Lorcan Coyle · Juan Ye · Emerson Loureiro · Stephen Knox · Simon Dobson · Paddy Nixon

A Proposed Approach to Evaluate the Accuracy of Tag-based Location Systems

Received: June 11 2007 / Accepted: July 3 2007

Abstract Location detection systems that use tags are a popular means of determining a user's location. These systems are characterised as requiring the user to carry an identity tag that is detected by sensors, which typically use some form of triangulation to determine location. Although estimates for precision for these systems are published by the respective manufacturers the customer experience can vary widely. This paper proposes an evaluation framework which will allow different systems to be compared more directly. This framework is specifically targeted at evaluating the experiences of tagging humans, which can cause particular difficulties due to the fact that many tag-based systems use communication frequencies that cannot pass easily through the human body.

Keywords Evaluation Frameworks · Location-based Systems

1 Introduction

Location based systems are becoming more commonly applied to a number of problems in ubiquitous computing and beyond [11,5]. When dealing with these systems it is important to have a gauge of their accuracy. Hightower and Borriello performed an evaluation of location systems and defined the quality of location systems by looking at the percentage of readings that fall within a certain distance of the true value [6]. The distances denote the accuracy, or grain size, of the position information GPS can provide and the percentages denote precision, or how often we can expect to get that accuracy. Dobson et al. derive three factors for accuracy of location information: precision, which captures inaccuracy due to the resolution of sensor readings; decay, which captures inaccuracy due to the staleness of sensor readings; and confidence, which measures the degree of belief on a sensor reading [2]. We use a combination of these definitions: using Hightower and Borriello's definition of accuracy and precision to calculate values for Dobson et al's precision component of location accuracy. When we use the term precision in this paper, we are using Dobson et al's meaning of the word.

This paper describes an evaluation framework for calculating the precision of a tag-based location system. Other papers have evaluated location based technology individually, but these evaluations are usually performed by the stakeholders, and the evaluation criteria are not always transparent. Additionally, it is often unclear whether the published precisions are determined from evaluations using tags alone or from tagged human subjects. Our work focuses on human subjects due to the inherent interference caused by the human body on the communications mechanisms used by the tested location systems. While the examples described in this paper focus on human tagging, the framework itself is agnostic to what is being tagged. The important point to take is that evaluations using this framework should take the intrinsic physical properties of what is being tagged into account when determining system precision.

Section 2 briefly surveys a number of tag-based location systems that could be evaluated using this framework. Section 3 describes the evaluation framework and Section 4 describes an exemplar evaluation that is underway to test the precision of an installation of a tag-based Location based system. Finally, Section 6 concludes the paper, lays out the hypotheses underpinning our evaluation, predicts some results and outlines some ongoing research related to this paper.

This work is partially supported by Science Foundation Ireland under grant numbers 05/RFP/CMS0062 "Towards a semantics of pervasive computing" and 04/RPI/1544 "Secure and predictable pervasive computing"

L. Coyle
Systems Research Group, School of Computer Science and Informatics, UCD Dublin, Ireland;E-mail: lorcan.coyle@ucd.ie

2 Survey of Tag-based Location Detection Systems

There are a number of popular tag-based location detection systems today. The most widely used is the GPS system, which allows a user to be positioned anywhere on the surface of the planet with a high degree of accuracy (approximately 3-10m precision). This is beyond the scope of this paper since its precision has been calculate effectively and its limitations are well understood. The most prominent of these limitations is the necessity to have clear line-of-sight to at least four satellites. The framework described in this paper is better suited to local location positioning systems offering a sub-room granularity.

Ubisense allows precise local positioning by tracking a tag (the Ubitag), which is attached to a person. The tag maintains radio contact with an installation of sensors. These sensors use ultra-wideband (UWB) technology to detect and react to the position of Ubitags. Ubisense uses both Time Difference Of Arrival and Angle of Arrival to calculate location. In a typical open environment, Steggles and Gschwind claim that location accuracy of about 15cm can be achieved across 95% of readings [10].

The Active Badge location system uses diffuse infrared technology to communicate between the tag and sensor network [12] and offers room-level precision; the Cricket system uses ultrasonic signals to calculate positional data, at a precision of approximately 1m (4sq feet) [9]. When a cricket tag receives signals from multiple beacons in the infrastructure it triangulates its position. When receiving signals from only a single beacon it can still provide location based on proximity to that beacon. Many of these systems use radio signal strength information (RSSI) to calculate location: these include SpotON [7]; LANDMARC, which uses active RFID [8]; RADAR, which uses signals from existing WiFi infrastructure [1]; and Feldmann et al.'s work using Bluetooth [3].

