information in the properly lit area.

Then, for each set of parameters we obtain the tone mapped image

$$I_{TM}^k = TMO(I;Par^k) \quad (10)$$

where $TMO(\cdot)$ is the application of the tone-mapping explained in Eq. 7 and $I$ is the linear version of the original image as stated before. Linearization is important since the selected tone-mapping operator works on linear data.

Finally, we combine the different tone-mapped images taking into consideration the respective weight maps to obtain our final image

$$I_{final}(x) = \sum_k I_{TM}^k(x) \cdot W^k(x). \quad (11)$$

## Results

We have compared our approach both quantitatively and qualitatively versus the current state-of-the-art methods: Li and Wu [15], Ying *et al.* [24], Dong *et al.* [4], the Fu *et al.* variational model [6], the Multi-scale Retinex approach (MSRCR) [17], the Naturalness preserving method (NPE) [23], the Fu *et al.* fusion scheme [7], and the LIME method [10]. We have considered two different datasets: the Vonikakis *et al.* dataset [22], and the Li and Wu [15] dataset. The former dataset consists of 23 images with size $2304 \times 1728$, while the latter consists of 38 images with sizes ranging from $1400 \times 1049$ to $250 \times 190$. All our results were obtained with a fixed set of parameters: $\eta = 1$, $\alpha = 0.5$, $\beta = 0.5$, $\tau = 2$, and $threshold = 0.75$. The standard deviation of the gaussian kernel was set to $1/3$ of the largest dimension of the image.

### Qualitative results

Figure 2 shows from left to right the original image, the NPE method [23], the fusion scheme method [7] and our result. We want to note that although all the methods are able to recover most of the information of the girl's head, our method is the only one that prevents a strong red cast to appear in her face. Figure 3 presents from left to right the original image, the Multi-scale Retinex approach [17], Fu *et al.* variational model [6], and our result. We can see that the Multi-Scale Retinex approach creates artifacts, specially in the bottom part of the image, while the varional model of [6] is not able to fully recover the face of the woman. In contrast, our method is able to properly recover her face without introducing artifacts in the image.

Figure 4 shows from left to right the original image, the Dong *et al.* method [4], the Fu *et al.* variational model [6], and our result. In this case, our method is able to represent a scene with natural and realistic colors while recovering details of the tree in the front, contrary to the results of the other two appraches, in which the t-shirt of the girl shows unrealistic effects. Finally, Figure 5 shows from left to right the original image, the LIME method [10], Ying *et al.* [24], and our result. in this case we can see that the LIME method washes the image and over-enhances the noise. Both the approach of Ying *et al.* and our result are able to provide a good result, however our result preserves better the shadow effect on the face of the man.

### Quantitative results

In this subsection we quantitatively compare our approach to the state-of-the-art approaches considering two different metrics proposed in [24]: the Lightness Order Error and the Visual Information Fidelity .

The Lightness Order Error metric evaluates the violation of the order statistics in the processed image. Ideally, the order between the pixels of the original image should be maintained in the processed image. This is crucial in two aspects. First, we do not want an original dark region to become lighter than an original light region. Second, the order between the pixels is also violated when the processed image presents halos or ghosting effects. The Lightness Order Error metric is defined as follows [24]

$$LOE = \frac{1}{|\Omega|} \int_\Omega \int_\Omega (U(L_o(x), L_o(y)) \oplus U(L_p(x), L_p(y)))dxdy \quad (12)$$

where $\Omega$ is the image domain, $\oplus$ is the exclusive or operator, $L_o$

***Figure 2.*** *From left to right: Original image, Naturalness preserving method (NPE) [23], Fu* et al. *fusion scheme [7], and our result.*

