IMAGE EXPOSURE ASSESSMENT: A BENCHMARK AND A DEEP CONVOLUTIONAL NEURAL NETWORKS BASED MODEL

Lijun Zhang[1], *Lin Zhang*[1,2*], *Xiao Liu*[1], *Ying Shen*[1], *Dongqing Wang*[1]

[1]School of Software Engineering, Tongji University, Shanghai, China
[2]Institute of Intelligent Automotive, Tongji University, Shanghai, China

ABSTRACT

In the camera equipment manufacturing industry, the exposure calibration is one of the basic steps for manufacturers to consider before launching their products to the market. To this end, a method that can objectively and automatically assess the exposure levels of images taken by the camera is highly desired. However, few studies have been conducted in this area. In this paper, we attempt to solve this issue to some extent and our contributions are twofold. Firstly, in order to facilitate the study of image exposure assessment, an Image Exposure Database ($IE_{ps}D$) is established. In this database, there are 15,582 images with various exposure levels, and for each image there is an associated subjective exposure score which could reflect its perceptual exposure level. Secondly, we propose a novel highly accurate DCNN-based model, namely $IE_{ps}M$ (Image Exposure Metric), to predict the exposure level of a given image.

***Index Terms*—** Exposure levels, image quality assessment, deep convolutional neural networks

1. INTRODUCTION

For many years, photogrammetry researchers, camera equipment manufacturers and recently, computer vision researchers have taken the camera calibration issue into consideration [1]. During this camera calibration process, camera exposure parameters need to be adjusted to eliminate the exposure distortion in images. However, currently the commonly used methods to determine whether an image is properly exposed or not are totally based on the experience of photographers, which is time-consuming, cumbersome, and cannot be implemented in systems where a real-time exposure level score is needed. Therefore, it is necessary to develop an image exposure metric to evaluate the image exposure level automatically, so as to become an objective index for camera exposure calibration.

In this paper, we attempt to address the problem of IEA (Image Exposure Assessment) to some extent. Our goal is to design an algorithm that could automatically and efficiently evaluate the exposure level of a given image and the evaluation results should be consistent with the human perception.

*Corresponding author. Email: cslinzhang@tongji.edu.cn

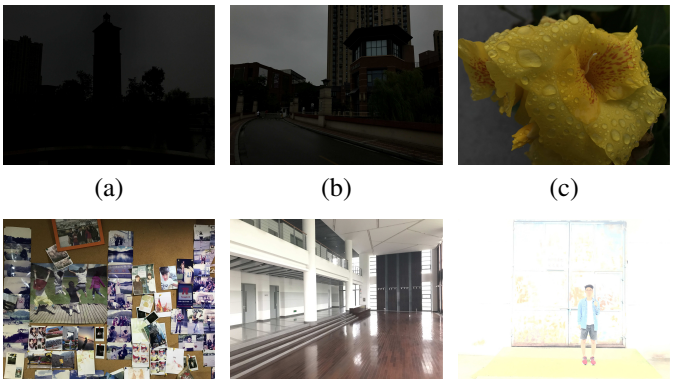

Fig. 1. (a)∼(f) are 6 images with various exposure levels. Their exposure scores predicted by our approach $IE_{ps}M$ are 1.0115, 2.0219, 2.9686, 4.1119, 5.2048, and 6.3953, respectively. The output range of $IE_{ps}M$ is from 1 to 7. "1" means the image is extremely underexposed, "4" means it is properly exposed, and "7" means it is extremely overexposed.

To demonstrate our goal more clearly, in Fig. 1, we show 6 images along with their exposure scores predicted by our proposed approach $IE_{ps}M$ (see Sect. 3 for details). It needs to be noted that the predicted exposure scores by $IE_{ps}M$ can vary from 1 to 7. "1" means that the image being examined is extremely underexposed, "4" means that it is properly exposed, and "7" means that it is extremely overexposed. Using $IE_{ps}M$, the predicted exposure scores of Figs. 1(a)∼1(f) are 1.0115, 2.0219, 2.9686, 4.1119, 5.2048, and 6.3953, respectively. It can be seen that the assessment results of the images' exposure levels by using $IE_{ps}M$ are highly consistent with the perceptual evaluation.

