Multiscale-SSIM Index Based Stereoscopic Image Quality Assessment

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

Abstract

Stereoscopic image quality typically depends on two factors: i) the quality of the luminance image perception, and ii) the quality of depth perception. The effect of distortion on luminance perception and depth perception is usually different, even though depth is estimated from luminance images. Therefore, we present a full reference stereoscopic image quality assessment (FRSIQA) algorithm that rates stereoscopic images in proportion to the quality of individual luminance image perception and the quality of depth perception. The luminance and depth quality is obtained by applying the robust Multiscale-SSIM (MS-SSIM) index on both luminance and disparity maps respectively. We propose a novel multi-scale approach for combining the luminance and depth scores from the left and right images into a single quality score per stereo image. We also explained that a small amount of distortion does not significantly affect depth perception. Further, heavy distortion in stereo pairs will result in significant loss of depth perception. Our algorithm performs competitively over standard databases and is called the 3D-MS-SSIM index.

Index Terms

Stereoscopic images, full-reference image quality assessment, depth perception, MS-SSIM.

I. Introduction

Stereoscopic content creation and rendering has a long and interesting history [1]. The earliest known stereoscopic camera (Fig.1) was invented by David Brewster in 1844. Since this invention, stereoscopic technology has come a long way and is now a standard medium for content creation and consumption in the movie and gaming industry. As the demand for 3D applications are increasing, there is a necessity to manage the huge volumes of 3D content being generated. Perceptual quality assessment then plays a crucial role in data management. Perceptual quality assessment is typically classified into three categories. i) Full reference (FR), ii) Reduced reference (RR) and iii) No reference (NR). In FR the quality of test stereo pair

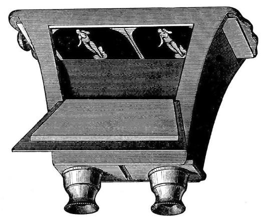

Fig. 1: The Brewster stereoscope. Image credit Wikicommons.

is evaluated in comparison with reference stereo pair, in RR we have partial information about reference stereo pair and in NR the quality of test stereo pairs are evaluated with out any reference stereo pairs.

In this paper, our focus is limited to full reference quality assessment. Literature suggests that 3D image quality assessment (IQA) methods can be grouped into two categories based on whether or not depth/disparity information is considered. Earlier methods for 3D IQA, with out disparity, were done by simply adopting conventional 2D full reference image quality assessment (FRIQA) methods on left and right images and pooled the scores by different processes. Campisi et al. [2] carried out a systematic study of stereoscopic image quality. 2D FRIQA methods were applied to left and right images and it was concluded that 2D FRIQA does not necessarily perform well on stereo images. Benoit et al. [3] explicitly considered disparity information in their analysis. They combined existing 2D FRIQA methods along with local and global disparity measure to come up with an objective quality metric. While our work is similar in approach to this paper, it differs significantly in the perceptual underpinnings. The cyclopean paradigm proven to be promising in FR stereo IQA. Maalouf et al. [4] proposed a RR 3D IQA where cyclopean images are constructed and human visual system (HVS) sensitivity coefficients of reference and test stereo pairs were compared to determine stereo image quality. Chen et al. [5] proposed a cyclopean model where existing 2D FRIQA methods were used to measure the quality of cyclopean image.

Other recent methods for stereo FRIQA algorithms have considered different approaches. Bensalma et al. [6] proposed an algorithm based on difference of binocular energy between reference and tested stereo pairs. Shao et al. [7] proposed a stereo FRIQA algorithm by considering the binocular visual characteristics. The major contribution of this paper is that the binocular perception and combination properties are considered in quality assessment. Wang et al. [8] proposed a subjective quality assessment on a database. Their results suggests that simply averaging the quality of left and right images predicts the quality of symmetrically distorted stereo images well, but generates substantial bias in the case of asymmetric distortions. Wang et al. [9] proposed a 3D gradient magnitude based stereo FRIQA where they calculate pointwise 3D gradient magnitude similarity (3D-GMS) along horizontal, vertical and viewpoint directions and quality score is obtained by averaging the 3D-GMS scores of all points. Lin et al. [10] present a model where they utilized binocular combination and binocular frequency integration for measuring the quality of stereoscopic images. Khan et al. [11] proposed a full reference metric for stereoscopic images based on the statistical modeling of luminance and disparity. They modeled marginal statistics of luminance and disparity using a univariate GGD model. In their approach, they considered the disparity maps only for reference stereoscopic images.

