Image Quality Assessment Using Sparse Representation in ICA Domain

Cheng Cheng[*] and Hanli Wang[*,+]
[*]Key Laboratory of Embedded System and Service Computing, Ministry of Education,
Tongji University, Shanghai 200092.
E-mail: chengcheng_lcc@163.com, hanliwang@tongji.edu.cn

***Abstract*— A novel metric for full-reference image quality assessment (IQA) is proposed in this paper. Based on the sparse representation in independent component analysis (ICA) domain, the image basis is generated from natural images adaptively, which coincides with the characteristics of human vision system (HVS). In order to extract the feature vector, a hybrid norm optimization strategy is introduced for achieving more stable computational performances. The proposed IQA metric is calculated as a correlation coefficient between the two feature vectors from reference and distorted images, respectively. Experimental results on the LIVE Database Release 2 demonstrate that the proposed metric can achieve competitive performances as compared to the well-known structural similarity (SSIM) metric.**

I. Introduction

Image quality assessment (IQA) has attracted much attention in the last decades and is becoming an important issue in many applications such as image acquisition, compression, transmission, and restoration [1]. There are generally two kinds of IQA methods: subjective assessment by humans and objective assessment by algorithms. The first one, such as the differential mean opinion score (DMOS), is time-consuming and can't be applied in real-time systems. Conversely, the objective assessment aims to provide computational models to measure the perceptual quality of an image which corresponds to the subjective assessment well. According to the availability of a reference image, the objective assessment is classified into three categories: full-reference (FR), no-reference (NR), and reduced-reference (RR) methods. In this work, we focus our study on FR IQA only.

Conventional FR IQA metrics, e.g., mean square error (MSE) and peak signal-to-noise ratio (PSNR), have been widely used due to their mathematical simplicity, but they are not in agreement with the perceived quality which reflects the mechanism of human visual system (HVS). Recent advances have resulted in the emergence of several powerful perceptual distortion measures that outperform the MSE and its variants. The structural similarity (SSIM) [2] metric is applied to measure image quality by capturing the similarity of images, which is assumed that natural images are highly structured and HVS can adaptively extract such structural information easily. In SSIM, a product of three factors of similarity, including luminance, contrast and structure, is considered to formulate the SSIM metric. Its variant, multi-scale structural similarity (MS-SSIM) index [3], which involves structural similarity information from different scales, provides a much finer model to mimic the human perception of image quality. Visual information fidelity (VIF) [4], another IQA metric based on the natural image statistics in wavelet domain, has been demonstrated to be equivalent to SSIM [5].

In the last decade, independent component analysis (ICA), as a statistical model for natural images, has been used to formulate an image as the mixture of image basis [6]. The features extracted from ICA are sparse in nature, and suitable for IQA because they resemble the simple cells in mammalian primary visual cortex. Therefore, a novel IQA metric with sparse representation in ICA domain is proposed in this work. Experimental results show that the IQA metric proposed here can achieve competitive performances compared with SSIM.

II. Preliminary

A. SSIM

As mentioned above, SSIM utilizes the structural information to evaluate the image quality. Three factors, such as luminance, contrast and structure comparison are integrated into the SSIM metric which is defined as

$$SSIM(\mathbf{x},\mathbf{y})=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}. \quad (1)$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ represent the reference and distorted images, respectively; μ_x and σ_x (μ_y and σ_y) denote the mean and standard deviation of the reference (distorted) image, respectively; and σ_{xy} is the covariance between $\boldsymbol{x}$ and $\boldsymbol{y}$. The constants C_1 and C_2 are included in (1) to avoid instability.

Local SSIM metric is estimated using a symmetric Gaussian weighting filter. By pooling the spatial SSIM values, the overall image quality is obtained as

$$SSIM(\mathbf{x},\mathbf{y})=\frac{1}{M}\sum_{j=1}^{M}SSIM(\mathbf{x}_j,\mathbf{y}_j). \quad (2)$$

[+] Corresponding author.

where M is the number of local windows over the image; $\mathbf{x}_j$ and $\mathbf{y}_j$ represent the image patches covered by the jth window from the reference and distorted images.

