Behavior Research Methods
https://doi.org/10.3758/s13428-022-01941-1

Comparison of visual SLAM and IMU in tracking head movement outdoors

Ayush Kumar[1] · Shrinivas Pundlik[1] · Eli Peli[1] · Gang Luo[1]

Accepted: 22 July 2022

Abstract

Tracking head movement in outdoor activities is more challenging than in controlled indoor lab environments. Large-magnitude head scanning is common under natural conditions. Compensatory gaze (head and eye) scanning while walking may be critical for people with visual field loss. We compared the accuracy of two outdoor head tracking methods: differential inertial measurement units (IMU) and simultaneous localization and mapping (SLAM). At a fixed location experiment, a gaze aiming test showed that SLAM outperforms IMU in terms of error (IMU: 9.6°, SLAM: 4.47°). In an urban street walking experiment conducted with five patients with hemifield loss, the IMU drift, quantified by root-mean-square deviation, was as high as 68.1°, while the drift of SLAM was only 5.3°. However, the SLAM method suffered from data loss due to tracking failure (~10% overall, and ~ 18% when crossing streets). Our results show that the SLAM and IMU methods have complementary properties. Because of no data gaps, the differential IMU method may be desirable as compared to SLAM in settings where the signal drift can be removed in post-processing and small gaze estimation errors can be tolerated.

Keywords Head scanning · IMU · Outdoor navigation · Visual SLAM

People's mobility is affected by where they look, which in turn requires both head and eye movements. While saccade amplitude can be quite large, the distribution of saccade amplitudes in real-world circumstances are largely below 15° (Bahill, Adler, & Stark, 1975), making head tracking a necessity to understand real-world gaze (eye + head) (Bahill, Adler, & Stark, 1975; Einhäuser et al., 2007; Rothkopf & Pelz, 2004).

To compensate for their visual field loss, visually impaired people while walking would presumably need to scan the environment more than the normally sighted. Numerous studies have analyzed the eye and head movements for people with normal vision and with vision loss in controlled laboratory setups (Barabas et al., 2004; Bowers, Ananyev, Mandel, Goldstein, & Peli, 2014; Cesqui, de Langenberg, Lacquaniti, & d'Avella, 2013; Essig et al., 2012; Grip, Jull, & Treleaven, 2009; Imai, Moore, Raphan, & Cohen, 2001; Kugler, Huppert, Schneider, & Brandt, 2014; Lin et al., 2007; MacDougall & Moore, 2005; Proudlock, Shekhar, & Gottlob, 2003). Tracking head movements while walking, particularly in an unconstrained outdoor environment, is challenging because usually it is not feasible to set up an external sensor system in large outdoor spaces. Therefore, only limited reports of real-world outdoor gaze behavior of visually impaired people are available, most of which were acquired by manually analyzing scene images captured with head-mounted video scene cameras (Geruschat, Hassan, Turano, Quigley, & Congdon, 2006; Hassan, Geruschat, & Turano, 2005; Marius't Hart & Einhäuser, 2012).

Typical head-mounted eye trackers provide information about eye-in-head position. Head movements should be measured with an additional sensor. Integration of an inertial measurement unit (IMU) with the mobile eye tracker is a potential solution (Tomasi, Pundlik, Bowers, Peli, & Luo, 2016). A limitation of the IMU-based approaches is that IMU sensors are susceptible to various environmental interference resulting in tracking errors, even though the state-of-the-art IMUs fuse multiple sensors using sophisticated algorithms. Given this limitation, single IMU sensor-based approaches were mostly tested in controlled environments over a short duration (Linnéa Larsson, Schwaller, Holmqvist, Nyström, & Stridh, 2014; L. Larsson, Schwaller, Nystrom,

✉ Gang Luo
gang_luo@meei.harvard.edu

[1] Schepens Eye Research Institute of Mass Eye & Ear, Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Published online: 11 August 2022

& Stridh, 2016; Stoll, 2015; Wang, Zeng, & Liu, 2016) and may not be suitable for long-duration outdoor walking studies. To counter the impact of environmental interference, Tomasi et al. (2016) proposed a differential IMU method to track head movement in an outdoor walking scenario, with one sensor attached to the head and one to the waist. Head orientation (yaw, pitch, and roll angles) were measured relative to the body by computing the difference between the two IMUs' signals. The differential IMU sensors approach was designed under the assumption that environmental factors would affect both sensors similarly, and therefore their impacts can be mitigated in the differential output. Tomasi et al. showed that the differential IMU method indeed improved the accuracy, but there were residual errors due to signal drift (which is discuss further below).

An alternative to the IMU-based approach is to use video from the head-mounted scene camera to compute the head pose by means of visual SLAM (simultaneous localization and mapping) algorithm. Visual SLAM processing involves using local image-based features matching among video frames at different time instants (Mur-Artal, Montiel, & Tardos, 2015; Mur-Artal & Tardós, 2017). Visual SLAM techniques have their own limitations, in particular tracking loss, and accumulating errors while building and localizing trajectory maps. Tracking losses are typically caused by large and fast changes in the camera position while walking. Lack of overlap between consecutive frames makes it difficult to extract sufficient features needed for building the trajectory map. Localization errors due to erroneous calculation of location accumulate in the absence of feature matching.

We compared the head tracking accuracy of the two head tracking methods based on data acquired during the same walking events: differential IMU sensors (Tomasi et al.) and an open-source state-of-the-art visual SLAM solution, OpenVSLAM. We evaluated the drift in the IMU and data loss in the SLAM during naturalistic outdoor walking scenarios. The goal of this study was to determine the pros and cons of each method.

Method

The data used in this work was obtained using the outdoor gaze tracking system described in Tomasi et al. (2016). The eye-tracking system from Positive Science (Positive Science, New York City, NY; 2013) included a MacBook Air laptop running proprietary software. The two commercial IMUs from VectorNav (VectorNav, Dallas, TX) were connected to a lightweight ASUS Eee PC notebook with an Intel Atom N455 processor that performed data logging. The two systems were used for simultaneous head tracking in naturalistic urban walking scenarios. The mobile eye tracker includes a miniature analog scene camera (30 Hz, 640 × 480 resolution, 60° field of view) to capture the front view. According to the manufacturer, the eye-tracker error is about 0.5°. In the original study (Tomasi et al., 2016) the scene videos were used only to visualize the eye movements by overlaying traces of eye movements on the scene video frames but were not included in the analyses. Here we processed the scene videos using monocular SLAM to compute the head movement.

IMUs

Angular values of head orientation were obtained using the differential IMU approach, originally described by Tomasi et al. (2016). Each IMU included a three-axis accelerometer, a three-axis gyroscope, and a three-axis magnetometer. The data from these three sensors were fused internally by a microprocessor running VectorNav's proprietary algorithms to output orientation angles (yaw, pitch, and roll). According to the datasheet of the manufacturer, the heading (root-mean-square, RMS) error is 2° under proper calibration, and a magnetic environment (which we do not have in our experiments) was observed during the IMU experiment, whereas a pitch/roll error of 1° under dynamic conditions was recorded, which is more relevant to our application.

SLAM

Simultaneous localization and mapping is used for estimating a 3D structural map of previously unknown environment by building a trajectory map. There are various modalities of SLAM. We used OpenVSLAM, a monocular modality of visual SLAM (also known as vSLAM) (Sumikura et al., 2019a, b). OpenVSLAM was selected based on its robustness and popularity. OpenVSLAM is a derivative of ORB-SLAM2 (Mur-Artal & Tardós, 2017), and can potentially overcome some of the limitations of ORB-SLAM2. Unlike other visual SLAM techniques, OpenVSLAM can store and load map databases for further localizations. Localization based on a prebuilt map improves the absolute trajectory error as well as reduces the tracking time. These techniques are usually faster in detecting key feature points (Rosten, Porter, & Drummond, 2008). Sample tracked key features for the image in Fig. 1a are shown in yellow. The OpenVSLAM system, in general, is compatible with various types of inputs such as monocular, stereo, or RGBD videos, and we have used its monocular version in our experiments.

