

Smartphone sensor reliability for augmented reality applications

Jeffrey R. Blum, Daniel G. Greencorn, and Jeremy R. Cooperstock

McGill University, Montréal, Québec, Canada
{jeffbl, dangreencorn, jer}@cim.mcgill.ca

Abstract. With increasing reliance on the location and orientation sensors in smartphones for not only augmented reality applications, but also for meeting government-mandated emergency response requirements, the reliability of these sensors is a matter of great importance. Previous studies measure the accuracy of the location sensing, typically GPS, in handheld devices including smartphones, but few studies do the same for the compass or gyroscope (gyro) sensors, especially in real-world augmented reality situations. In this study, we measure the reliability of both the location and orientation capabilities of three current generation smartphones: Apple iPhone 4 and iPhone 4s (iOS) phones, as well as a Samsung Galaxy Nexus (Android). Each is tested in three different orientation/body position combinations, and in varying environmental conditions, in order to obtain quantifiable information useful for understanding the practical limits of these sensors when designing applications that rely on them. Results show mean location errors of 10–30 m and mean compass errors around 10–30°, but with high standard deviations for both making them unreliable in many settings.

Keywords: GPS, location, compass, magnetometer, augmented reality, sensor fusion, smartphones

1 Introduction

Smartphone augmented reality applications typically require the user’s location (latitude, longitude) and orientation (relative to north) within certain bounds of accuracy. Some applications may require only relative device orientation, such as that produced by a gyroscope, which can determine changes in device orientation, albeit not relative to north. For example, our In Situ Audio Services application [2], which renders points of interest via spatialized audio to blind users, relies on both the location of the device and its orientation relative to north. Due to the unreliability of the sensors, the scene was frequently mis-rendered, with points of interest in the wrong direction and/or at the incorrect distance. The study described in this paper was motivated by these issues arising from smartphone sensor unreliability. Specifically, we needed to determine when and by how much these sensors failed to provide accurate GPS, orientation, and gyro data. To do so, we walked two separate areas in Montreal 18 times each, with three smartphones in different body position and device orientation combinations, resulting in a total of 108 device logs for analysis. Based on the results,

we are able to offer several recommendations for designers of other augmented reality smartphone applications. By way of example, we describe how we revised one of the rendering techniques of our application to take into account the high degree of sensor unreliability.

2 Previous Work

Accuracy has significant impact on real-world usability of mobile devices, e.g., for transit tracking systems [12], and there has thus been considerable work in evaluating the accuracy of smartphone sensors. Most such studies find that the GPS accuracy of a smartphone is considerably less than that of a dedicated GPS device designed solely for navigation purposes. These findings hold true for devices such as the Apple iPhone, which augments the standard GPS with WiFi and cellular tower information, as well as an online database of satellite locations, to implement augmented GPS (A-GPS) [16]. Accuracy is often measured by maintaining the device in a stationary position over a given sampling duration. This scenario is common in domains such as forestry, where the GPS unit can be left in position for an extended period to obtain a better fix, but is of limited relevance to an augmented reality application. For such tests in ideal open-sky conditions in a forest, even the best consumer-grade dedicated GPS systems yield average errors on the order of 2 m [15]. In more typical augmented reality application conditions, on-body location of the GPS receiver has been found to impact accuracy [13], as does the manner in which a smartphone is held [3]. Studies that examine the accuracy of various location sensors while the device is in motion often use one, presumably more accurate, device, such as a higher-end GPS unit, as the reference, although this can only establish relative error. This may be a reasonable approach in areas without tall buildings in good weather conditions, and when evaluating devices and approaches that are expected to have significantly higher error than the “reference” unit [4]. Other reports do not appear to use ground truth measurements, but instead rely solely on location accuracy estimates reported by the device itself [14]. To save power, the GPS sensor can be turned off between readings, although this may further decrease accuracy [7].