Each of these systems has some common factors that affect any evaluation of their precision. We address these factors in Section 4; our evaluation framework will attempt to deal with these factors and allow evaluations of multiple systems to be compared directly against each other.

3 Evaluation Framework

Any evaluation of a location-based system must take into account a number of factors that affect readings from location sensors. Some of these relate to the technology used and others relate to tag-based systems in general:

- **Tag State:** the tag state may have an impact on accuracy. To conserve power, many tags will enter an idle state after a period of inactivity. When performing evaluations this feature should be disabled. Also, powered tags should be fitted with fresh batteries to ensure that they operate at peak efficiency.
- **Interference:** depending on the tag communication system, if more than one tag is put in the same general area, they could interfere with each other. This is typically a feature when dealing with passive RFID tags.
- **Sensor Configuration:** The space with the highest precision will typically be that which is well covered by sensors. The physical configuration of these sensors is critical, e.g., when dealing with Ubisense, the best configuration is to place a sensor in each corner of a room. In this way, the space with the highest precision is in the middle of the room, i.e., is visible to all sensors. Correspondingly, the worst reading area is the wall or the corner of the room that is out of coverage of the sensors. A fair evaluation should compare systems at a range of areas from where they should be most accurate to where they will perform worst.
- **Environment:** the environment in which the tag is located will have an important impact on the detected precision. The presence of metallic structures or electronic equipment could lead to interference with the location detection hardware when using UWB systems, and the presence of ultrasound noise may interfere with the cricket system. Any evaluation should attempt to minimise the effect of these obstacles.
- **Height** When dealing with location detection systems care should be taken if 2D coordinates are required when dealing with systems that offer 3D coordinates. Depending on the application, often the height dimension is ignored and only x and y coordinates are used. However, this may cause problems if there is a difference in precision at different heights. If this is the case, care must be taken when positioning the tag on the human body. If evaluating a 3D location based system it would be possible to test the extent of this problem by positioning the tag at a different heights and comparing the generated precisions.
- **Frequency:** The frequency of sensor readings is an important consideration when gathering evaluation data. Generally this can be altered but as the frequency is increased, the cost in terms of power usage increases. An evaluation should use a frequency that will be usable in real world conditions.

The factor that is usually ignored in typical evaluations is the human body itself. If the location positioning system is explicitly for use in tracking people then its evaluation should use human test subjects. This is especially important when evaluating systems that use communications that have difficulty penetrating water (since people are composed of mostly water). This is an extension of the environment factor except that it is not useful to remove the human from the evaluation if tracking people is the target application.

Assuming that communications that must pass through the person are of lower quality than those that do not, tag orientation is critical. Positions will be calculated with a higher precision if a user is facing the sensor since the tag has a direct line of sight to the sensor. Any tag-based system that suffers from this limitation should be tested in different subject user orientations.

Precision can be measured at a single point by comparing the system's estimated location with the ground truth or actual value. If the tag remains in that location, the set of readings should be independent and normally distributed about a point. By taking a large sample of readings and calculating the mean and standard deviation it should be possible to compare these to the ground truth and from that calculate an error for that location (using Student's t-distribution [4]). This error reflects the precision of the system at that point in space. By calculating precision over a large area it is possible to calculate a usable measure of the precision of the overall system. Further statistical tests can be performed to assess the advertised precision and to compare the variance in precision between optimal locations and locations for which precision would be expected to be less.

We propose to calculate large-area precision by collecting a broad dispersal of samples throughout the area covered by location detection infrastructure. We assume that the tag can be sensed and located throughout the test area. A grid of squares will be marked in the test area. These squares will be just large enough for a person to have enough room to stand completely inside them (the grid is 30cm square). The position of the centre will be calculated accurately using laser range finders. Test subjects will wear a tag and walk the grid, stopping long enough in each square for enough readings to be recorded to calculate precision. These data will be matched to their corresponding grid ground-truth. Means will be calculated for each sample set and the overall set of means will be tested for offset (it is possible that the ground-truth generated does not match the ground-truth used in calibrating the original infrastructure). Precision will be calculated for each position in the grid and graphed. It should be clear from this graph where the precision is best. With good test coverage of the installation space, it should be possible to test the advertised precision for the system.

