1. Oleksii HYKA[1], 2. Andres VEJAR[1,2]

Netrix R&D Center (1), Wyższa Szkoła Ekonomii i Innowacji w Lublinie (2)
ORCID: 1.0000-0003-2129-5138; 2.0000-0002-2041-0387

Analysis of 3D head rotations for eye tracking calibration in large computer monitors

Abstract. In order to operate in the center of the vision field, the computer user maintains a frontal pose with respect to the screen. In this ideal situation, the computer user can visually explore the screen, moving only the eyes. The user activity changes when the computer screen is large, and it is necessary to change the pose and gaze orientation in order to explore the entire screen. The availability of high-end and high-size computer monitors (621mm width, 341mm height, or bigger), improves the possibilities of computer visualization but requires constant head pose changes to be performed by the computer user. In this study, we evaluate the 3D head rotation activity for the process of calibration on eye tracking. We compare the activity data for a fixed zigzag calibration pattern with respect to a random calibration pattern. The results show that the rotation data provide important information about the user screen exploration, and the incorporation of head rotations can contribute to reducing the quantity of information required to estimate the parameters of calibration models.

Streszczenie. Aby działać w centrum pola widzenia, użytkownik komputera utrzymuje pozycję frontalną względem ekranu. W tej idealnej sytuacji użytkownik komputera może wizualnie zbadać cały ekran, poruszając jedynie oczami. Aktywność użytkownika zmienia się, gdy ekran komputera jest duży i konieczna jest zmiana pozy i orientacji spojrzenia w celu eksploracji całego ekranu. Dostępność wysokiej klasy monitorów komputerowych (621mm szerokości, 341mm wysokości lub większych) zwiększa możliwości wizualizacji komputerowej, ale wymaga ciągłej zmiany pozycji głowy przez użytkownika komputera. W tej pracy oceniamy aktywność rotacji głowy w 3D dla procesu kalibracji na eye trackingu. Porównujemy dane aktywności dla stałego zygzakowatego wzorca kalibracji w odniesieniu do losowego wzorca kalibracji. Wyniki pokazują, że dane o rotacji dostarczają ważnych informacji o eksploracji ekranu przez użytkownika, a włączenie rotacji głowy może przyczynić się do zmniejszenia ilości informacji wymaganych do oszacowania parametrów modeli kalibracyjnych. (Analiza obrotów głowy 3D do kalibracji śledzenia ruchu gałek ocznych w dużych monitorach komputerowych)

Keywords: Visual detection, eye-tracking, calibration.
Słowa kluczowe: Detekcja wzrokowa, śledzenie wzroku, kalibracja.

Introduction

Eye tracking is an important problem in computer vision, and it is very dependent on the user's activity, any subtle movement or pose change will compromise the accuracy of eye-tracking systems. The end-user calibration process for eye tracking systems is one of the most important steps to customize the tracking algorithm to the user behaviour, and it can be crucial for a correct estimation and measurement. One of the methods for system calibration is to estimate a model to obtain a prediction of the screening point where the user is focusing his visual attention, generally by the use of linear regression.

To solve the calibration problem, it is required to track the head and eye positions in order to obtain the input data for the model estimation. Different methods for optimization and detection are available in the scientific literature [1-12]. In this work, we use the machine learning framework media pipe, face mesh and iris, which applies landmark detection with a mesh of 468 vertices to acquire the 3D geometry of the user's face from the video data frames.

The objective of this work is to perform a comparison of two calibration processes with random and zigzag-like calibration point sequences in a small and a large computer monitor. Head and eyes pose estimates can be collected as continuous angular measurements across multiple degrees of freedom (DOF). The DOF orientation can be described in terms of pitch, roll and yaw movements, as defined in flight dynamics. For our eye-tracking case, these movements are presented in Fig. 2 (right).

