Gaze-Sensing LEDs for Head Mounted Displays

Kaan Akşit   **Jan Kautz**   **David Luebke**

NVIDIA

2701 San Tomas Expy, Santa Clara, California 95050, USA
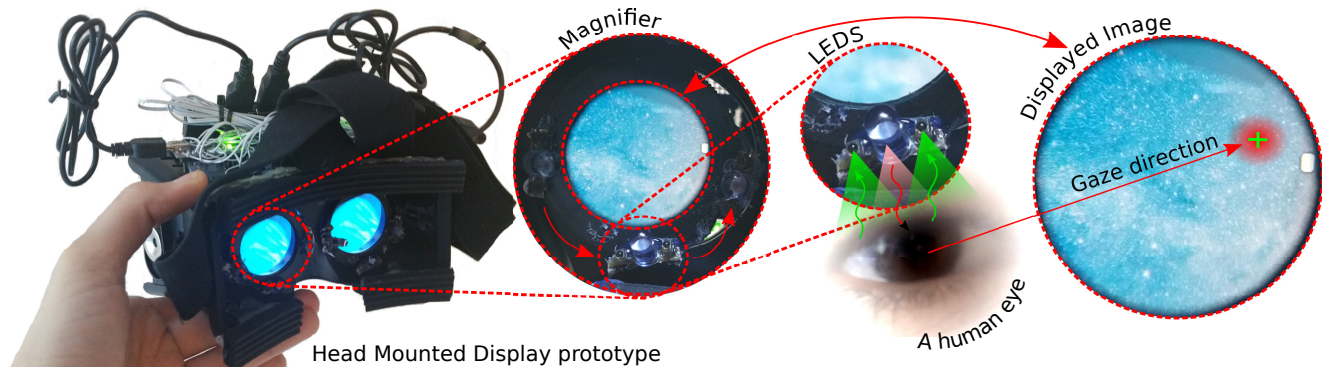
kaksit@nvidia.com

Figure 1. Adding gaze tracking to Head Mounted Displays (HMDs) by exploiting sensing capability of Light Emitting Diodes (LEDs): An off-the-shelf HMD system is modified to demonstrate gaze tracking support only with LEDs (left). LEDs are placed in front of a human subject's eyes for sensing and illumination. The prototype comprises only a smartphone, two microcontrollers, and a low number of LEDs with a supervised adaptive pattern recognition algorithm. The overall system estimates the gaze direction of a human subject in real-time over a smartphone screen seen through a pair of magnifier lenses.

ABSTRACT

We introduce a new gaze tracker for Head Mounted Displays (HMDs). We modify two off-the-shelf HMDs to be gaze-aware using Light Emitting Diodes (LEDs). Our key contribution is to exploit the sensing capability of LEDs to create low-power gaze tracker for virtual reality (VR) applications. This yields a simple approach using minimal hardware to achieve good accuracy and low latency using light-weight supervised Gaussian Process Regression (GPR) running on a mobile device. With our hardware, we show that Minkowski distance measure based GPR implementation outperforms the commonly used radial basis function-based support vector regression (SVR) without the need to precisely determine free parameters. We show that our gaze estimation method does not require complex dimension reduction techniques, feature extraction, or distortion corrections due to off-axis optical paths. We demonstrate two complete HMD prototypes with a sample eye-tracked application, and report on a series of subjective tests using our prototypes.

Author Keywords

gaze trackers; supervised learning; light emitting diodes

ACM Classification Keywords

H5.2 User Interfaces: Input devices and strategies

INTRODUCTION

Head Mounted Displays use a variety of sensors to provide immersive interaction with engaging virtual reality experiences. Emerging consumer HMDs for VR use gyroscopes, accelerometers, various optical sensors, and so on, either embedded in the headset or grouped into an external unit. These sensors track head orientation, user motions, whether or not a user is wearing the device, to provide user controls, and in short to enhance the user's experience in a virtual world.

Using gaze as an input modality [11, 20] can be natural, fast, and has the potential to enhance the user experience in an HMD system. While decoupled non-mouse input modalities as a pointing mechanism for large displays has proven awkward, gaze-supported target acquisition has been found fast and natural [34, 39, 42, 43]. Hence, industry has been pushing towards employing gaze tracking as a key part of future HMD systems. Current prototype solutions are typically expensive (e.g., FoveVR, Tobii-Starbreeze, SMI's HMD upgrade for Oculus Rift), but there is also a growing interest towards low-cost gaze trackers with some promising results [35, 36, 12, 21].

We believe major challenges remain for conventional gaze tracking hardware and software:

1. Conventional gaze trackers rely on imaging techniques, which have relatively high power demands and may not be truly suitable for low-power mobile solutions;

2. Imaging equipment typically introduces complexity in the software, adding an extra image processing block in the pipeline; and,

3. Conventional gaze trackers work with high-dimensional inputs (high-resolution images) and thus introduce latency at multiple stages, including the capture hardware, communication protocols, and image processing-based gaze estimation pipeline.

We are motivated to address these issues by decreasing hardware and software complexity, seeking the simplest possible solution that provides useful-quality eye tracking. We use Light Emitting Diodes (LEDs) as the core of our gaze tracking hardware design. Commonly used as illumination devices, LEDs in fact have well-known bi-directional characteristics [7, 33] with the ability to sense and emit light. Since they are also inexpensive, can perform color selective sensing, can illuminate and sense from the same physical location, and are easily controlled without a dedicated custom circuitry, LEDs constitute a good choice as a hardware solution.

In this paper, we describe how to take advantage of LED's bi-directional characteristics to allow both light capture and illumination. We position an array of LEDs in front of a human subject's eyes for the task of gaze estimation, in the context of an HMD system. As a human subject observes a scene projected at a fixed virtual plane, we illuminate the subject's eye with infrared (IR) light from different perspectives in a time multiplexed fashion. We capture and digitize the intensities of IR light reflecting off the eye from different perspectives with high refresh rates. Using our supervised adaptive pattern recognition implementation, we accurately estimate the gaze location of the human subject at the virtual image plane.

Contributions
All-in-one Sensor Technology: The core item in our design, an LED, can be used as an illuminator or as a sensor for integrating light intensity over a certain field of view (FOV) and exposure interval. To our knowledge, this is the first time that LEDs are used in a gaze tracker both for capture and illumination.

Less Hardware and Lower Cost: Our design employs fewer components. To our knowledge, our design has the simplest electronics design yet demonstrated. We believe that our hardware design leads the category of ultra-low cost gaze tracker. While decreasing cost, we show that accuracy and sampling rates can match existing devices.

