A VISUAL COMFORT ASSESSMENT METRIC FOR STEREOSCOPIC IMAGES

Feng Qi, Xiaopeng Fan, Debin Zhao

School of Computer Science and Technology, Harbin Institute of Technology

Tingting Jiang, Jian Zhang

School of Electronics Engineering and Computer Science, Peking University

ABSTRACT

Recent studies have shown that one of the main reasons inducing visual discomfort is accommodation-vergence conflict. To evaluate visual discomfort induced by this conflict, this paper proposes a stereo visual comfort assessment (SVCA) metric based on the measurement of accommodation and vergence for stereoscopic image. Here, accommodation corresponds to the monocular focusing process which is modeled by two-view images' joint entropy; vergence corresponds to binocular fusion which is modeled by two-view images' mutual information. The joint entropy and mutual information are calculated by the visual primitives extracted from two-view images. In this paper, accommodation-vergence conflict is expressed as the ratio of the mutual information over joint entropy. To evaluate the proposed metric, a subjective experiment is conducted to construct a ground truth database. The experimental results show that the proposed SVCA metric achieves a highly competitive performance with some state-of-the-art SVCA models.

Index Terms**—**Visual discomfort, stereoscopic visual comfort assessment, accommodation-vergence conflict, visual perceptual information, entropy of primitive.

1. INTRODUCTION

Stereoscopic visual comfort assessment (SVCA) becomes a completely novel problem for certain physiological symptoms, such as visual discomfort, eyestrain, headache, and dizziness. These symptoms potentially hamper the popularity of 3D applications. Similar with traditional 2D image quality assessment, SVCA can be classified into subjective evaluation and objective metric.

Subjective evaluation is a psychophysical method that tests whether subjects experience discomfort or fatigue symptoms such as eyestrain, double or blurred vision and headache when watching certain types of stereoscopic images or videos. Researchers have found that several factors may induce visual discomfort, including excessive screen disparity [1], accommodation-vergence conflict [2][3], binocular asymmetry [4], vertical disparities [5], and crosstalk artifacts [6]. Because human eyes are the ultimate receiver of stereoscopic images, subjective experiment is regarded as the most reliable way to evaluate stereoscopic image quality. In this paper, a subjective experiment is conducted to construct a ground truth database.

Compared with subjective evaluation, objective metric has lots of advantages, such as easy operation, inexpensive and good embedment in image processing algorithms. In objective metric, visual comfort is predicted by quantitative measurement, in which stereoscopic image analysis is usually performed to quantify comfort. Based on the research findings in the subjective experiments [1-6], several objective SVCA metrics have considered the founded discomfort induced factors. For excessive screen disparity factor, Yong *et al.* [7][8], Kim *et al.* [9] and Nojiri *et al.* [10] investigated the relationship between the disparity distribution and visual comfort for stereoscopic image or video. For accommodation-vergence conflict factor, Park *et al.* [11] proposed a 3D accommodation-vergence mismatch predictor algorithm using local 3D bandwidth. For binocular asymmetry factor, Yano *et al.* [12] detected visual discomfort image scenes based on the correlation of left and right images. For crosstalk artifact factor, Xing *et al.* [6] proposed an objective quality metric for predicting crosstalk perception by combining the structural similarity map and a filtered depth map. In addition, Jung *et al.* [13], Sohn *et al.* [14], Ide *et al.* [15] and Jones *et al.* [16] used their proposed SVCA metrics to lessen visual discomfort. Since the accommodation-vergence conflict is one of the main reasons inducing discomfort [2][3], this paper focuses on the prediction of stereoscopic visual discomfort induced by this conflict.

As oculomotor cues, accommodation, a monocular cue, refers to the variation of the lens shape and thickness (and thus its focal length), which allows the eye to focus on an object at a certain distance. Vergence, a binocular cue, refers to the muscular rotation of the eyeballs, which is used to converge both eyes on the same object [17]. Thus, accommodation corresponds to focusing process, which maintains clear image viewing; while vergence corresponds to fusion process, wherein the binocular retina images fuse into one perceptive image. In the proposed SVCA metric, accommodation is represented as the clarity of two-view images and is measured by their visual perceptual information which is modeled by joint entropy; vergence is represented as the fusion of visual perceptual information between two-view images which is modeled by mutual information. The joint entropy and mutual information are calculated by the visual primitives extracted from two-view images with a dictionary learning algorithm. Then, we express accommodation-vergence conflict as the ratio of the mutual information over joint entropy. To evaluate the proposed metric, a subjective experiment is conducted to construct a ground truth database.

