Sparsity Based Stereoscopic Image Quality Assessment

Sameeulla Khan Md and Sumohana S. Channappayya, *Member, IEEE*

Abstract

In this work, we present a full-reference stereo image quality assessment algorithm that is based on the sparse representations of luminance images and depth maps. The primary challenge lies in dealing with the sparsity of disparity maps in conjunction with the sparsity of luminance images. Although analysing the sparsity of images is sufficient to bring out the quality of luminance images, the effectiveness of sparsity in quantifying depth quality is yet to be fully understood. We present a full reference Sparsity-based Quality Assessment of Stereo Images (SQASI) that is aimed at this understanding.

I. Introduction

Stereo images when fused together provides a visual experience of natural 3D perception. This motivates the use of stereo content in entertainment industry. In the future we expect the parallel growth of it alongside with 2D, which leads to increase volume of stereo data (both images and video). Then compression plays an immediate and important role in this volume management which may reduce the quality of stereo images. Stereo may also find the applications where 2D images were used such as transmission over noisy channel, watermarking etc., which degrades the perceptual quality. Hence this necessitates the assessment of stereoscopic image quality. One way to assess image quality is via subjective assessment, where subjects are asked to rate the quality of perceived stereo images. However, it is a very time consuming process and hence not used on large data volumes. Therefore, we need to come up with an objective quality metric which automatically rates the quality of stereoscopic images.

Objective quality metric can be classified into three types depends on the availability of reference data. They are (i) Full reference (FR) (ii) Reduced Reference (RR) and (iii) No reference (RR). In FR we have the reference image and its corresponding distorted image is assessed with respect to the reference image. In RR we have some attributes of reference image

based on that the distorted image should be quantified its quality. In NR we don't have any reference image. The distorted image is assessed based on a model which is trained to give a quality score. In this paper we discuss full reference stereoscopic image quality assessment (FRSIQA) using sparse representations of luminance and disparity maps. First we briefly review the latest FRSIQA methods.

In [1], Cardoso et al. present a disparity weighting technique for evaluating stereoscopic image quality evaluation. They consider the absolute difference between left and right images as disparity map and its pixel value is considered as weights in evaluating weighted difference between reference and test stereo pairs. Feifei et al. [2] proposed a cyclopean based approach where they combined three features from binocular disparity, gradient magnitude, and phase congruency into one quality score. They include both the binocular combination characteristic and disparity effect in their proposed stereoscopic image quality assessment (SIQA).

Galkandage et al. [3] proposed an FRSIQA method based on HVS modelling and regression analysis. They have a new definition of complex cells based on additions of complex cells behaviour to binocular suppression and implementation of recurrent excitation. The outputs of complex cell models are formed into a number of objective scores based on luminance, chrominance and decomposed sub bands of spatial frequencies. Then a statistical model is proposed to find a relationship between objective and subjective scores. In [4] we proposed a statistical model on natural stereo pairs where we describe about how statistical model brings out a way to estimate quality of stereo image. Also we studied the statistical relationship between luminance and disparity subbands which indirectly describes the role of structural content. In this paper we consider only disparity maps of reference stereo pairs and thus avoid the issues in estimating disparity maps of impaired stereo pairs. Zhang et al. applied their previous 2D-MAD [5] model to stereo called 3D-MAD [6] in two stages. Firstly they apply on both left and right images separately and their quality scores are linearly combined. In the second stage they compare the cyclopean feature images of reference and test stereo pairs. Final quality is estimated from the above two stages. In [7] we proposed another FRSIQA in which we concluded that heavy loss in structural information leads to significant loss in depth perception. With this structural motivation we use MS-SSIM [8] on both luminance and disparity pairs and pooled the scores. We surveyed some other methods which is based on sparsity and will be discussed in the next section.

II. Sparse Representation Of Natural Images

Several experimental studies have shown that neurons encode sensory information using a small number of active neurons [9], [10], [11]. Sparsity of images arose from the idea that neurons in the early stages of visual pathway are more active than latter stages. For example, neurons in retina responds to simple stimulus where as neurons in visual cortex only responds to certain stimulus like edges [12], [13]. For a given image, its sparsely encoding of neurons will be different with respect to its distorted image. Hence we opt for sparsity of images, whose change in sparse level will serve as base for quality assessment of images. Even if we fix the sparsity of images i.e, the sparsity of reference and distorted image is same, the variation in coefficients will help us to evaluate the quality of an image. With these principle several 2D image quality assessment (IQA) were proposed ([14], [15], [16], [17], [18], [19], [20], [21] and [22]). There are also some sparsity based SIQA methods. Shao et al. proposed FRSIQA [23] using using binocular receive field properties. They learnt multi-scale dictionaries from where they compute sparse feature similarity and global luminance similarity for luminance quality. They calculate sparse energy and sparse complexity as basis of binocular combination. Finally Qi et al. proposed a RR-SIQA [24], where they adopted information theory on sparse matrices of reference and distorted stereo pairs. Neither of these methods have considered disparity maps for evaluating depth quality.

