Sparse Structural Similarity for Objective Image Quality Assessment

Xiang Zhang*, Shiqi Wang*, Ke Gu§, Tingting Jiang*, Siwei Ma* and Wen Gao*
*Nat'l Engineering Laboratory for Video Technology
*Cooperative Medianet Innovation Center
*Key Laboratory of Machine Perception (MoE)
*Sch'l of EECS, Peking University, Beijing, 100871, China.
§Shanghai Key Lab. of Digital Media Processing & Transmissions, Shanghai Jiao Tong University, Shanghai, China.

***Abstract*—In this paper, a novel full-reference (FR) image quality assessment (IQA) metric based on sparse representation is proposed. Sparse representation has been widely applied in many applications such as image denoising and restoration. It is a high-efficiency way in representing sparse and redundant natural images. Also it has been shown to be highly related to the human visual perception, which is characterized by a set of responses of neurons in visual cortex. In this paper, the sparse representation is applied in decomposing natural images into multiple layers depending on the visual importance. Inspired by these observations, a novel IQA metric called sparse structural similarity is proposed by measuring the fidelity of the stimulation of visual cortices. Experimental results on public databases indicate that the proposed method is effective in predicting subjective evaluation and as compared to state-of-the-art FR-IQA methods.**

***Index Terms*—Image quality assessment (IQA), orthogonal matching pursuit (OMP), sparse representation.**

I. Introduction

Image quality assessment (IQA) has shown its essentiality in a wide range of applications, such as image acquisition, compression, transmission, enhancement [1], [2] and analysis etc. IQA controls the perceptual quality performance of these systems. Objective IQAs can be classified into three categories depending on the availability of reference image: full-reference (FR), no-reference (NR) and reduced-reference (RR) methods. Pristine reference image is fully available in FR-IQAs, while NR- and RR-IQAs have no or part of the reference.

This paper will focus on the FR-IQA model. Given a distorted image and its corresponding reference image, a good FR-IQA method can effectively and efficiently predict the subjective perceptual score without any prior knowledge of specific distortion type. One of the most popular FR-IQA is mean square error (MSE) or peak signal-to-noise ratio (PSNR). It is appealing due to its high-efficiency and clear physical meaning. But it has long been criticized for its poor correlation with HVS [3]. Thus, more accurate criteria considering HVS characteristics have been studied, including visual signal-to-noise ratio (VSNR) [4], internal generative mechanism (IGM) [5], structure similarity (SSIM) [6] and its variants MS-SSIM [7], IW-SSIM [8]. SSIM, assuming that HVS tends to perceive structural information from natural scene, is a typical method for IQA and has become successful in many applications such as image/video coding [9], [10].

Sparse representation is efficient in dealing with rich, varied and directional information contained in natural scene. From the viewpoint of computational neuroscience, sparse representation leads to the emergence of receptive fields similar to the simple cells, which is a better model of visual cortex. Early work on optimizing sparse representation can be found in [11]. Sparse representation has shown its power in some IQA models [12]–[14]. In [15], a sparse feature based metric is presented and has shown good performance in matching subjective ratings. An image fidelity assessment algorithm is proposed by comparing the coefficients in sparse domain in [16], however it trains the content-adaptive dictionary for each input picture and has not considered the sparse structure in image representation.

In this paper, we develop a novel FR-IQA framework based on sparse representation. An online training algorithm [17] is applied for obtaining the global sparse dictionary using extensive natural images as training samples. Then, the Orthogonal Matching Pursuit (OMP) based algorithm [18] is employed in sparse representation, which is an effective tool for decomposing image into several layers with different visual importance. The proposed IQA method measures the similarity of sparse coefficients of each layer and weighted by their importance.

The rest of the paper is organized as follows. In Section II, we introduce the dictionary training process briefly and analyse the OMP algorithm in detail. In Section III, we propose the sparse representation based IQA model. Section IV shows the performance of the proposed FR-IQA algorithm. Finally, we conclude this paper in Section V.

