Look! Blickschulungsbrille: Technical Specifications

THOMAS C. KÜBLER, Look! ET UG (haftungsbeschränkt), Germany
ELENA WAGNER

Fig. 1. Look! Blickschulungsbrille

The *Look! Blickschulungsbrille* is a mobile eye-tracking device capable of pupil detection based eye-tracking as well as appearance based gaze estimation. For pupil detection, research institutions can directly access the algorithms implemented in EyeRecToo, such as *Purest* pupil detection and tracking as well as *Swirski* temporal eye modeling by ellipse shape.

For real-world studies in challenging conditions, i.e., when pupil detection is error-prone due to eyeglasses, make-up, irregularly shaped pupil or iris, convolutional neural networks are likely to outperform the traditional approach by estimating directly the gaze vector from a complete image of the eye.

The *Blickschulungsbrille* consists of three miniature USB cameras. We call the two cameras that produce an image of the left and right eye *eye cams* and the camera that points from the face of the wearer towards the scenery the *scene cam*. The scene cam records what is happening and the eye cams help us determine where in the scene the wearer is looking at. They are mounted on a 3D printed frame which also houses near-infrared filters combined with eye-safe near-infrared illumination. The whole device works via a single USB2.0 connector.

Technical documentation

How to cite:
Thomas C. Kübler and Elena Wagner. 2022. Look! Blickschulungsbrille: Technical Specifications.

1 TECHNICAL SPECIFICATIONS

The scene camera records a cyclops view from a central perspective located over the nose. Its resolution is 640×480 pixels. The default wide angle lens covers a relatively large field of view that spans 120x70°. Eye cameras operate at a resolution of 320×240 pixels. They can go 640×480, if required for some reason. Sampling frequency is 30 Hz, resulting in 33.33ms intervals between frames. For research use-cases the camera firmware can be upgraded to operate at 60 Hz, provided a device with sufficient real-time processing capabilities is connected to handle to additional load (most modern CPUs can easily cope with that).

2 GAZE ESTIMATION

At the core of the Look! software is a neural network that predicts jointly the gaze position on the scene camera image as well as the geometric orientation of both eyeballs. We utilize a variant of Mobilenet v3 tiny architecture that was downsized to provide minimal latency and CPU load without significant performance drops compared to full-resolution, large-scale networks.

It processes the eye images at a downsized resolution of 180×135 pixels. Depending on the variant of the network configured, we process:

- Both eyes jointly as a two-channel image.
- Both eyes separately with the same network weights and a mirrored right eye, followed by concatenation of the features before the final regression layers.

The neural network is executed either via OpenVINO (https://openvino.ai), which will optimize the network for execution on Desktop CPUs, or as a tensorflow-lite model, optimized and quantized for ARM processors such as the Raspberry Pi. The Look! recording unit runs a quantized tensorflow-lite model, optionally also on the Coral tensor processing unit, if available (https://coral.ai). That model is, besides optimizations for the specific architecture, identical to the Desktop version and produces very similar results.

3 RECOMMENDED METRICS

With the given resolution and sampling rate, we recommend the use of fixation based metrics (dwell time, dwell on AOI, gaze on target,...). It is possible to use time-based metrics (first glance on AOI, time to first fixation,...).

It is not possible to calculate saccadic velocity profiles (such as saccadic peak velocity, acceleration,...).

4 CALIBRATION

4.1 Calibration Markers

The Look! software detects both Aruco markers (as in EyeRecToo) and concentric markers (see Figure **??**). We recommend the use of concentric markers as it is easier to explain to the subjects where to look at. Also, the marker is considerably smaller than a reliably detectable Aruco marker. A calibration can be started by clicking the calibrate button or long pressing one of the remote control buttons. The calibration procedure needs to be manually finished in the same way.

Fig. 2. Concentric calibration marker

4.2 Fitting methods

In hybrid calibration mode, Look! will first try to estimate a simple offset to the gaze pixel coordinates predicted by the neural network. This simple calibration is indicated by a yellow marker on the detected calibration marker. Once the indicator turns green (after some seconds of collecting calibration data), an extensive calibration is performed: a full mapping function is determined via kernel ridge regression.