In an initial evaluation environment, we choose a single tag that is active and put in a static and stable position. The location data for this position can be physically measured beforehand. We propose to use laser range finding to calculate the baseline positions in the x-y plane. When the tag is positioned correctly, the feed of sensed positions is stored in a dataset.

The sample data collected for each point are processed as follows: the mean (μ) and standard deviation (σ) are calculated for both the x and y dimensions. For each square in the grid a precision is calculated using a 95% confidence interval. The sample is also used to test whether, and under which conditions, the precision meets the advertised system precision.

4 Proposed Evaluation of Ubisense

As an exemplar evaluation, we will perform an evaluation of a Ubisense installation in the Systems Research Group in University College Dublin. The optimal location for precision in this installation is an open-plan common area, which is used for coffee breaks. This area is empty of metal objects, with four sensors pointing inwards, as shown in Figure 1. The sensors are indicated by the red icons in the North-West, North-East, South-East, and South-West of the map. Each sensor has an orientation plane and Ubitags located normally to this plane will tend to have greatest visibility. With this configuration of sensors, a tag will have a direct line of sight to at least one sensor, no matter what the orientation. Ubisense has an advertised typical precision of 15cm in 3D space. This evaluation will test this claim in a small space. We will then perform similar small-scale evaluation of specific locations where the Ubisense would be expected to perform with less precision – spaces with less visible sensors, in areas with metallic obstructions, etc., to test the variance of precision with changes in the factors outlined in Section 3.

Our evaluation will test the Ubisense installation in a single horizontal plane. To control against user height in the final evaluation tags will be hung at different lengths on each user so that they each hang at 120cm from the floor.

Readings are taken by the test subject under supervision by clicking on their Ubitag. This signals to the evaluation system that readings should be recorded in the database. The tag will beep after 30 seconds worth of readings have been taken in a single position. Our preliminary testing used sensor readings taken at 10Hz so a 30 second sample tends to yield as much as 300 readings (it should be noted that our initial testing yielded far fewer readings). When these readings are taken the user moves to the next square and repeats the process. They are expected to enter each square in the grid in sequence and take readings when they are comfortable and relatively stable. It is impossible to completely steady the test subjects but these slight variances will introduce a relatively low error, and should be evened out somewhat by taking readings over time.

The evaluation will be repeated using different users. When comparing precision between data gathered from different users it will be important to perform controls against other variables. Since ultra-wideband has difficulty passing through the water present in the human body, we can expect body mass to have an influence. We will attempt to evaluate the effect of this variable by comparing the precision of measurements given different body masses. A final evaluation will test the ability of

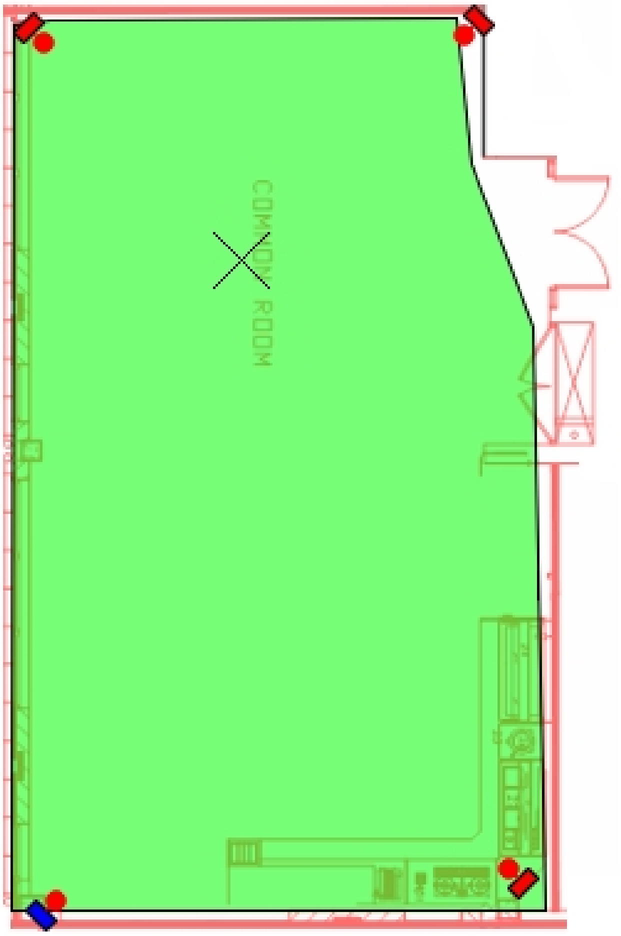

Fig. 1 Configuration of Ubisense in Exemplar Evaluation

Ubisense to maintain precision when multiple tags are present.

