# NO-REFERENCE IMAGE QUALITY ASSESSMENT FOR PHOTOGRAPHIC IMAGES OF CONSUMER DEVICE

*Yucheng Zhu, Guangtao Zhai, Ke Gu, and Zhaohui Che*

Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong University, Shanghai, China
{zyc420, zhaiguangtao, gukesjtuee, chezhaohui}@sjtu.edu.cn

## ABSTRACT

In this paper we study common, camera-specific kinds of distortions and propose a no-reference image quality assessment algorithm for photographic images produced by consumer devices. Those real consumer-type images, being different from simulated-distortion images, are with realistic artifacts and quality ranges. We find that the state-of-the-art no-reference image quality assessment approaches do not perform well on those photographic images, and propose an approach that achieves high prediction performance on a dataset of consumer-centric images. The proposed method, with no need for the original image, is able to reveal camera-specific problems and differentiate consumer cameras.

***Index Terms***— Consumer devices, camera-specific distortions, no-reference image quality assessment(NR IQA).

## 1. INTRODUCTION

With the development of the electronic technology, digital cameras entered people's life since last century. In 2013, the number of households in the United States owning a digital camera amounted to around 122.68 million, which accounted for 85 percent. Besides, with nearly everyone carrying smartphone in their pocket everywhere they go, it is convenient to take snaps anytime and anywhere. The arising question is how to objectively compare the quality of photographs produced by different digital devices, which is an important issue for consumers to determine which device to buy. And for manufacturers, it helps to better design the optical system and Image Signal Processor (ISP).

Different cameras with specific components and features will introduce different distortion combinations and thus affect consumers' decisions. These distortions are hard to simulate as they depend on the optics, the signal processor and the photographic content. Whether it is a mobile phone camera, a compact digital camera or a Digital Single Lens Reflex ( DSLR ), there are some basic component parts: the lense, shutter, aperture, flash, sensor, and ISP. And the features usually used for comparisons are: resolution and sensor's size, color reproduction, noise and distortion, optical and digital zoom, shutter delay, manual and auto focusing, image stabilisation, ISO, and white balance. A change in one component or feature will make difference. For example, some cameras have the anti-alias filter in the optical system which compromises on image quality, trading fine details for a lower risk of moire. Fig. 1 shows excellent examples of the resultant images with and without the aliasing filter. What's more, the metering determines the exposure accordingly and wrong evaluation will make some images come out too bright or too dark. Insufficient white balancing will cause global color errors such as a green overcast in the final image. Some mobile phone cameras, without IR filters, will identify the infrared as magenta. Considerable colored digital noise will appear in dim light shot, etc. These images with the listed kinds of distortions can reveal camera-specific problems and differentiate consumer cameras. However, these more practical and complex distortions cannot be unified into the domain of common

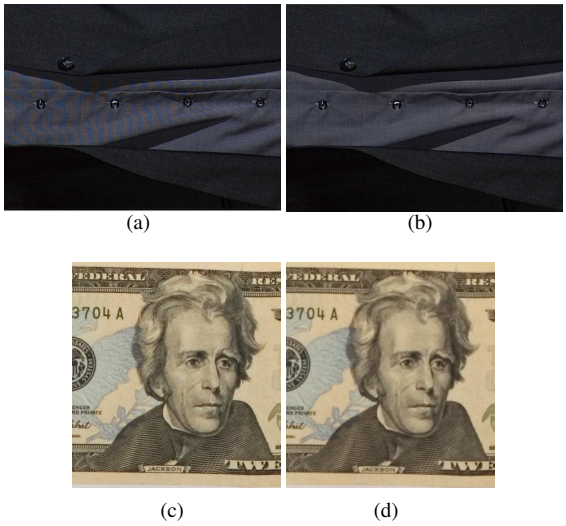
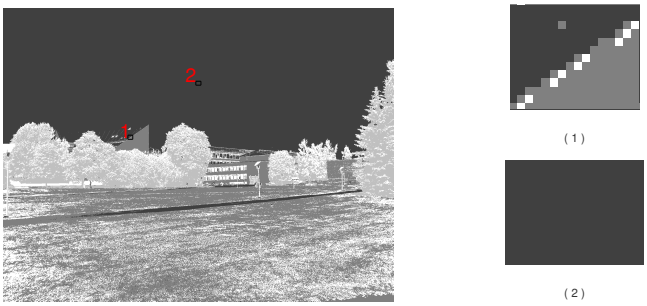



**Fig. 1**: Examples with and without the anti-alias filter: (a) is without the filter and with moire, (b) is with the filter and without moire. (c) shows more fine details without the filter, and (d) shows the softness resulted by the filter.