1.1. Related work

The most popular research area closely related with image exposure is referred to as High Dynamic Range (HDR) imaging or multi-exposure fusion, which is in order to enhance the dynamic range of an image by combining details in images taken under different exposures [2]. Another research area

978-1-5386-1737-3/18/$31.00

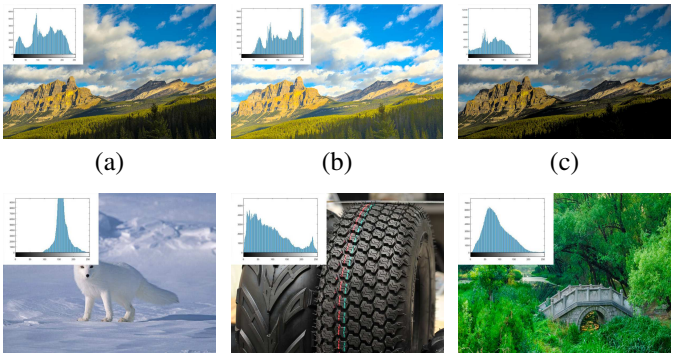

Fig. 2. (a)∼(c) show three images having the same contents but different exposure levels, along with their luminance histograms. (a) is properly exposed, while (b) and (c) are overexposed and underexposed, respectively. (d)∼(f) are three images that are all properly exposed; however, their luminance histograms are quite different from each other due to their different contents.

related to image exposure is to enhance image quality by optimizing the exposure decision in digital cameras [3]. In both of these two research areas, image exposure is considered as a controllable parameter that can be modulated by acquisition parameters, including aperture, ISO, and exposure time. The open issue is how to quantify the overall image exposure level by an automatic algorithm and how the quantification results correlate with the human perception.

At present, there exist some unsystematic methods to evaluate the image exposure level. In fact, the human experience suggests that the image histogram could reflect the image exposure level to some extent. According to this hypothesis, in [4], Efimov *et al.* evaluated the image exposure level by blocking the image and analyzing the luminance histogram of each image block. Based on the similar idea, Liu *et al.* [5] used the histogram distribution and relevant luminance features of images that are properly exposed to determine and adjust the exposure level of a given image. To better evaluate the presented method, Romaniak *et al.* [6] firstly proposed an exposure generation model by analyzing the mapping function between the image color saturation and the image exposure level to generate an experimental dataset. And then they designed an image exposure metric by analyzing the shape of the given image's histogram.

One shared drawback of the aforementioned IEA approaches [4–6] is that they are susceptible to the specific image content due to the histogram-based hypothesis. In this hypothesis, they simply associate image exposure levels with corresponding luminance histograms. As shown in Figs. 2(a)∼2(c), the three images have the same image content, while because of their different exposure levels, their histograms show different distribution patterns. For a properly exposed image Fig. 2(a), its histogram spreads over the whole range of the luminance, while histograms of over- and underexposed images are shifted to the right and the left side, respectively, as shown in Fig. 2(b) and Fig .2(c). Those existing methods [4–6] are just based on the hypothesis that the histograms have a strong direct link with image exposure levels. However, this hypothesis becomes questionable when applied to images of various contents. As shown in Figs. 2(d)∼2(f), the three images are all properly exposed, but their histogram distribution patterns are quite different due to their different image contents. That is to say, although the three given images are all properly exposed, the histogram-based hypothesis would lead to false judgements since the image histograms could be seriously influenced by specific image contents.

1.2. Our motivations and contributions

Having investigated the literature, we find that in the field of IEA, there is still a large room for further improvement in at least two aspects. Firstly, though the problem of IEA is of paramount importance and has great demand for camera exposure calibration, systematic studies in this area are quite rare. Hence, it is still a challenging open issue to develop a systematic method to assess image exposure. Secondly, for training and testing IEA algorithms, a public large-scale benchmark dataset, comprising images with different exposure levels and associated subjective scores, is indispensable. Unfortunately, such a dataset is still lacking in this area.

In this work, we attempt to fill the aforementioned research gaps to some extent. Our contributions are summarized as follows:

(1) To facilitate the study of IEA, we have established a benchmark dataset, namely $IE_{ps}D$ (*I*mage *E*xposure *D*atabase), and will make it publicly available. This dataset comprises 15,582 images with different exposure levels, including 1,512 collected real-world images and 14,070 artificial images generated using our proposed exposure simulation algorithm. Each image in $IE_{ps}D$ has an associated subjective score reflecting its exposure level. In our experiments, $IE_{ps}D$'s artificial images are used for training and validating IEA models, while its real-world images are used for testing. Please refer to Sect. 2 for more details about $IE_{ps}D$.

