A Model for the Appearance of Interocular Colorimetric Differences in Binocular XR Displays

Minqi Wang, Emily A. Cooper**, Lorenza Moro*, Bharath A. Narasimhan*, Hui Chen**

*Samsung Display America Lab, San Jose, CA

**University of California – Berkeley, Berkeley, CA

Abstract

Many extended reality (XR) devices present different views to the left and right eyes. Unwanted colorimetric differences between these views can cause perceptual artifacts that degrade binocular image quality. We present an image-computable model designed to predict the appearance of binocular views with colorimetric differences in XR displays. The model is fitted to data from a recent perceptual study in which people provided multidimensional responses about the appearance of stimuli simulating an optical see-through augmented reality device with interocular intensity differences. This work can be used to create preliminary assessments of binocular artifact appearance and inform XR display design.

Author Keywords

applied vision; XR; display perception; binocular vision; fusion; rivalry

1. Introduction

Many extended reality (XR) devices use separate light sources and optics for the two eyes, so an understanding of their visual quality requires an understanding of binocular vision. For example, variations between the two eyes' light sources and optics can introduce unintended colorimetric differences in the luminance, hue, and color saturation between the eyes. Large interocular differences can lead to annoyance and visual discomfort during viewing [1], however some differences can be tolerated without dramatic degradation to the user experience [2].

The human visual system receives information from the two eyes as separate channels, and combines them to achieve a unified binocular percept of the world. When the intensity or color differs between corresponding features in the two eyes' views, the binocular percept is often determined by the eye that is receiving the "higher energy" signal (i.e., higher luminance, contrast or color saturation) [3,4,5]. Thus, for example, a display defect present in only one eye may nonetheless create a binocular artifact if it results in a higher contrast pattern than what is seen by the other eye. However, the precise contribution of each eye's view to the binocular percept depends on various factors, such as the surrounding context [6]. We refer to this aspect of binocular appearance as the "binocular summation," because it can be modelled as a weighted summation of the two eye's inputs.

When differences between the two eyes exceed some threshold, additional perceptual phenomena are elicited that can lead to uniquely binocular perceptual artifacts [7,8]. For example, it is known from the vision science literature that people are able to detect colorimetric differences between the two eyes based on the appearance of luster (a shimmery, sometimes shiny appearance) [7,8] and rivalry (an alteration of perceptual appearance over time) [8]. We refer to these aspects of binocular appearance as the "binocular difference" because these phenomena tend to increase with increasing differences between the two eyes' views.

While computational models of binocular summation and binocular differencing based on human psychophysical studies exist, these models are limited in their utility for XR applications because they are based on data from simple stimuli that do not capture the complexity of natural vision [3,5,9,10]. For example, recent work modelling perception of uniform colored patterns found that the binocular color difference threshold has a range of 30-50 units in the CIELAB space [11,12]. However, with more natural images, the difference threshold can deviate from this range [13]. With the complexity of imagery encountered in XR applications, it is thus challenging to build a simple model that robustly accounts for perceptual features of interest across a broad set of potential experiences. In the fields of image compression and tone mapping, several groups have introduced image-computable algorithms for binocular summation and/or differencing that can be applied to arbitrary natural images [14,15,16,17]. However, these models have not yet been directly evaluated using human data that characterize binocular summation and differencing when people view colorimetric differences between the two eyes. In addition, many of these algorithms use sRGB values as inputs because the display properties were not quantified or reported for the dataset that was used to create the model [15,16,17], and thus cannot be used directly to address the appearance of colorimetric differences in XR display design.