Supervised Adaptive Gaze Estimation Algorithm: We use a supervised adaptive pattern recognition algorithm with our sensor technology. We show that our method has low computational demands and runs fast enough on a conventional mobile platform. We compare against the most common methods from the literature, and demonstrate equivalent or better accuracy with a much simpler hardware.

Complete Prototype: We demonstrate two different complete custom HMD systems, created by modifying off-the-shelf items. We also demonstrate example eye tracked applications with our prototype, and provide results of an informal subjective test with human subjects.

Benefits and Limitations
Benefits: Today's HMD manufacturers are competing to create the least bulky HMD. One way to achieve this goal is to reduce the bulk caused by each component of the system. Off-the-shelf LEDs are lighter and smaller than off-the-shelf cameras. Thus, our proposed method uses less volume and less weight than an alternative using single or multiple camera(s).

LEDs also offer an advantage in power consumption over camera based solutions. Possible heat dissipation related issues caused by cameras are also avoided with our low power consuming system. Additionally, our gaze estimation pipeline is computationally light-weight. Thus, our proposed technique is a good match for battery operated applications. Since we only employ a small number of sensors, compared to camera-based solutions (which have millions of sensing pixels), latency within the hardware and software is reduced and results in higher sensing rates.

Our method offers good accuracy. We show that mean angular error can be as low as 1.1^o with a median angular error of 0.7^o.

Limitations: While providing sufficient accuracy for a variety of VR applications, due to their poor sensing characteristics compared to photodiodes LEDs may not be suitable choice for applications requiring very high accuracy ($error < 0.5^o$), such as in psychological, neurological and reading research [10].

Our proposal requires a larger amount of calibration data than conventional imaging based gaze tracking systems, thus the initial phase of calibration is comparably longer.

Conventional HMDs can shift slightly on a subject's face, commonly due to fast head movements. In turn, this causes the sensing hardware to shift relative to the subject's eyes. Using the original calibration data makes gaze estimates less reliable. We share this common problem with other gaze trackers. We choose to recalibrate in such cases, which can be burdensome to users.

RELATED WORK
Sensing with LEDs: LEDs are known to be reliable light sensors, and have been used in other bi-directional systems such as visible light communication (VLC) systems [6, 37], temperature and pressure sensing systems [16], bidirectional reflectance distribution function (BRDF) measurement systems [2], color sensing and illumination mimicking applications [17, 18], and human shape sensing with VLC [19, 1].

To our knowledge, this is the first time LEDs are used as sensors in a gaze tracker application.

Gaze Tracker Hardware: The foundation of any gaze tracker is the hardware used to capture the raw data. The

"scleral search coil," for instance, requires the user to wear a contact lens with copper coils on top while exposed to an alternating magnetic field [31]. The scleral search coil is reported to provide 15 arc seconds of angular resolution with $1000\ Hz$ sampling capabilities. Despite this impressive accuracy, the scleral search coil fails to to provide a comfortable user experience and is not an option for consumer products. Another common methodology is electro-oculography (EOG) [22], which uses electrodes placed around the eye along the horizontal and the vertical axes. However, the technique is generally known to suffer from resolution limitations, and shares the same discomfort problem with a scleral search coil.

Gaze trackers using light are categorized as imaging and non-imaging optical devices. Relatively few imaging gaze trackers have explicitly tackled the problem related to power-consumption and computation-intensity: iShadow [25], an imaging system, reported a 70 mW power consumption and $30\ Hz$ estimation rate with an angular resolution of 3 degrees. OLED-on-CMOS technology [45] has been demonstrated for capturing an eye's images from a large spectrum of light ($600 - 900\ nm$) and displaying those images, however, in its current state, the technology's display resolution is very limited, and negatively effecting a display's fill factor.

Non-imaging gaze trackers provide another approach. Naugle and Hoskinson [27] demonstrated two different methods using binary gaze tracking with a pair of LEDs and a photodiode for a low field-of-view (FOV) ($< 20^o$) HMD. While this setup reduces the power consumption of the display by over 75 %, it only provides limited information on whether the user is actively wearing the display or not.

More closely related to our proposal, Topal et al. [41] demonstrate a low-computational overhead, non-imaging gaze tracker based on IR light-emitting diodes and IR sensors around an eyeglasses frame. Another similar non-imaging gaze tracker [8] contains a ring of LED emitters that are stimulated in sequence. Unlike our proposal, the tracker contains photodiode sensors, which are measured for each LED stimulation. These systems, however, require head pose to be completely fixed or stabilitized with a mechanism such as a bite-bar. In contrast, our prototype hardware does not depend on fixed head pose, does not use photodiodes nor sophisticated or custom light-sensing circuits, and is embedded in an HMD context without producing occlusion in front of the eye. Our algorithmic approach is also simpler and less demanding on the hardware, using Gaussian Regression Processes (GPR) [30] for gaze estimation rather than a model-based linear or non-linear mapping — thus avoiding tuning of algorithm parameters.

We believe our solution provides the simplest hardware design yet for a gaze tracker embedded inside an HMD.

Adaptive Gaze Estimation Algorithms: Because each human subject introduces multiple differences for the input of a non-imaging gaze tracker, we believe that the task of gaze estimation, in this case, is a better match for an adaptive supervised pattern recognition technique rather than finding a good model for noisy sensor input. Traditionally, supervision of such pattern recognition techniques in gaze trackers has been incorporated into a calibration procedure.

Within the realm of imaging gaze trackers, the relevance vector regression (RVR) [23] method was found to be more effective than support vector regression (SVR). RVR shares the same functional form as SVR. Unlike SVR, however, RVR tries to find the weights of the regressor from training data. We have experimented with SVR with radial basis function (RBF) as the kernel, and we found that GPR with a non-parametric similarity measure provides an accurate result without the comptuational burden of finding the correct weights using RVR. This is important since finding correct weights online may not be a feasible task for limited processing power on mobile devices.

Noris et al. [29] show that imaging gaze trackers can benefit from GPR for calibration-free operation. Their implementation used principle component analysis (PCA) for feature dimension reduction. As our technique depends on a low number of sensory inputs, we did not find dimension reduction necessary for our gaze estimation algorithm, and instead we use GPR directly for the task of gaze estimation.

Using cost effective hardware (albeit with a relatively power hungry imaging gaze tracker), Sewell and Komogortsev [38] demonstrated effective usage of a neural network for offline training, and demonstrated a relatively low error ($< 3.68^o$). Our method achieves a much lower error, and we also demonstrate that in specific applications, online training can be enabled.

