Illumination Quality Assessment for Face Images: A Benchmark and a Convolutional Neural Networks Based Model

Lijun Zhang[1], Lin Zhang[1](✉), and Lida Li[2]

[1] School of Software Engineering, Tongji University, Shanghai, China
{1632761,cslinzhang}@tongji.edu.cn
[2] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
cslli@comp.polyu.edu.hk

Abstract. Many institutions, such as banks, usually require their customers to provide face images under proper illumination conditions. For some remote systems, a method that can automatically and objectively evaluate the illumination quality of a face image in a human-like manner is highly desired. However, few studies have been conducted in this area. To fill this research gap to some extent, we make two contributions in this paper. Firstly, in order to facilitate the study of illumination quality prediction for face images, a large-scale database, namely, *F*ace *I*mage *I*llumination *Q*uality *D*atabase (FIIQD), is established. FIIQD contains 224,733 face images with various illumination patterns and for each image there is an associated illumination quality score. Secondly, based on deep convolutional neural networks (DCNN), a novel highly accurate model for predicting the illumination quality of face images is proposed. To make our results reproducible, the database and the source codes have been made publicly available at https://github.com/zhanglijun95/FIIQA.

Keywords: Illumination quality assessment · Illumination transfer · Convolutional neural networks

1 Introduction

Over the past few decades, with the rapid development of e-commerce, more and more commercial institutions are going to provide remote services for customers to initiate their commercial activities anywhere [1]. Among the basic information required by those institutions, a face image under uniform and adequate lighting environment of the remote user is usually a must. For example, in a remote bank system, the success of account establishment process will be effected by the illumination condition on the input face image collected by the user's equipment. In such a case, if the system has a module that could dynamically evaluate the input face image's illumination quality and give the user some hints for adjusting

D. Liu et al. (Eds.): ICONIP 2017, Part III, LNCS 10636, pp. 583–593, 2017.
https://doi.org/10.1007/978-3-319-70090-8_59

the ambient light accordingly, it would be quite helpful. Therefore, a method that can automatically monitor the illumination quality of face images is desired.

Thus in this paper, we focus on addressing the face image illumination quality assessment (FIIQA) problem. Our goal is to design an algorithm that could automatically and efficiently evaluate the illumination quality of a given face image and the evaluation results should correlate well with human judgements. To demonstrate our goal more clearly, in Fig. 1, we show 4 face images along with their illumination quality scores predicted by our proposed approach FIIQA_{DCNN} (see Sect. 3 for details). With FIIQA_{DCNN}, the predicted illumination quality scores of Fig. 1(a)–(d) are 0.0, 0.503, 0.670, and 1.0, respectively. It can be seen that the results are highly consistent with subjective evaluations.

Fig. 1. (a)–(d) are four face images. Their illumination quality scores predicted by our approach FIIQA_{DCNN} are 0.0, 0.503, 0.670, and 1.0, respectively.

1.1 Related Work

Studies particularly focusing on FIIQA are quite sporadic. In [2], Sellahewa and Jassim proposed an objective measure of face illumination quality to decide whether the face image should be preprocessed by illumination normalization. However, their quality index is adapted from a full-reference image quality assessment model UQI [3] and thus requires a reference image, which greatly reduces its practicability.

There are several studies focusing on face image quality assessment (FIQA). In these papers, usually various factors effecting quality are integrated together to induce a score as an overall quality index of the examined face image. These factors may include sharpness, noise, illumination, pose, etc. In [4], the authors considered several factors of face image quality and proposed an evaluation method for each factor separately. For illumination quality assessment, they partition the image into blocks and regard the weighted-average of blocks as the quality score. Such a simple model is not powerful enough to characterize the illumination quality of face images. In [5], a face selection technique using local patch-based probabilistic image quality assessment was proposed for video-based face recognition. However, this method is video-based and cannot be used for still images. In [6], Chen et al. proposed a learning-based approach for FIQA by fusing multiple features.

Actually, the FIIQA problem can also be considered as a special kind of NR-IQA problem. The aim of the NR-IQA research is to design an algorithm that can automatically evaluate the overall quality of a given image. In recent years, many eminent NR-IQA algorithms have emerged, such as BRISQUE [7], NIQE [8], SSEQ [9], LPSI [10], IL-NIQE [11], TCLT [12], OG-IQA [13], etc. Their performance for FIIQA have also been evaluated in our experiments (see Sect. 4 for details).

