



## Data Article

# Comprehensive VR dataset for machine learning: Head- and eye-centred video and positional data

Alexander Kress<sup>a,\*</sup>, Markus Lappe<sup>b</sup>, Frank Bremmer<sup>a</sup><sup>a</sup>Department of Neurophysics, Philipps University Marburg, Karl-von-Frisch Straße 8a, 35043 Marburg, Hesse, Germany<sup>b</sup>Institute of Psychology, University Münster, Fliegenerstraße 21, 48149 Münster, North Rhine-Westphalia, Germany

## ARTICLE INFO

## Article history:

Received 24 September 2024

Revised 30 October 2024

Accepted 25 November 2024

Available online 29 November 2024

Dataset link: [Comprehensive VR Dataset for Machine Learning: Head- and Eye-Centred Video and Positional Data \(Original data\)](#)

## Keywords:

Eye tracking

Head tracking

Deep learning

Spatial navigation

Foraging behaviour

Behavioural data

Naturalistic VR locomotion

## ABSTRACT

We present a comprehensive dataset comprising head- and eye-centred video recordings from human participants performing a search task in a variety of Virtual Reality (VR) environments. Using a VR motion platform, participants navigated these environments freely while their eye movements and positional data were captured and stored in CSV format. The dataset spans six distinct environments, including one specifically for calibrating the motion platform, and provides a cumulative playtime of over 10 h for both head- and eye-centred perspectives.

The data collection was conducted in naturalistic VR settings, where participants collected virtual coins scattered across diverse landscapes such as grassy fields, dense forests, and an abandoned urban area, each characterized by unique ecological features. This structured and detailed dataset offers substantial reuse potential, particularly for machine learning applications.

The richness of the dataset makes it an ideal resource for training models on various tasks, including the prediction and analysis of visual search behaviour, eye movement and navigation strategies within VR environments. Researchers can leverage this extensive dataset to develop and refine algorithms requiring comprehensive and annotated video and positional data. By providing a well-organized and detailed

\* Corresponding author.

E-mail address: [alexander.kress@physik.uni-marburg.de](mailto:alexander.kress@physik.uni-marburg.de) (A. Kress).

dataset, it serves as an invaluable resource for advancing machine learning research in VR and fostering the development of innovative VR technologies.

© 2024 The Authors. Published by Elsevier Inc.  
This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Neuroscience.
Specific subject area	Self-motion, navigation and visual search in Virtual Reality (VR)
Type of data	Table, Video, Raw, Processed.
Data collection	For data collection participants were tasked to collect coins in different virtual environments. Movement within the virtual reality was realized by the participant walking on a VR motion platform (Virtualiser Elite 2 by Cyberith). During the experiment the participants were wearing an HTC Vive Pro Eye VR Headset with a build in Tobii Eye tracker (120 Hz sampling rate). During the experiment head position and rotation, hip position and rotation as well as gaze position were recorded.
Data source location	<ul style="list-style-type: none"><li>• Institution: The Adaptive Mind – TAM DataHub</li><li>• City/Town/Region: Marburg</li><li>• Country: Germany.</li></ul>
Data accessibility	Repository name: Comprehensive VR Dataset for Machine Learning: Head- and Eye-Centred Video and Positional Data Data identification number: <a href="https://doi.org/10.60834/tam-datahub-6.2">10.60834/tam-datahub-6.2</a> Direct URL to data: <a href="https://doi.org/10.60834/tam-datahub-6.2">10.60834/tam-datahub-6.2</a> Instructions for accessing these data: Download the desired archive files and unpack all of them into the same directory.
Related research article	none.

1. Value of the Data

- This dataset is valuable for studying scenery-dependent self-motion and navigational behaviours and visual search in quasi natural settings.
- It is also useful for training computer vision models and computational models of visual search, as it includes annotated gaze data along with head and eye-centered video recordings.
- Researchers can leverage this data to improve algorithms for gaze estimation, gaze prediction and enhance the accuracy of head movement tracking.
- The inclusion of annotated gaze positions allows for detailed analysis and validation of theoretical models in neuroscience.
- This data can also be applied to develop and refine augmented reality and virtual reality applications by understanding user gaze and head movement patterns.