While the calibration is running, the calibration function will continuously and automatically update. It is recommended to calibrate until the marker detection and gaze estimation indicators overlap for all desired gaze directions.

Calibration be be applied either on the predicted gaze coordinate, on the predicted per eye orientation, or on the eye orientation calculated as *instantaneous gaze* via EyeRecToo.

4.3 Caveats

Note that the field of view of our scene camera is very large. It might not be necessary nor feasible to calibrate the complete image. However, gaze estimation quality deteriorates dramatically when leaving the calibrated area. It is thus important to calibrate at least that field of view which is expected to be relevant.

5 OPTIONAL EXTENSION MODULES

5.1 Streaming via WebRTC and MJPG

When streaming via MJPG, a web server provides access to the MJPG stream. An AJAX API allow to remote control the device. This way of streaming is recommended as performance requirements for video compression are critical even for powerful systems and JPEG compression can often be done in hardware. The Look! Recording unit utilizes this interface.

It is possible to configure the software to provide a WebRTC stream of the display as well as a stream of the data. Please note that h264 or VP9 video compression might put substantial additional load on the recording device. For this configuration a WebRTC data channel can be used to remote control the software.

6 FRAME

The frame is made of bio-compatible PA12 Nylon. Its flexibility allows fitting many different head shapes. Thereby, it presses lightly against the side of the head, fixing itself in position and avoiding slippage. The area where the pressure is produced (near the ears) is slightly thicker to distribute the force on a larger area and make wearing more comfortable. The frame is optimized to be worn on the tip of the nose. When worn correctly, the eye cameras should be at the bottom of the field of view of the wearer and not obstruct the important central field of vision. If this is not the case, the height of the nose piece can be adjusted.

Where the earpieces meet the front piece (where regular glasses have a joint), stability is of special importance to avoid that the frame bends diagonally. This would lead to a skewed positioning of the glasses on the nose and a less accurate, oddly looking fit. When adapting the frame, this area should therefore be handled with special care (e.g., no holes drilled there).

Two NIR-LEDs (QBLP630-IR3) are placed on each eye camera. The LEDs are connected to pins on the camera chip for power supply, so no extensive wiring is required. The NIR illumination used is comparatively weak - in fact often much weaker than natural sunlight. That is because we are working with methods that do not require a glint to be visible. The illumination is only required for situations in which ambient illumination is dim. For accurate eye modelling based on glints it might be necessary to add more and stronger LEDs.

6.1 Safety considerations

Look! Blickschulungsbrille bears the CE mark and Look! ET UG (haftungsbeschränkt) declares conformity with the relevant regulations:

(1) NIR illumination is classified as safe for continuous exposure following IEC/EN 62471: 2008 Photobiological safety of lams and lamp systems. The maximum intensity of the LEDs is stated as 1.6 mW/sR. We advise to keep a distance of at least 1 cm between LEDs and eye. Undamaged devices ensure an even larger distance through their geometry. Please do not use damaged devices that may allow the LEDs to get closer to the eye than intended. Broken camera arms can be repaired!
(2) RoHS2 Directive 2002/95/EC
(3) Electronics passed EMC/LVD tests following EN62368-1:2014+A11:2017
(4) PA12 Nylon is safe for skin contact.

Look! Recording unit bears the CE mark, documents can be requested.

7 AI TRAINING

The built-in domain-specific data augmentation [Eivazi et al. 2019] allows to perform training with relatively few recordings. For recording the training data, we made participants look at the center of a calibration marker [Romero-Ramirez et al. 2018]. These can easily be detected in the scene camera image and tuples of eye images and gaze targets in the scene camera image can be acquired.

Other approaches abstract over the scene camera's calibration [Tonsen et al. 2020]. We chose to predict scene camera pixel coordinates directly as well as gaze direction relative to each eye camera's coordinate system. The consequence of this decision is that the device requires a calibration. As we recommend a personal calibration anyways, both steps can simply be merged. The gaze estimate is adjusted to the subject's specific eye as well as to the specific scene camera's intrinsics. We implemented a calibration procedure following the CalibMe [Santini et al. 2017] approach. This calibration is applied on top of the neural network gaze estimate. That way we achieve similar accuracy as when using EyeRecToo, but the network is not as susceptible to losing track of the pupil as the pupil detection methods employed in EyeRecToo are.