This work was supported in part by the National Science Foundation of China under Grants 61422112, 61371146, 61521062, 61527804, the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201339 and National High-tech R&D Program of China under Grant 2015AA015905.


(a) Photograph of good quality (b) Local region

**Fig. 2**: Example of sharpness: ( a ) is one photograph of good quality. ( b ) shows two chosen regions in ( a ), region ( 1 ) has obvious edges, region ( 2 ) has no edges or textures, but both of them are good.




(a) Division map (b) Local region

**Fig. 3**: Output division map: ( a ) is the result of classification which shows the four main color clusters in the whole image, different grey levels represent different clusters. ( b ) shows two chosen regions, pixels in region ( 1 ) cross three clusters, all pixels in region ( 2 ) are in the same cluster.

distortions like white noise, gaussian blur, JPEG, etc. We call them camera-specific distortions in this paper.

The largest number of objective image quality assessment metrics are full-reference (FR) methods [1], [2], [3], [4], which assume that the original image signal is completely known. Those metrics are usually of high performance. However, the dependence of original images severely reduces its applicability, and FR IQA methods measure the fidelity to the reference which will take image enhancement as "distortions". To solve these problems, many blind quality measures have been developed during the last decade [5], [6], [7], [8]. Those NR IQA methods are applicable in many more practical scenarios and show high performance on image databases [9]. However, when measuring the above-mentioned camera-specific distortions, these NR IQA methods do not perform well. Those photographic images are of complex distortions and have a narrow range of quality. The correlations between predicted and subjective scores might be high when predicting images of single distortion and a wide range of quality, but the correlations are much lower when predicting photographic photos. [10] introduces an approach with training to solve the problem. Here we propose an effective method without training to solve the problem at hand.

Estimation of sharpness is an important part in our model. Photographs produced by low-resolution cameras will inevitably lack details. Improper white-balance or unsound filters will make images look like being covered by haze. Inaccurate metering will make images too bright or too dark. All of those will decrease sharpness. Measure of sharpness is important in photographic images. However, according to Segur [11], it is common for photography to have some areas with few textures and edges. These areas are not always distorted ones but are the actual appearance of scenes, e.g., region ( 2 ) in Fig. 2. To be more precise, we take the color information into consideration to give the pure-color area an impartial treatment. We classify pixels into several clusters by their color information, the number of clusters is designed to automatically change depending on the image content. We evaluate sharpness of local blocks and finally predict the overall accurate sharpness of images. By the observation of abundant noise appearing when there is dim light in the scene, we take the evaluation of noise among dark photographs as another important part of our model. We conduct experiments on CID2013 [12] database with real distortions to validate the proposed model.

The remainder of this article is organized as follows. Section 2 presents the proposed photographic image quality assessment method. In Section 3, the effectiveness of our algorithm is proved by comparison of its experimental results with those obtained through existing relevant models. Finally, several concluding remarks are presented in Section 4.

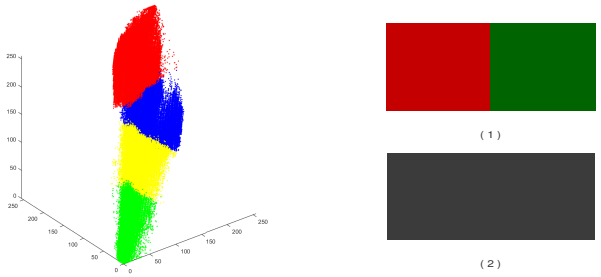
## 2. FRAMEWORK FOR NO-REFERENCE PHOTOGRAPH QUALITY ASSESSMENT

**Table 1**: Self-adaptive classification considering color information via kmeans.

| |
|---|
| Image classification ( input image $\mathbf{D}$, threshold $T$, output division map $M$, number of clusters to be classified $N$, cluster centroid locations $\mathbf{C}$. ) |
| 1. Let $N = 2$ be the initial number of clusters to be classified. |
| 2. Let $\mathbf{D}$ be the input of kmeans function. Kmeans partition the $H - by - 3$ data matrix into $N$ clusters where $H$ is the number of pixels and 3 represents the RGB color information. Each cluster contains pixels with similar color information. Calculate the proportion $P$ of the smallest cluster. |
| 3. Add one to the number of clusters to be classified $N$. |
| 4. Iterate between (2) and (3) until $P < T$. |
| 5. Return $M$, $N$, $\mathbf{C}$. |

### 2.1. Sharpness measure

Taking the measure of sharpness first, we choose to use color information partly because the grey-scale map is somewhat




(a) (b)

**Fig. 4**: Color map: ( a ) is the classification of pixels of the sample image in RGB color space. ( b ) shows two rectangles whose color maps are different but the grey-scale maps are the same.

inaccurate to evaluate sharpness. In Fig. 4 ( b ), two rects have different colors but their grey-scale maps are the same. What's more, color information is important in photography. In our model, we classify pixels into several clusters against color information, the classification will be a guide to further sharpness evaluations. Table 1 gives the detailed operations to make the process be self-adaptive.

