Software Based Visual Aberration Correction for HMDs

Feng Xu*  
TNList and School of Software  
Tsinghua University

Dayang Li†  
TNList and School of Software  
Tsinghua University

ABSTRACT

When using current head-mounted displays (HMDs), users with optical aberrations need to wear the equipment on the top of their own glasses. As both the HMDs and the glasses require to be tightly attached to faces, wearing them together is very inconvenient and uncomfortable, and thus degrades user experiences heavily. In this paper, we propose a real-time image pre-correction technique to correct the aberrations purely by software. Users can take off their own glasses and enjoy the virtual reality (VR) experience through an ordinary HMD freely and comfortably. Furthermore, as our technique is not related to hardware, it is compatible with all the current commercial HMDs. Our technique is based on the observation that the refractive errors majorly cause the ideal retinal image to be convolved by certain kernels. So we pre-correct the image on the display according to the specific aberrations of a user, aiming to maximize the similarity between the convolved retinal image and the ideal image. To achieve real-time performance, we modify the energy function to have linear solutions and implement the optimization fully on GPU. The experiments and the user study indicate that without any changes on hardware, we generate better viewing experience of HMDs for users with optical aberrations.

Keywords: visual aberration correction, image deconvolution, real-time rendering, image-based rendering

Index Terms: Computing methodologies—Computer graphics—Image manipulation—Image-based rendering; Human-centered computing—Human computer interaction—Interaction techniques

1 INTRODUCTION

Head-mounted displays (HMDs) have become the most popular virtual reality (VR) equipment in recent years. They generate immersive and realistic visual experiences for users by near-eye displays, real-time head tracking, and real-time image rendering techniques. However, HMDs need to be worn on users’ faces, and they are usually tightly attached to faces, which is very inconvenient for users with optical aberrations, who have to wear glasses to correct their refractive errors (myopia, hyperopia or astigmatism). Without their own glasses, those users can not see clear images in HMDs. But with the glasses, it is difficult to wear HMDs comfortably. Very recently, some HMDs, such as the Google Daydream, use lightweight fabric to reduce the problem. However, the problem still exists.

In this paper, we propose an HMD-based real-time solution to solve this problem purely by software. To be specific, we propose a technique to calculate pre-corrected images based on the specific refractive errors of a user. When seeing the pre-corrected images through an HMD without glasses, the user can get better viewing experience than directly viewing the original images. An illustration of this system is shown in Fig. 1. This kind of technique has already been investigated for viewing images on papers or traditional displays. The observation is that the aberrations majorly cause the ideal retinal images to be convolved by certain kernels. Based on this, pre-correction can be performed by deconvolving the displayed images with the kernels, and thus the observed retinal images should be similar to the ideal ones.

However, for pre-correcting the images in HMDs in real time, there are several key challenges. First, the pre-correction needs to be performed in real time as the rendering should respond to the head motion of a user with very low latency. The current state-of-the-art pre-correction technique [13] is very time-consuming as it solves an L1 optimization to estimate all the pixel values of one pre-corrected image. Second, the traditional pre-correctional model can not be directly used for HMD due to the different viewing scenarios where the pre-corrected images are displayed on two near-eye displays which cover the users’ whole viewing range. The imaging system of users with refractive errors needs to be revisited to get a correct model for the pre-correction on an HMD. Note that Oshima et al. [14] also try to improve viewing quality in HMD. However, they only try to make the viewing image sharper by a simple aperture model.

In this paper, we propose an end-to-end system which overcomes the two difficulties. First, to achieve real-time performance, we formulate the image deconvolution problem as an L2 optimization. In our experiments, the L2-based optimization generates similar results compared with the previous total variation (TV)-based solutions. Furthermore, we implement a fast Fourier transform (FFT)-based GPU solver to solve the L2 optimization in 0.08 ms for each displayed image, which guarantees the real-time performance of the whole system. To keep a reasonable color range, we propose a fast parameter selection scheme based on binary search. Second, for users with refractive errors, we build a clear relationship between the convolution kernels and the parameters of the refractive errors, as well as the viewing distance. Thus it is possible to derive the correct kernels when the user is viewing through an HMD, which is never fully considered by previous works focusing on viewing a photo or a traditional display. From the user study, our method gives better visual experiences to users with refractive errors.

