2012 International Conference on Indoor Positioning and Indoor Navigation, 13-15th November 2012

Using Natural Footstep-Accurate Traces for Indoor Positioning Evaluation

Tim Schwartz, Christoph Stahl, Angela Mahr, Christian Müller
DFKI GmbH
German Research Center for Artificial Intelligence
Saarbrücken, Germany
Email: {*firstname.lastname*}@dfki.de

Bahjat Saliba
Saarland University
Dept. for Computer Science
Saarbrücken, Germany
Email: Bahjat.Saliba@googlemail.com

***Abstract*—In this paper, an evaluation method using natural footstep-accurate traces as ground truth is proposed. Said ground truth was obtained by video-taping people who were naturally walking in the foyer of our company building. These video-clips were then manually analyzed, using the tiles of the foyer's floor as a coordinate system. The result of this analysis are not simply the positions of the filmed users as they progress through the foyer, but the coordinates of each single footstep of each user including timestamps for each step. To evaluate the position accuracy of an indoor positioning system, these traces were laid out on the floor and 're-walked' while carrying the hardware for the positioning system. A short sound was played according to the timestamps of each recorded step of the ground truth, to ensure that the walking speed was close to the real walking speed. The distances of these 'target system traces' to the ground-truth traces were then calculated and analyzed.**

***Keywords*—Positioning Evaluation, Natural Ground Truth, Active RFID.**

I. INTRODUCTION

With a plethora of indoor positioning systems available, their evaluation becomes more and more important. Common evaluation methods in literature can be classified into the following three categories:

1) *Static evaluation:* In a test field several points with known coordinates are chosen as evaluation points (or reference points). The entity (e.g. the needed hardware) of which the position is to be determined, is brought to these evaluation points and the distance to the known coordinates is calculated.
2) *Dynamic evaluation with predefined geometrical paths:* Easy-to-follow geometrical paths like lines, rectangles or circles are defined for the test field and then followed by a machine or a human. The obtained coordinates of the positioning system are then compared to the predefined paths.
3) *Dynamic evaluation using a reference positioning system:* A positioning system with a supposedly higher accuracy is used as a baseline for the evaluation of the target system (i.e. the system that is to be evaluated).

A static evaluation (*category 1*) is usually very easy to conduct, but also lacks to measure all the inaccuracies and errors that may occur if an entity is moving through an area. Therefore a dynamic evaluation should be preferred for any positioning system that is planned to be used with moving entities. Predefined geometrical paths (*category 2*) however do not correspond to the natural movement patterns of humans and could thus have an impact on the evaluation result. With these considerations in mind, using a reference positioning system (*category 3*) seems to be a good idea. However, since this reference system has to be evaluated first, a kind of 'chicken and egg' situation is created. Moreover, for both dynamic evaluation methods (*category 2 & 3*) and depending on the design of the evaluation, users might consciously or unconsciously adapt their movements to the output of the tested positioning system, e.g. walking slower or walking towards the positions indicated by the system.

In order to address these problems, we propose an evaluation approach that uses natural step-accurate traces as ground truth, which is obtained by manually analyzing video clips of people moving through an indoor area while following their everyday tasks. This method of ground-truth acquisition has the advantage that moving patterns are not influenced by carrying additional hardware or by the awareness of being part in a positioning experiment. In the rest of this paper, we will describe the details of performing such an evaluation and give a practical example of its outcome when applied to an indoor positioning system based on active RFID tags and infrared beacons.

II. RELATED WORK

Of all the papers dealing with indoor positioning systems and their evaluation, two stand out as they put their focus on the evaluation method itself rather than on the position system.

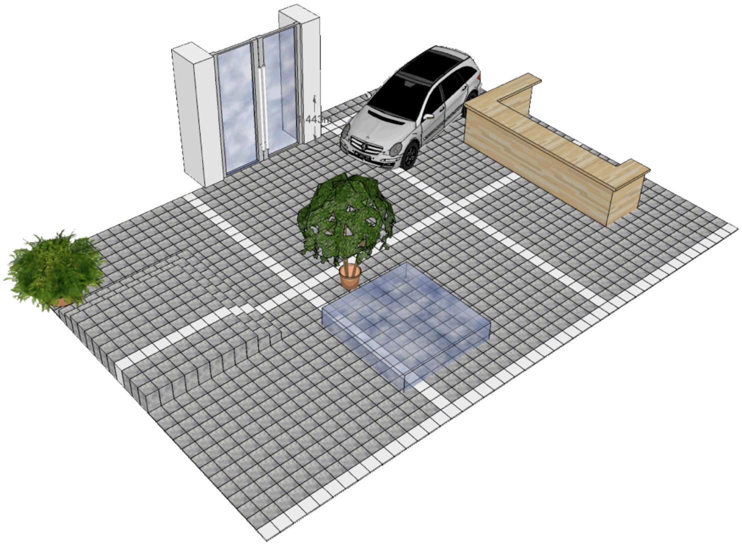
In [1] Stephan et al. describe the 'evaluation of indoor positioning technologies under industrial application conditions'. They tested two commercially available positioning systems, namely Ubisense[1] (see also [2]) and the MIT Cricket Indoor Location System system (see [3]), which is distributed by Crossbow Technology[2]. The evaluation method itself can