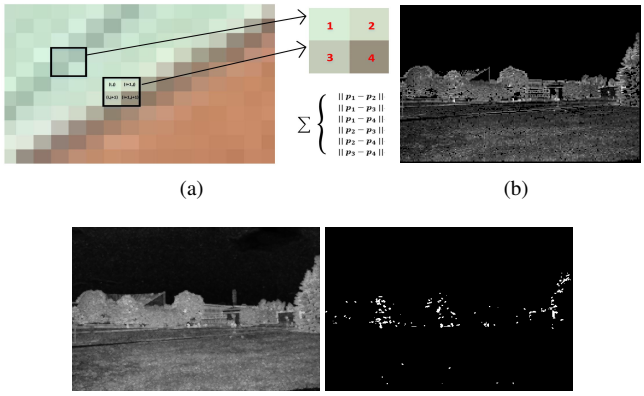
Fig. 3 is the result of color classification for Fig. 2. The input image is classified into four parts according to the main color information as the division map shows. The sky and cloud are classified into one cluster because of threshold, if we set a smaller value of threshold, the sky and cloud will be separated. Fig. 4 ( a ) shows the distribution and classification of pixels in RGB color space.

In the original RGB image we take the spatial-based measure of sharpness after each pixel has been classified into its own cluster. In a $m \times n$ sized local patch, we count the number of clusters. If there existing only one cluster like region ( 2 ) in Fig. 3, we will switch to next patch. In the multi-cluster patch $X$, we calculate in each smaller block of $X$ via

$$S(X) = \max_{\gamma_n \in X} \sum_{i,j \in \gamma_n} ||\mathbf{p_i} - \mathbf{p_j}||, \quad n = 1, 2, 3...M. \quad (1)$$

where $\gamma_n$ is one of smaller $2 \times 2$ sized blocks in $X$, $\mathbf{p}$ is the pixel with RGB information in $\gamma$. An example of the spatial measure of sharpness is shown in Fig. 5 ( a )

The max operator used in Eq. (1) and the threshold $T$ in Table 1 attempt to eliminate the negative effect of pixels near cluster borders. Euclidean distance of pixels near different color cluster centroid locations help to predict the sharpness, but the pixels near the border between two clusters will make the prediction inaccurate. When $T$ is too small, the number of clusters will increase which will intuitively bring about more borders and degrade the performance. When $T$ is too large, many pixels carrying useful information will be cut down. So we should set an appropriate value of threshold. Besides, the max operator ignores the pixels of small distance which reduces the influence of pixels near borders. And when the size of $X$ is appropriate, the max operator does not cut down useful pixels.




(a) (b)

(c) (d)

**Fig. 5**: ( a ) is the block diagram of the spatial measure of sharpness, ( b ) is the sharpness map in which brighter blocks denote greater perceived sharpness, ( c ) is the sharpness map generated by S3, and ( d ) shows the top one percent areas marked by white for calculation in S3.

We estimate the overall sharpness of image via

$$S = \frac{1}{N} \sum_{i=1}^{N} S(X_i) \quad (2)$$

where $N$ is the number of multi-cluster patches in image. Fig. 5 ( b ) is the resulting sharpness map, the single-cluster patches appearing black are assigned zero, those patches of abundant edges or textures like trees, grass, and pedestrian are marked by light gray as we expect. Fig. 5 ( c ) and ( d ) are results generated by S3 algorithm [7], the top one percent areas in ( d ) marked by white are calculated to represent the overall sharpness in S3, and FISH_bb [6] also use the same way to take the top one percent values into account which we think is inappropriate. Unlike these images with simulated distortions, the photographic image will be processed by ISP before its final display. The ISP shall conduct in-camera sharpening which enhances the sharpness of images by emphasizing the transitions between light and dark areas with halos, e.g., black areas in Fig. 5 ( a ). If only the top one percent areas are calculated, we will actually measure the halos and avoid most textures.