III. Proposed Approach

In [4] and [7] we proposed two different approaches for FRSIQA. The key takeaway from these two works is that both image and depth quality depends on loss in structural information. To effectively study the structural content we focused on algorithms that predominantly work on structures of natural images. In a departure from these methods, we now propose a sparse representation based approach. Learnt dictionaries will lead to sparse data representation. Therefore, we choose sparsity of images and disparity maps for SIQA. The proposed approach is moderately different from [23] and very different from [24]. Our approach is described in the following stages.

A. Dictionary Learning

For a pristine stereo pair both images perceptually look similar, hence it is enough to consider a single dictionary for both left and right images. The same dictionary can be used by the

corresponding distorted stereo pairs. Similarly in the case of disparity maps, left and right pristine disparity maps and the corresponding distorted disparity maps use a single dictionary. Hence for each stereo pair we have two dictionaries, one refers to luminance and second refers to disparity. Fig 1 shows how dictionary is obtained by considering salient patches. Figs 1a and 1b are reference left image and its corresponding disparity map respectively. We take all possible 8×8 overlapping patches from image and disparity maps. Each patch is vectorized and concatenated to form a matrix. Form all the vector patches we consider top 3000 salient patches. Now these patches were subtracted from its mean value to have zero-mean patches. Now these patches are considered as inputs for dictionary construction. In Figs. 1c and 1d white pixels indicates the locations at which salient patches are considered. For images we choose the saliency as having high entropy value. In Fig. 1c the locations of salient patches are spreading across the image. It means that it is considering salient patches from different areas of image that are distinct from each other which is an extreme requirement for dictionary construction. On the other hand we consider variance as saliency for disparity maps. In Fig. 1d the salient patches are clustered around the edges of disparity maps. As mentioned in previous para edges and structural information are important for depth perception, therefore we consider variance as saliency for disparity maps. Figs 1e and 1f shows the learnt dictionaries for image and disparity maps respectively.

B. Sparsity

After obtaining the dictionary we then consider non-overlapping distinct patches from reference and test stereo pairs and vectorized it. These patches again transformed to zero-mean patches. Using advanced OMP [25] algorithm we obtain the sparse matrices of reference and test stereo pairs with respect to dictionary learned from patches of left reference image. By a sparse matrix we mean the ordered collection of sparse vectors each of which correspond to a non-overlapping image patch. Obtaining sparsity of disparity maps is slightly different. In image perception we believe that each region of image is important and hence we consider all non-overlapping distinct patches of images. As we know that disparity maps are representatives for depth perception and the sense of depth perception is observed at salient regions like edges and other structural information. The loss of depth perception can be measured if we check the variation of sparsity of disparity maps at these salient regions. Hence, in reference disparity pairs (left and right disparity maps) we consider non overlapping patches and vectorize them.

(a) Reference image.

(b) Disparity map.

(c) Patch locations in luminance image.

(d) Patch locations in disparity map.

(e) Dictionary for image.