We propose a modelling framework that connects basic vision science models derived from simple stimuli and image-computable algorithms that can be applied to colorimetric data. To guide the design of our model, we leveraged a recently published perceptual dataset that evaluated binocular summation and differencing with imagery that simulated colorimetric differences in optical see-through augmented reality (AR) viewing [18]. This dataset enables us to create an image-computable framework that is relevant for XR applications. Because observer responses covered multiple features of the stimulus appearance, our model can evaluate aspects of binocular appearance resulting from both summation and differencing. Our proposed model performs two tasks: 1) it generates a binocular appearance image given a set of colorimetric differences in a pair of views (binocular summation), and 2) it computes a prediction of the likelihood that human observers will detect luster, rivalry, and related phenomena (binocular difference). While the dataset simulates AR viewing, we formulate the model with the aim that it can be used to inform design decisions for virtual reality (VR) as well. We assess our model performance using a held-out validation sample from this same dataset and show that the model predictions have a high correlation and low error with respect to the ground truth data.

ISSN 0097-996X/24/5501-0177-$1.00 © 2024

2. Methods

Dataset: The stimuli used in the perceptual study simulated optical see-through AR content, in which a semi-transparent circular AR *icon* [19] was superimposed on a natural image background [20] (Figure 1). These stimuli were created graphically by compositing each icon with a natural background image and the brightness of the icons was manipulated by adjusting the sRGB values before compositing. The stimuli were presented on a desk-mounted mirror haploscope with two LCD displays, such that each eye could receive a different view of the stimulus.

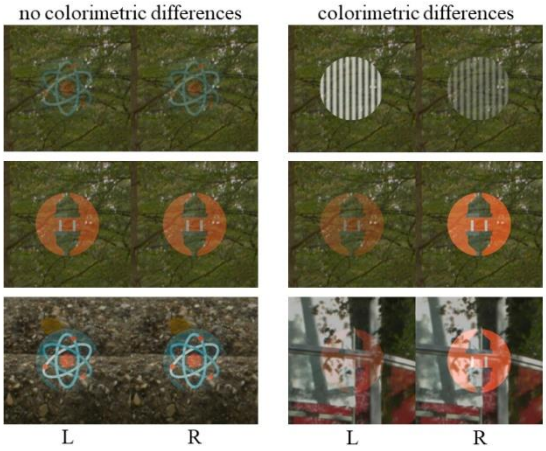

Figure 1. Six example stimuli showing the left (L) and right (R) eye views of the reference pair, which could contain colorimetric differences or not. For the stimuli simulating no colorimetric difference, the icons had the same sRGB values in the two eyes. Stimuli that simulated colorimetric difference had icons with different values between two eyes. Image credits: Freepik and SYNS [19,20].

During each trial, observers (n=31) were asked to perform a matching task between two stimuli containing a given icon and background combination. One stimulus contained the *reference icon*. The reference icon sometimes had the same appearance in both eyes, and sometimes it had a colorimetric difference such that one eye's icon was higher intensity than the other. At the same time, observers were shown an *adjustable* icon. This icon always had the same sRGB values in both eyes, and participants could use key presses to adjust the overall intensity of this icon to try to match it as closely as possible to the appearance of the reference. Once they found the best possible match, they were asked to judge whether the match was perfect or not (that is, was the adjustable icon identical to the reference, or did it contain some differences such as luster or rivalry). If it was not an exact match, that serves as an indication that people detected a binocular difference in the reference. With our model, we aimed to predict both the average perceived intensity of the best match across all observers (a measure of binocular summation) and the proportion of observers who indicated that the match was not perfect (as a measure of binocular difference).

We included 36 stimuli reported in [18] that had four AR icons with different intensity levels shown against the same background. In addition, we included another 18 stimuli with different backgrounds. In total, there were 54 stimuli with different icon and background combinations and various colorimetric differences between the two eyes. We split the dataset into two subsets, using 81% of the data (44 stimuli) as the model building set, and 19% (10 stimuli) as a validation set.

