LETTER

Color-Enriched Gradient Similarity for Retouched Image Quality Evaluation

Leida LI[†], ***Member***, **Yu ZHOU**[†], **Jinjian WU**[††], **Jiansheng QIAN**[†a)], ***and*** **Beijing CHEN**[†††], ***Nonmembers***

SUMMARY Image retouching is fundamental in photography, which is widely used to improve the perceptual quality of a low-quality image. Traditional image quality metrics are designed for *degraded* images, so they are limited in evaluating the quality of retouched images. This letter presents a RETouched Image QUality Evaluation (RETIQUE) algorithm by measuring structure and color changes between the original and retouched images. Structure changes are measured by gradient similarity. Color colorfulness and saturation are utilized to measure color changes. The overall quality score of a retouched image is computed as the linear combination of gradient similarity and color similarity. The performance of RETIQUE is evaluated on a public Digitally Retouched Image Quality (DRIQ) database. Experimental results demonstrate that the proposed metric outperforms the state-of-the-arts.

key words: *image quality assessment, image retouching, gradient similarity, color colorfulness, saturation*

1. Introduction

Image retouching is a fundamental operation in image editing and digital photography, which is usually used to improve the perceptual quality of a low-quality image. Many commercial softwares are now available for image retouching, among which is the most popular Adobe Photoshop. Accompanying the extensive use of image retouching is the lack of corresponding quality metrics for retouched images. The current image quality metrics are mainly designed for *degraded* images [1], [2], so they have limited capacities for evaluating the quality of retouched images. The quality assessment of retouched images is still an open problem.

Thus far, very few work has been done on the quality evaluation of retouched images [3], [4]. Vu *et al.* [3] proposed to measure contrast, sharpness and saturation changes between the original and retouched images. For contrast, both images were first converted into gray scale and divided into 8×8 patches with 50% overlap. Then the L2-norm difference between the root-mean-square luminance contrasts was computed as the contrast score. For sharpness, the spectral and spatial measure of local perceived sharpness (S3) maps [5] were adopted to generate the sharpness score. For saturation, they directly used the S component of HSV color space. The final quality score of a retouched image was obtained by pooling the three components. In [4], the same authors further improved this model by integrating it into the Most Apparent Distortion (MAD) [6] metric, which is a state-of-the-art image quality metric for degraded images. Although noticeable advances have been achieved by these metrics in evaluating the quality of retouched images, their performances are far from ideal.

Fig. 1 An example of image retouching: left: original, right: retouched.

It has been widely acknowledged that structures are crucial for image quality assessment [7]. This also applies to the quality assessment of retouched images. Besides the structure changes, retouched images are also characterized by enhancement in color. Specifically, retouched images are usually more colorful and higher in color saturation. Figure 1 shows an example of image retouching. It is clearly observed that the quality of the retouched image is improved in terms of both structure and color. These properties hold for the general retouched images. With these considerations, this letter presents a new RETouched Image QUality Evaluation (RETIQUE) algorithm. The proposed method consists of two modules, i.e., a structure module and a color module. The structure module is based on gradient similarity, which can measure the structure changes between the original and retouched images. The color module is measured using color colorfulness and saturation, which can effectively measure the color enhancements in retouched images. The overall quality score is computed as the linear combination of structure, colorfulness and saturation. The performance of the proposed method is evaluated on a public Digitally Retouched Image Quality (DRIQ) database. The experimental results show that the proposed method is effective in evaluating the quality of retouched images, and it outperforms the state-of-the-arts.

Manuscript received September 24, 2015.
Manuscript publicized December 9, 2015.
[†]The authors are with School of Information and Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China.
[††]The author is with School of Electronic Engineering, Xidian University, Xi'an 710071, China.
[†††]The author is with School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China.
a) E-mail: iqaqian@gmail.com (Corresponding author)
DOI: 10.1587/transinf.2015EDL8204

2. Proposed Retouched Image Quality Metric

In practice, image retouching is usually achieved by simultaneously adjusting several properties of an image, typically contrast, sharpness and color. Since contrast and sharpness are closely related to the structures of an image, we classify the changes in retouched images into structure and color. Accordingly, the proposed metric consists of a structure module and a color module.

2.1 Structure Module

As a special kind of image quality assessment, structure is important for retouched image quality modeling. In this letter, structure changes are measured by gradient similarity. For the original and retouched images, they are first converted into gray scale, which are denoted by $\mathbf{I}_1$ and $\mathbf{I}_2$, respectively. For $\mathbf{I}_1$, the gradient map is computed as:

$$\mathbf{G}_1 = \frac{|\mathbf{G}_x| + |\mathbf{G}_y|}{2}, \tag{1}$$

$$\mathbf{G}_x = [-1\ 0\ 1] * \mathbf{I}_1, \quad \mathbf{G}_y = [-1\ 0\ 1]^T * \mathbf{I}_1, \tag{2}$$

where $*$ denotes the convolution, and "T" is the transpose. Similarly, the gradient of $\mathbf{I}_2$ is computed and denoted by $\mathbf{G}_2$.

