Mobile gaze tracking system for outdoor walking behavioral studies

Matteo Tomasi

Schepens Eye Research Institute, Mass Eye and Ear, Harvard Medical School, Boston, MA, USA ✉

Shrinivas Pundlik

Schepens Eye Research Institute, Mass Eye and Ear, Harvard Medical School, Boston, MA, USA

Alex R. Bowers

Schepens Eye Research Institute, Mass Eye and Ear, Harvard Medical School, Boston, MA, USA

Eli Peli

Schepens Eye Research Institute, Mass Eye and Ear, Harvard Medical School, Boston, MA, USA

Gang Luo

Schepens Eye Research Institute, Mass Eye and Ear, Harvard Medical School, Boston, MA, USA

Most gaze tracking techniques estimate gaze points on screens, on scene images, or in confined spaces. Tracking of gaze in open-world coordinates, especially in walking situations, has rarely been addressed. We use a head-mounted eye tracker combined with two inertial measurement units (IMU) to track gaze orientation relative to the heading direction in outdoor walking. Head movements relative to the body are measured by the difference in output between the IMUs on the head and body trunk. The use of the IMU pair reduces the impact of environmental interference on each sensor. The system was tested in busy urban areas and allowed drift compensation for long (up to 18 min) gaze recording. Comparison with ground truth revealed an average error of 3.3° while walking straight segments. The range of gaze scanning in walking is frequently larger than the estimation error by about one order of magnitude. Our proposed method was also tested with real cases of natural walking and it was found to be suitable for the evaluation of gaze behaviors in outdoor environments.

Introduction

Analysis of gaze orientation is essential to answering questions about where people look and why (Eggert, 2007; Godwin, Agnew et al., 2009; Hansen & Qiang 2010; Itti, 2005; Land & Furneaux 1997; Land, Mennie et al. 1999). Where a person is looking is a result of both head and eye movements (Land, 2004). In many vision research studies involving stimuli presented on a fixed display, the subject's head is often stabilized using a chin rest, unless the eye tracking system is able to account for the head movement, when determining gaze points on a screen (Barabas, Goldstein et al., 2004; Cesqui, de Langenberg et al., 2013; Johnson, Liu et al., 2007).

In studies involving mobile subjects, two approaches have been used for gaze estimation: tracking of eye-in-head orientation with registration of the eye position on scene images from a head mounted camera (Fotios, Uttley et al., 2014; Li, Munn et al., 2008), or tracking the gaze orientation in a confined environment using multiple cameras (Land, 2004). In the former approach, the head orientation is typically not measured, and interpretation of gaze behaviors requires visual examination of the recorded scene videos and manual data entry of head orientation. Using the latter approach, tracking both the eye position (in head) and the head orientation, makes it possible to measure the gaze point in real world coordinates. However, this is usually performed only in confined indoor or within-car environments, where the head orientation (relative to the car) can be tracked by the sensing system set up in the enclosed environment (Barabas, Goldstein et al., 2004; Bowers, Ananyev et al., 2014; Cesqui, de Langenberg et al., 2013; Essig, Dornbusch et al., 2012; Grip, Jull et al., 2009; Imai, Moore et al., 2001; Kugler, Huppert et al.; Land 2009; Lin, Ho et al., 2007; MacDougall & Moore

2005; Proudlock, Shekhar et al., 2003). Due to technical difficulties in outdoor gaze tracking, there have been only a few behavioral studies (Geruschat, Hassan et al., 2006; Hassan, Geruschat et al., 2005; 't Hart & Einhauser, 2012) that investigated gaze behaviors while walking in uncontrolled outdoor open spaces. Furthermore, analyses of these studies were limited by the challenge of manual processing of large volumes of gaze data. For example, in studies of gaze behaviors during street crossings at intersections (Hassan, Geruschat et al., 2005), the head mounted scene videos were visually examined and gaze behaviors were manually classified coarsely as left, center, or right relative to the crosswalk. The quantitative analysis of a large amount of gaze tracking data is impractical with such a technique.

Quantitative analyses of gaze tracking data can be valuable and provide important behavioral insights that eye tracking data alone cannot. For instance, for people with peripheral visual field loss, such as tunnel vision and hemianopia, two hypotheses can be proposed regarding their gaze scanning strategies. On one hand, according to a bottom-up eye movement model, they might make smaller gaze movements than normally sighted observers due to the lack of visual stimuli from the periphery (Coeckelbergh, W. Cornelissen et al., 2002). On the other hand, according to a top-down eye movement model, they might make large gaze movements toward the blind area to compensate for their field loss (Martin, Riley et al., 2007; Papageorgiou, Hardiess et al., 2012). Luo et al. (2008) measuring only eye-in-head orientation, found that people with tunnel vision made saccades as large as normally sighted in outdoor walking and they frequently moved their eyes to presaccadic invisible areas. This finding suggests that top-down mechanisms play an important role in eye movement planning. However, a question remains unanswered—did the patients with tunnel vision and the normally sighted observers look toward the same areas? For example, it is possible that patients with peripheral vision loss tend to look down at the ground by lowering their head orientation more than normally sighted observers while maintaining the same eye scanning patterns, or they could be scanning laterally with their heads more than the normally sighted observers as a compensatory strategy? Measuring eye movements alone will miss these behaviors. Also, when considering gaze patterns in normally sighted people, Kinsman et al. (2012) suggest that, in the presence of the vestibular-ocular response (VOR), the head counteracts eye-in-head rotation and knowing the head rotation improves the detection of fixations. To address such questions, gaze orientation that combines eye and head has to be measured.

A system able to record gaze orientation in natural walking conditions would be a useful tool for quantitative studies; yet, none of the existing methods allow us to easily and reliably record and analyze gaze orientation

under these conditions. The main challenge is in tracking head orientation. Shoulder mounted cameras have been proposed to track the head in outdoor environments (Essig, Dornbusch et al. 2012; Imai, Moore et al. 2001). However, natural body wobbling while walking requires special handling of the camera because the sway motion introduces noise that is computationally and technically challenging to cancel. Sway motion cancellation has been performed in humanoid robots in which more information about the self-movement is available (Dune, Herdt et al. 2010) but not in real walking.

Computer vision techniques for orientation estimation, such as simultaneous localization and mapping (SLAM; Davison, Reid et al. 2007), have been proposed for robot navigation applications. However, even the latest SLAM-based methods have some crucial limitations that make them ill suited to estimate head movement using a head-mounted scene camera. These approaches are inherently iterative and require extensive amount of prior tuning, as they rely on recognition of stable landmarks for localization and mapping over multiple runs, and the estimation errors can accumulate over time. Such approaches face difficult challenges when applied to walking in natural environments (e.g., on a busy city street) characterized by uncontrolled changes, such as passing pedestrians and vehicles.

To solve the problem, an existing commercial eye tracking system, EyeSeeCam (Interacoustics, Eden Prairie, MN), mounts an eye tracker together with a motion sensor on the head. An application for this system was presented by Kugler et al. (2014), who used the EyeSeeCam device for measuring the gaze-in-space of subjects standing still on an emergency balcony for 30 s. Motion sensor drift can present a significant problem when sensors are used in uncontrolled situations, such as outdoor walking. Kugler et al. (2014) likely avoided these problems because the recording time was very short. In a similar manner, Larsson et al. (2014) proposed the integration of a mobile eye tracking system with one inertial measurement unit (IMU) for head movement compensation in an indoor scenario, again avoiding the many problems of sensor drift in outdoor recordings.