Eye tracking

Eye tracking is a composite problem. It requires the calculation of many sub-problems in a chain, also called a processing pipeline. The first step is the data acquisition of user activity using the appropriate sensors. In this work, we focus on the most important part of these sensors and data, the visible light cameras that produce the video stream. The following step is the decoding of the frames in order to obtain a data structure from the binary encoding formats of the digital cameras. The algorithms for the detection of the face and the detection of the eyes are the first image-processing algorithms of the pipeline for eye tracking. The solution to the problem of eye and face detection is a difficult task: it involves a considerable amount of computation time, it needs to be robust to variable ambient light conditions, and it needs to be able to take into account the inclusion of additional visual components, for example, the use of glasses or hats. The determination of a set of important features is one of the main steps to solving engineering problems using machine learning. A model with too many features can be memory and computationally intensive, and it can be impractical to apply a huge model to solve real-world problems. This is especially true when the algorithm is aimed to be executed in an embedded system. In the opposite case, a model with few features can produce strong errors in variable environments. Therefore, the selection of a correct model and the selection of a subset of features using dimensionality reduction techniques is generally a required step. Fig. 1 presents a typical eye-tracking calibration diagram.
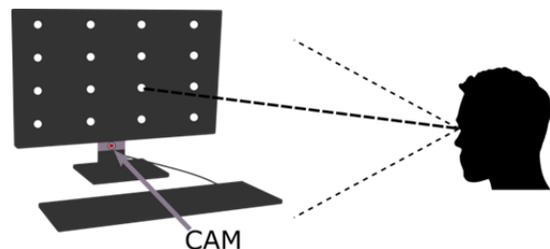

Fig.1. Eye tracking calibration diagram.

Face and eyes landmark detection

Landmark detection is a computer vision technique that allows to automatically detect key points of the contour of a template object in images. One important application of landmark detection is the detection of facial features. These

points can be used to calculate pitch, roll and yaw using the distance between facial landmark points. The available techniques for landmark detection can be divided into three categories: holistic method, constrained local model (CLM) and regression-based method. An example of facial landmark detection is presented in Fig. 2 (left).
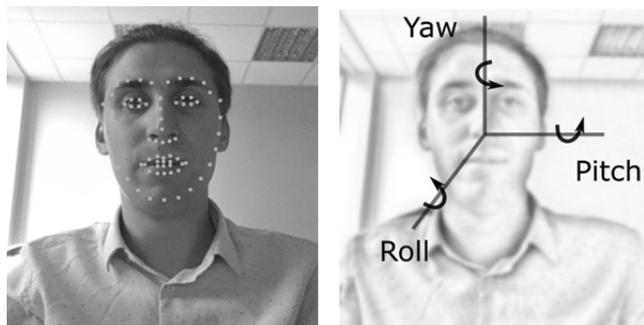

Fig.2. Facial landmarks (left), head rotations(right).

A comparison of facial landmark detection categories is presented in Table 1. The holistic method for whole face detection has good accuracy with respect to the cost of computation, and the CLM is performed for local patches and offers good performance and medium speed. The regression-based methods for local and whole-face detection show the best performance in accuracy and computation time. The regression-based method can be divided into direct regression, cascade regression and deep learning-based methods.

Table 1. Comparison of landmark detection methods.

Algorithm	Appearance	Shape	Performance	Speed
Holistic method	Whole face	Explicit	Poor generalization/good	Slow/fast
Constrained Local Method (CLM)	Local patch	Explicit	Good	Slow/fast
Regression-based method	Local patch/whole face	Implicit	Good/very good	Fast/very fast

Regression-based models can be classified into direct regression methods, deep learning-based regression methods and cascade regression methods. The direct regression method performs the prediction of the landmarks in one iteration, while cascade methods usually require an initial landmark location. Deep learning-based methods follow the cascade regression or direct regression. Regression-based face landmark detection methods have been the subject of intense research, and several scientific articles present methods with great accuracy. For example, in [13] a robust cascaded pose regression (RCPR) is proposed. The result of the RCPR method showed 80% precision and 40% recall.

Calibration of gaze estimation based on polynomial regressions

Gaze estimation is a popular topic in scientific research, spanning many domains such as human-computer interaction [16,18, 19] and computer vision [17, 20]. The gaze tracking can have a nonlinear behaviour with respect to eye movement and require adaptive regression for solving this problem. Polynomial regression is the most spread regression function used in gaze estimation. In [11], the authors present a polynomial regression used in commercial eye trackers:

$$\begin{pmatrix} PoR_x \\ PoR_y \end{pmatrix} = K \begin{pmatrix} 1 \\ v_x \\ v_y \\ v_x^2 \\ v_y^2 \\ v_x v_y \end{pmatrix}$$