A tradeoff between cost of components, power and accuracy is also evident for the three-axis magnetometers, commonly used to determine the orientation of a device relative to north [6]. Again, the often poor accuracy of the compass headings produced by the magnetometer can be a critical issue for augmented reality (AR) applications [1]. Body position, how the phone is held, and the effects of device movement when carried while walking can all impact compass readings [4]. Filtering the raw magnetometer data in order to remove sensor noise has been proposed [5, 8], but such solutions do not measure or improve the accuracy of the compass when it is subjected to distortions of the local magnetic field. One measure of the compass accuracy of a Nokia phone, taken while walking in an indoor corridor, found a mean error of 18.1° , with a standard deviation of 12.3° [10]. In such indoor spaces, attempts to use a built-in smartphone camera

to better detect motion or location have been tried, coupled with sensor fusion via a Kalman filter [11]. However, such camera-based solutions have the practical problem of greatly reduced battery life due to their power requirements. In addition, they require maintaining the device in a position that provides sufficient viewing area to compute optical flow, which is impossible, for example, if the smartphone is kept in a pocket. Another effort to overcome limitations of the compass sensor involves indoor uses of ultrasound signals with multiple receivers on the device to determine orientation. However, this solution requires the introduction of new infrastructure in the environment [9], which may be impractical for general deployment.

The results of the previous studies notwithstanding, there exists very limited quantitative data on the performance of current smartphone location and orientation sensors. In particular, there is a notable lack of such data for dynamic urban environments and under realistic conditions relevant to augmented reality applications. For these reasons, we undertook the experiment described in the remainder of this paper.

3 Experiment Design

An undergraduate student (the *walker*) walked the same path repeatedly under different conditions to log compass, orientation (particularly yaw) and location data under a variety of conditions. Each of two different paths, termed *Downtown* and *Commercial*, was walked half the time in a clockwise (CW) direction, and half in a counterclockwise (CCW) direction. Three smartphones (iPhone 4, iPhone 4s and Google Galaxy Nexus) were carried on each walk, each in one of three body positions: attached to a belt, hanging on the chest, or resting flat on the head, as shown in Figure 1. Each combination was repeated three times. This resulted in 2 walking locations \times 2 directions \times 3 carrying positions/orientations \times 3 repetitions = 36 data logs for each of the three devices.

The first two body positions were chosen to test the devices in the most practical carrying positions in real-world situations. The head position was included, regardless of practicality, since this exposes the device maximally to the sky, which we expected to result in the best location performance. In this case, we could determine whether the other two body positions further compromise location accuracy.

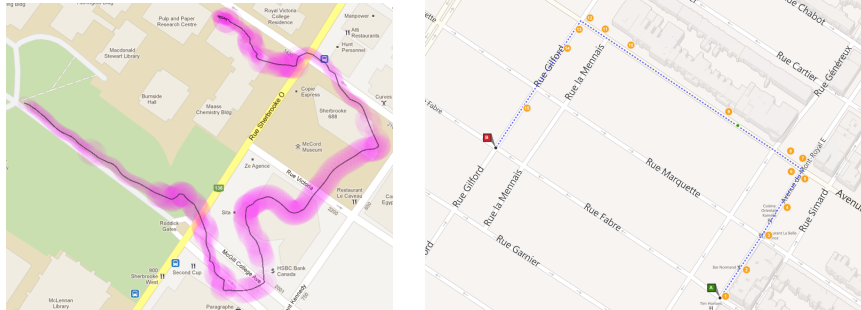
3.1 Walking Locations

The two areas walked are described in Figure 2. Each leg of each walk is analyzed independently in Section 4 to provide a comparison of the different conditions. Before each walk, the walker noted the weather conditions for the upcoming walk, shut down and restarted the application, and recalibrated the system to start with reasonably good initial sensor readings. Data plans and WiFi were enabled on all phones used in this study, and were thus using A-GPS and potentially



(a) Walker with three phones (b) Chest position (c) Belt position (d) Head position, under hat

Fig. 1: The three device positions/orientations. To avoid magnetic interference, the only metal used in the carriers was a belt buckle on the opposite side of the waist from the belt position.