Estimating head movements using visual SLAM involves three stages: (a) intrinsic camera calibration, (b) vSLAM execution, and (c) angle transformation. The first stage includes calculation of a configuration file to find out camera parameters as well as the optical distortion parameters needed to be fed into the vSLAM pipeline. We obtained

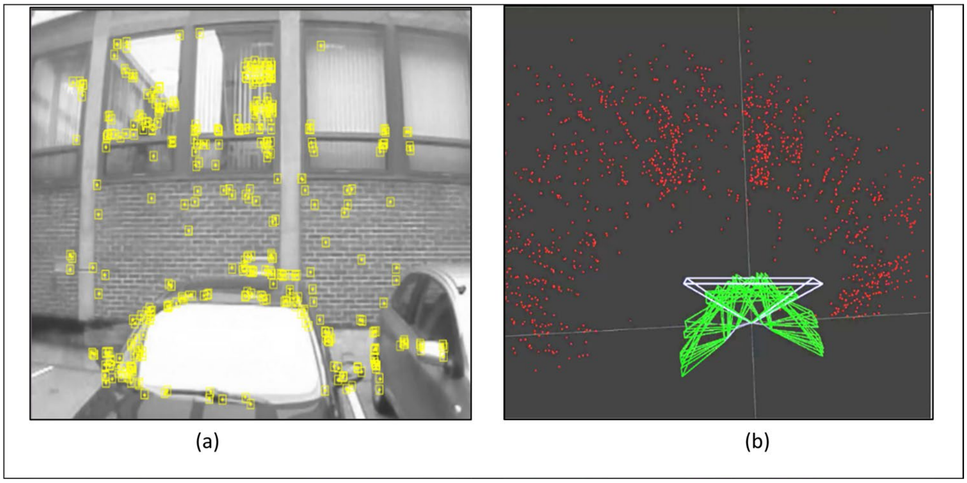


Fig. 1 Using vSLAM for estimating head movement. **a** An image frame with tracked features (yellow point inside a rectangle). **b** Key frames (green pyramids) and mapped 3D feature points cloud (red) for all the image frames used for building the map and localization

optical distortion parameters by applying a geometrical transformation computed offline using the camera calibration on the images of chessboard provided in the OpenCV library (Bradski & Kaehler, 2008).

The second stage of this process starts with the pose estimation pipeline of the SLAM that includes initializing, mapping, and localization. In the second stage, consecutive video frames are fed to the vSLAM system for feature extraction. The initial frame is usually considered as the origin of the map. The consecutive frames are used for extracting and matching (Fig. 1a) features (Rublee, Rabaud, Konolige, & Bradski, 2011). An initial map of the environment is built by matching features extracted from two consecutive frames. The initial map keeps on improving as consecutive frames keep adding new information to it in the form of additional tracked feature points (the tracked feature points are shown in red in the form of 3D point cloud viewed from above in Fig. 1b). Relative motion is calculated between two consecutive frames and the matching feature points between these two frames are then used for triangulation in the 3D world coordinates. In the mapping mode, the map is extended using the triangulated 3D points through the inserted key frames (KF) (green pyramids shown in Fig. 1b are the frames selected for extracting features during SLAM computation) (Mur-Artal & Tardós, 2017). Thus, the mapping and localization operations occur simultaneously, which helps us calculate the pose and camera trajectory (participant) as it moves (Fig. 1b). Limitations of vSLAM such as trajectory drift or scale drift, common in the case of monocular camera input, are resolved via a global optimization process (for further details see Kümmerle, Grisetti, Strasdat, Konolige, & Burgard, 2011).

In the third and last stage, we calculate the angle transformation of the camera pose from the trajectory acquired during localization of the map. The trajectory, output by vSLAM in the form of rotation matrix and camera coordinates, is then transformed into a set of Euler angles (Shoemake, 1994), which gives us the yaw angles measuring the horizontal head movement. (Campos, Elvira, Rodríguez, Montiel, & Tardós, 2021; Mur-Artal et al., 2015; Mur-Artal & Tardós, 2017; Sumikura et al., 2019a, b).

The code snippets of the whole system are available on GitHub[1] which we downloaded for our own use (Sumikura et al., 2019a, b). The makefile available for various component of the SLAM system in the code snipped allowed us to compile the complete OpenVSLAM system ready to be used. Running the OpenVSLAM system with the given command with the desired video and camera parameters pops up two OpenGL-based viewers called PangolinViewers, as shown in Fig. 1. One of the PangolinViewers shows the frames-based view (Fig. 1a) with overlaid featured points on the image frames, while the other shows the mapped feature points in the world coordinate system (Fig. 1b). It can run in both mapping as well as localization modes. The processing time during tracking is approx. 4.14 frames/second on our system (64-bit Windows operating system; Intel® Core™ i7-7700K CPU @ 4.40GHz; 16 GB RAM).

During our experiments, vSLAM suffered loss of tracking at multiple instances, mostly while encountering larger-magnitude head turns. To partially cope with the tracking loss, we restarted the OpenVSLAM system manually following tracking loss, by saving the intermediate trajectory map in the mapping stage. The saved pre-built maps were later used for localization. The complete camera motion trajectory was put together by merging the trajectory segments extracted

[1] https://github.com/OpenVSLAM-Community/openvslam

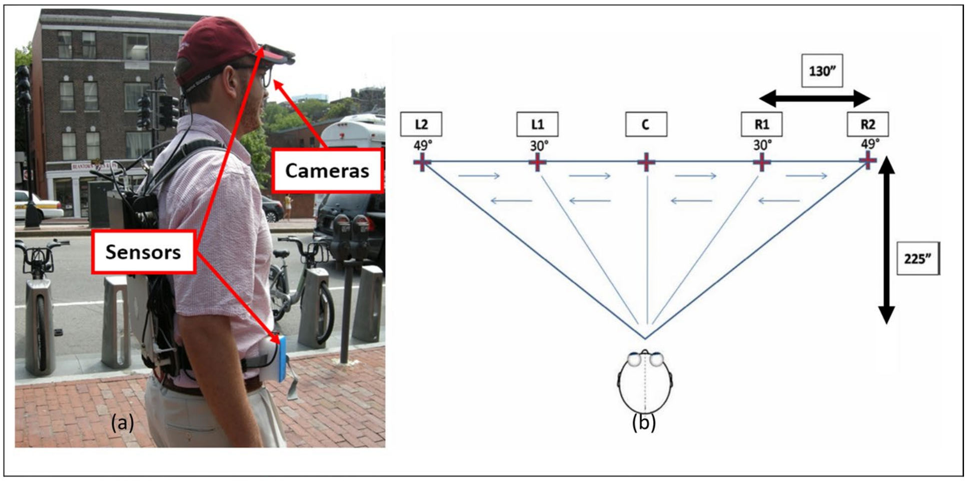


Fig. 2 **a** The spectacles-mounted eye tracker (eye-tracking and scene cameras) and the body and head tracking sensors. **b** An observer stands at a fixed location and looks sequentially and repetitively at each of the fixation points in the following order: C, L1, L2, L1, C, R1, R2, R1, and C

from each of the intermediate maps saved in the mapping stage. The ORB-SLAM2 on which OpenVSLAM is developed has an error with an average ± standard deviation (SD) of 1.9 ± 1.5, going as high as 6.1° on one of the sequences when tested on 11 different sequences of the famous KITTI dataset.

Synchronization

In the outdoor gaze tracking system, the eye tracker and the head tracking system operated independently. The data streams from these systems were synchronized offline after capture, as part of the data processing routine (details described in Tomasi et al., 2016), using custom code developed by us. The synchronized data consisted of global timestamps corresponding to the frames of the scene camera, eye movement coordinates, and the head movement angles. The same timestamps were used for synchronizing the SLAM data.

Testing conditions

The performance of SLAM and IMU sensors for head tracking was compared using data collected in two experiments conducted outdoors: (i) scanning targets at known angular eccentricities in a parking lot, and (ii) walking in a busy urban street. In the scanning experiment, two normally sighted participants performed the scanning tasks from a fixed position, under two conditions: (a) always standing at the fixed position across multiple scanning instances, and (b) walking around the parking lot for a while and then returning back to the fixed position to scan. Under the first condition, the IMU sensor on the waist remained almost stationary, and the second condition added complex movements to both IMU sensors. In the street walking experiment, there were five subjects with left homonymous hemianopia field loss (average ± SD age 56.7 ± 20.6 years) wearing the gaze tracking system who walked on a city street for about 0.6 miles in downtown Boston. Data from five of those participants were selected randomly for this evaluation study. This work reanalyzed the previously collected data, which was approved by the local institutional review board (IRB). Our study followed the tenets of Declaration of Helsinki, and the data were collected under an IRB protocol approved by the Schepens Eye Research Institute/Massachusetts Eye and Ear (MEE) Human Subject Committee. Written informed consent was obtained.

