# VISUAL COMFORT ASSESSMENT FOR STEREOSCOPIC 3D IMAGES BASED ON SALIENT DISCOMFORT REGIONS

*Cheolkon Jung, Hongmin Liu, and Yu Cui*

School of Electronic Engineering, Xidian University, Xi'an 710071, China
zhengzk@xidian.edu.cn

## ABSTRACT

We propose visual comfort assessment for stereoscopic 3D (S3D) images based on salient discomfort regions. Color-based saliency successfully represents visual attention because that the human visual system (HVS) focuses on the most salient region in an image. Disparity-based saliency effectively expresses visual comfort in S3D images. Based on two saliencies, we extract salient discomfort regions which mostly determine the overall visual comfort degree in S3D images. Then, we predict the visual comfort score from salient discomfort regions by a disparity feature vector which combines saliency-weighted disparity and maximum disparity. Finally, we provide visual comfort index maps of S3D images based on visual comfort scores, thus showing perceptually salient discomfort regions and their degrees. Experimental results demonstrate that the proposed method achieves significantly high accuracy in visual comfort assessment as compared to existing methods.

***Index Terms***—Stereoscopic 3D, disparity, visual comfort assessment, salient discomfort regions, visual saliency.

## 1. INTRODUCTION

Because stereoscopic 3D (S3D) contents provide an immersive viewing experience to viewers, more and more people are eager to experience them. However, S3D contents often cause visual discomfort and fatigue to viewers. Thus, their viewing safety should be ensured, which becomes a growing issue of concern. Visual discomfort is induced by many factors such as excessive disparity, mismatches between the left and right images, and depth cue conflicts [1-3]. Conventional methods have employed global disparity statistics such as mean or standard deviation of disparity for visual comfort assessment [4, 5]. However, in the global statistics, disparity information is often hidden in some regions, which leads to the inaccuracy of visual comfort assessment. Thus, Percival Zone of Comfort (PZC) model [6] is proposed to assess the visual comfort in S3D images. PZC model has provided the best disparity range for the disparity adjustment module to adjust disparity maps which gives viewers realistic 3D effects without any visual discomfort. We need quantitative evaluation metrics for visual comfort assessment to accurately assess the visual comfort in S3D images. Based on human visual perception, most people have interests in salient regions [7, 8]. When the disparity is excessively large and exceeds the visual comfort zone, it causes visual fatigue and people would be annoyed by visual discomfort. In this paper, we propose visual comfort assessment for S3D images based on salient discomfort regions. In [9], Jung et al. proposed a human visual attention model for visual comfort prediction based on both image-based saliency and disparity-based saliency. They used the disparity features to quantify some characteristics such as disparity magnitude and disparity gradient. Inspired by [9], we also predict the visual comfort degree of S3D images using color-based saliency and disparity-based saliency. However, in our proposed method, we differently use visual attention and visual comfort for visual comfort assessment based on its own merit. That is, the visual attention is related to image-based saliency and represents the regions where HVS focuses. Meanwhile, the visual comfort is related to disparity-based saliency, which the over-excessive disparity gives an annoying experience to viewers. We extract two simple disparity features of saliency-weighted disparity and maximum disparity to effectively reflect the image characteristics.

Fig. 1 illustrates the overall framework of the proposed method. As shown in the figure, the proposed method consists of four main steps: saliency map estimation, visual saliency map generation, region segmentation, and visual comfort prediction. In Step 1, we generate visual saliency maps in S3D images from color-based saliency and disparity-based saliency. As compared to the linear combination method in [9], the proposed method first normalizes the color-based saliency map, and then multiplies the weighted-saliency value to the disparity-based map. Thus, the proposed method can maintain both the original disparity range and the saliency information more effectively than [9]. In Step 2, we perform region segmentation to obtain salient discomfort regions which represent both visual attention and the visual comfort degree in S3D images.