---

[1] http://www.ubisense.net

[2] http://bullseye.xbow.com:81/Products/productdetails.aspx?sid=176

(a) DFKI foyer (b) DFKI foyer top view (c) DFKI foyer open architecture



(d) 3D model of the DFKI foyer

Fig. 1: The foyer of DFKI Saarbrücken was used a test field, since it provides a large area and a visual coordinate system through the tiles.

be classified as static, i.e. reference points were determined using a tachymeter, which resulted in a reported accuracy of ±2 millimeters. The determined positions of the Ubisense and Cricket systems were then compared to these reference points. The evaluation process however differs in an important aspect from other static approaches: the systems were tested under two different conditions – optimal conditions and a realistic scenario. The realistic scenario was carried out inside a complete production facility, including metal structures, piping, glass vessels and heavy machinery. While both systems operated roughly as advertised under optimal conditions, the accuracy was significantly lower in the realistic scenario. Although our evaluation is a dynamic one, it was also carried out under realistic conditions (although not in a factory setting but in an office building).

In [4] Rydell et al. describe a dynamic evaluation method using a camera-based reference system to test a foot-mounted inertial navigation system. Their method can thus be classified as being of the third category (as described in Section I). The camera for the reference system is worn by the user and connected to a small laptop, which is carried in a backpack. This reference system determines the current user position with the help of optical markers, which are distributed in the environment. The authors address the aforementioned 'chicken and egg' problem by using a *third* positioning system, the Vicon motion capture system[3], which – according to the authors – is limited to a single room. With the help of the Vicon system, the accuracy of the camera-based system was evaluated using one optical marker and testing different distances to that marker. At distances of 1 and 1.75 meters, the measured positioning error was below 10 centimeters. Additionally, the authors tested the reliability of the marker detection itself, and found out that at reasonable camera speeds all markers were detected and correctly identified. Unfor-

---

[3] http://www.vicon.com

tunately, the actual process of evaluating the target system with the help of the evaluated reference system is not further described. Although our evaluation method is also camera based, our camera is mounted in the environment and in order to obtain ground truth, participating users do not have to carry any hardware and do not even have to be aware that they are participating in a positioning experiment.

III. THE TARGET POSITIONING SYSTEM: LORIOT

Although the focus of this paper lies on the evaluation method itself and not on the evaluated positioning system, it might be helpful to understand the basic functionality of the tested system. The system is called LORIOT, which is an acronym for *L*ocation and *OR*ientation in *I*ndoor and *O*utdoor environmen*T*s. LORIOT is an onboard/egocentric Always Best Positioned system, which uses active RFID tags, infrared (IR) beacons and GPS (outside of buildings) to determine its own position. Here, *onboard* means that all necessary calculations are done on the user-device itself and *egocentric* means that the device performs all necessary measurements itself. The term *Always Best Positioned* describes the ability of the system to either work with RFID tags, IR beacons or GPS alone, or to automatically combine them to obtain a better positioning accuracy (more on this terminology can be found in [5]). Thus, in order to use the system indoors, the environment has to be instrumented with active RFID tags and/or IR beacons. Furthermore, the RFID tags are also used to store their own coordinates as well as the coordinates and identification codes of nearby IR beacons. Consequently, if an environment is instrumented with RFID and IR, the system can obtain all information necessary to position itself directly from the environment, whereas if only IR beacons are installed, a list containing their identification codes and coordinates has to be downloaded beforehand. LORIOT and the used algorithms (so-called geo-referenced dynamic Bayesian networks) are described in detail in [6].

One of the reasons for the evaluation was the claimed *Always Best Positioned* property of the system, i.e. it should be tested whether the system achieves a higher accuracy when RFID tags and IR beacons are combined as opposed to using just one technology.

