Closing The Gap: Visual Quality Assessment Considering Viewing Conditions

Yucheng Zhu, Guangtao Zhai, Ke Gu and Zhaohui Che
Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong University, Shanghai, China
{zyc420, zhaiguangtao, gukesjtuee, chezhaohui}@sjtu.edu.cn

***Abstract*—Most of existing visual quality assessment algorithms are tested on standard databases that are created in controlled viewing conditions (e.g. display device, viewing distance and lighting). This implies that all the recoded subjective scores are only valid for the specific settings used in the database. However, with the prevalence of mobile devices, the practical viewing environments can significantly vary from moment to moment. It is our daily experience that the same image can look drastically different on dissimilar devices under changed viewing distance and/or lighting conditions. In other words, a gap exists between the eyes and the visual contents behind the screen in current research of quality assessment. Therefore, in this work, we perform subjective quality evaluation with varied actual viewing conditions. To make the research reproducible, we build a prototype system to record what the eyes really see from the screen and construct the viewing environment-changed image database. The database will be made available to the public. Meanwhile we design a dedicated effective environment-assessing algorithm. We believe that this work will benefit the research of visual quality assessment towards more practical applications.**

Keywords—Image quality assessment (IQA), viewing environments, subjective / objective assessment.

I. Introduction

Living in an information age, people are becoming ever more demanding about display picture quality. There are many viewing conditions such as desktop computing, digital graphics design, digital photography, electronic interactive games; as well as applications in technical, medical, industrial and security applications where accurate imaging is an important factor. Although the programme sources, the transmission and the display technology are at the heart of the picture quality improvement. The viewing environment also has a direct and weighty impact on the picture quality. Viewing conditions can have a profound impact on how viewers perceive videos and images. These environmental factors can assist or impede viewing comfort.

The Society of Motion Picture and Television Engineers (SMPTE) is responsible for setting standards and practices for the motion picture and related industries. The organization's recommended practices document (SMPTE RP 166-1995) deals very specifically with viewing room conditions and the interaction of viewers with studio monitors and the environment in which they are used.

But except for the lucky minority who own a dedicated home theater room, decorated strictly according to all of the various SMPTE standards and practices that can improve the viewing comfort in a large extent. Most residential viewing environments can only be set up in a way that incorporates and satisfies some of the more important professional viewing environment conditions. What's more, with the prevalence of mobile devices, the practical viewing environments can significantly vary from moment to moment. So we should make out the influence of different environmental factors on the viewing comfort.

In order to thoroughly understand the impact of viewing conditions, we construct a new dedicated viewing environment-changed IQA database (VEID). Some recommendations regarding the viewing environment have been given for television pictures [1], multimedia [2], 3DTV [3], flat panel displays [4], and mobile devices [5]. And some databases [6]–[8] take those suggestions during subjective tests. However, our practices to evaluate the environment should be different. The database [9] changes the viewing distance in the subjective test and considers the effects of viewing distance and image resolution on viewing experience. We here choose the high-quality images to display and vary the viewing condition to do the assessment of the environment. We set up a prototype system to capture image contents. The subjective evaluations and ratings are made by observers under different viewing conditions.

Such assessment ought to be done by dedicated device with the development of robotization technique. Over the years, a number of researchers have contributed significant research in the design of objective IQA algorithms. Device using IQA algorithms can be set up to evaluate the effect of different surroundings on the viewing experience. It is effective, objective and more convenient.

The goal of IQA is to design algorithms for objective evaluation of quality in a way that is consistent with subjective human evaluation. The largest number of objective image quality assessment metrics are full-reference (FR) methods [10]–[13], which assume that the original image signal is completely known. Those metrics are usually of high performance. However, the dependence of original images severely reduces its applicability, and FR IQA methods measure the fidelity to the reference which will take image enhancement as "distortions". To solve these problems, many blind quality measures have been developed during the last decade [14], [15]. Those NR IQA methods are applicable in many more practical scenarios and show high performance on image databases.