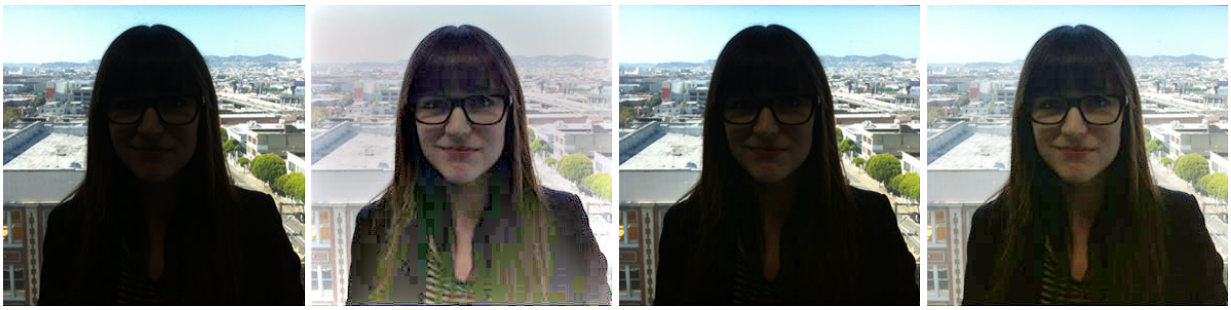

***Figure 3.*** *From left to right: Original image, Multi-scale Retinex approach (MSRCR) [17], Fu* et al. *variational model [6], and our result.*

is the lightness of the original image, $L_p$ is the luminance of the processed image and

$$U(p,q) = \begin{cases} 1 & \text{if } p \geq q \\ 0 & \text{if } p < q \end{cases} \tag{13}$$

Due to the high computational cost of the LOE metric, we follow the recommendations of [24] and we compute the results on a down-sampled version of the images containing the pixels of 100 rows and 100 columns. Results for this metric computed as the mean over all the images in the dataset are presented in Table 1. Our method outperforms all the others, being the method of Ying *et al.* [24] and the Fusion-based method of Fu *et al.* [7] the closest to our results.

We consider the Visual Information Fidelity (VIF) metric to measure the amount of distortion introduced by the different methods. The VIF quality metric (VIF) was proposed by Sheikh and Bovik in [20]. It can be understood as the ratio between the distorted image information and the reference image information. The reference image information is the mutual information between the reference image and the output of a process that estimates how this reference image is viewed by the Human Visual System (HVS). The distorted image information is the mutual information between the reference image and the output of the same HVS process, but considering the distorted image. Mathematically,

$$VIF = \frac{I(R;F)}{I(R;E)} \tag{14}$$

| | Vonikakis *et al.* | Li and Wu |
|---|---|---|
| Li and Wu [15] | 1059.30 | 845.11 |
| Ying *et al.* [24] | 257.89 | 269.25 |
| Dong *et al.* [4] | 876.22 | 827.03 |
| Fu *et al.* [6] | 562.84 | 313.00 |
| MSRCR [17] | 2717.70 | 2308.70 |
| NPE [23] | 818.50 | 845.11 |
| Fu *et al.* [7] | 480.84 | 497.67 |
| LIME [10] | 1293.80 | 1128.10 |
| Our approach | **211.23** | **244.81** |

**Table 1: Results for the Lightness Order Estimation metric computed as the mean over all the images in the dataset.**

where $R$ is the reference image, $D$ is the distorted image, $E = HVS(R)$ and $F = HVS(D)$ are the outputs of the HVS process explained above.

In our experiment, we follow the approach detailed in [24], and therefore we consider as reference image the output of the different methods and as distorted image the original one. In our case, a higher value on the $VIF$ metric indicates that the output of the method is respectful to the original content of the image. Results for this metric computed as the mean over all the images in the dataset are presented in Table 2. Our method is the most respectful one to the original content for the Vonikakis *et al.* dataset followed by the methods of Ying *et al.* and the Fusion-based method of Fu *et al.* [7]. In the Li and Wu dataset our method ranks fourth but the difference to the method of Ying *et al.* [24] is less

***Figure 4.*** *From left to right: Original image, Dong* et al. *[4], Li and Wu [15], our result.*

***Figure 5.*** *From left to right: Original image, LIME method [10], Ying* et al. *[24], our result.*

| | Vonikakis *et al.* | Li and Wu |
|---|---|---|
| Li and Wu [15] | 0.6335 | 0.7892 |
| Ying *et al.* [24] | 0.7781 | **0.9572** |
| Dong *et al.* [4] | 0.5158 | 0.6479 |
| Fu *et al.* [6] | 0.6682 | 0.8250 |
| MSRCR [17] | 0.4312 | 0.8456 |
| NPE [23] | 0.7065 | 0.9491 |
| Fu *et al.* [7] | 0.7369 | 0.9376 |
| LIME [10] | 0.3575 | 0.6104 |
| Our approach | **0.8866** | 0.8783 |

**Table 2: Results for the Visual Information Fidelity metric computed as the mean over all the images in the dataset.**

than 0.08.