The specific contributions of our technique are listed below:

- A system that corrects the specific refractive errors of users when they are using HMDs without their own glasses. To the best of our knowledge, this is the first system that uses purely software to correct the errors for HMD-based VR displays. And this technique is totally compatible to other commercial HMDs.

Figure 1: In an HMD-based VR application, our systems generates pre-corrected videos (on the left) according to the viewing content and the aberration parameters of the user in real time. The user with optical aberrations do not need to wear his/her own glasses but will still get a clear view as shown on the bottom right.
2 Related Works

Vision aberration is a common problem for human beings [3]. Besides wearing glasses, researchers in image processing and computational display are also working on handling this problem. For examples, some works focus on hardware and modify the display to pursue better quality for users with aberrations. Huang et al. [7] use multi-layer displays to correct the aberrations. Pamplona et al. [15] and Huang et al. [9] explore the theory of light field to compensate the aberrations. Barsky et. al. [2] propose both single layer and multi-layer solutions for the purpose of aberration correction. Even though these techniques solve the problem in both theory and practice, they require to specially design and fabricate the displays, and thus they are not able to work with current displays or to be largely used in our daily applications.

To be easily apply to our daily displays, some works also focus on pure software solutions. Montalto et. al. [13] is the latest work on this topic. It uses total variation to constrain the image deconvolution process and generates a pre-corrected image which is better for users with optical aberrations. The kernel used in the deconvolution is related to the specific aberrations of a user. Before this work, Yellott and Yellott [20] indicate the underlying theory of the properties of aberration correction by Fourier transform and also show the low contrast of the pre-corrected image is inevitable. Alonso and Barreto [1] transfer the input image into the frequency domain to do the pre-correction. Later, this work is improved by [12]. In general, all the pure software-based solutions require to solve an image deconvolution problem. Due to the computational cost, all these works focus on images but not videos and can not achieve real-time performance as we do.

Different aberration parameters correspond to different Kernels (also known as PSFs). Their relationship is a key for deconvolution-based correction techniques. Constructing the kernels is not an easy task and requires a lot of knowledge related to photics and ophthalmology. Some works have summarized the procedure of estimating kernels from wavefront aberration functions [6, 18, 19]. Recently, Krueger et. al. [11] make a step further to formulate the correlation between aberration parameters and the coefficients of Zernike polynomials, which is used to model the wavefront aberration function. However, the previous works neither consider the viewing distance into account nor propose a delicate process about calculating kernels from the aberration parameters as we do.

3 Method

For real-time pre-correction in VR, we follow the theory which models the retinal image $I^r$ in human vision system as the convolved version of the displayed image $I^d$:

$$I^r = I^d \otimes K.$$  (1)

Here, the blur kernel $K$ is related to the specific optical aberrations of a user. If we directly display the original image $I^o$, we get

$$I^o = I^d \otimes K.$$  (2)

So $I^r$ is a blurred version of $I^o$, which is the reason that users with aberrations always observe blurred scenes.

On the other hand, if we know the blur kernel $K$ of a specific user, we could calculate a pre-corrected image $I^p$ as follows:

$$I^p = I^o \otimes K^{-1}.$$  (3)

And then, if we display this pre-corrected image $I^p$ rather than $I^o$, we get

$$I^d = I^p \otimes I^r = (I^p \otimes K^{-1}) \otimes K = I^o.$$  (4)

In this case, the user with aberrations still gets the original $I^o$ image on retina. This pre-correction method is also used in image correction for 2D displays [6], but we will adapt it to handle real-time videos for near-eye VR displays.

3.1 Kernel Construction

In this sub-section, we will discuss how to construct $K$ for a specific user. The $K$ is related to the wavefront function of an eye, which is usually represented by Zernike polynomials in ophthalmology. So we will first introduce the Zernike polynomials, and then how to calculate the wavefront of the user with optical aberrations. Finally, we calculate $K$ according to the wavefront.