(2) Recent years, the deep convolutional neural networks (DCNN) have gained researchers much attention and achieved great success for numerous computer vision tasks. In this paper, we make an attempt to adopt DCNN to solve the IEA problem. Consequently, a novel DCNN-based approach, namely $IE_{ps}M$ (*I*mage *E*xposure *M*etric), is proposed to objectively assess the given image's exposure level. The efficacy and efficiency of $IE_{ps}M$ have been thoroughly evaluated in experiments.

The remainder of this paper is organized as follows. Section 2 presents the construction of $IE_{ps}D$. Section 3 describes the details of $IE_{ps}M$. Experimental results are reported in Section 4. Finally, Section 5 concludes the paper.

2. $IE_{PS}D$: AN IMAGE EXPOSURE DATABASE

In this section, the establishment of our image exposure database $IE_{ps}D$ is presented. Totally, $IE_{ps}D$ has 15,582 images with different exposure levels taken from a wide variety of shooting scenes, and for each image there is an associated subjective exposure score indicating its perceptual exposure level.

To fulfill this task, the construction of $IE_{ps}D$ mainly comprises three steps, including real-world images collection, artificial images generation, and subjective evaluation. Details are introduced in the following subsections.

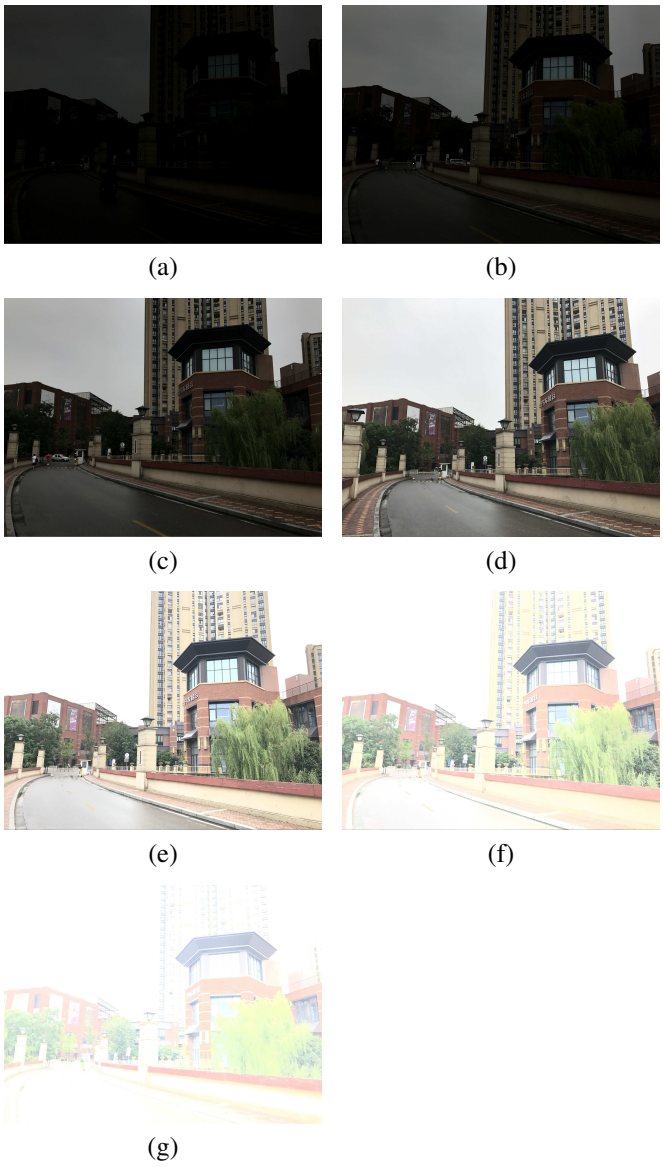

Fig. 3. (a)∼(g) are 7 images in $\mathcal{R}$ with different image exposure levels taken from one Outdoor Scenery scene. From (a)∼(g), the exposure levels are changing from "extremely underexposed" to "extremely overexposed".