1.2 Our Motivations and Contributions

Having investigated the literature, we find that in the field of FIIQA, there is still large room for further improvement in at least two aspects. Firstly, though the problem of FIIQA is of paramount importance and has great demand for institutions providing remote services with customers face images, the studies in this area are quite rare. Hence, how to assess the illumination quality of a given image is still a challenging open issue. Secondly, for training and testing FIIQA algorithms, a public large-scale benchmark dataset, comprising face images with associated subjective scores and covering various real-world illumination patterns, 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) A large-scale database, namely, Face Image Illumination Quality Database (FIIQD) is constructed. This dataset comprises 224,733 face images with various real-world illumination patterns, each of which has an associated subjective score reflecting its illumination quality. To our knowledge, this is the first large-scale dataset established for the study of illumination quality assessment of face images.
(2) Recent years, the deep convolutional neural networks (DCNN) have gained researchers' much attention and achieved great success for numerous computer vision tasks [14,15]. In this paper, we make an attempt to adapt DCNN to solve the FIIQA problem. Consequently, a novel FIIQA model based on DCNN is proposed, namely FIIQA$_{DCNN}$. Experimental results have shown that FIIQA$_{DCNN}$ has an extremely strong capability in predicting the illumination quality of a given face image.

To make the results fully reproducible, the collected dataset and the source codes of FIIQA$_{DCNN}$ are publicly available at https://github.com/zhanglijun95/FIIQA.

The remainder of this paper is organized as follows. Section 2 presents steps for FIIQD construction. Section 3 describes the details of FIIQA$_{DCNN}$. Experimental results are reported in Sect. 4. Finally, Sect. 5 concludes the paper.

2 FIIQD: A Face Image Illumination Quality Database

In this section, the steps for establishing FIIQD are presented. To fulfill this task, we adopt a semi-automatic strategy. The construction of FIIQD mainly comprises four steps, the construction of the image set with source illumination patterns, the subjective evaluation of illumination pattern images, the construction of target face set, and illumination transfer. The pipeline of FIIQD construction is shown in Fig. 2.

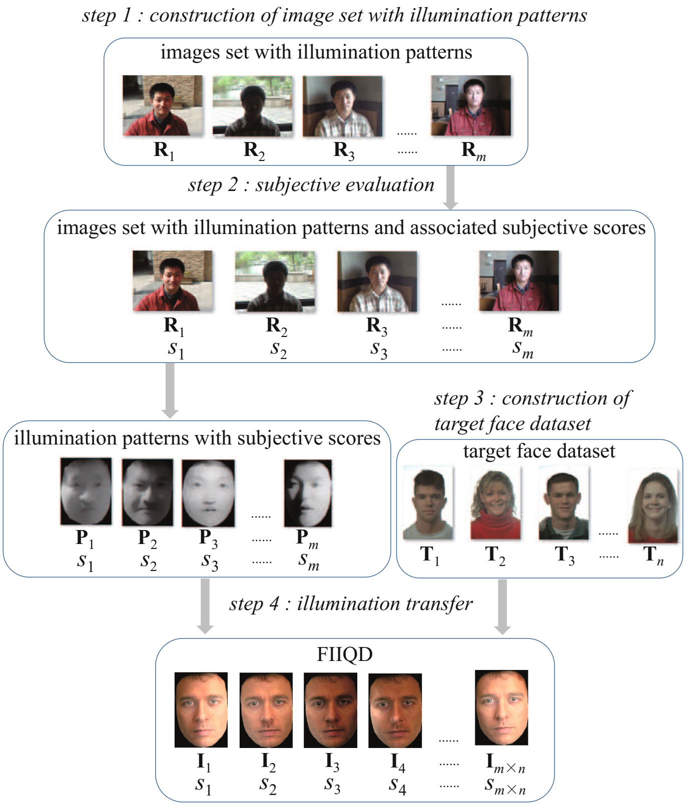

Fig. 2. The pipeline of FIIQD construction.

2.1 Step 1: The Construction of the Image Set with Source Illumination Patterns

In this step, our aim is to collect an image set containing face images with various real-world illumination patterns. These images are expected to provide source illumination patterns to be transferred to target face images so they should cover sufficient types of illumination conditions from real world.