(a) Reported data from one downtown walk, including McGill campus and downtown skyscrapers. Error estimates are shown in purple. (b) Commercial walk, outside downtown Montreal, with mostly sub-three story buildings nearby. Orange circles are ground truth points.

Fig. 2: The two areas walked during testing.

WiFi triangulation. This implies that we were not testing the GPS in isolation, but rather, the full location hardware and software stack of each device.

3.2 Ground truth

To facilitate data analysis, the walker clicked a button on the headphone cable upon reaching each of the location ground truth waypoints, and at the start and end of each leg. This helped exclude corners and intersections, which require head movement to cross safely. All three device clocks were tightly synchronized, which allows us to capture these clicks on one device as an accurate timestamp reference for all three. With the system's clock synchronization setting enabled on each device, the log files generated during a laboratory test indicated that

the timestamps associated with simultaneous button presses on the two iPhones occurred within 250 ms. When verifying manually before each walk in the Commercial set, we did not see differences greater than 1 s between any of the devices.

Compass ground truth was established by using a map screenshot of the area being walked and measuring the angle of each straight-line leg relative to north. For all of the data gathering and analysis, true north, as corrected from magnetic north via the magnetic declination, was used.

Distance error was calculated by comparing the most recently reported GPS location to the ground truth value at each waypoint. We used commercial online mapping systems to determine the ground truth waypoint locations along the path. Comparison against known survey points yielded accuracy for the manually selected map locations of within 1 m. This avoids experimental bias that could be introduced by comparing one GPS against another, especially knowing that even high-end units can have difficulty with limited view of the sky.

3.3 Sources of error

There are several sources of potential error that result from our choice of using real-world conditions for these tests. First, the walker could not hit every waypoint exactly, but reported being within approximately 1 m of the ground truth targets. Second, as the walker moves, he sways from side-to-side. Similarly, to avoid other pedestrians and maintain his safety around traffic, some changes in direction were inevitable, although he endeavoured to keep his body as straight as possible. Judging from the local minimum and maximum gyro values manually measured from the graphs of several iPhone walks, it appears that the walker typically swayed by roughly 8-17° to each side for the Belt and Chest positions, but only roughly 5° for the Head position. Third, especially in the chest position, where the phone is hanging in a pouch around the neck, we expect more motion due to the device swaying on its lanyard. In the Head and Belt positions as well, we expect some error in the manual alignment process at the beginning of the walk. A cursory examination of the data indicated that compass error can be consistently positive in one leg, and consistently negative in the next. This precludes the possibility of estimating device alignment error from an assumption of a zero-mean error distribution. A more rigid apparatus for holding the devices may reduce these errors, but would also make the test significantly less realistic.

4 Results

4.1 Individual log results

Before discussing the aggregate data across all walks, an explanation of the data from a single walk, illustrated in Figure 2a, may prove helpful.¹

¹ Our data, along with additional figures not shown here for space reasons, are available for review from isas.cim.mcgill.ca/Sensors/.

In addition to the location path, we can also plot various heading values over time, visualizing where the sensors deteriorate, and how they perform relative to each other. Figure 3 indicates compass values (cyan) against ground truth (black), the latter of which is constant (horizontal) for each straight-line leg of the walk. Actual compass error (red) is calculated as the absolute difference between these two, whereas reported compass error (grey) is an estimate of error magnitude by the sensor itself. As can be seen, the actual error fluctuates above and below the estimate. Gaps in the plot represent transitions between legs of the walk, during which we have no ground truth heading information.

Yaw (green), obtained from the gyro sensor, is not calibrated to north, so only represents relative variation. This data is generally expected to be flat, excepting body sway while walking. Slope in yaw indicates drift, observed in all legs of this walk. The reported course (purple) is derived from the direction of travel based on previous location updates. Both the iPhone and Android devices report course and speed based on location changes, but these appear to be of limited use even in the constrained straight-line testing we performed.

