Data-Driven and Optics-Inspired Decomposition of Global Pupil Swim in VR/AR for an Improved Perception Model of Motion Discomfort

Phoebe Lim Ching*, Tsz Tai Chan**, Yudong He*, Chuchu Qiu*, Richard H.Y. So*, Jerry Jia**

***Hong Kong University of Science and Technology, Hong Kong SAR, China**

****Meta Reality Laboratory, CA, USA**

Abstract

Users can observe dynamic distortion (or global pupil swim) during the usage of VR/AR headsets. Our earlier study correlated pupil swim to selected optic flow patterns corresponding to motion and mathematically modeled discomfort. This study decomposes global pupil swim as a linear sum of orthogonal basis patterns based on Zernike polynomials for improved applicability and for an improved perception model.

Author Keywords

Distortion; pupil swim; head-mounted display; virtual reality

1. Introduction

Users of virtual reality (VR) may experience image distortion while viewing a head-mounted display (HMD). This distortion can be partially corrected by presenting a pre-distorted image that offsets the distortion from the optics. However, this approach assumes that the user's eyes stay on the optical axis of the display. As users' eyes move across the display during a typical VR experience (e.g., vestibulo-ocular reflex, or VOR) the correction fails and images dynamically distorts (or exhibit global pupil swim, PS) [1]. The perceived distortion changes as a function of eye locations and the intrinsic optic design of the display system.

PS map can be plotted as a vector field map (also called optic flow, an example is shown in Figure 1), each vector in the map representing the angular shift of content from its expected location to perceived location at that field. Our earlier study [1] mathematically modeled perceived motion discomfort by comparing and correlating the complex optic flows with simple basis patterns that are linked to natural motions. The magnitudes of correlation to these patterns were used to predict human perception of PS. While it is greatly helpful in evaluating and validating different optical system designs, it is a mapping algorithm that can generate duplicated estimations in predicting discomfort. For example, the sum of correlated basis patterns is usually larger than the raw PS map. These duplicated representations will add noise in predicting perceptual discomfort and prevent detailed analyses of the associative causal effects.

The objective of this study is to decompose PS into a finite set of basis components (BCs) which satisfy the following constraints: (a) PS can be decomposed into a linear sum of BCs; (b) The components should be as orthogonal as possible; and (c) Reconstructions of PSs do not involve components with vector features that are not present in the original pupil swim. The last constraint aims to avoid the creation or inclusion of interim features that will be cancelled out during the summation process. In other words, we aim to reconstruct the PS through a progressive linear summation process [2].

Compared to our previous study, this new study: (1) enhances the math and perception model with orthogonal Zernike polynomials in a data-driven manner; (2) expands the scope of perception prediction to include more optical system designs; and (3) applies to more gaze directions /user observation modes.

2. Method of Decomposing PS maps

2.1. Defining PS basis components (BC)

PS maps (total 246) from 41 different optical designs and 6 positions of eye motion were decomposed into a linear sum of basis components. The PS maps are initially represented as a 2xN array corresponding to the horizontal and vertical magnitudes of each vector in an optic flow pattern (where N is the total number of vectors). For this study, N of 1764 (42x42) was used. These 2xN components are reshaped into a 1-dimensional matrix with length 2N to allow for the simple mathematical representation in eq. 1. Each case of PS indexed *i* (PS_i) is the sum of BC indexed *j* ($comp_j$) weighted by the fitted coefficients ($coef_{i,j}$), or the magnitude of the component. An example of such decomposition is illustrated in Figure 1.

$$PS_i' = \sum_j coef_{i,j} * comp_j \qquad (1)$$

The choice of BC patterns is of great importance. Our study used the orthogonal set of Zernike vector polynomials inspired by Zhao and Burge [3]. Orthogonality ensures that the absence or presence of a component does not affect the fitted coefficient of other components [4]. Moreover, lower order Zernike polynomials are consistent with optic flow observed during

