Using Online Manifold Learning for Color Image Quality Assessment

Meiling He[1], Mei Yu[1], Hua Shao[1], Hao Jiang[1]
Faculty of Information Science and Engineering[1]
Ningbo University
Ningbo, China
{hemeiling1991, yumei2, shaohua_nb, jhznjd}@126.com

Gangyi Jiang[1,2]
National Key Lab of Software New Technology[2]
Nanjing University
Nanjing, China
e-mail: jianggangyi@126.com

***Abstract*—The structure of the low-dimensional characteristics of images is manifold, which is precisely what the human visual system perceives. With this inspiration, a new Image Quality Assessment (IQA) metric called Online Manifold Learning based Quality (OMLQ) is proposed for color IQA in this paper. Online manifold learning is employed to construct a feature extraction matrix, which is used to obtain low-dimensional manifold vectors. In addition, visually important regions are detected to mimic the properties of the visual perception. The new IQA score is defined as the similarity of feature vectors between reference image and the corresponding distorted one. Extensive experiments performed on three publicly available benchmark databases demonstrate that the proposed IQA index OMLQ works better in terms of prediction accuracy than the other state-of-the-art indices.**

Keywords-color image quality assessment; visual saliency; human visual system; manifold learning.

I. INTRODUCTION

The quantitative evaluation of an image's perceptual quality is one of the most fundamental yet challenging problems in image processing system and vision research, confirmed by the idiom, "A picture is worth a thousand words" [1]. Objective image quality assessment is capable of approximating subjective opinion of an average human observer by employing an efficient computational model, which is suitable for different image content, different distortion types and different degree of distortion [2]. As conventional metrics, Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) are widely accepted due to their computational efficiency and definite physical meaning. However, MSE or PSNR do not correlate well with human beings' subjective scores with a variety of image content and distortion types involved since they do not consider the properties of Human Visual System (HVS) and just measure the pixel difference between reference and distorted image.

Structural Similarity (SSIM) [3] index based on the hypotheses that HVS is highly adapted for extracting structural information in images, can be considered a milestone of the development of Image Quality Assessment (IQA) models, it can provide a good prediction of the perceived quality score. In the following years, many SSIM extensions are proposed, such as Multi-Scale SSIM (MS-SSIM) [4], Complex Wavelet SSIM (CW-SSIM) [5], information weighted SSIM (IW-SSIM) [6], and so on. The research in [7] proposed a wavelet-based Visual Signal-to-Noise Ratio (VSNR) metric, which operates via a two-stage approach in the wavelet domain based on near-threshold and supra-threshold properties of human vision. Except for structural approaches, Sheikh et al. proposed the Visual Information Fidelity (VIF) index [8], which was an extension of its former version, namely the Information Fidelity Criterion (IFC) index [9]. VIF tries to quantify the amount of information shared between the reference image and the corresponding distorted one. Larson et al. asserted that the HVS performs different strategies for high-quality image and low-quality image. Inspired by this, they proposed a Most Apparent Distortion (MAD) model which shows remarkable and robust result [10]. In addition, a different IQA approach, based on Sparse Representation (SPARQ) index [2], is proposed for gray image. Most commonly used algorithms are just designed for gray image, but in RGB image graying process, there is part of information lost, resulting in inaccurate evaluation results.

For visual perception phenomenon, studies have shown that manifold is the basis of perception [11]. There exists massive redundancy in the high-dimensional digital image data, it is essential to reduce the dimension but still maintain essence of structure. Given a set of high-dimensional data points, manifold learning aims at discovering the nonlinear geometric properties embedded in high-dimensional data space of low-dimensional manifolds, which reflects the intrinsic nature of things. Deng et al. introduced a novel subspace learning algorithm, called Orthogonal Locality Preserving Projection (OLPP) [12], which can find the manifold structure of image. We can apply OLPP algorithm to the given image patches, mapping it to a low- dimensional manifold, so the feature extraction will be achieved.

Motivated by above consideration, this paper presents a novel IQA model for color image, called Online Manifold Learning based Quality (OMLQ). We use visual saliency (VS) model to strike a maximum combined saliency map and a maximum absolute difference map from RGB color space to detect visually important regions. OMLQ relates perceived quality of an image with the fidelity to the reference image in the form of manifold features that are extracted in the detected salient regions by a feature detector, i.e. feature extraction matrix obtained by online OLPP. Finally, the manifold features are used to predict an objective value.