Here we design one dedicated algorithm to evaluate the viewing environment. During the subjective test, digital camera is used to simulate the behavior of human eyes as much as possible. We propose our objective environment-assessing algorithm to simulate the brain to evaluate the comfort of the

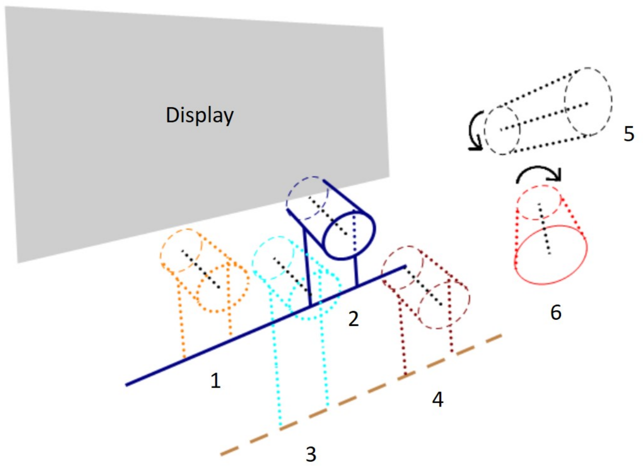

Fig. 1: The usage of our prototype system: The camera site can be altered according to need. We can control the slide rails to move the camera to position 1,2,3,4 in the picture. At the same time, the manipulator arms can rotate in the vertical direction and horizontal direction as 5 and 6 show.

viewing environment. Our idea is to simulate the process from eye to brain. The images to display are all of high quality and the distortions are introduced by changing different viewing environmental factors. We first predict what brain perceives with the free energy [16]. Then we calculate the feature points in the valid region. By assessing the captured images, designed algorithm gives the evaluation of different viewing conditions. It is one kind of algorithm which simulates human's assessing process.

The rest of this paper is organized as follows. Section II first presents the construction of VEID. In Section III, the proposed algorithm is explained in detail. The effectiveness of our algorithm is proved by experimental results in Section IV. Finally, concluding remarks are given in Section V.

II. Database Construction

For the convenience of the research about the effects of different environmental factors on the image quality, we construct the viewing environment-changed image database. We elaborately decorate the photostudio and design different environments to cover actual viewing conditions that are common in our daily life.

We build a prototype system to make the research reproducible. The whole system consists of slide rails, manipulator arms and a digital camera. The digital camera is used to capture the image on the display under different viewing conditions. The camera is set to auto mode to simulate the behavior of human eyes as much as possible. The manipulator arms hold the camera and control the rotation in horizontal and vertical directions. The slide rails support the manipulator arms and control the shift in horizontal directions. The usage of the system is shown in Fig. 1. The photos of the system and the scene are shown in Fig. 2.

The prototype system is used to capture the image on the display under different viewing conditions. We carefully prepare different viewing environments and take several environmental factors into account.

Fig. 2: Prototype system: photos of manipulator arms, digital camera, the slide rail, the 180-inch projection screen, 4k projector and TN LCD.

(1) Viewing Distance: The distance the viewer is from the screen will have a significant impact on perceived image quality. In ideal conditions, observers can perceive a sharp smooth image, free of visible pixels or image artifacts. However, different viewing distances usually happen in common experience. Here we set three distances from the screen and at each position we use the system to capture the image on the display.

(2) Viewing Angle: Seating locations relative to the screen must be emphasized. Different viewing angles of seatings from the centerline of the screen on the horizontal axis and the vertical viewing angles below the horizon will impede viewing comfort in different extent.

(3) Room Lighting: Room lighting must be controlled to avoid screen reflections. Reflections, haze, and diffused glare will interfere with the light coming from the display and this will result in direct compromises in picture quality. In our database, we set different levels of brightness and let the camera capture the display under these varied lighting environments.

(4) Bias Lighting: During the course of the typical movie or TV program, the images that appear on the screen range widely from very dark to very brightly lit scenes. This range of picture brightness causes the iris in our eyes to open wide during dark scenes and then shutter down during the brightest images. Over time, this leads to fatigue to our eyes and associated

Fig. 3: The fifteen captures of the TN LCD.