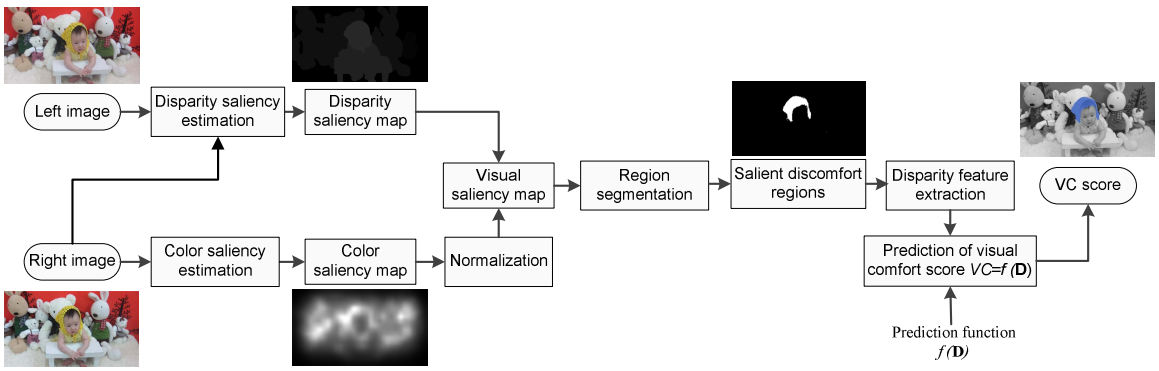



**Fig. 1.** Overall framework of the proposed method.

In Step 3, we predict visual comfort scores of S3D images using support vector regression (SVR), i.e. $VC=f(\mathbf{D})$ where **D** is a disparity feature vector. We also visualize the salient discomfort region with the visual comfort score. Experimental results demonstrate that the proposed method effectively evaluates visual comfort. As compared to existing methods, our main contributions are as follows:

1) We use salient discomfort regions for visual comfort assessment based color-based and disparity-based saliencies instead of their linear combination. Salient discomfort regions successfully reflect both visual saliency and visual comfort in S3D images.

2) We provide a disparity feature vector which consists of the saliency-weighted disparity and maximum disparity to predict visual comfort scores in S3D images. The disparity feature vector effectively considers the visual attention of HVS, which remarkably improves the prediction performance of visual comfort.

## 2. VISUAL SALIENCY MAP GENERATION

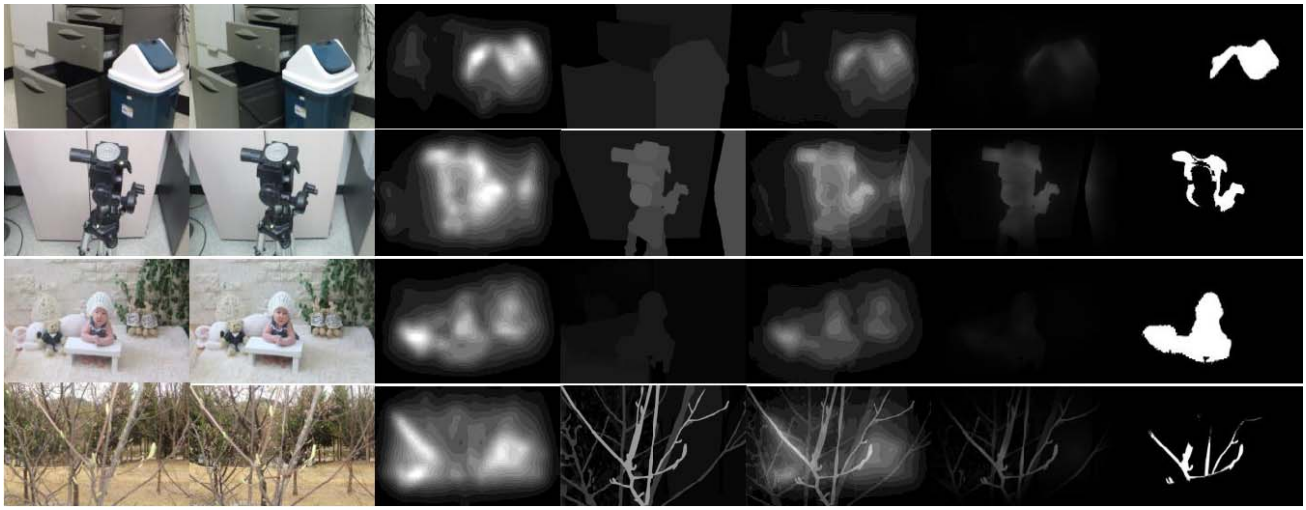
In general, there exist some areas which are attractive to the human eye in S3D images, i.e. visual attention. Color-based saliency effectively represents visual attention. In contrast, visual comfort is mostly related to disparity. That is, the larger S3D images contain excessive disparity, the more uncomfortable viewers perceive. In [10, 11], individual differences in human 3D perception have been explored. First, we estimate a color-based saliency map $S_C$ using graph-based visual saliency (GBVS) in [12]. GBVS is a bottom-up approach to visual saliency detection based on low-level features such as intensity, orientation, and color. Thus, GBVS effectively considers human eye fixations [13]. Fig. 2(c) shows $S_C$ obtained by GBVS [12]. In $S_C$, pixel values show salient degrees, i.e. black pixels are non-salient while white ones are salient. It has been reported that the density of fixations by HVS increases on objects with large disparity [14]. However, people suffer from fusing left and right views with excessively large disparity on retina. In this case, HVS would feel visual fatigue and discomfort. Thus, it is necessary to estimate a disparity-based saliency map $S_D$. We obtain $S_D$ by the depth estimation reference software (DERS) in [15]. Fig. 2(d) shows $S_D$ obtained by DERS. In the figure, the disparity saliency range is from 0 to 255. Then, we combine $S_C$ and $S_D$ to obtain visual saliency maps in a S3D image. In [9], Jung et al. obtained them by linear combination of $S_C$ and $S_D$. However, the simple linear combination weakens the saliency effects and changes the original disparity range significantly. Color-based saliency effectively represents visual attention, but has little to do with visual comfort in S3D images. Thus, we use $S_C$ as weights for $S_D$ instead of their linear combination to get visual saliency maps. Firstly, we normalize $S_C$ into range from 0 (non-salient) to 1 (most salient) as follows:

$$W_t(x,y) = \frac{S_C(x,y)}{255} \tag{1}$$

where $S_C(x,y)$ is the saliency value at the pixel position $(x,y)$ in $S_C$, and $W_t$ is the weight map to generate the visual saliency map. After normalizing the color saliency map, we regard normalized values as weights for $S_D$. As $W_t$ is close to 0, the pixel becomes non-salient. When the small weights are multiplied to $S_D$, the excessive disparities become smaller than the original ones. In contrast, as $W_t$ is close to 1, the pixel becomes more salient. When the large weights are multiplied to $S_D$, the disparity values are retained. The visual saliency map $S_{vs}$ is obtained as follows:

$$S_{vs}(x,y) = W_t(x,y) * S_D(x,y) \tag{2}$$

where $S_D(x,y)$ is the saliency value at the pixel position $(x,y)$ in $S_D$. The excessive disparity is related to the visual discomfort. Thus, it is required to maintain the original disparity range as much as possible, which is the motivation of (2). In the proposed method, $S_{vs}$ is used to obtain salient discomfort regions and disparity features for visual comfort prediction. Fig. 2(f) shows $S_{vs}$ obtained by (2), whereas Fig. 2(e) shows visual saliency maps by linear combination [16]. As can be observed, (2) is more effective in extracting salient discomfort regions in S3D images (Fig. 2(f)) than the linear combination.



(a) (b) (c) (d) (e) (f) (g)

**Fig. 2.** (a) Left view. (b) Right view. (c) Color saliency maps generated by GBVS [12]. (d) Disparity saliency maps generated by DERS [15]. (e) Visual saliency maps by linear combination [16]. (f) Visual saliency maps by (2). (g) The corresponding salient discomfort regions of (f).

## 3. SALIENT DISCOMFORT REGION DETECTION

Salient regions are more attractive to the human eye than others, and thus HVS first focuses on them. Salient discomfort regions are generated by $S_{vs}$ in Section 2, and can mostly represent degrees of visual saliency and visual comfort. To detect salient discomfort regions, we first perform region segmentation from $S_{vs}$. We adopt the region segmentation in [1] obtained by the winner-takes-all operation [8]. We select top 6%~10% values in $S_{vs}$ as candidates of salient discomfort regions. Salient discomfort regions *SDR* are obtained by:

$$Seg(SDR) = \begin{cases} 1 & if\ S_{vs} > T_r; \\ 0 & otherwise. \end{cases} \tag{3}$$

where $Seg(\cdot)$ is a segmentation function of salient discomfort regions, and $T_r$ is the threshold value. If the pixel value in visual saliency map $S_{vs}$ is larger than the threshold value $T_r$, the pixel value would become 1; and otherwise, the pixel value would become 0. Fig. 2(g) shows salient discomfort regions by (3).