Model: Figure 2 shows an overview of the modelling pipeline. The inputs to the binocular vision model are pixel-wise CIEXYZ values simulating the imagery seen by each eye. Because we do not have the full display metrology information for the dataset that we are using, we implemented a preprocessing stage that approximates the XYZ values for each eye. First, we recreated the 8-bit sRGB image bitmaps used in [18]. Next, we simulated the displayed image by combining these bitmaps with the measured white points for each of the haploscope LCD displays to approximate the corresponding XYZ maps for each eye using Matlab's rgb2xyz function. The convert these maps into a more perceptually meaningful color space (CIELAB), we needed to determine a binocular white point. This white point can be interpreted as the white that the brain is adapted to given the two eyes' inputs. For the model pipeline, we assume that the brain is adapted to the D65 illuminant (x,y= 0.31272,0.32903) and the higher luminance of the two displays. Next, we used Matlab's xyz2lab function to convert the left and right XYZ maps to CIELAB (L*a*b*) maps with respect to this binocular white. The choice of the binocular white point could be switched with another standard illuminant. For example, we also tested the model using Illuminant E and did not find it to affect the model's performance. Once in the L*a*b* space, the model has two paths to predict the binocular appearance associated with summation and differencing, which were hand-tuned to model the average best match resulting from the adjustment task and the probability that the match was imperfect, respectively.

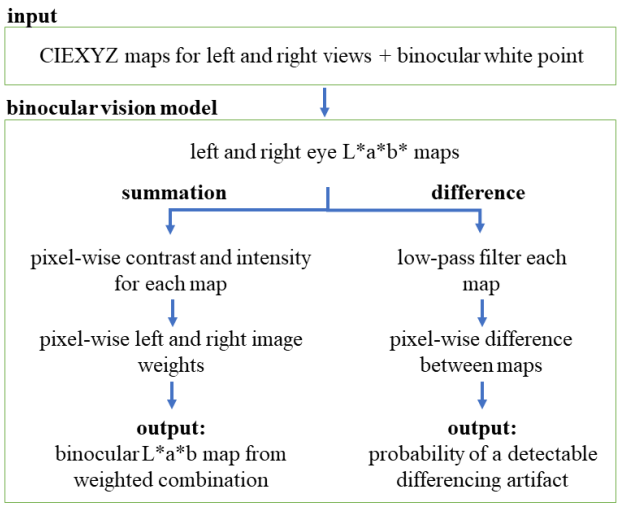

Figure 2. Overview of model pipeline.

The *summation* path combines the left and right eyes' images to create a prediction of the binocular image percept. This path comprises a pixel-wise weighted summation of the left and right eye's input, in which the weight for each eye depends on the local contrast and intensity (lightness and color saturation) [3,4,5]. To determine the local contrast in each eye's image, we computed local image gradients with a Prewitt filter, operating on the L*, a*, and b* channels separately. Gradient magnitudes less than 1 were clipped to 0 such that color changes less than 1 were considered to be uniform to reduce small noise. When both eyes' gradients were 0, the weights for both eyes were set to 0.5. We then summed across the three channels to obtain a single contrast value at each pixel. We found that implementing an averaging filter on the contrast map lead to a small improvement in the model accuracy. The final chosen size for the filter was 0.5 x 0.5 degrees, which in this case was 30 x 30 pixels. We speculate that this may be due to local contours modulating the binocular

balance of their proximal surroundings [5]. Prior work has only considered luminance in modeling stimulus intensity because the stimuli were grayscale [5], but empirical work has shown that color saturation also plays a role in binocular summation [4]. In our pipeline, we therefore summed the L* channel with the absolute values of the a* and b* channels as a measure of intensity since the absolute value of the a* and b* channels can be thought of as approximating the color saturation. The resulting weights for the left and right eye inputs at each pixel location are then determined as follows:

$$w_l(i,j) = \frac{I_l(i,j)C_l(i,j)}{I_l(i,j)C_l(i,j) + I_r(i,j)C_r(i,j)}$$
$$w_r(i,j) = \frac{I_r(i,j)C_r(i,j)}{I_l(i,j)C_l(i,j) + I_r(i,j)C_r(i,j)} \quad (1)$$

where I and C are the image intensity and contrast, the subscripts l and r denote left and right eye respectively, and i,j denote the pixel row and column. The binocular L*a*b* map (Lab_b) is determined by applying these weights (w_l, w_r) to the left and right eye's L*a*b* images (Lab_l, Lab_r) before summing:

$$Lab_b(i,j) = w_l(i,j)Lab_l(i,j) + w_r(i,j)Lab_r(i,j)\ . \quad (2)$$

Figure 3 shows a visualization of a left and right image pair, along with the weight assigned to each pixel and the resulting binocularly combined image.