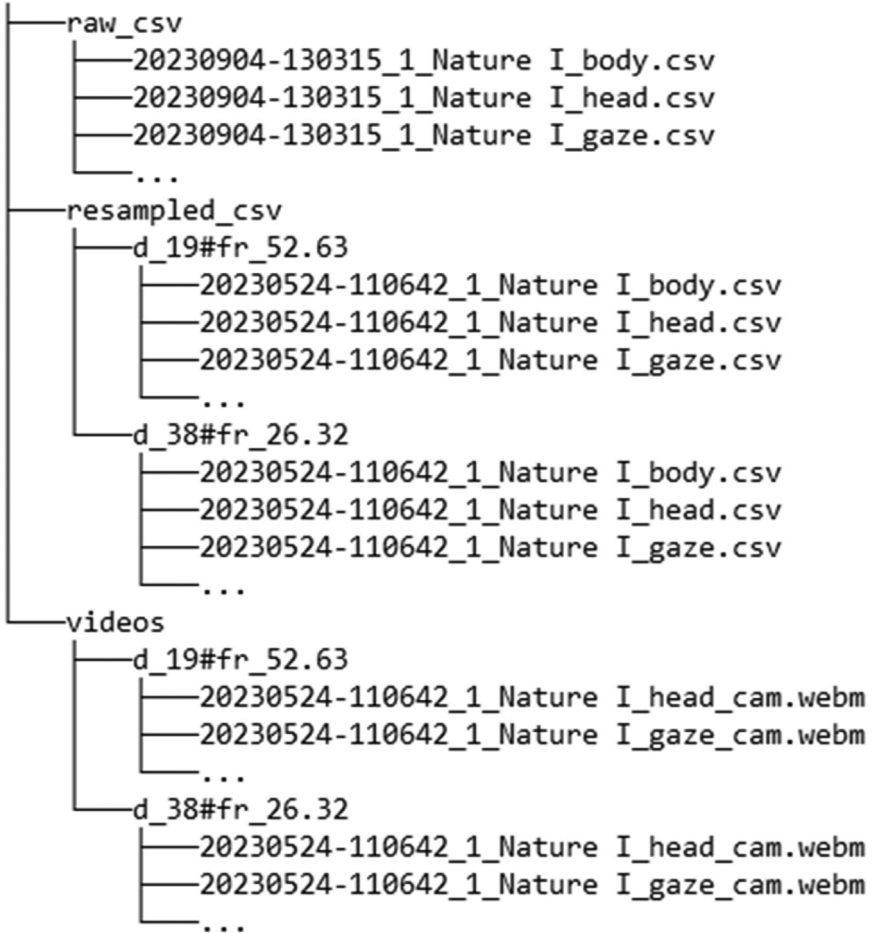
2. Background

The data was collected to train a Variational Autoencoder with the purpose of parameterizing visual self-motion stimuli as well as predict heading and gaze parameters from a given set of parameters. In addition, the data was also used to train a modified 3D ResNet to predict heading and gaze parameters in a goal driven approach to model parts of the primate dorsal visual pathway in a similar approach to DorsalNet [1].

### 3. Data Description

The dataset comprises head- and eye-centered videos recorded from human participants performing a search task in a Virtual Reality (VR) environment. In blocks of trials, participants navigated freely using a VR motion platform to enable a naturalistic mode of locomotion. Eye and positional data were recorded and saved in CSV format, corresponding to the recorded video files.

The dataset is organized into three main directories as illustrated in Fig. 1.



**Fig. 1.** Structure of the dataset. Raw csv, resampled csv and video files each have their respective directory. Resampled data is further subdivided into 52.63 Hz and 26.32 Hz sampling rate.

