Indoor Positioning and Floor Plan Based Ground Truth: Can You Really Click Where You Are?

Andrei Popleteev
SnT, University of Luxembourg
Luxembourg City, Luxembourg
andrei.popleteev@uni.lu

ABSTRACT

The increasing accuracy of indoor positioning systems requires an appropriately accurate evaluation, which compares system outputs with the known coordinates of test locations — the ground truth. Although ground truth data are rarely (if ever) tested, they are traditionally assumed to be perfectly accurate. However, even small errors introduced by inaccurate ground truth need to be taken into account for fair evaluation and comparison between modern high-resolution positioning systems.

In this paper we analyze the quality of ground truth data provided by clicking on an interactive floor plan (a method employed by such classical systems as RADAR and Horus). Experimental results show that this method has high precision but low accuracy, and high systematic errors make it unsuitable for evaluation of fine-grained localization systems.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; H.1.2 User Machine Systems: Human factors; H.3.4 Systems and Software: Performance Evaluation

Author Keywords

Indoor localization; ground truth; floor plan; accuracy; performance evaluation; benchmarking; uncertainty; resolution; level of detail; granularity; BIM; HCI

INTRODUCTION

With the advance of indoor positioning research, state-of-the-art systems start to surpass the 1-meter accuracy milestone and accuracies of the best systems differ only by few centimeters [3]. At this scale, benchmarking results and comparative rankings can be influenced by the fine nuances of the evaluation methodology — in particular, by the quality of the ground truth data.

MUM '16, December 12 - 15, 2016, Rovaniemi, Finland

ISBN 978-1-4503-4860-7/16/12…$15.00

DOI: http://dx.doi.org/10.1145/3012709.3012737

Ground truth (GT) is critically important for evaluation of indoor positioning systems, as their performance is measured by the differences between the GT and the system outputs. In automated tests, GT is provided by robot-based benchmarking platforms with well-documented accuracy [18, 7, 12]. More commonly, however, benchmarking is performed manually: an experimenter brings the test device into predefined locations as accurately as possible [14], or carries it around the testbed, providing the GT by marking their current position on an interactive floor plan.

While floor plan based GT has been widely used in experimental practice (for instance, in the classical RADAR [6] and Horus [20] papers) the accuracy of such GT data is unknown. This is a critical knowledge gap, since only few centimeters distinguish an award-winning and simply very good system, and this difference could be easily introduced by inaccurate GT. Thus, understanding the limits of the GT measurement methodology is crucial for adequate evaluation of indoor positioning systems.

This paper provides an evaluation of the errors of floor plan based ground truth. In particular, we experimentally quantify the human ability to accurately and consistently pinpoint indoor locations on an interactive floor plan. We distinguish the systematic and random error components in repetitive tests, and report them both in terms of physical and screen space.

BACKGROUND

Evaluation of indoor positioning systems includes an analysis of the system outputs with regard to the true position of the test device. The procedure can be an automatic or a manual one, depending on who or what moves the test device between the test points or along the test paths.

Automatic benchmarking solutions, powered by robotic platforms, acquire the GT data from robot's high-accuracy reference positioning system. Examples of such platforms include EVARILOS [18], RAWSEEDS [7] and ALPS [12]. Their GT accuracy is usually well-documented, with average errors in the range of 15 to 25 cm [14].