### 2.2. Compound noise estimation

Among those images of low luminance, single sharpness measure is not enough because of the emerging noise. Noise is an unfortunate byproduct of digital sensors. Abundant noise appears when there is dim light in the scene. When the luminance of image is small, we will measure the noise, where the luminance is estimated as the mean intensity. Realistic noises here are more complex than the mathematically convenient additive Gaussian model and we refer to the compound noise model in [13] to solve the problem. It is assumed

**Table 2**: PLCC and SROCC correlations between predicted and subjective MOS scores on the CID2013 database.

| Algorithm | PLCC | SROCC |
|---|---|---|
| Proposed (with noise estimation) | 0.818 | 0.793 |
| Proposed (without noise estimation) | 0.804 | 0.785 |
| FISH_bb [6] | 0.745 | 0.728 |
| S3 [7] | 0.725 | 0.708 |
| FISH [6] | 0.691 | 0.674 |
| SISBLIM [14] | 0.690 | 0.643 |
| NFERM [8] | 0.638 | 0.614 |
| FEDM [15] | 0.541 | 0.508 |
| ARISMC [16] | 0.497 | 0.429 |
| BRISQUE [5] | 0.476 | 0.443 |

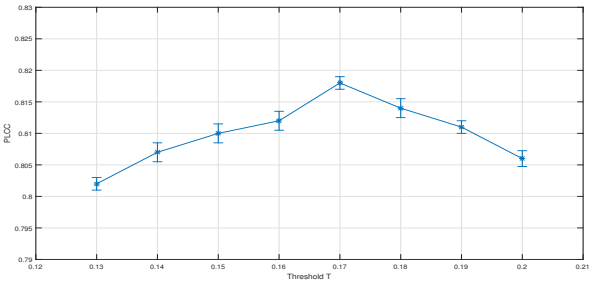



**Fig. 6**: Prediction accuracy of the proposed algorithm for the images in the CID2013 database with different values of threshold T. The horizontal axis is the threshold. The vertical axis is PLCC correlations. The error bars show the float of kmeans among 100 times.

in [13] that the kurtosis values tend to be invariant across scales for a natural image, and the scale invariance will be deteriorated by the added noise. For an input image signal $x$, the kurtosis of its noisy version $y$ can be expressed via

$$\mathbf{K}(y) = \left(\frac{\sigma^2(y) - \sigma^2(x)}{\sigma^2(y)}\right)^2 \mathbf{K}(x) + \left(\frac{\sigma^2(n)}{\sigma^2(y)}\right)^2 \mathbf{K}(n) \quad (3)$$

where $\mathbf{K}(x)$, $\mathbf{K}(y)$ and $\mathbf{K}(n)$ are the kurtosis values of $x$, $y$ and the noise $n$. $\sigma(x)$ , $\sigma(y)$ and $\sigma(n)$ are the standard deviation of $x$, $y$ and $n$. We change the form into compound noise. An additive and multiplicative compound noise model for noises of CCD and CMOS sensors was suggested

$$Y = X + (s_1 + s_2 X)N, \quad N \sim \mathcal{N}(0,1) \quad (4)$$

where $X$ and $Y$ are the original and noisy images, and $s_1$ and $s_2$ are the additive and multiplicative parameters. By experimental results we empirically set $K(n) = 1$ in (3). Therefore, the variance of noise $\hat{\sigma}_n^2$ can be estimated by minimizing:

$$\hat{\mathbf{K}}(x), \hat{\sigma}^2(n) = \arg \min_{\mathbf{K}(x), \sigma^2(n)} \quad (5)$$

$$\sum_{i \in \alpha} \left\| \hat{\mathbf{K}}(y_i) - \left(\frac{\hat{\sigma}^2(y_i) - \sigma^2(n)}{\hat{\sigma}^2(y_i)}\right)^2 \mathbf{K}(x) + \left(\frac{\sigma^2(n)}{\hat{\sigma}^2(y_i)}\right)^2 \right\|$$

where $\alpha$ denotes the set of selected frequency indexes, and $(\frac{\sigma^2(n)}{\hat{\sigma}^2(y_i)})^2$ reflects the influence of the multiplicative term of the noise.

We estimate the luminance of image, for those images of low luminance we take the noise level estimation and use term $S_n = S - \beta \times \hat{\sigma}(n)$ as the overall prediction where $\beta$ is derived from the observation of experimental results.