Fig. 4: The first row is of different distances. The second row is of different angles. The third is of different bias color temperature. The forth is of different bias brightness. The fifth is of different room lighting.

muscle groups, causing short-term discomfort and possibly long-term fatigue. Correctly implementing bias lighting into the typical viewing environment will minimize the extremes of iris movement and lead to more comfortable viewing. Besides, the color temperature of bias lighting also concerns. Lighting with different color temperatures will cause different effects on the experience. In our database, we set bias lighting of different colors and bright scale levels.

To discover the effect of environmental factors, we try to eliminate other adverse effects on image quality before capture.

TABLE I: The experimental conditions and some important device that are used during the construction of the database.

Viewing distance	2 / 4 / 6 times the image height
Viewing angle	0 / 30 / 45 horizontal degrees, 10 / 20 vertical degrees
Viewing lighting	no additional lighting fluorescent lamp above the screen fluorescent lamp above the seat fluorescent lamp behind the seat fluorescent lamp around the room
Viewing bias lighting	four gray scale levels (0, 100, 150, 200 among 0-255) three color temperature (2462k, 6676k, 11415k)
Digital camera	HERO3+ Black Edition resolution is 12MP with auto spot metering on remote wireless controlled
TN LCD	DELL P2815Q
4K projector	SONY VPL-VW1100

We use the 180-inch projection screen and the matched 4K projector to deliver images and exhibit outstanding picture quality overall. Images of high quality are placed onto the pure-color background which is used to simulate bias lightings. After the capture, we choose the 1000*600 pixel-sized region containing the image to display and cut the 1000*600 sized region out as one database image.

Using the above mentioned method, we construct the new dedicated viewing environment-changed image database (VEID). The proposed database consists of 140 images captured under different viewing conditions and 10 pristine images. There are five lighting modes, seven kinds of bias lighting, three distances and five different viewing angles in VEID. The total number of designed environments is twenty-five. Besides, we do fifteen captures of the twisted nematic liquid crystal display (TN LCD) from different angles for its popularity but small range of viewing angles and general ability of color reproduction. The fifteen images as shown in Fig. 3 of TN LCD can reflect the problem. Fig. 4 shows some other images in the database. Table I gives some important information during the construction of the database.

Reliable subjective evaluations to represent the ground truth are one of the most important components for the evaluation and benchmarking of new image quality assessment algorithms [17]. In our subjective test, the more effective way to carry out the perceptual quality evaluation is to provide observers with the same viewing environment as that of the capture. In other words, observers are instructed to sit under different viewing conditions to assess the content on the display. We only replace the camera with observers to do the evaluation under different viewing conditions. Fig. 5 shows the averaged scores under the twenty-five designed viewing conditions. We can conclude that the strong lighting which falls onto the screen will ruin the experience. And inappropriate bias lighting will also impede the viewing comfort. Although the database is almost constructed using the dedicated projector, the images that reflect the effects of the viewing environment on the viewing comfort also have sufficient reference for other kinds of display technology.