3.1.1 Zernike Polynomials

The Zernike polynomials are a set of functions that are orthogonal over a unit circle. They are useful for describing the shape of an aberrated wavefront function of the pupil of an optical system. There are several ways for representing Zernike polynomials. Here we adopt a double indexing scheme ($Z^m_n$, where $n$ is the order and $m$ is the frequency). Such a scheme is defined as [17]:

$$Z^m_n(r, \theta) = \begin{cases} N^m_n R^m_n(r) \cos(m\theta) & \text{for } m \geq 0 \\ -N^m_n R^m_n(r) \sin(m\theta) & \text{for } m < 0 \end{cases}$$  (5)

where $N^m_n$, $R^m_n$ and the sinusoidal functions stand for the normalization factor, radial component, and azimuthal component, respectively. In ophthalmology, they can be applied directly to the wavefront evaluation of the pupil. And the radial degree $n$ is used for classifying aberrations as lower-order ($n \leq 2$) and higher-order ($n > 2$).

3.1.2 Wavefront Function

The wavefront is the surface that is perpendicular to the light rays at the same phase. The rays are converged by the combined optical system of the eye [6]. For a user with optical aberrations, a complete wavefront function can be represented by both low-order and high-order Zernike terms. In this work, we only consider the low-order aberrations like defocus (including myopia and hyperopia, astigmatism), since previous work has indicated that low-order aberrations alone are responsible for about 90% of one’s total loss of visual acuity [16]. We obtain the aberrated wavefront function $W(x, y)$ from prescription data as [5]:

$$W(x, y) = \sum_{i=-1}^{1} c_{2i} Z_{2i}^2(x, y)$$  (6)

where

$$c_2^2 = \frac{R^2 C \sin(2A)}{4\sqrt{6}}$$  (7)

$$Z_{2,-2}^2(x, y) = 2\sqrt{6}xy$$

$$c_0^2 = \frac{R^2 (S + C/2)}{4\sqrt{3}}$$  (8)

$$Z_{2,0}^0(x, y) = \sqrt{3}(2x^2 + 2y^2 - 1)$$

$$c_2^2 = \frac{R^2 C \cos(2A)}{4\sqrt{6}}$$  (9)

$$Z_{2,2}^2(x, y) = \sqrt{6}(x^2 - y^2)$$
and \( c_5^2, c_2^0, \) and \( c_3^1 \) are the coefficients of the Zernike polynomials corresponding to oblique astigmatism \((Z_5^2)\), defocus \((Z_0^0)\), and vertical astigmatism \((Z_3^1)\) respectively. \( S \) and \( C \) are respectively the sphere and cylinder values from the prescription, specifying the optical powers in diopters. \( A \) is cylinder axis expressed in degrees, which is also from the prescription [4]. The \( S \) is related to defocus abberation \((S < 0 \text{ for myopia and } S > 0 \text{ for hyperopia})\), while the \( C \) and \( A \) is related to astigmatism abberation. \( R \) is the radius of the subject’s pupil (an aperture, in general) measured in mm, and \( c_5^2, c_2^0, \) and \( c_3^1 \) are in \( \mu m \).

**Viewing Distance** Previous works [5,11] directly use Eq. 7 to construct the coefficients of the Zernike polynomials, with the measured parameters \( S, C \) and \( A \) of a user. However, we observe that in Eq. 8 is actually not a constant, but related to the viewing distance \( d \) from the eye to the object where the user is focusing. So we propose to calculate \( S \) according to \( d \). Since for an HMD, users are always focusing on a virtual display with a fixed \( d \), we can pre-compute \( S \) without adding any extra work load to the online stage of our method.

