Ultrasound for gaze estimation

Andre Golard[1] and Sachin S. Talathi[1]

Facebook Reality Labs, Redmond WA 98052, USA
{agolard,stalathi}@fb.com

Abstract. Most eye tracking methods are light-based. As such they can suffer from ambient light changes when used outdoors. It has been suggested that ultrasound could provide a low power, fast, light-insensitive alternative to camera based sensors for eye tracking. We designed a bench top experimental setup to investigate the utility of ultrasound for eye tracking, and collected time of flight and amplitude data for a range of gaze angles of a model eye. We used this data as input for a machine learning model and demonstrate that we can effectively estimate gaze (gaze RMSE error of 1.021 ± 0.189 degrees with an adjusted R^2 score of 89.92 ± 4.9).

Keywords: Ultrasound · Eye tracking · Machine learning · CMUT.

1 Introduction

Most current eye tracking methodologies use video to capture the position of the iris and/or reflected lights sources – glints [7]. As such these methods can be affected by ambient light [2], which will be the case with for eye tracking applications in wearables such as augmented-reality (AR). Other light-based methods such as scanning lasers, dual Purkinje and directional light sensors can likewise be affected. Speed can also be limited to 100 Hz, especially in wearables, where operating a camera at high speed would imply high power consumption. At these speeds the camera-based sensors can capture fixations but not other eye motions such as saccades, which have been implicated as markers of neurological disorders [12]. Current devices capable of measuring saccades are designed for laboratory use, and tend to lack portability. The possibility of using ultrasound for eye tracking has been raised [10]. However there was no modeling and no experimentation.

A recent paper explored the possibility of using non-contact ultrasound sensors to track fast eye movements in the field [6]. The work focused on the development of finite element simulation model to investigate the use for ultrasound time of flight data to track fast eye motions . The simulation model is based on a setup made of four transducers positioned perpendicular to the cornea. Distances are measured with each transducer receiving the reflection of its own signal. For this to be possible the device needs to be precisely positioned relative to the eye. We are interested in applications for eye tracking in AR and virtual reality (VR), where user-specific placement of the sensors is not possible. It is

also to be noted that the modeling in [6] was done in the absence of occlusions. Occlusions are known to be problematic for eye tracking systems in general. [4]. Furthermore, the authors [6] chose to model standard 40kHz transducers. While these would be advantageous in terms of minimizing attenuation in air, such a system may be subject to interference from range-finding applications (typically in the 40-100kHz range). Common range finding systems lack the resolution and short distance sensing capabilities required for eye tracking.

Another concern for our application of interest is size. Capacitative Micromachined Ultrasonic Transducers (CMUTs) operating at 500kHz-2MHz [8] provide a range, resolution and size that is suitable for use in VR and AR devices. This type of transducer has found numerous medical applications in both imaging and therapy. These applications are for contact ultrasound. Here, we use the devices as airborne transmitters and receivers. In this mode, the difference in impedance between air and tissue means over 99 percent of the ultrasound signal will be reflected by the eye surface.

While our long-term goal is integration in a VR or AR form factor device and size was considered for the selection of transducers, related concerns (test bench size, power consumption) did not drive our experimental design prototype. We built a series of table top test benches to verify our ability to accurately measure distances in the appropriate range, characterize the transducers, and generate data to be used in a machine learning model to estimate gaze. As such we focus on empirically testing the hypothesis that ultrasound sensors can be used for gaze estimation in the presence of occlusions. We demonstrate that ultrasound time of flight and amplitude signals can be leveraged to train a machine learning model to track gaze in such conditions. Results show that the trained model produces a regression R^2 score of 89.92 % and a gaze RMSE error of 1.021 $\pm$ 0.189 degrees.

We note that while there exists a vast literature on eye tracking and ultrasound [11], none has focused on using ultrasound for eye tracking. To the best of our knowledge, this paper presents the first experimental study to empirically demonstrate the feasibility for gaze estimation using ultrasound sensors.

2 Materials and Methods

In this section, we describe the bench top experimental setup for data collection, the signal processing steps to extract the ultrasound time of flight and amplitude, and the machine learning framework adopted to train a gaze estimation model.

2.1 Bench-top setup

We designed a series of three test benches to evaluate distance measurements, signal attenuation, transducer directionality, and our ability to estimate gaze.