IV. EVALUATION DESIGN

As pointed out earlier, we aimed for a natural ground truth, i.e. the moving-paths of persons should not be influenced by any means. Early ideas to use special shoes, which leave printed marks on the floor, were therefore discarded, as telling people to wear these shoes could already influence their behavior. Likewise, any other 'instrumentation' of persons was also rejected. Using a smart floor that can electronically sense footsteps (e.g. the SenseFloor by Future Shape[4]) would have been an option, but was too expensive. We therefore opted to record video clips of moving people and to analyze these clips manually (described in detail in Section IV-A).

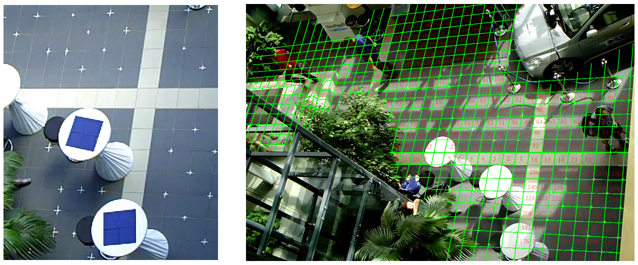

(a) Adhesive tape was used to enhance the visibility of the tile seams.
(b) Video-overlay representing the coordinate system.

Fig. 2: A grid overlay was used to annotate each step of a person with according coordinates.

In short, the evaluation consists of three steps:

1) *Obtaining ground truth* by recording video clips of people moving through a natural indoor environment and extracting foot-step accurate positions from these clips,
2) *Obtaining system traces* by marking these traces on the floor and following them while carrying the needed hardware for the target positioning system,
3) *Calculating the error distance* between ground truth and the obtained system traces. Each step will be described in detail in the following subsections.

A. Obtaining Ground Truth

In order to obtain the needed ground truth, an appropriate test field had to be found which allowed for the unobstructed recording of moving persons inside a building. With respect to the tested positioning system, the test field should also incorporate known error sources, which have a non-beneficial effect on the accuracy of the system. In the case of LORIOT for example, the main error source is possible overreach of far-away RFID tags, which are most likely to happen in large open areas without attenuating walls.

For our evaluation, we chose the main foyer of DFKI building in Saarbrücken as a test field for several reasons: it provides such a large, open area in which many people move while entering and leaving the building; the area is easily observable, since it is not obstructed by intermediate ceilings; and the floor tiling can be used as a visible coordinate system.

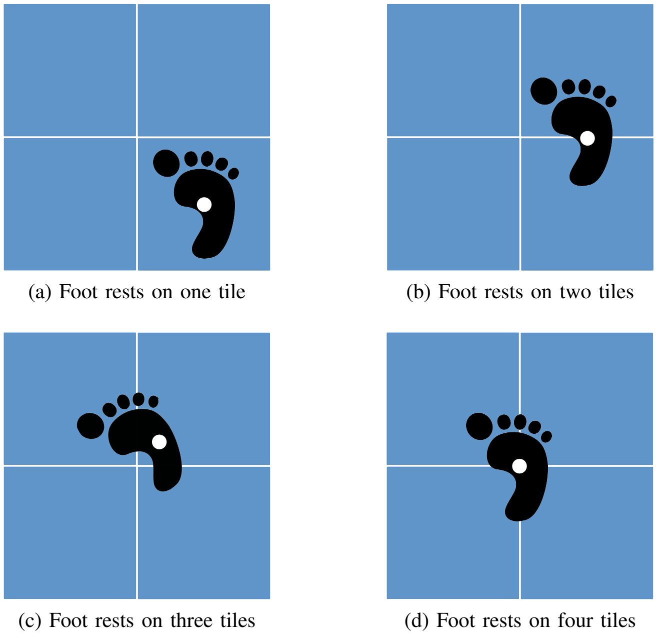
Figure 1 shows several views on the foyer, where Figure 1b was obtained while standing on the top 'balcony' as seen in Figure 1c. With the permission of DFKI's workers' counsel, we installed an HD video camera on that balcony, such that a large part of the foyer could be observed.

We filmed people walking through the area over the course of three days for approximately 0.5 hours per day around lunch time. The time of day was chosen because it ensures high pedestrian traffic and it protects the privacy aspects of the recorded workers (as this data does not allow conclusions on when somebody arrived to or left from work)[5].

---

[4] http://www.future-shape.com/en/technologies/23/sensfloor

[5] Moreover, the recorded video clips were erased after the positioning data was extracted



(a) Foot rests on one tile

(b) Foot rests on two tiles

(c) Foot rests on three tiles

(d) Foot rests on four tiles

Fig. 3: Four basic cases were considered for obtaining coordinates of each step of a person.

This ‘near birds eye view’ in combination with the visible tile seams on the floor allowed us to manually extract the positions of each step of the observed workers. An accurate 3D model of the foyer was created using Google SketchUp[6], which also represents each tile, as can be seen in Figure 1d.

