Quality Assessment of Tone-mapped Images Based on Sparse Representation

Lijuan Xie*, Xiang Zhang*, Shiqi Wang*, Xinfeng Zhang*, Siwei Ma*†
*Institute of Digital Media & Cooperative Medianet Innovation Center, Peking University, Beijing, China
†Peking University Shenzhen Graduate School, Shenzhen, China
Email: {xielijuan, x_zhang, sqwang, swma}@pku.edu.cn, xfzhang@jdl.ac.cn

***Abstract*—Recently, an increasing number of tone-mapping operators (TMOs) have been proposed in order to display high dynamic range (HDR) images on low dynamic range (LDR) devices. Developing perceptually consistent image quality assessment (IQA) measures for TMO is highly desired because traditional LDR based IQA methods cannot support the cross dynamic range quality comparison. In this paper, a novel objective quality assessment method is proposed on the basis of sparse-domain representation, which has been well advocated as a powerful tool in describing natural sparse signals with the over-complete dictionary. Specifically, two indices, incorporating both local and global features extracted from sparsely represented coefficients, are introduced to simulate the human visual system (HVS) characteristics on HDR images. The local feature measures the sparse-domain similarity between the pristine HDR and tone-mapped LDR images by leveraging the intrinsic structure with sparse coding. On the other hand, benefiting from the natural scene statistics (NSS), the global features are recovered from the sparse coefficients to account for the natural behaviors of tone-mapped images. Combining the local sparse-domain similarity and the global "naturalness" prior, validations on the public database show that the proposed sparse-domain model for tone-mapped images (SMTI) provides accurate predictions on the human perception of tone-mapped images.**

***Keywords*—High dynamic range, image quality assessment, tone-mapping operators, sparse representation.**

I. Introduction

Traditional device referred technology uses low dynamic range (LDR) image formats to accommodate the capabilities of display devices. This leads to irrecoverable losses of information as the visible range of human visual system (HVS) is much larger than the range achievable by the traditional cameras or displays. High dynamic range (HDR) technology overcomes such limitations and is able to offer high levels of immersion by adapting to the broad range of luminance levels that can be perceived by the HVS. However, capture and display devices for HDR images are too expensive at current stage, precluding the development of this field. To overcome this issue, a large number of tone-mapping operators (TMOs) have been developed to fill the gap between real-world scene and display luminance. Existing tone-mapping algorithms can be classified into two categories, i.e. the global and local operators. Global one performs the identical operations for all pixels regardless of any local variance [1].In contrast, the local operator alters the transform metric according to the regional characteristics of local image patches [2].

Traditionally, much efforts have been devoted to evaluating the TMOs based on subjective evaluation [3], which is expensive and time consuming in general. Developing trusted objective TMO image quality assessment (IQA) method is highly desired for automatically controlling and optimizing the TMOs. In view of this, the tone-mapped image quality assessment methods (TMQI) that combines the structural similarity metric and the first/second order image information is developed to measure the quality of tone-mapped images in [4], [5]. However, the correlation between subjective scores and TMQI models implies that there is still large room for improvement on developing accurate objective IQA models to evaluate TMOs.

Sparse representation is efficient in dealing with rich, varied and directional information contained in natural scene [6] and image restoration [7], [8]. In our previous work, it has been demonstrated to be closely related to the human perception [9]–[13]. Inspired by this, an objective quality assessment metric for tone-mapped images based on sparse-domain representation is proposed. The model is composed of two parts, the local sparse-domain similarity and global naturalness measurement. Sparse similarity preserves local behavior between the pristine HDR and generated LDR images by leveraging the intrinsic structure during sparse coding. The global measurement is fruited by the natural scene statistics, seeking to capture the natural behaviors of tone-mapped images from the coefficient distribution in sparse coding.

The rest of this paper is organized as follows. Section II explains the procedure of dictionary learning and sparse decomposition. In Section III, the proposed SMTI algorithm is described and analyzed. Section IV shows the performance comparing with the state-of-the-art algorithms. Finally, Section V concludes this paper.

II. Unified Sparse Decomposition

In this section, we will briefly introduce the process of sparse decomposition using a pre-trained dictionary from HDR images. It is worth noting that both the HDR and LDR images are projected to the unified sparse space to make sure they are represented under the identical bases.