II. Proposed Approach

The frame work for the proposed approach is shown in Fig.2. The proposed approach is based upon the intuition that stereoscopic image quality assessment will depend upon two factors: i) perceptual annoyance of left and right luminance images, and ii) loss in depth perception. Perceptual annoyance in luminance can be evaluated by applying conventional 2D IQA techniques on both the left and right luminance images. Loss in depth perception can be quantified by measuring the dissimilarity of distorted disparity map w.r.t the original disparity map. Our algorithm is described in the following subsections.

A. *Luminance Quality Evaluation*

In early work of stereoscopic image quality assessment (SIQA), conventional 2D IQA methods were used without considering disparity information. This approach partially achieved their goals. This partial success led us to believe that 3D IQA will depend on the individual quality of left and right images, but not completely so. To assess the quality of luminance images we rely on the robust MS-SSIM [12] index. The MS-SSIM index is applied to both the left and right images. Therefore we come up with two quality scores per stereo pair. The luminance quality scores are obtained as follows:

$$s_l^i = Q(I_l^o, I_l^t) \quad ; \quad s_r^i = Q(I_r^o, I_r^t), \tag{1}$$

where I_l^o, I_r^o are the reference stereo pair, I_l^t, I_r^t are the test stereo pair and Q indicates the MS-SSIM index. The subscripts l and r indicates left and right, the superscripts o and t indicates original (reference) and test (distorted) and the superscripts i refers to luminance image.

B. Depth Quality Evaluation

It has been shown in the literature [2] [5] that disparity plays an important role in stereo quality evaluation, and especially so in the case of asymmetric distortions. We describe our depth quality perception strategy next. We first compute disparity map using SSIM-based disparity estimation algorithm which was also used in [5]. The strength of dissimilarity of distorted disparity map with respect to the original disparity map will quantify the loss in depth perception. We claim that loss in structural information of luminance images will affect depth perception. This is because the sense of depth can be perceivable at the edges, lines and other structural features. With the loss in structural information, the depth perception is also reduced. This intuition can also be supported with the fact that complex cells, where binocular disparities are measured in the HVS, will respond primarily to oriented edges and gratings [13] [14]. Also disparity map retains the same structural information from both luminance images, it would be worthwhile to evaluate the loss in structural information from disparity maps. Therefore, to quantify the loss in depth perception we need to measure the amount of loss in structural information of distorted disparity map w.r.t the reference disparity map. The aforementioned reasons motivate us to use the MS-SSIM [12] metric to measure the dissimilarity of disparity map of distorted stereo pair w.r.t the reference disparity map. We compute the disparity maps w.r.t both the left and right luminance images (for both reference and distorted stereo pair), and therefore have two scores for each distorted stereo pair. The depth quality scores from disparity maps are obtained as follows:

$$s_l^d = \sqrt{Q(D_l^o, D_l^t)} \quad ; \quad s_r^d = \sqrt{Q(D_r^o, D_r^t)}, \tag{2}$$

where D_l^o, D_r^o are the reference disparity maps, D_l^t, D_r^t are the test disparity maps and the super script d refers to disparity map. We observed that for most stereo pairs, small amounts of distortion doesn't significantly affect depth perception. The loss in depth perception is perceivable only when the distortion is high enough to degrade the structural information. Since disparity estimation algorithms are sensitive to noise, the resultant disparity maps estimated for test stereo pairs are noisy. This affects the value of MS-SSIM index and therefore, the resultant value