5 Preliminary Results

As a simple demonstration of our approach, we took readings at a point X (marked on the map in Figure 1), which is 5.63m East of the left wall and 8.32m North of the bottom wall. Four sets of readings were taken, one for each of the four cardinal orientations — North, South, East, and West (North is facing up on the map). Thirty seconds worth of readings were recorded at each orientation and fed into the scatterplot in Figure 2. As can be seen from the scatterplot, the readings for each orientation are quite tightly bound together. This is to be expected, since at those orientations the Ubitag can be easily seen by at least two sensors.

Table 1 shows the means for each of the orientations samples as well as the absolute distance between those means and the actual ground truth. From this sample it can be seen that orientation is an important factor in precision; greater precision is observed when facing South or East than North or West for this location. It should be noted that at this location the observed precision values are not within those advertised by Ubisense. Our final evaluation will sample readings from throughout the area in Figure 1 and measure the varience in precision of both position and orientation.

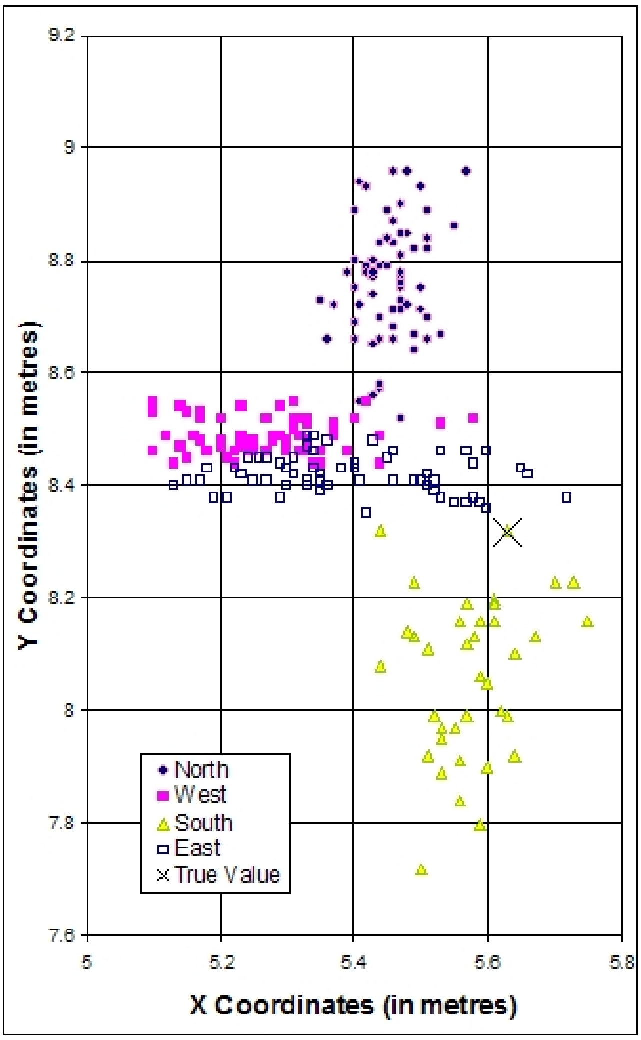

Fig. 2 Scatterplot of Location Coordinates for Different Orientations

Orientation	Mean X	Mean Y	Error
True Value	5.63	8.32	0.00m
North	5.45	8.76	0.48m
West	5.27	8.49	0.40m
South	5.58	8.07	0.26m
East	5.39	8.42	0.26m

Table 1 Table of Means for Each Orientation with Distance Error

6 Conclusions and Future Work

This paper describes an evaluation framework for calculating the precision of a location detection system that uses tags. Section 4 describes our aims for the evaluation of Ubisense, which is ongoing. We will test the following hypotheses and anticipate the following results:

- We will test the variance of precision throughout our installation. We anticipate that precision will be high in areas of good sensor coverage and low in poorly covered areas. We anticipate that the optimal precision will approach the advertised precision of 15cm (except that our evaluation will use 2D space rather than in 3D space as advertised).
- We will test the variance of precision with body orientation. Since the human body will block much Ubisense communication we believe that human orientation will be critical to precision. The more sensors that are in direct line-of-sight to the tag the better for precision.
- As an extension of the previous evaluation we will test the variance of precision with human body mass. We believe that precision will drop as body mass increases.

