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

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

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

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

Abstract

VR HMD users can observe dynamic distortion (or global pupil swim). Our earlier study correlated pupil swim to selected optic flow patterns and mathematically modeled discomfort. This study decomposed global pupil swim as a linear sum of orthogonal basis patterns for improved prediction of its perceptual effects 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 are fixed at the center of the display. As users' eyes move across the display during a typical VR experience, the correction fails and images are exposed to dynamic distortion (or global pupil swim, PS). The perceived distortion changes as functions of eye locations as well as the intrinsic optic distortion of the display.

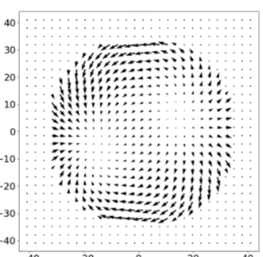

Figure 1. A pupil swim represented as a vector field map to illustrate the direction and magnitudes of distortions. Vectors are scaled by 12 for better illustration.

PS map can be illustrated as a vector field map (also called optic flow, example given in Figure 1), representing the angular shifts of projected images from their expected locations to perceived locations. Our earlier study [1] mathematically modeled perceived motion discomfort by comparing and correlating the complex optic flows with simple basis patterns. The magnitudes of correlation to these patterns were used to predict human perception of PS. Despite its tremendous value in evaluating and validating different optical system designs, it is a mapping algorithm that can generate duplicates/ over estimation in predicting discomfort complaint. For example, the sum of correlated basis patterns is usually larger than the raw pupil swim map. These duplicated representations will add noise in predicting perceptual discomfort score and prevent detailed analyses of the associative causal effects.

The objective of this study is to decompose PSs into a finite set of components which satisfy the following constraints: (a) PS can be decomposed into a linear sum of these basis components; (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 summative process. In other words, we aim to reconstruct the PS through a progressive summative process. This will increase the linearity of the model.

From our previous work [1], we believe that perception of PS can be attributed to a few key basis components. The contribution of this new study is to determine whether these components, which linearly sum to the original PS, also additively contribute to perception score following Weber-Fechner law.

2. Method of Decomposing Dynamic Distortion

2.1. Defining pupil swim (PS) basis components

PS maps (total 246) from 41 different displays and 6 positions of eye motion were decomposed into a linear sum of components. The PS maps are initially represented as a 2xN matrix 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 but this can change with the FOV of the PS and the density of the vector. The 2xN matrix is also referred to as the component indexed j ($comp_j$). 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 component 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 2.

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

The choice of basis component patterns is of great importance. We considered 3 methods: 1) basis patterns identified from previous study that has visual and motion meanings; 2) gradients of Zernike polynomials to account for optical aberration [2]; 3) basis patterns obtained from principal components analysis (PCA).

Four components previously derived from 2D translation and rotation motions along the surface of the displays were adopted [1] (Figure 3-a). In addition, another four motion-related components derived by 3D titling transformations projected through the 246 PSs were added (Figure 3-b). These four components were calculated using cluster analyses to maximize their coverage.

Zernike polynomials were considered because they represented a complete basis set of polynomials orthogonal in a unit circle. In theory, any optical aberration can be modeled as a linear combination of Zernike polynomials. In particular, the vector

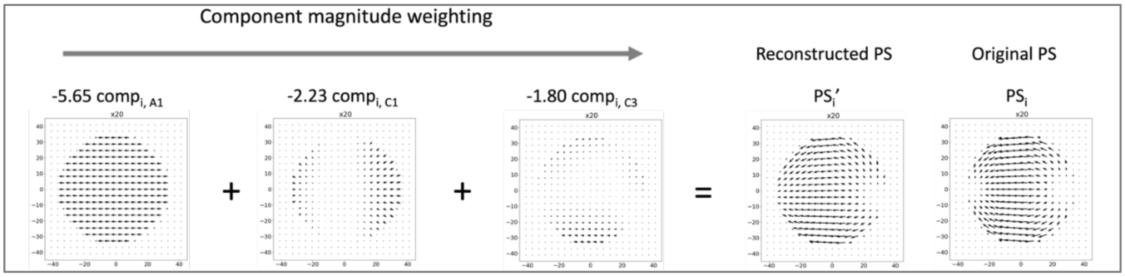

Figure 2. The components of a pupil swim (PS) and its reconstructed sum, compared with the original pupil swim. Vectors are scaled up 20 times for better illustration, mean reconstruction errors are 0.07 degrees before scaling.