PoR$_x$ and PoR$_y$ correspond to the point of regard in screen. The coordinates of v_x and v_y are normalized

vectors that generally track the iris and K is the unknown coefficients matrix. In [12] it is proposed a model of polynomial regression that provides high accuracy:

$$PoR_x = \begin{aligned} & K_{x0} + K_{x1}v_x + K_{x2}v_y + K_{x3}v_xv_y + K_{x4}v_x^2 \\ & + K_{x5}v_x^2v_y + K_{x6}v_x^3 + K_{x7}v_x^3v_y, \end{aligned}$$

$$PoR_y = \begin{aligned} & K_{y0} + K_{y1}v_x + K_{y2}v_y + K_{y3}v_xv_y + K_{y4}v_x^2 \\ & + K_{y5}v_y^2 + K_{y6}v_x^2v_y, \end{aligned}$$

where $K_{x0} \cdots K_{x7}$ and $K_{y0} \cdots K_{y6}$ are the model arameters to estimate.

As we can appreciate, the head rotations can be incorporated into the observation model. We are currently developing and testing a model using both the iris coordinates and head rotations as input data for the regression.

Head position detection

Head position detection is important to reduce the input data space for eye-tracking. The head detection is performed considering the detection of face features in the image frame. For accelerated face and eye detection, the Viola-Jones algorithm [14] has proven to be fast and accurate. The algorithm combines features from the face as shape and edge and other statistical models, i.e. adaptive boosting (AdaBoost). The algorithm of Viola-Jones uses the concept of integrating the image through the graph. The pixel sum of all regions in the image can be obtained by one traversal exploration of the image. This technique reduces the overall computation time. In [15], an improved method for face and eye detection based on the Viola-Jones algorithm was presented. This face detection method scans an image with windows of different sizes. Then, the system uses two Viola-Jones detectors to receive each window for a binary classification as a face or non-face. These detectors contain multiple cascaded classifiers trained by the AdaBoost algorithm. The idea of these detectors is to detect a frontal and a profile face separately. In the case of false detection, the window will be labelled as non-face for both detectors.

Experimental setup for calibration data acquisition

The proposed setup for data collection contains a camera laptop, a large 28-inch QHD monitor with a 621x341mm LCD matrix and small 15.6 inch HD monitor with a 194x344mm LCD matrix. To performing the calibration procedure a program using the Python language was develop to display the calibration dots in the required order. To simplify calibration process the point are generated with a very small offset from the monitor borders, allowing the computer user to follow the calibration in a comfortable manner. In Fig. 3 are presented two calibration patterns use, a zigzag-like sequence (left) and one instance of a random sequence (right). To calculate the head rotations, (yaw, pitch and roll) we selected six 3D facial landmark points, two for each axis. Facial landmark detection is performed using the mediapipe library.
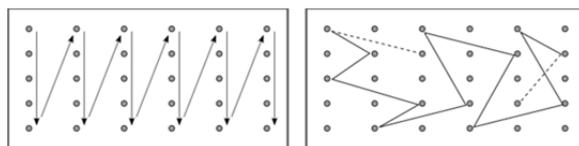

Fig.3. Calibration patterns, sequential (left), random (right).

Results

We perform measurements using two types of calibration patterns, one that cover the screen in a zigzag-like pattern and another that considers random sequences

over the base calibration points. The two types of calibration patterns are presented in Fig. 3. The data was collected with calibration measurements in a small (194x344 mm) and in a large (621x341 mm) computer monitor. It is important to note that the area of the large monitor is more than three times bigger than the area of the small monitor.
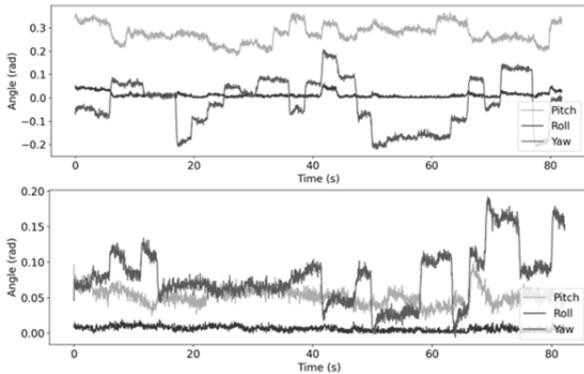

Fig.4. Head rotation time series (yaw, pitch, roll) for two instances of random sequence calibration patterns in large (upper) and small (lower) monitors.