 ISBN: 978-1-61208-641-5

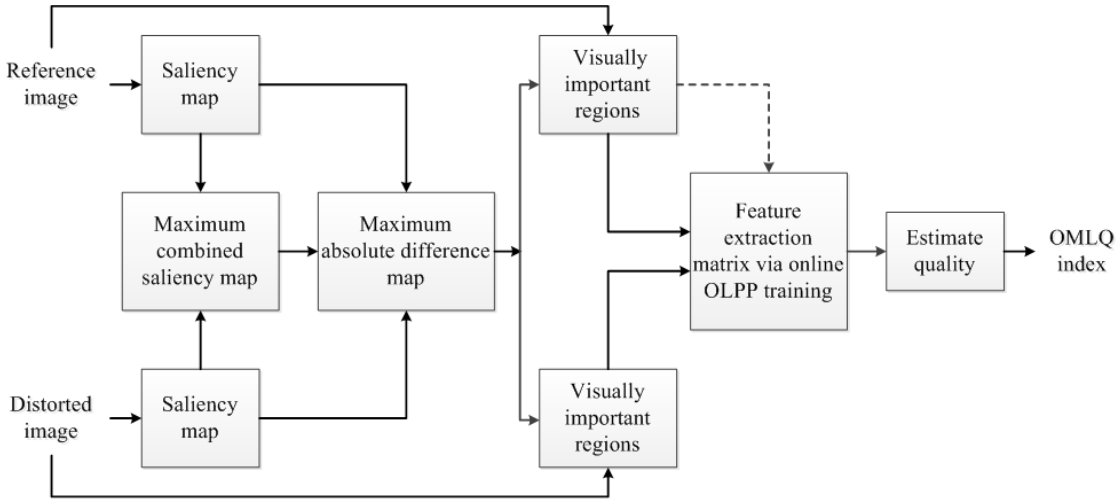

Figure 1. Illustration for the computational process of the proposed model OMLQ

II. THE PROPOSED APPROACH

Because HVS has the ability to capture nonlinear manifold structure, we propose a metric based on online manifold learning for color image quality assessment. The procedures to compute OMLQ are illustrated in Figure 1.

It is acknowledged that not every pixel in an image receives the same level of visual importance. The relationship between VS and IQA has been investigated by some researchers and it is broadly recognized that incorporating VS information appropriately can benefit IQA metrics. In this work, we experiment with the VS model in [13] to detect the visually important regions.

1) Let M^r and M^d denote the saliency maps pertaining to reference image I^r and distorted image I^d, computed by VS model mentioned above. A maximum combined saliency map $M^{\max}$ with the same size of M^r and M^d is created with $\max(M^r, M^d)$. It is well known that an average observer perceives the world in color instead of black and white, so we directly deal with the RGB image. Before the computation, the I^r, I^r, M^r, M^d, and $M^{\max}$ should be first divided into non- overlapping 8×8 patches and then each of them is vectorized and arranged by scanning the numerical values in columns, which forms the matrices: reference image patch matrix $\mathbf{X}^r$, distorted image patch matrix $\mathbf{X}^d$, reference saliency patch matrix $\mathbf{S}^r$, distorted saliency patch matrix $\mathbf{S}^d$, and the maximum combined saliency patch matrix $\mathbf{S}^{\max}$, respectively. Since a color image has three channels, the length of $\mathbf{X}^r$ and $\mathbf{X}^d$ is 8×8×3=192. The length of $\mathbf{S}^r$, $\mathbf{S}^d$, and $\mathbf{S}^{\max}$ is 8×8=64.

2) Let $\mathbf{S}_j^{\max}$ denote the value of the *j*th column in the maximum combined saliency patch matrix $\mathbf{S}^{\max}$, then the saliency of the *j*th patch of $M^{\max}$ is expressed by

$$d_j = \sum_{i=1}^{N} S_{ij}^{\max} \quad (1)$$

where N denotes the number of pixels in a patch, $\mathbf{S}_{ij}^{\max}$ denotes the value of the ith row and jth column in $\mathbf{S}^{\max}$. Suppose $t_1 = \lambda_1 \cdot k$, where λ_1 to the range (0,1] indicates the scale factor of the selected maximum combined saliency patch, t_1 represents the number of selected salient patches. and k denotes the number of patches extracted from each image. We select the first largest t_1 saliency value d_j corresponding to the reference image patch matrix $\mathbf{X}^{r*}$ and the distorted patch matrix $\mathbf{X}^{d*}$. Similarly we can get the corresponding reference saliency patch matrix $\mathbf{S}^{r*}$ and the distorted saliency patch matrix $\mathbf{S}^{d*}$.