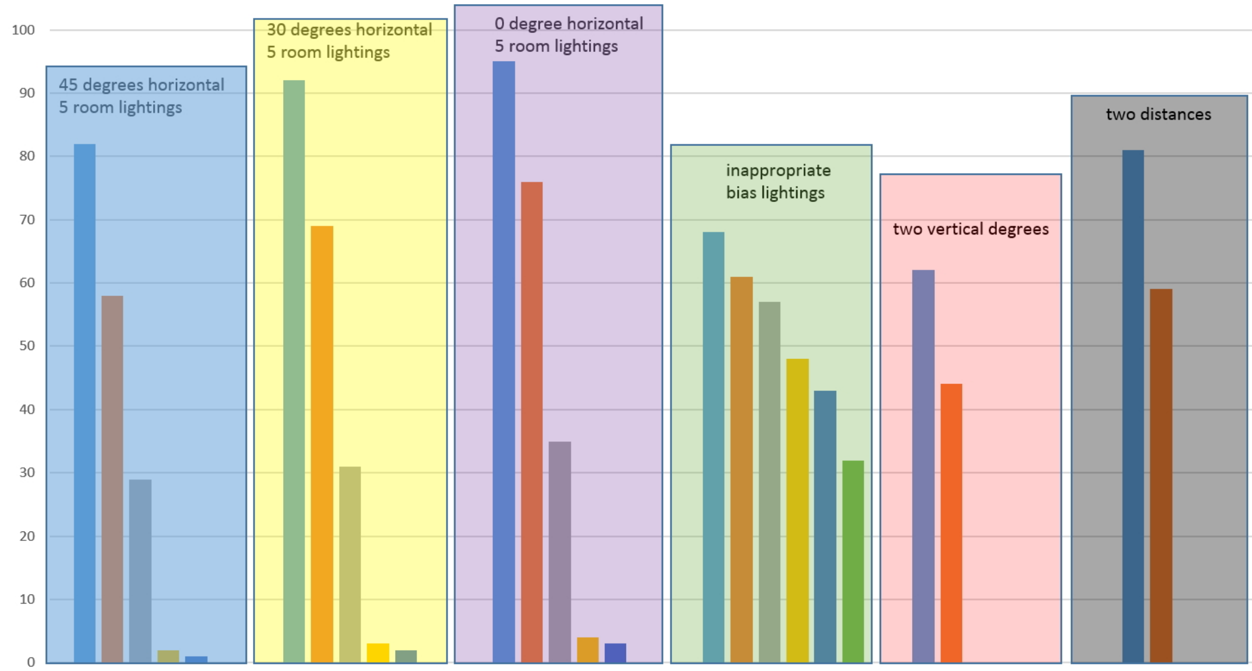

Fig. 5: Each bar averages the subjective scores of images for one viewing condition. The x-axis represents the twenty-five different viewing environments. They are further divided into six groups and each group focuses on different factors. Within one group from left to right are viewing conditions from good to bad.

III. Objective Evaluation

We design the objective method to assess different viewing environments. We first want to predict what human brain perceives. Given an image signal I, the free-energy principle suggests that the cognitive process is governed by an internal generative model $\mathcal{G}$ in the brain [16], [18]. Given different scenes or images, the model $\mathcal{G}$ will adapt itself through varying a parameter vector θ . Then the free energy caused by image I can be computed by integrating the joint distribution $P(I,\theta|\mathcal{G})$ over the space of model parameter θ. The minimization of free energy equals the maximization of the model evidence.

$$-\log P(I|\mathcal{G}) \quad = \quad -\log \int P(I,\theta|\mathcal{G})d\theta. \qquad (1)$$

We can use an auxiliary posterior distribution $Q(\theta|I,\mathcal{G})$ to calculate the surprise of I in (1). Referring to [16], we can drop the latent model assumption $\mathcal{G}$ in our analysis for simplicity, since the behavior of the model can be characterized by parameter θ. By letting the auxiliary term into (1) and using Jensen's inequality we have

$$-\log P(I) \quad \leq \quad -\int Q(\theta|I) \log \frac{P(I,\theta)}{Q(\theta|I)} d\theta. \qquad (2)$$

The right hand side of (2) is defined as the free energy:

$$F(I,\theta) \quad = \quad -\int Q(\theta|I) \log \frac{P(I,\theta)}{Q(\theta|I)} d\theta. \qquad (3)$$

By noticing that $P(I,\theta) = P(\theta|I)P(I)$, we can write (3) into

$$\begin{aligned} F(I,\theta) &= \int Q(\theta|I) \log \frac{Q(\theta|I)}{P(\theta|I)P(I)} d\theta \\ &= -\log P(I) + \int Q(\theta|I) \log \frac{Q(\theta|I)}{P(\theta|I)} d\theta \\ &= E_Q[-\log P(I|\theta)] + KL\left(Q(\theta|I) \| P(\theta)\right). \quad (4) \end{aligned}$$

TABLE II: Self-adaptive classification considering color information via kmeans.