To enhance the visibility of the tile seams, white adhesive tape was applied at selected spots (as can be seen in Figure 2a). The tile seams and marked spots were used to overlay a grid of green lines on the video clips, to enhance the visibility of each tile. The grid also contained a unique ID for each tile.

To derive numerical coordinates for each single footstep of the recorded persons, the enhanced videos were manually analyzed. The quality of the video clips was high enough to discriminate four basic cases for each step, depending on how many tiles a person’s foot is resting on. Figure 3 shows these four different cases, where such a depicted cluster of four tiles measures 60 × 60 centimeters (including the seams). The actual coordinates were then derived by using the coordinates of the middle point of each covered tile and calculating the geometric middle according to the formula:

$$x = \frac{1}{n}\sum_{i=1}^{n} x_{\mathrm{id}_i},\ y = \frac{1}{n}\sum_{i=1}^{n} y_{\mathrm{id}_i} \qquad (1)$$

where n is the number of tiles covered and x_{id_i} and y_{id_i} are the x and y coordinates of the middle point of a tile with identification id_i. The white dots in Figure 3 indicate the resulting coordinates for each case. This leads to an average position accuracy of 15 cm per foot step (shoes for adults vary between 20 and 34 cm in length).

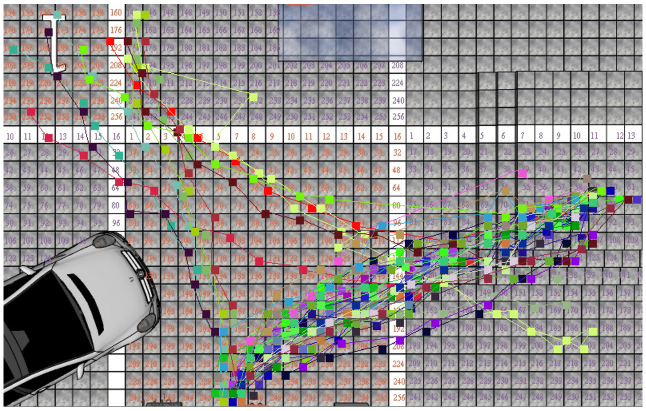


Fig. 4: Visualization of all traces obtained in one day.

Using this method, we extracted the coordinates and time-stamps for every single footstep of a recorded person. This lead to 119 highly accurate ground-truth traces. A tool was implemented to visualize the recorded traces and to perform evaluation-calculations (see also [7]). Figure 4 shows the visualization of all traces obtained in one day as an example.

B. Obtaining System Traces

To keep the ground-truth traces as natural as possible, none of the recorded persons wore a mobile device. Thus, the acquisition of the system traces, i.e. LORIOT’s calculated positions, had to be done in a separated step. In this step 58 active RFID tags were placed on the floor of DFKI foyer, with a distance of 105 centimeters between two adjacent tags. Coordinates of each tag were stored on their internal memories using the same coordinate system as in the ground-truth acquisition process. In addition, 10 IR beacons were placed in the environment using microphone stands.

From the 119 available ground-truth traces, 16 were randomly chosen. These traces were laid out one after the other, according to the coordinates obtained in the ground-truth acquisition process.

Figure 5 shows one such trace. Each trace was then followed step by step while carrying a mobile device with LORIOT running. Each trace was followed two times with two different speeds:

1) Original speed of the recorded trace. This was accomplished by playing back beeps according to the original time-stamps of the trace.
2) In a very slow speed, where after each step a pause of approximately one second was made.

The LORIOT system was modified to log all calculated positions, their time-stamps and raw sensor-data into text files. This process led to 32 log files including derived positions and all measured raw sensor-data.

From each log file, five system traces were derived by using LORIOT’s positioning algorithm in varying conditions: considering only IR beacons, considering only RFID tags

---

[6] Now part of Trimble http://www.sketchup.com

Fig. 5: Traces were laid out on the floor and followed while carrying a mobile device running LORIOT.

without caching, considering RFID tags & IR beacons without caching, considering only RFID tags including caching and considering RFID tags & IR beacons including caching. Caching here means that RFID data that had already been read (coordinates of tags, see Section III) will be retrieved from a cache instead of reading it again. The cache conditions were included in the evaluation because they may have an impact on positioning performance, as it speeds up the RFID inquiry process (because only the RFID's ID has to be read), but could also lead to a stronger influence of overreach errors (the longer reading process may have a higher probability for reading errors, which in turn would lead to an exclusion of the tag in the position determination). The answer to that question can be found in Section V-A. All in all, this led to 160 system traces that were compared to their respective ground truth.