B. ICA

ICA is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals, which arises from the study of blind source separation [6]. The standard linear ICA model can be formulated as

$$\mathbf{x}=\mathbf{A}\mathbf{s}. \tag{3}$$

where $\mathbf{x}=(x_1, x_2,\dots, x_n)^T$ is an observed vector, $\mathbf{s}=(s_1, s_2,\dots, s_m)^T$ is a vector of the independent latent variables, and $\mathbf{A}$ is an unknown mixing matrix. Using the observations of $\mathbf{x}$ alone, the task of ICA is to estimate both $\mathbf{s}$ and $\mathbf{A}$ simultaneously based on the assumption of independence of $\mathbf{s}$. In this work, FastICA [7], which is an efficient algorithm based on maximum negentropy, is employed to extract meaningful components to construct the proposed IQA metric.

III. Proposed IQA Metric based on ICA

A. ICA for image basis

ICA is a suitable computational technique for modeling the statistical structure of natural images. There is no unifying definition of natural images in literature. Simply they can be defined as photographs of typical environment where we live, which only form a tiny subset or manifold with low dimension in the space of all possible signals.

For simplicity, only monochrome image $\mathbf{I}(x,y)$ is studied here, where $\mathbf{I}(x,y)$ means the intensity value at the location (x,y). From the viewpoint of ICA, an image can be modeled as a linear weighted sum of features, which are assumed to be fixed and denoted by $\mathbf{A}_i(x,y)$, $i=1,\cdots,n$. For any incoming image, the coefficient of each feature is denoted by s_i. Algebraically, we can write the decomposition as

$$\mathbf{I}(x,y)= \sum_{i=1}^{n}\mathbf{A}_i(x,y)s_i\ . \tag{4}$$

where we assume for simplicity that the number of features n equals the number of pixels. And thus the system in (4) can be inverted, indicating that for a given image $\mathbf{I}$, the coefficients s_i can be computed as

$$s_i= \sum_{x,y}\mathbf{W}_i(x,y)\mathbf{I}(x,y)\ . \tag{5}$$

for certain weights $\mathbf{W}_i$. The terminology is not quite strict here, either $\mathbf{A}_i$ or s_i can be called feature. They will not bring any confusion because $\mathbf{A}_i$ is an image vector and s_i is a number associates with $\mathbf{A}_i$. The weights vector $\mathbf{W}_i$ is called a feature detector.

In [6], the relationship among the concepts of independence, non-gaussianity and sparseness is well studied. It is claimed that under quite loose assumptions, these three concepts are equivalent. In ICA, the weights $\mathbf{W}_i$ can be obtained by finding a local maximum of sparseness for s_i, then the image basis $\mathbf{A}_i$, $i=1,\cdots,n$, can be computed as the inversion of the feature detectors $\mathbf{W}_i$, $i=1,\cdots,n$. After applying ICA on 29 benchmark distortion-free images from the LIVE Database Release 2 with different size of image patches, two groups of features are obtained and shown in Fig. 1. There are 256 features on the left of Fig. 1, which constitute a basis in the 16×16 image patch space. In addition, only 256 representative features from the 32×32 image patch space are shown on the right of Fig. 1. By comparison, it turns out that the features obtained by maximization of sparseness look like the Gabor filters which have been widely used to model the simple cell receptive fields in physiology and anatomy. It is clear that the features in Fig. 1 have three basic localization properties, that is they are localized in (x,y)-space, frequency space and orientation space. It is interesting that the image basis is quite stable with different patch sizes, which may indicate that the sparse features in natural image form a manifold with low dimension in the overall patch space.

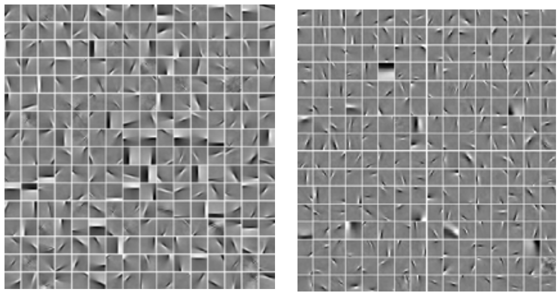

Fig. 1 Image basis obtained from ICA (Left: with patch size of 16×16; Right: with patch size of 32×32).