## Conclusions

We have presented a method that, given a LDR backlit image, recovers the information that was not visible in the dark regions without losing information in the already well-exposed parts of the image. Our method first computes a set of weight maps from the iterates of the gradient descent of a energy functional. Then, it computes a set of tone-mapped results using the method of [3]. Finally, the tone-mapped images are combined with the weight maps to output a final image. We have shown that our method outperforms the state-of-the-art both visually and considering the LOE metric. Also, our approach is comparable with the state-of-the-art for the VIF metric. Further work will deal with the automatic adaptation of the different parameters in an image-based manner to improve the current results.

## References


[1] M. Bertalmío, V. Caselles, and E. Provenzi. Issues about retinex theory and contrast enhancement. *International Journal of Computer Vision*, 83(1):101–119, June 2009.

[2] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi. Perceptual color correction through variational techniques. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 16(4):1058–1072, April 2007.

[3] Praveen Cyriac, David Kane, and Marcelo Bertalmío. Optimized tone curve for in-camera image processing. *Electronic Imaging*, 2016, 2016.

[4] Xuan Dong, Guan Wang, Yi Pang, Weixin Li, Jiangtao Wen, Wei Meng, and Yao Lu. Fast efficient algorithm for enhancement of low lighting video. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6, July 2011.

[5] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.*, 27(3):67:1–67:10, August 2008.

[6] X. Fu, D. Zeng, Y. Huang, X. P. Zhang, and X. Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2782–2790, June 2016.

[7] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82 – 96, 2016.

[8] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío. Enhanced Variational Image Dehazing. *SIAM Journal on Imaging Sciences*, 8(3):1519–1546, January 2015.

[9] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío. Fusion-based variational image dehazing. *IEEE Signal Processing Letters*, 24(2):151–155, Feb 2017.

[10] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26:982–993, 2016.

[11] Jinggang Huang and David Mumford. Statistics of natural images and models. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1. IEEE, 1999.

[12] David Kane and Marcelo Bertalmío. System gamma as a

function of image-and monitor-dynamic range. *Journal of vision*, 16, 2016.

[13] Edwin H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, dec 1977.

[14] Joon-Young Lee, Kalyan Sunkavalli, Zhe Lin, Xiaohui Shen, and In-So Kweon. Automatic content-aware color and tone stylization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2470–2478, 2016.

[15] Z. Li and X. Wu. Learning-based restoration of backlit images. *IEEE Transactions on Image Processing*, 27(2):976–986, Feb 2018.

[16] Sylvain Paris, Samuel W. Hasinoff, and Jan Kautz. Local laplacian filters: Edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph.*, 30(4):68:1–68:12, July 2011.

[17] Ana Belén Petro, Catalina Sbert, and Jean-Michel Morel. Multiscale Retinex. *Image Processing On Line*, pages 71–88, 2014.

[18] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[19] Daniel L Ruderman. The statistics of natural images. *Network: computation in neural systems*, 5(4):517–548, 1994.

[20] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, Feb 2006.

[21] Chun Ming Tsai and Zong Mu Yeh. Contrast compensation by fuzzy classification and image illumination analysis for back-lit and front-lit color face images. *IEEE Transactions on Consumer Electronics*, 56(3):1570–1578, 8 2010.

[22] Vassilios Vonikakis, Rigas Kouskouridas, and Antonios Gasteratos. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, May 2017.

[23] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Processing*, 22(9):3538–3548, 2013.

[24] Zhenqiang Ying, Ge Li, and Wen Gao. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *CoRR*, abs/1711.00591, 2017.

[25] Lu Yuan and Jian Sun. Automatic exposure correction of consumer photographs. In Andrew W. Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *ECCV (4)*, volume 7575 of *Lecture Notes in Computer Science*, pages 771–785. Springer, 2012.

[26] S. W. Zamir, J. Vazquez-Corral, and M. Bertalmío. Gamut extension for cinema. *IEEE Transactions on Image Processing*, 26(4):1595–1606, 2017.

[27] S.W. Zamir, J. Vazquez-Corral, and M. Bertalmío. Gamut mapping in cinematography through perceptually-based contrast modification. *IEEE Journal of Selected Topics in Signal Processing*, 8(3):490–503, June 2014.

[28] Karel Zuiderveld. Graphics gems iv. chapter Contrast Limited Adaptive Histogram Equalization, pages 474–485. Academic Press Professional, Inc., San Diego, CA, USA, 1994.