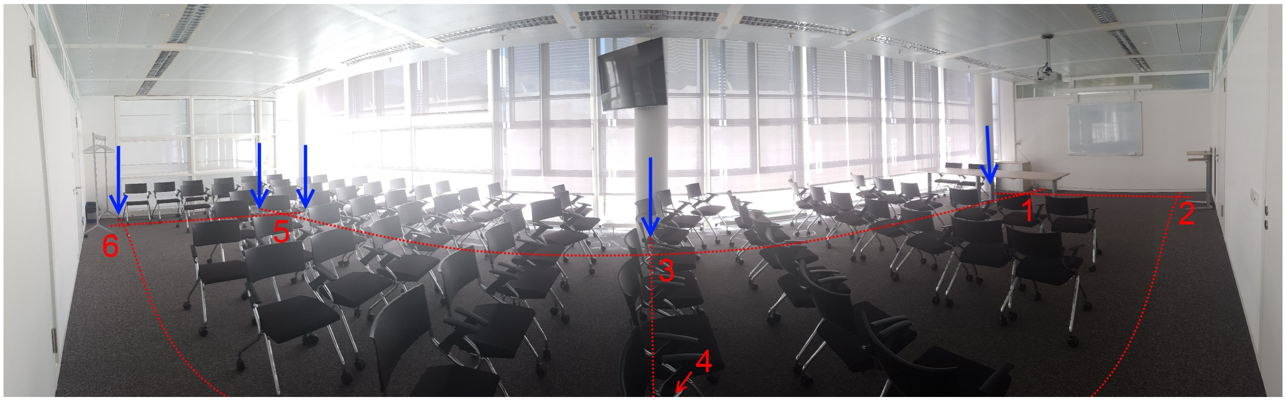

Figure 1. Experimental testbed. Numbered test points are defined by the virtual grid (red dotted lines) aligned with the environmental landmarks (pillars and window frames, indicated by blue arrows). Only the landmarks were present in the testbed, the test points were not marked. (Please note that the straight grid lines are displayed bent due to the perspective distortion of the photo.)

Manual evaluation, in turn, is performed personally by the experimenters and represents the *de facto* standard of the evaluation procedure. Since the task requires more time than expertise, it is typically delegated to students or research assistants. Unfortunately, the accuracy of such human-defined GT is not well documented [14]. Even worse, it is rarely even taken into account: experiment descriptions either provide ballpark estimates, assume the GT error to be negligible, or even omit the GT methodology altogether [4].

The quality of user-provided spatial data has previously been discussed in the context of geographic information systems [8]. However, in this paper we focus on a smaller-scale scenario, exploring the human limits of floor plan based localization in indoor positioning experiments. This issue has recently been addressed in [14], where the author took a mechanistic perspective on the experimenters, evaluating them as a "human-based indoor positioning system" [14]. The study analyzed the positioning performance of people with a handheld smartphone, measuring their ability to place that smartphone exactly into the test location defined by visual clues such as floor markers, ceiling markers and environmental landmarks. The paper reported absolute errors of 22 cm for markers and 36 cm for environmental landmarks.

This work extends [14] and focuses on interactive floor plans — another method of defining the GT, which has been used in evaluation of such systems as Horus [20], RADAR [6] and Zee [15]. There, the experimenters moved the test device inside the indoor environment and periodically labeled their GT location on a computer-based floor plan. Unfortunately, information about error characteristics of the floor plan based GT is currently limited to the following estimate of the Horus' authors: "We expect an error of about 15–20 cm due to the inaccuracies in clicking the map." [20].

The following sections address the gap and provide an analysis of the floor plan based GT error characteristics, based on empirical data.

EXPERIMENTAL SETUP

The experiment has been conducted in a 16×6 m conference room shown in Figure 1. We specified six test points located in the nodes of the virtual grid defined by building pillars and window frames (see Figure 1).

The floor plan of the environment was scaled and rotated to align with the satellite view of the building and augmented with the pan-and-zoom functionality provided by the Leaflet library [5]. This interactive floor plan was displayed on a MacBook Pro (Retina, 13-inch, Early 2015) laptop as shown in Figure 2. By default, the complete floor plan was visible on the screen at the scale of 1:108 (1 cm of the screen displaying 108 cm of the testbed space, or about 1.2 cm of testbed space per pixel).

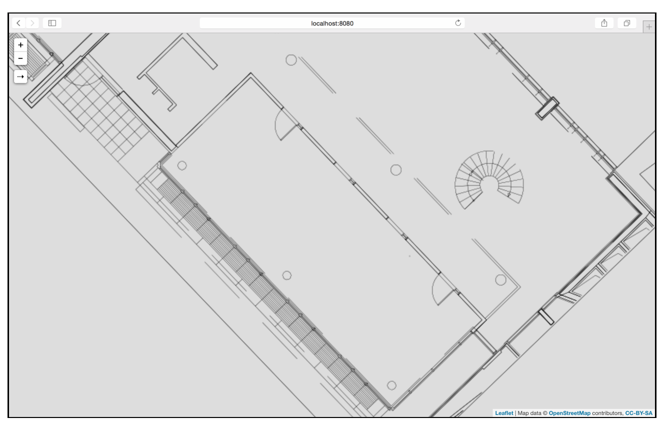

Figure 2. User interface screenshot.

Interaction with the floor plan was performed using the laptop's touchpad. The laptop-based setup was chosen instead of a more modern touchscreen interface in order to provide pixel-perfect pointing and to reduce input errors [9] (in particular, those introduced by the so-called "fat finger" problem [16]). While the pointing accuracy could be further improved by mouse-based interaction [10], mouse manipulations require a suitable flat surface which is typically not available during the localization experiments.