Table 1. Shooting scenes for real-world images collection

	Building	Scenery	Human	Plant
Indoor (with man-made light)	6 *	6	3	3
Indoor (without man-made light)	6	6	3	3
Outdoor	12	12	6	6

* Choose 6 different locations of Indoor Building with man-made light to form 6 scenes.

2.1. Real-world images collection

As stated in Sect. 1, since there is no systematic method in the field of IEA, we are motivated to fill the research gap to some extent by building up $IE_{ps}D$ at first. In [6], Romaniak *et al.* designed an image exposure generation model for artificially generating images with various exposure levels. However, they failed to take real-world images into account. In our dataset, we include real-world images with different exposure levels, which can help evaluate the effectiveness of IEA algorithms on real data.

At this step, we used digital cameras to collect real-world images with diverse exposure levels. In general, exposure is the total amount of light allowed to fall on a photographic medium during the process of taking a photograph and there are basically three methods to control exposure in digital cameras. The first way is by opening or closing the aperture. The larger the hole of the iris, the more light is reaching image sensors during a fixed period of time. The second way is by changing the light signal boost, which is commonly referred to as ISO sensitivity. The last way is to change the exposure time. For different shooting scenarios, the most straightforward strategy to obtain similar exposure levels is to only adjust one of the above factors and keep other camera parameters fixed.

Furthermore, it is necessary to collect real-world images from all kinds of common scenes, as many as possible, in order to verify the algorithm's robustness in dealing with various image contents. Therefore, we draw up a collection plan as shown in Table 1, to include common scenes. Specifically, we selected 4 types of shooting objects and 3 types of shooting occasions, which altogether form 72 different scenes, to cover the most common application scenarios. For each occasion, we took the distance-view, close-shot, medium-shot and macro-mode into account. For each scenario, we captured 21 photos under 7 exposure levels, 3 images each, by adjusting the exposure time. Finally, we collected 1,512 (72×21) real-world images and we denote the dataset formed by these real-world images by $\mathcal{R}$. 7 sample images of $\mathcal{R}$, taken from the scene, are shown in Fig. 3. From Figs. 3(a)∼3(g), the exposure levels are changing from "extremely underexposed" to "extremely overexposed".

2.2. Artificial images generation

The on-spot data collection is quite time-consuming and laborious. However, in order to obtain a prediction model with a high generalization capability, a large-scale dataset is indispensable. To cope with this contradiction, a novel method for simulating images with various exposure levels from source properly exposed images is proposed.

Suppose that I is a source properly exposed image. Its variant $\hat{I}$ having a different exposure level could be generated by adjusting I's illumination and saturation channels. To manipulate its illumination and saturation channels separately, I is firstly transferred from the RGB space to the HSV space. Denote by I_v and I_s the illumination channel and the saturation channel of I, respectively. Similarly, we denote by $\hat{I}_v$ and $\hat{I}_s$ the illumination channel and the saturation channel of $\hat{I}$, respectively. $\hat{I}_v$ is obtained by adjusting I_v as,

$$\hat{I}_v(\mathbf{x}) = I_v(\mathbf{x}) + C_1 \tag{1}$$

where $\mathbf{x}$ indicates the spatial location and C_1 is a tunable parameter. For simulating overexposed images, C_1 should be positive, while for simulating underexposed images, C_1 should be negative. In the meanwhile, we also need to adjust I_s to $\hat{I}_s$ according to the mapping function as suggested by Romaniak *et al*. [6],

$$\hat{I}_s(\mathbf{x}) = \left(\frac{e^{\ln I_s(\mathbf{x})^{2.5} - \ln(1-I_s(\mathbf{x})^{2.5}) + C_2}}{1 + e^{\ln I_s(\mathbf{x})^{2.5} - \ln(1-I_s(\mathbf{x})^{2.5}) + C_2}}\right)^{0.4} \tag{2}$$

where C_2 is a tunable parameter. For simulating overexposed images, C_2 needs to negative, while for simulating underexposed images, C_2 needs to be positive.