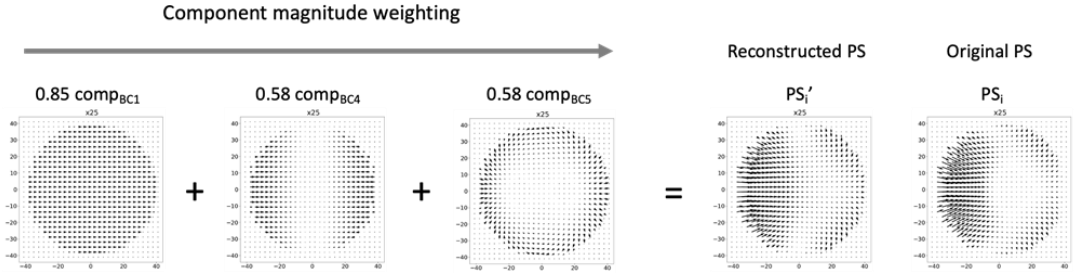

Figure 1. Decomposition of a PS (Lens G, gaze forward) and comparison between reconstructed PS and the original PS. Vectors are scaled up 25 times for better illustration. The median reconstruction errors were 0.02 degrees before scaling up.

motions [5]. Our earlier study [1] showed that these motion-based optic flow patterns are reliable predictors for discomfort. Hence, the choice of Zernike polynomials is consistent to the earlier study. Zernike polynomials can be derived in a continuously increasing order and complexity. However, for this application, it is not desirable to use many components, and the very high-order polynomials do not resemble PS maps. An iterative data-driven approach for selecting components from a pool of Zernike vector polynomials was therefore developed and tested.

At each iteration, the 246 pupil swim maps were decomposed into a set of components (the decomposition approach will be discussed further in the succeeding section). For the first iteration, the lowest order Zernike polynomials were used. The 246 residual maps calculated as the differences between the original PS map and the partial reconstructed map were then clustered.

The center of the largest cluster was understood to represent the most common pattern of residuals among the PS maps. This cluster center was used to select a single Zernike polynomial as a basis component by optimizing the linear fit of a single polynomial to the residual map at the cluster center. The selected BC (now a particular Zernike polynomial) would then be included in further iteration in which the 246 components were once again decomposed into a set of components. Eventually, as more components were added and a larger percentage of PS were accounted for, the largest cluster of residuals became the cluster of null vectors. The next largest cluster that was above a minimum residual magnitude threshold would be selected. The iterations were concluded using residual magnitude as the stopping criteria.

To minimize the inclusion of basis components with features that will be cancelled out during the summative process, correlated BCs (especially negatively correlated) were grouped into a single BC based on the weighted average of its magnitude in the previous decomposition of PS. The 6 BCs derived by this procedure are shown in Figure 2. They correspond to Zernike polynomials of S2, S2-mirrored; S5-mirrored; weighted sum of S4-mirrored and S6-mirrored; weighted sum of S8 and T7-mirrored; and S8-mirrored.

It is worth noting that data-driven approach using principal component analysis (PCA) to decompose the 246 PSs have been examined but found unsatisfactory. Patterns not found in the original PSs were frequently found in the decomposed PCs. One main reason is the orthogonality constraint by PCA was too strong. The current iterative methods optimally balanced the need for orthogonality and the conformity to optical flow patterns consistent with Zernike polynomials.

2.2. Reconstruction of PS with the basis components: Sequential fitting vs. concurrent fitting

In decomposing PS into a linear sum of vector field patterns, there was a tendency for BC or groups of BCs to be larger than the original PS. Larger magnitude is defined as the tendency of some components to have larger vectors, or to be denser than the original PS. This is caused by the potential for patterns to offset each other.

A sequential fitting process was used wherein one component would be fitted to the pupil swim at each time. In effect, the magnitude of each component would be constrained to the magnitude of the original PS. After one component is fitted, the next component is fitted to the residual from the previous fitting.

At each iteration, the residuals from previous iterations (res_i) were fitted with a new component which was selected by the minimization problem in equations 2-4. The constraint in equation 4 was added given that a positive or negative coefficient can significantly alter the perception of the component (for example, it could mean the difference between translation along and opposite to the motion, where the latter was perceived more strongly based on our earlier study [1]). The objective function fits a coefficient (α_i) to each potential component indexed j. Thus the optimized coefficient α_i* is a function of j. This links to linear sum in eq. 1 through eq. 5. In addition to sequential fitting, regularization was introduced to further reduce the collinearity between features. In this case, the L1 regression coefficient (l_1) was applied to improve the sparsity of components. This prevented less important components from being used in decomposition.

$$\min_j \min_{\propto_i} \left\{ e_{i,j}^T e_{i,j} + l_1 |\propto_i| \right\} \tag{2}$$

$$e_{i,j} = res_i - \propto_i comp_j \tag{3}$$

$$\propto_i \geq 0 \tag{4}$$

$$coef_{i,j} = \propto_i^* \tag{5}$$

The distributions of the median reconstruction errors for the 246 PSs are plotted in Figure 3. The number of components used in the fitting ranged from 2 to 4 for each pupil swim.