We present a novel approach for tracking gaze relative to the heading direction while walking outdoors realized via a portable system comprised of a head mounted eye tracker and a pair of off-the-shelf motion sensors for head movement tracking. In contrast to the mobile tracking devices implemented by Kugler et al. (2014) and Larsson et al. (2014), which used only one motion sensor, we use a pair of motion sensors to address the problem of sensor drift that occurs during extended periods of outdoor recording. The system was designed to measure gaze and head scanning behaviors of visually impaired people as well as normally sighted people, and was tested on busy city streets in order to demonstrate its potential use for

a)   **b)**
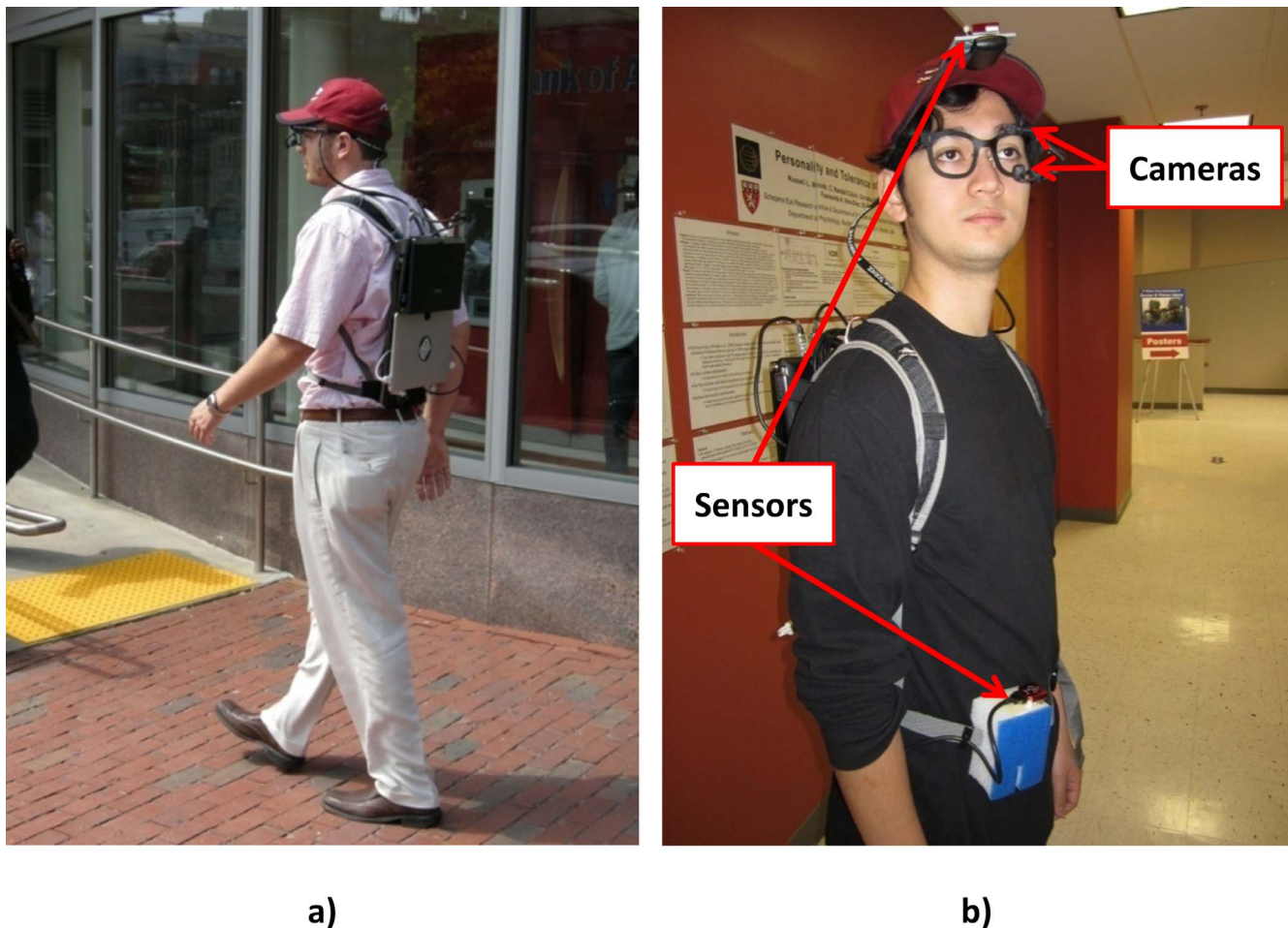
Figure 1. The mobile eye and head tracking system. (a) Rear view showing the two data logging computers on the backpack for recording eye and head movements. The eye movement logger also records the video from the scene camera. (b) Front view of the system showing the eye tracker headgear (eye tracking and scene cameras) and the body and head tracking sensors.

addressing the challenges in recording and analyzing human gaze patterns in outdoor environments.

Materials and methods

Our portable gaze tracking system includes a commercial eye tracking system from Positive Science (Positive Science, New York City, NY; 2013) and two commercial IMUs from VectorNav (VectorNav, Dallas, TX). We integrated these components in a lightweight backpack system (Figure 1), which is reasonably comfortable for the wearer.

While the head tracking system is capable of measuring yaw, pitch, and roll of the head relative to the body trunk, in this paper, we focus on the yaw measurement because of its importance in street walking and road crossing. It would be straightforward to extend our technique to the pitch movements (vertical movements of the head and eyes). Our gaze

tracker records two videos (one directed to the eye and the other directed to the forward scene) and the motion sensor output; all these data streams are logged at a frequency of 30 Hz.

In the following subsections we detail each part of the system and the integration of all collected data during the post-processing.

Eye tracker

The Positive Science eye tracker (Figure 1b) includes an eye tracking camera and a scene (60° field of view) camera, both mounted on one side of a spectacle frame to track one of the eyes. This arrangement enables the mounting of prescription lenses in front of both eyes as well as optical visual field expansion devices (e.g., peripheral prism glasses for hemianopia; Bowers, Keeney et al., 2008; Giorgi, Woods et al., 2009; Peli, 2000) fitted to the other eye not being tracked.

We use a 13-point calibration for eye tracking at the beginning and the end of each recording session to detect any displacement of the headgear during the walk. The calibration is performed at a distance of 4 m using fixation targets posted on a whiteboard covering the full scene camera image. If the point correspondences between the two calibrations match, it means that there were no shifts of the cameras. In such cases, we can use the same calibration for the whole sequence. Otherwise, the sequence is split in two unequal parts, in which the first calibration is applied to the initial part of the data, and the second calibration to the latter part of the data. In order to localize the split point in the acquired data, we perform brief calibration checks along the outdoor walking route at convenient locations, where the subject looks at a specified point and confirms it verbally. The experimenter asks the subject to fixate on specific points and audio is recorded to help with identifying the calibration points when playing back the scene video for postprocessing. Whenever the eye tracker output confirms an accurate location of the point of regard (POR) we assume that the prewalk calibration is valid for the video otherwise we verify again with the calibration performed at the end. The best calibration is applied for the segment. If none of the calibration is found valid, the corresponding segments are discarded from analyses.

Head tracking system

We use IMUs for head tracking, which require much simpler processing than video-based head tracking cameras. The IMU is comprised of accelerometers, gyroscopes, magnetometers, and a microcontroller. Gyroscopes are notorious for drifting over time. The magnetometer, acting as a reference, is used to correct the gyroscopic drift over time using Kalman filters. With sensor fusion technology, this type of IMU normally can perform well without severe drifting in suitable environments. However, they are not completely interference-proof, especially in complicated environments, such as busy downtown areas, where a strong presence of electromagnetic and metallic objects disturbs the local Earth magnetic fields. We use a few strategies, described below, to address the problems of sensor drift and magnetic interference. Specifically with a pair of matched IMUs, external electromagnetic interferences can be largely canceled.

After a qualitative evaluation of various commercial IMUs in terms of accuracy and internal processing capabilities (as detailed by the manufacturer), we chose the VN-100 model (VectorNav, Dallas, TX). While the results reported here are associated with this particular choice, the method we propose can be applied with other commercial IMUs. As Figure 1b shows, one

sensor is secured on a hat visor that the subject wears, while the other is attached to the belt and used as a reference. The sensors are connected to a laptop in the backpack.

With a two-sensor system, where one of the sensors is mounted on the waist, we are able to measure horizontal head rotations relative to the body by computing the difference between the two sensors. For this mode of use it is important to use sensors that are as matched as possible to reduce uncorrelated noise across the sensors and benefit from the differential signal. Correlated external interference, affecting both sensors in a similar manner because of their proximity, will be cancelled out when the differential signal is computed providing a cleaner signal of head orientation relative to the body (see Results).