## 4. PREDICTION OF VISUAL COMFORT SCORE

To assess visual comfort in S3D images, we extract two perceptually significant disparity features as follows. In [17, 18], the previous studies on subjective visual comfort assessment have reported that visual discomfort increases as disparity increases. Indeed, disparity plays an important role in generating 3D effects, which quantifies the visual discomfort in S3D images. Thus, we compute the mean of saliency-weighted disparity, $D_{MSD}$, in salient discomfort regions to reflect the disparity characteristics. $D_{MSD}$ in salient discomfort regions is obtained by:

$$D_{MSD} = \frac{1}{N} \sum_{x \in SDR} \sum_{y \in SDR} S_{vs}(x,y) \tag{4}$$

where *SDR* stands for salient discomfort regions and *N* denotes the number of pixels in *SDR*. The maximum disparity is also a significant factor to affect visual comfort of S3D images. Thus, we extract the maximum disparity $D_m$ as another feature to predict the visual comfort as follows:

$$D_m = \max\left(S_{vs}(x,y)\right) \tag{5}$$

Thus, a disparity feature vector **D** is obtained by concatenating $D_{MSD}$ and $D_m$ as follows:

$$\mathbf{D} = \left[D_{MSD}, D_m\right] \tag{6}$$

Then, we compute visual comfort score using support vector regression (SVR). The overall visual comfort score of a S3D image is calculated by the extracted disparity feature vector **D** and a prediction function *f*. The visual comfort score is obtained by $VC=f(\mathbf{D})$ where *f* has been trained by SVR [19]. *f* takes a two-dimensional disparity feature vector as an input, and produces the output visual comfort score. The prediction function is constructed as follows:

$$VC = f(\mathbf{D}) = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) \cdot K\left(D_i, D\right) + b \tag{7}$$

where $\alpha$ and $\alpha^*$ are the Lagrange multipliers; and $K(D_i, D)$ is a kernel function which represents a nonlinear transformation of input features. An $\varepsilon$-insensitive loss function is used in solving SVR, and ignores errors that are smaller than a certain threshold, i.e. $\varepsilon > 0$. We adopt a radial basis function (RBF) kernel in (7) as follows:

$$K\left(D_i, D\right) = e^{-\gamma \left\| D_i - D \right\|^2} \tag{8}$$

where $D$ is the input feature vector; $D_i$ is the $i$-th support vector; and $\gamma$ is the variance of the kernel function.

## 5. EXPERIMENTAL RESULTS

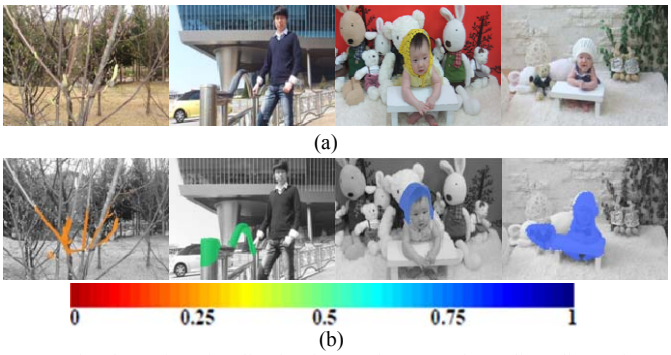



**Fig. 3.** Visual comfort visualization in S3D images using salient discomfort regions. (a) S3D image. (b) Visual comfort index map of *SDR*.