II. Multi-layer Sparse Representation for Natural Images

Images captured in natural scene are instinctively sparse and redundant due to the local and non-local similarities among natural images. Sparse representation mimics the characteristics of human visual system by extracting sparse structures from images. In this section, the training process of constructing a global dictionary for sparse representation is briefly introduced, then the typical OMP method applied in sparse reconstruction is analysed in detail.

A. Dictionary Training

In sparse-land model [19], each signal $x \in \mathbb{R}^d$ can be represented by a linear combination of a few items (namely primitives) in D ($D \in \mathbb{R}^{d\times k}$). D is an over-complete dictionary containing k primitives. It is crucial that the dictionary D should be well designed to adapt with the image content. Such that $\forall x, \exists \alpha \in \mathbb{R}^k$ satisfying $x \approx D\alpha$ and $\|\alpha\|_0 \ll d$, where the notation $\|\bullet\|_0$ represents the l_0 norm. Let the input image be X, which can be partitioned into many non-overlapped patches $x_1, x_2, \cdots, x_i, i = 1, 2, \cdots, M$. The objective function is formulated as follows,

$$(D, \{\alpha_i\}) = \arg\min_{D,\{\alpha_i\}} \sum_k \|x_i - D\alpha_i\|_2^2, \text{s.t.} \|\alpha_i\|_0 < L, \quad (1)$$

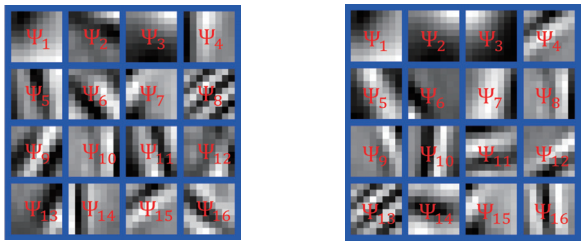
where L controls the sparse level. The K-SVD algorithm [20] is a typical method for training the content-adaptive sparse dictionary, which is performed in an iterative batch way by accessing the whole training set at each iteration in order to minimize the cost function under the constraints. However it cannot effectively handle with the very large training data, and also the computation complexity is considerable for realtime applications such as image quality assessment, image/video coding etc. To address these issues, an online training algorithm [17] is employed for obtaining the global dictionary for general input natural images in this work.

B. Sparse Representation

Sparse representation is referred to calculating appropriate coefficients α with respected to the trained dictionary D. Thus it can be formulated as below:

$$\alpha_i = \arg\min_{\alpha_i} \sum_k \|x_i - D\alpha_i\|_2^2, \text{s.t.} \|\alpha_i\|_0 < L. \quad (2)$$

The OMP method [18] is applied to obtain sparse representation vectors α_i for each patch x_i, which works in a greedy fashion, that choosing the primitive from D which is most similar with the residual at each iteration. Note that the residual of first iteration is the original patch itself. That is to say, in the first iteration, OMP chooses the primitive that is highly correlated with the original signal in a nearest neighbour way. Then the original signal is subtracted by the chosen primitive to update the residual.



(a) Patch 1 in *Lena* image (b) Patch 2 in *Lena* image

Fig. 1. Sixteen primitives picked by *OMP* algorithm arranged in raster order.

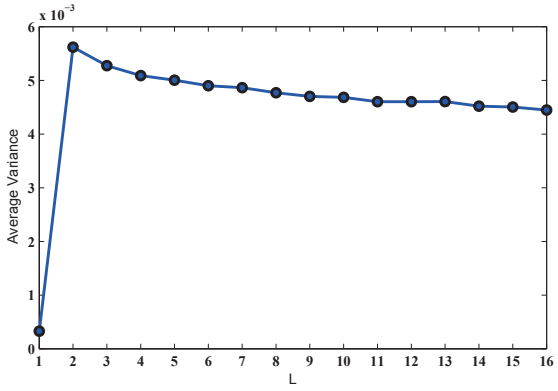


Fig. 2. Plot of the average variance V_j in terms of L.