8 DEMONSTRATION OF FLEXIBILITY

Researchers are unlikely to fiddle around with their expensive eye-tracking gear, cut off pieces, replace modules or change their hardware. In this section we want to give an impression on how we utilize the ability to build our own, relatively inexpensive eye-tracking devices as well as the ability to easily modify them and what kinds of problems such modifications could possibly solve.

8.1 Angle of the scene camera

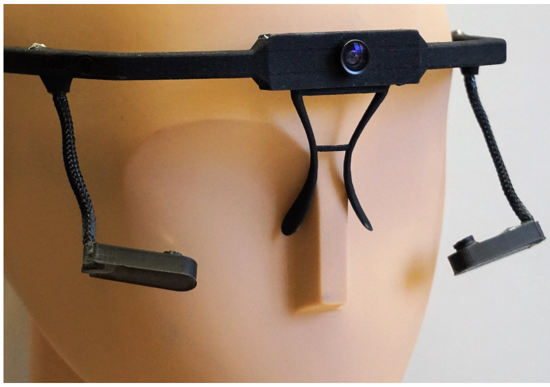
When recording tasks where the participants have to manipulate something with their hands, many scene cameras reach the limits of their field of view. We found some scene cameras not to cover that area at all. Generally, gaze accuracy deteriorates towards that region (as calibration is hard to perform, glints disappear on the sclera and camera viewing angle on the pupil is extreme). For these recordings we found it necessary to slightly tilt the scene camera of the TuEye device towards the ground. Doing so is relatively easy with the provided 3D sketches. Additionally, we produced a variant with adjustable scene camera angle (figure 3b).

8.2 Flexible eye cameras

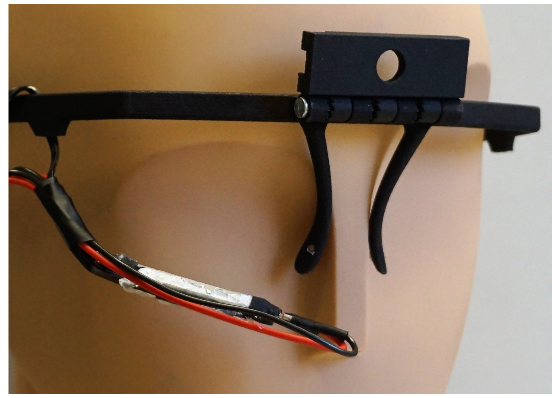
Having a rigid frame reduces mechanical wearing and makes the device easy to use. With a fixed eye camera position deep learning based methods can be trained on a smaller subset of images and a bad camera adjustment becomes unlikely. However, the flexibility gained by being able to move the eye cameras might be important in some cases. The frame of habitual eyeglasses might be in the way, the negative effect of eyelashes and dropping eyelids or reflections on eyeglasses can be reduced or completely avoided by a good, user-specific camera positioning. This can be achieved either by the use of multiple cameras (e.g., Tobii), an adjustable nose piece that preserves the geometry between eye and scene cameras (TuEye), or adjustable eye cameras (PupilLabs core).

To gain more flexibility in eye camera placement we replaced the connector between the eye camera and the frame with a flexible aluminium wire (Figure 3a). This setup allows for full freedom in the adjustment of the eye cameras without joint mechanics limitations. That makes finding a good camera position and orientation comparatively easy.

We found that the wire can withstand a reasonable number of readjustments without breaking when tightly fixed to the eye camera with epoxy and not bent too far in a single spot.



(a) TuEye variant with flexible eye cameras that can be adjusted to work well with eyeglasses and the most variance in head shape.



(b) TuEye variant with an acrylic diffuser as illumination pattern as well as an adjustable scene camera to capture the hand action space.