Here, based on $\mathbf{S}^{r*}$ and $\mathbf{S}^{d*}$, the difference between a pair of vectors is measured by the mean absolute error, then the absolute difference value e_j is defined as

$$e_j = \frac{1}{n}\sum_{i=1}^{n} |S_{ij}^{r*} - S_{ij}^{d*}|, \quad j = 1,\mathrm{L}\ ,t_1 \quad (2)$$

Let $t_2 = \lambda_2 \cdot t_1$, where λ_2 is the scale factor of the selected patch and t_2 denotes the number of selected maximum absolute difference patches. We take the first t_2 largest e_j corresponding $\mathbf{X}^{r*}$ and $\mathbf{X}^{d*}$ as the final visually important regions, denoted by $\mathbf{Y}^r$ and $\mathbf{Y}^d$. The goal of feature extraction is the acquisition of a feature detector, which is applied to evaluate the image quality. Many researchers usually receive a learner through off-line mode, which requires a lot of training samples and has significant limitations in real-time applications. Therefore, we apply online learning adaptively updating feature detector. First, each sample vector is centered by subtracting the mean pixel value of each patch. All the sample vectors construct a matrix Y as online OLPP learning input. Next, following the intuition that the image

 ISBN: 978-1-61208-641-5

data may be generated by sampling a probability distribution that has support on or near a submanifold of ambient space, we apply OLPP algorithm to project the sample vectors into a subspace to obtain manifold features. The procedure of online OLPP learning is stated by.

1) **PCA Projection:** By throwing away dispensable components, we preserve the maximum amount of the sample vectors and discard redundant information after the matrix Y is projected into the PCA subspace. PCA can be done by eigenvalue decomposition of a covariance matrix. After the decomposition, let $\boldsymbol{\psi} = diag(\psi_1,\mathrm{L}\ ,\psi_M)$ and $\mathbf{E} = diag(e_1,\mathrm{L}\ ,e_M)$ indicate the M largest eigenvalues and the corresponding eigenvectors for the covariance matrix. In our work, M is fixed at 8. This means the dimension of each whitened vector will be reduced from 192 to M=8. The whitened matrix, **W**, is given by

$$\mathbf{W} = \boldsymbol{\Psi}^{-1/2} \times \mathbf{E}^{\mathrm{T}} \tag{3}$$

Eventually, sample data Y can be whitened into $\mathbf{Y}^w$ by the following implementation

$$\mathbf{Y}^{\mathrm{w}} = \mathbf{W} \times \mathbf{Y} \tag{4}$$

2) **Constructing the Adjacency Graph:** Let G represent a graph with m nodes. The a-th node corresponds to whitened sample data y_a^w. We connect them when node a and node b are adjacent, i.e., y_a^w is among k nearest neighbors of y_b^w.

3) **Choosing the Weights:** If node a and b are connected, set $S_{ab} = e^{-\left\|y_a^w - y_b^w\right\|^2}$, otherwise, set $S_{ab} = 0$. The weight matrix S of graph G exactly explains the local structure of image manifold.

4) **Computing the Orthogonal Basis Function:** We define Φ as a diagonal matrix, which is expressed by $\Phi_{aa} = \sum_{a=1}^{N} S_{ab}$. We also define Laplacian matrix L, i.e. $\mathbf{L} = \boldsymbol{\Phi} - \mathbf{S}$. Let $\{\mathbf{p}_1,\mathrm{L}\ ,\mathbf{p}_n\}$ be the orthogonal basis vectors, then

$$\mathbf{P}^{(n-1)} = [\mathbf{p}_1,\mathrm{L}\ ,\mathbf{p}_{n-1}] \tag{5}$$

$$Q^{(n-1)} = [\mathbf{P}^{(n-1)}]^T (\mathbf{Y}^w \boldsymbol{\Phi} \mathbf{Y}^{wT})^{-1} \mathbf{P}^{(n-1)} \tag{6}$$

The orthogonal basis vectors are computed as follows.