The gradient similarity between $\mathbf{G}_1$ and $\mathbf{G}_2$ is defined as:

$$\mathbf{GS}(x, y) = \frac{2\mathbf{G}_1(x, y)\mathbf{G}_2(x, y) + c}{\mathbf{G}_1^2(x, y) + \mathbf{G}_2^2(x, y) + c}, \tag{3}$$

where $1 \leq x \leq M$, $1 \leq y \leq N$, $M \times N$ denotes the image size, c is a small constant to ensure numerical stability. Then the gradient similarity score is calculated as:

$$Q_G = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} \mathbf{GS}(x, y). \tag{4}$$

Figure 2 (b) and (e) show the gradient maps of an original image and the corresponding retouched image. It is clear that the gradient map of the retouched image is enhanced. It should be noted that gradient is effective in representing contrast and sharpness changes [8]–[10], which are common in image retouching.

2.2 Color Module

Besides structure, color is crucial in retouched image quality assessment. Retouched images are usually richer in color. In the proposed method, color colorfulness and saturation are considered in the color module.

For colorfulness, we adopt the Color Colorfulness Index (CCI) [11], which is based on an opponent color space:

$$\mathbf{RG} = \mathbf{R} - \mathbf{G}, \quad \mathbf{YB} = \frac{1}{2}(\mathbf{R} + \mathbf{G}) - \mathbf{B}, \tag{5}$$

where $\mathbf{R}$, $\mathbf{G}$ and $\mathbf{B}$ denote the red, green and blue components of an image. Then CCI is computed as:

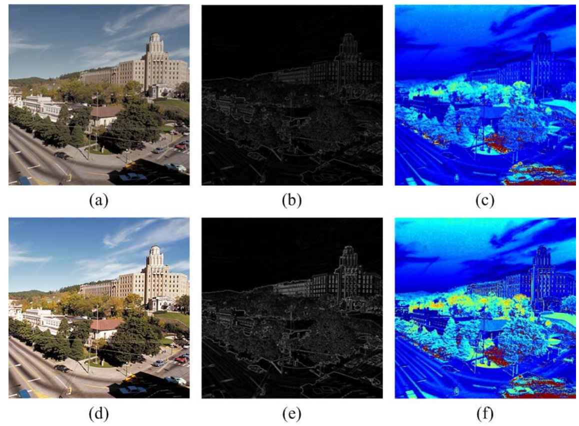

Fig. 2 Illustration of gradient, colorfulness and saturation between the original and retouched images. (a) Original image, CCI=0.0115; (b) gradient map of (a); (c) saturation map of (a); (d) retouched image, CCI=0.0257; (e) gradient map of (d); (f) saturation map of (d).

$$C = \sigma_{RGYB} + 0.3 \cdot \mu_{RGYB}, \tag{6}$$

with

$$\sigma_{RGYB} = (\sigma_{RG}^2 + \sigma_{YB}^2)/2, \ \ \mu_{RGYB} = \sqrt{\mu_{RG}^2 + \mu_{YB}^2}, \tag{7}$$

where σ_{RG}^2 and σ_{YB}^2 denote the variances of $\mathbf{RG}$ and $\mathbf{YB}$, μ_{RG} and μ_{YB} denote the corresponding mean values.

Let the CCI values of the original and retouched images be denoted by C_1 and C_2, then the similarity of colorfulness is computed as:

$$Q_C = \frac{2C_1C_2 + c}{C_1^2 + C_2^2 + c}. \tag{8}$$

Another property we considered in the color module is saturation. To compute the saturation similarity, both images are converted into HSI color space. Then the saturation components of them are denoted by $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively. The saturation similarity map is obtained by:

$$\mathbf{SS}(x, y) = \frac{2\mathbf{S}_1(x, y)\mathbf{S}_2(x, y) + c}{\mathbf{S}_1^2(x, y) + \mathbf{S}_2^2(x, y) + c}. \tag{9}$$

The saturation similarity score is computed as:

$$Q_S = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} \mathbf{SS}(x, y). \tag{10}$$

Figure 2 shows the saturation maps of the original and retouched images, as well as the CCI values. It is observed that both colorfulness and saturation are enhanced after retouching. Therefore, it is reasonable to include CCI and saturation in the proposed method.

2.3 Pooling

With the gradient similarity score Q_G, colorfulness similarity score Q_C and saturation similarity score Q_S, the final quality score of the retouched image is defined as follows:

Table 1 Performances of different image quality metrics in the DRIQ database.