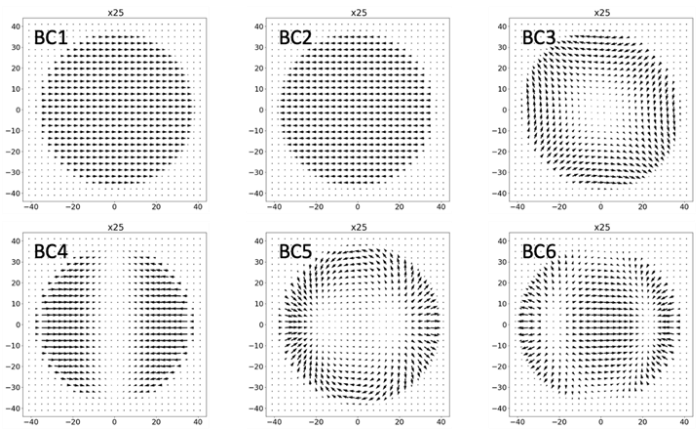

Figure 2. The basis components (BCs) derived from the iterative data-driven selection process.

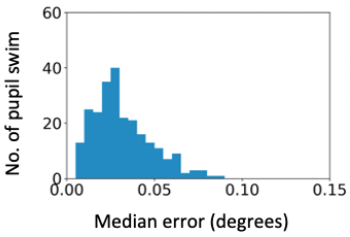

Figure 3. Median reconstruction errors (in degrees) of the 246 pupil swims using the BCs shown in Figure 2.

3. User Study and Perception Model

3.1 Method for the user study

Our earlier study showed that certain types of optic flow may be more disturbing than others[1]. In the previous study, only 4 lens were investigated, and the gaze direction was forward only. This new user study aims to examine more representative lens and two different gaze directions (forward and downward). This was intended to substantiate the findings of the previous study and to expand the scope of model prediction. 41 lens designs with different PS were grouped into 6 clusters, and lens positioned closest to the center of each cluster were extracted as Lens E to J. It turned out that Lens E was Lens D in [1] and Lens F was Lens C in [1], respectively. This suggested that the arbitrary selection of lens in [1] were indeed partially representative. The pupil swim maps of the selected lenses are shown in Figure 4.

To maintain consistency with our previous study [1], the study followed similar experimental designs. Vector maps of display distortions associated with a PS and its components were dynamically overlaid on VR content and presented through a modified Oculus Rift CV1 headset. Participants were asked to fixate on objects at specific regions of the display, and then rotate their heads while keeping their eyes fixated on the object. This effectively created the PS and its perceptual effects. The motion was repeated according to an 80bpm metronome (one beat for rightward rotation, one beat for returning to original position.) Exposure time was capped at 60 seconds, with participants being free to give responses earlier if they were able to. Similar to [1], the participants were asked to report their perception of each simulated map by answering the following questions: (a) In this trial, suppose you are exposed to this visual environment for about 20 minutes, how would you assess the scene in terms of discomfort, dizziness, and disorientation? (Rate 0-5)

(b) In this trial, comparing with extremely realistic virtual world, how would you assess the scene in terms of image deformation or disorientation? (Rate 0-3).

In total, 20 participants were each presented with 36 PS maps: the 6 lenses presented at 3 magnitudes in 2 gaze directions. The 3 magnitudes consist of 1x (intrinsic PS magnitude), 0.2- and 0.3-degrees average vector magnitude. One participant dropped out due to sickness. This suggests a dropout rate of 5% which is lower than the 15.6% reported in a survey of 46 experiments on VR motion sickness [7][8].

