HMDPose: A large-scale trinocular IR Augmented Reality Glasses Pose Dataset

Ahmet Firintepe
BMW Group Research, New Technology, Innovations
Garching (Munich), Germany
TU Kaiserslautern
Germany
Ahmet.Firintepe@bmwgroup.com

Alain Pagani
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
Alain.Pagani@dfki.de

Didier Stricker
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany
TU Kaiserslautern
Germany
Didier.Stricker@dfki.de

ABSTRACT

Augmented Reality Glasses usually implement an Inside-Out tracking. In case of a driving scenario or glasses with less computation capabilities, an Outside-In tracking approach is required. However, to the best of our knowledge, no public datasets exist that collects images of users wearing AR glasses.

To address this problem, we present HMDPose, an infrared trinocular dataset of four different AR Head-mounted displays captured in a car. It contains sequences of 14 subjects captured by three different cameras running at 60 FPS each, adding up to more than 3,000,000 labeled images in total. We provide a ground truth 6DoF-pose, captured by a submillimeter accurate marker-based tracker. We make HMDPose publicly available for non-profit, academic use and non-commercial benchmarking on ags.cs.uni-kl.de/datasets/hmdpose/.

CCS CONCEPTS

• **Computing methodologies → Tracking**.

KEYWORDS

AR Glasses, Tracking, Deep Learning, Object Pose Estimation, Dataset

ACM Reference Format:
Ahmet Firintepe, Alain Pagani, and Didier Stricker. 2020. HMDPose: A large-scale trinocular IR Augmented Reality Glasses Pose Dataset. In *26th ACM Symposium on Virtual Reality Software and Technology (VRST '20), November 1–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3385956.3422121

1 INTRODUCTION

Augmented Reality has been continuously developing in industry and research over decades and has reached commercial maturity. Almost all high-end smartphones are delivered with AR applications and AR-dedicated hardware. At the same time, AR glasses development flourished with more companies developing their own AR-enabled devices. The deployment of AR glasses inside the car enable an abundance of driving relevant use cases. Visualizations for the driver like an AR navigation in front of the eyes is one useful example. In this case, tracking based on cameras built inside the glasses is difficult, as the cameras observe parts of the static car interior as well as from the dynamic outside world. Thus, tracking with additional cameras deployed inside the car can help. Outside-In head tracking is popular [6–9], but cannot replace the glasses pose as the head and glasses relation can vary per person and can change during driving. Head tracking relies on crucial facial features like the eyes [5, 11] which are covered by the glasses. Further, in-car AR glasses tracking adds the challenge of tracking functionality in adverse and changing lighting conditions. Infrared cameras capture images less prone the lighting changes and ensures that tracking can work at all times. To enable development of algorithms for this purpose, public datasets are required, which are not available yet. To solve this issue, we propose HMDPose, a very large-scale dataset containing more than 3,000,000 IR images of car passengers with various types of AR glasses. It provides images from three different camera angles with 60 FPS each, including sub-millimeter accurate ground truth 6DoF-pose annotations (Figure 1).

Figure 1: Example triple of the Mini Augmented Vision glasses as one of the four glasses in the recordings.

2 HMDPOSE DATASET

Hardware Setup. We deployed a marker-based solution for acquiring the 6DoF-Pose as ground truth and captured the images with three infrared cameras (Figure 2).

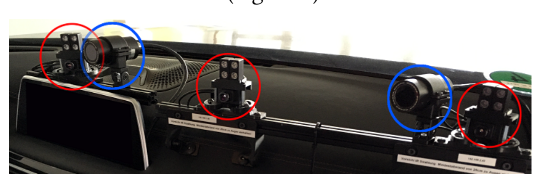

Figure 2: The hardware setup on the car dashboard for the recording. There are three IR-cameras with additional flashers on top for the image recording (red). Two marker-base cameras acquire the ground truth poses (blue).

Three IR-cameras record the images with the resolution of 1280×752 pixel each for the dataset. Each camera records with 60 FPS. In addition, the marker-based cameras track the markers on the glasses, recording ground truth 6DoF-poses in 150 Hz.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
VRST '20, November 1–4, 2020, Virtual Event, Canada
© 2020 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7619-8/20/11.
https://doi.org/10.1145/3385956.3422121

Calibration & Synchronisation. Calibration to one common coordinate frame is required for accurate and synchronous pose recording. Through calibration, we can obtain the extrinsics of each camera system in relation to the set coordinate frame. This is done with a static checkerboard inside the car, where given points on the checkerboard and their relation to the common coordinate frame are known. By localizing the same points in camera coordinates, we can calibrate the cameras to the coordinate frame. Our coordinate frame is set to the middle of the cars frontal vehicle axle. The x axis points towards the back of the car, z points upwards and y to the right from the drivers perspective. All cameras and poses are given in respect to this coordinate system. All coordinate systems in place are visualized in Figure 3. The center of the local coordinate systems of the individual glasses are defined as the closest marker to the center of the respective glasses. In addition, a synchronized recording of the frames and poses is performed.

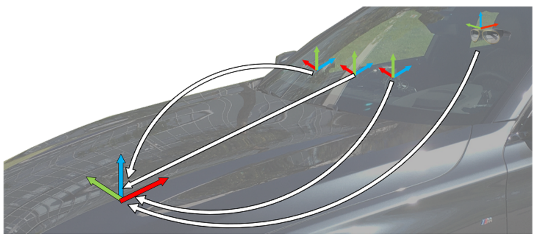

Figure 3: All coordinate systems and the their transformations. The corresponding x,y and z axis are colored red, green and blue, respectively. The common coordinate frame in the front of the car, the three IR cameras pointing towards the passenger and the glasses pose orientation similar to the common coordinate frame are pointed out.