C. Multi-layer Sparse Decomposition

Considering primitives as image blocks, it is intuitive that the primitive which is most similar with the image patch is picked first (i.e. Ψ_1), and more image details ($\Psi_2 \sim \Psi_L$), e.g. edges and textures, are added for recovering the image. As shown in Fig. 1, sixteen picked primitives ($\Psi_1 \sim \Psi_{16}$) are arranged in raster order. They are picked out to reconstruct two 8×8 patches of *Lena* image respectively. The first chosen primitive, located at the top-left corner, is visually smoother than other primitives picked in later iterations.

More generally, let $\Psi_(i, j)$ denote the j^{th} selected primitive of the i^{th} patch, and $V_j = \frac{1}{N}\sum_{i=1}^{N} Var(\Psi_{i,j})$ is the average variance of all the j^{th} primitives in an image, where $Var(*)$ represents the variance operation and N is the number of patches. In Fig. 2, the $V_j (1 \le j \le 16)$ curve of *Lena* image is plotted. It is obvious that the average variance of the first primitive (i.e. V_1) is much lower than that of other primitives. It can be interpreted by the fact that natural images contain large percentage of smooth area. Such that the first chosen primitives are always smooth, and the average variance of first primitives is rather small. And the variances from V_2 to V_{16} are relatively stable and gradually decline as shown in Fig. 2. This observation is highly related to the visual perception that the basic component (e.g. what the object is it) is perceived before details (e.g. what does the object look alike).

The OMP algorithm has shown to be an effective tool in decomposing image into a hierarchical representation, typically containing a basic layer and several detail layers as depicted in Fig. 3. Each layer is constructed by the primitives of i^{th} iteration and its coefficients. Particularly, the basic layer utilizes the several previous primitives for reconstruction and describes the major structural information contained in images. However, HVS is relatively insensitive with the last detail layers relating to the non-structural information. Thus we conclude that the OMP algorithm reconstructs image using a set of primitives selected one by one from trained dictionary, and these primitives are naturally ordered by perceptual importance. This greedy characteristic of OMP algorithm is the

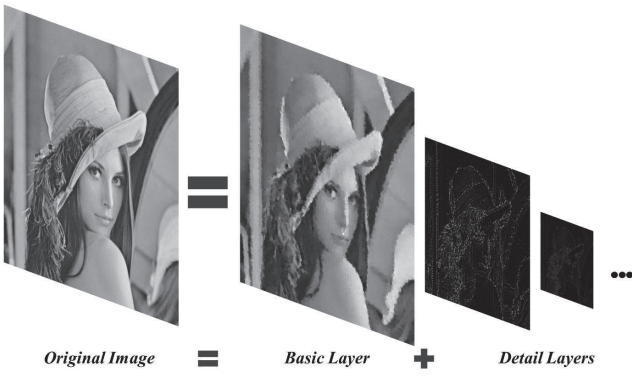


Fig. 3. Illustration of multi-layer image decomposition using sparse theory.

foundation of the proposed IQA method.

III. Sparse Structural Similarity for Image Quality Assessment

In this section, we detail the framework of the proposed IQA scheme as illustrated in Fig. 4. On the sender side, original image is firstly partitioned into 8×8 non-overlapped patches. And the dictionary for sparse representation is trained by K-SVD algorithm using these patches. Then two OMP-motivated methods, namely *modified* OMP as well as *constrained* OMP, are specifically defined for original and distorted images respectively. By these processes, the sparse structures including the selected primitives and its corresponding coefficients are extracted. The local quality map (LQM) is calculated by comparing the coefficients between original and distorted patches. Lastly we get the ultimate quality score using the average weighted by distortion level of each patch.

A. Local Quality Map

Firstly, original and distorted images are partitioned into several non-overlapped patches, denoted by $\{x_i\}$ and $\{x'_i\}$ respectively. For each original patch $\{x_i\}$, the *modified* OMP algorithm is used to approximately reconstruct it by a set of primitives $\{\Psi_{i,1}, \ldots, \Psi_{i,L}\}$ and corresponding coefficients $\{\alpha_{i,1}, \ldots, \alpha_{i,L}\}$. This process can be formulated by following equation:

$$\hat{x}_i = \sum_{j=1}^{L} \alpha_{i,j} \Psi_{i,j}, \tag{3}$$

where $\hat{x}_i$ represents the reconstituted original patch and L controls sparseness level in OMP algorithm.