Results

Scanning experiment

In the first experiment we compared the performance of differential IMU-based and SLAM-based methods in terms of horizontal gaze angles relative to known ground truth. Eye-movement angles, calculated by the eye tracker, were summed with the horizontal angles measured from each of the two head tracking methods (IMU and SLAM) to obtain the gaze angles. Details of the experimental set up can be found in Tomasi et al. (2016). Briefly, participants wearing eye tracker spectacles and head tracking IMU sensors (Fig. 2a) sequentially fixated pillars on the wall of the parking lot while standing at a fixed location, as shown in Fig. 2b. The pillars' position served as the ground truth with eccentricities of ±30° and ±49°, respectively, from the central pillar (0°). We mark angles to the left of the central pillar (C) as negative. A single gaze-scanning sequence consisted of gazing at targets in the following sequence: C, L1, L2, L1, C, R1, R2, R1, and C. This experiment involved two conditions: (a) standing still, where the participants stood at the fixed location and repeatedly scanned the targets, and (b) intervening walks condition, where the participant walked around in the parking lot with multiple body turns between successive instances of target scanning from the same spot.

Eye-movement data and IMU signals were logged in laptop computers in the participants' backpacks. Post-processing combined eye movement angles (relative to head) and the head orientation from the IMU signals. The videos from the scene camera were processed using SLAM to obtain the camera pose, which corresponded to the head orientation. Calibration of the eye tracker, IMU system reset, and synchronization steps between eye tracker and the IMU system were performed at the start and end of the recording session for each participant, including the reset events in between.

Horizontal head movement based on the two methods for one subject in the standing still condition is shown in Fig. 3a, and the combined gaze movement traces are shown in Fig. 3b. The spikes in the plot of gaze angles in Fig. 3b are due to artifacts in eye movement measures (blinks). Prominent decay of the IMU signal is visible in the head and gaze estimates from the IMU. The solid black oval labeled (A) in Fig. 3a and b marked one such decay for fixation point C, which is also visible for other fixation points. Only minor variations can be seen in the SLAM signal during these instances.

The second condition (intervening walk condition) involved various head movements, complete body turns and lateral displacements between consecutive scanning instances to determine the effect of users' movements on the IMU drift, and consequently on the estimated head orientation. Gaze angles output by IMUs for five consecutive scanning instances and during the intervening walks for one of the participants is shown in Fig. 4 (the intervening walks are marked as gray zones such as "X").

The mean absolute gaze estimation errors (MAGEE) were calculated for each fixation location (Fig. 5). MAGEE at a given fixation location is the average absolute error with respect to the ground truth gaze angle over the entire duration of fixation (shown by horizontal green line segments in Fig. 3). The ground truth gaze angles are manually annotated by viewing the video in which the eye-in-head positions were rendered (this allowed us to determine the video frames corresponding to each fixation).

The effect of tracking method (SLAM vs. IMU), fixation location (L2, L1, C, R1, R2), walking condition (standing still vs. intervening walks), and participant (subject #1 and #2) on MAGEE was tested using repeated-measures ANOVA. Significant differences in MAGEE were found between tracking methods, $df = (1,4)$, $F = 276$, $p < 0.001$, fixation locations, $df = (4,16)$, $F = 31$, $p < 0.001$, and the two participants, $df = (1,4)$, $F = 47$, $p = 0.002$. There was no significant difference in MAGEE between the walking conditions. Combining the two walking conditions and both participants, MAGEE was larger for higher eccentricities for both IMU and SLAM methods and was lower for the SLAM method than that for the IMU method at each fixation location (Fig. 5a). The estimated marginal means (95% confidence interval) across all fixation locations for the IMU and SLAM methods were 9.6° (8.7°–10.5°) and 4.5° (3.6°–5.4°), respectively.

Significant interaction effects were found between tracking method and fixation location, $df = (4,16)$, $F = 8.9$, $p = 0.001$, and between tracking method, fixation location, and participants, $df = (4,16)$, $F = 6.7$, $p = 0.002$. Given these interactions, we further analyzed MAGEE separately for SLAM and IMU methods to determine the effect of fixation location and participant. For SLAM, it was found that fixation location had a significant effect on MAGEE, $df = (4,16)$, $F = 16$, $p < 0.001$, i.e., the error was larger at far eccentricities. There was no significant effect of participant ($p = 0.14$) or walking condition ($p = 0.68$) on the MAGEE of the SLAM method.

For the IMU measurements (Fig. 5b), there was a significant effect of participant, $df = (1,4)$, $F = 24$, $p = 0.008$, in addition to fixation location, $df = (4,16)$, $F = 20$, $p < 0.001$. This means that MAGEE for IMU was not only higher at higher eccentricities, it was also significantly higher for participant 1 (11.7°) than participant 2 (7.6°) across all fixation locations. There was also a significant interaction between participant and walking condition factors, $df = (1,4)$, $F = 22$, $p = 0.009$, because of the large difference in MAGEE between the two participants in the standing still condition (participant #1: 12.8°, participant #2: 6.6°). We also measured the standard deviation of head angles without eye movements at different fixation for both IMU and SLAM. At 0 it is (SLAM: 2.77, IMU: 6.78), at −30 it is (SLAM: 7.07, IMU: 3.95), at +30 it is (SLAM: 3.1, IMU: 5.72), at −49 it is (SLAM: 7.61, IMU: 4.28), and at +49 it is (SLAM: 3.05, IMU: 5.83).

To summarize, our analyses found that MAGEE with SLAM was significantly lower than the IMU method. The significant difference between the participants was primarily driven by the difference in the participants' MAGEE for the IMU method. Since the more accurate SLAM data did not show any significant difference between the two participants, this suggests that the between-participant difference in IMU data was not due to participant per se. Based on our observation in the street walking experiment presented below, we postulate that it may be because random factors (likely from environment interference) occurred in the two IMU recording sessions (taken at different days, where the number and positions of cars in the lot had been different). One interesting phenomenon related to IMUs that could be