To better understand \( S \), we first present the well-known thin lens formula:

\[
\frac{1}{d} = \frac{1}{f} - \frac{1}{d'},
\]

which is also the imaging model of a human eye. Here, the distance from the object to the lens and from the lens to the image are denoted as \( d \) and \( d' \), respectively. The \( f \) is the focal length. The left side in Eq. 10 is determined by the scene while the right side is related to the eye. In ideal cases, Eq. 10 should be satisfied, but for people with myopia or hyperopia, Eq. 10 is not satisfied for some ranges of \( d \). And the \( S \) is used to measure the aberration and is defined as

\[
S = \frac{1}{d} - \left( \frac{1}{f} - \frac{1}{d'} \right),
\]

For an aberrationless eye, it should be able to focus on objects in the range of \([0.125, +\infty)\), so \( \left( \frac{1}{f} - \frac{1}{d} \right) \) should be able to be adjusted in the range of \([0,8] \) accordingly [8]. However, for myopia, this value is not able to reach 0, and \( S \) will be negative and its maximum absolute value happens when \( d = +\infty \). The optist-measured \( S \), denoted as \( S_m \), is actually the \( S \) at this situation. Similarly, for hyperopia, \( \left( \frac{1}{f} - \frac{1}{d} \right) \) can not reach 8, and \( S \) will be positive and its maximum absolute value (also denoted as \( S_m \)) happens when \( d = 0.125 \).

Since we know that \( S_m \) is the extreme case of \( S \) for a user \((d = +\infty \text{ for myopia and } d = 0.125 \text{ for hyperopia})\), we derive the correlation between the actual \( S \) and \( d \) for myopia and hyperopia as:

\[
S = \begin{cases} 
S_m + \frac{1}{2} & \frac{1}{d} < |S_m| \\
S_m - \left( 8 - \frac{1}{2} \right) & 8 - \frac{1}{2} < |S_m| \\
0 & \text{otherwise}
\end{cases}
\]

This also explains why myopia still gets clear vision at near distance while hyperopia does the opposite. For myopia, when \( d \) is small, and \( \frac{1}{d} \) is larger than \( |S_m| \), the eye is able to focus on the object and \( S = 0 \); with \( d \) increasing, \( \frac{1}{d} \) becomes smaller than \( S_m \), and then \( S \) decreases until reaching \( S_m \) which indicates the biggest aberration, happening when \( d \) goes to \(+\infty\).

### 3.1.3 Kernel Calculation

The above method outputs a wavefront function which will be used to calculate the kernel \( K \), also called point spread function (PSF). \( K \) is computed given the pupil size as follows:

\[
K_{(x,y)} = \left| FT \{ P_{(x,y)} e^{-i2\pi W_{(x,y)}} \} \right|^2.
\]

So \( K \) is given by the Fourier transform of the generalized pupil function, a composition of the amplitude function \( P_{(x,y)} \) which describes the transmission through the pupil and is defined as 1 within the pupil area and 0 elsewhere, and the interferograms \( e^{-i2\pi W_{(x,y)}} \) which is from all possible pair of points across the wavefront surface \( W_{(x,y)} \).

### 3.2 Pre-corrections

In this subsection, we will discuss the method to perform Eq. 3, with a given kernel \( K \) estimated in the previous subsection. In theory, the deconvolution problem in Eq. 3 is an ill-posed problem. To solve it, previous work has proposed to involve Hyper-Laplacian Priors [10]. Considering the real-time requirement for VR application, we use a gradient-based prior represented by \( L_2 \) norm, which provides a direct analytic solution to speed up the algorithm.

The pre-corrected image is calculated as:

\[
P^p = \text{argmin} \sum_{i=1}^{N} \left( \lambda (I^p \otimes K - I^o)^2 + \sum_{j=1}^{2} (I^p \otimes f_j)^2 \right).
\]

\( \lambda \) is a weight to balance the two terms. Here we use two first-order derivative filters \( f_1 = [1 - 1] \) and \( f_2 = [1 - 1]^{T} \) in the regularization.

Given the quadratic equation, we get the ideal \( P^p \) as follows:

\[
P^p = F^{-1} \left( \frac{\lambda F(K) \ast F(I^o)}{F(f_1) \ast F(f_1) + F(f_2) \ast F(f_2) + \lambda F(K) \ast F(K)} \right)
\]

where \( F \) denotes the Fourier Transform and \( F^{-1} \) denotes the inverse Fourier Transform, * is the complex conjugate, \( \ast \) is the component-wise multiplication, and the division is also calculated component-wise.