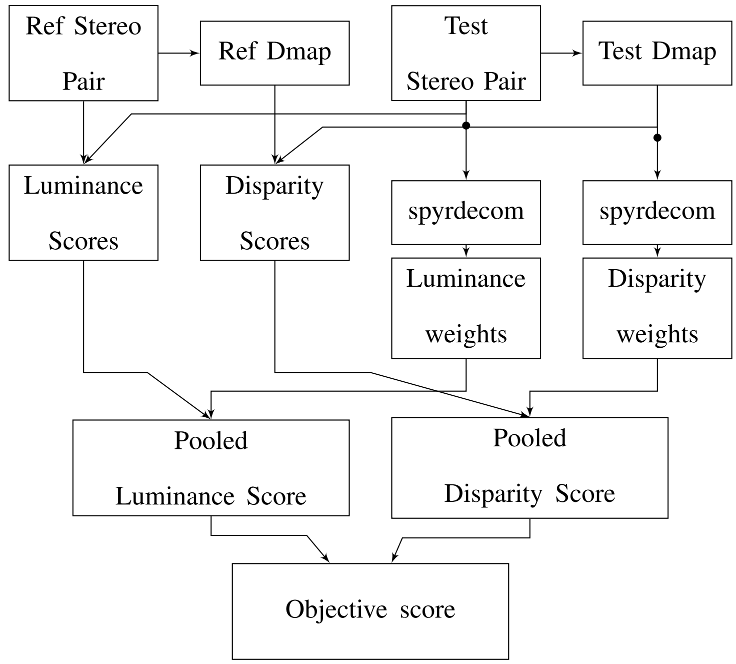

Fig. 2: Flowchart of proposed algorithm.

should be enhanced. We would like to recall that the MS-SSIM indices lie in the range [0, 1]. Therefore, to accurately capture the effects of distortion on disparity map, we apply the square root operator on disparity scores.

C. Stereo Quality Evaluation

We consolidate the left and right scores obtained from previous subsections into a single score of luminance and disparity. This consolidation is inspired from the fact that binocular strength is a convex combination of monocular stimulus strength [15]. The weights in the convex combination depends on dominance of one eye over other and this is particularly visible in the case of asymmetric distortions. Similarly we believe that the weights used in consolidation are also convex weights. These weights are computed from monocular strengths of stereo pairs. To compute convex weights, we need to adapt the strategy that comes from the theories of human visual system (HVS). So we use multi-scale multi-orientation steerable pyramid decomposition (spyrdecom) [16] [17] whose space scale orientation models the band pass filtering that occurs in area V1 of primary visual cortex [18]. We believe that the monocular strengths of each luminance image can be obtained from mean square (MS) of subband of spyrdecom. The convex weights are obtained as follows. Firstly we perform one scale six orientations $(0^0, 30^0, 60, 90^0, 120^0, 150^0)$

spyrdecom. As mentioned in [19] [20], to be consistent with luminance perception in the HVS, the pyramid decomposition is performed on the logarithm of distorted stereo pair and disparity maps. After spyrdecom, per each stereo pair we obtain six subbands. We compute MS of each subband as given in (3).

$$\mathbf{q}(j) = \frac{1}{|\mathbf{X}_j|}\sum_{\forall m,n}(\mathbf{X}_j(m,n))^2, \;\; j = 1,..,6. \tag{3}$$

Where $\mathbf{X}_j$ indicates j^{th} subband and $|\mathbf{X}_j|$ indicates its cardinality. Hence for six subbands we have six length MS vector. For distorted stereo pair and disparity maps we have the following MS vectors.

- $\mathbf{q}^i_{lt}$, $\mathbf{q}^i_{rt}$, $\mathbf{q}^d_{lt}$, $\mathbf{q}^d_{rt}$ are the 6 length MS vectors for distorted left and right luminance and disparity maps respectively.

We compute mean values of MS vectors as follows:

$$e^i_l = \frac{1}{6}\sum_{j=1}^{6}\mathbf{q}^i_{lt}(j) \;\; ; \;\; e^i_r = \frac{1}{6}\sum_{j=1}^{6}\mathbf{q}^i_{rt}(j), \tag{4}$$

$$e^d_l = \frac{1}{6}\sum_{j=1}^{6}\mathbf{q}^d_{lt}(j) \;\; ; \;\; e^d_r = \frac{1}{6}\sum_{j=1}^{6}\mathbf{q}^d_{rt}(j). \tag{5}$$

From the mean values obtained in (4) & (5) we compute scalar weights shown in (6) & (7). These weights facilitate the convex combination of scores obtained in (1) & (2).

$$E^i_l = \left(\frac{e^i_l}{e^i_l + e^i_r}\right) \quad ; \quad E^i_r = \left(\frac{e^i_r}{e^i_l + e^i_r}\right), \tag{6}$$

$$E^d_l = \left(\frac{e^d_l}{e^d_l + e^d_r}\right) \quad ; \quad E^d_r = \left(\frac{e^d_r}{e^d_l + e^d_r}\right). \tag{7}$$