polynomials based on gradients of the original Zernike polynomials by Zhao and Burge [2] seemed useful for this study. We found that the lower-order vector polynomials corresponded to simple motion patterns in Figure 3-a. However, the higher-order vector polynomials were dissimilar to the PSs from the 41 displays. Most of the vector polynomials exhibited rotational symmetry, and could not represent the asymmetrical PS maps commonly found in displays.

To determine suitable basis components to capture the asymmetric characteristics found in most PSs, PCA was employed to decompose the residue after fitting a linear model with the 8 motion-based components [3]. The initial fitting with motion-based components was regularized with L1 regularization to prevent the model from fitting the residues with sub-optimal components. This effectively provided the stopping criteria. According to the data representation for PSs and components, the residues are 1-dimensional matrices of length 2N. By applying regular PCA on this matrix, the x-magnitude and y-magnitude were analyzed as separate features instead of a single vector. Although some studies have demonstrated that 2D PCA is more efficient for decomposition, this was considered unnecessary for the current study, as more elementary components could be fitted to a wider variety of PS. [4] The first 4 principal components were selected (Figure 3-c).

2.2. Sequential fitting vs. concurrent fitting

In decomposing PS into a linear sum of vector field patterns, there was a tendency for components or groups of components 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 objective function in eqs. 2-3. The objective function fits a

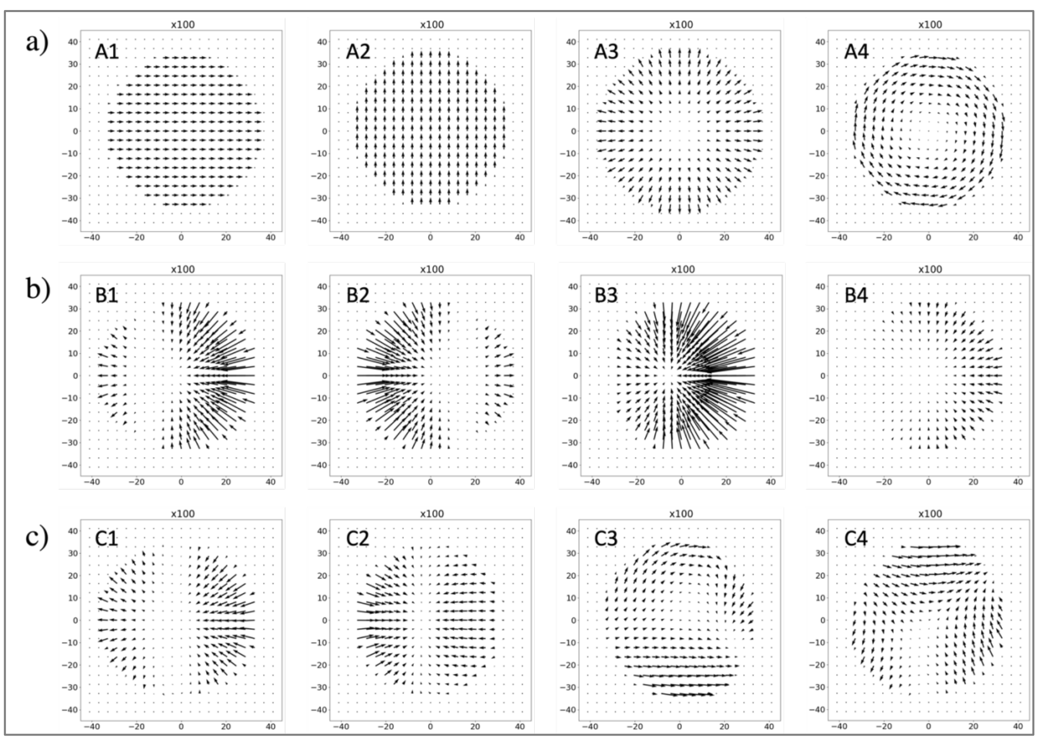

Figure 3. Components based on (a) simple motion or lower-order Zernike polynomials, (b) complex pupil swim from observing plane tilting, and (c) principal components derived from the residues after fitting with components in (a) and (b). Vectors have been scaled 100 times for better illustrations.