SYSTEM OVERVIEW

Gaze tracking devices for HMDs estimate where a user is gazing relative to a virtual image plane as seen through the HMD's lenses. The task of gaze estimation is a layered problem that requires the integration of sensors, optical design, image/signal processing, and electronics design. The following sections describe our approach in addressing each of these tasks.

Sensing with LEDs

At the core of our design, LEDs are used to capture light and to illuminate the eye. LEDs with infrared light emission are typically used in HMD gaze tracker hardware, since humans are insensitive to IR illumination [14]. A human eye's cornea has similar absorption and reflection characteristics in the near IR as in visible light [44]. Furthermore, IR LEDs have a narrow bandwidth (typically $\sim 50\ nm$), avoiding cross-talk with other wavelengths.

LEDs provide illumination when a forward voltage is applied to their two electrical terminals. However, LEDs can also act as photodetectors [7] by following three steps: ① Apply a reverse voltage pulse for a short time duration. ② Discharge LED's capacitance immediately afterwards. ③ Measure the voltage across LED to determine how much discharge of capacitance took place after a certain time. Figure 2 illustrates each of the explained steps.
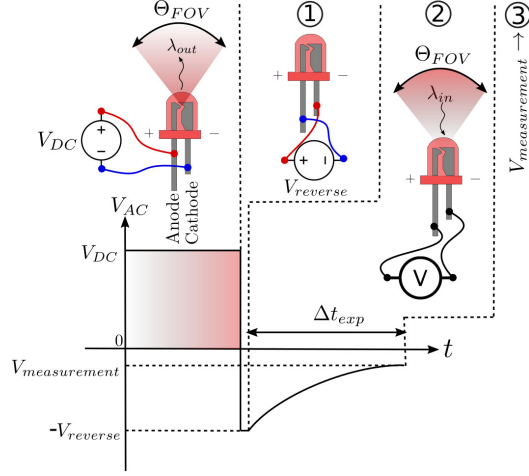
Figure 2. **Illustration of different modes of a bi-directional LED (from left to right): (1) applying a forward voltage of V_{DC}, in which the LED emits light with a wavelength of λ_{out} and an emission cone angle, Θ_{FOV}; (2) applying a reverse voltage pulse, $V_{reverse}$, for a short time duration, discharging LED with incoming light that has a wavelength of λ_{in} for a specific time, Δt_{exp}, with a reception cone angle of Θ_{FOV}; and (3) measuring a voltage, $V_{measurement}$, from the LED.**

These steps can be easily implemented by wiring the LED to a microcontroller for full control over charge, discharge, and measure. The microcontroller is controlled by and relays measurements to a host. Typically, LEDs are most sensitive to wavelengths λ_{in} that are shorter than their emission spectrum (so $\lambda_{in} < \lambda_{out}$) [16]. Thus, larger exposure times are required if LEDs with the same emission spectrum are used. To achieve lowest possible latency with a given configuration, we choose to use different LEDs that have intersecting emission and sensing spectra in the IR range.

Eye safety is also important when a user is exposed to infrared radiation; Δt_{exp} and maximum irradiance of an LED must be considered according to safety regulations for infrared light sources [3].

Positioning of LEDs: To use LEDs for capture and illumination, LEDs must be placed at specific locations in front of an eye, or optical path must be relayed to have the same effect. Currently the most common configuration for the optics of a commercial HMD, as shown in Figure 3, uses a pair of magnifier lenses placed in front of a display to create a virtual image at some distance in front of a user. Such an HMD setting typically includes a distance from eye, called *eye relief* d_{relief} of $25 - 30\ mm$, and an additional spacing d_{object} of $35 - 50\ mm$ between the magnifier lens and display, which is determined by the focal length of the magnifier lens. This leaves two obvious options of where to place the LEDs – between the lens and the eye or between the magnifier lens and the display.

From the stand point of integrating our solution, the easiest place to put the LEDs would be directly in front of the user's eyes, as shown in Figure 3. This arrangement also has the advantage that it minimizes light loss due to scattering off the other optical elements of the HMD. However, LEDs positioned on-axis in front of an eye would occlude parts of the
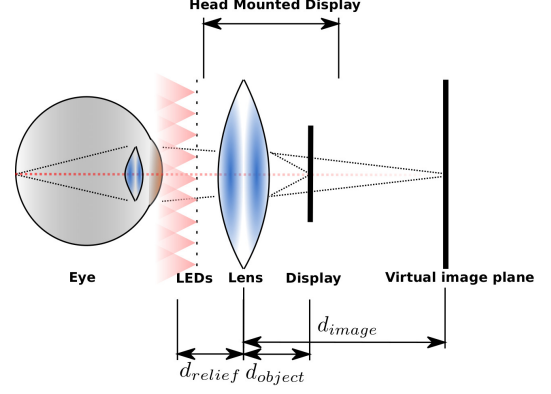

Figure 3. **A sketch showing our proposed configuration within a conventional HMD.**

image. Placing LEDs in an off-axis arrangement avoids such occlusions.

We rely on the experimental study from Nguyen et al. [28] to determine the best positions. A maximal amount of corneal and pupil reflection can be achieved, when the sensing LEDs are close to the illuminating LEDs. We thus place LEDs side-by-side in a ring around the magnifier lens. Front view in Figure 5 shows a layout for positioning of LEDs in our prototype. We arrange the LEDs as groups of two sensing LEDs with an illuminating LED in between.

Estimation Using Captured Data

Our capture hardware hosts a fixed number M of sensing LEDs. The capture hardware transmits a data capture vector to the host (PC, smartphones, and etc.). This vector contains measured data as

$$\mathbf{s}(t) = [s_0(t), s_1(t), \ldots, s_M(t)], \qquad (1)$$

where each $s_m(t)$ represents the output of the m-th LED connected to the capture hardware. We construct a calibration matrix that represents the relation between a captured vector and the gaze locations as follows. First, pre-defined locations are highlighted on the virtual image plane in random order, and users are asked to dwell on each of the highlighted locations for a certain duration Δt_{fix}. During this time, we sample $\mathbf{s}$ multiple times at fixed intervals Δt_{v}, and store the mean of the measurements as $\bar{c}_p = \sum_t \mathbf{s}(t)$ for each predefined location p. To ensure a meaningful calibration, we check that the variance of all the measurements for a location p is below a certain threshold, otherwise we discard the measurements for that location. Figure 4 summarizes the procedure for creation of a calibration matrix.