The remaining parts of the paper are organized as follows. Section 2 describes the framework of the proposed visual perceptual information-based SVCA metric. Section 3 describes the subject experiment. Section 4 provides the experimental results and Section 5 concludes the paper.

2. VISUAL PERCEPTUAL INFORMATION-BASED SVCA METRIC

The framework of the proposed visual perceptual information-based SVCA metric is shown in Fig. 1. For a given stereoscopic image, the visual primitives are extracted from two-view images. Then, to measure the accommodation and vergence of the stereoscopic image, the visual primitives are used to calculate the joint entropy and mutual information. Finally, the comfort score is computed by the ratio of the mutual information over joint entropy.

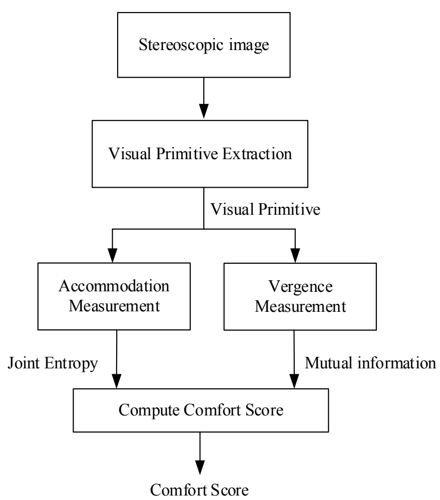

Fig. 1. Framework of the proposed SVCA metric.

2.1. Visual Primitive Extraction

Sparse representation is usually used to obtain a compact representation from the observed signal. The compact representation of an image can efficiently describe the perceived signal information. In this subsection, we employ a dictionary learning algorithm to extract visual primitives from two-view images.

In sparse representation, for a given image, the basic unit of sparse representation is the patch. The vector representations of the original image and an image patch of size $\sqrt{B_s} \times \sqrt{B_s}$ at location $i, i = 1, 2, ..., n,$ are mathematically denoted by $X \in \mathbb{R}^N$ and $\boldsymbol{x}_i \in \mathbb{R}^{B_s}$, where N is the number of the image vectors, B_s is size of each patch vector, and n is the number of patches in an image [18]. Then, we have

$$\boldsymbol{x}_i = R_i X, \tag{1}$$

where $R_i \in \mathbb{R}^{B_s \times N}$ is a matrix operator that extracts patch $\boldsymbol{x}_i$ from X. Note that patches are usually overlapped, and such patch-based representation is highly redundant. Therefore, the recovery of X from $\{\boldsymbol{x}_i\}$ becomes an over-determined system, from which it is straightforward to obtain by the least-square solution [19],

$$X = \left(\sum\nolimits_{i=1}^{n} R_i^T R_i\right)^{-1} \sum\nolimits_{i=1}^{n} \left(R_i^T \boldsymbol{x}_i\right), \tag{2}$$

which is nothing but an abstraction strategy of averaging all the overlapped patches.

Given patches $\{\boldsymbol{x}_i\}$, the purpose of dictionary learning is to search the best possible dictionary to sparsely represent $\{\boldsymbol{x}_i\}$. The dictionary learning process is formulated as:

$$\boldsymbol{D}, \{\boldsymbol{a}_i\} = \underset{\boldsymbol{D}, \{\boldsymbol{a}_i\}}{\operatorname{argmin}} \sum_i \left\|\boldsymbol{x}_i - \boldsymbol{D}\boldsymbol{a}_i\right\|_2^2 \text{ s.t. } \left\|\boldsymbol{a}_i\right\|_0 < T, \tag{3}$$

where $D \in \mathbb{R}^{B_s \times M}$ is an over-complete dictionary matrix that contains M visual primitives of an image as columns. Here, M is set to 256. The vector $\boldsymbol{a}_i \in \mathbb{R}^M$ contains the representation coefficients for each patch. $\left\|\boldsymbol{a}_i\right\|_0$ is the ℓ_0 norm, which counts the nonzero entries of the vector $\boldsymbol{a}_i$, and T is the constraint of the non-zero number of $\boldsymbol{a}_i$ and is set to 3.

Generally, Eq. (3) can be solved by the K-SVD algorithm [20].