1. raw\_csv: Contains the original body, head, and gaze data as recorded during the experiment.
2. resampled\_csv: Includes data resampled to 26.32 Hz and 52.63 Hz.
3. videos: Contains video files recorded from the replay of the resampled data.

Each CSV file is structured identically, with separate files for body, head, and gaze data. The filenames include the recording start date and time, recording block number and environment name. Table 1 details the CSV column descriptions. Note that all data is recorded in global co-ordinates, with pitch and yaw of gaze given in eye-in-head coordinates. Since Unity was used

**Table 1**

Description of the columns in the csv data. Note, that all data is recorded in global coordinates, while the pitch and yaw of the gaze is given in eye in head coordinates.





Column	Description	UNIT
TIME	Time in seconds since start of recording	s
POS.X	X position relative to ground plane	Degree [°]
POS.Y	Y position relative to ground plane (up)	Degree [°]
POS.Z	Z position relative to ground plane	Degree [°]
ROT.X	Pitch	Degree [°]
ROT.Y	Yaw	Degree [°]
ROT.Z	Roll	Degree [°]

for the creation of the different VR environments, presentation, and data recording, the y-axis denotes the upwards direction.

The dataset includes annotated video data from six environments, one of which was used solely for motion platform calibration. The total playtime across all environments is 10:21:50 h for both head- and eye-centered data. Table 2 provides the cumulative playtime as well as a short description of the scenery for each environment. The environments were designed to

**Table 2**

Cumulative video playtime for each environment. The given playtime is equal for head- and eye-centred videos. A short description of each scenery is given in the last column.

Environment	Sample image	Cumulative play time	Description
ENDLESS HALLWAY		00:02:06	A straight corridor with striped walls, that infinitely repeats itself.
NATURE I		02:16:22	A serene and open grassy field environment with sparse trees and bushes.
NATURE II		06:13:29	An expanded and denser version of Nature I, featuring additional trees, vegetation and other points of interest.
HDR FORREST		00:36:44	A darker, more realistic forest environment with less vegetation on eye level.

(continued on next page)

**Table 2** (continued)

Environment	Sample image	Cumulative play time	Description
FLODDED GROUNDS		01:02:29	An urban environment characterized by waterlogged areas, featuring flooded fields and marshlands as well as a selection of buildings with an explorable interior.
FANTASY FORREST		00:10:37	A dense forest featuring small, steep hills and vibrant colours

simulate a diverse collection of natural sceneries, including grassy fields, dense forests, and an abandoned urban area.

Heading direction was exclusively forward, as the participants did at no point walk backwards. The distribution of heading directions collapsed across environments can be seen in Fig. 2.