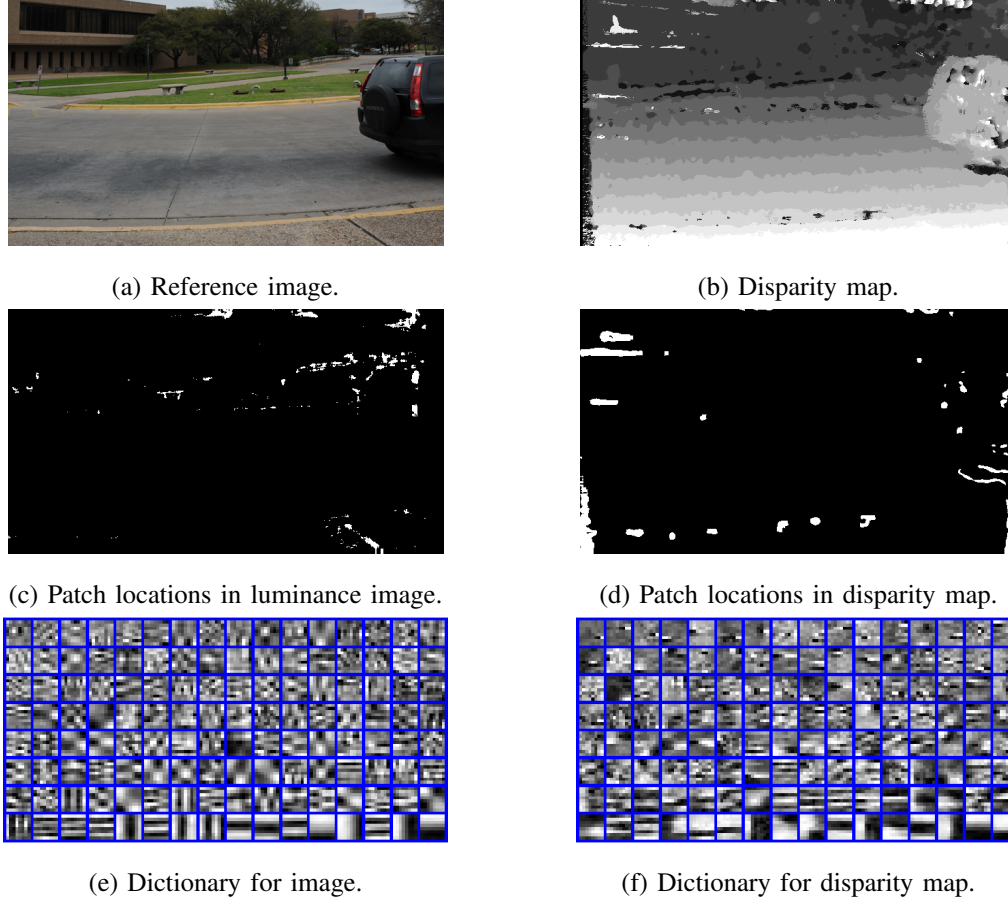(f) Dictionary for disparity map.

Fig. 1: Illustration of dictionary learning.

Then we take the top 3000 salient patches (saliency with respect to variance) and for distorted disparity pair we consider the same locations where we had taken the patches from reference. Later these patches are transformed to zero-mean patches. Now with these vector patches and the dictionary learnt from the patches of left disparity map, we obtain sparsity of reference and test disparity pairs.

C. Quality Estimation

Prediction of quality depends on the sparse matrices of reference and distorted stereo and disparity pairs. Let $\mathbf{X}_l^r$, $\mathbf{X}_r^r$, $\mathbf{X}_l^t$, and $\mathbf{X}_r^t$ be sparse matrices of reference and test stereo pairs. The superscript r and t indicates reference and test (distorted) and subscript l and r represents left and right respectively. Similarly let $\mathbf{D}_l^r$, $\mathbf{D}_r^r$, $\mathbf{D}_l^t$ and $\mathbf{D}_r^t$ be the sparse matrices of reference and test disparity pairs. We compare the reference and test sparse matrix of image (and disparity

map) column-wise as described next. The computation of luminance quality is slightly different from depth quality. In computing luminance quality we consider structural comparison and non structural comparison as implemented in [20].

Let $\mathbf{a}_r^i$ and $\mathbf{a}_t^i$ be the ith column vectors of $\mathbf{X}_l^r$ and $\mathbf{X}_l^r$ respectively. We compute structural similarity as given in (1).

$$\rho_l^i(\mathbf{a}_r^i,\ \mathbf{a}_t^i) = \frac{|\mathbf{a}_r^{i^T}\mathbf{a}_t^i| + k}{\|\mathbf{a}_r^i\|_2\ \|\mathbf{a}_t^i\|_2 + k}, \tag{1}$$

where k is a constant to avoid zero-by-zero condition. The non structural similarity can be obtained from (2).

$$\eta_l^i(\mathbf{a}_r^i,\ \mathbf{a}_t^i) = 1 - \left| \frac{\|\mathbf{a}_r^i\|_2 - \|\mathbf{a}_t^i\|_2 + k}{\|\mathbf{a}_r^i\|_2 + \|\mathbf{a}_t^i\|_2 + k} \right|. \tag{2}$$

Then the luminance quality of left image is given by

$$S_l(\mathbf{X}_l^r, \mathbf{X}_l^t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\rho_l^i\eta_l^i}, \tag{3}$$

where N is the total number of columns of sparse matrix. Similarly we also obtain luminance quality score for right image. Hence for a test stereo pair we have left and right quality scores as S_l and S_r respectively.