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. 4. In addition to sequential fitting, regularization was introduced to further reduce the collinearity between features.

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

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

$$coef_{i,j} = \propto_i^* \tag{4}$$

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.

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

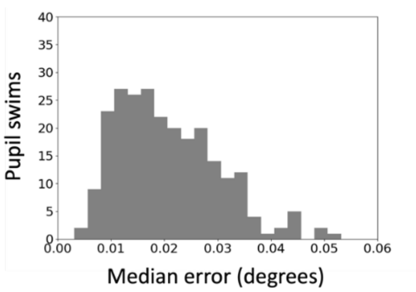

Figure 4. Median reconstruction errors (in degrees) of the 246 pupil swims using 2 to 4 components fitted from the 12 components shown in Figure 3.

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] The 12 components derived above are expected to have different perceptual weights as well. This section discusses the user study to determine the perceptual weight of each component and to build a complete perceptual model. A pilot study is being conducted to test two hypotheses: (1) decomposed components of a disturbing PS will also cause disturbance; and (2) the level of disturbance of a PS can be predicted by a weighted function of the perceptual weights of its decomposed components.

To maintain consistency with our previous study [1], pilot studies 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 the center 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.

For the pilot study, two distinct cases of PS were selected (Figure 5). Based on the selected cases, this study will focus on components $A1_+$, $C1_-$, $C2_+$, $C3_-$ and $C4_-$ defined in Figure 2 (B1 and B2 were ignored as their magnitudes were too small – see Figure 5). These components will be presented to the participant as visual stimuli at 5 different scaling factors. By decomposing the pupil swim into a linear sum of components, it is possible to have negative and positive magnitudes for each component. The sign of the coefficient affects the meaning of the component (negative coefficient for a 'zooming in' optic flow becomes 'zooming out'.) Thus, the user study considers two versions of the component, negative and positive-coefficient versions. Perception data on different scaling factors will be used to predict the perception of a component at various magnitudes, obtained from decomposition.

In total, ten participants will be presented with 52 simulated PS maps: the 5 components x 5 scaling factors x 2 signs and the 2 pupil swim cases (8 participants were used in [1] and a power analysis will be conducted to confirm the number of participants needed for the statistical analyses). The participants will be 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 1-5) (b) In this trial, comparing with extremely realistic virtual world, how would you assess the scene in terms of image deformation or disorientation? (Rate 1-3). This allows for comparison with the results of [1], which used the same means of measuring perception.

3.2. Result of the user study

In this study, we placed a heavy emphasis on identifying components that were as orthogonal as possible to additively reconstruct the pupil swim. We hypothesize that the perceptual weight of each component will also additively contribute to the perception when these components are summed. It is possible that the components offset each other in a way that cannot be explained by numerical decomposition; it's also possible that the combined perceptual weights of the components may be less than the perception score of the original pupil swim. Regardless, the results are relevant towards characterizing the perception of PS. (We are currently collecting the data for the user study. The data will be included in the final manuscript submitted by March 15, 2022.)

3.3 Perception model

The aim of the study is to use the perception scores of individual components to build a predictive model for the perceptual effect of pupil swim in different displays. By collecting data on the perception scores of each component at different scaling factors, it is expected that we can eventually predict the perception score for that component ($score'_{i,j}$) at a specified magnitude ($coef_{i,j}$). Eq. 5 is fitted using the perception data collected for each component. The log-scale is based on Weber-Fechner law, which is extensively used to relate the magnitude of physical stimuli with human perception. Specifically, Weber-Fechner takes the logarithm of the ratio of stimulus intensity to a threshold value corresponding to no perception. This study assumed that the threshold value was 0, i.e., when no stimuli was presented. Hence eq. 5 was written such that the score would be 0 when $coef_{i,j}$ is 0.