Image classification (input image $\mathbf{D}$, threshold T, output division map M, number of clusters to be classified N, cluster centroid locations $\mathbf{C}$ and the final separated binary map W)
1. Let $N = 2$ be the initial number of clusters to be classified.
2. Let $\mathbf{D}$ be the input of kmeans function. Kmeans partition the $H - by - 3$ data matrix into N clusters where H is the number of pixels and 3 represents the RGB color information. Each cluster contains pixels with similar color information. Calculate the proportion P of the smallest cluster.
3. Add one to the number of clusters to be classified N.
4. Iterate between (2) and (3) until $P < T$.
5. Return M, N, $\mathbf{C}$.
6. Take several sampling points at the four corners of M.
7. Classify all the pixels with the same cluster as the sampling points as the background region. Classify the others as the valid region.
8. Estimate the valid region. Return to step 2 and apply zigzag change to T if it is not a quadrangle. Otherwise output this as W.

Here the term $KL\left(Q(\theta|I) \| P(\theta)\right)$ measures the distance between the recognition densities and the true prior of the model parameters. The term $E_Q[-\log P(I|\theta)]$ is the averaged entropy of predicting I.

We hope to compute the free energy. For operational amenability we can hypothesize the generative model $\mathcal{G}$ to be a 2D linear autoregressive (AR) model for its high description capability for natural images. The AR model is defined as

$$x_n \quad = \quad \boldsymbol{\chi}^k(x_n)\boldsymbol{\alpha} + \varepsilon_n \qquad (5)$$

where x_n is the n_{th} pixel, $\boldsymbol{\chi}^k(x_n)$ is a row-vector that consists of k nearest neighbors of x_n, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, ..., \alpha_k)^T$ is the vector of AR coefficients and the ε_n is the error term.

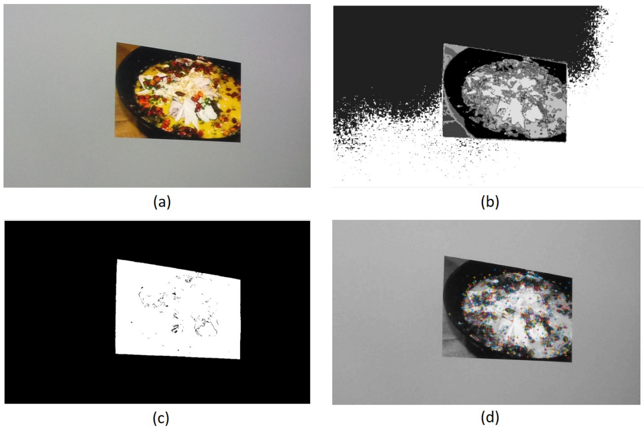

Fig. 6: Results of the division: (a) is the input image, (b) is the division map according to color information. (c) is the separated binary map and the white area is the valid region. (d) shows the extracted feature points.

Under the large sample condition, the free energy equals the total description length of image I. So we estimate the AR coefficients by minimizing the description length

$$\hat{\boldsymbol{\alpha}} \quad = \quad \arg\min_{\alpha}(-\log P(I|\boldsymbol{\alpha}) + \frac{k}{2}\log N) \qquad (6)$$

where N is the data sample size. We fix the order and the training set size of the model and thus turn the comparison process into residual minimization

$$\hat{\boldsymbol{\alpha}} \quad = \quad \arg\min_{\alpha} ||x - \mathbf{X}\boldsymbol{\alpha}||_2 \qquad (7)$$

where $x = (x_1, x_2, ..., x_N)^T$ and $\mathbf{X}(n,:) = \boldsymbol{\chi}^k(x_n)$. And the parameter can be solved as $\hat{\boldsymbol{\alpha}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T x$. In this case, the parameter θ of the model can be well described by $\hat{\boldsymbol{\alpha}}$. And then we can get the prediction of brain perceived image through the model.