Table 1. The plan for collecting images with source illumination patterns

Index	Scene	Illumination patterns	Time slots
1	Office	Toplight	Day/Night
		Lamp	
		Natural light	
2	Cafe	Window natural light	Day/Night
		Sunshade natural light	8 am/2 pm/5 pm
		Inside light	Sunny/Cloudy
3	Mall	Atrium light	Day/Night
		Shop light	
		Corridor light	
4	Library	Window natural light	Day/Night
		Toplight	3F/8F/14F
5	Home	Toplight	Day/Night
		Lamp	
6	Outside	Natural light	Day/Night
			8 am/2 pm/5 pm
			Sunny/Cloudy

Taking the abovementioned requirements and the typical application scenarios into consideration, we worked out a plan for collecting images with source illumination patterns as showed in Table 1. We selected 6 types of scenes to cover the most common application scenarios. For each scene, we define several different illumination situations by combining different illumination patterns with different time points of a day. And we captured at least 16 photos, 2 every 45° for each illumination situation. At last, we collected 499 images with various source illumination patterns and selected 200 from them, whose qualities are good enough, to form the image set $\mathcal{R}$ with illumination patterns. 6 samples from this set are showed in Fig. 3.

2.2 Step 2: Subjective Evaluation of Images in $\mathcal{R}$

In this step, the illumination quality of images in $\mathcal{R}$ is evaluated by subjective judgements. To achieve this goal, a single-stimulus continuous quality evaluation [16] was conducted. Then, we performed some postprocessing steps to the

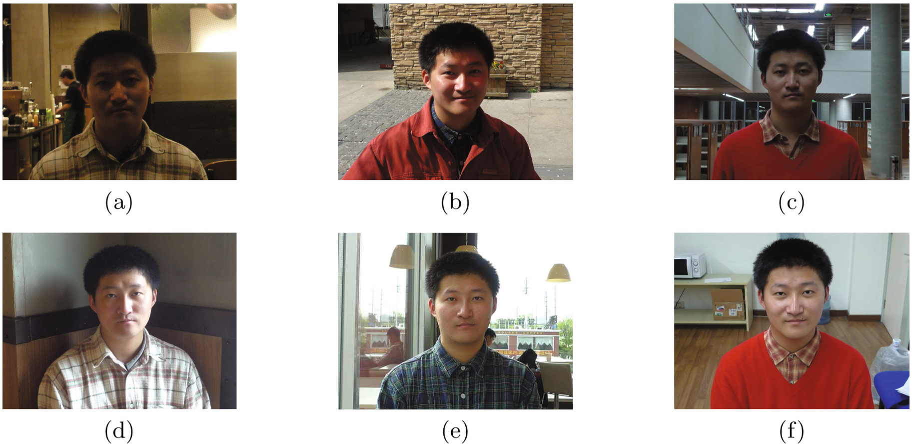

Fig. 3. (a)–(f) are six sample images with illumination patterns. Their associate subjective scores of illumination quality are 0.083, 0.175, 0.341, 0.667, 0.833 and 1.0, respectively.

raw scores. At first, we filtered out those heavily biased subjective scores that satisfy

$$d_{ij} - \overline{d_j} > T \cdot \sigma_j \tag{1}$$

where d_{ij} is the illumination quality score of the image $\mathbf{R}_j \in \mathcal{R}$ given by the ith evaluator, $\overline{d_j}$ is the mean score of $\mathbf{R}_j$, T is the threshold constant and σ_j is the standard deviation of $\mathbf{R}_j$'s scores. Then, to eliminate the effect of different subjective evaluation standards of evaluators, the raw scores d_{ij} were converted as,

$$z_{ij} = \frac{d_{ij} - \overline{d_i}}{\sigma_i} \tag{2}$$

where $\overline{d_i}$ is the mean score of the ith evaluator and σ_i is the standard deviation of his scores for all images in $\mathcal{R}$. We regard the mean evaluation score of $\mathbf{R}_j$ as its final subjective illumination quality score,

$$s_j = \frac{1}{N_j} \sum z_{ij} \tag{3}$$

where N_j is the number of the subjective scores for $\mathbf{R}_j$.

Now, for each image $\mathbf{R}_j \in \mathcal{R}$, an associated subjective score s_j reflecting its illumination quality is obtained.

2.3 Step 3: Target Face Set Construction

In this step, we built a target face image set including 1134 face images under uniform illumination from 1014 unique subjects. We established this set by selecting suitable face images from existing face datasets, such as YaleB [17], PIE [18], FERET [19], etc. In consideration of diversity, the subjects in target face image set have wide distributions over several attributes, such as face shape, race, skin color, gender, generation, etc.