- Compute $\mathbf{p}_1$ by the eigenvector corresponding to the smallest eigenvalue of $(\bar{\mathbf{Y}}^w \boldsymbol{\Phi} \bar{\mathbf{Y}}^{wT})^{-1} \bar{\mathbf{Y}}^w \mathbf{L} \bar{\mathbf{Y}}^{wT}$.
- Compute $\mathbf{p}_n$ by the eigenvector corresponding to the smallest eigenvalue of $\mathbf{M}^{(n)}$

$$\mathbf{M}^{(n)} = \{\mathbf{I} - (\mathbf{Y}^w \boldsymbol{\Phi} \mathbf{Y}^{wT})^{-1} \mathbf{P}^{(n-1)} [\mathbf{Q}^{(n-1)}]^{-1} [\mathbf{P}^{(n-1)}]^T\} \cdot (\mathbf{Y}^w \boldsymbol{\Phi} \mathbf{Y}^{wT})^{-1} \mathbf{Y}^w \mathbf{L} \mathbf{Y}^{wT}$$

5) **Feature Extraction Matrix by OLPP Embedding:** Suppose the best projection matrix $\mathbf{J}_{OLPP} = [\mathbf{p}_1,...,\mathbf{p}_l]$. After the learning process, the feature is transformed from the whitened space to original space by

$$\mathbf{D} = \mathbf{W} \times \mathbf{J}_{OLPP} \tag{7}$$

where D is the feature extraction matrix through online OLPP learning, which is used to extract image features that capture intrinsic manifold structure in an image.

After the online OLPP learning step, the manifold feature vectors, $\mathbf{u}_i$ and $\mathbf{v}_i$, can be extracted by a multiplication operation.

$$\mathbf{u}_i = \mathbf{D} \times \mathbf{y}_i^r, \mathbf{v}_i = \mathbf{D} \times \mathbf{y}_i^d \tag{8}$$

Since the size of D is 8×192, the length of $\mathbf{u}i$ and $\mathbf{v}i$ is 8. For simplicity, we use a vector pair to represent the features of a reference patch together with its distorted patch. Therefore, all of the feature vectors of $\mathbf{Y}^r$ and $\mathbf{Y}^d$ are concatenated to form two matrices, U and V, respectively.

Finally, we defined perceived quality score as the feature similarity by

$$Score = \frac{1}{K \bullet M} \sum_{i=1}^{K} \sum_{j=1}^{M} \frac{2\mathrm{U}_{ij}\mathrm{V}_{ij} + C}{(\mathrm{U}_{ij})^2 + (\mathrm{V}_{ij})^2 + C} \tag{9}$$

where K denotes the number of image patches in visually important region, i.e., the number of manifold features is reserved, M represents the dimension of manifold features. C is a positive constant that supplies numerical stability.

TABLE I. PERFORMANCE COMPARISON UNDER DIFFERENT TYPES OF DISTORTION ON LIVE DATABASE

	JP2K	JPEG	WN	GB	FF	ALL
SROCC	0.9558	0.9724	0.9574	0.9473	0.9514	0.9523
PLCC	0.9524	0.9709	0.9645	0.9492	0.9433	0.9506
RMSE	8.4314	7.5546	5.7865	5.8132	8.9312	8.4433

TABLE II. PERFORMANCE COMPARISON FOR SEVEN IQA METRICS ON THREE TEST DATABASES

		PSNR	SSIM	IFC	VIF	VSNR	SPARQ	OMLQ
SROCC	LIVE	0.8756	0.9479	0.9259	**0.9636**	0.9274	0.9310	0.9523
	CSIQ	0.8057	0.8756	0.7671	0.9195	0.8106	0.9460	**0.9465**
	TID	0.5531	0.7749	0.5675	0.7491	0.7046	0.7920	**0.8356**
PLCC	LIVE	0.8723	0.9449	0.9268	**0.9604**	0.9231	0.9280	0.9506
	CSIQ	0.8000	0.8613	0.8384	0.9277	0.8002	0.9390	**0.9433**
	TID	0.5734	0.7732	0.7340	0.8084	0.6820	0.8200	**0.8228**
RMSE	LIVE	13.3600	8.9455	10.2641	**7.6137**	10.5060	10.1850	8.4433
	CSIQ	0.1575	0.1344	0.1431	0.0980	0.1575	0.0900	**0.0871**
	TID	1.0994	0.8511	0.9113	0.7899	0.9815	0.7680	**0.5975**