As the kernel \( K \) does not change for a specific user and is estimated by previous method, a large part of the calculation in Eq. 17 can be pre-performed. We denote this part as:

\[
M = \frac{\lambda F(K) \ast}{F(f_1) \ast F(f_1) + F(f_2) \ast F(f_2) + \lambda F(K) \ast F(K)}
\]

Then in the online stage, the pre-correction of an image can be computed by only 1 FFT, 1 inverse FFT, and 1 per element matrix-multiply:

\[
P^p = F^{-1} (M \ast F(I^o)).
\]

This can be performed very efficiently using GPU.

However, there is no guarantee that the solved pixel values of \( P^p \) still lie in the meaningful range (denoted by \([0,1]\) as usual). There might be negative or extreme large values in \( P^p \). Clipping or normalizing the pixel values are two ways to solve the problem. However, if the obtained pixel range is far from the meaningful range, they both generate artifacts. We solve this problem inspired from our experiments, where we found that \( \lambda \) value heavily influence the range of pixel values in \( P^p \).

Fig. 2 shows how the range increases with \( \lambda \) under different kernels. In practice, we propose to use binary search to decide the value of \( \lambda \). First, we set an acceptable pixel value range according to large amount of experiments, typically 1 - 4. Then we search for the value satisfying this range with maximum \( \lambda \).

### 3.3 Implementation Details

We implement our whole system in GPU and adapt it into Oculus Rift DK2. It runs on a desktop PC with one NVIDIA GTX 1080. Our system is implemented by C++ and CUDA to ensure the efficiency. We have also developed a matlab version for comparisons with other techniques. To generate RGB color images which contain 3 channels, we perform pre-correction per channel so the algorithm runs 3 times for every frame.
4 Experiments

As a real-time technique, we first discuss the performance, especially the online performance, by comparing with some alternative solutions. Then, we show our results with different aberration parameters of different users. To evaluate the effectiveness of our technique, we also compare it with the state-of-the-art 2D image correction technique. We achieve similar quality on 2D images but ours is much faster and is adapted to VR applications. Finally, we discuss our results shown in the accompanying video and our user study.

4.1 Performance

We test the performance of our online system with some alternative solutions using different deconvolution techniques. For fair comparison, we run all methods on Matlab, with a 3.40 GHz Intel E3-1231 CPU and 16GB RAM. We perform the algorithms with various kernels and various image sizes, and the numbers in Tab. 1 are the mean time costs.

<table>
<thead>
<tr>
<th>Image Size</th>
<th>FIDHLP(L1)</th>
<th>FIDHLP(L2)</th>
<th>Ours(L2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200x200</td>
<td>0.174</td>
<td>0.172</td>
<td>0.014</td>
</tr>
<tr>
<td>400x400</td>
<td>0.365</td>
<td>0.365</td>
<td>0.080</td>
</tr>
<tr>
<td>512x512</td>
<td>0.519</td>
<td>0.524</td>
<td>0.136</td>
</tr>
<tr>
<td>1024x768</td>
<td>1.442</td>
<td>1.448</td>
<td>0.430</td>
</tr>
<tr>
<td>400x400</td>
<td>3.782</td>
<td>3.766</td>
<td>1.167</td>
</tr>
</tbody>
</table>

Fast image deconvolution with Hyper-Laplacian Priors (FIDHLP) [10] iteratively performs two steps in the minimization. They support both $L1$ and $L2$ regularization in deconvolution. On the other hand, our algorithm uses an $L2$ regularization and only uses an analytic step, so it performs faster than the other methods.

The GPU-based implementation is also evaluated. Here we compare the C++ version w/ and w/o GPU (Tab. 2). We can clearly see that the GPU version is extremely fast and real-time performance is achieved.

<table>
<thead>
<tr>
<th>Image Size</th>
<th>CPU</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>400x400</td>
<td>0.012531</td>
<td>1.198e-5</td>
</tr>
<tr>
<td>800x600</td>
<td>0.02568</td>
<td>1.190e-5</td>
</tr>
<tr>
<td>1280x720</td>
<td>0.05841</td>
<td>0.817e-5</td>
</tr>
<tr>
<td>1920x1080</td>
<td>0.128654</td>
<td>0.808e-5</td>
</tr>
</tbody>
</table>

We also compare the pre-correction quality of FIDHLP (L1&L2) and our method. As shown in Fig. 3, all the three pre-corrected images performs better than the original image, judging by the quality of the simulated retinal images. The two results based on L2 are quite similar, while the one based on L1 is slightly different. But it is difficult to say which one is better.