2.4 Step 4: Illumination Transfer

In this step, we transfer the illumination patterns from images in $\mathcal{R}$ to the images in the target face set by implementing the illumination transfer algorithm proposed in [20]. This step results images in the final FIIQD. Sample images in FIIQD are showed in Fig. 4. The first row are images with source illumination patterns; the first column are target images; the others are the corresponding transferred images in FIIQD. We obtain a database with 224,733 face images with various illumination patterns at last.

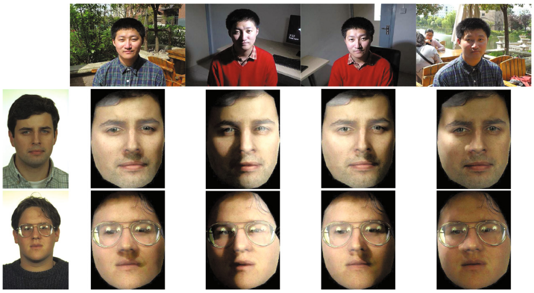

Fig. 4. Samples of the FIIQD.

Suppose that $\mathbf{R}_j \in \mathcal{R}$ is an image with a source illumination pattern and its subjective score is s_i. For the images whose illumination patterns are transferred from $\mathbf{R}_i$, their illumination quality scores are assigned as s_i.

3 FIIQA$_{DCNN}$: A Face Illumination Quality Assessment Model Based on DCNN

In this paper, we propose an FIIQA method, FIIQA$_{DCNN}$, based on DCNN and to our knowledge, our work is the first one to introduce DCNN to FIQA, not alone FIIQA.

In FIIQA$_{DCNN}$, we adopt the Deep Residual Networks in [15], which won the 1st places in ImageNet classification, ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation in ILSVRC and COCO 2015 competitions. The key idea of [15] 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 block's input tensor.

In FIIQA$_{DCNN}$, we select ResNet-50 as our model for its suitable depth and complexity to solve our problem. ResNet-50 has 53 convolution layers, 2 pooling layers and 1 fully connected layer. And the output number of the last fully connected layer is the number of classes, which is 200 in our case since we have 200 different illumination patterns in FIIQD.

In our implementation, we resize the input images into the size 224×224 and train the ResNet-50 model on the FIIQD training set. The learning rate starts from 0.01 and is divided by 10 when the error reaches plateaus, and the models are trained for up to $500K$ iterations. And we keep the weight decay and momentum the same with the original network settings.

4 Experimental Results and Discussion

4.1 Database Partition and Experimental Protocol

In experiments, we used FIIQD, constructed in Sect. 2, to train FIIQA$_{DCNN}$ and compare it with the other state-of-art algorithms. We partitioned the whole database into 3 subsets whose face identities are independent from each other. Table 2 shows the details of each subset, 71% used for training, 14% used for validation, and the remaining 15% used for testing.

Table 2. Partition of FIIQD

Name	#images	#face identities	Ratio
Training set	159159	709	71%
Validation set	30930	141	14%
Testing set	34644	164	15%

To evaluate the performance of our method, we adopted 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 quality estimation for both indices.

4.2 Comparisons with FIQA Methods

In this experiment, the performance of two state-of-the-art competing FIQA methods were evaluated. The first one is the RQS [6], which evaluates the overall quality of a face image. The other one is the method proposed in [4], which can evaluate the illumination quality of a face image. In Table 3, we list the two correlation coefficients, SROCC and KROCC, achieved by each method on the testing set of FIQD.

Table 3. Performance comparison with FIQA methods

Method	SROCC	KROCC
RQS	0.125	0.086
Method in [4]	0.6873	0.5031
$\mathbf{FIIQA_{DCNN}}$	**0.9477**	**0.8915**

4.3 Comparisons with NR-IQA Methods

In this experiment, we compared FIIQA_{DCNN} with some state-of-the-art NR-IQA methods, including BRISQUE [7], NIQE [8], SSEQ [9], LPSI [10], IL-NIQE [11], TCLT [12], OG-IQA [13]. As no source codes of LPSI has been released yet, we implemented it by ourselves and tuned all parameters to achieve its best results. For the rest of the competing methods, we used the source codes provided by their authors. In Table 4, we list SROCC and KROCC achieved by each method on the testing set of FIIQA.