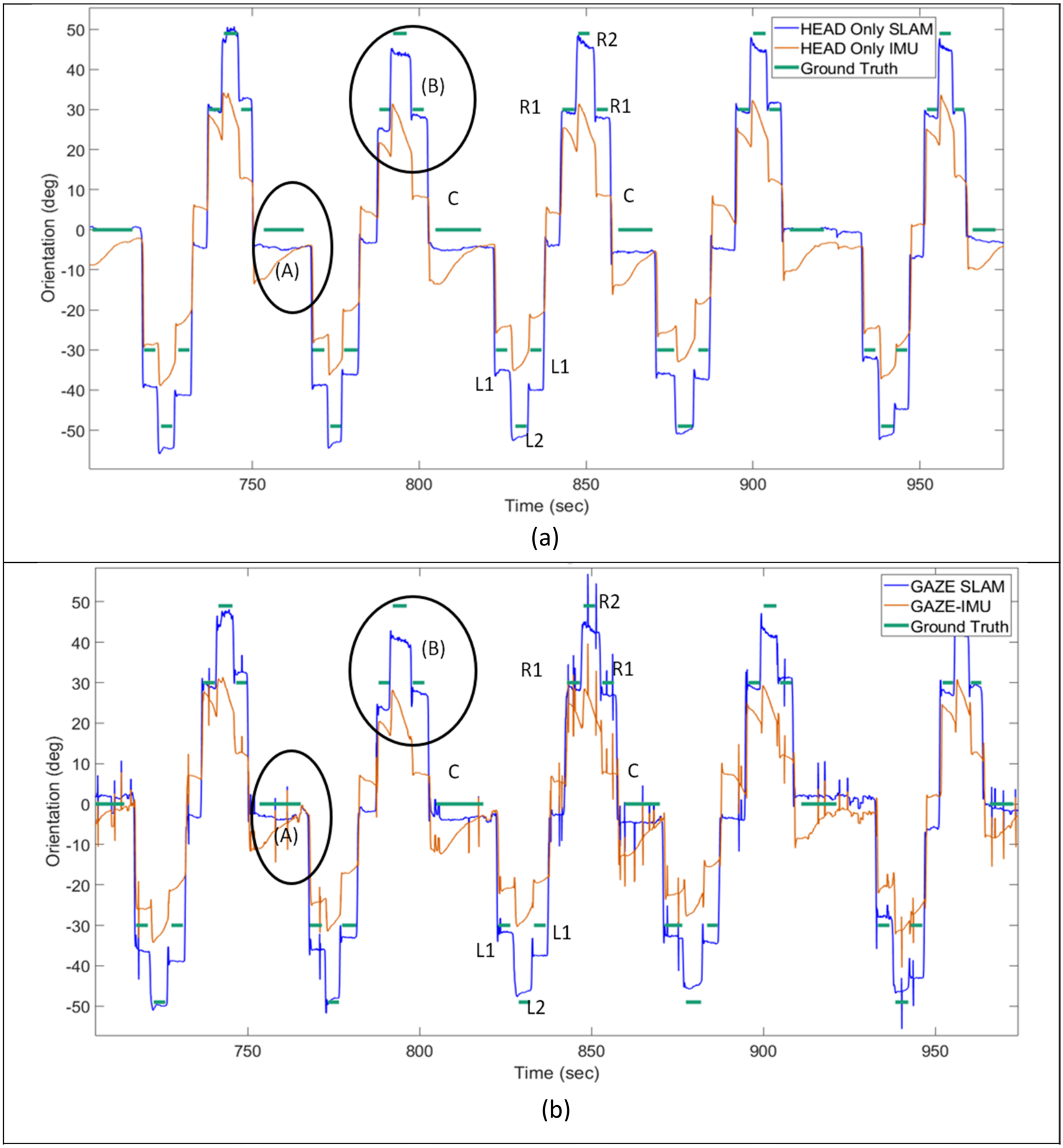


Fig. 3 **a** Horizontal head movement measures using both the SLAM (solid blue line) and IMU (solid orange line), respectively, are shown along with the ground truth fixation angels (solid green line) for one of the participants in the standing still condition in Experiment 1. **b** Horizontal gaze angles, which are combination of head and eye movements, for the same head movement data shown in **a**. The negative values show orientation toward the left side (L1) and (L2) and positive towards the right side (R1) and (R2). Proximity to the green horizontal line segments indicates greater accuracy. The IMU shows a lower gain, and the SLAM shows asymmetry between the left and right response. Some prominent drift variations in IMU can be seen at locations annotated with solid black ovals labeled as (A) and (B) on the plot

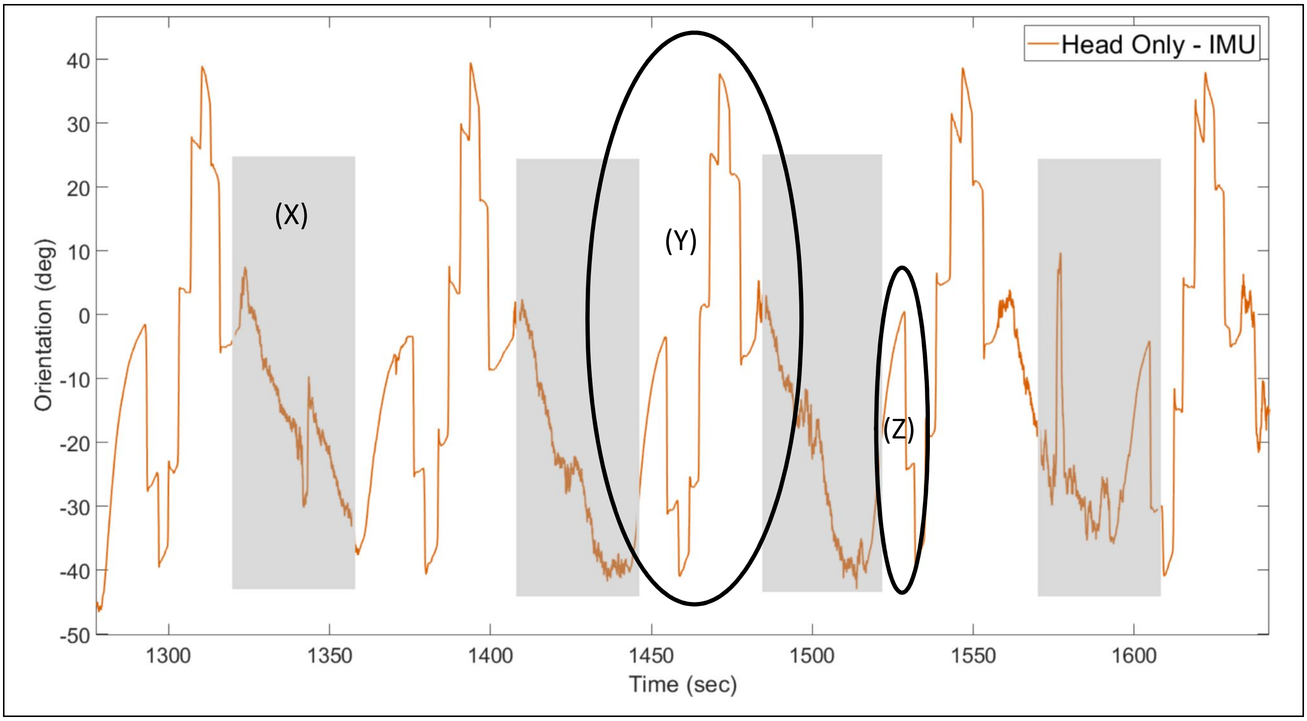


Fig. 4 Horizontal head orientation angle from the IMU in the scanning experiment involved with walk sessions (marked by gray blocks, such as (X), between measurements). After walking, making various body and head turns, the observer returned to the fixed observation location and performed the target scanning task, which is indicated by the pedestal-like patterns of the head movement. There was a large signal drift, marked by solid black ovals (Y) and (Z), when the participant returned to the baseline position to repeat the scanning instance. Here, the oval (Y) marks the full scanning cycle, whereas (Z) marks the initial scan to the left side after returning to the baseline scanning position

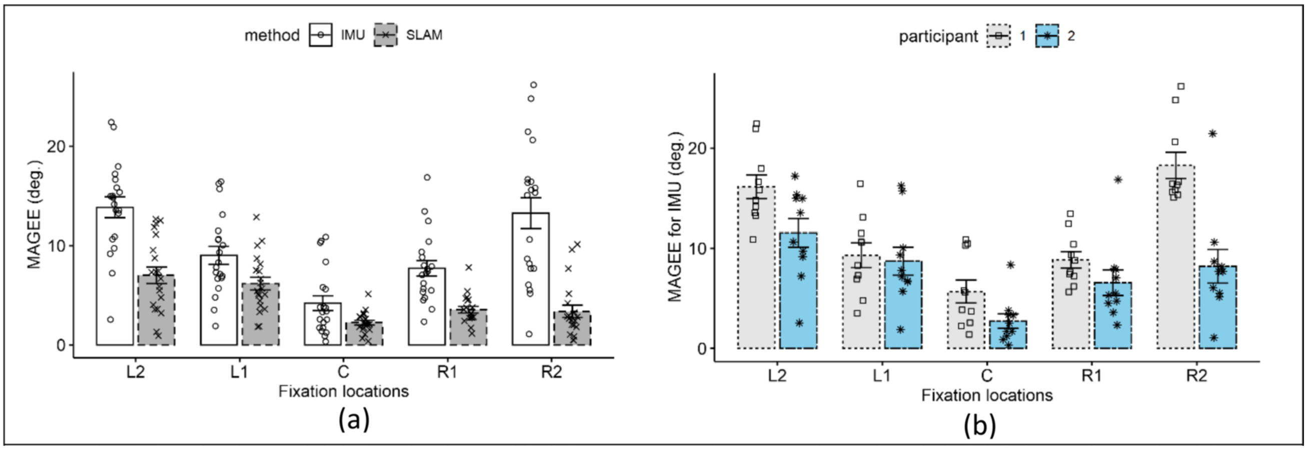


Fig. 5 **a** Mean absolute gaze estimation error (MAGEE) for SLAM and IMU methods at the different fixation locations: L2, L1, C, R1, and R2. Overall, the error was larger for larger eccentricities and was significantly larger for IMU than that of SLAM. **b** MAGEE recorded with IMU method at different fixation locations for two participants. MAGEE for participant #1 was significantly larger than participant #2. Error bars represent standard error of mean

seen was the decay after the intervening walks, as shown in the solid black oval (Z) in Fig. 4. Even after the participant returned to the designated location and started to fixate at the central pillar (C), there was a delay in IMU response, and the estimated orientation did not recover to 0° for an average of about 10 seconds. Such a delay did not occur for other pillars on the right side or left side, because the intervening impact of walking had already faded away. MAGEE for IMU