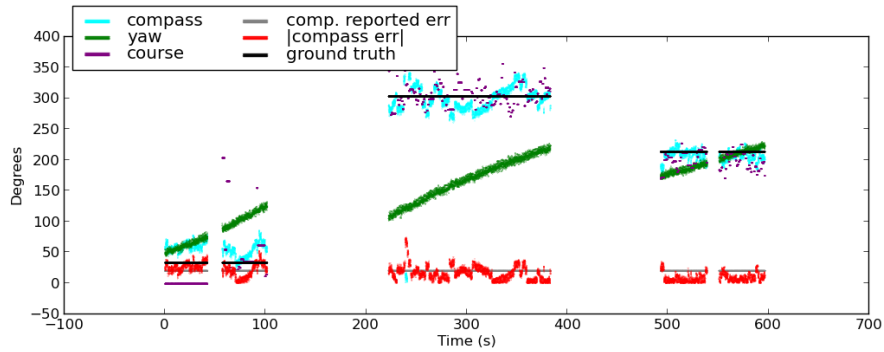


Fig. 3: Sensor values over time for a single device log file. Values are compass heading (cyan), heading ground truth (black), compass error estimate reported by the system (grey), yaw (green), course (purple), and $|\text{ground truth} - \text{compass reading}|$ (red).

Aggregating the data across walks and segmenting it into the legs (for headings) or individual measured ground truth waypoints (for locations), allows us to discern general trends in the sensor performance.

4.2 Heading accuracy

The only sensor on a smartphone with any knowledge of north is the magnetometer. However, large metal objects such as cars, electric power lines, and other interference can cause distorted readings. In early testing, we noted a 30° change in an iPhone 4's heading information when walking near a large vehicle.

To summarize our data, we pooled all of the values for a given condition, then calculated the mean and sample standard deviation, shown as error bars. For heading data, since the samples did not arrive on a regular period, we re-sampled the data at 25 samples/sec to make the sets consistently sampled per unit time. Without this, a single sample, valid for more than a second due to lack of device motion, would be weighted far less in the analysis than a series of samples within the same second on another device. Due to differences in walking speed across walks, we also subsampled the compass updates in order to balance the number of samples in each set for analysis. On Android, we simply registered for TYPE_ROTATION_VECTOR updates, which is north aligned, with no parameter for their frequency. Separate yaw values are not provided. On both platforms, compass or yaw changes greater than 0.2° were logged, which resulted in anywhere from less than one to over 30 unique samples/s, depending on the amount of motion. Most notably, the Head position appears to be more stable than Chest or Belt, and on the iPhone devices has dramatically fewer compass values, in some cases, well under one sample / second. For the two iPhone devices, the yaw value was updated on a timer, at a maximum of 10 samples/s in the Downtown walks, and 20 samples/s in the Commercial walks.

Aggregate data for the Downtown walks (omitted for space) indicate compass errors with a mean near 10° in the open area in the middle of McGill's campus, rising to nearly 30° near large skyscrapers. Compass errors in the Commercial walk (Figure 4) had means more consistently around or below 15° .

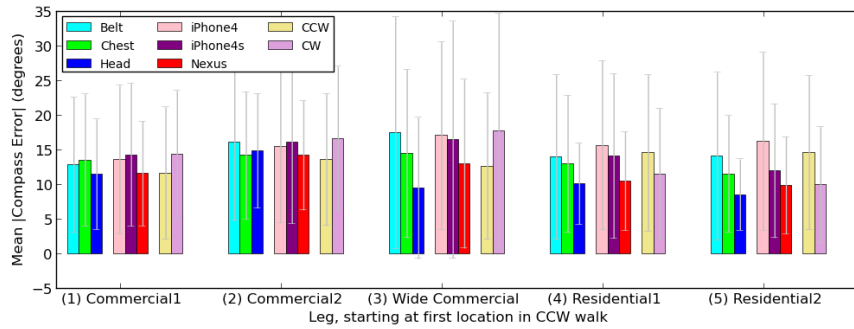


Fig. 4: Commercial walks compass error.