C. *Calculation of the Error Distance*

As indicated above, the extracted traces from the ground-truth acquisition contain highly accurate data for each single footstep. LORIOT on the other hand, does not measure footsteps, but was designed to estimate the position of the user's whole body. The question arises what the position of a user is, if the positions of his feet are known. For the evaluation, it was assumed that the user's position is on the straight line between two successive foot positions.

This consideration is important, since LORIOT computes a new position every time a new measurement is taken, meaning that time-stamps of derived positions do not necessarily coincide with time-stamps of ground truth traces. Thus, a way had to be found to find the user's ground truth position at an arbitrary time-stamp.

Figure 6 exemplifies the situation. The two footprints indicate two subsequent footsteps of a ground truth, TS_R and TS_L are the time stamps for the right and left foot. The blue dot shows the position calculated by LORIOT, derived at time-stamp TS_{LORIOT}. According to the exemplary given time stamps, LORIOT's position was derived 0.325 seconds after the right foot reached the ground and 0.375 seconds before the left foot will reach the ground in the ground truth. The user's position in the ground truth is thus somewhere in between.

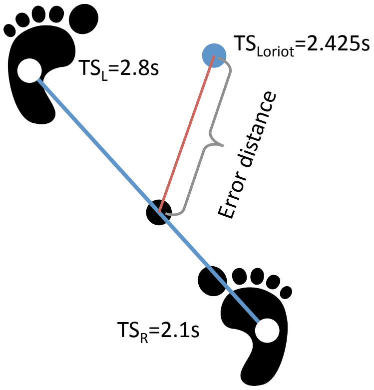


Fig. 6: Ground truth time-stamps of single footsteps and LORIOT time-stamps of user positions do not necessarily coincide.

To interpolate where the user's position was in the ground truth at time T_{LORIOT}, the current velocity v is calculated by dividing the distance between the two footsteps with the time difference between the two footsteps:

$$v = \frac{\sqrt{(x_L - x_R)^2 + (y_L - y_R)^2}}{\text{TS}_L - \text{TS}_R} \tag{2}$$

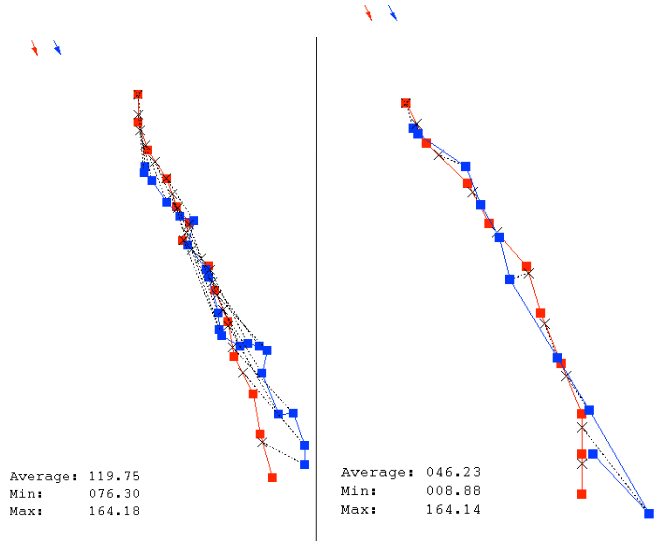
where (x_L, y_L) and (x_R, y_R) are the coordinates of the left and right foot. By multiplying this velocity with the time difference between TS_{LORIOT} and TS_R, the distance d, which the user has covered since putting their right foot down can be derived as follows:

$$d = v \times (\text{TS}_{\text{LORIOT}} - \text{TS}_R) \tag{3}$$

The user's position $P_{\text{groundtruth}}$ at time TS_{LORIOT} in the ground truth is estimated to be at distance d from the right footstep on the line between the two footsteps. The positioning error is thus the distance from LORIOT's derived position to $P_{\text{groundtruth}}$. $P_{\text{groundtruth}}$ is indicated as a black dot in Figure 6.

V. EVALUATION RESULTS

Figure 7 shows two comparisons of system traces with their respective ground truth as an example: Trace 2 in the *only RFID, with cache* condition and Trace 3 in the *RFID & IR, with cache* condition. The red squares represent the footsteps of the ground truth. The blue boxes depict the user position as derived by LORIOT. The black crosses show the interpolated user position on the ground truth. Each interpolated user position is connected via a black dotted line with the corresponding system position. The red and blue arrows show the general walking direction of the ground truth and system trace respectively. The average positioning error as well as the minimum and maximum positioning error of the trace is printed in the bottom left corner.



(a) Trace 2 in the *only RFID, with cache* condition.
(b) Trace 3 in the *RFID & IR, with cache* condition.