TABLE I
PREDICTION PERFORMANCE OF CONVENTIONAL METHODS [4, 5, 23] AND THE PROPOSED METHOD IN TERMS OF PCC AND SROCC

| Method | PCC | SROCC |
|---|---|---|
| Mean disparity in [4] | 0.550 | 0.665 |
| Range of disparity in [4] | 0.843 | 0.745 |
| H. method 1 in [5] | 0.621 | 0.593 |
| H. method 2 in [5] | 0.815 | 0.804 |
| H. method 3 in [5] | 0.814 | 0.804 |
| Global variance of disparity in [23] | 0.720 | 0.723 |
| Proposed method | 0.921 | 0.881 |

To evaluate the performance of the proposed method, we use S3D image database provided by the KAIST IVY Lab [20]. The database consists of total 120 S3D images with a resolution of 1920×1080 pixels. We also provide subjective visual comfort scores which ranges from 1 to 5 with five visual comfort levels: 1 is extremely uncomfortable, 2 is uncomfortable, 3 is mildly comfortable, 4 is comfortable, and 5 is very comfortable. In our experiments, the prediction model parameters ($C$, $\gamma$) in (8) are obtained by the cross validation in a series of 60 natural scene images. The regression of the prediction function is performed based on using the LIBSVM [21]. By a grid search, we choose the best parameters for regression as $C$=64, $\varepsilon$=0.1, and $\gamma$=1. Fig. 3 shows the visual comfort index map of salient discomfort regions obtained by using the proposed method. In the figures, the predicted visual comfort index is normalized in range from 0 (red color) represents "extremely uncomfortable" to 1 (blue color) represents "very comfortable". Experimental results demonstrate that the proposed method provides accurate information on visually salient regions and their visual comfort in S3D images. To evaluate the correlation between the subjective visual comfort score and the predicted visual comfort score, we measure the Pearson linear correlation coefficient (PCC) and the Spearman rank order correlation coefficient (SROCC) that have been commonly used to evaluate image quality metrics [22]. From the experimental results, PCC and SROCCC are 0.921 and 0.881, respectively. TABLE I lists the prediction performance comparison between the proposed method and conventional ones of [4], [5], and [23].

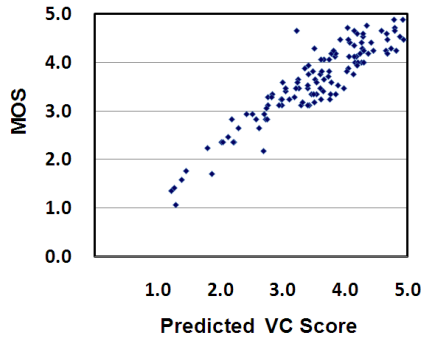



**Fig. 4.** Scatter plot between mean opinions score (MOS) and predicted visual comfort score by the proposed method.

In [4], mean and range of disparity are obtained to quantify the disparity characteristics of a scene. The range of disparity is calculated by the difference between the minimum and maximum 10% of disparity over all pixels for each image. PCC and SROCC of the mean disparity are 0.55 and 0.665, respectively, while those of the range of disparity are 0.843 and 0.745, respectively. PCC and SROCC of a global variance of disparity in [23] are 0.720 and 0.723, respectively. We also provide PCC and SROCC of three methods (H. methods 1, 2, and 3) in [5] as shown in Table I. From Table I, it can be observed that the proposed method significantly improves the prediction performance of visual comfort and achieves the best prediction performance. Fig. 4 shows a scatter plot between mean opinions score (MOS) and predicted visual comfort score obtained by the proposed method. As can be observed, the predicted visual comfort score by the proposed method has close correlation with MOS. Consequently, we can safely conclude that the proposed method can effectively assess the visual comfort score in S3D images.

## 6. CONCLUSIONS

In this paper, we have proposed visual comfort assessment for S3D images based on salient discomfort regions. Visual attention is mostly related to salient regions where HVS focuses in S3D images, while visual comfort is closely related to disparity because excessively large disparity gives an uncomfortable experience to viewers. Thus, we detect salient discomfort regions for visual comfort assessment to effectively consider both factors of visual saliency and discomfort. Experimental results demonstrate that the proposed method substantially improves the prediction performance of visual comfort.

## 7. ACKNOWLEGEMENT

This work was supported by the National Natural Science Foundation of China (No. 61271298) and the International S&T Cooperation Program of China (No. 2014DFG12780).