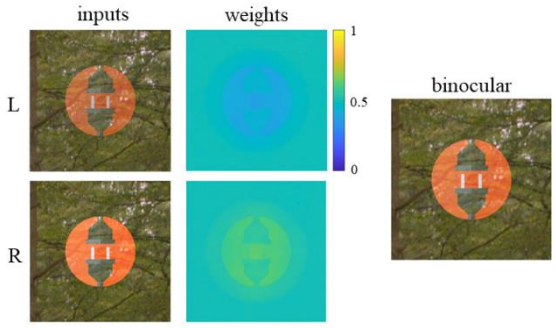

Figure 3. The left and right input images are shown, along with the assigned weight for each pixel. On the right is the resulting binocular image determined through weighted summation. The computations were done in L*a*b* space and converted to sRGB here for visualization.

To determine the image selected as the best match to the reference icon in the perceptual task, the model iterates through all possible settings for the adjustable stimulus via a grid search method (using 50 steps from the minimum to the maximum possible icon intensity), and compares the resulting binocular image with the binocular image for the reference. The model match is the option with the smallest mean difference between the reference and adjustable binocular L*a*b* map as shown in Figure 4.

In parallel, the model uses the difference in the two eye's views of the reference to predict the probability that this best match was not a perfect match (i.e., the probability that luster, rivalry, or other binocular differences were detected). Once in the L*a*b* space, we used the existing standard color difference formula, dE_{ab}, to compute the difference between the two eyes, which is the Euclidean distance between the left and right eye color vectors. As in the summation calculation, we found that applying an averaging filter on the L*a*b* maps prior to computing dE_{ab} improved model performance. Contrary to traditional color difference formula that considers each pixel as a sample, this averaging suggests that a certain spatial area of binocular difference may be necessary for the visual system to detect the mismatch between the two eyes' inputs. The averaging filter used was 1 x 1 degrees (60 x 60 pixels).

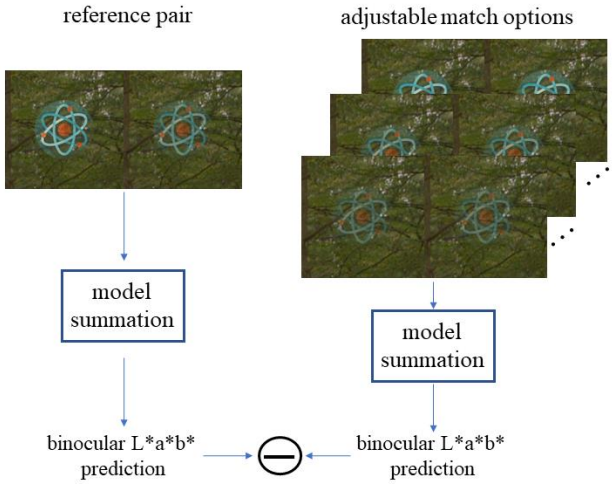

Figure 4. To predict the best perceptual match, both the reference pair and pairs of the adjustable stimuli were run through the pipeline separately and the binocular L*a*b* maps were subtracted to find the pair with the smallest mean difference.