A. Dictionary Training

Based on sparse-land model [14], each image patch $\boldsymbol{x} \in \mathbb{R}^d$ can be represented by a linear combination of a few primitives in an over-complete dictionary $\boldsymbol{D}$ ($\boldsymbol{D} \in \mathbb{R}^{d \times k}$). As such, $\forall \boldsymbol{x}$, $\exists \boldsymbol{\alpha} \in \mathbb{R}^k$ satisfying $\boldsymbol{x} \approx \boldsymbol{D}\boldsymbol{\alpha}$ and $\|\boldsymbol{\alpha}\|_0 \ll d$, where the notation $\|\bullet\|_0$ represents the l_0 norm. Let the input image be X, which can be partitioned into non-overlapped patches $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,

978-1-4799-5341-7/16/$31.00 ©2016 IEEE

$\cdots$, $\boldsymbol{x}_i$, $i = 1, 2, \cdots, N$. The objective function of dictionary training can be formulated as follows,

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

where L controls the sparse level. The typical KSVD algorithm [15] is used to learn the content-adaptive dictionary.

To apply the design philosophy of sparse representation in the scenario of tone mapping, the patches from HDR images are firstly extracted for dictionary learning. In regarding to the LDR image, the same dictionary is applied to ensure the HDR and LDR share the same over-complete dictionary in representation. Alternatively, a global dictionary can be obtained offline from a large dataset of images for training, leading to less computational complexity in real applications. The performance comparison between the two strategies will be given in Section IV.

B. Unified Sparse Decomposition

Sparse decomposition aims at approximately recovering the image signals using a few number of primitives and appropriate coefficients with respected to the trained dictionary $\boldsymbol{D}$. This can be formulated as follows:

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

It is a NP hard problem but can be well solved by matching pursuit family. The orthogonal matching pursuit (OMP) algorithm [16] is applied in this work.

The HDR and LDR images are both partitioned into 4×4 non-overlapped patches, denoted as $\{\boldsymbol{x}_i\}$, $\{\boldsymbol{x}'_i\}$ respectively. For each HDR patch $\boldsymbol{x}_i$, the OMP method can choose the suitable primitives $\{\Psi_{i,1}, \ldots, \Psi_{i,L}\}$ and the relevant coefficients $\{\alpha_{i,1}, \ldots, \alpha_{i,L}\}$. For the corresponding LDR patch $\boldsymbol{x}'_i$, the sparse coefficient is obtained using the same primitives with the HDR patch as follows,

$$\boldsymbol{\alpha}'_i = ((\boldsymbol{\Psi}_r^T \boldsymbol{\Psi}_r)^{-1} (\boldsymbol{\Psi}_r)^T) \boldsymbol{x}'_i \quad (3)$$

where $\boldsymbol{\Psi}_r = [\Psi_{i,1}, \ldots, \Psi_{i,L}]^T$. As such, the HDR and its tone-mapped LDR patches are projected into the unified sparse space which is learnt from HDR ones. In the following, the sparse coefficient matrices for the HDR and LDR images are denoted as $\boldsymbol{A_{hdr}}$ and $\boldsymbol{A_{ldr}}$ respectively.

III. Sparse-domain Quality Measurement

The proposed SMTI model is defined as two-stage framework by taking advantages of the sparse coefficients derived in Section II. The first stage is to calculate the sparse-domain similarity that predicts the local quality between the pristine HDR and the generated LDR images by leveraging the intrinsic structure during sparse coding. The second stage is measuring the overall feature to capture the experience of HVS characteristics according to the sparse coefficient distribution of the LDR images. It's worth noticing that images with RGB color space is firstly converted to the Yxy space and only the Y component is taken into account in the implementation.

A. Sparse-domain Similarity

As sparse coefficients inherits the different range between HDR and LDR images in representing signals and the variation of range in a local window is negligible, we firstly perform a local coefficients comparison instead of obtaining similarity map using original coefficients $\boldsymbol{A_{hdr}}$. Then, motivated by the method calculating block similarity in [17], we transform them into a unified space through L2 norm. Specifically, the sparse coefficients are pre-filtered by a local standard deviation (std) function, followed by the normalization with respect to its coefficient energy,

$$\boldsymbol{A'_{hdr}} = \frac{\boldsymbol{A_{hdr}} * f_\sigma}{\|\boldsymbol{A_{hdr}} * f_\sigma\|_2}, \quad (4)$$

where $*$ is the 2-D convolution operator and f_σ is a 3×3 std kernel to restrict the coverage area within 12×12 pixels. $\boldsymbol{A'_{hdr}}$ is the normalized coefficient matrix of the HDR image. Along this vein, $\boldsymbol{A'_{ldr}}$ can be calculated in a similar way.