The participants — seven graduate students with computer science background — were instructed to mark the locations of test points on the floor plan using the laptop's touchpad. All the volunteers were informed about the purpose of the study and were asked to show their best-effort performance. Each participant completed three passes over the testbed, thus providing three estimates for each point. None of the participants changed the default zoom level of the floor plan (despite being explicitly advised about the possibility), likely because zooming in would hide some of the visual references (in particular, window frames) beyond the screen boundaries.

Ground truth positions of the test points within the testbed were established using laser rangefinder [1]; the calculated distances enabled us to identify test points' locations on the floor plan with a pixel-level accuracy (that is, with 1.2 cm granularity). Both GT and participants' inputs were recorded using the WGS-84 coordinate system [2] with 7-digit precision, which allows for 1.1 cm spatial resolution. Distances between the coordinates defined in WGS-84 system were calculated using Karney's geodesic method [11], which provides sub-millimeter accuracy. Combining the above factors — ± 0.2 cm rangefinder error, 1.2 cm floor plan granularity, and 1.1 cm coordinate system resolution — we estimate the GT error to be within 2.5 cm.

RESULTS

Positioning error of all the participants is shown in Figure 3. Only half of user estimates were within 40 cm from their true coordinates (3.7 mm on the screen), while 95% of the clicked points were up to 91 cm away, and the maximum error reached 109 cm (8.4 mm and 10 mm on the screen, respectively).

A closer look into the performance of each participant reveals the importance of the personal factors (see Figure 4). Since we could not control such parameters as spatial awareness and attention level of the volunteers (beyond asking them for a best-effort performance), there were significant differences between their personal results ($p < 0.01$ in Welch's t-test [19] used throughout this section).

As any measurement error, the errors reported so far can be divided into two components: systematic error and random error [17]. The *systematic error* reflects the accuracy of the user input, that is the difference between the mean of the user-provided estimates and the true coordinates. The *random error*, in turn, reflects the precision (spread) of the user input, — that is, the participant's ability to consistently click the same point in repeated tests. This value is estimated as the standard

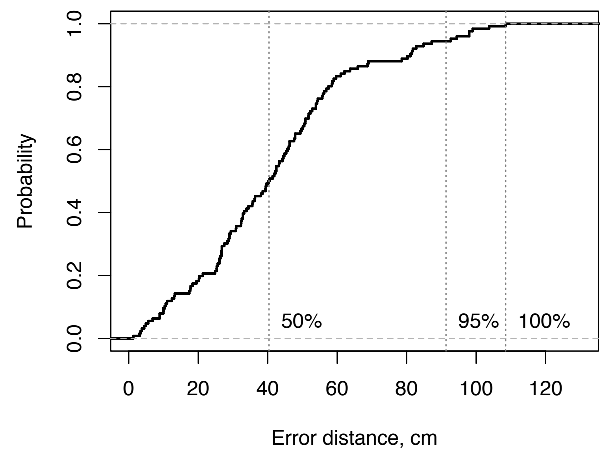

Figure 3. Positioning error of all the participants.

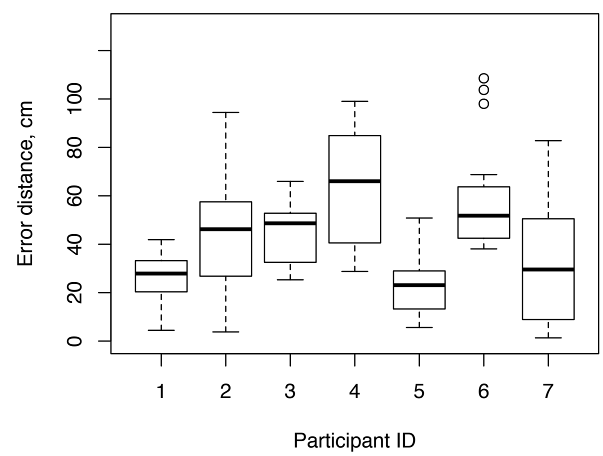

Figure 4. Positioning error of each participant.

distance deviation (SDD), that is the average distance between the three estimates and their mean [13, Eq. 8.7].