In terms of electronics and data acquisition, all test benches are based on a CMUT evaluation kit from Fraunhofer IPMS (Dresden, Germany). This test

kit is comprised of CMUT transducers (1.74 MHz), an amplifier, bias-tee, and associated software. These transducers fit our size and power requirements.

We first verified our ability to measure distances, as well as the signal decay due to attenuation in air given that ultrasound signal attenuation is significant at MHz frequencies [1]. We used a setup consisting of a pair of transducers aimed at a flat target attached to a linear translation stage (Test bench 1, Figure 1A).

Next we tested the emission properties of the transducers. Our CMUTs are comprised of an array of cells connected to a single electrode and a single counter electrode. As such they act as a fixed phased array, which is expected to exhibit directionality. We tested this using a fixed transducer and one on a rotating stage (Test bench 2, Figure 1B). The Tx transducer was rotated in 1 degree increments and the amplitude of the Rx signal was recorded.

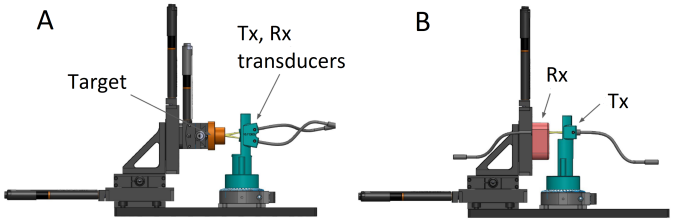

Fig. 1. CAD schema for attenuation and directionality test benches. Tx refers to transducer in transmit mode, Rx receive mode.

Our third test bench is designed for gaze estimations (Figure 2A). The transducer side is on the right. We used a pair of transducers (one in transmit mode and one receiver) mounted on rotating stages to allow us to mimic multiple locations around a ring (or glasses frame). We acquired data for all transmit and receive locations covering 360 degrees in 10 degree increments (Figure 3C).

On the target side (left part of Figure 2A), a standard sphere on sphere model eye (cornea radius 7.8mm, sclera radius 11.925mm, offset 5.6mm) was mounted on a goniometer (Thor Labs). Gaze angles were set in one degree increments between ± 5 degrees in both up/down (ϕ) and left/right (θ) directions.

Occlusions (known to affect eye trackers) were added for realism. They consisted of a partial scanned face printed in flexible material (A40 durometer Polyjet) with a cavity to accommodate the model eye (Figure 2B). This was mounted in front of and against the model eye and allowed the eye to move freely.

Our test signal consisted of a train of seven oscillations at 1.74 MHz, repeated at 2 kHz. The transmitter was moved to positions around a 180 degree arc opposite the receiver (-90, -80, ... , 80, 90), Figure 2C. Fifty runs were recorded for each transducer position. The series was repeated for all goniometer positions. The received signal was digitized at 80 MHz.

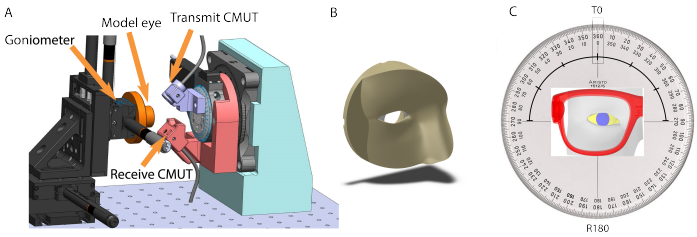

Fig. 2. A: CAD schema for the experimental bench-top setup and B: occlusions C. Transducer rotation. The receiver is fixed and the transmitter rotates around an arc. 30 degree steps are shown.