The measurement was performed with a controlled distance between the monitor and the user head.
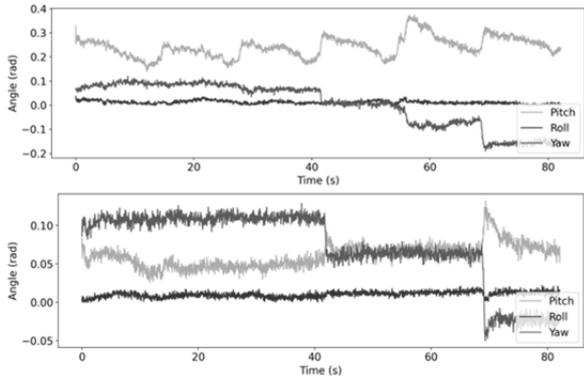

Fig.5. Head rotation time-series (yaw, pitch, roll) for a zigzag-like sequence calibration pattern in large (upper) and small (lower) monitors.

Fig. 4 are presented the time series of measurements for the yaw, pitch and roll rotations on a large and small monitor with random sequence calibration patterns. As we can see, the rotations in a small screen are minimal to comparing with a large screen. Similar results are achieved for zigzag-like sequence measurements presented in Fig. 5, where we can see how the head rotations are minimal in the case of the small screen. Therefore we can appreciate that a constant pose adjustment needs to be executed in the case of a large screen. For the two cases, the rolling activity is negligible. For the following results, we consider only the most important rotations, pitch and yaw.

For the yaw and pitch rotation data, a cubic smoothing spline approximation was performed to estimate the average trajectory for the time-series measurements for the two calibration types in small and big monitors—the result of the cubic spline approximation and the point cloud presented in Fig. 6 and 7.

In Fig. 8 and 9 are presented the histograms for random and sequential calibration points of yaw measurements in large (left) and small (right) monitors. The histograms with zigzag-like sequence calibration points clearly show distributions of horizontal movements of the head. For the large monitor, the user activity can be classified into 4 classes of head movement while for the same zigzag pattern in the small screen, three classes of head

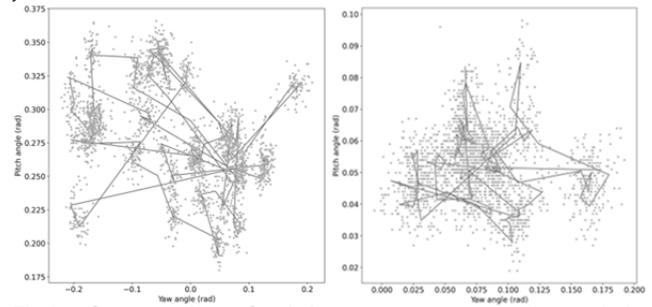movements are enough to follow the entire calibration process.

Fig.6. Scatter plot of pitch versus yaw rotations during measurements of two random calibration sequences in large (left) and small (right) monitors. A cubic smoothing spline approximation is calculated over the point cloud to estimate the average trajectory.
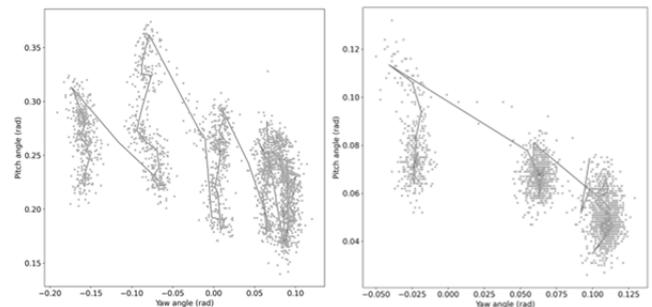

Fig.7. Scatter plot of pitch versus yaw rotations during measurements of two zigzag-like calibration sequences in large (left) and small (right) monitors. A cubic smoothing spline approximation is calculated over the point cloud to estimate the average trajectory.