By adjusting the parameter settings, a series of I's variants with different exposure levels can be simulated. In our experimental settings, 7 exposure levels were created. That is to say, from each source properly exposed image, we generated 7 its variants (including itself) having different exposure levels, ranging from "extremely underexposed" to "extremely overexposed". In Fig. 4, (a) shows one sample source image while (b)∼(g) are its variants with different exposure levels simulated by our proposed algorithm. Figs. 4(b)∼4(d) are underexposed images while Figs. 4(e)∼4(g) are overexposed ones. It can be seen that the simulation results using our algorithm correlate well with the perceptual evaluation.

For establishing $IE_{ps}D$, 2,010 properly exposed images were collected from the Internet, which cover all the scenes listed in Table 1. Taking them as source images, 14,070 (2010 $\times$ 7) images covering a wide range of exposure levels were simulated using our algorithm. We denote by $\mathcal{A}$ the dataset formed by them.

2.3. Subjective evaluation

After collecting and generating images with various image exposure levels, the exposure scores of images, which could reflect the corresponding exposure levels, were evaluated by human subjective judgements. We firstly offered a uniform standard for evaluators to give the images exposure scores based on their own perceptions.

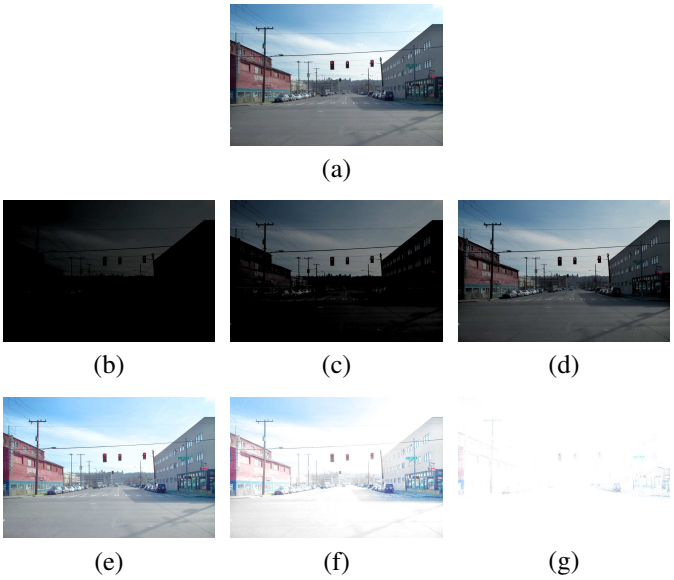

Fig. 4. (a)∼(g) are 7 images in $\mathcal{A}$ with different image exposure levels simulated from one properly exposed image (a) using our proposed exposure level simulation algorithm. (b)∼(d) are underexposed images and (e)∼(g) are overexposed ones.

In our experiment, we employed a 7-point system, with which the subjective scores given by an individual were integers ranging from 1 to 7. The closer the score is to 4 point, the more likely the image is of proper exposure, while the score below 4 means it is underexposed and the score beyond 4 means it is overexposed. For each image, there are 10 raw subjective evaluations. Sample images with typical different exposure levels were demonstrated to the evaluators before the subjective test. A single-stimulus continuous quality evaluation [7] was conducted.

Then, we performed some postprocessing steps to the raw scores. At first, we filtered out those heavily biased subjective scores that satisfy

$$d_{ij} - \overline{d_j} > T \cdot \sigma_j \tag{3}$$

where d_{ij} is the exposure score of the image I_j given by the ith evaluator, $\overline{d_j}$ is the mean score of I_j, T is the threshold constant and σ_j is the standard deviation value of I_j's scores. Then, to eliminate the influence of different subjective evaluation standards of evaluators, the raw scores d_{ij} were converted as,

$$z_{ij} = \frac{d_{ij} - \overline{d_i}}{\sigma_i} \tag{4}$$

where $\overline{d_i}$ is the mean score of the ith evaluator and σ_i is the

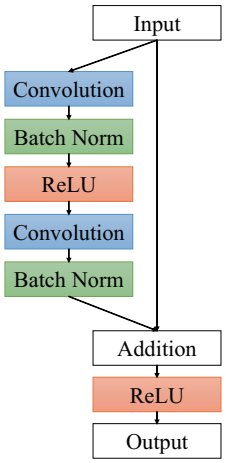

Fig. 5. A residual block.

standard deviation of his/her scores for all images. We regard the mean of the evaluation scores of I_j as its final subjective exposure score s_j,

$$s_j = \frac{1}{N_j} \sum z_{ij} \tag{5}$$

where N_j is the number of the subjective scores for I_j.