After obtaining the dE_{ab} map, a binocular difference metric (d) was defined based on the maximum difference value between the two eyes. We applied a hand-tuned transformation to dE_{ab} to constrain it to be between 0 and 1 to maximize the fit with the perceptual data, as follows:

$$d = \begin{cases} 0, & max(dE_{ab}) < 4.1 \\ 1, & max(dE_{ab}) > 29.9 \\ 0.5(\ln(max(dE_{ab})) - 1.4), & otherwise \end{cases} \quad (3)$$

3. Model Performance

Figure 5 shows the human data (x axis) against the model fit (y axis). For the matching task (top plot), the axes reflect the average normalized icon brightness value that provided the best match to the reference stimulus. For the question about whether the match is exact, the axes reflect the proportion of observers that said they could not find a match through adjustment, meaning that there were residual differences in appearance between the reference and the adjustable stimuli. Table 1 shows the fitting performance for each subset of the data and for the full dataset in terms of correlation with the human data (r) and the mean error (ME). The model shows good performance for the both the model building and the validation subsets.

Qualitatively, the model is slightly under predicting the values for the matching task, suggesting that it is performing slightly more binocular averaging than the human observers. The addition of a nonlinearity that biases percepts more towards the higher contrast/intensity image may improve the model fits. For the exact match judgement, the model is capturing the overall trend and correlates well with the human data. One potential use of this model is to estimate the maximum acceptable color difference for

a given display system by solving Equation 3 for max(dE_{ab}). For example, targeting a proportion (d) of less than 0.1 would mean that the maximum dE_{ab} should be less than 5. However, it is possible that more variance could be explained by considering other features of the binocular difference, in addition to the maximum difference.

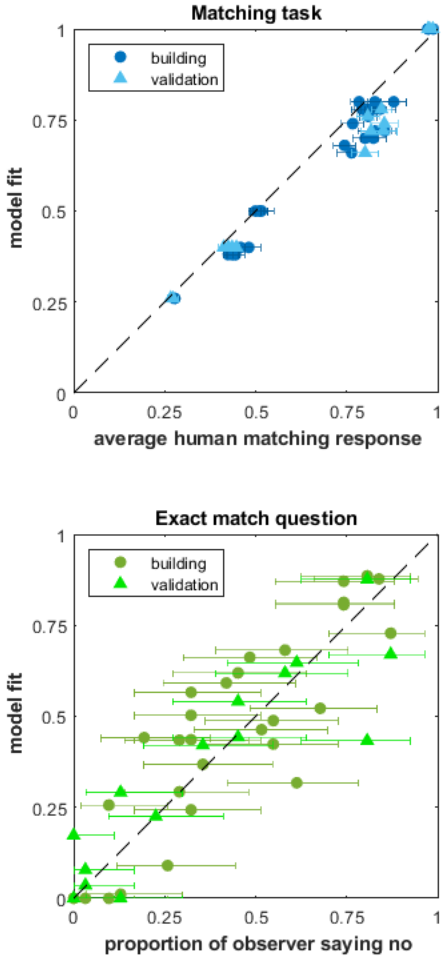

Figure 5. Data and model fits for the matching task (top) and the exact match question (bottom). Each symbol represents a unique stimulus. The dark circles are the model building stimuli and light triangles symbols are the validation stimuli. The black dashed line is the identity line. The error bars represent the 95% confidence intervals for the human data.

Table 1. Pearson's correlation coefficient (r) and mean error (ME) between the model fit and the human data for the model building subset, validation subset, and all data combined.

Matching Task			
	Model building	Validation	Combined
r	0.98	0.98	0.98
ME	0.04	0.05	0.05
Exact Match Question			
	Model building	Validation	Combined
r	0.92	0.90	0.91
ME	0.09	0.09	0.09

4. Conclusion

Binocular appearance models are important tools for developing XR technologies. Such models should consider the properties of the XR display and optics, the complexity of the content that people view of these displays, and the multifaceted nature of binocular perception. Towards this goal, we leveraged a new perceptual dataset to develop an initial model that uses display metrology and a perceptual colorspace to predict perceived binocular appearance and performance on the detection of binocular differences. Because the modelling pipeline is highly modular, alternative implementations could be tested to potentially improve model performance in the future.

5. Acknowledgements

This research is supported by Samsung Display America and NSF funding (Award #2041726). We would like to thank Luke Hellwig for helpful discussions.