For distorted images, the problem can be formulated as a simplified optimization problem,

$$\alpha'_{i,j} = \arg\min_{\alpha'_{i,j}} \sum_{k} \left\| x_i - \Psi_r \alpha'_{i,j} \right\|_2^2. \tag{4}$$

We call this method as guided decomposition because the sparse primitives denoted by Ψ_r is composed of that used for pristine image,

$$\Psi_r \triangleq \left[\Psi_{i,1}, \ldots, \Psi_{i,L}\right]^T. \tag{5}$$

It is a subproblem in OMP when recalculating the coefficients at each iteration and can be solved by

$$\alpha'_{i,j} = \left((\Psi_r^T \Psi_r)^{-1} (\Psi_r)^T\right) x_i \tag{6}$$

Before calculating the similarity term, the normalization process is typically performed to the sparse coefficients of both original and distorted images. It is performed by subtracting the average and dividing the standard deviation as follows,

$$\widetilde{\alpha}_{i,j} = \frac{\alpha_{i,j} - AVG_j}{STD_j}, \tag{7}$$

where $AVG_j = \frac{\sum_{i=1}^{M} \alpha_{i,j}}{M}$ and $STD_j = \sqrt{\frac{(\alpha_{i,j} - AVG_j)^2}{M}}$. Accordingly, the local quality S_i can be calculated by the normalized coefficients as follows,

$$S_i = \frac{\sum_{j=1}^{L} \omega_j \frac{2\widetilde{\alpha}_{i,j}\widetilde{\alpha}'_{i,j} + C_1}{\widetilde{\alpha}_{i,j}^2 + \widetilde{\alpha}'^2_{i,j} + C_1}}{\sum \omega_j}, \tag{8}$$

where C_1 is a constant to avoid instability, and ω_j denotes the weighting on each sparse level. Considering the observation that the visual importance is decreasing with the increasing value of sparse level L, it is highly recommended that ω_j is a monotonic decreasing function. In this work, ω_j is empirically designed as a Laplacian function as follows,

$$\omega_j = \frac{1}{2\sigma} e^{-\frac{(j-1)^2}{\sigma}}, \tag{9}$$

where σ is the only model parameter that controls the divergence degree between different sparse level.

B. Pooling Strategy

Pooling process has been considered to be a crucial strategy and is applied in many popular IQAs, by which an ultimate quality assessment score is obtained by calculating the weighted average of local quality scores.

Considering the fact that the more degraded image regions lead to worse subjective experience. The weighting strategy is modeled as an exponential function of the distortion as follows,

$$W_i = e^{C_2(1-S_i)}, \tag{10}$$

where C_2 is a parameter that guides the relative importance on the overall score. The ultimate score, dubbed as sparse structural similarity (SSS), is a weighted average as follows,

$$SSS = \frac{\sum_{i=1}^{N} W_i S_i}{\sum W_i}. \tag{11}$$

IV. Experimental Results

To evaluate the performance of proposed SSS scheme, we compare our work with some state-of-the-art FR-IQAs, including PSNR, VSNR [4], SSIM [6], MS-SSIM [7], GSM [21], IGM [5], GMSM/GMSD [22], VSI [23] and LTG [24], on public databases. The basic information of these datasets are given as follows:

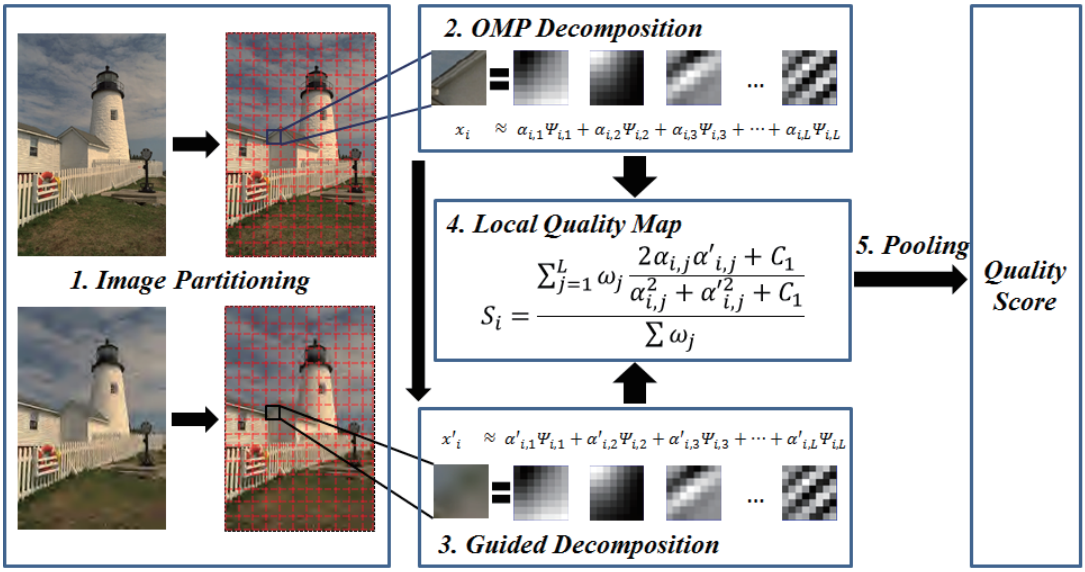


Fig. 4. Flowchart of the proposed IQA algorithm which follows the steps: 1) image partition; 2) OMP decomposition for pristine image patches; 3) guided composition for distorted image patches using the same sparse primitives of original one; 4) pooling strategy for generating the overall quality score.

- The LIVE database [25] has 29 reference images and 779 distorted images, undergone 5 different distortion types including JPEG2000, JPEG, white noise, Gaussian blur, and fast fading channel distortion.
- The LIVEMD database [26] is the first image database for multiple distortions. It has two image subsets which are created by adding different levels of noise/JPEG to blurred images, respectively. There are 225 images generated from 15 pristine images in each subset.
- The CSIQ data base [27] totally consists of 866 images, which are created from 30 original images by using six types of distortions at four to five distortion levels. In this work, the 750 distorted images are selected for evaluation excluding the contrast distortion type. Actually the images with contrast changes are not really distorted and even more pleasing to human eyes.
- The TID2008 database [28] is the largest database including 1700 distorted images generated from 25 references with 17 distortion types at 4 distortion levels. Here, we pick 300 images corrupted by three common distortion types including: a) Gaussian blur; b) JPEG compression; c) JPEG2000 compression.

Five criteria are employed in evaluating the performance of IQAs, which are Pearson linear correlation coefficient (PLCC) for measuring prediction accuracy, Spearman rank-order correlation coefficient (SRCC) and Kendall rank-order correlation coefficient (KRCC) for measuring prediction monotonicity, Mean Absolute Difference (MAD) and Root mean squared error (RMSE) for measuring prediction consistency. A good IQA model is expected to acquire one for PLCC, SRCC and KRCC, while acquire zero for MAD and RMSE.

In order to quantify the accuracy of predicted quality, we use a five-parameter logistic function for non-linear mapping:

$$p\left(o\right)=\beta_1\left(\frac{1}{2}-\frac{1}{1+e^{\beta_2\left(o-\beta_3\right)}}\right)+\beta_4 o+\beta_5, \qquad (12)$$

where o and p represent the objective score and predicted subjective score respectively.

The simulation results are reported in Tab. I&II. It has been shown that the proposed scheme outperforms other state-of-the-art FR-IQAs on CSIQ database, and shows competitive performance on other databases. It can be proven that the SSS algorithm can accurately estimate the perceptual quality of natural scenes for most typical distortions. Also the scatter plots of several IQAs and the proposed SSS on CSIQ database are plotted in Fig. 5. Those algorithms include PSNR, SSIM, MS-SSIM, IW-SSIM [8], GSM, GMSM and VSI. It can be concluded that the convergency and monotonicity of the SSS overmatches other methods.