Now, for each image I_j in $IE_{ps}D$, an associated subjective score s_j reflecting its exposure level is obtained.

3. $IE_{PS}M$: A DCNN-BASED IMAGE EXPOSURE METRIC

In this paper, we propose a DCNN-based image exposure metric, namely $IE_{ps}M$. Inspired by the great performance in categorizing different illumination patterns on faces [8] which, as we do, focuses on the illumination component of the image, we employ the Deep Residual Networks [9] as well. The key idea of [9] is to take a standard feed-forward ConvNet and add skip connections that bypass (or shortcut) a few convolution layers at a time. Each bypass gives rise to a residual block in which the convolution layers predict a residual that is added to the blocks input tensor. A residual block is shown in Fig. 5.

In $IE_{ps}M$, we select ResNet-50 for its suitable depth and complexity to solve the problem of IEA. And we adjust the neural network structure by changing the final output number into 1 and the loss layer into EuclideanLoss so that it could solve the regression problem. As required by the structure of ResNet, the input image needs to be resized to 224×224 before being fed into ResNet.

$IE_{ps}M$ was trained on the training set of $IE_{ps}D$. The base learning rate was 0.001 and learning rate policy was "inverse policy" to decrease the learning rate with iterations. With respect to the optimization solver, we resorted to ADAM [10].

Table 2. Performance comparison with IEA algorithms

METHOD	SROCC	KROCC
Method in [6]	0.8669	0.7276
$\boldsymbol{IE_{ps}M}$	**0.9540**	**0.8532**

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

Our established dataset $IE_{ps}D$ as described in Sect. 2 was explored to train $IE_{ps}M$ and to evaluate the prediction performance of different IEA algorithms. It needs to be noted that in order to test the validity of our proposed exposure level simulation algorithm, when training $IE_{ps}M$, only the images of subset $\mathcal{A}$ (images in $\mathcal{A}$ were simulated from source properly exposed images) were used. Specifically, images in $\mathcal{A}$ associated with 70% of the source images were used as the training set and the remaining ones in $\mathcal{A}$ were used as the validation set. We regarded the real-world image dataset $\mathcal{R}$ of $IE_{ps}D$ as the test set to validate the IEA algorithms' effectiveness for dealing with real-world images.

To evaluate the performance of competing IEA methods, we adopt two correlation coefficients to measure the monotonic coherency between the prediction results and the subjective scores: Spearman rank-order correlation coefficient (SROCC) and Kendall rank-order correlation coefficient (KROCC). A value closer to 1 indicates a better result of exposure level estimation for both indices.

4.1. Comparisons with other exposure metrics

As mentioned in Sect. 1.1, studies particularly focusing on the IEA problem are quite sporadic. And considering the similar basic hypothesis used in [4–6], we chose to implement the algorithm in [6] for comparison because of its integrity and reproducibility. In Table 2, we list the two correlation coefficients, SROCC and KROCC, achieved by each method on the test set of $IE_{ps}D$.

4.2. Comparisons with NR-IQA methods

To some extent, the IEA problem can also be considered as a special kind of NR-IQA (No-Reference Image Quality Assessment) problem. NR-IQA algorithms aim to automatically evaluate the overall quality of a given image. Hence, in our experiment, we also compared $IE_{ps}M$ with several state-of-the-art NR-IQA approaches, including BRISQUE [11], NIQE [12], SSEQ [13], LPSI [14], ILNIQE [15], TCLT [16], and OG-IQA [17]. The evaluation results on the test set of $IE_{ps}D$ in terms of SROCC and KROCC are summarized in Table 3.