Metric	PSNR	SSIM [7]	VIF [13]	MAD [6]	GSM [8]	FSIM [14]	GMSD [15]	Ref. [3]	Ref. [4]	GS	COLOR	RETIQUE
PLCC	0.2797	0.7029	0.6780	0.6977	0.7154	0.7740	0.7962	0.8498	0.8880	0.8191	0.8640	**0.9220**
SRCC	0.2515	0.6950	0.6679	0.6856	0.7123	0.7528	0.7762	0.8343	0.8681	0.7992	0.8568	**0.9145**
RMSE	1.9582	1.4507	1.4992	1.4612	1.4252	1.2915	1.2341	1.0751	0.9377	1.1701	1.0271	**0.7896**

$$Q = 1 - (\alpha \cdot Q_G + \beta \cdot Q_C + \gamma \cdot Q_S), \tag{11}$$

where α, β and γ are parameters that are used to adjust the relative importance of gradient, colorfulness and saturation. It is easy to know that the higher the Q value, the better the quality of the retouched image.

3. Experimental Results

3.1 Experimental Settings

The performance of the proposed method is evaluated in DRIQ database [12]. DRIQ consists of 26 original images and the corresponding 78 retouched images, which are obtained using Photoshop by adjusting contrast, sharpness, brightness, color, or combination of these properties of the original images [12]. Difference mean-opinion-score (DMOS) is used as the subjective score. For performance measurement, we adopt the three commonly used criterions, namely Spearman rank order correlation coefficient (SRCC), Pearson linear correlation coefficient (PLCC) and root mean squared error (RMSE). SRCC is used to measure prediction monotonicity, while PLCC and RMSE are used to measure prediction accuracy. Before computing these criterions, a four-parameter logistic mapping is conducted between the subjective and predicted scores:

$$f(x) = \frac{\lambda_1 - \lambda_2}{1 + e^{(x-\lambda_3)/\lambda_4}} + \lambda_2, \tag{12}$$

where $\lambda_i, i = 1, 2, 3, 4$, are the parameters to be fitted. In implementation, other parameters in the proposed method are set as follows: $\alpha = 0.4, \beta = 0.3, \gamma = 0.3, c = 0.0005$, which are determined by experiments.

3.2 Performance Evaluation

Figure 3 shows the scatter plots between the subjective scores and predicted scores by the proposed metric. It is observed that the sample points all gather closely around the fitting curve, which indicates that the predicted scores are very consistent with the subjective scores.

Table 1 summarizes the experimental results of different quality metrics in DRIQ database. For comparison, we test seven traditional image quality metrics and two retouched image quality metrics. The seven traditional metrics include PSNR, SSIM [7], VIF [13], MAD [6], GSM [8], FSIM [14] and GMSD [15], and the two retouched image quality metrics are Refs. [3] and [4]. In order to investigate the individual contributions of structure module and color module in the proposed method, they are further tested separately, which are denoted by GS and COLOR in Table 1.

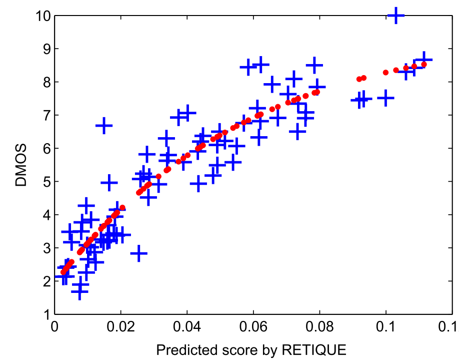

Fig. 3 Scatter plots of the subjective scores versus the predicted scores predicted by RETIQUE in DRIQ database.

It is observed from the table that traditional quality metrics are very limited in evaluating the quality of retouched images. The best results are delivered by GMSD [15], which are still below 0.8 for both PLCC and SRCC. The two retouched image quality metrics in [3] and [4] produce better results. By contrast, the proposed RETIQUE achieves the best performance, and both PLCC and SRCC are higher than 0.91. The individual performances of the proposed structure and color modules also produce very promising results. Specifically, the color module alone outperforms Ref. [3]. By combining these two modules, the performance improves significantly. This further demonstrates that structure and color are both indispensable for the quality evaluation of retouched images.

Figure 4 shows several retouched images with quite similar subjective scores, together with the predicted scores by RETIQUE. It is observed that the predicted scores are also similar. Furthermore, with the slight increase of subjective scores, the predicted scores also increase slightly. This indicates that the proposed method can distinguish tiny quality differences between images.