2.2. Measurement of Accommodation and Vergence

Although the two views' image are highly similar in natural viewing, our eyes can still perceive their slight differences and fuse them into a 'cyclopean' 3D image with accommodation and vergence. Inspired by the concept of entropy of primitive [29], we quantify the visual information for stereoscopic image which is used to represent the measurement of accommodation and vergence.

The extracted visual primitives are the basic visual perceptual elements of an image. A set of visual primitives ($\boldsymbol{D}$) approximates the stereoscopic image I as following:

$$I = \left(\sum\nolimits_{i=1}^{n} R_i^T R_i\right)^{-1} \sum\nolimits_{i=1}^{n} \left(R_i^T \boldsymbol{D}\boldsymbol{a}_i\right), \tag{4}$$

where $\boldsymbol{a}_i$ denotes the *i-th* column of $\boldsymbol{A}$, $\boldsymbol{A}$ is the coefficient matrix of $\boldsymbol{D}$. n is the number of patches in I. Assume $\boldsymbol{d}_k$ is the *k-th* visual primitive of $\boldsymbol{D}$, $k = 1, \cdots, M$, M is the number of visual primitives in $\boldsymbol{D}$. The total number of times that $\boldsymbol{d}_k$ is used to reconstruct the patches in both views is calculated by:

$$nz_{d_k} = \left\| \boldsymbol{a}^k \right\|_0, \tag{5}$$

where $\boldsymbol{a}^k$ is the coefficient vector of $\boldsymbol{d}_k$. Note that $\boldsymbol{a}^k$ denotes the *k-th* row of coefficients matrix $\boldsymbol{A}$. The total number of times that $\boldsymbol{d}_k$ is used to reconstruct the patches in I_L is calculated by:

$$nzl_{d_k} = \left\| \boldsymbol{a}_L^k \right\|_0, \tag{6}$$

where $\boldsymbol{a}_L^k$ is the coefficient vector of $\boldsymbol{d}_k$ corresponding to I_L. The total number of visual primitives that are used to reconstruct I_L is calculated by:

$$nz_L = \sum_{k=1}^{M} nzl_{d_k}. \tag{7}$$

The probability of visual primitive $\boldsymbol{d}_k$ for I_L is calculated by:

$$p_k^L = \frac{nzl_{d_k}}{nz_L}. \tag{8}$$

The probability of visual primitive $\boldsymbol{d}_k$ to reconstruct I_R can be calculated in the same way.

According to Shannon theory, to represent the amount of information, entropy is a measure of the uncertainty of a single random variable, while mutual information is a measure of one random variable contains about another [21]. Suppose stereoscopic image as random variable of visual primitives, the entropy of visual perceptual information of I_L is defined as:

$$H(I_L) = -\sum_{k=1}^{M} p_k^L \log\left(p_k^L\right). \tag{9}$$

The entropy of visual perceptual information of I_R can be calculated in the same way.

The atom $\boldsymbol{d}_k$ is used to reconstruct the *i-th* patch in left view and the *j-th* patch in right view are calculated by:

$$nz_{d_k}(i,j) = \operatorname{sgn}\left|\boldsymbol{a}_i[k]\right| \cdot \operatorname{sgn}\left|\boldsymbol{a}_j[k]\right|, \tag{10}$$

where $\boldsymbol{a}_i[k], (i = 1, \cdots, \frac{n}{2})$ is the coefficient of $\boldsymbol{d}_k$ in *i-th* patch of left view's image, and

$$\operatorname{sgn}\left|\boldsymbol{a}_i[k]\right| = \begin{cases} 1, \boldsymbol{a}_i[k] \neq 0 \\ 0, \boldsymbol{a}_i[k] = 0 \end{cases}. \tag{11}$$

Similarly, $\boldsymbol{a}_j[k], (j = \frac{n}{2}+1, \cdots, n)$ is the coefficient of $\boldsymbol{d}_k$ in *j-th* patch of right view's image. The total number of times that $\boldsymbol{d}_k$ is used to reconstruct the patches in both the left view and the right view is calculated by:

$$nz_{d_k} = \sum_{i=1}^{n/2} \sum_{j=(n/2+1)}^{n} nz_{d_k}(i,j). \tag{12}$$

The total number of visual primitives that are used to reconstruct the both views' image I is calculated by:

$$nz_I = \sum_{k=1}^{M} nz_{d_k}. \tag{13}$$

The probability of visual primitive $\boldsymbol{d}_k$ for I is calculated by:

$$p_k = \frac{nz_{d_k}}{nz_I}. \tag{14}$$

The mutual information of visual perceptual information for I_L and I_R is defined as:

$$MI(I_L; I_R) = \sum_{k=1}^{M} p_k \cdot \log\left(\frac{p_k}{p_k^L \cdot p_k^R}\right). \tag{15}$$