4.3. Discussion

The results listed in Table 2 lead us to the following conclusions. First, $IE_{ps}M$ achieves a high SROCC (above 0.95) on

Table 3. Performance comparison with NR-IQA methods

METHOD	SROCC	KROCC
NIQE [12]	0.0638	0.0455
IL-NIQE [15]	0.0658	0.0490
SSEQ [13]	0.4460	0.3276
LPSI [14]	0.5175	0.4133
TCLT [16]	0.5792	0.4390
BRISQUE [11]	0.5930	0.4536
OG-IQA [17]	0.7991	0.5883
$\boldsymbol{IE_{ps}M}$	**0.9540**	**0.8532**

the test set, which suggests that it has a high accuracy for predicting the perceptual exposure levels for real-world images. Second, it needs to be stressed that when training $IE_{ps}M$, only the subset $\mathcal{A}$, whose images were generated using our exposure level simulation algorithm, was used and none improperly exposed real-world images was involved. This fact indirectly proves the validity of our algorithm for simulating different exposure levels. Third, $IE_{ps}M$ performs much better than the method in [6], which is in line with the shortcoming analysis in Sect. 1.1 about the histogram-based hypothesis.

The superiority of our method $IE_{ps}M$ over the other competitors in NR-IQA can be clearly observed from Table 3. The evaluation results also indicate that though the predictions of NR-IQA algorithms can reflect the general quality of images, they cannot faithfully measure the distortions aroused by improper exposures. Thus, it is more suitable to solve the IEA problem with a specifically designed method rather than a general-purpose one.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we focus on addressing the problem of Image Exposure Assessment. Our contributions are twofold. First, to facilitate the study of exposure levels prediction for images, we have constructed a database, namely $IE_{ps}D$. It comprises 15,582 images and each one is assigned an exposure score evaluated through human judgements. Second, we proposed a DCNN-based model $IE_{ps}M$ to predict the image's exposure level whatever the content of the image is. Experiments conducted on $IE_{ps}D$ show that the proposed metric $IE_{ps}M$ outperforms all its competitors by a large margin, making it quite attractive for real applications. In the future, we would try to embed the proposed method in camera equipments.

6. ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China under grant no. 61672380 and in part by the Fundamental Research Funds for the Central Universities under grant no. 2100219068.

7. REFERENCES

[1] M.H. Saeifar and M.M. Nia, "Camera calibration: An overview of concept, methods and equations," *IJERA*, vol. 7, no. 7, pp. 49–57, Jul. 2017.

[2] P. Ke, C. Jung, and Y. Fang, "Perceptual multi-exposure image fusion with overall image quality index and local saturation," *MMSys*, vol. 23, no. 2, pp. 239–250, 2017.

[3] H.J. Park and H.H. Dong, "The optimum exposure decision for the enhanced image performance using a digital camera," in *IST*, 2010, pp. 333–336.

[4] S. Efimov, A. Nefyodov, and M. Rychagov, "Block-based image exposure assessment and indoor/outdoor classification," *GraphiCon*, pp. 23–27, Jun. 2007.

[5] M. Liu, P. Yuan, and R.S. Turner, "Automatic analysis and adjustment of digital images with exposure problems," U.S. Patent 7,646,931, Dec. 2008.

[6] P. Romaniak, L. Janowski, M. Leszczuk, and Z. Papir, "A no reference metric for the quality assessment of videos affected by exposure distortion," in *ICME*, 2011, pp. 1–6.

[7] B.T. Itu-R, "Methodology for the subjective assessment of the quality of television pictures," *ITU*, Apr. 2005.

[8] L. Zhang, L. Zhang, and L. Li, "Illumination quality assessment for face images: A benchmark and a convolutional neural networks based model," in *ICONIP*, 2017, pp. 583–593.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[10] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2014.

[11] A. Mittal, A.K. Moorthy, and A.C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. IP*, vol. 21, no. 12, pp. 4695–4708, 2012.

[12] A. Mittal, R. Soundararajan, and A.C. Bovik, "Making a completely blind image quality analyzer," *IEEE SPL*, vol. 20, no. 3, pp. 209–212, 2013.

[13] L. Liu, B. Liu, H. Huang, and A.C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

[14] Q. Wu, Z. Wang, and H. Li, "A highly efficient method for blind image quality assessment," in *ICIP*, 2015, pp. 339–343.

[15] L. Zhang, L. Zhang, and A.C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. IP*, vol. 24, no. 8, pp. 2579, 2015.

[16] Q. Wu, H. Li, F. Meng, K.N. Ngan, B. Luo, C. Huang, and B. Zeng, "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Trans. CSVT*, vol. 26, no. 3, pp. 425–440, 2016.

[17] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A.C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Processing: Image Communication*, vol. 40, no. C, pp. 1–15, 2016.