It deserves to be mentioned that the Gabor-like features in Fig. 1 are extracted by maximization of sparseness with the feature coefficients s_i, which corresponds to the firing rate of cells in primary visual cortex [6]. The point is that firing of cells consumes energy which is one of the major constraints on biological design of organism. The sparse regularization means that most cells do not fire more than their spontaneous firing rate. Thus, sparse representation is energy efficient, and the proposed ICA based model is more natural than the Gabor representation in natural image analysis.

B. Feature Extraction with l_1 Norm Regularization

Under the assumption that natural images can be formulated as the weighted mixture of image basis $\mathbf{A}_i$, the feature coefficient s_i obtained by ICA contains the local structure information of the natural image, which is a proper index for the construction of IQA metric.

Regarding the calculation of feature coefficients with the image basis, after computing the determinant of the image basis, we find that the image basis used to analyze the natural image is extremely unstable, because its determinant is as tiny as zero. This indication could partly verify our guess that the intrinsic dimension of natural image may be much lower than that of the original signal space.

To address this problem, a hybrid optimization strategy is introduced in this work. Generally speaking, a feature vector $\boldsymbol{s}$ is said to be sparse when most of its components are zero or very small in magnitude. Mathematically, l_0 norm is the proper index for measuring sparseness of a vector. But for its difficulty in manipulation, it is replaced by l_1 norm minimization, which is easy to be done with linear programming or second-order cone programming. The l_1 norm minimization can be found in many literatures on compressed sensing and sparse signal recovery in recent years. It has been proved that with overwhelming probability, the solution to l_1 norm minimization is equivalent to the one to l_0 norm minimization [8].

Therefore, in order to extract the feature vector $\boldsymbol{s}$ from natural images, we formulate this problem as follows

$$\boldsymbol{s}^*=argmin_s\ ||\boldsymbol{s}||_1+\lambda||\boldsymbol{I}\text{-}\boldsymbol{A}\boldsymbol{s}||_2. \tag{6}$$

where the l_1 norm term depicts the sparseness of the feature vector $\boldsymbol{s}$; the l_2 norm term indicates the reconstruction error; $\boldsymbol{A}=(\boldsymbol{A}_1, \cdots ,\boldsymbol{A}_n)$ with $\boldsymbol{A}_i$ as its columns; λ is the weight controlling the importance of the error term.

In general, the hybrid norm optimization is always used to solve underdetermined systems. As we have seen that the intrinsic dimension of natural images is not as large as the one in patch space, so the hybrid optimization strategy is suitable despite the system is well-defined in (6). Moreover, such optimization processing can provide more stable performance than inverting the feature matrix directly. In section IV, the MATLAB toolbox l_1-MAGIC is chosen to solve this hybrid optimization problem in (6).

C. IQA with Sparse Representation

As discussed above, the decomposition coefficients $\boldsymbol{s}$ by the feature matrix in natural images with no distortion is sparse in its components. When the feature matrix $\boldsymbol{A}$ is used to decompose any type of distortion images, the sparseness of the extracted coefficients is changed to a certain extent. Hence, the feature vector $\boldsymbol{s}$ is a good index to construct the IQA metric.

To provide a quantitative description of the change in sparseness with distorted images in LIVE Database Release 2, we model $\boldsymbol{s}$ with the generalized Gaussian distribution (GGD), which has been widely used in image modeling and RF IQA [9]. Mathematically, GGD is a double-parameter distribution with the form as

$$p(s)=\frac{\beta}{2\alpha\Gamma(1/\beta)}\,exp(-(|s|/\alpha)^{\beta}). \tag{7}$$