The pre-defined calibration points can be at any location. Typically, we use a grid of 2×2 to 5×5 evenly spaced locations. Starting from this small number of calibration points, we need to derive the gaze location given a set of new measurements $\mathbf{s}(t)$. At a later stage, calibration data is enlarged online through a gaming application, in which a user's task is to gaze and destroy opponents. Using collected data, we have evaluated two different regression methods: support vector regression (commonly used in prior research), and gaussian process regression (our choice).
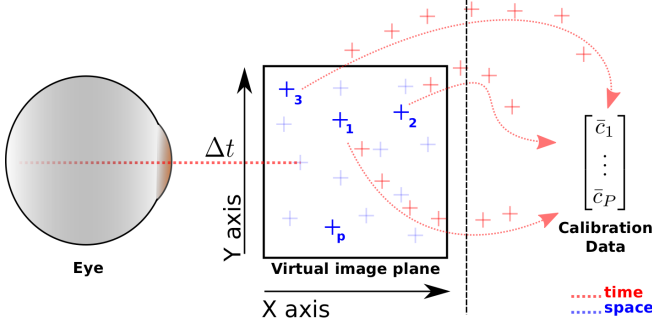
Figure 4. A sketch showing construction of a calibration matrix in a simplified way. A user gazes at a certain point at a virtual image plane for a time period of Δt_{fix}. During that time data is captured, and the arithmetic mean value of the capture, $\bar{c}_p$ is stored at a corresponding row of the calibration matrix.

Support Vector Regression: SVR has a generalized form as

$$\begin{bmatrix} e_x \\ e_y \end{bmatrix} = k^T \begin{bmatrix} u_x \\ u_y \end{bmatrix}^T, \qquad (2)$$

where e_x and e_y represents estimated gaze position along x and y, k^T represents a vector that contains the similarity measures between the captured $\mathbf{s}(t)$, and the calibration vectors $\bar{c}_p$. Finally, u_x and u_y represent vectors that correspond to the x and y position of each $\bar{c}_p$.

A regression technique depends on a distance measure, that indicates how similar the captured data is to the stored calibration data. Such a comparison using any method will provide a similarity vector as in

$$k = \begin{bmatrix} \kappa(\mathbf{s}(t), \bar{c}_1) \\ \vdots \\ \kappa(\mathbf{s}(t), \bar{c}_P) \end{bmatrix}, \qquad (3)$$

where $\kappa(a, b)$ donates the used distance function to determine similarity in between vector a and b, k represents the distance vector. Our choice of distance measure for SVR calculations is the commonly used radial basis function (RBF) [23, 29].

Gaussian Process Regression: Through our experiments, we found GPR to be a robust and accurate regression method. GPR takes the following general form:

$$\begin{bmatrix} e_x \\ e_y \end{bmatrix} = k^T C^{-1} \begin{bmatrix} u_x \\ u_y \end{bmatrix}^T, \qquad (4)$$

with variables as described above. The covariance matrix C is calculated as

$$C = \begin{bmatrix} \kappa(\bar{c}_0, \bar{c}_0) & \dots & \kappa(\bar{c}_0, \bar{c}_p) \\ \vdots & & \\ \kappa(\bar{c}_p, \bar{c}_0) & \dots & \kappa(\bar{c}_p, \bar{c}_p) \end{bmatrix}. \qquad (5)$$

Distance Measures: Comparing a vector with another vector can be accomplished in multiple ways. Although, we have evaluated multiple different distance measures (*Cosine, Minkowski, Manhattan, Canberra*) [4, 15, 30], we found the Minkowski distance measure to be the most effective to be

used with the GPR algorithm:

$$\kappa(a, b) = \left(\sum_{i=1}^{n} w_i \, |a_i - b_i|^m \right)^{1/m}, \qquad (6)$$

where a and b are two vectors to be compared, w_i is the weighting factor for a certain channel in the captured data, m is the degree of the norm, and i is the index of the element. We find $m = 2$ and setting $w_i = 1$ to yield good results; more details will be presented in Section .

For the SVR algorithm, we employed the RBF kernel:

$$\kappa(a, b) = e^{-\frac{||a-b||}{2\sigma^2}}, \qquad (7)$$

in which σ represents a free parameter. We use a grid-search to find the optimal σ.

IMPLEMENTATION

Two different off-the-shelf HMDs are transformed into a gaze sensing HMD using our methodology. In this section, we introduce both design choices with practical aspects. Our prototypes are shown in Figure 5.

Our first prototype consists of 3 LEDs per eye functioning as light sources, 6 LEDs per eye functioning as light sensors, a smartphone, an Arduino Nano microcontroller (uC) per eye, a controller, and a VR headset as a housing. Our gaze tracking algorithm runs on a smartphone in synchronism with two uCs driving LEDs.

Our second prototype consists of 6 LEDs per eye functioning as both light sensors and light sources, an Arduino Nano microcontroller (uC) per eye, a HDMI supported 2K resolution display, a controller, and a VR headset as housing. This time, our gaze tracking algorithm runs on a desktop computer.

Optics: Our both prototypes use a pair of magnifier lenses in front of a display, as shown in Figure 3. Considering the magnification of the lens, and the distances between LCD, lenses, and eyes, and the display's pixel density, each pixel corresponds to a visual angle of 0.12 degrees in our both prototypes.

Sensing Electronics: LEDs are hooked to the two uCs for both prototypes, so that they can be programmed accordingly. In the case of our first prototype, illuminator LEDs are attached to digital input/output (IO) pins, and the sensing LEDs' anodes were attached to digital IOs, and their cathodes were attached to Analog-to-Digital (ADC) pins of the uC. In the case of our second prototype, LEDs are only attached in the same way as the sensing LEDs in our first prototype. Each time a LED is requested to sense, it follows the pattern of ①-②-③ from Section and Figure 2. LEDs have a soft-coded mechanism that adjust Δt_{exp} per LED basis, so that saturation caused by varying light conditions can be avoided for each LED.