Subsequently, the similarity map between HDR and LDR images can be obtained,

$$\boldsymbol{F} = \frac{2\boldsymbol{A'_{hdr}} \cdot \boldsymbol{A'_{ldr}} + C_1}{\boldsymbol{A'_{hdr}} \cdot \boldsymbol{A'_{hdr}} + \boldsymbol{A'_{ldr}} \cdot \boldsymbol{A'_{ldr}} + C_1}, \quad (5)$$

where $\cdot$ is the dot product operator, and C_1 is a stabilizing constant.

Finally, the sparse-domain similarity index can be calculated by a pooling process as follows,

$$S = \frac{\text{tr}\,(\boldsymbol{W}^T \boldsymbol{F})}{\|\,\boldsymbol{W}\,\|}. \quad (6)$$

$\boldsymbol{W} = \left[\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(i)}, \ldots, \boldsymbol{w}^{(N)}\right]^T$ is the weighting matrix consisting of N row vectors where N is the number of the patches. Each row vector weights an image patch.

In this comparison, all the patches shares the same weighing strategy. Considering the perceptual importance of different sparse levels, the weighting vector $\boldsymbol{w}^i$ is designed to be a monotonic decreasing function as follows,

$$w_j^i = \frac{1}{2\eta} \exp\left[-\frac{(j-1)^2}{\eta}\right], \quad (7)$$

where η is the only parameter that controls the divergence degree between sparse levels and set as 2 empirically.

B. Naturalness Measurement

Benefiting from the natural scene statistics, a global no-reference (NR) metric with respect to tone-mapped images is developed to measure the perceived naturalness. First, plenty of natural LDR images representing various scene types are decomposed by the OMP algorithm, obtaining sparse coefficients matrices respectively. The measurement factor for arbitrary sparse level j is then defined as follows,

$$M_j = \frac{\sum_{i=1}^N \alpha_{i,j}}{N}, \quad (8)$$

where $\alpha_{i,j}$ is the sparse coefficient for image patch x_i in sparse level j. Subsequently, the representative for natural image is

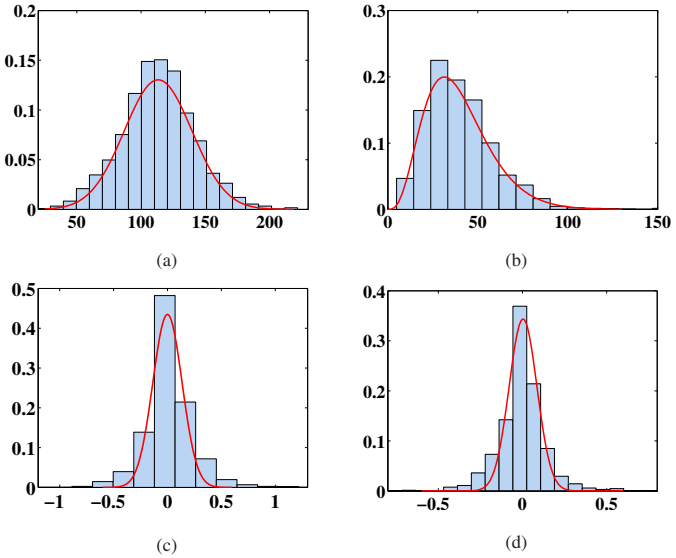

Fig. 1. The histograms of (a)M_1 (fitted by Gaussian Kernel), (b)M_2 (fitted by Beta PDF), (c)M_3 (fitted by Gaussian Kernel) and (d)M_4 (fitted by Gaussian Kernel) of natural images. M_j is the mean of sum of the j-th dimension in sparse coefficients.

simplified to an L-parameter vector from the redundancy pixel matrix, denoted as $\boldsymbol{M} = [M_1, M_2, \ldots, M_j, \ldots, M_L]$.

Fig. 1 shows the statistics histograms of the basic four characterization factors based on 2300+ natural LDR images, respectively. By observation, the effect of $M_j(j{>}3)$ is not significant. It inspires us to employ the basic three sparse levels $(M_j, j = 1, 2, 3)$ to avoid redundant calculations while keeping best performance.