Table 4. Performance comparison with NR-IQA methods

Method	SROCC	KROCC
BRISQUE	0.0487	0.0333
NIQE	0.0260	0.0173
IL-NIQE	0.0459	0.0314
SSEQ	0.1185	0.0811
LPSI	0.1255	0.0847
TCLT	0.1600	0.1094
OG-IQA	0.1757	0.1209
$\mathbf{FIIQA_{DCNN}}$	**0.9477**	**0.8915**

4.4 Discussion

Based on the results listed in Table 3, we could have the following findings. At first, FIIQA_{DCNN} performs the best among all of the methods and achieves a high SROCC around 0.95. Secondly, the method in [4] performs much better than RQS. The major difference between them is that ROS evaluates the face image quality as a whole and does not consider the illumination factor separately while the method in [4] is specially designed for measuring illumination quality. This fact indicates that it is better to solve the FIIQA problem separately from the FIQA problem.

The superiority of our method FIIQA_{DCNN} over the other competitors in NR-IQA can be clearly observed from Table 4. And it indicates that the predictions of NR-IQA can reflect the general quality of an image, but cannot reflect the illumination quality of a face image. It is more suitable to solve the problem using concrete analysis of face illumination.

5 Conclusions and Future Work

In this paper, we focus on addressing the problem of face image illumination quality assessment. Our contributions are twofold. First, we have constructed a large-scale database, namely FIIQD. It contains over $224K$ face images with various illumination patterns and each one is assigned an illumination quality score as ground truth. Second, we are the first to employ DCNN models to predict illumination quality of face images. Experiments conducted on FIIQD show that the proposed FIIQA model FIIQA$_{DCNN}$ outperforms all its competitors by a large margin, making it quite attractive for real applications.

Acknowledgments. This work was supported in part by the Natural Science Foundation of China under grant no. 61672380 and in part by the ZTE Industry-Academia-Research Cooperation Funds under grant no. CON1608310007.

References

1. Xia, D., Cui, D., Wang, J., Wang, Y.: A novel data schema integration framework for the human-centric services in smart city. ZTE Commun. **13**(4), 25–33 (2015)
2. Sellahewa, H., Jassim, S.A.: Image-quality-based adaptive face recognition. IEEE Trans. IM **59**(4), 805–813 (2010)
3. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. IP **13**(4), 600–612 (2004)
4. Truong, Q.C., Dang, T.K., Ha, T.: Face quality measure for face authentication. In: Dang, T.K., Wagner, R., Küng, J., Thoai, N., Takizawa, M., Neuhold, E. (eds.) FDSE 2016. LNCS, vol. 10018, pp. 189–198. Springer, Cham (2016). doi:10.1007/978-3-319-48057-2_13
5. Wong, Y., Chen, S., Mau, S., Sanderson, C., Lovell, B.C.: Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In: IEEE CVPR Workshops, pp. 74–81 (2011)
6. Chen, J., Deng, Y., Bai, G., Su, G.: Face image quality assessment based on learning to rank. IEEE SPL **22**(1), 90–94 (2015)
7. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Trans. IP **21**(12), 4695–4708 (2012)
8. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a 'completely blind' image quality analyzer. IEEE SPL **20**(3), 209–212 (2013)
9. Liu, L., Liu, B., Huang, H., Bovik, A.C.: No-reference image quality assessment based on spatial and spectral entropies. Sig. Process. Image Commun. **29**(8), 856–863 (2014)
10. Wu, Q., Wang, Z., Li, H.: A highly efficient method for blind image quality assessment. In: IEEE ICIP, pp. 339–343 (2015)
11. Zhang, L., Zhang, L., Bovik, A.C.: A feature-enriched completely blind image quality evaluator. IEEE Trans. IP **24**(8), 2579–2591 (2015)
12. Wu, Q., Li, H., Meng, F., Ngan, K.N., Luo, B., Huang, C., Zeng, B.: Blind image quality assessment based on multichannel feature fusion and label transfer. IEEE Trans. CSVT **26**(3), 425–440 (2016)
13. Liu, L., Hua, Y., Zhao, Q., Huang, H., Bovik, A.C.: Blind image quality assessment by relative gradient statistics and adaboosting neural network. Sig. Process. Image Commun. **40**, 1–15 (2016)

14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE CVPR, pp. 770–778 (2016)
16. Assembly, ITU Radiocommunication: Methodology for the Subjective Assessment of the Quality of Television Pictures. International Telecommunication Union (2003)
17. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. PAMI **23**(6), 643–660 (2001)
18. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination and expression database. IEEE Trans. PAMI **25**(12), 1615–1618 (2003)
19. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. PAMI **22**(10), 1090–1104 (2000)
20. Chen, X., Chen, M., Jin, X., Zhao, Q.: Face illumination transfer through edge-preserving filters. In: IEEE CVPR, pp. 281–287 (2011)