In the small screen, as the result of small and short head movements, the resolution of detection is limited by the pixel size. This quantization is visible in Fig. 6 and 7.
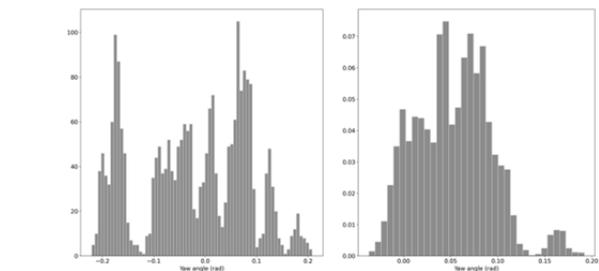

Fig.8. Histogram for two instances of random sequence calibration points of yaw rotations in large (left) and small (right) monitors.
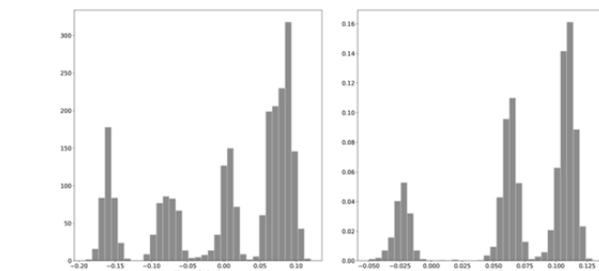

Fig.9. Histogram of zigzag-like sequence calibration points of yaw rotations in large (left) and small (right) monitors.

A global mean and standard deviation calculation was performed for three sets of data and is presented in Table 2. We can appreciate the small standard deviation in the small monitor, given that fewer rotations from the mean position are required. The mean rotation angle variate between the large and the small monitor, given that a different pose need to be performed for the computer user

given the variation on the locations of the small and the large screen.

Table 2. Mean and standard deviation of yaw, pitch and roll for large and small computer monitor.

Large screen (621x341mm, 39.1 pixel/mm^2)				Small screen (344x194mm, 31.1 pixel/mm^2)			
	Yaw	Pitch	Roll		Yaw	Pitch	Roll
mean	0.009	0.272	0.018	mean	0.053	0.046	0.009
std	0.101	0.037	0.013	std	0.04	0.012	0.006

Additionally, the center of the user's iris was tracked during the zigzag-like calibration process with large and small monitors. Fig. 10 is presented the histogram of the iris displacement speed. The speed calculation $\Delta L/\Delta t$ was performed by estimating the distance between two consecutive iris 3D location measurements ΔL and the time between measurements Δt.

Based on this histogram, we can appreciate that the user needs to accommodate the large screen performing faster iris displacements on the large monitor. This implies that the combined effect of rotation and iris movement allows the computer user to explore large screens in a similar way to small screens.
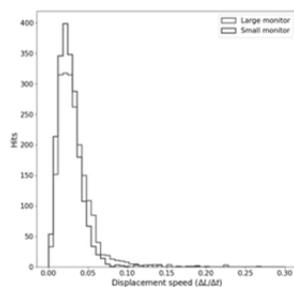

Fig.10. Histogram of displacement eyes speed.

Conclusions and Further Work

In this paper, we track and analyze 3D head rotations and iris movements for eye-tracking calibration considering small and large computer monitors. We perform an online head rotation calculation and collect measurements for random and zigzag-like calibration pattern sequences with a large and small monitor. The time series of yaw, pitch and roll rotations was presented, and a cubic smoothing spline approximation was calculated to estimate the trajectory of yaw and pitch rotations. Based on our measurements, we can classify user activity with small and large computer monitors. The analysis of the data allows us to validate the incorporation of head rotation data into the calibration models. A further contribution in preparation will analyze the calibration results in a regression model that consider only iris data versus a regression model considering iris data and head rotation data.

Authors
mgr inż. Oleksii Hyka, Netrix S.A, Research and Development Center, Związkowa 26, 20-148 Lublin e-mail: oleksii.hyka@netrix.com.pl dr inż. Andres Vejar, Netrix S.A., Research and Development Center, Związkowa 26, 20-148 Lublin, University of Economics and Innovation in Lublin, ul. Projektowa 4, 20-209 Lublin, e-mail: andres.vejar@netrix.com.pl.

REFERENCES

[1] Rymarczyk T., Kłosowski G., Hoła A., Sikora J., Tchórzewski P., Skowron Ł., Optimising the Use of Machine Learning Algorithms in Electrical Tomography of Building Walls: Pixel Oriented Ensemble Approach, Measurement, 188 (2022), 110581.