The joint entropy of visual perceptual information for I_L and I_R is defined as:

$$H(I_L, I_R) = H(I_L) + H(I_R) - MI(I_L; I_R). \tag{16}$$

2.3. Visual Comfort Calculation

The human visual perception of stereoscopic images is a dynamic interaction of accommodation and vergence [22]. Accommodation corresponds to the focusing process, while vergence corresponds to the fusion process. Intuitively, both the focusing and fusion processes are driven by the visual information in perceptual cognitive process. In the proposed SVCA metric, the focusing process and fusion process are modeled by joint entropy and mutual information, respectively. In natural viewing, the focusing and fusion process can maintain a consistent adjustments with the changes of viewing distance. Whereas, in stereoscopic viewing, this consistence is broken. The inaccurate accommodation or vergence will induce the rapid change in accommodation and vergence interaction and create the accommodation-vergence conflict, which cause more visual discomfort [23]. In this paper, accommodation-vergence conflict is expressed as the ratio of the mutual information over joint entropy:

$$R_{AV} = \frac{MI(I_L; I_R)}{H(I_L, I_R)}. \tag{17}$$

Then, R_{AV} is the score of visual discomfort. A larger R_{AV} means the better similarity between the two views' image and a more comfortable experience of the stereoscopic image.

3. SUBJECTIVE EXPERIMENT

To evaluate the proposed SVCA metric, in line with the recommendation of ITU-R BT.500-11 [24], a subjective experiment was conducted to construct the stereoscopic image database. We choose sixteen subjects and a shutter 3D display system in the subjective experiment. All of the subjects had normal stereoacuity and passed the Titmus Stereo test [25]. The setup information of subjective test are listed in Table I. (Please visit our website [26] for more details).

The experimental stimuli consist of 80 stereoscopic images (see Fig. 2 for example images) which are collected from the internet. All of them have the corresponding subjective scores.

Table I. Setup of the subjective test

Test standard	ITU-R BT.500-11
Test method	DSCQS
subjects	16
Age range	20-34
Display	ViewSonic VX2268wm
Display Size	22 inch
Display resolution	1680×1050
Refresh Rate	120 Hz
Response Time	less than 2 ms
Display Card	NVIDIA GeForce GTS 450
Display Card Interface	DVI-D DualLink
Glasses	Nvidia® 3D Vision shutter stereo glasses
Glasses Refresh Rate	60 Hz
Crosstalk Levels	0.714% (left), 0.769% (right)

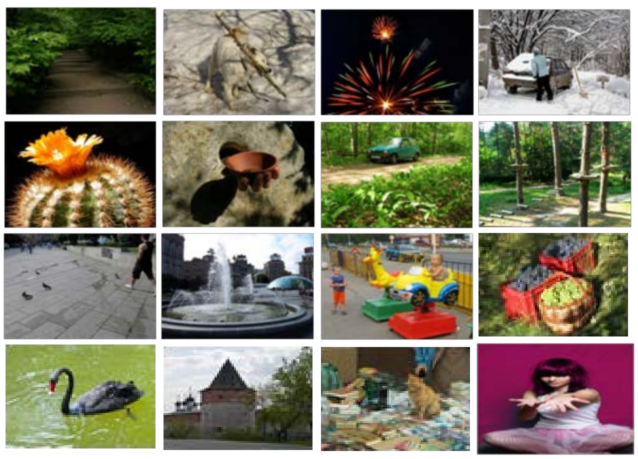

Fig. 2 Examples of stereoscopic images used in the subject experiment.

Because this paper only focuses on accommodation-vergence conflict in stereoscopic images, other discomfort-inducing factors (*e.g.*, excessive disparity, viewing distance, and crosstalk) are eliminated as much as possible. Especially, to avoid the discomfort-induced by the excessive disparity, the crossed and uncrossed disparity ranges of the stereoscopic visual stimuli belong to (-1,+1) degree [27]. In our experiment, the maximum disparity is set to less than 20 mm. All of the experimental stimuli have passed the artificial check.