## 4. Experimental Design, Materials and Methods

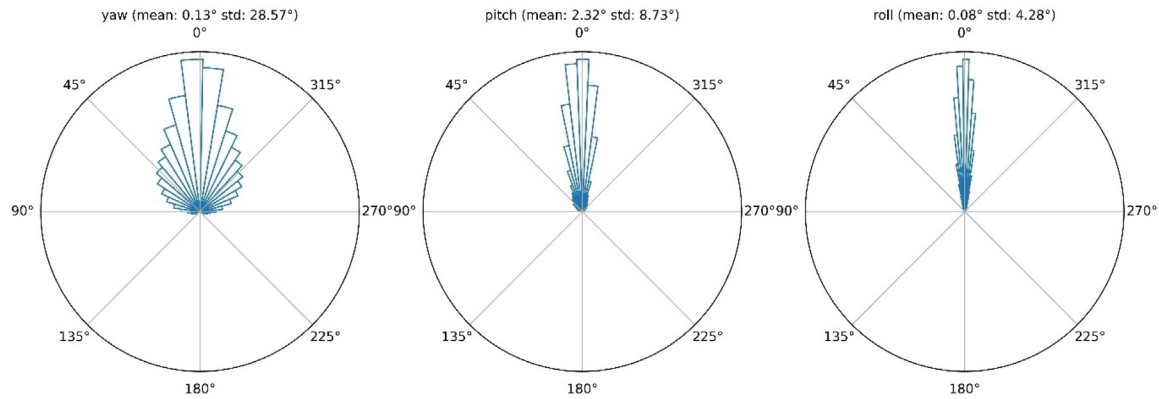
### 4.1. Participants

The study sample comprised 14 participants, consisting of seven males and six females, all recruited from a university student population. Consequently, the cohort consisted entirely of young adults, with ages ranging from 20 to 30 years. All participants possessed normal or corrected-to-normal vision. Individuals requiring corrective lenses were able to use their glasses in conjunction with the virtual reality (VR) headset, ensuring consistent visual acuity across participants.

### 4.2. Experimental design and data collection

The data collection for this study was designed to simulate naturalistic environments within a Virtual Reality (VR) framework. The creation and presentation of all scenes to the participants were performed using the Unity game engine (version 2021.3.1f1 [2]), with freely available assets for the maps sourced from the Unity Asset Store [3,4]. Participants navigated these environments using a VR motion platform [5], with their primary task being the collection of virtual coins dispersed throughout varied landscapes.

Participants were immersed in five distinct environments, each crafted to present unique ecological features: (1) a small grassy field adorned with sparse trees and bushes; (2) an expanded version of the grassy field featuring denser tree populations and notable points of interest; (3) a darker, more lifelike forest; (4) a densely vegetated, hilly forest; and (5) an abandoned urban area with accessible buildings. Additionally, a straight corridor was utilized for system testing and calibration purposes.



**Fig. 2.** Distribution of heading direction collapsed across environments. Shown, are the yaw, pitch and roll of the head relative to the current heading direction. Mean and standard deviation of the distributions are given in the heading of the respective subplot.

In each environment, coins were strategically placed at diverse locations. The number of coins varied between six and eleven per environment, all of which were instantiated at the beginning of each trial, allowing participants to collect them in any order. To mitigate motion sickness, participants were given the flexibility to take breaks as needed.

Prior to each trial, a thorough calibration process was conducted. The calibration involved a two-step procedure: First, the eye tracker was calibrated using the HTC Vive Pro Eye Headset's built-in 5-point calibration system [6]. Second, the forward direction of the motion platform was aligned. This was achieved by instructing the participants to stand upright and look into the direction their hips were facing, which was then synchronized with their head orientation in Unity, ensuring precise directional control.

Post-calibration, the motion platform enabled highly accurate and naturalistic 360° movement. Throughout each trial, positional and rotational data were recorded and stored in CSV format. The recorded parameters included body position in world space, head position in world space, and gaze direction relative to the head orientation.

Following the recording sessions, the collected data underwent resampling at frequencies of 26.32 Hz and 52.63 Hz. The resampled data were re-imported into Unity, where the participants' movements were replayed for further analysis. During this replay, both head and gaze camera videos were recorded, with the output saved as uncompressed WebM files. The head camera was positioned centrally within the head, aligned with the participant's head orientation, while the gaze camera was co-located but adjusted according to the instantaneous gaze direction.

#### 4.3. VR motion platform

Due to the significant impact of the gait cycle on head position, a VR motion platform allowing unrestricted vertical hip movement was utilized, enabling participants to move freely within the virtual environment. Specifically, we used the Virtualizer ELITE 2 © [5], an advanced omnidirectional VR treadmill designed to enhance immersive virtual reality experiences. It features an actively powered motion platform that supports user gait through dynamic adjustments, significantly reducing physical effort and enabling to capture the influence gait has on head position and retinal optic flow. The system is equipped with six optical motion sensors operating at a sample rate of 1000 Hz which are located in the base of the platform, that track the speed and direction of the feet. In addition, an optical rotation sensor for precise hip orientation tracking and an optical height sensor to monitor user hip height are used in the device to accurately define the forward direction of motion within the virtual environment. These sensors enable detailed and accurate capture of the participants' movement direction, speed, and user orientation, supporting a variety of actions such as walking, running, crouching, and backward movements.