Fig. 7: Two example results from the evaluation. The red boxes depict the ground-truth steps. The blue boxes represent the positions derived by LORIOT. The black crosses show the interpolated user steps, which are connected by black dotted lines with their respective user position.

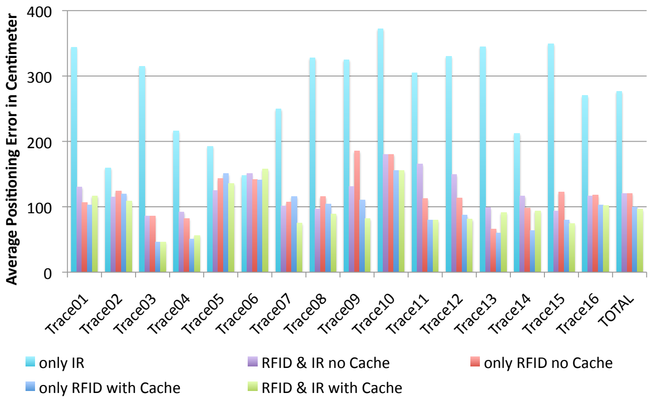


Fig. 8: The average positioning error of all traces with original velocities and with respect to the five tested conditions.

The average positioning error in centimeters for each trace and each condition is summarized in Figure 8. The last column, labeled TOTAL, shows the average error over all traces for each condition. Table I summarizes the key values for each condition. The entries are ordered top to bottom by their average positioning error over all traces (from lowest to highest). The standard error as well as the 95% confidence interval is given for each condition. A repeated measures ANOVA was performed over the differences of each trace and for each condition, and showed an overall significance with $F(4, 180) = 47.3, p < .001$.

Condition	**Average** in cm	Std. Error in cm	95% Confidence Interval Lower Bound in cm	Upper Bound in cm
RFID & IR with cache	**96.31 (1)**	4.00	88.42	104.20
only RFID with cache	**99.79 (2)**	4.13	91.64	107.94
RFID & IR no cache	**120.13 (3)**	5.80	108.68	131.57
only RFID no cache	**120.36 (4)**	4.97	110.55	130.17
only IR	**276.67 (5)**	14.03	248.99	304.36

TABLE I: Comparison of positioning errors when following the ground truth in original velocity. The numbers in parenthesis show the ranking of each value.

With the help of the conducted evaluation, the following questions regarding LORIOT could be answered.

A. *How is the accuracy influenced if the caching algorithm is enabled or disabled?*

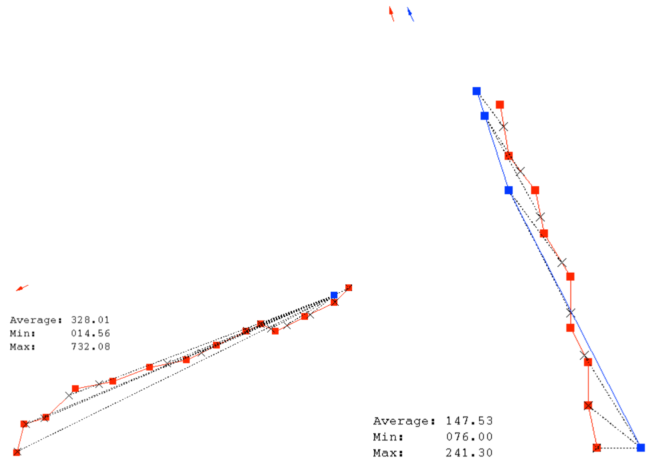
Table I shows that both cached conditions (*'only RFID with cache'* and *'RFID & IR with cache'*) outperform all other conditions. With 99.79 centimeters, the average positioning error in the *'only RFID with cache'* condition is 20.57 centimeters lower than in the *'only RFID no cache'* condition. A Bonferroni adjusted pairwise comparison shows that this difference is significant with $p < .001$. The difference between the average positioning error of the two *RFID & IR* conditions amounts to 23.82 centimeters in favor of the *with cache* condition and is also significant with $p < .001$. It can thus be concluded that the caching algorithm improves the positioning accuracy by approximately 20 centimeters in average.

B. *How is the accuracy influenced if only IR beacons are considered in the positioning evaluation?*

When only considering IR beacons, LORIOT could only achieve an average accuracy of 2.77 meters, which is the highest measured average positioning error measured in this evaluation. The difference to all other conditions is significant with $p < .001$ for all pairwise comparisons.