For computing depth quality scores it is slightly different from luminance score. Consider left reference and test disparity maps. As we know that edges and structures are responsible for depth perception, hence structural similarity need to be measured between reference and test disparity pairs. Therefore we retain (1) for structural similarity calculation. Let ρ_{dl}^i be the structural similarity obtained from (1). We know that disparity maps are not natural images hence for measuring non structural information we treat disparity as a normal data matrix. In [26], Song et al. propose several distance measures for comparing sparse vectors. One method is using radial basis function kernel defined in (4). This comparison is used for measuring non structural similarity between disparity maps.

$$\eta_{dl}^i(\mathbf{a}_r^i,\ \mathbf{a}_t^i) = \exp\left(\frac{-\|\mathbf{a}_r^i - \mathbf{a}_t^i\|_2^2 + k}{\|\mathbf{a}_r^i\|_2\ \|\mathbf{a}_t^i\|_2 + k} \right). \tag{4}$$

The depth quality with respect to left disparity maps is now estimated as given in (5).

$$S_{dl}(\mathbf{D}_l^r, \mathbf{D}_l^t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\rho_{dl}^i\eta_{dl}^i}. \tag{5}$$

Similarly the depth quality with respect to right disparity map is S_{dr}.

1) Stereo Quality Score: The left and right luminance and depth scores need to be combined to have a single luminance and depth scores. We empirically found that using weighted geometric combination of left and right scores yield better performance. The weights are obtained from mean square values of sparse matrices. Let m_l, m_r, m_{dl} and m_{dr} are mean square values of $\mathbf{X}_l^t, \mathbf{X}_r^t, \mathbf{D}_l^t$ and $\mathbf{D}_r^t$ respectively. The luminance and disparity weights are obtained as shown in (6) & (7) respectively.

$$w_l = \frac{m_l}{m_l + m_r} \quad ; \quad w_r = \frac{m_r}{m_l + m_r} \tag{6}$$

$$w_{dl} = \frac{m_{dl}}{m_{dl} + m_{dr}} \quad ; \quad w_{dr} = \frac{m_{dr}}{m_{dl} + m_{dr}} \tag{7}$$

From (6) & (7) the weighted geometric combination of left and right scores are given in (8).

$$S = S_l^{w_l}.S_r^{w_r} \quad ; \quad S_d = S_{dl}^{w_{dl}}.S_{dr}^{w_{dr}} \tag{8}$$

The overall quality of stereo image is given in (9)

$$Q = S.\sqrt{S_d}. \tag{9}$$

2) Settings: We would like to mention the settings we used in computing quality metric. The number of non zero elements in each patch is made 15 and 5 for image and disparity map respectively. The size of luminance and disparity map dictionaries are 64×128. The size of data matrix obtained from salient patches of image and disparity maps is of dimension 64×3000.
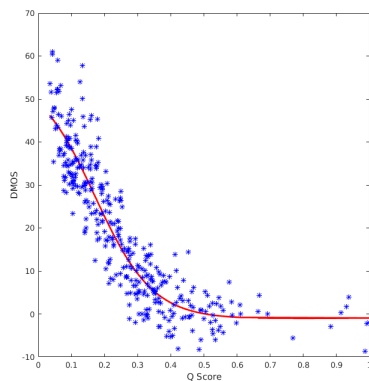
(a) Scatter plot of SQASI versus DMOS over LIVE-I.
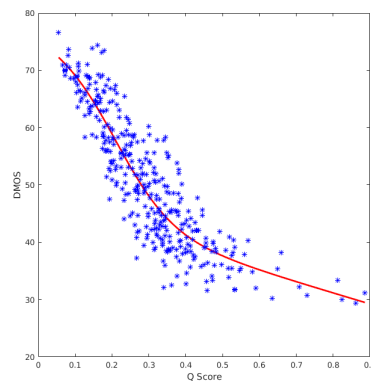
(b) Scatter plot of SQASI versus DMOS over LIVE-II.

Fig. 2: Scatter plots over the LIVE-I and LIVE-II databases.