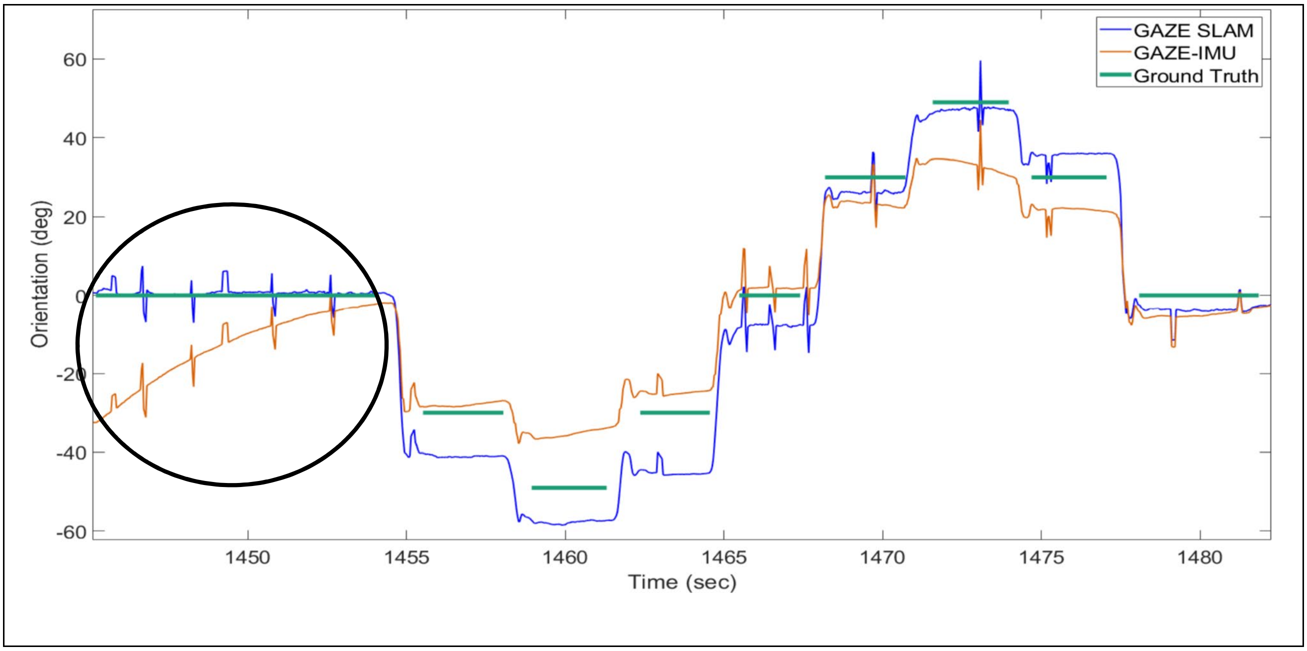


Fig. 6 A time-expanded view of a gaze tracking of one left to right fixation sequence from Fig. 4 (oval-shaped Y) where IMU in orange shows drift after being back from the walk. SLAM is shown in blue, and green horizontal lines represents the ground truth of the fixations

(8.84°) and SLAM (1.97°) differed substantially only for the central fixation location (location C) in the intervening walk condition (when the participant came back to the central position "C" after the walk).

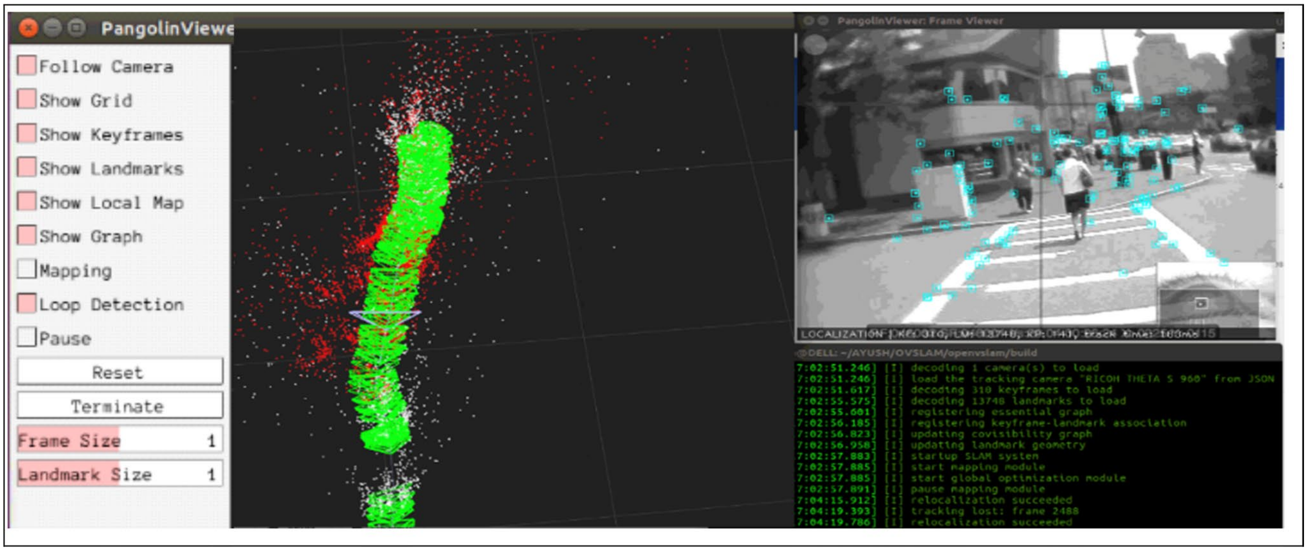
We confirmed the return of participants to the location aligned with "C" by watching the video frame by frame manually, whereas head orientation estimated by the SLAM returned to its position (0°) immediately as shown in one of the sequences in Fig. 6, which is a zoomed version of the plot in Fig. 4 marked with (Y). When SLAM, in blue, is at "C" (0°), at the same time the IMU (in orange) is off by 36°. As can be seen, SLAM is resistant to such impact throughout in both the conditions.

Street walking experiment

In the second experiment we recorded head and eye movements of five patients (all male, average ± SD age 56.7 ± 20.6 years) with homonymous hemianopia walking in a busy urban street. The eye and head tracker and the calibration procedures were the same as those used in the scanning experiment. Periodic eye calibrations and IMU system resets were performed at intermediate waypoints during the street walk, to reduce the impacts of tracking failures due to various causes. All the walking sessions were done on a nearly straight segment of Cambridge Street in downtown Boston (Fig. 7), which includes street crossings with and without pedestrian signals. The participants walked on both the eastbound and westbound directions of the same segment, so the traffic and buildings, etc., appeared on both their left and right sides. The total walking distance was 0.6 miles. We have excluded data related to the calibration stops from our analysis. Therefore, each walking sequence was analyzed in multiple sessions, depending on the number of calibrations performed during the session.

Figure 8a shows the horizontal head orientation angle plots for both IMUs (in orange) and SLAM (in blue) for one of the sessions of a walking sequence (walking between two successive calibrations is counted as one session) of a patient. In Fig. 8a, one can clearly see the effect of drift in the IMU head measurements (in orange), as compared to the SLAM data which has maintained a straighter path. We can see the drift signal better after both head movement signals were filtered using a Butterworth low-pass filter of order 1 (Fig. 8b). This processing is valid because the walking route in each session was largely straight without turns and average head position is assumed to remain zero. Following low-pass filtering of the head horizontal signal, the amount of drift was quantified using root-mean-square deviation (RMSD) with respect to 0° which was the expected mean head orientation during the walk. For this walking session, the RMSD for IMU was 35.4° while for SLAM it was 3.2°. However, other walking sessions of the same participant showed much lower drift (Figs. 9 and 10) in both IMUs and SLAM.

We analyzed the data from the five left hemianopia patients. Each of them walked for 19 minutes on average that resulted in total of 17 sessions (between calibrations).