where α and β are two parameters which control the scale and shape of the distribution, respectively. In (7), when β=2, $p(s)$ corresponds to the Gaussian distribution. While β=1, we call it Laplacian, which is a typical sparse distribution. In general, the smaller the parameter β is, the sparser the data taken from this distribution are. To test the degree of the change in sparseness, we estimate the shape parameter β with the coefficient s extracted from both the distortion free and degraded image with the same image basis. It turns out that the average of the shape parameter β estimated from 29 distortion free images is round 0.5, and the one estimated from 982 degraded images from 5 different distortion types, including JPEG2000 compression (JP2K), JPEG compression (JPEG), white Gaussian noise (WN), Gaussian blur (GBLUR) and transmission errors in the JPEG2000 bit stream using a fast fading Rayleigh channel model (FF), in LIVE Database Release 2 is round 1.5. These two parameters are smaller than 2, so both of the distributions are super-Gaussian in statistics. But the degree of their sparseness is quite different.

So we can see that the feature vector $\boldsymbol{s}$ has taken much information which can be used to discriminate the difference between the distortion free and degraded images. Here, we use the correlation coefficient to measure the **D**ifference of **F**eature **V**ectors as an IQA metric, which is expressed as

$$DFV(\boldsymbol{s}_r^i,\boldsymbol{s}_d^i)=\frac{\langle \boldsymbol{s}_r^i\ \boldsymbol{s}_d^i\rangle}{||\boldsymbol{s}_r^i||_2||\boldsymbol{s}_d^i||_2}. \tag{8}$$

where $\boldsymbol{s}_r^i$ and $\boldsymbol{s}_d^i$ denote the feature vectors from the reference distortion free and degraded images, respectively. Both of them are extracted from the same image patch which is indexed by i. Essentially, DFV measures the cosine distance between the two feature vectors. To obtain an overall IQA metric, we can average the whole DFV indexes over all image patches. Accordingly, the overall image quality can be measured as

$$DFV(I_r,I_d)=\frac{1}{M}\sum_{i=1}^{M}DFV(s_r^i,s_d^i)\ . \tag{9}$$

where the letter M means the number of image patches.

IV. Experimental Results

The proposed IQA metric in (9) is tested on the LIVE Database Release 2, which contains 29 high resolution original images and 982 degraded images with different distortion levels from 5 types of distortion, including JP2K, JPEG, WN, GBLUR and FF. The goal of objective IQA is to design computational models that can predict perceived image quality accurately. Therefore, the most important criterion to evaluate the usefulness of the perceived IQA algorithm is whether its prediction correlates well with the human subjectivity. Two types of performance indexes are suggested in [10]. They are correlation coefficient (CC) which indicates prediction accuracy, and Spearman rank-order correlation coefficient (SROCC) which indicates prediction monotonicity. In addition, the MSE is also calculated. In this paper, the prediction between DFV (or SSIM) and DMOS is obtained by a cubic polynomial regression and the size of image patch is chosen to be 16×16.

Figure 2 shows the scatter plots of DMOS versus model prediction by the IQA metrics of DFV and SSIM with those 5 distortion types mentioned above. And Table I gives the performance comparison including MSE, CC and SROCC. As far as CC and MES are concerned, DFV is superior to SSIM except for the distortion type of JP2K. When SROCC is considered, SSIM is superior to DFV except for the distortion type of WN.

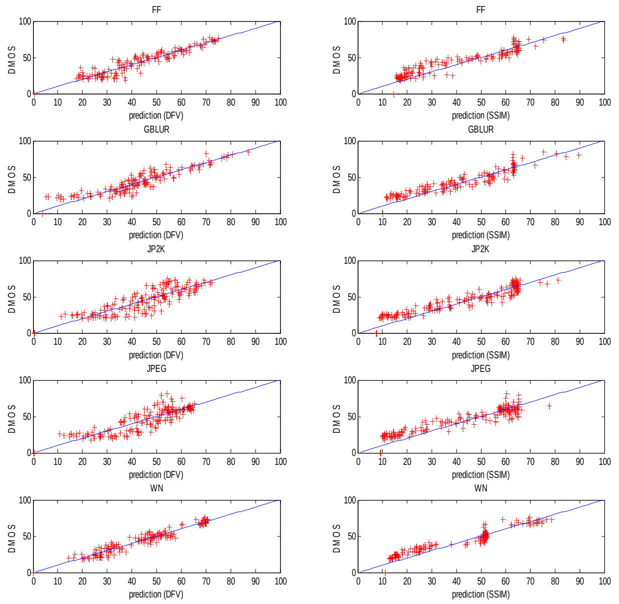

Fig. 2 Scatter plots of DMOS versus model prediction by DFV and SSIM with 5 distortion types.