Identical two LEDs typically have a limited overlap in their emission-sensing spectra. Thus, leading to low resolution capture with less sampling frequency when identical LEDs are used. In our first prototype, we choose to dedicate two
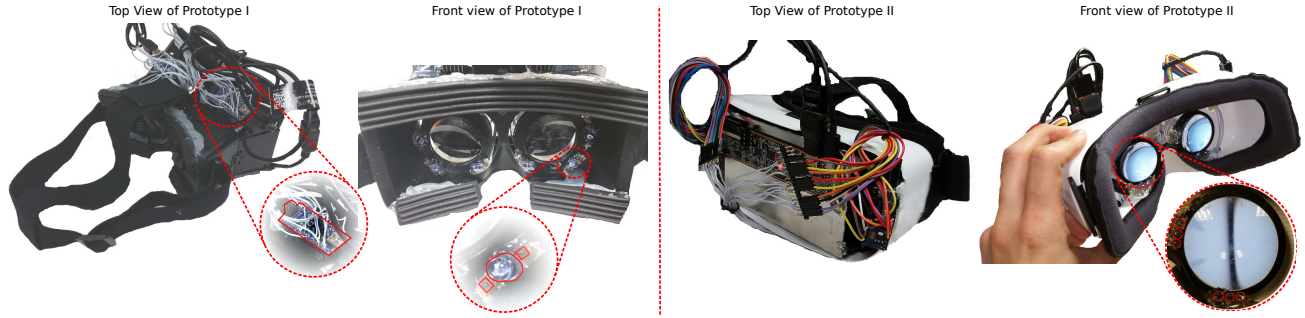
Figure 5. Left: A pair of photographs showing top and front view of our first prototype. Zoomed region in top view shows a microcontroller. Zoomed region in front view shows three LEDs, center one that is highlighted in red is used in illumination source mode, and neighbouring two that are highlighted in green are in receiver mode in our driving scenario. Note that all LEDs in our prototype can switch to both modes upon request. Right: A pair of photographs showing top and front view of our second prototype. Zoomed region in front view highlights six LEDs that are used for both sensing and illumination.

different LEDs to specific tasks to maintain good resolution. However, with our second prototype, we also evaluated a configuration that uses the identical LEDs for illumination and sensing.

In our first prototype, an illuminator LED is turned on shortly before taking a measurement from its pair of neighbouring sensing LEDs. In our second prototype, all the remaining LEDs are illuminating the scene at the time of a capture from a LED. In our prototypes, used uC only allows a time multiplexed capture routine -one capture from a single LED at a time-. However, a simultaneous capture from different LEDs can be achieved with a different uC that has discrete ADCs for each analog pin. Thus, the effective sampling rate would increase significantly and latency would reduce even further.

User-interface application: The uC works hand-in-hand with the user-interface application using a USB connection. Our application deals with a number of predefined tasks: (1) collecting measurements from each LED by requesting them from the two uCs, (2) updating the user interface, (3) producing the actual gaze estimation using GPR, and (4) keeping logs related to captured data (event time stamps, and so forth).

EVALUATION

Success of a regression algorithm for the task of gaze estimation is heavily affected by a number of factors, such as noise levels in the captured signal, positioning of LEDs in front of an eye, the used distance measures, and the number of stored points in a calibration matrix. We evaluate our method with respect to each factor.

Our first prototype uses 18 illuminating LED and 12 sensing LED in total. Our first prototype is supplied by a voltage source with galvanic isolation (smartphone's battery), thus leading to more reliable analogue to digital conversions. On the other hand, our second prototype uses 12 LEDs in total for both illumination and capture. Our second prototype interfaces with a personal computer that is typically known to suffer more from noise caused by electromechanical parts, power-supply and supply-line. Noises can propagate through data and power lines. Thus, it requires more levels of filtering in circuitry and in digital means, which are causing la-

tency both in digital estimation and analogue signal. We try to tackle noise problems at most with an additional layer of first order Infinite Impulse Response (IIR) digital low-pass filters, and a physical USB isolator. At extreme, a custom dedicated circuitry for analogue to digital conversions can outperform other options.

We report our findings on accuracy, and number of calibration data using our first prototype. We extend our subjective tests using our second prototype. Thus, we also report on practicality, and inter-intra personal differences using the second prototype. We also share our findings on sampling rates of both prototypes. Our evaluation method depends on the experimental procedure summarized in Table 1. We have conducted a series of subjective tests using this procedure, and recorded each session for analysis.

Instructions, levels and hints for a session

L1	User wears the headset.
L2	User starts the application by pressing "X" button in the controller.
L3	User is verbally asked about image quality. If not satisfactory, session is terminated. Otherwise user continues with the session.
L4	Application shows a visual containing a rabbit character together with a text that states press "X" button and follow the rabbit.
L5	Rabbit on the screen stops after a while, a text in the visual appears, which commands a user to gaze at the rabbit constantly until it disappears.
L6	Another visual appears with an instruction, which commands a user to gaze at a red dot constantly, press "X" button on controller, and gaze constantly until red dot turns to green. Once it is green, another red dot appears at a different location, and user repeat his previous actions until there is no more red dot in the screen.
L7	User follows the instructions from L6 for each event objects.
L8	Another visual appears with a new instruction, which states that there will be a ghost character among with multiple characters in the next screen. Task is to destroy ghost, the user has to gaze at it for a certain time period. If user is gazing at the ghost, the ghost turns blue or other objects are changing colors; the user is asked to keep on gazing at the ghost and press "X" button on the controller.
L9	As the user is gazing at each object, the calibration data is augmented with the new incoming measurements from each failure case.

Table 1. Our experimental procedure, including the guidance provided to our participants.
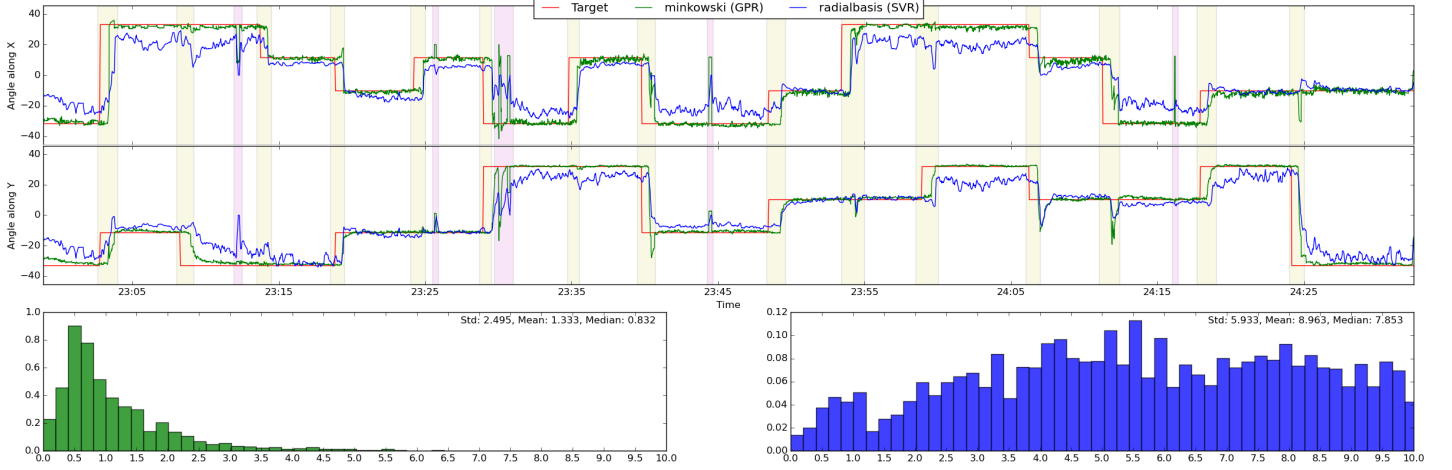
Figure 6. Top row shows angular positions of a target on a screen along X axis in red color among with the outputs of GPR with Minkowski and SVR with RBF in green and blue color respectively. Middle row shows angular positions of a target on a screen along Y axis in red color among with the outputs of GPR with Minkowski and SVR with RBF in green and blue color respectively. Both in top and middle rows, Regions highlighted with Magenta color shows blink events, and regions highlighted with yellow color shows Saccadic Reaction Time (SRT) with a gaze transition (saccade). Bottom row shows normalized histograms of angular errors for GPR with Minkowski and SVR with RBF in green and blue color respectively. There are 82 calibration points in this sample dataset (16 from calibration grid, 66 from VR gaming). Captured data used in making is from our first prototype.