However, even when a differential signal is used, the outputs of both sensors may still drift relative to each other because of uncorrelated noise. In these cases, the drift is not necessarily predictable and it might depend on uncontrolled external factors that affect the two sensors in a different way or internal noise, which is different in both sensors. In order to reduce errors from uncorrelated external or internal noise, we included in our experimental protocol several reset operations performed at preset checkpoints along the outdoor route. The location of the checkpoints depends on the distance walked, and it is preferable to set checkpoints after turning points on the walking path. At the checkpoints, the subject is asked to pause briefly and look straight ahead. A heading reset command is issued (to the logging computer) to realign the orientations of the two sensors and therefore minimize the impact of possible differential drifts. The quick verification of the eye tracker calibration, described earlier, is performed at the same time.

With the software development kit provided for the IMUs, we have developed customized software for logging the inertial data. This software also deals with basic sensor diagnostics, parameter setting, and user interaction. The software runs on a lightweight ASUS Eee PC notebook with an Intel Atom N455 processor, mounted on the backpack (Figure 1a). The output provided is a text file containing timestamp in seconds, orientation in degrees, acceleration in meters per second squared, and angular velocity in degrees per second for both sensors. The software keeps track of the reset points and can record a key code from the experimenter to identify significant events.

Post-processing

We collected the data from two different computers because of bandwidth limitations. A few steps are

required for synchronization. Here, we briefly summarize the procedure used to generate the final gaze estimations.

Eye tracking from raw videos: Eye + scene camera

Yarbus eye-tracking software (Positive Science) is used to generate eye position coordinates for every scene camera video frame. The calibration data is marked in the software by playing back the recorded videos from both eye and scene cameras, and interactively marking the calibration points on the scene videos. During the playback, we confirm that the subject was looking at the correct calibration point with the help of the recorded verbal instructions. We also ensure by visual inspection that a viable corneal reflection is present in the eye video. The software then maps the eye movements in the video obtained by the infrared eye camera to the corresponding locations in the scene video based on the above calibration. Thus, the software generates an output video and a text file where information about eye position (in terms of scene image coordinates expressed in pixels), time, and pupil size are logged. Yarbus software also provides a measure of reliability for each tracked point based on the presence of the corneal reflection and the quality of the pupil detection.

Correction of detected eye position

The wide-angle (60°) scene camera lens results in a significant spatial distortion of the displayed scene. The Yarbus software does not correct for these distortions while mapping the eye movements to the scene. The output of the eye tracking system is calibrated in terms of pixel coordinates of fixation on the scene image coordinates, not the actual scene or real world coordinates. However, since we wish to convert the eye position on the scene image to visual angles in the real world, the effect of the image distortion has to be corrected. We later correct for the optical distortion imposed to the scene video by applying a geometrical transformation to the raw eye tracking data based on the intrinsic camera parameters and lens distortion parameters, computed offline using the camera calibration function provided in the OpenCV library (OpenCV Development Team, 2013). Using these computed parameters, the eye tracking coordinates in the scene camera image coordinates (in pixels) are transformed to their corresponding undistorted eye-in-head orientation angular values.

Synchronization of eye and head signal

At the beginning as well as the end of the recording session, we ask the person to perform rapid head turns to the right and back to the left in order to create easily identifiable anchor marks for the synchronization of eye movement output with the head signal. The sharp peaks of the head turns as observed in the scene camera are represented by a unique indexed frame in the video. Similarly, the peak corresponding to this rapid head movement is identified in the head sensor signal. Once these anchor points are manually identified, the corresponding time offsets from the start in both the data streams are fed to the customized gaze processing software to synchronize them with one frame precision. The software generates a new text file including synchronized eye and head position data for further analysis.

Combination of synchronized and corrected output into gaze orientation

Once the visual angle conversion factor is calculated, it is applied to the entire eye position dataset to obtain visual angles with respect to the neutral point in the scene camera frame. As there is no specific location in the scene camera that can be identified as the straight ahead orientation, the neutral point (0°) for the eye positions is arbitrarily calculated as the average of all the fixation points on the scene images recorded within 30 ms after the reset operations carried out within a recording session. Note that the subject is instructed to look straight ahead during the reset operations. During the reset, the head orientation signal is also set to zero so that the neutral point is represented by the heading direction. Obtaining the gaze angle is then a matter of summing the head rotation angle estimated by the sensors and the visual angle due to the eye movements relative to the head. The head and the eye angles can be directly summed in our application because the distance between the eye and sensors center of rotation (order of centimeters) is small when compared with the typical viewing distance used when walking (order of meters).

Motion sensors drift correction

Since there is always a possibility of drift in the motion sensor data, we check the deviation of the head orientation from 0° at reset/checkpoints. We assume that any error that we measured between a reset value and the following reset measure was due to an accumulated drift. If this residual drift is greater than 10° the whole segment is considered unusable and discarded.

Experiments and results

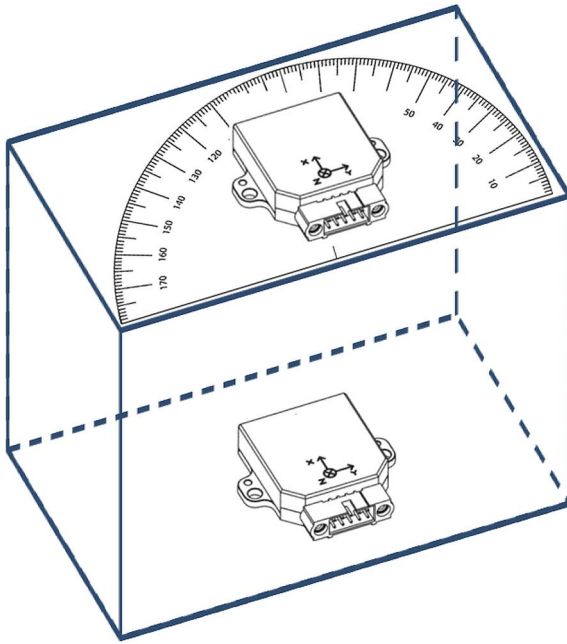We conducted three experiments to evaluate the performance of our system. The experiments were

Figure 2. A pair of sensors mounted on a rigid body such that one of the sensors is fixed to the box (representing the body motion sensor), while the other sensor (representing the head motion sensor) can be rotated horizontally (yaw) with respect to the other. A protractor is mounted underneath the rotating sensor to indicate the amount of rotation (performed manually).

designed to test each component in isolation. First we tested the paired IMU method and validated the benefit and limitation of the differential signal, which allowed us to determine the needed frequency of reset/check-points in our protocol. The second experiment evaluated the accuracy of the head orientation measurements, and the third was designed to test the combination of head orientation estimation with the eye tracker output. All the experiments were conducted outdoors in a busy urban environment, the Government Center area in Boston, MA.

Experiment 1

To evaluate the stability of the relative measurement between the two IMUs while walking outdoors, the two sensors were mounted on a rigid cardboard box, as illustrated in Figure 2. We measured a range of movements from −60° (left) to +60° (right). During the first experiment, an experimenter walked outdoors for 20 min carrying the entire assembly in his hands. The walking route was a 150-meter straight section of a sidewalk in a busy downtown area. The experimenter walked back and forth along the route making 180° turns of heading at both ends. Starting from the beginning of data collection, the top sensor (together

with the protractor) was manually rotated by 20° every 30 s during the walk. Because we wanted to evaluate the long term drift of the IMUs, we didn't reset the heading during the test.