The minimal positioning error was 14 centimeters and the maximum was 7.32 meters. Both values were achieved in Trace 8, which is shown in Figure 9a. Only one IR beacon was received in this test and thus only one position was fixed by LORIOT. Analyzing all IR only traces shows that in 11 out of the 16 traces only one IR beacon was detected during the test walks. Two IR beacons were detected in three traces. Three and four beacons where detected in only one trace each. In Trace 6, four beacons were detected and, with 1.48 meter, this trace also shows the lowest average positioning error for all traces in the *'only IR'* condition.

The low accuracy in the *'only IR'* condition was to be expected and is due to the comparably sparse instrumentation of the test field with IR beacons. IR beacons are advantageous



(a) Result for Trace 8 in the *'only IR'* condition.

(b) Results for Trace 6 in the *'only IR'* condition.

Fig. 9: The worst (a) and best (b) result for the *'only IR'* condition. In Trace 8 only one IR beacon was detected. Trace 6 contains 4 detected IR beacons.

at precise points of interest, like exhibits in a museum, particular shelves in a retail environment or decision points in a narrow corridor. Furthermore, the *'only IR'* condition provides a special case since without active RFID tags no coordinate information can be stored in the environment. Thus, a list containing the beacon IDs and their coordinates has to be stored on the mobile device. Installing only IR beacons in a large area with nearly no walking restrictions is therefore only recommended for special applications, like museums or shops.

C. How is the accuracy influenced if only RFID tags are considered in the position estimation?

The *'only RFID with cache'* condition shows the second best accuracy, with an average positioning error of 99.79 centimeters. The minimum positioning error in this condition was 3.88 centimeters (Trace 15) and the maximum was 276.88 centimeters (Trace 5). The *'only RFID no cache'* condition ranked second to last, with an average positioning error of 120.36 centimeters and minimum and maximum error of 65.91 (Trace 13) centimeters and 185.67 (Trace 9) centimeters respectively. The average is still 156.32 centimeters better than the *'only IR'* condition and this difference is significant with $p < .001$. Since caching already proved to be advantageous, it can be concluded that LORIOT can achieve a positioning accuracy of approximately 1 meter in an environment that is densely instrumented with only active RFID tags.

D. How is the accuracy influenced if RFID tags and IR beacons are considered in the position estimation?

Table I shows the lowest average positioning error in the case of combined RFID and IR instrumentation and with enabled caching. With 96.31 centimeters, the average positioning error is approximately 3 centimeters lower than RFID alone (with enabled caching). However, a pairwise Bonferroni adjusted comparison shows that this difference is not significant. The difference of 0.23 centimeter when comparing *only RFID* and *RFID & IR*, both with caching, is negligible and also not significant. These low, not significant differences can also be contributed to the sparse IR beacon instrumentation as well as to the high walking speed of the ground truth, which makes it less probable that an IR beacon will be properly detected.

Condition	Average in cm	Minimum in cm	Maximum in cm
RFID & IR, with cache	**24.81 (1)**	13.21 (2)	44.26 (3)
only RFID, with cache	**25.00 (2)**	12.60 (1)	39.10 (1)
only RFID, no cache	**30.33 (3)**	16.05 (3)	43.80 (2)
RFID & IR, no cache	**31.39 (4)**	19.34 (4)	48.77 (4)

TABLE II: Comparison of positioning errors when following the traces in slow velocity.

E. What is the influence of walking speed on the position accuracy?

To answer this question, the raw sensor data log-files of the slowly walked traces where analyzed. Because of the different velocities of the ground traces and the re-walked traces, there is no direct relation between their time-stamps, and thus the calculation of the error distance had to be adapted accordingly.

For the slow velocity traces, for every calculated user position the nearest footstep in the ground truth was found and the distance to that footstep was taken as the positioning error. If a footstep in the ground truth had already been used as reference point, it was not used again and only footsteps with a higher time-stamp than the last footstep were allowed. This method is thus analogous to a comparison of graphical similarity.

Table II summarizes the average, minimum and maximum positioning error for each of the four conditions. The results when walking slowly are greatly improved. The best result was achieved with RFID & IR and enabled caching. This condition led to an average positioning error of only 24.81 centimeters. The highest average positioning error was measured in the condition where RFID and IR was used without caching and amounts to 31.39 centimeters.

A part of this improvement is due to relaxed measurement of the positioning error. To test if the improvement can be attributed to the different measurement method alone, the traces that were followed based on the time-stamps of the ground truth were re-analyzed using the same method.

Table III shows the results of the analysis. The results are indeed an improvement over the time-stamp based analysis, but not as good as the measurements that were based on the slow velocity traces. In the worst case (*'RFID & IR, no caching'*), the average positioning error is 83.73 centimeters. Compared to the 31.39 centimeters when walking slowly, this average is approximately two times higher.