☑ Follow Camera
☑ Show Grid
☑ Show Keyframes
☑ Show Landmarks
☑ Show Local Map
☑ Show Graph
□ Mapping
☑ Loop Detection
□ Pause
Frame Size 1
Landmark Size 1

```
@DELL: ~/AYUSH/OVSLAM/openvslam/build
7:02:51.246] [I] decoding 1 camera(s) to load
7:02:51.246] [I] load the tracking camera "RICOH THETA S 960" from JSON
7:02:51.617] [I] decoding 310 keyframes to load
7:02:55.575] [I] decoding 13748 landmarks to load
7:02:55.601] [I] registering essential graph
7:02:56.185] [I] registering keyframe-landmark association
7:02:56.823] [I] updating covisibility graph
7:02:56.958] [I] updating landmark geometry
7:02:57.883] [I] startup SLAM system
7:02:57.885] [I] start mapping module
7:02:57.885] [I] start global optimization module
7:02:57.891] [I] pause mapping module
7:04:15.912] [I] relocalization succeeded
7:04:19.393] [I] tracking lost: frame 2488
7:04:19.786] [I] relocalization succeeded
```

Fig. 7 Patient walking across a road with pedestrian signals towards east of Cambridge Street (inset) and the trajectory being located using SLAM (green). The trajectory image shows all the key frames (green pyramids, as shown in Fig. 1) and mapped 3D feature points (red and white) cloud for all the image frames used for building the map in world coordinate, where red points are the active features for the current key frame. The tracked features are shown as dots within squares (blue) for the frame in the inset image

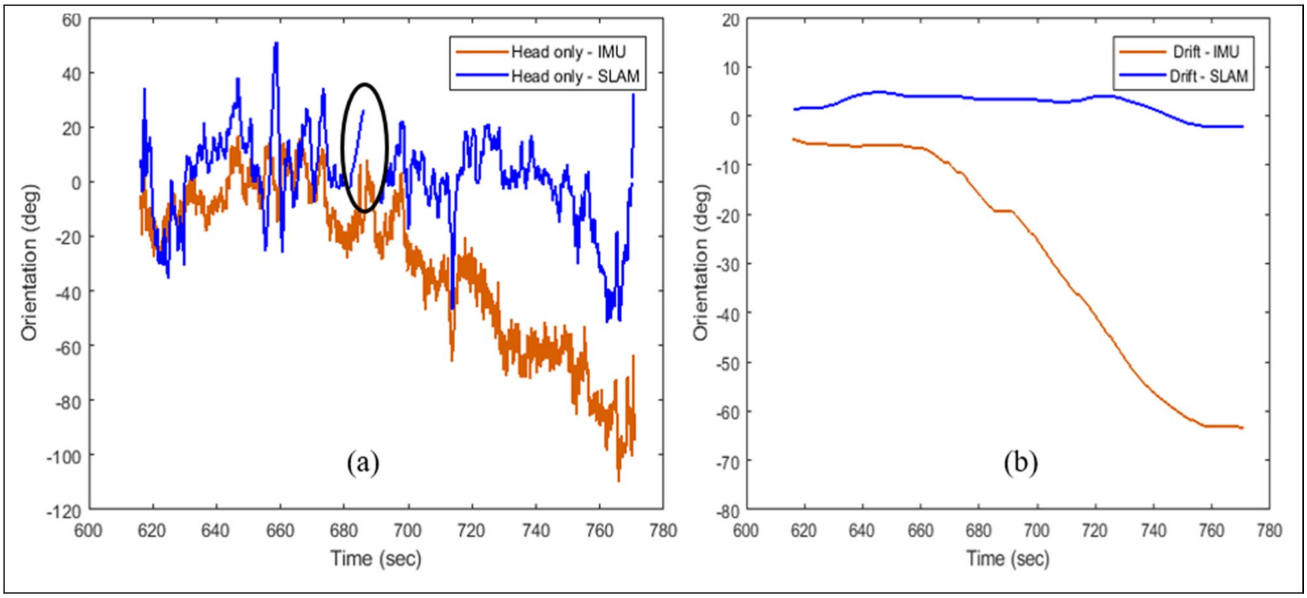


Fig. 8 **a** Horizontal head orientation for IMU (orange) and SLAM (blue) where large drift can be seen for the IMU. Black oval on the plot shows an instance of tracking loss in SLAM (discontinuous blue line). **b** Drift signal extracted using low-pass filter version of the IMU in both SLAM and IMU

The average RMSD with respect to 0° reference for IMU and SLAM was 16.2° and 2.4°, respectively.

The plot in Fig. 11 shows the variability in the drift experienced during the recording using IMU sensors (going as high as 68.13° for patient 3, IMU: average ± SD 16.2 ± 18.0, SLAM: average ± SD age 2.3 ± 1.3) compared to much lower RMSD for the SLAM. It should also be noted that IMU drift in some walking sessions are comparable to SLAM. In 10 out of 17 sessions, the IMU drift was higher than the largest SLAM drift. The paired *t*-test for the same session shows a statistically significant difference between RMSD values for IMU and SLAM, $p = 0.005$, $t(16) = 3.21$.

Although SLAM was not affected by the drift problem, it was affected by tracking losses, especially during larger

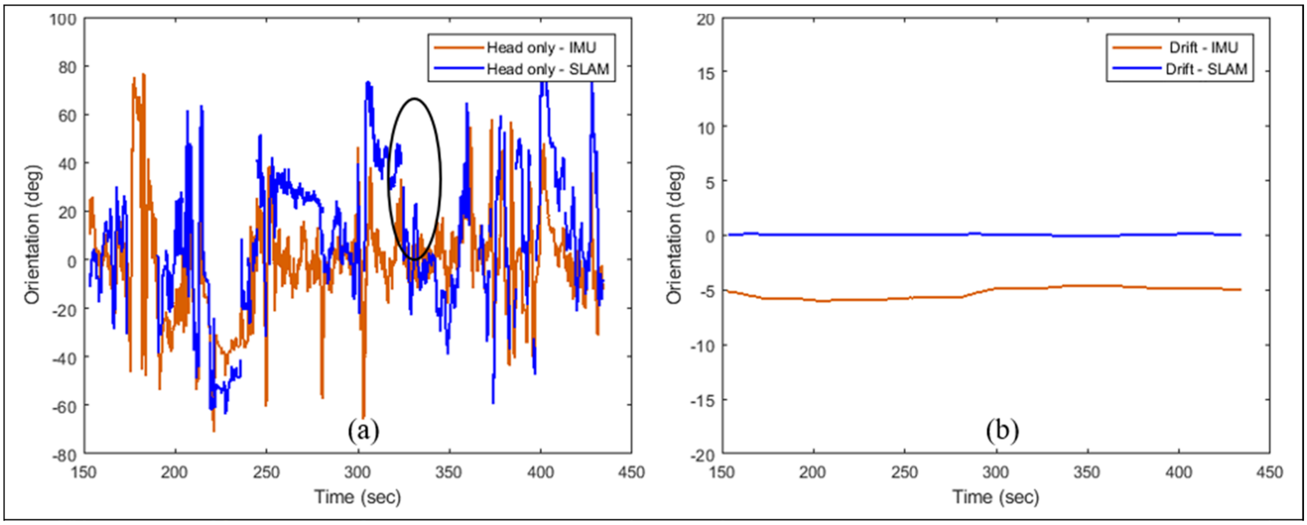


Fig. 9 **a** Horizontal head orientation for IMU (orange) and SLAM (blue) in another (first) segment of the walk. Black oval shows an instance of tracking loss in SLAM (discontinuous blue line). **b** Filtered signal using low-pass filter in both SLAM and IMU, where there is not much drift in either plot