While the distribution of the systematic error (shown in Figure 5) was similar to the general error distribution (Figure 3), the random error component was much more constrained. Indeed, according to Figure 6, most of the participants were consistent in their estimates: in 95% of the cases the SDD was below 19 cm (1.8 mm on the screen).

Interestingly, the spreading of user estimates significantly varied among the test points (Figure 7): the random error in points #2 and #4 was significantly ($p < 0.01$) higher then in other points. This can be explained by the fact that points #2 and #4 were away from the visual references defining the test points (pillars and window frame). At the same time, there were no significant differences ($p > 0.01$) in systematic errors between any of the test points.

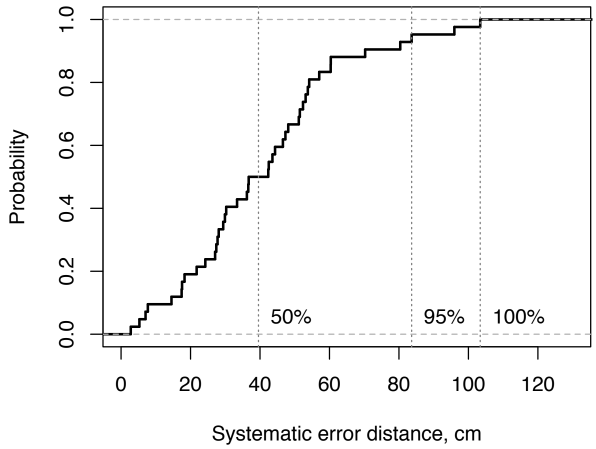

Figure 5. Systematic error component — all participants.

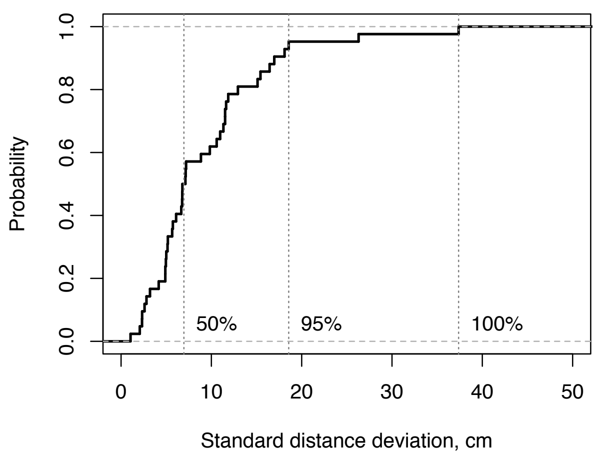

Figure 6. Random error component — all participants.

Limitations

There are a number of factors that limit the generalization of the reported results. Firstly, our experiment was performed with a limited number of participants in a small-scale testbed. Secondly, the input method (touchpad) was assumed to be error-free. Finally, the actual error values heavily depend on the specific details of the specifics of the experimental setup, such as input method, display (size and resolution), and floor plan itself (its type, resolution, level of detail and even factual correctness).

For instance, augmenting the floor plan with auxiliary markers (such as coordinate grid) might considerably improve the accuracy of GT data. However, since detailed analysis of these factors was beyond the scope of this short paper, their impact on the GT error remains open for future work.

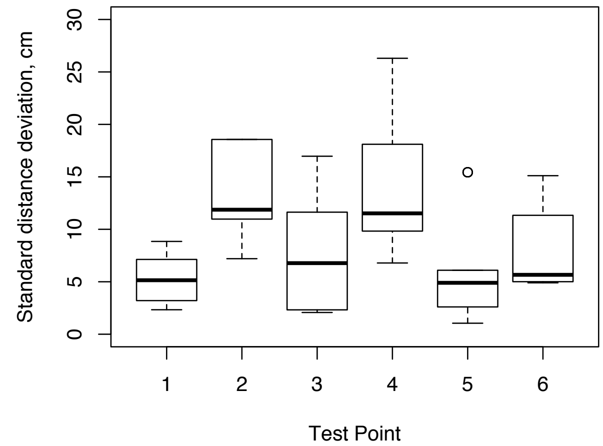

Figure 7. Random error (precision) for each test point.

CONCLUSION

We evaluated the quality of indoor positioning ground truth data based on an interactive floor plan. The results suggest that the experimenters struggle to correctly identify real-world test points on a floor plan: in our experiments, median systematic error reached 0.4 m, while its absolute values sometimes exceeded 1 m (3.4 mm and 10 mm on the screen, respectively). Nevertheless, the participants were surprisingly consistent in clicking the same locations in repeated test (95% of deviations were within 1.8 mm on the screen).