In addition, the distribution curve of M_1 and M_3 fits well with Gaussian kernel defined in Eqn. 9 and M_2 is satisfied with Beta function well given by Eqn. 10.

$$P_g(M_j) = \exp\left[-(\frac{M_j - \mu}{\sigma})^2\right] \tag{9}$$

$$P_b(M_j) = \frac{(1 - M_j)^{\beta-1}(M_j^{\gamma-1})}{B(\gamma, \beta)} \tag{10}$$

where $B(\gamma, \beta)$ is the Beta function. We perform a training to obtain the parameters, and the values are set as follows.($\mu_1 = 113.3$, $\sigma_1 = 37.22$, $\gamma_2 = 4.095$, $\beta_2 = 35.06$ and $\mu_3 = 0.008426$, $\sigma_3 = 0.2262$.)

Considering the various significance of sparse level, different scaling factors should be defined to adjust the respective impact factor. To measure the importance of each sparse level, an experiment on the basis of relevancy statistics comparing to subjective data is carried out and the result is shown in Fig. 2(a). Higher KRCC & SRCC score stands for stronger relevancy. It demonstrates that the correlation is decreasing while the indexing of sparse level rising and the first sparse level achieves significantly higher performance than the other two. Hence, the naturalness is defined as follows,

$$R = P_g(M_1)\ P_b(M_2)^a\ P_g(M_3)^b, \tag{11}$$

where the scale of M_1 is set to 1 as reference, and the other two exponents decrease progressively, $0{<}a, b{<}1$. We set $a = 0.9077$ and $b = 0.1106$ as the final parameters.

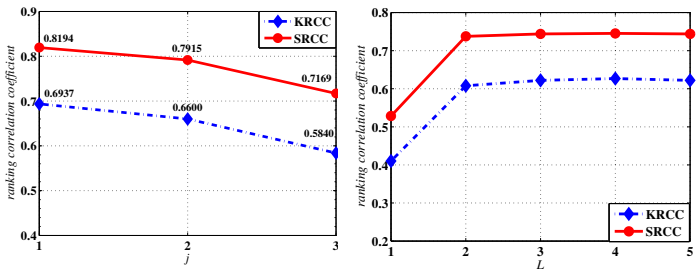

Fig. 2. Ranking correlation coefficient for (a) when naturalness is set to $P(M_j)$ and (b) sparse-domain similarity in terms of the sparse level L. The red solid and blue dashed lines indicate SRCC and KRCC curves, respectively.

C. Quality Prediction Model

The local sparse-domain similarity index described in Sec. III-A and global naturalness measurement introduced in Sec. III-B leads to the final SMTI model. Eventually, the overall weighting metric is defined as

$$\text{SMTI} = \lambda S^m + (1 - \lambda)R^n \tag{12}$$

where m and n are scale parameters to measure the sensitivity of two components, and $0{<}\lambda{<}1$ determines the weight of the relative importance. In the implementation, we choose $\lambda = 0.9012$, $m = 0.6983$ and $n = 0.1060$ as the final parameters.

IV. Experimental Results

To evaluate the performance of the proposed method, two criteria, which are Spearman rank-order correlation coefficient (SRCC) and Kendall rank-order correlation coefficient (KRCC), are employed under the public database [4]. A good IQA model is expected to achieve high SRCC and KRCC values.

Two experiments have been carried out to enhance the performance and save execution time. The first one is evaluating the influence on the SRCC & KRCC based on the variation of the sparse level L in Eqn. (4). Fig. 2(b) provides the ranking correlation coefficient. The performance increases and converges quickly after $L = 2$ and to the maximum at $L = 4$. This motivates us to adopt fewer primitives to avoid redundant calculations while preserving the relatively superior quality prediction performance.

The second experiment is discussing that the dictionary used in previous is adaptively learned for each HDR image. To further extend the generality, the global dictionary methodology is applied and examined. Specifically, the global dictionary is trained over all the natural images from LIVE database [18]. From Table I, it can be concluded that the global trained dictionary leads to slight performance decreases. However,

TABLE I. Average performance comparison of applying global and adaptive dictionary in terms of the KRCC and SRCC metrics.