2.2 Data Analysis

Feature Engineering In Figure 3a, we show one raw trace, $x_i^r(t,\theta,\phi)$ ($i \in [0,49]$ and $r \in [-90,-80,\cdots 80,90]$), for the Ultrasound signal captured at the receiver, in response to a single test signal emitted by the transmit CMUT transducer. Figure 3b, shows the average of ten traces, defined as $\bar{x}_k^r(t,\theta,\phi) = 0.1\sum_j^{j+10} x_j(t,\theta,\phi)$ ($k \in [0,4]$). The ultrasound time of flight, $\tau_k^r(\theta,\phi)$, and amplitude, $a_k^r(\theta,\phi)$, signal is estimated for each $\bar{x}_k^r(t,\theta,\phi)$ as follows: the signal, $\bar{x}_k^r(t,\theta,\phi)$ is band-pass filtered in the frequency range, [1.6 MHz, 1.9 MHz] using a Butterworth filter of order 4 to generate the filtered version, $f(\bar{x}_k^r)(t,\theta,\phi)$. In Figure 3c, we show the trace for $f^2(\bar{x}_k^r)(t,\theta,\phi)$. The ultrasound time to peak $\tau_k^r(\theta,\phi)$ and the amplitude, $a_k^r(\theta,\phi)$ is obtained by considering a time window of 45 μs around the time instance of peak value for $f^2(\bar{x}_k^r(t,\theta,\phi))$ and finding the first instance of the peak value for $\bar{x}_k^r(t,\theta,\phi)$ within the considered time window. The detected peak value represents the amplitude signal $a_k^r(\theta,\phi)$ and the time to peak recorded as the ultrasound time of flight signal, $\tau_k^r(\theta,\phi)$. In summary, for each position $\mathbf{Y} = (\theta,\phi)$ of the model eye on the goniometer, we obtain a set k=5 feature vectors $\mathbf{X} \in R^{36} = \{a^r,\tau^r\}_{r=[-90,-80\cdots 80,90]}$. Our goal for ultrasound based eye tracking is to learn a regression model, $H : \mathbf{X} \rightarrow \mathbf{Y}$; that is, given the ultrasound sensor time of flight and amplitude data, estimate two-dimensional eye gaze coordinates.

Gradient Boosted Regression Trees From a machine learning perspective, the task of learning a gaze estimation model H is categorized as a supervised regression problem. Gradient Boosting Regression Trees (GBRT) are a powerful class of boosting algorithms for classification and regression tasks, which combine output from several weak learners into a powerful estimator. Specifically, GBRT considers additive models of the form: $F_m(x) = F_{m-1}(x) + h_m(x)$, where h_m are the basis functions modeled as small regression trees of fixed size. For each boosting iteration, a new boosting tree is added to the GBRT model, F. For our problem, we train two separate GBRT models to independently estimate the response: $\mathbf{Y} = (\theta,\phi)$ as function of the input features, $\mathbf{X} = (\tau^r, a^r)$. Assuming the GBRT model is comprised of M regression trees with T_m leaf nodes per

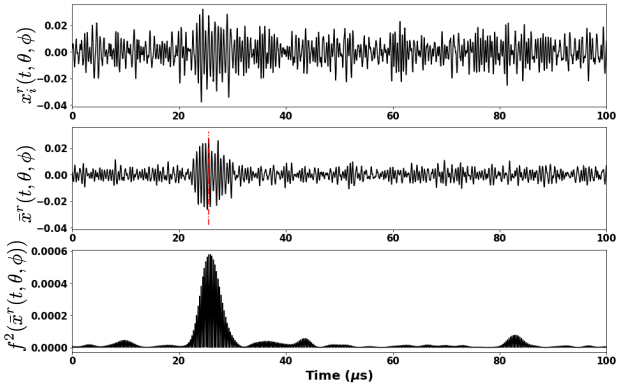

Fig. 3. Example of recorded raw time trace of ultrasound sensor signal. The top row shows an example of time trace recorded at the receive Ultrasound CMUT sensor in response to a single burst of test signal. The middle row shows averaged signal computed from the response to a set of 10 bursts of test signal. Finally the last row shows the squared filtered response signal out of a Butterworth filter. The red line indicates the time period of time-to-peak signal detection.

regression tree, the GBRT model for each of the gaze regressor is given as: $F^y(X, w^y) = w_0^y + \sum_{m=1}^{M} \sum_{j=1}^{T_m} w_{jm}^y I(X \in R_{jm}^y)$, where $y = \{\theta, \phi\}$ and R_{jm}^y represents the j^{th} disjoint partitioning of the input space for the m^{th} regression tree for the regressor variable, y. The GBRT model weights are estimated from data as follows: $w^* = \arg\min_w \frac{1}{N} \sum_i^N L(y_i, F(\mathbf{X}_i, w))$ where, L is the squared error loss function. For an exhaustive description of GBRT, see [5, 9].

3 Results

In this section we present findings from our experiments conducted using the three bench-top setups described in Section 2.1.