We repeated the experiment three times. Figure 3 shows the noisiest one out of the three collected data sets using the two-sensor system shown in Figure 2. In Figure 3b, the differential signal accurately represents the rotation of the rotating sensor while the raw data from each sensor were very noisy. As mentioned in the Methods section, the noise across sensors due to external effects, which is highly correlated, can be effectively cancelled with the differential output. Figure 3a shows the difference between the two sensors over the entire period of one data collection session with a clear stepwise pattern in the relative orientation output. The data sample (Figure 3b) demonstrates relatively stable behavior up to 300 s in the differential measure. After this point, the measure started to drift (the drift sections are highlighted with ellipses). The drift error between protractor shifts was no more than 20°. The worst drift was in the middle of data collection (from 400 to 800 s), where the sensors drifted very quickly, as shown in Figure 3a. Following that drifting period, the relative orientation measure appeared to be correct again without any intervention. We think this residual drift might have been caused by unpredictable magnetic interferences that affected the two sensors in a different way (i.e., the body sensor received stronger interference from the ground than the head sensor). Because the interference was uncorrelated and disrupted the drift by different amounts in the two sensors, it was not fully cancelled by the differential output function (see example in Figure 3c). However, these drifts did not always happen. In another recording session carried out in the same environment, we did not observe such an event. The differential signal, shown in Figure 4, was relatively stable for more than 11 min with only small levels of drift observed.

This evaluation experiment underscores the need for a carefully designed protocol with frequent reset points in order to limit the deleterious issue of differential sensor drift. For our particular outdoor route, if we had performed a reset every 5 min, the likelihood of a drift-free recording was 70% for each 5-min segment based on data combined across the three repetitions of the experiment. We suggest that a similar test should always be conducted to determine the frequency of heading reset needed for each experimental situation and environment.

Experiment 2

In the second experiment, we tested the system drift with natural head movements. The subject stood on a busy downtown sidewalk. One of the sensors was
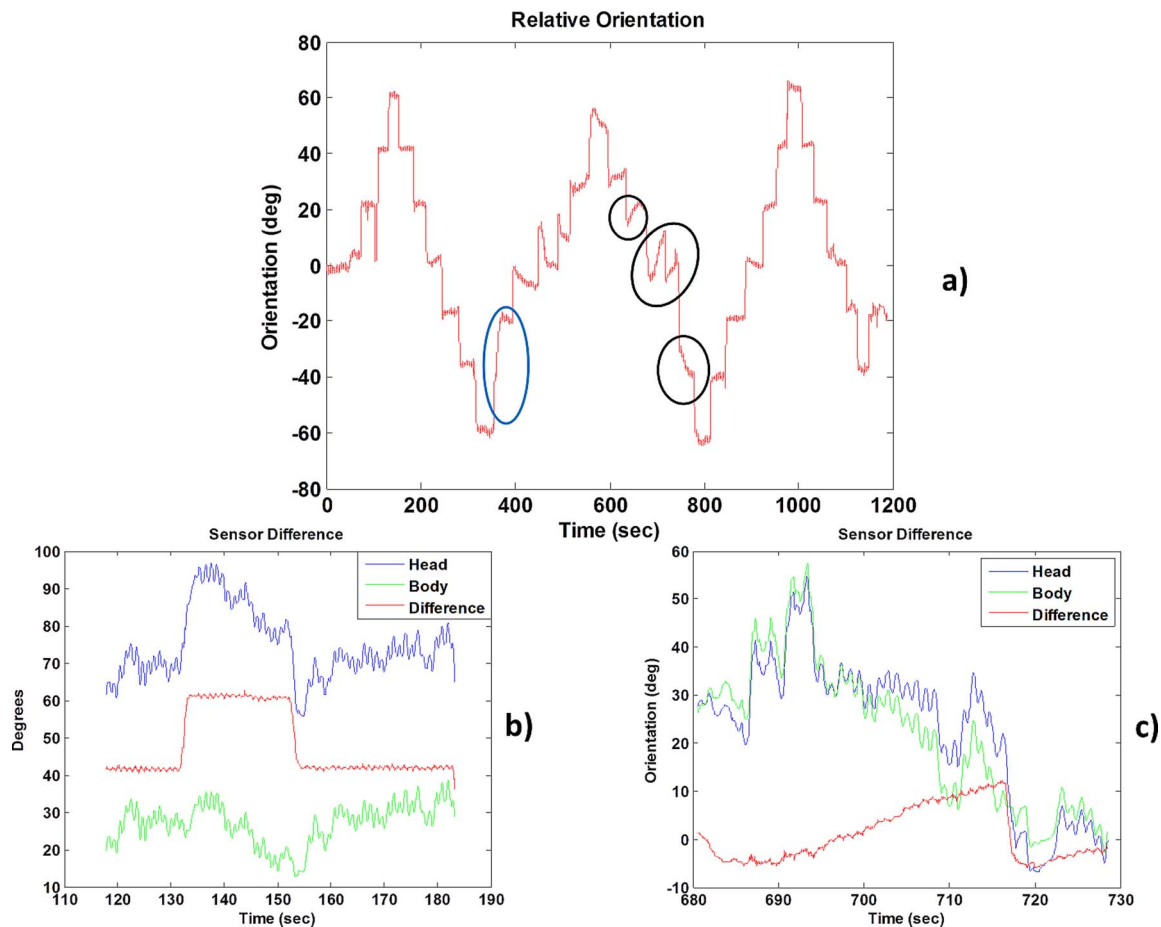
Figure 3. Figure 3. Relative orientation in degrees between the two inertial measurement units (IMUs) in Experiment 1a. (a) Steps of 20° are clearly shown in the relative orientation output. Note that after 300 s the system starts drifting (experiencing uncorrelated noise), and there is an error of about 20°. The step in the blue ellipse should be 20° increment and not 40°, and the sections of the output highlighted by the black ellipses should appear flat in the plot (constant); only a few of them are marked for the sake of clarity. (b) Orientation output of the head and body sensors for a 60-s segment of the recording around the peak at 140 s on plot (a). The net head rotation is obtained by subtracting the output of the body motion sensor from the head motion sensor. Although both appear to be very noisy, the noise is highly correlated, and the differential signal provides relative orientation, which is close to the ground truth. (c) Orientation output of the head and body sensors for a 50-s segment of the recording, starting at 680 s on plot (b), showing details of uncorrelated noise in the two sensors. In this case, the differential output does not help, and the measured net head rotation increases when it should have been flat.

mounted on a hat worn by him (Figure 1). The reference sensor was fixed on a box (as the lower one in Figure 2), and was held steadily in his hands against his body. During the 13 min of data collection, the person moved his head naturally (in a random manner) and watched the street scene. After about every minute he took off the head-mounted sensor and manually aligned it with the reference sensor so that the relative orientation was 0°. At this point, a key was pressed to register the moment as a checkpoint in the sensor log file. No heading reset was performed at these check-points.

Figure 5a shows the yaw angles for the head-mounted sensor extracted from 13 min of recorded data

between two of the reset points (at 1200 and 2100 s). Sixteen checkpoints are marked by red asterisks at 0°. The large and rapid fluctuations in the relative orientation measure are due to the arbitrary head movements as well as the abrupt movement of the head-mounted sensor when it was taken off the head to be aligned with the reference sensor. Figure 5b shows only the relative orientation at the 16 checkpoints. A slow drift from negative to positive can be seen in the figure. The standard deviation of the error was 2.6°, with a maximum error of 10.6° and a mean of 3.5°.

This experiment confirms the conclusion of the first experiment that the drift needs to be addressed by using frequent checkpoints and reset points.
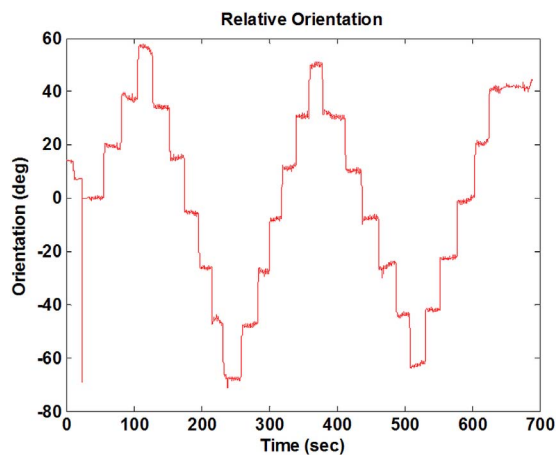
Figure 4. Relative orientation in degrees between the two IMUs in Experiment 1b. Repetition of Experiment 1a using the same conditions on another day over a period of time of 700 s without a reset showed that the system is more stable than the previous data collection shown in Figure 3.