Acquisition & Dataset analysis. We acquire the 6DoF-pose of four different glasses in the reference coordinate frame and the three IR images synchronously (Figure 1 and 4). As a conventional looking glasses, we selected the the North Focals Gen 1 [2]. Our representative for activity related glasses are the Everysight Raptor glasses [4]. All-day-use HMDs are represented by the Mini Augmented Vision glasses [1]. We chose Microsoft Hololens 1 [3] as an industrial, bigger, more powerful AR HMD. The recording was done in a driver simulation. We prerecorded a 1km long route with a 360° camera strapped on the given prototype. We showed different angles of this material on three screens simultaneously, which where placed around the car to trigger natural head movements (Figure 5). With each participant, we recorded four sequences for each HMD on the recorded, five minute long route. On this route, multiple points of interests (POIs) were defined, where the participants were instructed to look at. Between the instructed parts, the subjects were free to either follow the road or look around.

(a) (b) (c)

Figure 4: Example, cropped single IR images of the three other glasses in the recordings: (a) North Focals Generation 1, (b) Everysight Raptor, (c) Hololens 1.

Figure 5: The simulation setup, playing a route which was prerecorded with 360° camera and edited accordingly.

The dataset has a 35.7 % female and 64.3 % male ratio. The mean age is 33.5 years with a standard deviation of 10.73. Figure 6 shows the data distribution in orientation per axis. The orientation strongly suggests the frontal orientation during driving, visible on all axis.

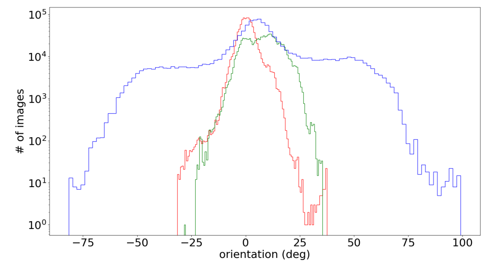

Figure 6: Glasses orientation distribution around all three angles in degrees. red: roll, green: pitch, blue: yaw.

Postprocessing. As we record the frames of the different glasses with markers on them to obtain an accurate ground truth pose, the markers are visible on the images after recording. In a postprocessing step, we remove the markers from the glasses (Figure 7).

(a) (b)

Figure 7: Example of the Mini Augmented Vision glasses before and after postprocessing. (a) Before postprocessing, where markers are circled red and reflections are highlighted in blue. (b) After postprocessing without markers but visible reflections.

We do this by utilizing the known marker positions of the marker-based system and combine it with the detection of bright circles. If there are bright circles detected on an image matching one of the given marker positions, the marker is being removed. We use the OpenCV inpainting method based on the method of Telea [10].

3 CONCLUSION

In this paper, we introduced HMDPose, a new large-scale AR glasses pose dataset. We captured infrared images from three different positions of 14 subjects, wearing four different AR glasses. The recording resulted in more than 3,000,000 infrared images enclosed with ground truth glasses pose. The labels are recorded with a sub-millimeter accurate marker-based tracking system. Future work will focus on training and comparing glasses pose estimation approaches.

REFERENCES

[1] 2015. MINI Augmented Vision. https://www.press.bmwgroup.com/global/article/detail/T0212042EN/mini-augmented-vision:-a-revolutionary-display-concept-offering-enhanced-comfort-and-safety-exclusive-prototype-of-augmented-reality-eyewear-underlines-the-innovative-flair-and-creativity-of-the-mini-brand?language=en

[2] 2020. Explore Focals - North. https://www.bynorth.com/focals

[3] 2020. HoloLens (1st gen) hardware. https://docs.microsoft.com/en-us/hololens/hololens1-hardware

[4] 2020. Raptor – Everysight. https://everysight.com/product/raptor/

[5] J. M. D. Barros, B. Mirbach, F. Garcia, K. Varanasi, and D. Stricker. 2018. Fusion of Keypoint Tracking and Facial Landmark Detection for Real-Time Head Pose Estimation. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2028–2037. https://doi.org/10.1109/WACV.2018.00224

[6] Guido Borghi, Marco Venturelli, Roberto Vezzani, and Rita Cucchiara. 2017. POSEidon: Face-From-Depth for Driver Pose Estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[7] Ahmet Firintepe, Selim Mohamed, Alain Pagani, and Didier Stricker. 2020. The More, the Merrier? A Study on In-Car IR-based Head Pose Estimation. In *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE.

[8] Anke Schwarz, Monica Haurilet, Manuel Martinez, and Rainer Stiefelhagen. 2017. DriveAHead-a large-scale driver head pose dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1–10.

[9] Mohamed Selim, Ahmet Firintepe, Alain Pagani, and Didier Stricker. 2020. AutoPOSE: Large-Scale Automotive Driver Head Pose and Gaze Dataset with Deep Head Pose Baseline. In *International Conference on Computer Vision Theory and Applications (VISAPP)*. SCITEPRESS Digital Library.

[10] Alexandru Telea. 2004. An Image Inpainting Technique Based on the Fast Marching Method. *Journal of Graphics Tools* 9, 1 (2004), 23–34.

[11] T. Vatahska, M. Bennewitz, and S. Behnke. 2007. Feature-based head pose estimation from images. In *2007 7th IEEE-RAS International Conference on Humanoid Robots*. 330–335.