We begin by presenting our findings on the CMUT sensor characterization. Data collected using test bench setup 1, allowed us to investigate the decay characteristics of the ultrasound signal in air, see Figure 4A. As expected, the ultrasound signal decays exponentially as a function of distance. An extrapolated fit shows it decays to zero. The distance axis shows the distance between the pair of transducers and the target (Figure 1A). Actual travel distance is twice this measurement. The range is similar to the distances for transducers mounted on eye glasses frames, our use case scenario.

Data collected using test bench 2 (Figure 1B) allowed us investigate whether the CMUT transducers exhibit directionality. Our findings are reported in Figure 4B. The CMUT transducers indeed exhibit directionality with an emission cone of 10 degrees. This applies to the transducers in both transmit and receive mode.

Based on the above findings we conclude that the strength of ultrasound signal at the receiver CMUT transducer will depend on two factors: distance

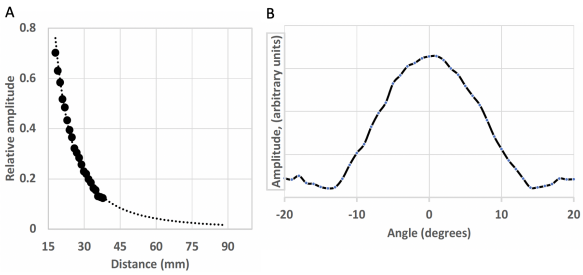

Fig. 4. CMUT sensor characterization

and incident angle. As such we believe that the amplitude of the ultrasound signal at the receiver contains relevant information to contribute to our ability to estimate gaze and as shown below, our findings indeed support this claim.

We next report findings from training a GBRT model on data collected using the third test bench setup (see Figure 2). For each model eye position on the goniometer, θ, ϕ, for a fixed receiver transducer position (180 degrees) and for a set of 19 transmit transducer positions, we fire the ultrasound test signal 50 times, at 2 kHz and record the raw receiver signal (see figure 3 top row). In order to increase the strength of ultrasound response at the receiver we average 10 traces of the raw response signals at a time, to effectively generate 5 averaged ultrasound response signals, in effect acquiring data at 200 Hz. The averaged response signal is passed through a Butterworth bandpass filter and we extract two ultrasound signal features: time of flight (τ) and the amplitude at peak (a), as explained in Section 2.2. In total for each model eye position, we generate a total of 45 samples for each model eye position on the goniometer over the duration of the study. For the set of 36 model eye positions, we produce a total of 1620 data samples.

We train a GBRT model on these data samples, performing a 5-fold cross-validation study. The model performance is reported using an adjusted R^2 score [3] and the gaze RMSE error in degrees. Hyper-parameter search on the GBRT model parameters that produced the best adjusted R^2 score for 5-fold CV are as follows: (a) Number of regression trees: 750 (b) Tree depth: 5 and (c) Learning rate: 0.085. We obtain gaze RMSE error of 1.021 ± 0.189 and mean adjusted R^2 score of 89.922 % with a standard deviation of 4.9965, suggesting that almost 90 % of the data fit the regression model. Residuals analysis confirmed that the estimates obtained using the GBRT model are un-biased (data not shown). This analysis offers an empirical evidence for our claim that ultrasound sensors can be used for gaze estimation in the presence of occlusions.

In Figure 5A and 5B, we show feature importance for the GBRT tree models trained to estimate the model eye gaze coordinates, θ (horizontal gaze) and ϕ (vertical gaze). We can see that the top two features for both horizontal and vertical gaze GBRT model are time of flight ultrasound signal followed by an amplitude feature. It has been our observation that while the time of flight

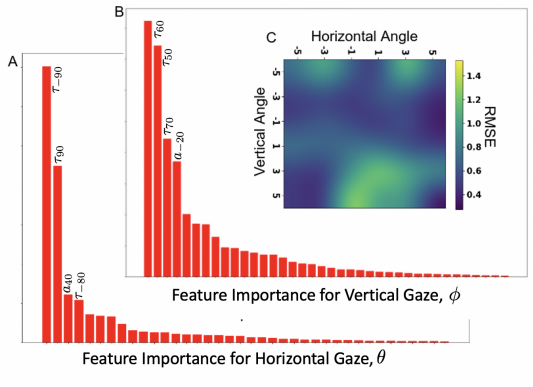

Fig. 5. Feature importance and mean-accuracy of GBRT models to estimate gaze

component of ultrasound signal contains dominant information signal to estimate gaze (95 % contribution to the regression score), the amplitude signal is also an important contributor for GBRT model to produce an adjusted-R^2 score close to 90 %. In order to test this observation, we trained GBRT model using just the ultrasound time-of-flight feature and another GBRT model using just the ultrasound amplitude feature. The findings are: GBRT model trained using time-of-flight features, produces an adjusted R^2 score of 85.38 ± 5.177, where as the GBRT model trained using only the amplitude feature produces an adjusted R^2 score of 78.64 ± 8.177. In Figure 5C, we show the mean-RMSE error (across all CV-folds) for the GBRT model. The error is biased towards the lower half of vertical gaze, primarily resulting from occlusions.