The computational complexity is mainly dominated by the OMP algorithm, whose cost is $O(Ldk)$ [18]. Such that the overall complexity of the proposed algorithm is $O(LdkN)$, where N is the number of the image patches. The time can be further saved with parallel implementation because image patches are independent to each other.

Moreover, Fig. 6 gives the relationship between the PLCC score and the iteration L evaluated on different databases, from which it can be observed that the PLCC increases monotonously and converges quickly to the maximum with the iteration times L. It reveals that the coefficients at latter iterations have almost no contribution to the PLCC improvement. It accords well with the discussions in Section II that the structural information is mostly recovered by the first few layers, and the distortions in latter non-structural layers are negligible to subjective perception.

TABLE I

PERFORMANCE COMPARISON BETWEEN THE PROPOSED SSS AND OTHER COMPETING FR-IQAS ON LIVE AND LIVEMD DATABASES. THE BOLDFACE FONTS IN EACH COLUMN INDICATE THE TOP 5 METHODS.

	LIVE database (779 images) [25]					LIVEMD database (450 images) [26]				
IQA metrics	PLCC	SRCC	KRCC	MAD	RMSE	PLCC	SRCC	KRCC	MAD	RMSE
PSNR	0.8723	0.8756	0.6865	10.5093	13.3597	0.7414	0.6771	0.5003	10.2256	12.6920
VSNR [4]	0.9236	0.9279	0.7624	8.0446	10.4760	0.8123	0.7719	0.5769	8.7654	11.0308
SSIM [6]	0.9042	0.9104	0.7311	9.2279	11.6694	0.6683	0.6459	0.4633	11.7764	14.0687
MSSSIM [7]	0.9489	0.9513	0.8044	6.6969	8.6181	0.8747	0.8363	0.6442	7.4754	9.1650
GSM [21]	0.9512	**0.9561**	**0.8150**	**6.5787**	8.4323	**0.8808**	**0.8454**	**0.6550**	**7.1738**	**8.9562**
IGM [5]	**0.9570**	**0.9581**	**0.8250**	**6.0101**	**7.9242**	**0.8859**	**0.8562**	**0.6681**	**7.1120**	**8.7744**
GMSM [22]	**0.9556**	**0.9595**	**0.8242**	**6.1976**	**8.0486**	**0.8787**	0.8442	**0.6574**	**7.1813**	**9.0278**
GMSD [22]	**0.9603**	**0.9603**	**0.8268**	**5.7892**	**7.6214**	**0.8810**	**0.8448**	0.6547	**7.0778**	**8.9492**
VSI [23]	0.9482	0.9524	0.8058	6.8593	8.6812	0.8577	0.8121	0.6189	7.8094	9.7221
LTG [24]	**0.9536**	**0.9585**	**0.8212**	**6.4729**	**8.2231**	0.8774	**0.8444**	**0.6570**	**7.2111**	9.0740
SSS	**0.9533**	0.9541	0.8086	6.5897	**8.2497**	**0.8811**	**0.8490**	**0.6569**	7.2516	**8.9447**

TABLE II

PERFORMANCE COMPARISON BETWEEN THE PROPOSED SSS AND OTHER COMPETING FR-IQAS ON CSIQ AND TID2008 DATABASES. THE BOLDFACE FONTS IN EACH COLUMN INDICATE THE TOP 5 METHODS.