Prior to modeling, data from the 19 subjects were analyzed with ANOVA and the main effects of lens, magnitude and gaze directions on the reported scores were significant with a confidence level of 98% ($p<0.02$).

3.2 Perception model

A perception model for predicting user discomfort from a specific PS was trained with data from the user study, and data from our earlier study [1]. Four participations found to be non-sensitive were excluded in a data-driven way. Equation 6 explains how the perception model used fitted coefficients (coefi,j) from PS decomposition as inputs to predict discomfort scores. As component BC6 was only fitted to a single lens (Lens J), it was replaced with the average of BC1-6 as the new BC."

Similar to the previous study, a gaze-centric eccentricity weight was applied in the model as well as gaze compensation to model the eye fixation mechanism [1, 4]. In addition, a log-scale was applied based on Weber-Fechner law [1].

$$Score = a\left(\sum_j w_j \log(1 + coef_{i,j}) + w_0\right) \qquad (6)$$

The perception model is trained by fitting perception weights for each component w_j and the individual subject scaler a. The subject scaler allows the model to account for inter-subject variability in the user study data. Figure 5 shows a comparison of the performance of the new model against the previous model using the previous data set. While both models can predict the test Lens B, the new model no longer over-predicts. Given the fact that the BCs are now non-arbitrary and are orthogonal, future work to fine tune the new model can lead to better results.

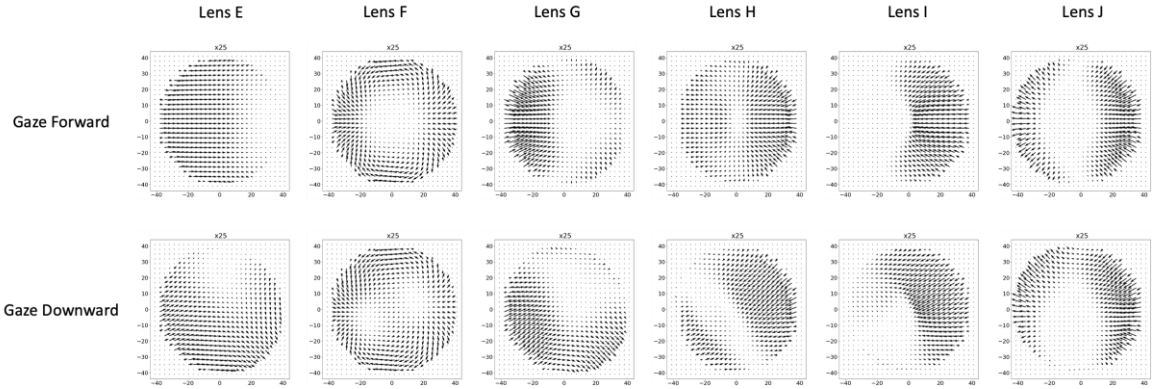

Figure 4. The 12 PSs used in the pilot study (6 lens x 2 gaze directions). The 6 lenses are representative of 41 optical designs.

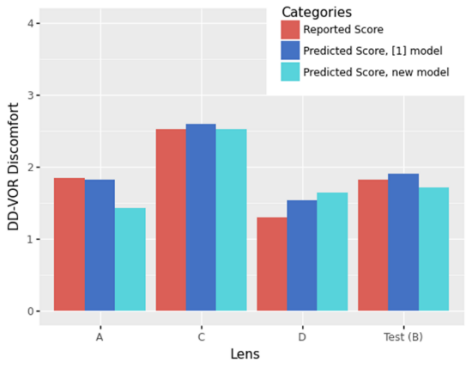

Figure 5. Comparison of mean reported scores, predicted scores for test Lens B and fitted scored for Lens A, C and D using previous model reported in [1] and the current model trained on phase 1 data (gaze forward with intrinsic unscaled PSs)

Figure 6 shows the predictions of the new model on two new lens unseen by the model. As these two new lenses are representative of two clusters consist of 19 optical designs, findings are much more significant than our previous study [1]. Inspections of the figure indicate that the new model could predict and differentiate the scores associated with the two test lenses in both gaze-forward and gaze-downward conditions.