## 8. REFERENCES


[1] W.J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono, "Stereoscopic 3D-TV: visual comfort," IEEE Trans. Broadcasting, vol. 57, no. 2, pp. 335-346, 2011.

[2] L.M.J. Meesters, W.A. IJsselsteijn, and P.J.H. Seuntiens, "A survey of perceptual evaluations and requirements of three-dimensional TV," IEEE Trans. Circuits Sys. Video Technol., vol. 14, no. 3, pp. 381-391, Mar. 2004.

[3] M. Lambooij, W.A. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: a review," J. Imaging Sci. Technol., vol. 53, no. 3, pp. 1-14, 2009.

[4] M. Lambooij, W.A. IJsselsteijn, and I. Heynderickx, "Visual discomfort of 3D TV: Assessment methods and modeling," Displays, vol. 32, no. 4, pp. 209-218, 2011.

[5] D. Kim and K. Sohn, "Visual fatigue prediction for stereoscopic image," IEEE Trans. Circuits Sys. Video Technol., vol. 21, no. 2, pp. 231-236, 2011.

[6] H. Pan, C. Yuan, and S. Daly, "3d video disparity scaling for preference and prevention of discomfort," Stereoscopic Displays and Applications XXII, SPIE, vol. 7863, 2011.

[7] H. Sohn, Y.J. Jung, S. Lee, and H.W. Park, and Y.M. Ro, "Attention model-based visual comfort assessment for stereoscopic depth perception," Proc. Int. Conf. Digital Signal Processing, pp. 1-6, 2011.

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp.1254-1259, 1998.

[9] Y. J. Jung, H. Shon, S.-I. Lee, H. W. Park, and Y. M. Ro, "Predicting visual discomfort of stereoscopic images using human attention model," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 12, pp. 2077–2082, 2013.

[10] C. Shigeru, "3-D consortium safety guidelines for popularization of human-friendly 3-D," Eizo Joho Media Gakkai Gijutsu Hokoku, vol. 30, no. 34, pp. 21–24, 2006.

[11] M. Lambooij, W. Ijsselsteijn, M. Fortuin, B. Evans, and I. Heynderickx, "Susceptibility to visual discomfort of 3-D displays by visual performance measures," IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 12, pp. 1913–1923, 2011.

[12] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," Proc. Advanced Neural Inform. Process. Syst., vol. 19, pp. 545-552, Dec. 2007.

[13] L. Jansen, S. Onat, and P. Konig, "Influence of disparity on fixation and saccades in free viewing of natural scenes," Journal of Vision, vol. 9, no. 1, pp. 1-19, 2009.

[14] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," IEEE Trans. Image Process., vol. 19, no. 1, pp. 185-198, 2010.

[15] M. Tanimoto, T. Fujii, and K. Suzuki, "Depth estimation reference software (DERS) 5.0," ISO/IEC JTC1/SC29/WG11 M16923, Oct. 2009.

[16] Y. J. Jung, H. Sohn, S. I. Lee, and F. Speraza, "Visual importance- and discomfort region-selective low-pass filtering for reducing visual discomfort in stereoscopic displays," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 8, pp. 1408-1421, 2013.

[17] S. Yano, M. Emoto, and T. Mitsuhashi, "Two factors in visual fatigue caused by stereoscopic HDTV images," Displays, vol. 25, no. 4, pp. 141-150, 2004.

[18] M. Wopking, "Viewing comfort with stereoscopic pictures: an experimental study on the subjective effects of disparity magnitude and depth of focus," J. SID, vol. 3, no. 3, pp. 101-103, 1995.

[19] V. N. Vapnik, Statistical learning theory, New York: Wiley, 1998.

[20] IVY Lab stereoscopic image database [Online]. Available: http://ivylab.kaist.ac.kr/demo/3DVCA/3DVCA.htm, 2012.

[21] C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector machines, 2001. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[22] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. on Image Processing, vol. 5, no. 11, pp. 3411-3452, 2006.

[23] J. Choi, D. Kim, B. Ham, S. Choi, and K. Sohn, "Visual fatigue evaluation and enhancement for 2D-plus-depth video," Proc. IEEE ICIP, pp. 2981-2984, 2010.