TABLE I PERFORMANCE COMPARISON WITH CC, MSE AND SROCC.

Indexes	IQA	FF	GBLUR	JP2K	JPEG	WN
CC	DFV	0.9692	0.9559	0.9452	0.9480	0.9827
	SSIM	0.9121	0.9379	0.9503	0.9374	0.9444
MSE	DFV	5.4403	6.3848	7.9654	7.7191	4.0701
	SSIM	9.0558	7.5434	7.5966	8.4498	7.2299
SROCC	DFV	0.9371	0.9225	0.8372	0.8922	0.9475
	SSIM	0.9474	0.9495	0.9356	0.9246	0.9401

V. DISCUSSIONS AND CONCLUSIONS

In this paper, we have proposed a novel metric for FR-IQA, which is based on the sparse representation in ICA domain. Compared with the common method from harmonic analysis, the proposed decomposition method is induced from the natural image statistics, which coincides with the evidence from HVS. As demonstrated in the experimental results, the proposed metric DFV could achieve competitive performance with SSIM. As the image basis is fixed in the next experiments, the time consumption in our algorithm comes from the problem in (6). Compared to the simplicity of SSIM, the algorithms in l_1-MAGIC are based on linear programming, whose complexity is bounded by $O(n^3)$, where n is the size of image blocks. Exciting result comes from the algorithm in [11], which claims that the complexity can be reduced to $n\log(n/k)$, where k is the degree of sparseness. Though our algorithm is performed on gray images, by pooling scores from 3 channels in any color space, it's generalization to color images is considered to be easy. It is believed that the research on FR IQA has already plateaued, while RR and NR IQA are premature [12]. Compared to the SSIM on pixel level, the proposed DFV metric is based on feature level. So it is more suitable to study DFV metric in the case of RR and NR IQA problems, which leads to our future work.

VI. ACKNOWLEDGMENT

This work was supported in part by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, the Program for New Century Excellent Talents in University of China under Grant NCET-10-0634, the Shanghai Pujiang Program under Grant 11PJ1409400, the National Natural Science Foundation of China under Grant 61102059, and the 2010 Innovation Action Plan of Science and Technology Commission of Shanghai Municipality under Grant 10DJ1400300.

REFERENCES

[1] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*. New York: Morgan & Claypool, 2006.

[2] 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, Apr. 2004.

[3] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, pp. 1398-1402, Nov. 2003.

[4] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol.15, no.2, pp. 430-444, Feb. 2006.

[5] K. Seshadrinathan and A. C. Bovik, "Unifying analysis of full reference image quality assessment," in *IEEE Int. Conf. on Image Process.*, pp. 1200-1203, Oct. 2008.

[6] A. Hyvaerinen, J. Hurri, and P. O. Hoyer, *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision.* London: Springer, 2009.

[7] A. Hyvaerinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no.3, pp. 626-634, May, 1999.

[8] D. Donoho, "For most large underdetermined systems of linear equations the minimal l_l-norm solution is also the sparsest solution," *Comm. on Pure and Applied Math*, vol. 59, no.6, pp. 797-829, 2006.

[9] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Proc. SPIE Human Vision and Electronic Imaging*, pp. 149-159, Jan. 2005.

[10] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II," http://www.vqeg.org/.

[11] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the l_1 norm," in *49th Annual IEEE Symposium on Foundations of Computer Science*, pp. 199-207, Oct. 2008.

[12] A. K. Moorthy and A. C. Bovik, "Visual quality assessment algorithms: what does the future hold?" *Multimedia Tools and Applications*, vol. 51, no. 2, pp. 675-696, 2011.