We conclude that such a high systematic error considerably limits the applicability of floor plan based ground truth methodology, making it suitable only for evaluation of coarse-grained positioning systems.

ACKNOWLEDGMENTS

The author would like to thank the anonymous reviewers for their remarkably helpful and constructive feedback.

This study was supported by the National Research Fund of Luxembourg (INDOORS project, grant number C14/IS/8311593).

REFERENCES

1. *Bosch PLR 50C Digital Laser Measure.* `https://www.bosch-do-it.de/media/media/garden/gardenmedia/manuals/1083577_160992A1F3_201508pdf..pdf`
2. 1997. *Department of Defense World Geodetic System 1984: Its Definition and Relationships With Local Geodetic Systems.* Technical Report TR8350.2. National Imagery and Mapping Agency.
3. 2016. Microsoft Indoor Localization Competition. Online: `http://research.microsoft.com/en-us/events/msindoorloccompetition2016/default.aspx.` (2016). (Collocated with IPSN 2016).
4. Stephan Adler, Simon Schmitt, Katinka Wolter, and Marcel Kyas. 2015. A Survey of Experimental

Evaluation in Indoor Localization Research. In *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN-2015)*.

5. Vladimir Agafonkin. 2016. Leaflet - a JavaScript library for interactive maps. `http://leafletjs.com`. (2016).
6. P. Bahl and V. N. Padmanabhan. 2000. RADAR: An in-building RF-based user location and tracking system. In *Proceedings of 9th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2000)*, Vol. 2. 775–784.
7. Simone Ceriani, Giulio Fontana, Alessandro Giusti, Daniele Marzorati, Matteo Matteucci, Davide Migliore, Davide Rizzi, Domenico G Sorrenti, and Pierluigi Taddei. 2009. Rawseeds ground truth collection systems for indoor self-localization and mapping. *Autonomous Robots* 27, 4 (2009), 353–371.
8. Michael F. Goodchild. 2008. Spatial accuracy 2.0. In *Proceedings of the 8th International Symposioum on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*.
9. Niels Henze, Enrico Rukzio, and Susanne Boll. 2011. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *Proceedings of MobileHCI–2011*. ACM.
10. Morten Hertzum and Kasper Hornbæk. 2010. How age affects pointing with mouse and touchpad: A comparison of young, adult, and elderly users. *Intl. Journal of Human–Computer Interaction* 26, 7 (2010), 703–734.
11. Charles F. F. Karney. 2013. Algorithms for geodesics. *Journal of Geodesy* 87, 1 (2013), 43–55. DOI:`http://dx.doi.org/10.1007/s00190-012-0578-z`
12. Patrick Lazik, Niranjini Rajagopal, Oliver Shih, Bruno Sinopoli, and Anthony Rowe. 2015. ALPS: A Bluetooth and Ultrasound Platform for Mapping and Localization. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 73–84.
13. Saroj K. Pal. 1998. *Statistics for Geoscientists: Techniques and Applications*. Concept Publishing Company.
14. Andrei Popleteev. 2016. HIPS: Human-based Indoor Positioning System. In *Proceedings of the 7th International Conference on Indoor Positioning and Indoor Navigation (IPIN-2016)*. IEEE.
15. Anshul Rai, Krishna Kant Chintalapudi, Venkata N Padmanabhan, and Rijurekha Sen. 2012. Zee: Zero-effort crowdsourcing for indoor localization. In *Proceedings of Mobicom-2012*. ACM, 293–304.
16. Katie A Siek, Yvonne Rogers, and Kay H Connelly. 2005. Fat finger worries: how older and younger users physically interact with PDAs. In *Proceedings of INTERACT-2005*. Springer, 267–280.
17. John Robert Taylor. 1999. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, p. 94.
18. Tom Van Haute, Eli De Poorter, Filip Lemic, Vlado Handziski, Niklas Wirström, Thiemo Voigt, Adam Wolisz, and Ingrid Moerman. 2015. Platform for benchmarking of RF-based indoor localization solutions. *Communications Magazine, IEEE* 53, 9 (2015), 126–133.
19. Bernard L Welch. 1947. The generalization of "Student's" problem when several different population variances are involved. *Biometrika* 34, 1/2 (1947), 28–35.
20. Moustafa Youssef and Ashok Agrawala. 2005. The Horus WLAN location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services (MobiSys-2005)*. 205–218.