	CSIQ database (750 images) [27]					TID2008 database (300 images) [28]				
IQA metrics	PLCC	SRCC	KRCC	MAD	RMSE	PLCC	SRCC	KRCC	MAD	RMSE
PSNR	0.7970	0.8019	0.6005	0.1210	0.1591	0.6165	0.6438	0.4772	0.7892	1.0725
VSNR [4]	0.7979	0.8092	0.6214	0.1176	0.1588	0.6692	0.7032	0.5289	0.7033	1.0121
SSIM [6]	0.8158	0.8397	0.6364	0.1166	0.1524	0.6829	0.6663	0.4904	0.7885	0.9950
MSSSIM [7]	0.9009	0.9160	0.7432	0.0866	0.1143	0.8603	0.8736	0.6814	0.5501	0.6945
GSM [21]	0.8979	0.9133	0.7413	0.0842	0.1160	0.8605	**0.8747**	**0.6930**	0.5161	0.6939
IGM [5]	**0.9292**	**0.9422**	**0.7906**	**0.0714**	**0.0974**	**0.9034**	**0.9055**	**0.7297**	**0.4515**	**0.5842**
GMSM [22]	0.9132	0.9303	0.7653	0.0776	0.1074	0.8505	0.8655	0.6754	0.5492	0.7165
GMSD [22]	**0.9536**	**0.9568**	**0.8127**	**0.0592**	**0.0793**	**0.8830**	**0.8983**	**0.7160**	**0.4903**	**0.6395**
VSI [23]	**0.9273**	**0.9426**	**0.7859**	**0.0715**	**0.0986**	**0.8841**	**0.9101**	**0.7302**	**0.4806**	**0.6365**
LTG [24]	**0.9531**	**0.9592**	**0.8179**	**0.0589**	**0.0798**	**0.8947**	**0.9167**	**0.7406**	**0.4630**	**0.6085**
SSS	**0.9609**	**0.9577**	**0.8194**	**0.0578**	**0.0755**	**0.8834**	0.8663	0.6865	**0.5045**	**0.6382**

V. CONCLUSION

In this paper, a novel FR-IQA scheme based on sparse representation is proposed. Sparse representation is an emerging and powerful tool in describing structural and directional information from natural scenes. We demonstrate the process of OMP method in image reconstruction, by which image is decomposed into multiple layers with different visual importance. Based on this observation, a novel IQA method, dubbed sparse structural similarity (SSS) is proposed. Image quality is estimated by calculating the fidelity of sparse coefficients in different layers. The pooling stage employs a distortion guided weighting strategy to generate the ultimate objective score. Experimental results on public databases show competitive performance with several state-of-the-art FR-IQAs. Our future work will focus on developing a general IQA framework based on this work that can be used for both full-reference and reduced-reference IQAs.

ACKNOWLEDGMENT

This work was partially supported by National Basic Research Program of China (973 Program) under contract 2015CB351803 and the Natural Science Foundation of China under contracts 61322106, 61390514, 61421062 and 61210005.

REFERENCES

[1] K. Gu, G. Zhai, X. Yang, W. Zhang, and C. Chen, "Automatic contrast enhancement technology with saliency preservation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, pp. 1–1, 2014.

[2] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–1, 2015.

[3] B. Girod, "Psychovisual aspects of image processing: What's wrong with mean squared error?" in *Workshop on Multidimensional Signal Processing*, 1991.

[4] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, 2007.