Therefore, left and right scores of test luminance pairs and disparity maps are consolidated as shown in (8) and (9):

$$S^i = E^i_l s^i_l + E^i_r s^i_r, \tag{8}$$

$$S^d = E^d_l s^d_l + E^d_r s^d_r. \tag{9}$$

The overall objective quality score of the stereo image is empirically obtained as:

$$S = S^i\sqrt{S^d}. \tag{10}$$

TABLE I: DMOS vs 3D MS-SSIM

	LIVE Phase-I		LIVE Phase-II	
Distortion	**LCC**	**SROCC**	**LCC**	**SROCC**
WN	0.9502	0.9430	0.9633	0.9572
JPEG2000	0.9362	0.8987	0.8728	0.8537
JPEG	0.7143	0.6576	0.9057	0.8775
BLUR	0.9445	0.9344	0.9752	0.9233
FF	0.8307	0.7624	0.9320	0.9164
OVERALL	**0.9318**	**0.9254**	**0.9313**	**0.9323**
ASYMM	–	–	0.9217	0.9183
SYMM	–	–	0.9372	0.9269

III. Results And Discussion

For the performance evaluation of the proposed algorithm, we used the LIVE 3D Image Quality Assessment database – both Phase-I [21], and Phase-II [5], [22], the IRCCYN [3] and MICT [23] database. LIVE Phase I & II databases spans five distortion categories – JPEG and JPEG2000 compression, additive white Gaussian noise (WN), Gaussian blur (Blur) and Rayleigh fast fading channel distortion (FF). These databases represent subjective quality as the difference of mean opinion score (DMOS) associated with each of its distorted stereo pairs. Phase-I consists of 20 reference stereo pairs with 365 distorted pairs. Each distortion type consists of 4 levels of distortion strengths. Phase-II consists of 8 reference stereo pairs with 360 distorted stereo pairs, which classified into asymmetric(240 stereo pairs) and symmetric (120 stereo pairs) distortions. Each distortion type consists of 9 levels of distortion strengths. The IRCCYN database consists of 6 stereo pairs each of 15 distortions. It includes blur and JPEG2000 distortions. The MICT stereo image database has 480 JPEG distorted stereo images and 10 pristine stereo images. It includes mostly asymmetrically distorted stereo images. The objective quality scores are fitted to subjective scores (DMOS) using a standard logistic function. All scores are reported post logistic fitting.

Table I shows the performance of proposed approach on LIVE (Phase-I & II) database. It shows how the proposed approach works on different distortions in the LIVE databases and also shows the performance on asymmetric distortions specified in LIVE-II. Table II shows the performance on MICT and IRCCYN databases. Table III shows the comparison of 3D MS-SSIM

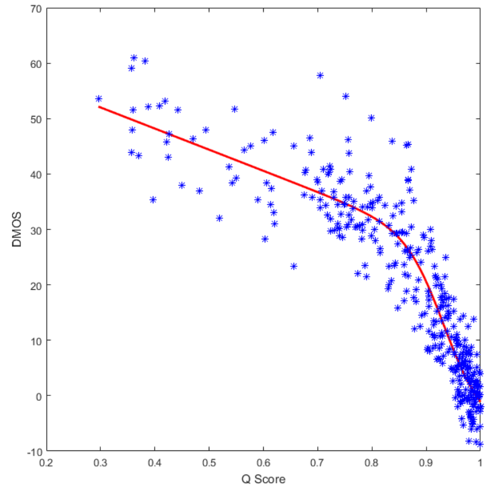

Fig. 3: Scatter plot of 3D-MSSSIM versus DMOS over LIVE-I.