## 3. EXPERIMENTS AND ANALYSIS

We use CID2013 [12] to validate the effectiveness of our method. This database includes six datasets with their respective subjective mean-opinion-scores (MOS). Every dataset includes six different scenes derived from strictly defined clusters. Every scene was captured by 12-14 different cameras. This database is able to differentiate consumer cameras, reveal camera-specific problems and represent views that typical consumer camera users might capture with their cameras. So the faithful prediction will provide manufacturers and consumers with objective guide.

Person linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC) are used to evaluate performance of our approach. PLCC can be considered as a measure of prediction accuracy, while SROCC measures the monotonicity by ignoring the relative distance between the data. The higher SROCC and PLCC values indicate better performance in terms of correlation with human opinion.

We repeat our algorithm 100 times and report the median result of the performance across these 100 iterations because the results of kmeans change little each time. We compare our algorithm against the recent no-reference IQA approaches FISH_bb [6], S3 [7], FISH [6], SSIBLIM [14], NFERM [8], FEDM [15], ARISMC [16],BRISQUE [5]. Table 2 shows that our method outperforms the state-of-art NR IQA. Model with noise estimation performs better because it improves the accuracy of predicting low-quality images within the database. Experimental results in Fig. 6 prove that too large or small thresholds will decrease the prediction accuracy, but the calculated PLCC correlations are not very sensitive to T as the figure shows. It is deserved to be mentioned that our approach with no need for training is easy to use with.

## 4. CONCLUSION

In this paper we propose a no-reference image quality assessment algorithm for photographic images with realistic, camera-specific distortions and quality ranges produced by consumer devices. The proposed algorithm solves the pure-color problem and dim-light noise that are common in photography. It is an approach with no need for training and the information of the original image which is easy to use with. However, we do not solve all the introduced camera-specific distortions. Moire is such a distortion we can find in CID2013 but it's not included in current framework, which will be studied on in the future.

## 5. REFERENCES


[1] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[2] Hamid Rahim Sheikh and Alan C Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.

[3] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang, "An efficient color image quality metric with local-tuned-global model," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 506–510.

[4] Guangtao Zhai, Wenjun Zhang, Xiaokang Yang, Susu Yao, and Yi Xu, "Ges: a new image quality assessment metric based on energy features in gabor transform domain," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*. IEEE, 2006, pp. 4–pp.

[5] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 12, pp. 4695–4708, 2012.

[6] Phong V Vu and Damon M Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation," *Signal Processing Letters, IEEE*, vol. 19, no. 7, pp. 423–426, 2012.

[7] Cuong T Vu, Thien D Phan, and Damon M Chandler, "S3: A spectral and spatial measure of local perceived sharpness in natural images," *Image Processing, IEEE Transactions on*, vol. 21, no. 3, pp. 934–945, 2012.

[8] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang, "Using free energy principle for blind image quality assessment," *Multimedia, IEEE Transactions on*, vol. 17, no. 1, pp. 50–63, 2015.

[9] Hamid R Sheikh, Zhou Wang, Lawrence Cormack, and Alan C Bovik, "Live image quality assessment database release 2," 2005.

[10] Michele Saad, Philip Corriveau, Ramesh Jaladi, et al., "Objective consumer device photo quality evaluation," *Signal Processing Letters, IEEE*, vol. 22, no. 10, pp. 1516–1520, 2015.

[11] Risë Segur, "Using photographic space to improve the evaluation of consumer cameras," in *IS AND TS PICS CONFERENCE*. SOCIETY FOR IMAGING SCIENCE & TECHNOLOGY, 2000, pp. 221–224.

[12] Tuomas Virtanen, Mikko Nuutinen, Mikko Vaahteranoksa, Pirkko Oittinen, and Jukka Hakkinen, "Cid2013: a database for evaluating no-reference image quality assessment algorithms," *Image Processing, IEEE Transactions on*, vol. 24, no. 1, pp. 390–402, 2015.

[13] Guangtao Zhai and Xiaolin Wu, "Noise estimation using statistics of natural images," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 1857–1860.

[14] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang, "Hybrid no-reference quality metric for singly and multiply distorted images," *Broadcasting, IEEE Transactions on*, vol. 60, no. 3, pp. 555–567, 2014.

[15] Guangtao Zhai, Xiaolin Wu, Xiaokang Yang, Weisi Lin, and Wenjun Zhang, "A psychovisual quality metric in free-energy principle," *Image Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 41–52, 2012.

[16] Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang, "No-reference image sharpness assessment in autoregressive parameter space," *Image Processing, IEEE Transactions on*, vol. 24, no. 10, pp. 3218–3231, 2015.