4. EXPERIMENTAL RESULTS

For the nonlinear regression, we used the following mapping function as suggested by Wang *et al.* [28]:

$$f(R_{AV}) = \frac{\tau_1 - \tau_2}{1 + \exp(-\frac{R_{AV} - \tau_3}{\tau_4})} + \tau_2, \quad (18)$$

where τ_1, τ_2, τ_3 and τ_4 are the regression coefficients, and $\exp$ is the exponential function. Three popular evaluation criteria are chosen to compare R_{AV} with *MOS*, including Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC) and root mean square error (RMSE). A good objective method should have high PLCC and SROCC values but low RMS value.

4.1. Selection of number of visual primitives

According to the relationship between visual perceptual information and visual primitives, an image reconstructed with more visual primitives could provide more visual perceptual information. As indicated in [29], the number of visual primitives is proportional to the quality of the reconstruction image and the quality of a reconstructed image is good enough by 14 visual primitives in their experiment. To obtain the best number of the visual primitives, we use the method in [30]:

$$t = \arg\min\{l\} \text{ s.t. } \frac{H_l(I_L, I_R) - H_{l-1}(I_L, I_R)}{H_{\max}(I_L, I_R) - H_{\min}(I_L, I_R)} < \varepsilon, \quad (19)$$

where ε is a threshold, and is set to 0.5, empirically.

4.2. Comparison with the state-of-the-art SVCA metrics

We attempted to develop two SVCA metrics for comparison. Table II shows the performance comparison results on our database. Yano *et al.* [12] metric measures visual fatigue based on the correlation of two-view images. Kim *et al.* [9] metric estimates the disparity of stereoscopic image to predict visual discomfort. Both the proposed and Kim *et al.*'s metrics achieve good performance. Highest PLCC value of Kim *et al.*'s metric indicates well prediction accuracy, while highest SROCC and lowest RMSE values of the proposed metric indicate good prediction monotonicity and less prediction offset.

Table II: Performance comparison on our database

	PLCC	SROCC	RMSE
Yano *et al.* [12]	0.5492	0.4643	0.3245
Kim *et al.* [9]	**0.7851**	0.7156	0.2149
Proposed method	0.7542	**0.7468**	**0.1494**

5. CONCLUSIONS

This paper proposes a SVCA metric based on perceptual information of stereoscopic images to evaluate visual discomfort induced by the accommodation-vergence conflict. In the proposed SVCA metric, accommodation and vergence are modeled by the two-view images' joint entropy and their mutual information, respectively. A subjective experiment is conducted to construct a ground truth database. The experimental results show that the proposed SVCA metric has a competitive performance with two state-of-the-art SVCA models. Many other visual discomfort factors, such as excessive screen disparity, binocular asymmetry and range of depth of focus, need to be investigate in the future work.

6. ACKNOWLEDGEMENTS

This work has been supported in part by the Major State Basic Research Development Program of China (973 Program 2015CB351804), the National Science Foundation of China under Grant (No. 61472101, No. 61272386).

7. REFERENCES

[1] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *Journal of Imaging Science and Technology,* vol. 53, no. 3, pp. 1–14, June, 2009.

[2] DM. Hoffman, AR. Girshick, K. Akeley and MS. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue", *Journal of Vision*, vol. 33, no. 8, pp.1–30, 2008.

[3] T. Shibata, J. Kim, DM. Hoffman and MS. Banks, "The zone of comfort: predicting visual discomfort with stereo displays," *Journal of Vision*, vol. 11, no. 8, pp.1–29, 2011.

[4] F.L Kooi and A. Toet, "Visual comfort of binocular and 3D displays", *Displays, vol.* 25, no.2, pp. 99–108, 2004.

[5] L.B. Stelmach, W.J. Tam, F. Speranza, R. Renaud and T. Martin, "Improving the visual comfort of stereoscopic images", *in Proc. Stereoscopic Displays Virtual Reality Syst. X,* vol. 5006, pp.269–282, 2003.

[6] L. Xing, J. You, T. Ebrahimi and A. Perkis. "Assessment of stereoscopic crosstalk perception," *IEEE Trans. Multimedia,* vol. 14, no. 2, pp. 326–337, April, 2012.

[7] Y.J. Jung, S. Lee, H. Sohn, H.W. Park, Y.M. Ro, "Visual comfort assessment metric based on salient object motion information in stereoscopic video," *in Proc. of SPIE-IS&T Electronic Imaging, Special Section On Stereoscopic Displays And Applications,* vol. 011008, pp. 011008-1–011008-16, 2012.