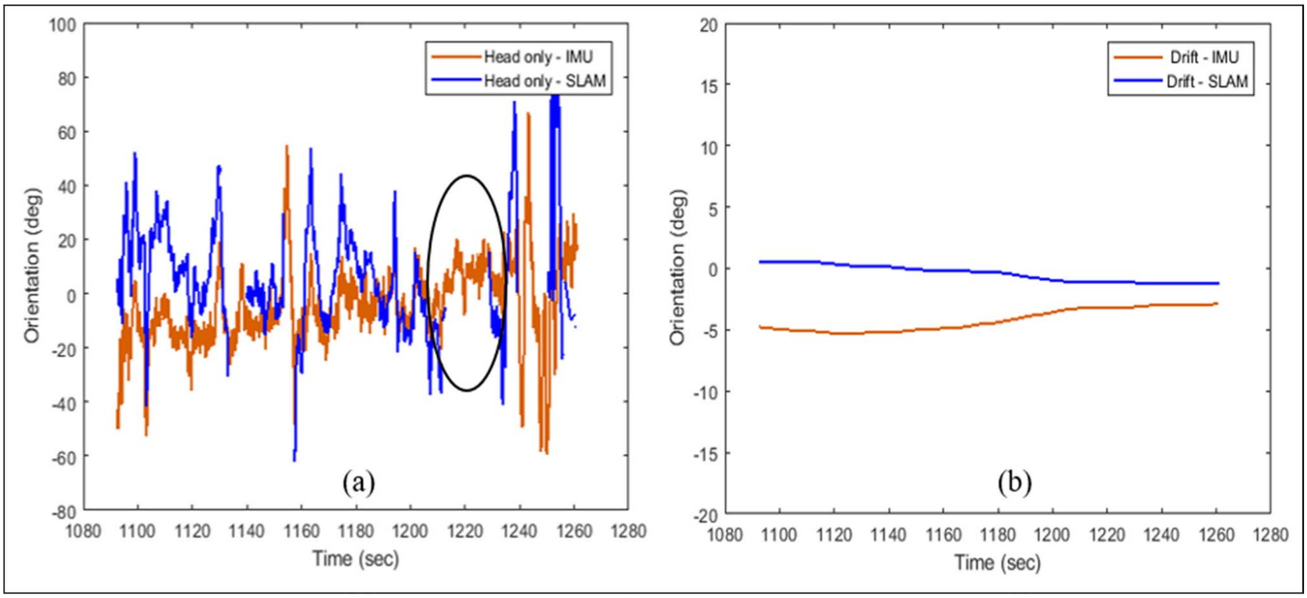


Fig. 10 **a** Horizontal Head only orientation for both IMU (orange) and SLAM (blue) where there is not much drift in both the plots in the last segment of the walk. Black oval shows an instance of tracking loss in SLAM (discontinuous blue line). **b** Filtered drift signal using low pass filter in both SLAM and IMU

head movements. We had loss of tracking while using the SLAM at multiple instances because of factors like faster head turns, not enough features, etc., resulting in loss of data as marked by the black ovals in Figs. 8, 9, and 10. In cases of tracking loss, we restarted the SLAM from the same point where tracking was lost, which resulted in multiple sequences of SLAM trajectory for each participant. All sessions were then merged using the timestamps on the generated trajectory map during the localization. For those 17 sessions of the five patients, the SLAM processing broke nine times on average (across all sessions) for each patient due to loss of tracking and had to be re-initialized. In total, such tracking failures resulted in data loss for approximately 7.2 minutes out of 70 minutes, about 10.3% of the walking data analyzed.

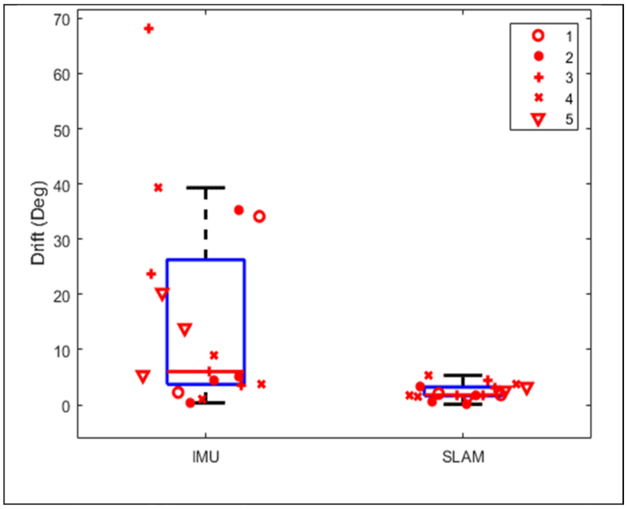


Fig. 11 Spread of RMSD for drift signals in walking sequences for both IMU and SLAM where each symbol denotes the RMSD for one sequence for a patient

Discussion

Measuring head movements is important for studying the gaze-scanning behaviors of patients with visual impairments, because of the relatively higher contribution of head movements in the overall gaze. Previous studies analyzed head movements by processing the data manually (Geruschat, Hassan, Turano, Quigley, & Congdon, 2006; Hassan, Geruschat, & Turano, 2005) or semi-manually (Marius't Hart & Einhäuser, 2012). Those analyses were either coarse, qualitative, and/or time-consuming. For instance, Marius't Hart et al. (2012) had to manually label vanishing points once every 30 video frames. Continuous automatic analysis of large amount of head movement data is desirable in behavior research.

In this study, we compared the accuracy and the accumulated drift in horizontal head orientation angles measured in an outdoor environment using differential IMU and visual SLAM approaches. The work could be extended to vertical and rotational movements in the future. The results from our experiments show that visual SLAM was more accurate and relatively resistant to drift as compared to the differential IMU method in both semi-controlled and uncontrolled outdoor settings. The main limitation of SLAM in our application was that it was associated with data loss due to tracking failures.

The drift in the case of IMUs was substantial in both the experiments. In the scanning experiment, the error could be attributed to large head turns (Figs. 5 and 6). The cause of this drift in both circumstances may be due to damping in the internal circuitry of the IMU sensor. In the case of the outdoor walking experiment, the IMU signal might have drifted due to external magnetic interference present in the environment. The observed signal drift was independent of the location during the walk, making it difficult to implement precautionary measures to mitigate the drifting errors. Tomasi et al. (2016) suggested periodically resetting the IMU. There were five predefined calibration points including two at the beginning and the end, respectively. The three predefined checkpoints result in four different sessions for each participant. Some participants did not walk the whole path, resulting in fewer calibration points and sessions, whereas some avoided stopping at any calibration checkpoint and continued walking, which resulted in longer walking durations in each session and fewer sessions. Some predefined calibration points do coincide with the larger turn where participants were asked to return. We found this measure helpful, but it did not completely prevent the drift.

We also inspected the data loss with SLAM when the participants were at locations that included either crossing the road at traffic signals or waiting for the walk sign to come on. Participants made larger head scanning for traffic inspection at those locations. SLAM failed to provide an output head orientation for approximately of 18.2% of the frames associated with crosswalk instances on an average. Due to this severe data loss, SLAM alone may not be suitable for head movement tracking at crosswalks where large head-scanning magnitudes are expected.

An interesting finding in our results was the difference in performance of SLAM in the scanning experiment and street walking experiment. While SLAM outperformed IMUs in terms of both accuracy and drift with no data loss in the scanning experiment, it suffered data loss in the street walking experiment due to tracking failure. The repetitive fixation locations (spatial map) in the parking lot provided a well-established map for the SLAM system. This consequently helped in generating more accurate results even in the presence of a wide head-scanning angle, with no data loss. Furthermore, the head movements in the parking lot were not rapid, as the participant held the gaze steady at each fixation. This was not the case in the street walking experiment, where there were rapid head turns, along with moving objects occluding the landmarks in the scene, leading to SLAM tracking failures. During the outdoor walking scenario, SLAM tracking failures were also seen because of the lack of unique features in consecutive frames, for example, when the participants look downward towards the road/pavement. This resulted in the requirement of manual intervention to restart the tracking process, which limits full automation of the processing when using SLAM. Data loss in SLAM due to tracking failure can be further attributed to factors like texture-less surface, which might hinder generation of enough unique features for SLAM algorithms. However, it is possible that the performance of SLAM might be improved by using an HD camera. As Sumikura et al.

showed, by using cameras (HD RICOH THETA series, Insta360 series) that can capture omnidirectional imagery, the tracking performance of SLAM can be further improved (OpenVSLAM, Sumikura et al.). Also, SLAM can be computationally intensive. On our system, the tracking frame rate for SLAM was about 4 frames/second.

While we did not quantitatively compare OpenVSLAM with other SLAM algorithms, other papers have done such comparisons. We chose OpenVSLAM based on the comparison results presented in the OpenVSLAM paper by Sumikura et al. It outperformed or performed similarly to a very popular visual SLAM algorithm, ORB-SLAM, in terms of trajectory error for several datasets. As the goal of this study is to compare VSLAM and IMU, we mainly focused on the comparison between the two types of technologies rather than subtype variants.

Considering the limitations of both the IMUs and SLAM-based approaches, SLAM is preferable when it comes to accuracy, whereas IMUs are preferred if data loss is of primary concern. IMUs do have advantages over visual SLAM in scenarios which hinder the image capture ability of conventional video cameras, such as in low ambient light. Drifts in IMUs might be removed by filtering in the case where participants are walking along a straight path such as our walking course. Ideally, both accuracy and data loss prevention are equally important, and therefore it could be beneficial to fuse information from IMUs and SLAM in a hybrid or in an adaptive sense. The integrated localization system of IMU together with SLAM, the monocular inertial SLAM, may have better performance as opposed to each of them individually (Campos et al., 2021; Juan-Rou & Zhan-Qing, 2020; Poulose & Han, 2019; Tiefenbacher, Schulze, & Rigoll, 2015).