TABLE I: DMOS vs SQASI

Distortion	LIVE Phase-I		LIVE Phase-II	
	LCC	**SROCC**	**LCC**	**SROCC**
WN	0.9250	0.9225	0.9277	0.9228
JP2K	0.9415	0.9067	0.8222	0.8025
JPEG	0.7444	0.6963	0.7978	0.7735
BLUR	0.9497	0.9184	0.9781	0.9431
FF	0.8502	0.7732	0.9415	0.9285
OVERALL	**0.9230**	**0.9238**	**0.8990**	**0.8898**
ASYMM	—	—	0.8693	0.8499
SYMM	—	—	0.9384	0.9346

TABLE II: DMOS vs SQASI

Database	**LCC**	**SROCC**	**RMSE**
IRCCYN	0.7224	0.6505	15.25
MICT	0.7358	0.7412	15.75
MICT_Asym	0.7236	0.7262	15.91
MICT_Sym	0.8929	0.8754	11.08

IV. RESULTS AND DISCUSSION

The proposed method is implemented on four databases namely LIVE Phase-I (LIVE-I) [27], LIVE Phase-II (LIVE-II) [28], IRCCYN [29] and MICT [30]. LIVE-I database consists of 20 pristine stereo pairs and 365 distorted stereo pairs. For each pristine stereo pair it has five different types of distortions namely White Noise (WN), JPEG2000 (JP2K), JPEG, Blur, Fast Fading (FF). Each distortion type has at most four different strengths of distortion levels. LIVE-II database consists of 8 pristine stereo pairs and 360 distorted stereo pairs. Like LIVE-I it also have the same five types of distortions and each distortion has nine different strengths of distortion levels. Out of 360 distorted stereo pairs in LIVE-II, 120 pairs are symmetrically distorted and 240 are asymmetric. IRCCYN has six pristine stereo pairs and for each pair it has 15 different types of distortions namely Gaussian blur, JPEG compression and downscale-upscale. Totally it has 90 distorted stereo pairs. MICT database consists of 10 pristine stereo pairs and a total of 480 JPEG compressed stereo pairs where each left and right image had 7 compression level and totally each pristine stereo pair had 48 impaired stereo pairs. Out of 480 stereo pairs 60 were

TABLE III: Comparison With Recent FRSIQA

(a) LCC Comparison

Algorithm	Database			
	LIVE-I	LIVE-II	IRCCYN	MICT
Khan [4]	0.9275	0.9019	0.8504	—
Khan [7]	0.9318	0.9313	0.8143	0.7725
Shao [23]	0.9350	0.8628	—	0.9355
Qi [24]	—	0.9150	—	—
Proposed	0.9230	0.8990	0.7224	0.7358

(b) SROCC Comparison

Algorithm	Database			
	LIVE-I	LIVE-II	IRCCYN	MICT
Khan [4]	0.0.9223	0.8920	0.8413	—
Khan [7]	0.9254	0.9323	0.7893	0.7681
Shao [23]	0.9251	0.8494	—	0.9391
Qi [24]	—	0.8670	—	—
Proposed	0.9238	0.8898	0.6505	0.7412

symmetrically compressed and 420 were asymmetrically compressed.

The performance analysis was carried out using standard measures namely Pearson linear correlation coefficient (LCC), Spearman rank order correlation coefficient (SROCC) and Root mean square error (RMSE). For better performance LCC and SROCC should be high (close to unity) and RMSE should be low. All scores are reported post logistic fitting. The performance of SQASI over LIVE-I and LIVE-II are shown in Table I. For IRCCYN and MICT database the performance of SQASI is shown in Table II. From Table I & II it proves that weighted geometric combination of left and right scores provides better efficiency especially in the case of asymmetric distortions. Figs 2a & 2b shows the scatter plot of SQASI over LIVE-I and LIVE-II database respectively.

Table IIIa & IIIb shows the comparison of SQASI with some of latest FRSIQA methods. We compare our algorithm with our previous work in FRSIQA i.e., [4], [7] and with already existing sparse based methods on FRSIQA i.e [23], [24]. Form tables IIIa & IIIb, we see that the proposed algorithm is competitive on the LIVE-I & LIVE-II databases. The performance on the MICT and IRCCYN databases however is sub-par.

V. CONCLUSION

In this paper we presented our preliminary work on sparsity based stereo image quality assessment. We built two dictionaries per stereo pair, one with respect to the luminance image and the other with respect to the disparity map. These dictionaries serve as representatives for that particular stereo pair and disparity pair, and also its distorted versions. We quantify the loss in luminance and depth quality by measuring the structural and non structural difference

between sparse matrices of image and disparity maps. The results are promising and we plan to explore this approach in greater depth in the future.