The captured images are not suitable for direct calculation, because they are also surrounded by background regions. The size of images in database is 1000*600 pixels but the valid region to be assessed is only the area recording the picture displayed on the screen. These pictures are captured under different viewing conditions, so the quality of the recorded picture can reflect the effect of environment. What we should do is to separate the background region and the valid region. We design an self-adaptive algorithm to do the classification. Table II gives the detailed operation.

Fig. 6 (b) is the result of color classification for Fig. 6 (a). The input image is classified into ten clusters according to the main color information as the division map (b) shows. Then after sampling at the corners of (b), the further separation as (c) shows between background and valid region is processed. To be more accurate and to eliminate the effect of the edge, we further apply the erosion and dilation operations. We first use the positive one to do the dilation to fill the black holes. And then we do the erosion to shrink the valid region to remove the effect of the edge.

When observers watch the screen displaying the same content but under different conditions. Observers will make out that the content on the screen is unchanged but is subject to a range of transformation, including scaling, affine stretch, change in brightness and contrast. Because some unchanged features that brain perceives will help observers to recognize the image. Mikolajczyk (2002) found that the maxima and minima of $\sigma^2\nabla^2 G$ produce the most stable image features compared to a range of other possible image functions, such as the gradient, Hessian, or Harris corner function [19]. According to research [20], about 80 percent of keypoints in the transformed image are of the stable features that can be detected at a matching location and scale. In other words, the extracted feature points are almost of good properties. So the number of this kind of keypoints can be regarded as evaluation basis. Besides, the computation is also convenient. The $\sigma^2\nabla^2 G$ can be computed from the difference of Gaussian.

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2\nabla^2 G \qquad (8)$$

Keypoints are defined as the scale-space extrema. Each sample point is compared to its eight neighbors in the current image and eighteen neighbors in the above and below images. We take the effect of different lighting conditions into account by thresholding the value of the sample point. Fig. 6 (d) shows the extraction of keypoints.

So after the simulation of brain function to get the prediction of brain perceived image, we evaluate the keypoints of the valid region. We calculate the number of keypoints in one set of images which record the same picture but affected by different environmental factors. The number of keypoints is only affected by the environmental factors and thus it can be used to assess the quality of environments.

IV. Experimental Results

We apply our objective method to five datasets (one image captured under various viewing conditions are classified as one dataset) in VEID to validate its effectiveness of evaluating the viewing environment. The images in the same dataset record the same picture but are captured under twenty-five different viewing conditions. Our proposed method is suitable to evaluate images of the same recorded content. It is inappropriate for our method to evaluate the images which are captured under different environments and meanwhile record different contents. So the evaluation of environmental factors is carried out among each of the five datasets. We display scatter plots of MOS versus predicted MOS on five datasets in Fig. 7. From the figure we can see that the objective evaluation is consistent with the subjective evaluation of different environments.

Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SROCC) are used to evaluate performance of our approach. PLCC can be considered as a measure of prediction accuracy, while SROCC measures the monotonicity by ignoring the relative distance between data. The higher SROCC and PLCC values indicate better performance in terms of correlation with human opinion.

Table III gives the SROCC and PLCC results on the five datasets. The table shows that evaluation of the environments by our objective method is consistent with the subjective assessment. Considering that the recorded images in our database are almost under the transformation of scaling and affine stretch, so common objective IQA approaches are not suitable to evaluate these images. And here we do not study