 ISBN: 978-1-61208-641-5

III. EXPERIMENTAL RESULTS

Three publicly benchmark databases including LIVE [14], CSIQ [15] and TID2008 [16] are involved. Each database consists of hundreds of degraded images with Mean Opinion Score (MOS) or Differential Mean Opinion Score (DMOS). It is customary to nonlinearly map the metric scores to the ones that have a linear relationship with the subjective scores. Three commonly used performance metrics, including Spearman Rank Order Correlation Coefficient (SROCC), Pearson Linear Correlation Coefficient (PLCC), and Root Mean Squared Error (RMSE) are adopted to evaluate the IQA model. As mentioned previously, there are three parameters, i.e., the scale factors λ_1 and λ_2 for the detection of visually important regions, and stability parameter C for feature similarity, which are determined by training. The training set consists of all images from LIVE database. For each λ_1 and λ_2, which are changed from 0.3 to 0.8 in step of 0.1, and for C, which is changed from 0.01 to 0.1 in step of 0.01, the best values are found by maximizing the SROCC value of OMLQ metric on the training set. When λ_1=0.7 , λ_2=0.6 , and $C = 0.05$ are used, SROCC reaches the peak by performance tuning. To validate the performance of OMLQ on different distortion types, the individual experimental results on LIVE database are summarized in Table I. For each distortion type, we can see this metric has good performance whether it is an individual distortion or a crossover distortion test.

We have evaluated the performance of the proposed metric with other six IQA metrics: PSNR, SSIM [3], IFC [8], VIF [7], VSNR [6] and SPARQ [2]. Table II lists the performance indicators on the three databases, where the best value across the seven IQA results is highlighted in boldface. It is clear that our metric outperforms other IQA metrics in CSIQ and TID2008 databases. Although the result in LIVE database is inferior to VIF, its SROCC value and PLCC value have already been reached 0.95, which means the proposed metric can accurately predict the perceptual image quality.

IV. CONCLUSION

In this paper, we proposed a novel metric for color IQA. It is based on the assumption that an image's VS map has a close relationship with its perceptual quality and online learning can exploit the low-dimensional manifold embedded in high-dimensional data. Our contribution in this work is that we apply manifold learning to IQA. The proposed OMLQ was thoroughly tested and compared with six state-of-the-art IQA indices on three publicly benchmark databases. The results demonstrated that OMLQ could yield much better results in terms of prediction accuracy than all competing methods.

REFERENCES

[1] S. C. Pei and L. H. Chen, "Image quality assessment using human visual DOG model fused with random forest," IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 3282-3292, Nov. 2015.

[2] T. Guha, E. Nezhadarya, and R. K. Ward, "Sparse representation-based image quality assessment," Signal Processing: Image Communication, vol. 29, no. 10, pp. 1138-1148, Nov. 2014.

[3] Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp.600-612, Apr. 2004.

[4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-Scale Structural Similarity for Image Quality Assessment," the Thirty-Seventh Asilomar Conference on Signals, Systems, and Computers, vol. 2, pp. 1398-1402. Nov. 2003.

[5] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: a new image similarity index," IEEE Transactions on Image Processing, vol. 18, no. 11, pp. 2385-2401, Nov. 2009.

[6] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Transactions on Image Processing, vol. 20, no.5, pp. 1185-1198, May 2011.

[7] D. M. Chandler and S S Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," IEEE Transactions on Image Processing, vol. 16, no. 9, pp. 2284 – 2298, Sep. 2007.

[8] H. R. Seheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," IEEE Trans. on Image Processing, vol. 14, no. 12, pp. 2117-2128, Dec. 2005.

[9] H. R. Seheikh and A. C. Bovik, "Image information and visual quality," IEEE Transactions on Image Processing, vol. 15, no. 2, pp. 430-444, Feb. 2006.

[10] E.C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," Journal of Electronic Imaging, vol. 19, no. 1, pp. 143-153, Jan. 2010.

[11] H.S. Seung, "The manifold ways of perception," Science, vol. 290, no. 5500, pp. 2268-2269, 2000.

[12] C. Deng, X. He, J. Han, and H.-J. Zhang, "Orthogonal Laplacian faces for face recognition," IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3608-3614, Nov. 2006.

[13] L. Zhang, Z. Gu, and H. Li, "SDSP: A novel saliency detection method by combining simple priors," IEEE International Conference on Image Processing (ICIP), Melbourne, VIC, pp. 171–175, Sep. 2013.

[14] H. R. Seheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.

[15] L. Zhang, Y. Shen, and H. Li, "VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment," IEEE Transactions on Image Processing, vol. 23, no. 10, pp. 4270-4281, Aug. 2014.

[16] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli, and Federica Battisti, "TID2008 - A database for evaluation of full-reference visual quality assessment metrics," Advances of Modern Radio electronics, vol. 10, no. 4, pp. 30-45, 2009.

 ISBN: 978-1-61208-641-5