Given natural body sway and alignment deviations from a straight-line path, these error means may seem tolerable for many purposes. However, the standard deviations raise questions as to their reliability. Thus, it is useful to determine whether the devices' estimates of their own accuracy are generally correct, as this would allow us to warn the user when values should not be trusted.

Reported error estimate accuracy iPhone devices provide an estimate of compass error in degrees, whereas Android devices only provide coarse levels of

sensor reliability and are thus excluded from this evaluation. Figure 5 allows us to comment on the accuracy of the error estimates, e.g., if the compass indicates it is within 30° of the actual value, is the reading actually within 30° ? If not, how often does this happen, and how far off is it? Bar values represent the mean number of degrees by which the actual error exceeds (positive bars) or is within (negative bars) the estimated error. The numbers below each set of bars represent the percentage of samples within the reported error. For example, in the first bar, 83% of the samples were within the reported error estimate, and under that estimated error by a mean of approximately 13° . The remaining 17% of the samples exceeded the estimated error by a mean of approximately 9° .

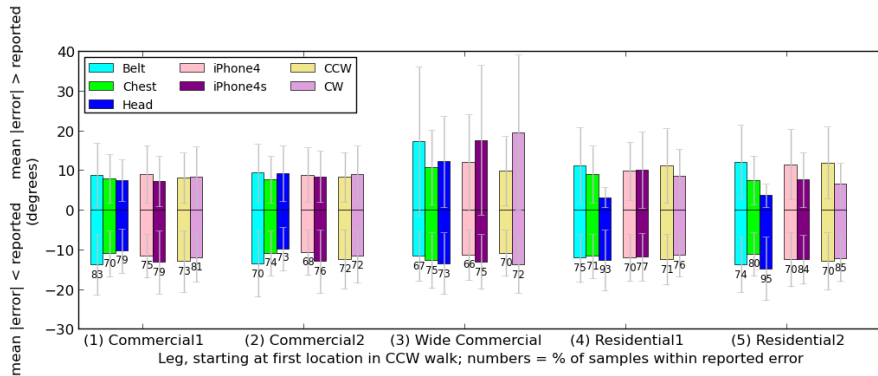


Fig. 5: Commercial walks: compass ground truth accuracy vs. reported error estimates.

Although the $10\text{-}20^\circ$ by which the actual error is seen to exceed the estimates may well be within the achievable tolerance of our experimental setup, the large standard deviations in some areas imply that there will be frequent cases with larger errors. Note that in some legs (e.g., leg 3), the compass error exceeds its estimate 25% or more of the time. The Downtown walks (not shown for reasons of space) performed more poorly outside of the central campus location, with periods where less than 50% of the samples were within the reported error, as well as error exceeding the estimate by more than 50° in some cases.

Yaw drift We had originally hoped that we could use the orientation sensors, primarily the gyro, to obtain better heading estimates by fusing them with the compass (magnetometer) data. We thus implemented a sensor fusion algorithm that recalibrated the gyro to north whenever a compass update was received with reported accuracy at 30° or less. When the compass error exceeded this threshold, we simply ignored further compass readings until the reported error returned below 30° . The hope was that the gyro would be sufficiently accurate to provide interim values. However, the observed yaw drift, which seems to accelerate in each leg, as seen in Figure 6, makes this futile. Downtown drifts range up to

almost $3^\circ/\text{s}$, and up to over 4° in the Commercial walks. Downtown walks were roughly 6.5 minutes each, and Commercial walks were approximately 9 minutes from the time the walker clicked the button to begin the walk, up until the final click at the end. Due to initial calibration time, the sensors may have been started earlier. Even worse, Figure 7a illustrates that gyro drift progressively increases with time. At least on iPhone, shutting down the application allows the gyro to be turned off, and we speculate that it recalibrates on restart. We believe that as a result, yaw drift is typically minimal in the first leg.