Experiment 3

This experiment tested the integrated gaze tracking system including the eye tracker in a mobile outdoor setting. The purpose was to evaluate the accuracy of the gaze tracking system in various natural conditions that may occur during outdoor walking scenarios. It involved three different conditions, (a) standing still, (b) walking with body lateral shifts only but without turns, and (c) walking with multiple body turns.

Data were collected in front of a building from a distance of 5.7 m (225 in.). Five pillars of the building served as the targets. A person sequentially looked at each of these pillars for a few seconds in turn by making natural head and eye movements (Figure 6). When the subject was looking at a pillar, the experimenter pressed a button on the logger to flag that time. This enabled us to compare the gaze tracking system output with the ground truth angles computed using the measured distances.

Standing still

The first condition with the standing observer consisted of 6 min. of data collection. The person looked at the 0 position and reset the sensor, then looked at each of the pillars in a predetermined order 0, L1, L2, L1, 0, R1, R2, R1, 0 (Figure 6). He repeated this sequence and looked at each pillar five times. The reset was used only at the beginning of the recording. When looking at each of the pillars, the subject was instructed to move both his head and his eyes naturally.

Figure 7a shows the plot of the head orientation (blue) and the overall gaze recorded (red) for the standing observer condition during 6 min. The ground-truth values for the gaze angles at various intermediate checkpoints are shown as green points. The overall gaze angle at any given moment is the sum of the head and the eye tracking system estimates. Positive values are movements to the right and negative to the left. A staircase-like pattern is seen in the gaze orientation plot
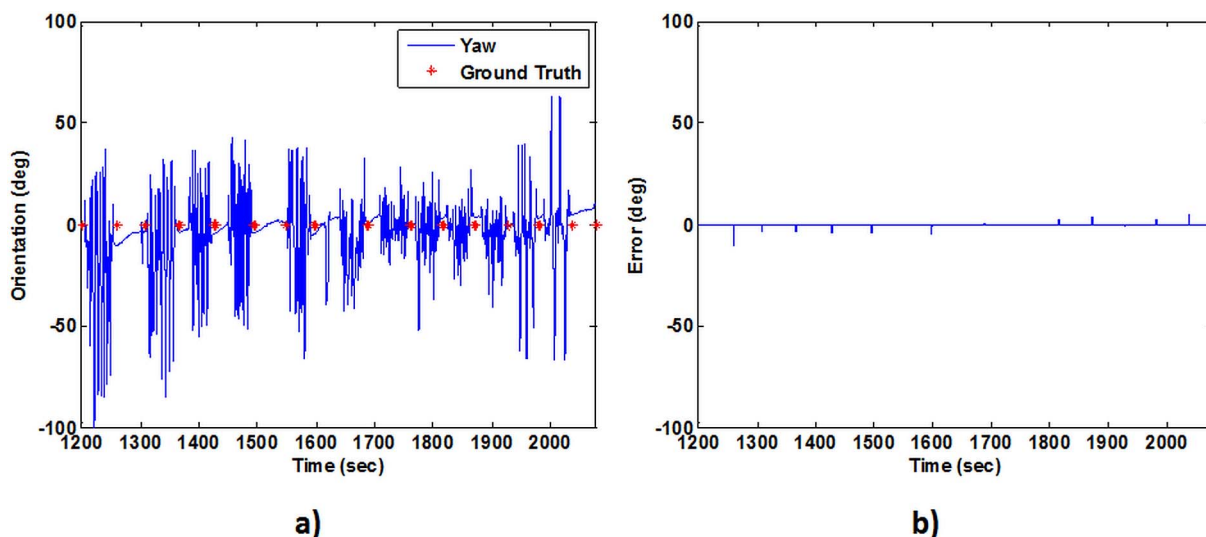

Figure 5. Outputs for the head tracking experiment using a head mounted sensor and one stationary sensor as a reference in Experiment 2. (a) The continuous data output for the head movement in the yaw direction recorded over a period of about 13 min between two reset points (at 1200 and 2100 s). Also overlaid, are the ground truth orientation values at each of the recorded checkpoints (red stars). (b) Only the orientation values corresponding to the checkpoints are shown. The orientation values indicate the deviation from the ground truth values (0°), and show a gradual drift from negative to positive over time. Negative values represent movement to the left.
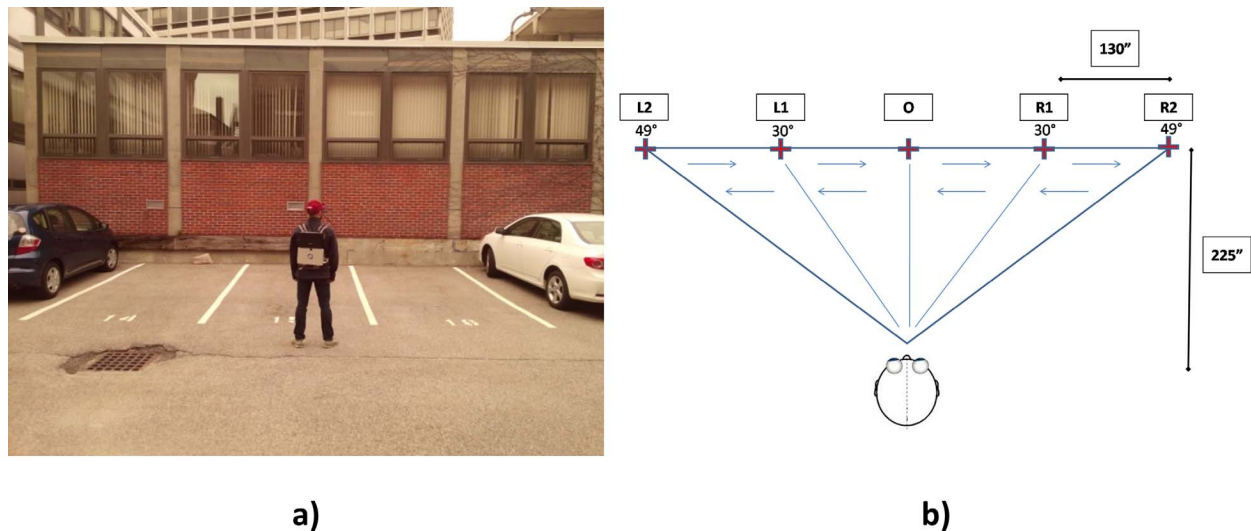
Figure 6. (a) Setup for Experiment 3 for gaze tracking with a stationary subject. (b) The person looks sequentially and repetitively at the points in the following order: 0, L1, L2, L1, 0, R1, R2, R1, 0. At each of the fixation points, a checkpoint is registered in the log, so as to facilitate the computation of the ground truth.

(Figure 7a) reflecting the systematic pattern of gaze shifts across the five pillars (Figure 6). The eye tracking output has a higher variance compared to the head orientation output. However, the head orientation is accompanied by a small amount of drift (seen prominently for the extreme right and left points in the scene).

Figure 7b shows the gaze estimation errors (black circles) at the each ground truth point. The gaze estimation error is the difference of the estimated gaze angle (red segment) and the ground truth (green crosses) angles for the scene locations. The average absolute gaze estimation error for all the checkpoints was 1.4° with a standard deviation of 0.9°. The system behaved in a stable manner for the standing observer condition over 6 min. The maximum error recorded over the full set of checkpoints (including the far periphery) was 3.6°.