It is expected that the sum of the component scores will add up to the perception score of the complete pupil swim as in eq. 6. At the moment, eqs. 5 and 6 are hypothetical and subject to further modification.

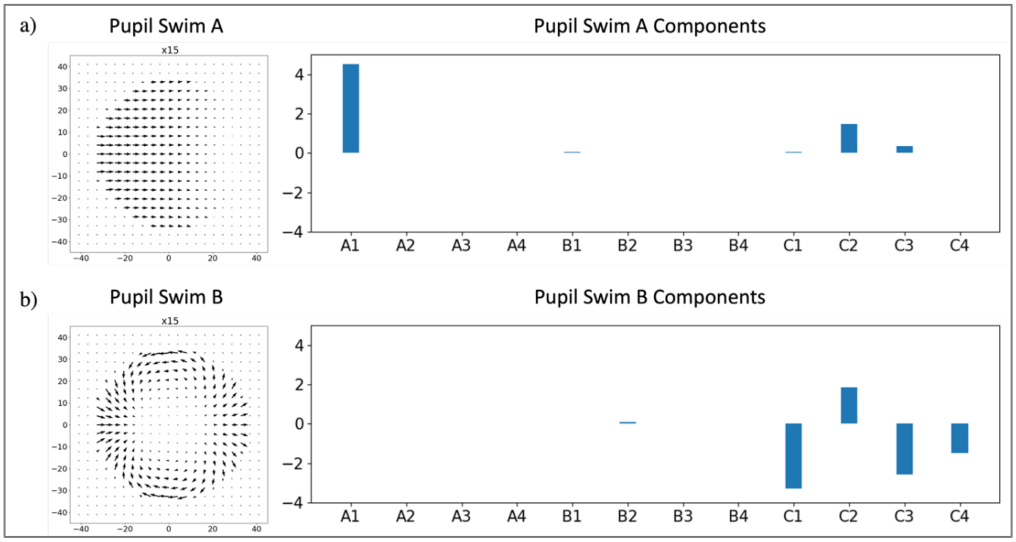

Figure 5. Selected PS cases for pilot study: the original PS and their component weightings.

$$score'_{i,j} = \mathrm{w_j}\log(1 + coef_{i,j}) \quad (5)$$

$$score'_i = \sum_j score'_{i,j} \; for \; \forall \; j \; where \; coef_{i,j} > 0 \quad (6)$$

The perception model will be validated and w_j will be determined by comparing the subject-reported score (questions a and b in Section 3.1) for PSs with the sum of predicted perception scores for each component.

4. Conclusion

Our news study decomposes pupil swim into a linear sum of optic flow patterns. Decisions on the components selected and design of the 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 component is proportional to its perceptual contribution.

The data collected in this study is being used to build an improved perception model for pupil swim. Data on user perception of the components from data-driven and optics-based methods were used to scale perception scores according to component magnitude from real PS. Following calibration and testing 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.

Note to the reviewers:

We are currently collecting the data for the user study described in Section 3.1. The data will be included in the final manuscript submitted by March 15, 2022. Upon completion of the pilot study, more user studies with the remaining 7 components and other PSs will be conducted to validate the model.

5. Impact of The Research

The mathematical perception model from our previous work has been tremendously useful in designing VR/AR products. It allows us to predict motion discomfort of optical system designs 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 this first math model for perception of motion discomfort due to pupil swim. This new work took a different approach (more mathematically sound and rigorous) from our earlier study by using a more comprehensive yet finite (12) set of components and deriving component magnitude relative to the magnitude of the original pupil swim. The individual components derived from this study have perceptual meaning, which align with the perception score reported by users. The work represents a significant contribution to an important emerging area.

6. Acknowledgements

This study is partially funded by Meta Reality Laboratory.

7. 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. 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.
3. Beghi A, Cenedese A, Masiero A. A comparison between Zernike and PCA representation of atmospheric turbulence. 2007 46th IEEE Conf Decis Control 2007:572-7.
4. Yang J, Zhang D, Frangi AF, Yang J. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Mach Intell 2004:131-7