Accuracy: The accuracy of a gaze tracker system refers to the difference between the true and the measured gaze direction. Given a set of sample input data, we estimate the gaze locations using both GPR with Minkowski distances and SVR with RBF as the regressor. We show the angular locations of the target gaze locations along with the estimates from both methods in Figure 6. Our error metric here is visual angular errors, which is calculated using Euclidean distances in between output of estimations and gaze targets. Figure 6 contains histograms of visual angular errors recorded over a fixed time duration for both techniques using a set of captured data from our first prototype. We have conducted a comparison of multiple different distance measures using GPR and SVR, a detailed analysis of these can be found in the supplementary material.

We point out that regions highlighted with magenta color in Figure 6 show blinks, which lead the estimated gaze direction to have sharp jumps. Regions highlighted with yellow color in Figure 6 show regions, where a gaze target changes position in space. Humans are known to have a response time before an event of gaze change (Saccade) [26], which is known as Saccadic Reaction Time (SRT). SRT may vary from individual to individual, affected by biological factors such as age and health conditions. Angular errors have been calculated by excluding highlighted regions. We have performed a user study among 5 subjects with varying age, and eye prescription. Using GPR with Minkowski distances, we report our findings for each subject in terms of accuracy in Table 2.

Commercial gaze trackers (SMI, FOVE VR, etc.) for HMDs claim to have an accuracy ranging from $0.2^o - 1.5^o$ in a controlled laboratory environment. We show that our mean angular accuracy can be as low as 1.10^o with a much simpler hardware setup in casual use case. We believe that our approach is a promising alternative to existing methods.

	Mean	Median	Standard Deviation
Subject 1	1.34	0.83	2.49
Subject 2	2.10	1.23	2.80
Subject 3	1.90	1.84	1.43
Subject 4	1.40	1.04	1.10
Subject 5	1.10	0.66	2.14
Total Mean	1.57	1.12	2.00

Table 2. Mean, median, and Standard Deviation values of each subjects using data collected from each subject. Angular errors are reported only for the case of using GPR with Minkowski distances.

Number of calibration points and LEDs: The number of calibration points and LEDs are other important variables in our design space. Thus, we investigate the resulting angular error when varying their number. Given a small number of calibration points, we have started our evaluation using only four channels with the highest variation in digital signals and have iteratively increased the number of channels. We have repeated this evaluation utilizing an increasing number of calibration points. Our findings suggest that smaller numbers of LEDs can produce similar angular accuracy levels. Adding more LEDs seem to only marginally enhance accuracy. However, we would like to highlight that an LED's contribution to the gaze estimate depends on its physical location and direction relative to a particular subject. I.e., depending on a subject's physiognomy (smaller vs. larger eyes, distance between eyes, and so forth), a particular sensing LED may contribute more or less. Thus, a larger number of LEDs compensate for such differences and add robustness.

Practicality and inter-intra personal differences: We extended our experiments using our second prototype. Total number of independent different subjects are 10 males and 4 females. At each experiment, a subject has to complete four independent sessions, in which a user takes brakes and wears the HMD at the beginning of each session. A subject typi-

cally complete four sessions within $45-60$ minutes in total. First two sessions follow the usual routine from Table 1 with a task of choosing a predefined target object from multiple choices (varying randomly from three to eight choices). At each task, a target object appears at a new random location chosen from uniformly distributed sample locations. Total number of target selection task in a single session is 50. Definition of success at a task is being able to choose the target object by gazing at the target object at least for 3 seconds. On the other hand, definition of a failure case can be expressed as follows: If our system fails to detect the case of gazing correctly, user continuously gaze at the target, and push a button to give a feedback to the system. System uses this as an online training data, and takes advantage of the new data for the upcoming tasks within a session. Third and fourth sessions follows the same routine as in first two sessions. However, those sessions does not go through the calibration phase. A third session is initiated with the calibration data captured from the same user during a second session. A fourth session is initiated with the captured data from an another subject during an another session in the past. Figure 7 summarizes results of our early extended experiments.
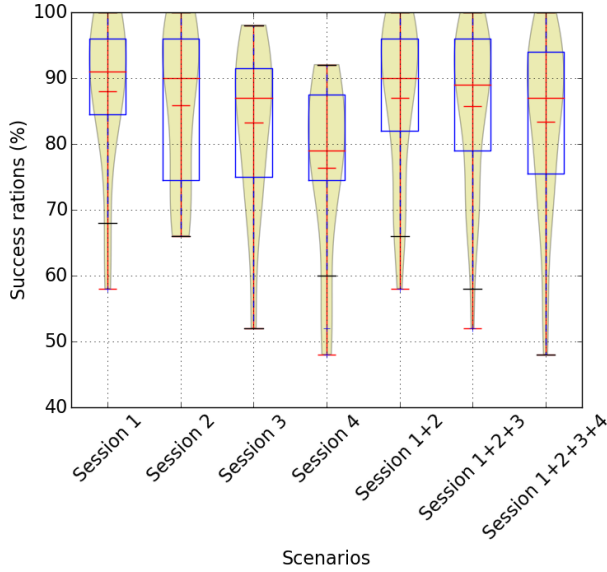

Figure 7. Violin-box diagrams showing distribution of success rations from different scenarios. Small and large red vertical lines indicates mean, and median values respectively. Dashed bars indicates 1.5 times interquartile range (IQR), and the blue boxes indicates first and third quartiles from a distribution.