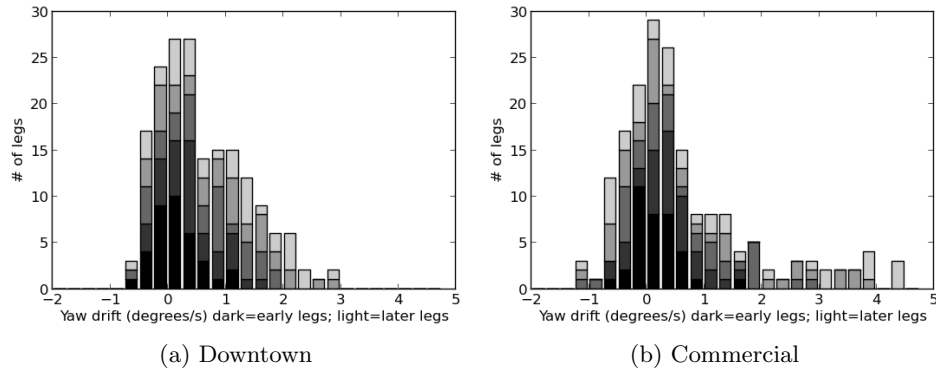


Fig. 6: Histogram of yaw drifts for the two iPhone devices, calculated by a linear regression of the yaw values within each leg. If the yaw sensor behaved perfectly, the drift would always be zero. Note that in both sets, drifts tend to be biased in a clockwise direction (to the right of zero), for which we have no explanation.

We suspect that Google’s Android API only provides device orientation data fused with the magnetometer to better correct for gyro drift. Since the Nexus heading does not seem to perform appreciably better than the iPhones, the benefit of this fusion may be marginal when trying to find a north-calibrated reading. However, such fusion may indeed stabilize the gyro values significantly, useful for applications where only a relative yaw value is needed, not calibrated to north. Apple offers the same option of a fused gyro/compass orientation relative to north starting in iOS5, possibly because they came to the same realization.

4.3 Location accuracy

Unlike heading, location data (Figure 8) was only sampled at 11 and 15 specific ground truth locations for the Downtown and Commercial walks respectively. In good conditions, we generally received updates separated by 1–3 m once the values had stabilized, although they sometimes rise higher. We also noted a “caterpillaring” effect in which an iPhone device would deliver a series of sparser location updates, then cluster a number more closely together, despite

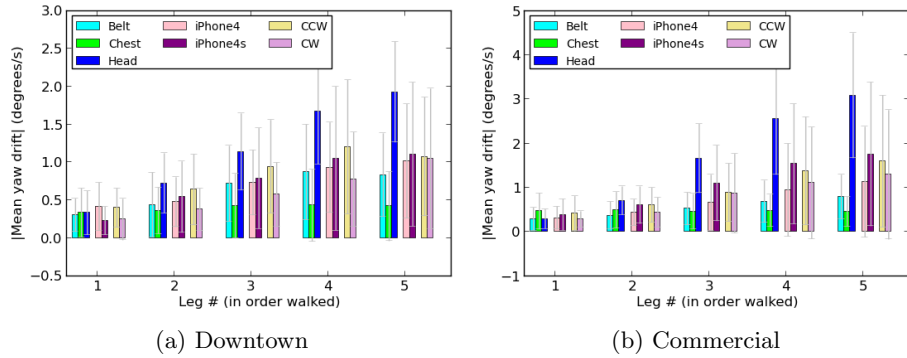


Fig. 7: Yaw drift for the two iPhones. Note that these graphs are arranged by leg in the *order walked* instead of by physical location, since the primary effect of yaw drift is associated with time rather than location. Thus, leg 1 is the first location walked in either CW or CCW direction, with the results averaged over the different corresponding segments. Interestingly, head position performs worse than the other positions/orientations, for which we do not have an explanation.

the walker moving at a consistent speed. These may be due to some sort of smoothing algorithm, with the system giving a series of more sparse points as it tries to “catch up” to a new position, since we see this frequently with the reported location often lagging behind the walker’s actual position.