The susceptibility of IMUs to various outdoor environmental factors could be countered with the use of a differential IMU system as proposed by Tomasi et al. (2016), followed by low-pass filtering. The overall pre- and post-processing involved in the IMU-based head movement tracking is highly computationally efficient as compared to the SLAM algorithm. A few degrees' advantage in accuracy of SLAM over IMU could be well ignored in studies concerning mainly large head movements, such as outdoor walking or street crossing.

Conclusion

Due to technical challenges, there is only limited research on head and gaze movements of visually impaired people performing mobility tasks in unrestricted outdoor conditions. We evaluated the visual SLAM method for tracking the head movements and compared the accuracy with the differential IMUs method with patients with hemianopia, who are expected to perform large head scanning to one side to compensate for their visual field loss. Visual SLAM outperformed the IMUs in terms of accuracy in all the experimental conditions and was resistant to drifting errors common with the IMUs. However, SLAM suffered from tracking loss (about 10% overall and 18.2% at street crossing). Trade-offs related to accuracy and data loss should be considered when comparing two complementary approaches for tracking unrestricted head movement in outdoor naturalistic settings. The signal drift problem of the differential IMU method did not always happen, but it might occur in walking. However, in our case where participants are walking along a straight path, the problem of IMU drift can be mitigated using low-pass filtering. By accounting for sensor drift, IMU-based head orientation may be used for studying straight-line walking behaviors.

Declarations

Conflict of interest All authors declare no conflicts of interest to disclose. Supported in part by a grant from NIH (R21EY031444), US Department of Defense grant (W81XWH-16-1-0033), and NIH core grant (P30- EY003790).

References

Bahill, A. T., Adler, D., & Stark, L. (1975). Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science, 14*(6), 468–469.

Barabas, J., Goldstein, R. B., Apfelbaum, H., Woods, R. L., Giorgi, R. G., & Peli, E. (2004). Tracking the line of primary gaze in a walking simulator: Modeling and calibration. *Behavior Research Methods, Instruments, & Computers, 36*(4), 757–770. https://doi.org/10.3758/bf03206556

Bowers, A. R., Ananyev, E., Mandel, A. J., Goldstein, R. B., & Peli, E. (2014). Driving with hemianopia: IV. Head scanning and detection at intersections in a simulator. *Investigative Ophthalmology & Visual Science, 55*(3), 1540–1548.

Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*: " O'Reilly Media, Inc.".

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*.

Cesqui, B., de Langenberg, R., Lacquaniti, F., & d'Avella, A. (2013). A novel method for measuring gaze orientation in space in unrestrained head conditions. *Journal of Vision, 13*(8). https://doi.org/10.1167/13.8.28

Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., & König, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems, 18*(3), 267–297.

Essig, K., Dornbusch, D., Prinzhorn, D., Ritter, H., Maycock, J., & Schack, T. (2012). *Automatic analysis of 3D gaze coordinates on scene objects using data from eye-tracking and motion-capture systems.* Paper presented at the Proceedings of the Symposium on Eye Tracking Research and Applications.

Geruschat, D. R., Hassan, S. E., Turano, K. A., Quigley, H. A., & Congdon, N. G. (2006). Gaze behavior of the visually impaired during street crossing. *Optometry and Vision Science, 83*(8), 550–558.

Grip, H., Jull, G., & Treleaven, J. (2009). Head eye co-ordination using simultaneous measurement of eye in head and head in space movements: Potential for use in subjects with a whiplash injury. *Journal of Clinical Monitoring and Computing, 23*(1), 31–40.

Hassan, S. E., Geruschat, D. R., & Turano, K. A. (2005). Head movements while crossing streets: Effect of vision impairment. *Optometry and Vision Science, 82*(1), 18–26.

Imai, T., Moore, S. T., Raphan, T., & Cohen, B. (2001). Interaction of the body, head, and eyes during walking and turning. *Experimental Brain Research, 136*(1), 1–18. https://doi.org/10.1007/s002210000533

Juan-Rou, H., & Zhan-Qing, W. (2020). *The Implementation of IMU/Stereo Vision Slam System for Mobile Robot.* Paper presented at the 2020 27th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS).

Kugler, G., Huppert, D., Schneider, E., & Brandt, T. (2014). Fear of heights freezes gaze to the horizon. *Journal of Vestibular Research, 24*(5–6), 433–441. https://doi.org/10.3233/VES-140529

Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., & Burgard, W. (2011). *g 2 o: A general framework for graph optimization.* Paper presented at the 2011 IEEE International Conference on Robotics and Automation.

Larsson, L., Schwaller, A., Holmqvist, K., Nyström, M., & Stridh, M. (2014). *Compensation of head movements in mobile eye-tracking data using an inertial measurement unit.* Paper presented at the Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication.

Larsson, L., Schwaller, A., Nystrom, M., & Stridh, M. (2016). Head movement compensation and multi-modal event detection in eye-tracking data for unconstrained head movements. *Journal of Neuroscience Methods, 274*, 13–26. https://doi.org/10.1016/j.jneumeth.2016.09.005

Lin, C.-S., Ho, C.-W., Chan, C.-N., Chau, C.-R., Wu, Y.-C., & Yeh, M.-S. (2007). An eye-tracking and head-control system using movement increment-coordinate method. *Optics & Laser Technology, 39*(6), 1218–1225.

MacDougall, H. G., & Moore, S. T. (2005). Functional assessment of head-eye coordination during vehicle operation. *Optometry and Vision Science, 82*(8), 706–715. https://doi.org/10.1097/01.opx.0000175623.86611.03

Marius't Hart, B., & Einhäuser, W. (2012). Mind the step: Complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Experimental Brain Research, 223*(2), 233–249.

Mur-Artal, R., & Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics, 33*(5), 1255–1262.

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics, 31*(5), 1147–1163.

Poulose, A., & Han, D. S. (2019). Hybrid indoor localization using IMU sensors and smartphone camera. *Sensors, 19*(23), 5084.

Proudlock, F. A., Shekhar, H., & Gottlob, I. (2003). Coordination of eye and head movements during reading. *Investigative Ophthalmology & Visual Science, 44*(7), 2991–2998. https://doi.org/10.1167/iovs.02-1315

Rosten, E., Porter, R., & Drummond, T. (2008). Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(1), 105–119.

Rothkopf, C. A., & Pelz, J. B. (2004). *Head movement estimation for wearable eye tracker.* Paper presented at the Proceedings of the 2004 symposium on Eye tracking research & applications, San Antonio, Texas. https://doi.org/10.1145/968363.968388

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). *ORB: An efficient alternative to SIFT or SURF.* Paper presented at the 2011 International conference on computer vision.

Shoemake, K. (1994). Euler angle conversion *Graphics gems IV* (pp. 222–229): Academic Press Professional, Inc.

Stoll, J. (2015). Measuring gaze and pupil in the real world: Object-based attention, 3D eye tracking and applications.

Sumikura, S., Shibuya, M., & Sakurada, K. (2019a). *OpenVSLAM: A versatile visual SLAM framework.* Paper presented at the Proceedings of the 27th ACM International Conference on Multimedia.

Sumikura, S., Shibuya, M., & Sakurada, K. (2019b). OpenVSLAM: A versatile visual SLAM framework. https://github.com/OpenVSLAM-Community/openvslam

Tiefenbacher, P., Schulze, T., & Rigoll, G. (2015). *Off-the-shelf sensor integration for mono-SLAM on smart devices.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Tomasi, M., Pundlik, S., Bowers, A. R., Peli, E., & Luo, G. (2016). Mobile gaze tracking system for outdoor walking behavioral studies. *Journal of Vision, 16*(3), 27–27.

Wang, Y., Zeng, H., & Liu, J. (2016). *Low-cost eye-tracking glasses with real-time head rotation compensation.* Paper presented at the 2016 10th International Conference on Sensing Technology (ICST).

Open practices statement None of the data or materials for the experiments reported here is available to public at the time of this publication, and none of the experiments was preregistered. We plan to make the video data, sensor data, and patient data open when we write a dataset paper in the near future.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.