4 Discussion

This study is the first experimental demonstration of ultrasound eye tracking. We show that ultrasonic transducers can effectively produce signals useful to resolve eye gaze within the range tested, ± 5 degrees in both up/down (θ) and left/right (ϕ) directions. This range reflects the full deflection of our goniometer. We plan on expanding the range in future studies.

Our GBRTs show that both amplitude and time of flight contribute to our ability to estimate gaze. This is a new finding as previous modeling work dealt with time of flight alone. Two factors contribute to amplitude: attenuation and the incident angle of the incoming sound. One way to compensate for attenuation is to use the time-gain correction built in our amplifier, increasing gain over time to compensate for the signal attenuation with longer distances. When we did this (data not shown) our model performed slightly worse. This indicates that attenuation plays a role in our ability to estimate gaze, and would favor the use of high frequency transducers.

For this proof of concept we chose to average ten individual tests prior to filtering the signal and extracting peak and amplitude. This reduces the eye tracking acquisition speed from a maximum of 2kHz to just 200Hz, which may not be sufficient to track saccadic eye motion. While this study focused on primarily testing the hypothesis that ultrasound signals can be leveraged to estimate gaze, in future works we will explore avenues to investigate the use for ultrasound in tracking fast eye motion. Specifically, we plan on using a fast-moving model eye coupled with multiple receivers operating at 2kHz. The GBRT models will be adapted so we can test the potential of ultrasound for fast eye tracking to resolve saccades.

Our application is eye tracking for virtual and augmented reality. In addition to sampling speed, power consumption is an important factor to consider. The transducers are very low power, in the milliwatt range. Our current system utilizes a high speed A/D converter. This can be replaced with a low power peak detection circuit. On the compute side, GBRTs are considered low compute.

In summary, this study presents data driven proof-of-principle findings to support the claim that ultrasound sensors operating in the MHz range could provide an alternative to camera-based sensors for eye tracking.

References

1. Blackstock, D.T.: Fundamentals of Physical Acoustics. John Wiley Sons (2000)
2. Cheng, D., Vertegaal, R.: An eye for an eye: A performance evaluation comparison of the lc technologies and tobii eye trackers. p. 61. ETRA (2004)
3. Dodge, Y.: The Concise Encyclopedia of Statistics. Springer (2010)
4. Hansen, D.W., Ji, Q.: In the eye of the beholder: A survey of models for eyes and gaze. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(3), 478–500 (March 2010)
5. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer (2011)
6. Kaputa, D., Enderle, J.: An ultrasound based eye tracking system. Journal of Biomedical Engineering and Medical Devices **1**(1), 1–4 (April 2016)
7. Kar, A., Corcoran, P.: A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms. IEEE Access **5**, 16495–16519 (2017)
8. Khury-Yacub, B., Oralkan, O.: Capacitive micromachined ultrasonic transducers for medical imaging and therapy. J Micromech Microeng **21**(5), 054004–054014 (May 2011)
9. Ridgeway, G.: Generalized boosted models: A guide to gbm package (2007), http://cran.r-project.org/web/packages/gbm
10. Scally, B.M., Perek, D.R.: Ultrasound/radar for eye tracking (May 5 2017)
11. Sánchez-Ferrer, M.L., Grima-Murcia, M.D., Sánchez-Ferrer, F., Hernández-Peñalver, A.I., Fernández-Jover, E., del Campo, F.S.: Use of eye tracking as an innovative instructional method in surgical human anatomy. Journal of Surgical Education **74**(4), 668–673 (2017)
12. Termsarasab, P., Thammongkolchai, T., Rucker, J.C., Frucht, S.J.: The diagnostic value of saccades in movement disorder patients: a practical guide and review. Journal of Clinical Movement Disorders **2**(14) (2015)