According to our early investigation in Table 7, task of choosing by gazing at a target is likely to be more successful with a calibration phase at the beginning of a session. We also observed that system can be initiated with a calibration data from a previous session with lower success rations. Thus, requiring further online training during usage. At the extreme, our findings suggest that it is possible to initiate the system with a calibration data from a different person with least amount of success rations and maximum amount of online training during usage.

Sampling Rates and latency: Given our current hardware and uC, we are able to sample up to $250\,Hz$ with our first prototype coupled to a mini computer (Raspberry Pi model B). On the other hand, our second prototype coupled to a personal computer samples at $100\,Hz$. A better uC with a dedicated ADC for each analogue pin can sample all channels simultaneously, thus such a uC can increase sampling rates up to kHz range. We observe that the smartphone's USB does not provide a reliable serial communication at $250\,Hz$. Thus, we intentionally use only a $100\,Hz$ sampling rates with our first prototype. Our GPR implementation can easily run at above $100\,Hz$ even on the smartphone (our datasets all contain less than 100 calibration points).

DISCUSSION

We believe that gaze tracking is a critical feature that will contribute to the success of HMDs in the consumer market. We propose a new sensor technology that enables gaze tracking at low cost and in a small form factor, ideal for the HMD use case.

Feedback from subjects: In informal feedback, subjects found our system to be accurate for the given tasks shown during a subjective test.

One common problem reported by the subjects was sudden jumps in gaze estimation after and during a blink. Similarly, slightly closing an eye lid was found to cause a shift in the gaze estimation. In the future work we plan to address these problems by detecting blinks from the LED inputs as well as predicting the eye lid status. Some subjects experienced lower spatial accuracy in certain regions within the screen. We believe this to be due to differences in eye sizes, relative eye locations, and eye distances. Simply adding more channels to each eye should help to address this problem. Finally, the initial calibration phase with a four-by-four grid was found to be cumbersome, and we will continue to investigate on how to further decrease the number of calibration points. Self-calibration through saliency maps [40] may be an option.

Future Work: Kalman filtering is a common method to enhance accuracy (but may add latency), as are visual anchors [13]. Expanding the subject pool can also help us in clustering subjects for estimating a good initial calibration data for each person. We will evaluate these methods in the near future. An analogy from coded aperture optics is an inspiring direction for our future work: merging our technique with backlights of Spatial Light Modulators (SLMs) could potentially provide an interesting all-in-one solution for sensing and illumination, similar to bi-directional displays [9, 45].

Our subject pool contains subjects with and without eye prescriptions. During a subjective test, subjects with eye prescriptions were asked to take off their glasses and wear contact lenses instead. If our technique was to be used in a non-HMD scenario, corrective glasses would need to be supported.

In a near future, VR users may also benefit from wearable facial recognition techniques using photo-reflective sensors [24] or from vital sign monitoring, such as heart-rate and

blood oxygenation, which have been estimated remotely using photodiodes [5] in the past. A similar methodology using LEDs could be explored through simple changes in our circuitry. Thus, those methodologies may allow us to improve in accuracy by considering effects of facial changes due to mood, respiration or heart-rate. Saccadic reaction times (SRTs) are well explored by the medical community as a basis for health condition prediction techniques [26]. Our current prototype can also be trained to detect SRTs (yellow regions in Figure 6) in the future. Rogers et al. [32] recently demonstrated that blinks and head movements captured using infrared, accelerometer, and gyroscope sensors in a HMD can be used for the task of user identification. All these sensors are readily available in our prototypes. User identification dimension can be added to our demonstration by merging [32]'s findings with our proposal.

CONCLUSION

Gaze sensing HMDs promise to provide a dataset for a set of useful tasks: foveated rendering, predicting a user's mood and health, predicting visual saliency at a personal level, and providing a VR experience that is unique, that is a natural fit for you.

We describe a novel gaze tracker methodology using only LEDs. Our proposal's main contribution is exploiting the sensing capability of LEDs with a GPR implementation using Minkowski distances inside a HMD system. We provide a detailed description of our algorithms and our hardware, which we believe is the simplest gaze-detecting design to date. We have shown two unique prototypes that validates our proposal in the context of a traditional off-the-shelf HMD. As gaze-sensing HMDs are poised to enter the market, we hope to inspire more research on simple, low-power non-imaging approaches such as ours.

Acknowledgements

REFERENCES

1. An, C., Li, T., Tian, Z., Campbell, A. T., and Zhou, X. Visible light knows who you are. In *International Workshop on Visible Light Communications Systems* (2015), 39–44.

2. Ben-Ezra, M., Wang, J., Wilburn, B., Li, X., and Ma, L. An led-only brdf measurement device. In *IEEE Conference on Computer Vision and Pattern Recognition* (2008), 1–8.

3. Boucouvalas, A. IEC 825-1 eye safety classification of some consumer electronic products. *IEE Colloquium on Optical Free Space Communication Links* (1996), 13:1–13:6.

4. Bray, J. R., and Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs 27*, 4 (1957), 325–349.

5. Cennini, G., Arguel, J., Akşit, K., and van Leest, A. Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optics Express 18*, 5 (2010), 4867–4875.

6. Corbellini, G., Aksit, K., Schmid, S., Mangold, S., and Gross, T. Connecting networks of toys and smartphones with visible light communication. *IEEE Communications Magazine 52*, 7 (2014), 72–78.

7. Dietz, P., Yerazunis, W., and Leigh, D. Very low-cost sensing and communication using bidirectional LEDs. In *UbiComp* (2003), 175–191.

8. Grover, D., Delbruck, T., and King, M. An eye tracking system using multiple near-infrared channels with special application to efficient eyebased communication. In *European Conference on Eye Movement* (2005), 36–39.

9. Hirsch, M., Lanman, D., Holtzman, H., and Raskar, R. BiDi screen: a thin, depth-sensing LCD for 3D interaction using light fields. *ACM Transactions on Graphics (TOG) 28*, 5 (2009).

10. Holmqvist, K., Nyström, M., and Mulvey, F. Eye tracker data quality: what it is and how to measure it. In *Symp. on Eye Tracking Research and Applications* (2012), 45–52.

11. Jacob, R., and Karn, K. S. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind 2*, 3 (2003), 4.

12. Johansen, S. A., San Agustin, J., Skovsgaard, H., Hansen, J. P., and Tall, M. Low-cost vs. high-end eye tracking for usability testing. In *ACM CHI* (2011), 1177–1182.