Walking with body lateral shifts only but without turns

In the other two experimental conditions, we evaluated the impact of movement and large body turns on system accuracy. Our hypothesis was that the motion sensors could be adversely affected by large
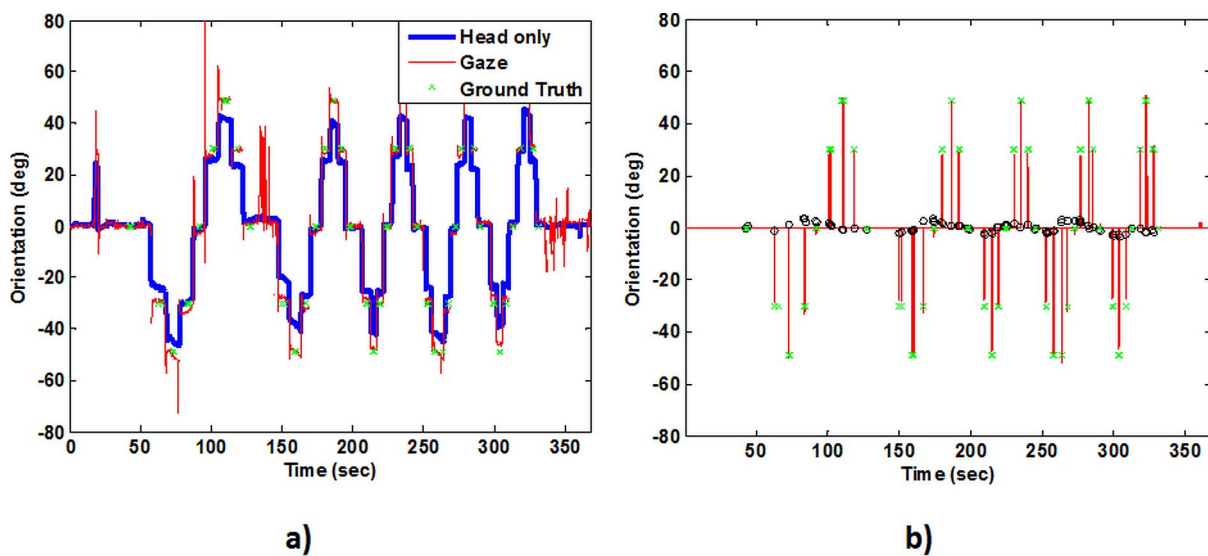

Figure 7. Outputs for gaze tracking in the static observer condition of Experiment 3. (a) Continuous data output for the head movement (blue line) and the gaze (red line) in the yaw direction. (b) The same output, but only at the checkpoints. The red segments represent the gaze estimation, and black circles the measurement error from the ground truth (green crosses).
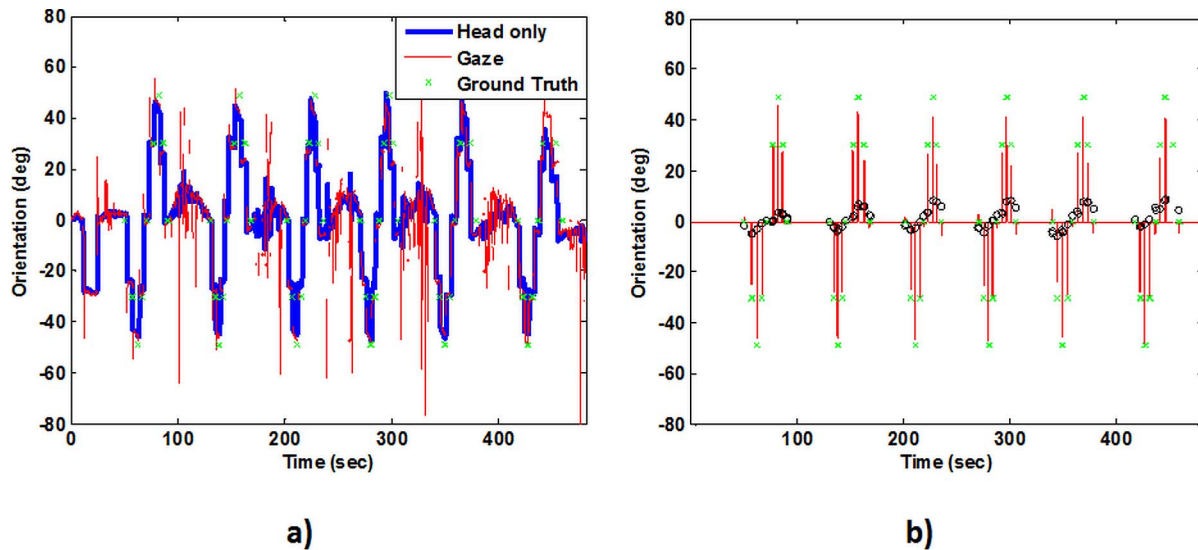
Figure 8. Outputs for the gaze tracking for the straight walking condition (without turns). (a) The continuous data output for the head movement (blue line) and the gaze (red line) in the yaw direction. The very noisy head data are during the walking segments and were not part of the analyses. (b) Data values at the checkpoints only: red segments represent the gaze measurements and black circles represent the deviation from the ground truth (green crosses).

turns and hence, the sensor drift could become worse. We first evaluated the accuracy of gaze estimation with only straight walks (condition b) in which the person moved back and forth but always faced the same direction (the wall with the pillars). In this way, we introduced the typical oscillatory nature of body movement when walking, but without any full body turns. The subject started from the same position as in Figure 6 and performed a full scan (starting from 0, going to the extreme left, then back to the extreme right, and then back again to 0). Between scanning cycles, the person moved backwards for about 5 m while still facing forward, and then walked back to the original scanning position without turning to perform the next round of scanning. No resets were applied between scans. The results of 8 min of data collection are shown in Figure 8.

As expected, the gaze estimation errors were larger in this case than those in the standing observer condition. Compared with the ground-truth values of the gaze at the checkpoints, there was an average absolute gaze estimation error of 3.3° with a standard deviation of 2.3°. The maximum absolute checkpoint error was 8.8°.

Walking with multiple body turns

The last condition (c) involved arbitrary head movements, complete body turns, and lateral displacements. The experiment was conducted using the same setup shown in Figure 6, but the subject walked around the parking lot and on the sidewalk for approximately 50 m before going back to the scanning position for the next scan. The sequence was repeated

a total of seven times during 18 min of recording. The walks included multiple body turns, so both the head sensor and the reference sensor on the waist rotated a great deal. The natural walking behavior was reflected in the gaze estimation output shown in Figure 9a, in which random angles were recorded between the checkpoints and an uncontrolled walking speed was adopted by the subject between consecutive scans (note the uneven separations of the ground truth checkpoints in the figure). The errors (black circles) from the ground truth (green crosses) are reported in Figure 9b. Again, no heading resets were applied between loops.

As expected, this condition generated a larger error as compared to the standing observer and no-turn walking conditions due to more unstable sensor drift. For this last condition, the average absolute error was 7.1° with a standard deviation of 5.2°.

Case study

To illustrate the value of the proposed system the gaze scanning behaviors of four persons (including one normally sighted and three visually impaired with hemianopia) walking on a busy street in downtown Boston were recorded. Each walk session lasted for about 10 min with four reset points along the entire length of the route; we placed checkpoints after a designated (or required) body turn.

From these pilot measures, we were able to quantify the contribution of head turns to the overall gaze
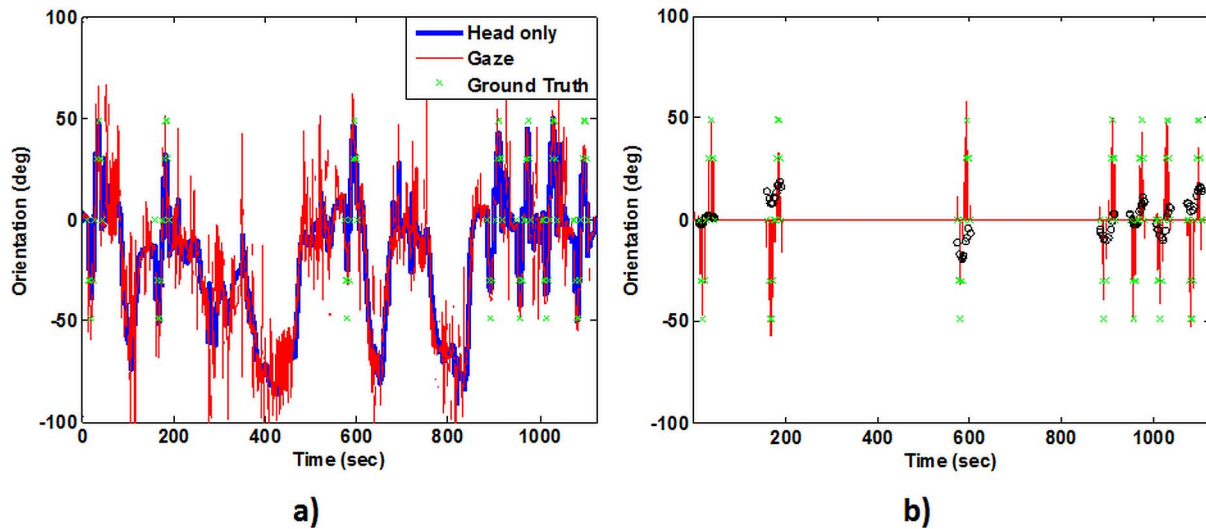
Figure 9. Outputs for the gaze tracking experiment while walking with body turns. (a) The continuous data output for the head movement (blue line) and the gaze (red line) in the yaw direction. The walking speed between checkpoints was not kept constant by the subject. (b) Data values at the checkpoints only: red segments represent the gaze estimation, and black circles the deviation from the ground truth (green crosses).