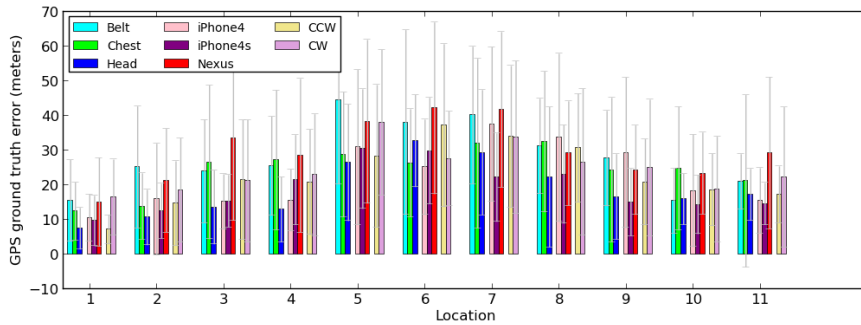


Fig. 8: Downtown walks: location ground truth errors.

As with heading, areas with taller buildings unsurprisingly cause the most difficulty for location accuracy. In these cases, locations may exhibit mean errors in the 30 m range, with maximum error values beyond 60 m in the vicinity of tall skyscrapers. The Commercial walks (not shown for reasons of space), with buildings generally no taller than three stories, exhibit mean error more consistently around 10–15 m, apart from location 10, which was particularly

troublesome for all of the devices. As expected, the Head position tended to outperform the Chest and Belt positions.

Given the relatively large standard deviations, it is crucial to inform the user when the location sensors are having difficulty. Alternatively, such knowledge would allow the application to render information in a more appropriate manner. As before, we evaluate reliability of the location sensing by comparing actual ground truth error with the reported error estimate from the device. Disappointingly, the actual error frequently exceeds the estimated error. For the Downtown walks, the location was within the reported error only 20.4% of the time, and still less than half the time at 46.3% for the Commercial walks.

5 Conclusions

The results of our experiment demonstrate important accuracy issues for the location and orientation sensors found in current smartphones. Their location sensors (using A-GPS) exhibit errors with means of 10–30 m, depending on the surrounding buildings. Compass error frequently exhibits a mean around 10° of ground truth, which is quite good given the other sources of error in our experiment, but in some areas, increases to approximately 30° . The relatively large standard deviations around these means will result in frequently poorer performance experienced by users, e.g., with location errors of 60 m within the standard deviation when walking near skyscrapers. Compounding these problems, reported error is frequently underestimated, reducing trustworthiness. Improving the results with sensor fusion, e.g., by leveraging the gyro to improve heading estimates, is problematic due to difficulty in determining a “good” initial compass reading. In addition, we observed significant gyro drift over time, accelerating to over $4^\circ/\text{s}$ in some cases. Practically, we conclude that augmented reality applications that rely on better location or heading accuracy than indicated by these results will be difficult at best to realize on current smartphone hardware if relying on raw sensor values. Further mining across our data set may also reveal additional correlations that would help point to the specific conditions under which the sensors are inaccurate.

Until smartphone sensors improve, we are exploring alternative rendering methods that can cope with poor accuracy and error estimation, as well as algorithms such as “snapping” the user to a nearest street to overcome location errors. In many cases, these subjectively appear to improve accuracy, but formal testing is required. Although such methods show promise, they also generate new failure cases, such as worsening the location error by snapping to an incorrect street when nearing an intersection. Snapping can be effective for those location errors lateral to the street, but not for errors in the direction of travel, which appear to occur more frequently, often with the reported location lagging the user’s actual position. In terms of implications to our own research we are now aware of the importance of carefully picking the areas in which we test our application, and treat the error estimates for both orientation and location sensors with a dose of skepticism.