3.3 Improving Traditional Image Quality Metrics Using the Color Module

The color module in the proposed metric can be used to adapt the current image quality metrics to evaluate the quality of retouched images. To this end, we conduct an experiment to integrate the proposed color module into several traditional image quality metrics (based on Eq. (11)) and investigate whether their performances can be improved. The tested metrics indlude SSIM [7], VIF [13], MAD [6], GSM [8], FSIM [14] and GMSD [15]. The performances of the improved metrics are all tested in the DRIQ database. Table 2 summarizes their performances before and after incorporating the color module, together with a statistics of

(a) DMOS=3.1705, Q=0.0148 (b) DMOS=3.2050, Q=0.0159

(c) DMOS=3.3870, Q=0.0164 (d) DMOS=3.4335, Q=0.0179

Fig. 4 Images in DRIQ database together with their quality scores.

Table 2 Performances of trational image quality metrics in DRIQ database before and after incorporating the proposed color module (CCI+S), together with a statistics of the performance gains in percentage.

Metric	PLCC			SRCC		
	Before	After	Gain (%)	Before	After	Gain (%)
SSIM [7]	0.7029	0.8642	↑ 22.95	0.6950	0.8565	↑ 23.24
VIF [13]	0.6780	0.7915	↑ 16.74	0.6679	0.7807	↑ 16.89
MAD [6]	0.6977	0.8656	↑ 24.06	0.6856	0.8558	↑ 24.82
GSM [8]	0.7154	0.8641	↑ 20.79	0.7123	0.8561	↑ 20.19
FSIM [14]	0.7740	0.9009	↑ 16.40	0.7528	0.8922	↑ 18.52
GMSD [15]	0.7962	0.8654	↑ 8.69	0.7762	0.8568	↑ 10.38

the performance gains relative to the original ones.

It is observed from Table 2 that by incorporating the proposed color module, all tested metrics achieve a significant performance gain, in terms of both prediction accuracy (PLCC) and monotonicity (SRCC). This further demonstrates that the proposed color module is effective in evaluating the quality of retouched images.

4. Conclusion

In this letter, we have presented a quality metric for retouched images. The proposed metric consists of a structure module and a color module. As a special kind of image quality assessment, measuring structure change is essential. Meantime, color change is also important for retouched image quality assessment. In this work, structure change is measured using gradient and color change is measured based on color colorfulness and saturation. The experimental results based on a retouched image database have demonstrated that the proposed metric is effective in retouched image quality assessment. We have adopted the proposed color module to improve traditional image quality metrics with very promising results.

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities (2015XKQY18), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology (CICAEET).

References

[1] D.M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," ISRN Signal Process., vol.2013, Article ID 905685, pp.1–53, 2013.
[2] W.S. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," J. Vis. Commun. Image R., vol.22, no.4, pp.297–312, 2011.
[3] C. Vu, T. Phan, and D. Chandler, "Can current image quality assessment algorithms predict visual quality of enhanced images?," Online: http://vision.okstate.edu/driq/HVEI_VPC_mini_final.pdf
[4] C.T. Vu, T.D. Phan, P.S. Banga, and D.M. Chandler, "On the quality assessment of enhanced images: A database, analysis, and strategies for augmenting existing methods," Proc. IEEE Southwest Symposium on Image Analysis and Interpretation, pp.181–184, 2012.
[5] C.T. Vu, T.D. Phan, and D.M. Chandler, "S_3: A spectral and spatial measure of local perceived sharpness in natural images," IEEE Trans. Image Process., vol.21, no.3, pp.934–945, March 2012.
[6] E.C. Larson and D.M. Chandler, "Most apparent distortion: Full reference image quality assessment and the role of strategy," J. Electronic Imaging, vol.19, no.1, pp.001006:1-21, March 2010.
[7] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.
[8] A.M. Liu, W.S. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," IEEE Trans. Image Process., vol.21, no.4, pp.1500–1512, 2012.
[9] L. Zhang, Y. Shen, and H.Y. Li, "VSI: A visual saliency induced index for perceptual image quality assessment," IEEE Trans. Image Process., vol.23, no.10, pp.4270–4281, 2014.
[10] P. Zhao, L. Li, and H. Cai, "Saliency guided gradient similarity for fast perceptual blur assessment," IEICE Trans. Inf. & Syst., vol.E98-D, no.8, pp.1613–1616, 2015.
[11] S.S. David. Hasler, "Measuring colorfulcolor in real images," Proc. SPIE Human Vision and Electronic Imaging, pp.87–94, 2003.
[12] C. Vu, T. Phan, and D.M. Chandler, Digitally Retouched Image Quality database, Online: http://vision.okstate.edu/index.php?loc=driq
[13] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," IEEE Trans. Image Process., vol.15, no.2, pp.430–444, 2006.
[14] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process., vol.20, no.8, pp.2378–2386, 2011.
[15] W. Xue, L. Zhang, X. Mou, and A.C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," IEEE Trans. Image Process., vol.23, no.2, pp.684–695, Feb. 2014.