	Global Dictionary	Adaptive Dictionary
KRCC	0.7461	**0.7508**
SRCC	0.8543	**0.8606**

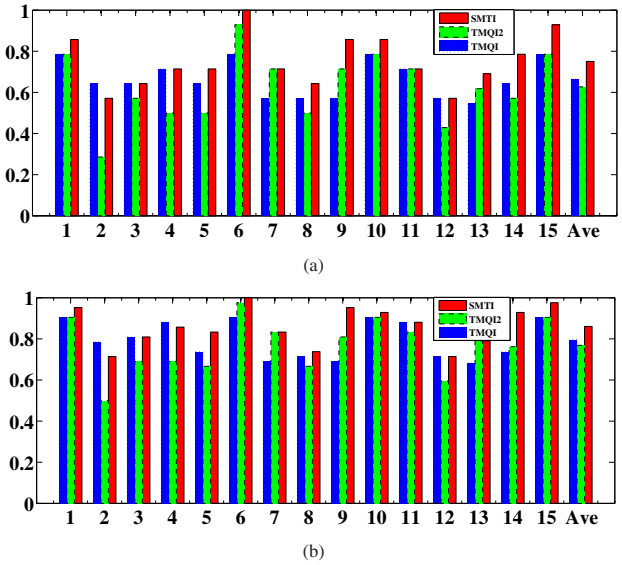

Fig. 3. Breakdown comparison of 15 test image groups among the proposed SMTI, TMQI and TMQI-II. The red one is for SMTI, blue one is for TMQI and green one is for TMQI-II. The last group shows the average scores.

the offline global dictionary training may further enable its applications in real scenarios.

In order to evaluate the overall performance of the proposed method, we compare our work with the TMQI [4] and TMQI-II [5]. The experimental results are given in Table II, indicating that the proposed SMTI method achieves higher performance compared to TMQI and TMQI-II. To demonstrate the robustness of SMTI, the breakdown prediction performance of SRCC & KRCC in 15 test groups and the average score is demonstrated in Fig. 3. It is shown that the SMTI outperforms the others in most of the cases.

TABLE II. PERFORMANCE COMPARISON OF SMTI, TMQI AND TMQI-II ON PUBLIC DATABASE [4].

	KRCC			SRCC		
	TMQI	**TMQI-II**	**SMTI**	**TMQI**	**TMQI-II**	**SMTI**
Average	0.6649	0.6269	**0.7508**	0.7963	0.7686	**0.8606**
Min	0.5455	0.2857	**0.5714**	0.6826	0.5000	**0.7143**
Max	0.7857	0.9286	**1.0000**	0.9048	0.9762	**1.0000**

V. CONCLUSION

In this paper, an objective image quality assessment method is proposed to evaluate the quality of tone-mapped images. The novelty of the paper lies in employing the sparse-domain coefficients to compare the cross dynamic range similarity and evaluate the naturalness. Firstly, with a dictionary trained from HDR images, the pristine HDR and tone-mapped LDR images are projected to the unified sparse space to compare the local sparse-domain similarity. Second, a novel naturalness measure is developed based on the features extracted from the coefficients in sparse representation. Experimental results show that the proposed method achieves higher correlation with the subjective data. In the future, TMO optimization algorithms will be further investigated based on the design philosophy of the proposed SMTI model.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (61322106, 61421062), National Basic Research Program of China (973 Program, 2015CB351800), and Shenzhen Peacock Plan, which are gratefully acknowledged.

REFERENCES

[1] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," in *Computer Graphics Forum*, vol. 22, no. 3, 2003, pp. 419–426.

[2] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, 2002, pp. 267–276.

[3] M. Barkowsky and P. Le Callet, "On the perceptual similarity of realistic looking tone mapped high dynamic range images," in *IEEE International Conference on Image Processing (ICIP)*, 2010, pp. 3245–3248.

[4] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, 2013.

[5] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3086–3097, 2015.

[6] X. Zhang, S. Wang, K. Gu, T. Jiang, S. Ma, and W. Gao, "Sparse structural similarity for objective image quality assessment," in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*,, 2015.

[7] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, 2014.

[8] J. Zhang, D. Zhao, R. Xiong, S. Ma, and W. Gao, "Image restoration using joint statistical modeling in a space-transform domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 915–928, 2014.

[9] S. Ma, X. Zhang, S. Wang, J. Zhang, H. Sun, and W. Gao, "Entropy of primitive: From sparse representation to visual information evaluation," *IEEE Trans. Circuits Syst. Video Technol.*, DOI: 10.1109/TCSVT.2015.2511838.(to appear).

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

[11] X. Zhang, R. Xiong, W. Lin, S. Ma, J. Liu, and W. Gao, "Video compression artifact reduction via spatio-temporal multi-hypothesis prediction," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6048–6061, 2015.

[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] X. Zhang, S. Wang, S. Ma, R. Xiong, and W. Gao, "Towards accurate visual information estimation with entropy of primitive," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1046–1049.

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

[15] 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.

[16] 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.

[17] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4613–4626, 2013.

[18] 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