6. References

[1] Chen, K., Chen, Z., Zhou, D., Tai, Y., & Shi, J. (2020). Visual comfort evaluated by hue asymmetries in stereoscopic images. Journal of the Society for Information Display, 28, 843 - 853.

[2] Chen, D., Chen, Z., Yang, T., Huang, X., & Yun, L. (2022). Visual comfort evaluation for stereoscopic videos with different binocular color allocation scheme. SPIE/COS Photonics Asia, 12318.

[3] Ding, J., Klein, S.A., & Levi, D.M. (2013). Binocular combination of phase and contrast explained by a gain-control and gain-enhancement model. Journal of Vision, 13(2), 13.

[4] Kingdom, F.A., & Libenson, L. (2015). Dichoptic color saturation mixture: Binocular luminance contrast promotes perceptual averaging. Journal of Vision, 15(5), 2.

[5] Ding, J., & Levi, D.M. (2017). Binocular combination of luminance profiles. Journal of Vision, 17(13), 4.

[6] Wang, M., Ding, J., Levi, D.M., & Cooper, E.A. (2022). The effect of spatial structure on binocular contrast perception. Journal of Vision, 22(12), 7.

[7] Wendt G, & Faul F. (2020). The role of contrast polarities in binocular luster: Low-level and high-level processes. Vision Research, 176, 141-155.

[8] Malkoc, G., & Kingdom, F. A. (2012). Dichoptic difference thresholds for chromatic stimuli. Vision Research, 62, 75–83.

[9] Legge G.E., & Rubin G.S. (1981). Binocular interactions in suprathreshold contrast perception. Perception & Psychophysics, 30(1): 49–61.

[10] Georgeson, M.A., Wallis, S.A., Meese, T.S., & Baker, D.H. (2016). Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. Vision Research, 129, 98-118.

[11] Xiong, Q., Liu, H., Chen, Z., Tai, Y., Shi, J., & Liu, W. (2021). Detection of binocular chromatic fusion limit for opposite colors. Optics Express, 29(22), 35022-35037.

[12] Asano, Y, & Wang, M. (2023). An investigation of color difference for binocular rivalry and a preliminary rivalry metric, ΔE*bino. Color Research & Applications, 1-14.

[13] Sun, P.L., Chang, T.Y. and Luo, R.M. (2012), 54.4: Binocular color-rivalry thresholds of complex images. SID Symposium Digest of Technical Papers, 43: 733-736.

[14] Yang, X.S., Zhang, L., Wong, T., & Heng, P. (2012). Binocular tone mapping. ACM Transactions on Graphics, 31, 1 - 10.

[15] Chen, M., Su, C., Kwon, D., Cormack, L.K., & Bovik, A.C. (2013). Full-reference quality assessment of stereopairs accounting for rivalry. Signal Processing: Image Communication, 28, 1143-1155.

[16] Wang, J., Rehman, A., Zeng, K., Wang, S., & Wang, Z. (2015). Quality prediction of asymmetrically distorted stereoscopic 3D images. IEEE Transactions on Image Processing, 24, 3400-3414.

[17] Yang, J., Lin, Y., Gao, Z., Lv, Z., Wei, W., & Song, H. (2015). Quality index for stereoscopic images by separately evaluating adding and subtracting. PLOS ONE, 10.

[18] Wang, M., Ding, J., Levi, D. M., & Cooper, E.A. (2023). The effect of interocular contrast differences on the appearance of augmented reality imagery. ACM Transactions on Applied Perception, 21, 1-23.

[19] Freepik Company S.L. (2019). Freepik Mobile App Icon Images. https://www.freepik.com/ [Accessed: April 19, 2019].

[20] Adams, W.J., Elder, J.H., Graf, E.W., Leyland, J., Lugtigheid, A.J., & Muryy, A.A. (2016). The Southampton-York Natural Scenes (SYNS) dataset: Statistics of surface attitude. Scientific Reports, 6, 35805.