4.2 Evaluations

We evaluate the effectiveness of our pre-correction for users with different aberration parameters. Fig. 4 and 5 show the different results caused by changing the three aberration parameters $S$, $C$ and $A$, respectively. We can see that even though the pre-corrected image looks different from the original image, the simulated retinal image is reasonable, which means users with the corresponding aberration parameters will get better viewing experience. Notice that with the increasing of the absolute values of the aberration parameters $S$ and $C$, the simulated retinal image becomes worse, which indicates that our technique still can not handle very high aberrations.

4.3 Comparisons

We compare the result of our method with the state-of-the-art visual correction technique [13] in the accompanying video. Since the method [13] only handles ordinary 2D images, we perform the comparisons just on videos. From the comparison, we can see that our method generates similar quality compared with the method [13], indicating that using an $L2$ term with an online adapted $A$ also gives good results. Notice that the method [13] does not consider VR applications and does not provide solutions to map aberration parameters to kernels for pre-correcting VR frames. Also, their technique is far from real-time.

4.4 Results

We present sequential results in the accompanying video where we can see that the pre-correction is fully performed in real time, when a user is using an HMD to view a VR scene. The corresponding simulated retinal image shows that the user will get a much clear viewing experience even without glasses. Fig. 1, 4 and 5 also demonstrate this point.
4.5 User Study

In our user study, there were 21 participants (12 males and 9 females) who had optical aberrations with known parameters. The ages of the participants ranged from 18 to 28. And the participants were with myopia or both myopia and astigmatism. The myopia parameter $5\sigma$s of the participants ranged from $-1.0$ to $-6.0$, and the astigmatism parameter $C$ ranged from $0$ to $-3.0$. We are sorry that we have not recorded the exact values of each participant.

In the user study, we asked the participants to score a stereo pair (watched though our HMD) by clearness. We did not define clearness for the participants and the participants understood clearness by their own. We asked the participants to score 0 for totally unclear scenes and 5 for totally clear scenes. For one scene, each participant watched the original stereo pair and the pre-corrected pair (pre-corrected with the parameters of the participant) consecutively but in a random order. And each participant watched 10 scenes, including 3 landscapes, 3 portraits, 2 words and 2 indoor scenes, also in a random order.

As both the scores for our pre-correction and the original images satisfy normal distributions (Shapiro-Wilk test with $W(21) = 0.87$, $p \leq 0.01$ for our pre-correction and $W(21) = 0.87$, $p \leq 0.01$ for the original images), we performed a two-tailed $t$-test on the data. The test showed that the mean results are significantly different ($t(40) = 6.82$, $p < 0.0001$). We find that our pre-corrected images have higher mean score ($M = 3.13$, $SD = 0.64$) than the mean score of the original images ($M = 1.87$, $SD = 0.55$).

4.6 Limitations

As indicated by [2], as all displays have limited dynamic ranges, our pre-correction can never achieve the same quality compared with wearing the user’s glasses. Also, in practice, the quality of our technique will be further reduced if the parameters of the aberrations are not accurate. In our user study, we expect “clearness” is judged by the quality of a whole image, not just the sharpness of image edges, which is not guaranteed. More concrete questions should have been asked to the participants to pursue a more solid conclusion.

5 Conclusions

This paper proposes a real-time system that achieves visual correction for near-eye VR applications. With this technique, users with optical aberrations can view clearer VR content without wearing their own glasses, which is very convenient for HMD users. Our technique is purely based on software, and thus it can be used in all commercial HMDs. The system is also flexible to be used for different users who can either input their aberration parameters or change the parameters online until getting a good view. The system is performed in real time with a highly efficient GPU solver.

Acknowledgments

This work was supported by the NSFC (No.61727808, 61671268). We wish to express our thanks to the reviewers for their insightful comments.

References