Acknowledgements. This work was possible thanks to the financial support of the Québec Secrétariat du Conseil du trésor through the *Appui au passage à la société de l'information* program, as well as additional funding from a Google Faculty Research Award. We gratefully acknowledge Mathieu Bouchard's Android work, as well numerous discussions with him and other Shared Reality Lab members concerning experiment design, implementation and analysis.

References

1. S. Azenkot, R.E. Ladner, and J.O. Wobbrock. Smartphone haptic feedback for nonvisual wayfinding. In *Computers and Accessibility*, pages 281–282. ACM, 2011.
2. J. Blum, M. Bouchard, and J. Cooperstock. What's around me? Spatialized audio augmented reality for blind users with a smartphone. In *MobiQuitous*, volume 104 of *LNCS*, pages 49–62. Springer, 2011.
3. H. Blunck, M. Kjærgaard, and T. Toftegaard. Sensing and Classifying Impairments of GPS Reception on Mobile Devices. In *Pervasive Computing*, volume 6696 of *LNCS*, pages 350–367. Springer Berlin Heidelberg, 2011.
4. I. Constandache, R. Choudhury, and I. Rhee. Towards mobile phone localization without war-driving. In *INFOCOM*, pages 2321–2329. IEEE, 2010.
5. B. Gotow, K. Zienkiewicz, J. White, and D. Schmidt. Addressing Challenges with Augmented Reality Applications on Smartphones. In *MOBILWARE*, pages 129–143, 2010.
6. W. Jones. A compass in every smartphone. *Spectrum, IEEE*, 47(2):12–13, 2010.
7. M. Kjærgaard, J. Langdal, T. Godsk, and T. Toftkjær. EnTracked: energy-efficient robust position tracking for mobile devices. In *Mobisys*, pages 221–234, NY, 2009. ACM Press.
8. R. Ozcan, F. Orhan, M. Fatih Demirci, and O. Abul. An Adaptive Smoothing Method for Sensor Noise in Augmented Reality Applications on Smartphones. In *Mobile Wireless Middleware, Operating Systems, and Applications*, pages 209–218. Springer Berlin Heidelberg, 2012.
9. N. Priyantha, A. Miu, H. Balakrishnan, and S. Teller. The cricket compass for context-aware mobile applications. In *MobiCom*, pages 1–14, NY, 2001. ACM.
10. L. Ruotsalainen, H. Kuusniemi, and R. Chen. Heading change detection for indoor navigation with a Smartphone camera. In *Indoor Positioning and Indoor Navigation*, pages 1–7. IEEE, September 2011.
11. L. Ruotsalainen, H. Kuusniemi, and R. Chen. Visual-aided Two-dimensional Pedestrian Indoor Navigation with a Smartphone. *Global Positioning Systems*, 10(1):11–18, June 2011.
12. A. Thiagarajan, J. Biagioni, T. Gerlich, and J. Eriksson. Cooperative transit tracking using smart-phones. In *Embedded Networked Sensor Systems*, pages 85–98, New York, NY, 2010. ACM.
13. C. Vaitl, K. Kunze, and P. Lukowicz. Does On-body Location of a GPS Receiver Matter? In *Body Sensor Networks*, pages 219–221. IEEE, June 2010.
14. S. von Watzdorf and F. Michahelles. Accuracy of positioning data on smartphones. In *Workshop on Location and the Web*, pages 2:1–4, New York, NY, 2010. ACM.
15. M. Wing. Consumer-Grade GPS Receiver Measurement Accuracy in Varying Forest Conditions. *Research Journal of Forestry*, 5(2):78–88, February 2011.
16. P. Zandbergen. Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. *Transactions in GIS*, 13:5–25, June 2009.