Condition	Average cm	Minimum cm	Maximum cm
RFID & IR, with cache	**57.48 (1)**	27.24 (1)	135.76 (2)
only RFID, with cache	**61.69 (2)**	28.77 (3)	159.63 (3)
only RFID, no cache	**73.11 (3)**	33.44 (4)	127.73 (1)
RFID & IR, no cache	**83.73 (4)**	27.30 (2)	219.23 (4)

TABLE III: Comparison of positioning error when comparing the graphical similarity of the system to the ground truth.

The lowest achieved average positioning-error was 57.48 centimeters and was measured with RFID & IR and enabled caching. This positioning error is also approximately two times higher than the best average when walking slowly.

It can thus be concluded that the accuracy of LORIOT is higher at slow walking speeds.

F. How accurate is LORIOT *on average?*

Considering the above results, LORIOT achieves its highest accuracy with enabled caching and with either RFID alone or with combined RFID and IR instrumentation. The average positioning error over all traces of *'only RFID with cache'* and *'RFID & IR with cache'* results in 98.05 centimeters at normal walking speed. The accuracy is higher at slow walking speeds. As a slower walking speed can be expected if a person is walking through unknown territory, while exploring their surroundings or when trying to find their way, this higher accuracy will most likely be available, when a person is using a location-based service.

VI. CONCLUSION AND FUTURE WORK

We presented an evaluation method based on natural footstep accurate traces as ground truth. We have shown how this ground truth can be obtained by manually analyzing recorded video clips and how the corresponding traces of a target positioning-system can be recorded and compared to ground truth. Furthermore we presented the results of such an evaluation on the example of an onboard/egocentric Always Best Positioned system.

The advantage of the presented evaluation method is that it enables the creation of an unbiased ground truth, i.e. participating users do not have to be aware that they are taking part in a positioning experiment. The 'chicken and egg' problem of having to evaluate a reference system before the actual target system can be evaluated is avoided by manually extracting the position of each single step of an observed person. One may argue, that the manual extraction process is error prone and that the resolution is limited, but taking into account that most positioning systems do not position the feet of users, but rather the position of the whole body, the question arises how to determine this position, if only the positions of the feet are known. The proposed method incorporates this question into the evaluation process itself, by interpolating the ground-truth data to the time stamps of the target-system traces. We tend to think that our evaluation method is thus more human-centric.

On the downside however, the process of manually analyzing video clips is very tedious and time consuming work. Automating this work would be highly desirable, but could also lead back to the 'chicken and egg' problem, i.e. the automated process has to be evaluated first. Nevertheless, since we decoupled the process of obtaining ground truth and obtaining target system traces, an erroneous ground truth may not have a big impact on the evaluation outcome, so long as it isn't 'unwalkable'.

For future work, we will research on how to ease the whole process and will develop easy to use tools to conduct further evaluations. We will also be highly interested in constructive criticism on our evaluation method from the research community and are looking forward to discussions on how to further improve it.

ACKNOWLEDGMENT

This work was partially supported by *EIT ICT Labs*.

REFERENCES

[1] P. Stephan, I. Heck, P. Kraus, and G. Frey, "Evaluation of Indoor Positioning Technologies under industrial application conditions in the SmartFactoryKL based on EN ISO 9283," in *Proceedings of the IFAC Symposium on Information Control Problems in Manufacturing (INCOM'09)*. IFAC, 2009, pp. 870–875.

[2] P. Steggles and S. Gschwind, "The Ubisense Smart Space Platform," in *Adjunct Proceedings of the International Conference on Pervasive Computing (Percom'05)*. Austrian Computer Society, 2005, pp. 73–76.

[3] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-Support Sytem," in *Proceedings of the 6th International Conference on Mobile Computing and Networking (MOBICOM '00)*. Boston, MA: ACM, July 2000, pp. 32–43.

[4] J. Rydell and E. Emilsson, "(Positioning Evaluation)^2," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN'11) (Short Paper)*, 2011, pp. 1–4 (online publication, no pagination).

[5] T. Schwartz, "The Always Best Positioned Paradigm for Mobile Indoor Applications," Ph.D. dissertation, Saarland University, 2012.

[6] T. Schwartz, C. Stahl, J. Baus, and W. Wahlster, "Seamless Resource-Adaptive Navigation," in *Resource-Adaptive Cognitive Processes*, ser. Cognitive Technologies, M. W. Crocker and J. Siekmann, Eds. Berlin, Heidelberg, New York: Springer, 2010, pp. 239 – 265.

[7] B. Saliba, "An Evaluation Method For Indoor Positioning Systems On the Example of LORIOT," Master's thesis, Saarland University, 2011.