13. Kumar, M., Klingner, J., Puranik, R., Winograd, T., and Paepcke, A. Improving the accuracy of gaze input for interaction. In *Symp. on Eye Tracking Research Appl.* (2008), 65–68.

14. Kunka, B., and Kostek, B. Non-intrusive infrared-free eye tracking method. In *Signal Processing Algorithms, Architectures, Arrangements, and Applications* (2009), 105–109.

15. Lance, G., and Williams, W. Computer programs for hierarchical polythetic classification ("similarity analyses"). *The Computer Journal 9*, 1 (1966), 60–64.

16. Lange, V., Lima, F., and Kühlke, D. Multicolour led in luminescence sensing application. *Sensors and Actuators A: Physical 169*, 1 (2011), 43–48.

17. Li, S., and Pandharipande, A. Color sensing and illumination with led lamps. In *IEEE International Conference on Consumer Electronics* (2014), 1–2.

18. Li, S., and Pandharipande, A. LED-based color sensing and control. *IEEE Sensors Journal 15*, 11 (2015), 6116–6124.

19. Li, T., An, C., Tian, Z., Campbell, A. T., and Zhou, X. Human sensing using visible light communication. In *Int. Conf. on Mobile Computing and Networking* (2015), 331–344.

20. Majaranta, P., and Bulling, A. Eye tracking and eye-based human–computer interaction. In *Advances in Physiological Computing*. 2014, 39–65.

21. Mantiuk, R., Kowalik, M., Nowosielski, A., and Bazyluk, B. *Do-it-yourself eye tracker: Low-cost pupil-based eye tracker for computer graphics applications*. Springer, 2012.

22. Marg, E. Development of electro-oculography: Standing potential of the eye in registration of eye movement. *AMA Archives of Ophthalmology 45*, 2 (1951), 169–185.

23. Martinez, F., Carbone, A., and Pissaloux, E. Gaze estimation using local features and non-linear regression. In *IEEE Int. Conf. on Image Processing* (2012), 1961–1964.

24. Masai, K., Sugiura, Y., Suzuki, K., Shimamura, S., Kunze, K., Ogata, M., Inami, M., and Sugimoto, M. Affectivewear: towards recognizing affect in real life. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ACM (2015), 357–360.

25. Mayberry, A., Hu, P., Marlin, B., Salthouse, C., and Ganesan, D. ishadow: Design of a wearable, real-time mobile gaze tracker. In *MobiSys*, ACM (2014), 82–94.

26. Mazumdar, D., Pel, J. M., Panday, M., Asokan, R., Vijaya, L., Shantha, B., George, R., and Van Der Steen, J. Comparison of saccadic reaction time between normal and glaucoma using an eye movement perimeter. *Indian journal of ophthalmology 62*, 1 (2014), 55.

27. Naugle, E., and Hoskinson, R. Two gaze-detection methods for power reduction in near-to eye displays for wearable computing. In *Int. Conf. on Wireless and Mobile Comp., Net. and Comm.* (2013), 675–680.

28. Nguyen, K., Wagner, C., Koons, D., and Flickner, M. Differences in the infrared bright pupil response of human eyes. In *Symp. on Eye Tracking Research & Appl.* (2002), 133–138.

29. Noris, B., Benmachiche, K., and Billard, A. Calibration-free eye gaze direction detection with gaussian processes. In *VISAPP* (2008), 611–616.

30. Rasmussen, C. E. *Gaussian processes for machine learning*. MIT Press, 2006.

31. Robinson, D., et al. A method of measuring eye movemnnt using a scieral search coil in a magnetic field. *Bio-medical Electronics, IEEE Transactions on 10*, 4 (1963), 137–145.

32. Rogers, C. E., Witt, A. W., Solomon, A. D., and Venkatasubramanian, K. K. An approach for user identification for head-mounted displays. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers* (2015), 143–146.

33. Round, H. J. A note on carborundum. *Electrical World 49*, 6 (1907), 309.

34. San Agustin, J., Hansen, J. P., and Tall, M. Gaze-based interaction with public displays using off-the-shelf components. In *ACM International Conference Adjunct Papers on Ubiquitous Computing* (2010), 377–378.

35. San Agustin, J., Skovsgaard, H., Hansen, J. P., and Hansen, D. W. Low-cost gaze interaction: ready to deliver the promises. In *ACM CHI* (2009), 4453–4458.

36. San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D. W., and Hansen, J. P. Evaluation of a low-cost open-source gaze tracker. In *Symposium on Eye-Tracking Research & Applications* (2010), 77–80.

37. Schmid, S., Ziegler, J., Gross, T. R., Hitz, M., Psarra, A., Corbellini, G., and Mangold, S. (in)visible light communication: combining illumination and communication. In *ACM SIGGRAPH 2014 Emerging Technologies* (2014), 13.

38. Sewell, W., and Komogortsev, O. Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. In *ACM CHI* (2010), 3739–3744.

39. Stellmach, S., and Dachselt, R. Look & touch: gaze-supported target acquisition. In *ACM CHI* (2012), 2981–2990.

40. Sugano, Y., and Bulling, A. Self-calibrating head-mounted eye trackers using egocentric visual saliency. In *ACM Symposium on User Interface Software & Technology* (2015), 363–372.

41. Topal, C., Gunal, S., Koçdeviren, O., Dogan, A., and Gerek, O. N. A low-computational approach on gaze estimation with eye touch system. *IEEE Transactions on Cybernetics 44*, 2 (2014), 228–239.

42. Turner, J., Alexander, J., Bulling, A., Schmidt, D., and Gellersen, H. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Human-Computer Interaction – INTERACT 2013*. 2013, 170–186.

43. Turner, J., Bulling, A., Alexander, J., and Gellersen, H. Cross-device gaze-supported point-to-point content transfer. In *Symp. on Eye Tracking Research and Appl.* (2014), 19–26.

44. van den Berg, T. J., and Spekreijse, H. Near infrared light absorption in the human eye media. *Vision Research 37*, 2 (1997), 249–253.

45. Vogel, U., Kreye, D., Richter, B., Bunk, G., Reckziegel, S., Herold, R., Scholles, M., Törker, M., Grillberger, C., Amelung, J., Graupner, S.-T., Pannasch, S., Heubner, M., and Velichkovsky, B. M. 8.2: Bi-directional OLED microdisplay for interactive HMD. *SID Symposium Digest of Technical Papers 39*, 1 (2008), 81–84.