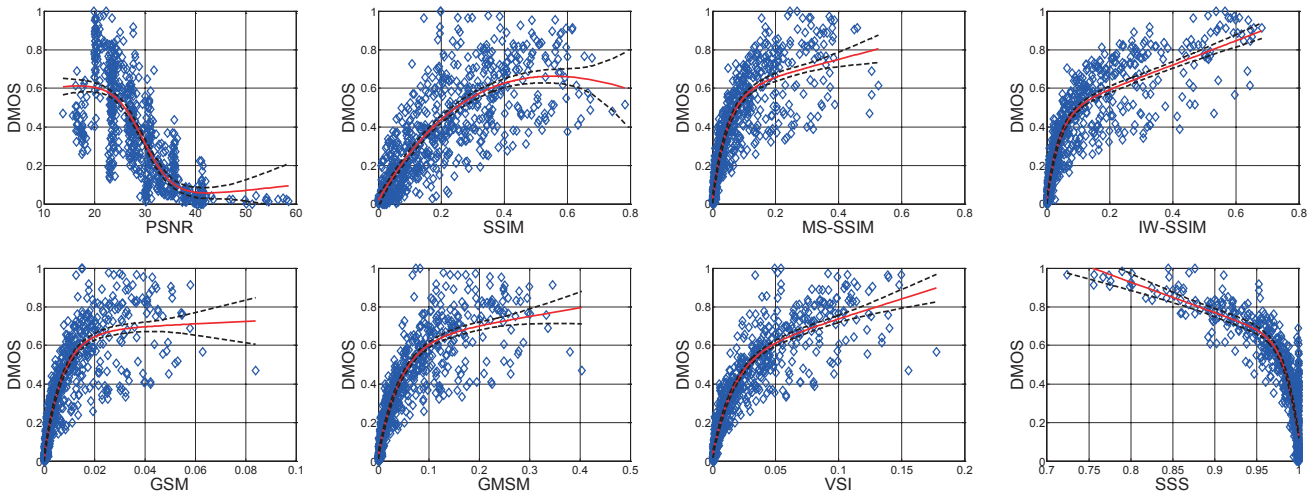


Fig. 5. Scatter plots of DMOS versus PSNR, SSIM, MS-SSIM, IW-SSIM, GSM, GMSM, VSI, and the proposed SSS on CSIQ database. The red solid lines are curves fitted with the logistic function and the black dash lines are 95% confidence intervals.

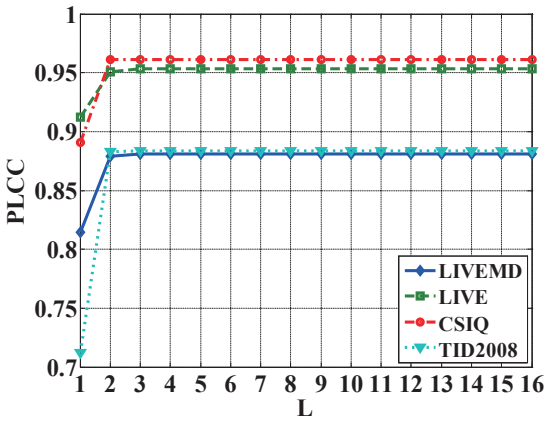


Fig. 6. PLCC Plots in terms of iteration L on different databases.

[5] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 43–54, 2013.

[6] W. Zhou, 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.

[7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Asilomar Conference on Signals, Systems and Computers*, vol. 2. Ieee, 2003, pp. 1398–1402.

[8] W. Zhou and L. Qiang, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, 2011.

[9] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516–529, April 2012.

[10] ——, "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1418–1429, April 2013.

[11] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[12] X. Zhang, S. Wang, S. Ma, S. Liu, and W. Gao, "Entropy of primitive: A top-down methodology for evaluating the perceptual visual information," in *Visual Communications and Image Processing (VCIP)*, 2013, pp. 1–6.

[13] S. Wang, X. Zhang, S. Ma, and W. Gao, "Reduced reference image quality assessment using entropy of primitives," in *Picture Coding Symposium (PCS)*, 2013, Conference Proceedings, pp. 193–196.

[14] X. Zhang, S. Wang, S. Ma, and W. Gao, "Towards accurate visual information estimation with entropy of primitive," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015.

[15] C. Hua-Wen, Y. Hua, G. Yong, and W. Ming-Hui, "Sparse feature fidelity for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 4007–4018, 2013.

[16] T. Guha, E. Nezhadarya, and R. Ward, "Learning sparse models for image quality assessment," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 151–155.

[17] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.

[18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[19] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[20] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[21] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images." in *NIPS*, 1999, pp. 855–861.

[22] 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, 2014.

[23] L. Zhang, Y. Shen, and H. Li, "VSI: a visual saliency-induced index for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4270–4281, 2014.

[24] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 506–510.

[25] H. Sheikh, Z.Wang, L. Cormack, and A. Bovik, "LIVE image quality assessment database release 2." [Online]. Available: http://live.ece.utexas.edu/research/quality

[26] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2012, pp. 1693–1697.

[27] E. C. Larson and D. M. Chandler, "Categorical image quality (CSIQ) database." [Online]. Available: http://vision.okstate.edu/?loc=csiq

[28] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.