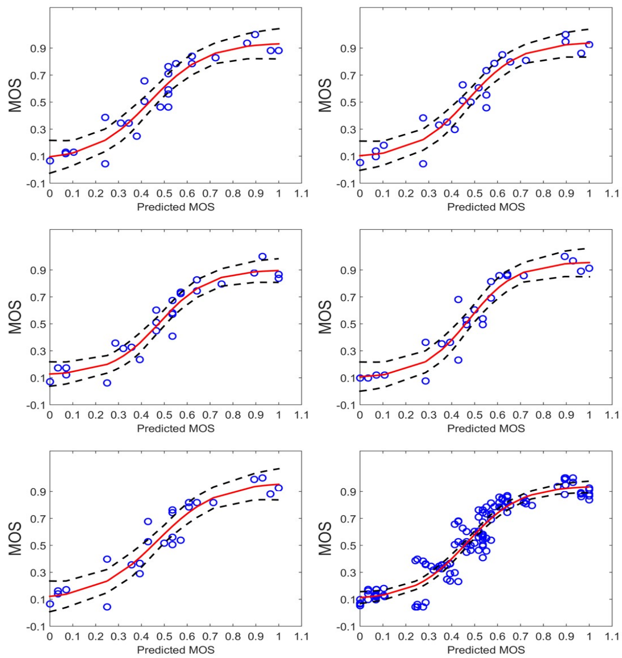

Fig. 7: Scatter plots of MOS vs. Predicted MOS on five datasets. The (red) lines are curves fitted with the logistic function and the (black) dash lines are 95% confidence intervals. The last is the overall graph.

TABLE III: The experimental results on the five datasets. One image captured under various viewing conditions are classified as one dataset.

VEID Database	PLCC	SROCC
Dataset1	0.948	0.912
Dataset2	0.959	0.923
Dataset3	0.949	0.921
Dataset4	0.949	0.903
Dataset5	0.943	0.924

the modifications of these methods to make them suitable for VEID because of the limited time and space.

We further explore about the environmental factors. We can draw some conclusions from the observation of the results. Small viewing angles from the centerline on the horizontal axis or the vertical axis have a small effect on the viewing experience. The strong lighting falls onto the screen area or in the path from display to viewer will impede the viewing comfort. The bias lighting that is restricted to the area behind the display illuminates the wall facing the viewer and it is beneficial in significantly improving the viewing comfort and reducing the eye strain.

V. Conclusion

In this paper, we set up a prototype system to emulate the eyes' function of capturing light signals under different environmental conditions and make the research reproducible. We construct the VEID with the dedicated system, photostudio and well-designed viewing environments to discover the impacts of environmental factors on viewing experience. Finally we propose an objective method to assess the viewing condition and demonstrate its effectiveness.

Acknowledgment

This work was supported in part by the National Science Foundation of China under Grants 61422112, 61371146, 61521062, 61527804, the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201339 and National High-tech R&D Program of China under Grant 2015AA015905.

References

[1] I. R. Assembly, *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union, 2003.

[2] R. ITU, "Methodology for the subjective assessment of video quality in multimedia applications," 2007.

[3] I. Union, "Subjective methods for the assessment of stereoscopic 3dtv systems," *Recommendation ITU-R BT*, vol. 2021.

[4] I. BT2022, "General viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays," *Int. Telecommun. Union*, 2012.

[5] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. De Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 652–671, 2012.

[6] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," 2005.

[7] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 626–639, 2012.

[8] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *Cybernetics, IEEE Transactions on*, vol. 46, no. 1, pp. 284–297, 2016.

[9] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *Broadcasting, IEEE Transactions on*, vol. 61, no. 3, pp. 520–531, 2015.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[11] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1500–1512, 2012.

[12] K. Gu, G. Zhai, X. Yang, and W. Zhang, "An efficient color image quality metric with local-tuned-global model," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 506–510.

[13] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 43–54, 2013.

[14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 12, pp. 4695–4708, 2012.

[15] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Using free energy principle for blind image quality assessment," *Multimedia, IEEE Transactions on*, vol. 17, no. 1, pp. 50–63, 2015.

[16] G. Zhai, X. Wu, X. Yang, W. Lin, and W. Zhang, "A psychovisual quality metric in free-energy principle," *Image Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 41–52, 2012.

[17] S. Winkler, "Analysis of public image and video databases for quality assessment," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 616–625, 2012.

[18] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[19] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Computer VisionECCV 2002*. Springer, 2002, pp. 128–142.

[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.