#### 4.4. HTC Vive Pro Eye VR headset

The HTC Vive Pro Eye VR headset [7] was employed for its advanced eye-tracking capabilities, providing detailed and accurate data essential for studying user interactions within virtual environments. This headset features dual OLED displays with a combined resolution of  $2880 \times 1600$  pixels ( $1440 \times 1600$  pixels per eye) and a 90 Hz refresh rate, delivering a 110-degree field of view. The integrated Tobii eye-tracking technology operates at a frequency of 120 Hz, with an accuracy between  $0.5^\circ$  and  $1.1^\circ$  in the center  $20^\circ$  field of view, and includes a 5-point calibration system. The eye tracker captures gaze origin, gaze direction, pupil position, pupil size, and eye openness, facilitating precise user attention mapping. However, for this dataset only gaze direction was recorded.

## Limitations

Not applicable.

## Ethics Statement

This research involved human participants and was conducted in full accordance with the principles outlined in the Declaration of Helsinki. Informed consent was obtained from all participants prior to their inclusion in the study. Consistent with the requirements of the Deutsche Forschungsgemeinschaft (DFG) [8], the study design incorporated a research ethics self-assessment [9], thereby obviating the need for formal ethical committee approval.

## CRediT Author Statement

**Alexander Kreß:** Conceptualization, Methodology, Software, Writing - Original Draft, Visualization, **Frank Bremmer:** Supervision, Funding acquisition, Conceptualization, Writing - Review & Editing. **Markus Lappe:** Supervision, Conceptualization, Writing - Review & Editing.

## Data Availability

[Comprehensive VR Dataset for Machine Learning: Head- and Eye-Centred Video and Positional Data \(Original data\)](#) (The Adaptive Mind Datahub).

## Acknowledgments

This work was supported by German Research Foundation ([Deutsche Forschungsgemeinschaft](#), DFG) CRC/TRR 135 - project number [222641018](#), and Hessian Ministry of Higher Education, Research, Science and the Arts - HMWK (Cluster project The Adaptive Mind and LOEWE - Exploration, project number LOEWE/5/A005/519/06/00.006(0007)/E39).

I also thank Anneke Below for helping with data acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] P.J. Mineault, S. Bakhtiari, B.A. Richards, C.C. Pack, Your head is there to move you around: goal-driven models of the primate dorsal pathway, *bioRxiv* (2021) 451701, doi:[10.1101/2021.07.09.451701](#).
- [2] Unity Technologies (2024), Unity Game Engine. <https://unity.com/en/products/unity-engine>.
- [3] Unity Technologies (2024), Unity Asset Store. <https://assetstore.unity.com/>.
- [4] Unity Technologies (2024), Terms of Service. <https://unity.com/de/legal/as-terms>.
- [5] Cyberith GmbH (2024). Virtualizer Elite 2 Specs. Retrieved from <https://www.cyberith.com/virtualizer-elite/on> 1 August 2024.
- [6] I. Schütz, K. Fiehler, Eye tracking in virtual reality: Vive pro eye spatial accuracy, and calibration reliability, *J. Eye Movem. Res.* 15 (3) (2022) 3, doi:[10.16910/jemr.15.3.3](#).
- [7] HTC Corporation (2024), VIVE Pro Eye Specs. Retrieved from <https://www.vive.com/sea/product/vive-pro-eye/specs/> on 1 August 2024.



- [8] Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). FAQs - humanities and social sciences: proposal funding process. Retrieved from <https://www.dfg.de/en/research-funding/proposal-funding-process/faq/humanities-social-sciences> on 25 October, 2024.
- [9] RatSWD [Rat für Sozial- und Wirtschaftsdaten], Forschungsethische grundsätze und prüfverfahren in den sozial- und wirtschaftswissenschaften, RatSWD Output 9 (5) (2017) 25–27, doi:[10.17620/02671.1](https://doi.org/10.17620/02671.1).