[8] Y.J Jung, H. Sohn, S. Lee, H.W. Park, Y.M. Ro, "Predicting visual discomfort of stereoscopic images using human attention model," *IEEE Trans. Circ. Sys. Video Tech.,* vol.223, no.12, pp. 2077–2082, Dec, 2013.

[9] D. Kim and K. Sohn , "Visual fatigue prediction for stereoscopic image", *IEEE Trans. Circ. Sys. Video Tech.,* vol.21, no.2, pp. 231–236, Feb, 2011.

[10] Y. Nojiri, H. Yamanoue, A. Hanazato and F. Okano, "Measurement of parallax distribution, and its application to the analysis of visual comfort for stereoscopic HDTV", *in Proc. of SPIE-IS&T Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems X*, Vol. 5006, pp. 195–205, May, 2003.

[11] J. Park, S. Lee and A.C. Bovik, "3D visual discomfort prediction: vergence , foveation, and the physiological optics of accommodation", *IEEE Journal of Selected Topics in Signal Processing*, vol.8, no.3, pp.415–427, June, 2014.

[12] S. Yano, S. Ide, T. Mitsuhashi and H. Thwaites, "A study of visual fatigue and visual comfort for 3D HDTV//HDTV images," *Displays*, vol. 23, no. 4, pp.191–201, 2002.

[13] Y.J. Jung, H. Sohn, S. Lee, F. Speranza and Y.M. Ro, "Visual importance- and discomfort region-selective low-pass filtering for reducing visual discomfort in stereoscopic displays," *IEEE Trans. Circ. Syst. Video Tech.*, vol. 23, no. 8, pp. 1408–1421, Aug, 2013.

[14] H. Sohn, Y.J. Jung, S. Lee, F. Speranza and Y.M. Ro, "Visual comfort amelioration technique for stereoscopic image: disparity remapping to mitigate global and local discomfort causes," *IEEE Trans. Circ. Syst. Video Tech.*, Vol.24, no.5, pp. 745–758, 2014.

[15] S. Ide, H. Yamanoue, M. Okui, F. Okano, M. Btou and N. Terashima, "Parallax distribution for ease of viewing in stereoscopic HDTV", *in Proc. of SPIE-IS&T Electronic Imaging, Stereoscopic Displays and Virtual Reality Systems IX*, vol. 4660, May, 2002.

[16] G.R. Jones, D. Lee, N.S. Holliman and D. Ezra, "Controlling perceived depth in stereoscopic images," *in Proc. of SPIE-IS&T Electronic Imaging, International Society for Optics and Photonics*, pp. 42–53, 2001.

[17] S. Winkler and D. Min, "Stereo/multiview picture quality: Overview and recent advances," *Signal Processing: Image Communication,* vol. 28, no. 10, pp. 1358-1373, 2013.

[18] 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.

[19] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.,* vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

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

[21] T.M. Cover, J.A. Thomas, "Elements of Information Theory," Wiley, pp. 13, 2006.

[22] R. Suryakumar, "Study of the dynamic interactions between vergence and accommodation," *PhD Thesis, Faculty of Science, U. Waterloo, Ontario, Canada*, 2005.

[23] J. Kim, D. Kane and MS. Banks, "The rate of change of vergence-accommodation conflict affects visual discomfort," *Vision Research*, vol. 105, pp. 159–165, 2014.

[24] ITU-R BT-500.13, "Methodology for the subjective assessment of the quality of television pictures," 2012.

[25] J. Cooper and J. Warshowsky, "Lateral displacement as a response cue in the Titmus Stereo test," *American journal of optometry and physiological optics*, vol. 54, no. 8, pp. 537, 1977.

[26] [Online]. http://www.escience.cn/people/qifeng/index.html.

[27] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," *Displays* vol. 25, pp. 141–150, 2004.

[28] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680-1689, 2006.

[29] J. Zhang, S.W. Ma, R.Q. Xiong, D.B. Zhao, W. Gao, "Visual primitive coding and visual quality assessment," *in Proc. PCM*, vol. 76774, pp. 674–685, 2012.

[30] S.Q. Wang, X. Zhang, S.W. Ma, W. Gao, "Reduced Reference Image Quality Assessment Using Entropy of Primitives," *in Proc. PCS*, pp.193–196, 2013.