orientation. Figure 10 shows a typical gaze-scanning event for the normally sighted subject. It can be seen that the contribution of the head rotation (blue line) to the total gaze orientation (red line) is large for wide quick gaze scanning (larger than 20°, 0.4–1.5 s) and a slow head turn (1.5–2.5 s), whereas the head rotation contributes much less for narrow quick gaze scanning (less than 15°, 0–0.4 s), and depends only on eye rotation. With a conventional head-mounted eye tracking system there would have been little or no information about the subject's true gaze scanning behaviors because the eye is often at low deviation positions (green line); for example, even when he scanned to the far right side (positive values), at about 0.8 s, gaze shift was more than 25° but eye deviation was less than 5°.
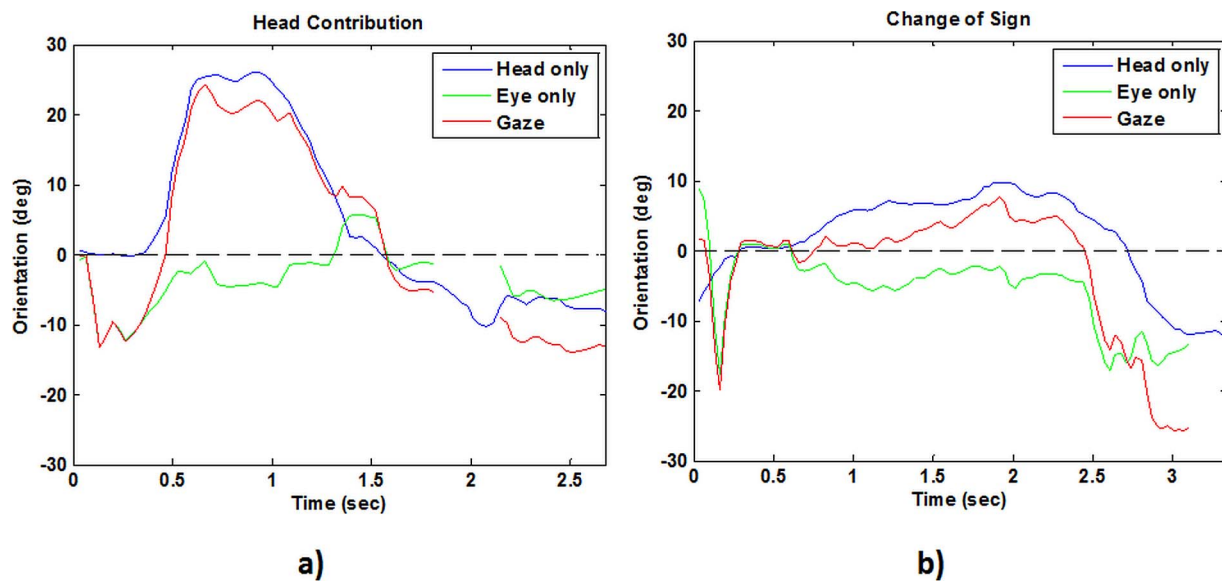

Figure 10. Two examples of scanning behaviors during a 10 min walk. Head, eye, and final gaze movements are shown. (a) An example of large head contribution (blue line) to the final gaze orientation (red line). The eye contribution (green line) in this case is small. If only eye movements had been tracked we might have erroneously concluded that this subject did not scan. (b) The head and the eye orientation moving in opposite directions (opposite signs). The gaze orientation stays positive (rotation to the right) between 0.7 and 2.6 s, although the eye tracking is showing the opposite behavior (negative, rotation to the left).

ID	Average head contribution	Opposite behavior points	Total valid points
NV1	46%	5,954 (16%)	36,058
LV2	43%	3,763 (17%)	21,128
LV3	37%	7,118 (15%)	47,702
LV4	46%	3,773 (16%)	23,251

Table 1. Summary of the behaviors illustrated in Figure 10 for data collected during the 10-min walk by four subjects. The average head contribution is the percentage of contribution of the head to the gaze values. Opposite head-eye rotation points are the number of points where the gaze orientation has a different sign as compared to the eye position. In parentheses we show the percentage of valid samples over total points (last column).

We were also able to document a behavior in which the eye and head turn in opposite directions. Figure 10 shows such examples: from 0.4 to 1.3 s in Figure 10a, and from 0.7 to 2.6 s in Figure 10b. The head rotation appears to be positive (rotation to the right) while the eye direction negative (rotation to the left). If the head movements were not considered, the eye tracking data alone could have been misleading.

In Table 1 (column 2) we report the average percentage of head contribution to the gaze movement for the full 10 min walk for each subject. The average head orientation contribution was in the range of 37%–46%. This was calculated with the ratio of the head scan over the total gaze scan on all the valid points: valid points are considered if we have both eye-in-head and head orientation from the system. Column 3 in Table 1 summarizes the number of eye/head opposite data points for the four subjects, which was more than 15% of the valid samples.

Discussion

Here we summarize several points that may be important for the success of future data collection using the proposed method based on a set of evaluation experiments of our gaze tracking system.

It is important to use at least two motion sensors for drift and interference reduction. The sensor attached to the body trunk acts as a reference so as to enable the computation of head movement with respect to the current heading direction. In this paper we showed how each sensor is affected by external interference and how the resultant drift could be corrected by the use of a differential output.

Although the error is reduced with the use of two sensors, we still have some residual error that may not be ignorable for some studies. We showed that this error increases when the person is walking with sudden changes in the heading direction. In order to reduce this accuracy loss, it is a good strategy to plan for frequent heading resets, especially after large body turns. If we perform a reset after each body turn, the segments between reset points can be considered straight. Our experiments demonstrated that the system has an average error of 3.3° for this condition. In the case study for a real walk that included 180° body turns, heading resets were always executed after turns. The situation in the case study then became equivalent to the experiment without body turns.

For all the experiments that were conducted, we report the results only for the yaw head movement but our method could be applied to the pitch head movements in the same way. Our primary application of the mobile gaze tracking system will be for the study of lateral gaze scanning by patients with homonymous hemianopia (the loss of the same half of the visual field in both eyes) when walking, hence our focus on yaw movements. However, for other applications, vertical gaze movements while walking may also be of interest.

We showed the large contribution of head position to gaze position in the outdoor walking data for four different persons. The cases demonstrate that the head plays a large role in the gaze scanning pattern, on average above 40% of the gaze movement. This observation can be expected since people walking on streets can freely move the head. This finding is consistent with a previous study by Freedman (2008) who reported that the head largely contributes to the gaze shift in cases of large movements (>30°) or in a particular starting position of the eyes. These results underscore the importance of tracking head movements when studying gaze behaviors in outdoor environments.