One limitation of our approach is that precision is only calculated for static positions. The main difficulty with calculating precision for moving objects is that the ground truth must be closely tied to individual sensor readings. We will extend our framework to calculate precision for moving objects by using pressure sensitive pads to determine when a user is located at a defined point during a walk. By walking subjects through these defined cue points we will be able to build datasets for these points relating to speed and trajectories.

Although the evaluation presented here is preliminary, it supports our argument that orientation is an important consideration when determining precision for Ubisense. We believe that a more detailed evaluation of our Ubisense installation will yield factors that will support the evaluation of other location based systems including Bluetooth-based systems, an active RFID installation, and a location system based on activity sensors and calendar information. By applying this framework to these diverse systems we will be able to generalise it and use the evaluation results to perform more direct comparisons between those technologies. Parallel work is underway on the representation of uncertain context data, the implementation of a unified model for dealing with data from diverse locaction models efficiently [13], and tools for modelling and reasoning about uncertain location data [14].

References

1. P. Bahl and V. N. Padmanabhan. RADAR: An in-building RF-based user location and tracking system. In *INFOCOM (2)*, pages 775–784, 2000.
2. S. Dobson, L. Coyle, and P. Nixon. Hybridising events and knowledge as a basis for building autonomic systems. *IEEE TCAAS Letters*, 2007. To appear.
3. S. Feldmann, K. Kyamakya, A. Zapater, and Z. Lue. An indoor bluetooth-based positioning system: Concept, implementation and experimental evaluation. In W. Zhuang, C.-H. Yeh, O. Droegehorn, C.-T. Toh, and H. R. Arabnia, editors, *International Conference on Wireless Networks*, pages 109–113. CSREA Press, 2003.
4. W. S. Gosset. On the probable error of a mean. *Biometrika*, 6(1), 1908.
5. M. Hazas, J. Krumm, and T. Strang, editors. *Location- and Context-Awareness, Second International Workshop, LoCA 2006, Dublin, Ireland, May 10-11, 2006, Proceedings*, volume 3987 of *Lecture Notes in Computer Science*. Springer, 2006.
6. J. Hightower and G. Borriello. Location systems for ubiquitous computing. *IEEE Computer*, 34(8):57–66, August 2001.
7. J. Hightower, R. Want, and G. Borriello. SpotON: An indoor 3d location sensing technology based on RF signal strength. UW CSE 00-02-02, University of Washington, Department of Computer Science and Engineering, Seattle, WA, February 2000.
8. L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil. Landmarc: indoor location sensing using active rfid. *Wirel. Netw.*, 10(6):701–710, 2004.
9. A. Smith, H. Balakrishnan, M. Goraczko, and N. B. Priyantha. Tracking moving devices with the cricket location system. In *2nd International Conference on Mobile Systems, Applications and Services (Mobisys 2004)*, Boston, MA, June 2004.
10. P. Steggles and S. Gschwind. Ubisense-a smart space platform. Technical report, Ubisense, May 2005.
11. T. Strang and C. Linnhoff-Popien, editors. *Location- and Context-Awareness, First International Workshop, LoCA 2005, Oberpfaffenhofen, Germany, May 12-13, 2005, Proceedings*, volume 3479 of *Lecture Notes in Computer Science*. Springer, 2005.
12. R. Want, A. Hopper, V. Falcao, and J. Gibbons. The Active Badge Location System. Technical Report 92.1, Olivetti Research Ltd., ORL, 24a Trumpington Street, Cambridge CB2 1QA, 1992.
13. J. Ye, L. Coyle, S. Dobson, and P. Nixon. A unified semantics space model. In *Proceedings of the 3rd international symbosium on Location- and Context-Awareness (LoCA 2007)*, Oberpfaffenhofen near Munich, Germany, September 2007. To appear.
14. J. Ye, L. Coyle, S. Dobson, and P. Nixon. Using situation lattices to model and reason about context. In *Fourth International Workshop on Modeling and Reasoning in Context (MRC 2007).*, August 2007. To appear.