[2] Koulountzios P., Rymarczyk T., Soleimani M., Ultrasonic Time-of-Flight Computed Tomography for Investigation of Batch Crystallisation Processes, Sensors, 21 (2021), No. 2, 639.

[3] Kłosowski G., Rymarczyk T., Niderla K., Rzemieniak M., Dmowski A., Maj M., Comparison of Machine Learning Methods for Image Reconstruction Using the LSTM Classifier in Industrial Electrical Tomography, Energies 2021, 14 (2021), No. 21, 7269.

[4] Rymarczyk T., Król K. Kozłowski E., Wołowiec T., Cholewa-Wiktor M., Bednarczuk P., Application of Electrical Tomography Imaging Using Machine Learning Methods for the Monitoring of Flood Embankments Leaks, Energies, 14 (2021), No. 23, 8081.

[5] Majerek D., Rymarczyk T., Wójcik D., Kozłowski E., Rzemieniak M., Gudowski J., Gauda K., Machine Learning and Deterministic Approach to the Reflective Ultrasound Tomography, Energies, 14 (2021), No. 22, 7549.

[6] Kłosowski G., Rymarczyk T., Kania K., Świć A., Cieplak T., Maintenance of industrial reactors supported by deep learning driven ultrasound tomography, Eksploatacja i Niezawodnosc – Maintenance and Reliability; 22 (2020), No 1, 138–147.

[7] Gnaś, D., Adamkiewicz, P., Indoor localization system using UWB, Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12 (2022), No. 1, 15-19.

[8] Styła, M., Adamkiewicz, P., Optimisation of commercial building management processes using user behaviour analysis systems supported by computational intelligence and RTI, Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12 (2022), No 1, 28-35.

[9] Korzeniewska, E., Krawczyk, A., Mróz, J., Wyszyńska, E., Zawiślak, R., Applications of smart textiles in post-stroke rehabilitation, Sensors (Switzerland), 20 (2020), No. 8, 2370.

[10] Sekulska-Nalewajko, J., Gocławski, J., Korzeniewska, E., A method for the assessment of textile pilling tendency using optical coherence tomography, Sensors (Switzerland), 20 (2020), No. 13, 1–19, 3687.

[11] Morimoto, C. H., & Mimica, M. R., Eye gaze tracking techniques for interactive applications. Computer vision and image understanding, 98 (2005), No.1, 4-24.

[12] Blignaut, P., A new mapping function to improve the accuracy of a video-based eye tracker. In Proceedings of the south african institute for computer scientists and information technologists conference , 2013, 56-59.

[13] Burgos-Artizzu, X. P., Perona, P., & Dollár, P., Robust face landmark estimation under occlusion. In Proceedings of the IEEE international conference on computer vision , 2013, 1513-1520.

[14] Viola, P., & Jones, M., Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, 1 (2001).

[15] Kaddouhi El, Saaidi S., Abarkan A., Eye detection based on the Viola-Jones method and corners points. Multimedia Tools and Applications, 76 (2017), No. 21, 23077-23097.

[16] Kong, A., Ahuja, K., Goel, M., & Harrison, C., EyeMU Interactions: Gaze+ IMU Gestures on Mobile Devices. In Proceedings of the 2021 International Conference on Multimodal Interaction, 2021, 577-585.

[17] Ahuja, K., Islam, R., Parashar, V., Dey, K., Harrison, C., & Goel, M., Eyespyvr: Interactive eye sensing using off-the-shelf, smartphone-based vr headsets. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2 (2018), No. 2, 1-10.

[18] Ahuja, K., Banerjee, R., Nagar, S., Dey, K., & Barbhuiya, F., Eye center localization and detection using radial mapping. In 2016 IEEE International Conference on image processing (ICIP), 2016, 3121-3125.

[19] Klamka, K., Siegel, A., Vogt, S., Göbel, F., Stellmach, S., & Dachselt, R., Look & pedal: Hands-free navigation in zoomable information spaces through gaze-supported foot input. In Proceedings of the 2015 ACM on international conference on multimodal interaction, 2015, 123-130.

[20] Papoutsaki, A., Gokaslan, A., Tompkin, J., He, Y., & Huang, J. (2018, June). The eye of the typer: a benchmark and analysis of gaze behavior during typing. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, 2018, 1-9.