REFERENCES

[1] J. V. de Miranda Cardoso, C. D. M. Regis, and M. S. de Alencar, "Disparity weighting applied to full-reference and no reference stereoscopic image quality assessment," in *2015 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 477–480, Jan 2015.

[2] F. Li, L. Shen, D. Wu, and R. Fang, "Full-reference quality assessment of stereoscopic images using disparity-gradient-phase similarity," in *Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on*, pp. 658–662, July 2015.

[3] G. C, C. J, D. S. V, and D. S, "A full-reference stereoscopic image quality metric based on binocular energy and regression analysis," in *2015 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–5, July 2015.

[4] S. K. Md, B. Appina, and S. S. Channappayya, "Full-reference stereo image quality assessment using natural stereo scene statistics," *IEEE Signal Processing Letters*, vol. 22, pp. 1985–1989, Nov 2015.

[5] 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. 011006–011006, 2010.

[6] Y. Zhang and D. M. Chandler, "3d-mad: A full reference stereoscopic image quality estimator based on binocular lightness and contrast perception," *IEEE Transactions on Image Processing*, vol. 24, pp. 3810–3825, Nov 2015.

[7] S. K. Md and S. S. Channappayya, "Multiscale-ssim index based stereoscopic image quality assessment," in *Proc. of 22nd National Conference on Communications in IEEE*, March 2016.

[8] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, pp. 1398–1402, Ieee, 2003.

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

[10] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Current opinion in neurobiology*, vol. 14, no. 4, pp. 481–487, 2004.

[11] P. Reinagel, "How do visual neurons respond in the real world?," *Current opinion in Neurobiology*, vol. 11, no. 4, pp. 437–442, 2001.

[12] H. Barlow, "Single units and sensation: a neuron doctrine for perceptual psychology?," 1972.

[13] B. A. Olshausen, "Principles of image representation in visual cortex, in the visual neurosciences," 2003.

[14] L. He, D. Tao, X. Li, and X. Gao, "Sparse representation for blind image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1146–1153, June 2012.

[15] H. w. Chang, M. h. Wang, S. q. Chen, H. Yang, and Z. j. Huang, "Sparse feature fidelity for image quality assessment," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 1619–1622, Nov 2012.

[16] H. W. Chang, H. Yang, Y. Gan, and M. H. Wang, "Sparse feature fidelity for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 22, pp. 4007–4018, Oct 2013.

[17] 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, Dec 2013.

[18] K. V. S. N. L. M. Priya, K. Manasa, and S. S. Channappayya, "A statistical evaluation of sparsity-based distance measure (sdm) as an image quality assessment algorithm," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2789–2792, May 2014.

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

[20] 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, 2014.

[21] K. V. S. N. L. M. Priya and S. S. Channappayya, "A novel sparsity-inspired blind image quality assessment algorithm," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, pp. 984–988, Dec 2014.

[22] X. Zhang, S. Wang, K. Gu, T. Jiang, S. Ma, and W. Gao, "Sparse structural similarity for objective image quality assessment," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, pp. 1561–1566, Oct 2015.

[23] F. Shao, K. Li, W. Lin, G. Jiang, M. Yu, and Q. Dai, "Full-reference quality assessment of stereoscopic images by learning binocular receptive field properties," *IEEE Transactions on Image Processing*, vol. 24, pp. 2971–2983, Oct 2015.

[24] F. Qi, D. Zhao, and W. Gao, "Reduced reference stereoscopic image quality assessment based on binocular perceptual information," *IEEE Transactions on Multimedia*, vol. 17, pp. 2338–2344, Dec 2015.

[25] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *CS Technion*, vol. 40, no. 8, pp. 1–15, 2008.

[26] Y. Song and D. Roth, "Unsupervised sparse vector densification for short text similarity," *Proc. North Am. Chapter Assoc. Computat. Linguistics*, pp. 1275–1280, 2015.

[27] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013.

[28] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," *Signal Processing: Image Communication*, vol. 28, no. 9, pp. 1143–1155, 2013.

[29] B. Alexandre, L. C. Patrick, C. Patrizio, and C. Romain, "Quality assessment of stereoscopic images," *EURASIP journal on image and video processing*, vol. 2008, 2009.

[30] R. Akhter, Z. P. Sazzad, Y. Horita, and J. Baltes, "No-reference stereoscopic image quality assessment," in *IS&T/SPIE Electronic Imaging*, pp. 75240T–75240T, International Society for Optics and Photonics, 2010.