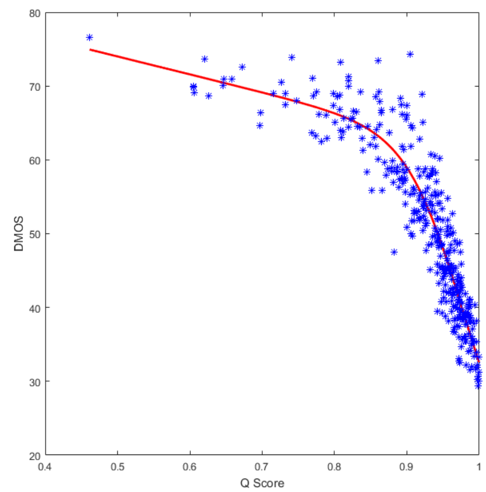

Fig. 4: Scatter plot of 3D-MSSSIM versus DMOS over LIVE-II.

with existing state-of-art algorithms over LIVE-II database. Table IV compares the performance of our algorithm on asymmetric distortions with respect to the state-of-the-art methods over LIVE-II database. The performance was measured using standard measures: Spearman's rank order correlation coefficient (SROCC), Pearson's linear correlation coefficient (LCC) and root-mean-squared error (RMSE) between predicted scores and the DMOS scores. The values are reported after logistic regression. Higher SROCC and LCC indicates good correlation with human scores, while lower values of RMSE indicate better performance. Figs. 3 and 4 shows the scatter plots of the proposed approach on LIVE-I & LIVE-II databases respectively. The major advantage

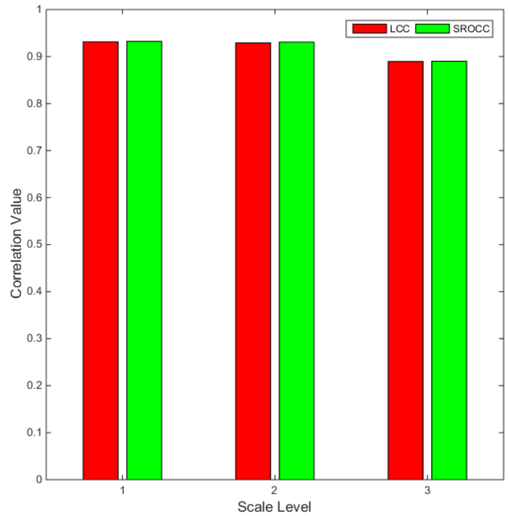

Fig. 5: Performance of 3D-MSSSIM as a function of number of scales over LIVE-II.

of our algorithm is low complexity, where the complexity lies only in obtaining convex weights.

We also studied the performance of our algorithm at 2nd and 3rd scales over LIVE Phase-II database and is shown in Fig. 5. From the bar plot it is clear that the better performance is obtained in the first scale itself thus minimizing the computational complexity.

While our method is similar in approach to the method proposed by Benoit et al. [3], there are important differences that we bring out in the following. Benoit et al. do not use conventional 2D IQA methods for measuring distortions in disparity maps since they are of the opinion that distortion maps are not natural images. We apply the MS-SSIM index on the disparity map since we believe that it has the ability to measure changes to structural information in the disparity maps. Next, in their pooling strategy, they simply average the left and right luminance scores while we propose a convex combination. This convex combination leads to better performance on asymmetrically distorted images. Further, their method considers only one disparity map (only w.r.t left), but we consider two disparity maps to better capture asymmetric distortions. We compare our method with their in Table V. It is clear that the proposed method outperforms their method on all but one databases. For the reasons mentioned above, our method performs well on the LIVE Phase-II and MICT databases that contain a majority of asymmetrically distorted images. Therefore, the proposed approach is particularly suited for the cases where we find asymmetric distorted stereo pairs.

TABLE II: DMOS vs 3D MS-SSIM

Database	LCC	SROCC	RMSE
IRCCYN	0.8143	0.7893	12.80
MICT	0.7725	0.7681	14.77

TABLE III: Comparison with state-of-the-art over LIVE Phase-II.

Algorithm	LCC	SROCC	RMSE
Chen [5]	0.901	0.893	10.58
Lin [10]	0.900	0.889	–
Fezza [24]	0.938	0.930	5.55
Khan [11]	0.902	0.892	4.87
Proposed	**0.931**	**0.932**	**4.11**

TABLE IV: Comparison with state-of-the-art on asymmetric distortions over LIVE Phase-II database.