4. Conclusion

Our new study decomposes pupil swim (PS) into a linear sum of optic flow patterns consistent with the Zernike polynomials. Selections on the basis components and the design of decomposition approach were made with the intention to minimize duplication and offsetting between components. This approach had two benefits: First, it creates a perceptually meaningful way of decomposing optic flow patterns. Second, the magnitude of each basis component is proportional to its perceptual contribution.

The data collected in this study was used to build an improved perception model for PS. Data on user perception of PS was used to fit perception weights for the basis components. Following calibration on a wider range of PSs from various kinds of eye movement, the perception model may be applied to testing optical designs as part of the prototyping process.

The mathematical perception model from our previous work has been greatly useful in designing VR/AR products [1]. It allows us to predict discomfort from optical systems without the need of making physical prototypes, significantly speeding up iterations. Like the early days of Color Science when people started to mathematically model color perception, we have created and substantiated this first math model for perception of motion discomfort due to pupil swim. Compared to our earlier study, this new model is more numerically sound in its identification of components and decomposition. The basis components identified are aligned with user perception.

5. Acknowledgements

This study is partially funded by Meta Reality Laboratory and Hong Kong University Grants Council.

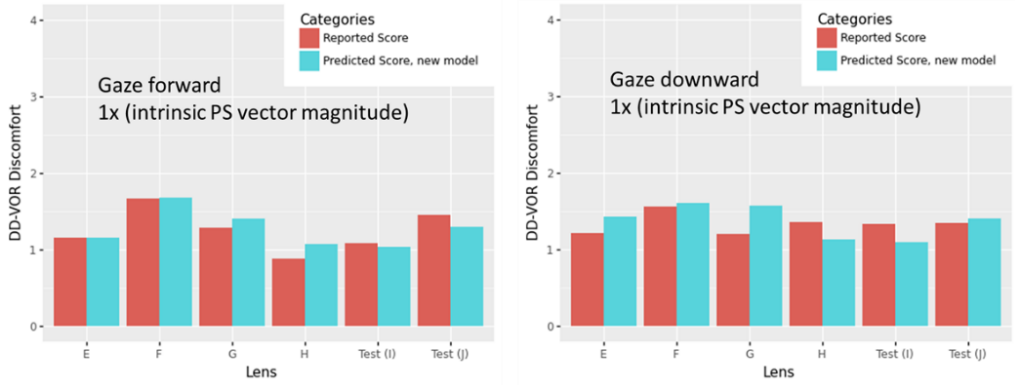

Figure 6. Comparison of mean reported scores, fitted scores and predicted score (test Lenses I & J) for gaze forward and downward conditions with horizontal VOR of 16 degrees and intrinsic unscaled magnitude of the PSs of the lenses. Models trained on data collected in [1] and this study.

6. References

1. Chan TT, Wang Y, So RHY, Jia J. Predicting subjective discomfort associated with lens distortion in VR headsets during vestibulo-ocular response to VR scenes. IEEE Trans Vis Comput Graph, doi: 10.1109/TVCG.2022.3168190.
2. Oppenheim, A.V., Willsky, A.S. and Hamid, S. (1997) Signals and Systems. ISBN 0-13-814757-4. Chapters 2 & 6.
3. Zhao C, Burge J. Orthonormal vector polynomials in a unit circle, Part I: basis set derived from gradients of Zernike polynomials. Opt Express 2007;15(26):18014-24.
4. Lakshminarayanan V, Fleck A. Zernike polynomials: a guide. J Mod Opt 2011; 58(7): 545–61.
5. Opoku-Baah C, Erkelens I, Qian R, Sharma R. A Binocular Model to Evaluate User Experience in Ophthalmic and AR Prescription Lens Designs. IEEE ISMAR-Adjunct 2022; 628-633, doi: 10.1109/ISMAR-Adjunct57072.2022.00130
6. McColgin F H. Movement thresholds in peripheral vision. J Opt Soc Am 1960; 50(8): 774.
7. Saredakis, D., Szpak, A., Birckhead, B., Keaga, H.A.D., Ruzzo, A. and Loetshcer, T. (2020) Factors associated with VR sickness in HMDs: a systematic review and meta-analysis. Frontiers in human neuroscience, 14.
8. Lo and So (2001) Cybersickness in the presence of scene rotational movements along different axes. Applied Ergonomics, 32, pp.1-14.