In our case study we also found an occasional discrepancy between the orientation of the head and the eyes. A person's eyes might be pointing to one side, but due to the head position, the overall gaze could still be on the opposite side. This result is in line with the vestibulo-ocular reflex (VOR) and the compensatory rotations of eye and head while walking (Freedman, 2008). In fact, because of the VOR, the eyes tend to move in the opposite direction to the head to stabilize the image on the retina during nonscanning head movements. Another possible reason accounting for the discrepancy is the presence of multiple saccades per single head movement as previously reported by Fang, Nakashima et al. (2015). The authors demonstrated that, during the same head movement, subjects can perform different eye movements and those are not always in the same

direction as the rotation of the head. Such differences between eye and head rotation patterns cannot be easily understood without the head tracking information.

It is also worth mentioning that the characteristics of the terrain on which the person is walking may play an important role in the gaze distribution ('t Hart & Einhauser, 2012; Vargas-Martín & Peli, 2006). Hart and Einhauser conclude that the contributions of eye-in-head orientation as compared to head-in-world orientation to gaze are complementary: eye movements are adjusted to maintain some exploratory gaze and to compensate for head movements caused by the terrain. This is an important motivation for the adoption of mobile gaze trackers similar to our system in future studies.

We reported the results of the system evaluation experiments with a few subjects. Based on our results we hypothesize that the accuracy can be affected primarily by interferences in the environment. The validity of the method can be extended to all subjects with a normal gait but we foresee some problem in the data collection of subjects affected by abnormal hip movement or mobility limitation, such as a limp.

Conclusion

A mobile gaze tracking system has been proposed for outdoor walking studies. The approach can be useful for behavioral gaze scanning studies. There are no existing commercial systems for calculating gaze (head + eye) movements and visual angles in open outdoor environments that are using more than one motion sensor to compensate for sensor drift. Gaze movement information in terms of visual angles is often needed for scanning behavior studies. We developed a method to calculate the head + eye orientation using off-the-shelf components, a portable eye tracker, and a pair of motion sensors. We are able to merge eye orientation with the head orientation to obtain gaze relative to the body. Outdoor testing using the proposed gaze tracking system shows an average error of $3.3° \pm 2.3$ when walking along straight segments. When walking included some large body turns and reset was not applied, the average error increased to $7.1°$. Based on our experimental results, some suggestions have been proposed to ensure the accuracy of gaze orientation tracking. We conclude the system may be sufficiently reliable for scanning behavior studies involving large gaze scanning movements in open-space mobility conditions.

Keywords: gaze tracking, outdoor walking, head movement

References

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*, 757–770.

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*, 1540–1548. [PubMed] [Article]

Bowers, A. R., Keeney, K., & Peli, E. (2008). Community-based trial of a peripheral prism visual field expansion device for hemianopia. *Archives of Ophthalmology, 126*, 657–664.

Cesqui, B., de Langenberg, R. v., Lacquaniti, F., & d'Avella, A. (2013). A novel method for measuring gaze orientation in space in unrestrained head conditions. *Journal of Vision, 13*(8):28, 1–22, doi:10.1167/13.8.28. [PubMed] [Article]

Coeckelbergh, T. R. M., Cornelissen, F. W., Brouwer, W., & Kooijman, A. C. (2002). The effect of visual field defects on eye movements and practical fitness to drive. *Vision Research, 42*, 669–677.

Davison, A. J., Reid, I. D., Molton, N. D., & Stasse, O. (2007). MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*, 1052–1067.

Dune, C., Herdt, A., Stasse, O., Wieber, P. B., Yokoi, K., & Yoshida, E. (2010, Oct 18–22). Cancelling the sway motion of dynamic walking in visual servoing. In Ren C. Luo & Hajime Asama (Eds.), *Proceedings of 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 3175–3180). Taipei, Taiwan: IEEE.

Eggert, T. (2007). Eye movement recordings: Methods. *Developments in Ophthalmology, 40*, 15–34.

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 (pp. 37–44). Santa Barbara, California, ACM.

Fang, Y., Nakashima, R., Matsumiya, K., Kuriki, I., & Shioiri, S. (2015). Eye-head coordination for visual cognitive processing. *PLoS One, 10*(3), e0121035.

Fotios, S., Uttley, J., Cheal, C., & Hara, N. (2014). Using eye-tracking to identify pedestrians' critical visual tasks, Part 1. Dual task approach. *Lighting Research and Technology,* doi:10.1177/1477153514522472.

Freedman, E. G. (2008). Coordination of the eyes and head during visual orienting. *Experimental Brain Research, 190*, 369–387.

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*, 550–558.

Giorgi, R. G., Woods, R. L., & Peli, E. (2009). Clinical and laboratory evaluation of peripheral prism glasses for hemianopia. *Optometry and Vision Science, 86*, 492–502, doi:10.1097/OPX.0b013e31819f9e4d.

Godwin, A., Agnew, M., & Stevenson, J. (2009). Accuracy of inertial motion sensors in static, quasistatic, and complex dynamic motion. *Journal of Biomechanical Engineering, 131*(11), 114501.

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*, 31–40.

Hansen, D. W., & Qiang, J. (2010). In the eye of the beholder: A survey of models for eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*, 478–500.

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

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*, 1093–1123.

Johnson, J. S., Liu, L., Thomas, G., & Spencer, J. P. (2007). Calibration algorithm for eyetracking with unrestricted head movement. *Behavior Research Methods, 39*, 123–132.

Kinsman, T., Evans, K., Sweeney, G., Keane, T., & Pelz, J. (2012). Ego-motion compensation improves fixation detection in wearable eye tracking. *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 221–224). Santa Barbara, CA: ACM.

Kugler, G., Huppert, D., Schneider, E., & Brandt, T. (2014). Fear of heights freezes gaze to the horizon. *Journal of Vestibular Research, 24*, 433–441.

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research, 159*, 151–160.

Land, M. F. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience, 26*, 51–62.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 352*(1358), 1231–1239.

Land, M. F., Mennie, N., & Rusted, J. (1999). Eye movements and the roles of vision in activities of daily living: Making a cup of tea. *Perception, 28*, 1311–1328.

Larsson, L., Schwaller, A., Holmqvist, K., Nystrom, M., & Stridh, M. (2014). Compensation of head movements in mobile eye-tracking data using an inertial measurement unit. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 1161–1167). Seattle, WA: ACM.

Li, F., Munn, S., & Pelz, J. (2008). A model-based approach to video-based eye tracking. *Journal of Modern Optics, 55*, 503–531.

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*, 1218–1225.

Luo, G., Vargas-Martin, F., & Peli, E. (2008). Role of peripheral vision in saccade planning and visual stability: Learning from people with tunnel vision. *Journal of Vision, 8*(14):25, 1–28, doi:10.1167/8.14.25. [PubMed] [Article]

MacDougall, H. G., & Moore, S. T. (2005). Functional assessment of head-eye coordination during vehicle operation. *Optometry and Vision Science, 82*, 706–715.

Martin, T., Riley, M. E., Kelly, K. N., Hayhoe, M., &

Huxlin, K. R. (2007). Visually-guided behavior of homonymous hemianopes in a naturalistic task. *Vision Research, 47*, 3434–3446.

OpenCV Development Team. (2013). OpenCV webpage. Retrieved from http://opencv.org/.

Papageorgiou, E., Hardiess, G., Mallot, H. A., & Schiefer, U. (2012). Gaze patterns predicting successful collision avoidance in patients with homonymous visual field defects. *Vision Research, 65*, 25–37.

Peli, E. (2000). Field expansion for homonymous hemianopia by optically induced peripheral exotropia. *Optometry and Vision Science, 77*, 453–464.

Positive Science. (2013). Positive Science webpage. Retrieved from http://www.positivescience.com.

Proudlock, F. A., Shekhar, H., & Gottlob, I. (2003). Coordination of eye and head movements during reading. *Investigative Ophthalmology & Visual Science, 44*(7), 2991–2998. [PubMed] [Article]

't Hart, B. M., & Einhauser, 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*, 233–249.

Vargas-Martín, F., & Peli, E. (2006). Eye movements of patients with tunnel vision while walking. *Investigative Ophthalmology & Visual Science, 47*(12), 5295–5302. [PubMed] [Article].

VectorNav Technologies. (2013). *VN100 User Manual.* Retrieved from http://www.vectornav.com/support/manuals.