Fig. 3

8.3 Diffuse illumination

As of now, head-mounted eye-tracking devices mostly utilize direct illumination via near-infrared LEDs. This wavelength is invisible to the human eye, yet provides a relatively constant illumination of the eye and therefore solid contrast between pupil, iris and sclera. Often the resulting glints are used to construct a geometrical model of the eyeball (or to compensate for device slippage), however current methods can work without them [Swirski and Dodgson 2013; Tonsen et al. 2020] and utilize the LEDs only as a source of illumination.

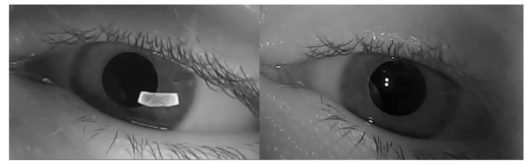
Less explored is the use of diffuse illumination, i.e., scattering the source of light over a larger area. Doing so might reduce shadows due to eyelashes and increase the ease of adjusting cameras and illumination to individual eyes.

Figure 4b demonstrates the different images produced by a diffuse bar illuminator versus two separate LEDs. At first glance such an approach seems to make pupil detection even harder. In our experiments we found that even edge based methods perform similar for both approaches. We expect additional potential of this approach when utilizing deep learning based gaze estimation methods as similar illumination approaches are applied during cornea topography mapping. It might be possible to build more accurate cornea models by combining the illuminator projection over multiple eye orientations. The curvature of the reflected illuminator bar could also be a helpful indication of cornea shape that could potentially be utilized by deep learning gaze estimation approaches. We did however not test this.

8.4 Line scanning sensors

Besides cameras, which are expensive and energy hungry, photodiode arrays are an interesting way to infer gaze location [Tonsen et al. 2017]. Here, we demonstrate how the TuEye frame can be modified to house three line-scanning sensors (TSL1401CL), NIR-pass filters as well as their cabling. We placed the line scanners diagonally in order to try to overcome the accuracy issues of other photodiode based devices in vertical gaze estimation.

(a) Sketch of the diffuser element and LED placement within the diffuser. The part consists of opaque acrylic glass with the front (facing away from the eye) covered by aluminium foil.



(b) Resulting projection of the diffuser on the cornea (left) and traditional glints as produced by the standard TuEye (right).

Fig. 4

We evaluated the device on 6 subjects. Two of them participated both with and without eyeglasses, leading to 8 recordings that we consider as separate subjects in the following. Gaze prediction was performed with a fully connected neural network of 3 layers with 96/24/2 neurons and an input vector dimension of 384 that contained the concatenated readings of the three sensors. The network was trained on 20 calibration points within 34° field of view range and evaluated on another 12 points. A chin rest was used in order to fix the gaze targets relative to the subject's head. Average gaze estimation error on the evaluation points was 7.6° when trained on all subjects jointly. With networks trained specifically for one subject, we reached a mean error of 3.4-12.0° (mean 6.3°). These results are preliminary and the number of subjects very low, however they demonstrate the potential of similar approaches for future generations of mobile eye-tracking devices.

ACKNOWLEDGMENTS

REFERENCES

Shaharam Eivazi, Thiago Santini, Alireza Keshavarzi, Thomas Kübler, and Andrea Mazzei. 2019. Improving real-time CNN-based pupil detection through domain-specific data augmentation. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. 1–6.

Francisco J Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. 2018. Speeded up detection of squared fiducial markers. *Image and vision Computing* 76 (2018), 38–47.

T. Santini, W. Fuhl, and E. Kasneci. 2017. CalibMe: Fast and Unsupervised Eye Tracker Calibration for Gaze-Based Pervasive Human-Computer Interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*.

Lech Swirski and Neil Dodgson. 2013. A fully-automatic, temporal approach to single camera, glint-free 3D eye model fitting. *Proc. PETMEI* (2013), 1–11.

Marc Tonsen, Chris Kay Baumann, and Kai Dierkes. 2020. A High-Level Description and Performance Evaluation of Pupil Invisible. *arXiv preprint arXiv:2009.00508* (2020).

Marc Tonsen, Julian Steil, Yusuke Sugano, and Andreas Bulling. 2017. Invisibleeye: Mobile eye tracking using multiple low-resolution cameras and learning-based gaze estimation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–21.