Algorithm	LCC	SROCC	RMSE
Bensalma [6]	0.766	0.721	6.51
Shao [7]	0.609	0.630	8.03
Chen [5]	0.879	0.854	7.93
Wang [9]	0.728	0.695	6.92
Fezza [24]	0.915	0.920	5.95
Khan [11]	0.889	0.868	4.64
Proposed	**0.921**	**0.918**	**3.93**

IV. Conclusions and Future Work

In this work, we presented a perceptually inspired stereo image quality assessment algorithm called 3D MS-SSIM. Our algorithm emphasises that stereo FRIQA can rely on 2D IQA of the luminance images. At the same time, we also showed that heavy loss of structural information in stereo pairs results in significant loss in depth perception. Loss in depth perception is quantified by measuring dissimilarity of distorted disparity maps w.r.t reference maps. We first obtain the luminance and disparity scores using the MS-SSIM index and propose a perceptually inspired pooling method to arrive at a single score per stereo image pair. Our algorithm is competitive with the state-of-the-art methods on standard databases and is particularly effective on asymmetric distortions. As future work, we plan on extending this method to stereo video quality assessment.

TABLE V: Comparison of 3D MS-SSIM with Benoit [3]

	Benoit		**3D MS-SSIM**	
Database	**LCC**	**SROCC**	**LCC**	**SROCC**
LIVE Phase-I	0.8946	0.8975	0.9318	0.9254
LIVE Phase-II	0.7571	0.7291	0.9313	0.9323
IRCCYN	0.8459	0.8444	0.8143	0.7893
MICT	0.6364	0.6353	0.7725	0.7681

REFERENCES

[1] "History of 3D Technology, http://www.visionnw.com/history-of-3d-technology.html."

[2] P. Campisi, P. Le Callet, and E. Marini, "Stereoscopic images quality assessment," in *Proceedings of 15th European Signal Processing Conference (EUSIPCO07)*, 2007.

[3] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP journal on image and video processing*, vol. 2008, 2009.

[4] A. Maalouf and M.-C. Larabi, "Cyclop: A stereo color image quality assessment metric," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 1161–1164, IEEE, 2011.

[5] 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.

[6] R. Bensalma and M.-C. Larabi, "A perceptual metric for stereoscopic image quality assessment based on the binocular energy," *Multidimensional Systems and Signal Processing*, vol. 24, no. 2, pp. 281–316, 2013.

[7] F. Shao, W. Lin, S. Gu, G. Jiang, and T. Srikanthan, "Perceptual full-reference quality assessment of stereoscopic images by considering binocular visual characteristics.," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1940–1953, 2013.

[8] J. Wang and Z. Wang, "Perceptual quality of asymmetrically distorted stereoscopic images: the role of image distortion types," in *Proc. International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2014)*, pp. 29–31.

[9] S. Wang, F. Shao, F. Li, M. Yu, and G. Jiang, "A simple quality assessment index for stereoscopic images based on 3d gradient magnitude," *The Scientific World Journal*, vol. 2014, 2014.

[10] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3d image compression by binocular integration behaviors," *Image Processing, IEEE Transactions on*, vol. 23, pp. 1527–1542, April 2014.

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

[12] 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.

[13] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat," *Journal of neurophysiology*, vol. 28, no. 2, pp. 229–289, 1965.

[14] B. A. Wandell, *Foundations of vision.* Sinauer Associates, 1995.

[15] W. J. Levelt, *On binocular rivalry*, vol. 2. Mouton The Hague, 1968.

[16] E. P. Simoncelli and W. T. Freeman, “The steerable pyramid: A flexible architecture for multi-scale derivative computation,” in *Image Processing, International Conference on*, vol. 3, pp. 3444–3444, IEEE Computer Society, 1995.

[17] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 587–607, 1992.

[18] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *JOSA A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[19] Y. Liu, L. Cormack, and A. Bovik, “Statistical modeling of 3-d natural scenes with application to bayesian stereopsis,” *Image Processing, IEEE Transactions on*, vol. 20, pp. 2515 –2530, sept. 2011.

[20] D. Field, “What is the goal of sensory coding?,” *Neural computation*, vol. 6, no. 4, pp. 559–601, 1994.

[21] 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.

[22] M.-J. Chen, L. K. Cormack, and A. C. Bovik, “No-reference quality assessment of natural stereopairs,” *Image Processing, IEEE Transactions on*, vol. 22, no. 9, pp. 3379–3391, 2013.

[23] 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.

[24] S. Fezza and M.-C. Larabi, “Stereoscopic 3d image quality assessment based on cyclopean view and depth map,” in *Consumer Electronics ??? Berlin (ICCE-Berlin), 2014 IEEE Fourth International Conference on*, pp. 335–339, Sept 2014.