Towards Qualitative Positioning for Pervasive Environments

Ian Anderson *Member, IEEE,* Henk Muller

***Abstract*— In this paper we present a strategy and set of algorithms for developing qualitative positioning services that are specifically optimised for the environment where they are to be deployed. We argue that for many context-aware applications this may be more appropriate than more common quantitative location systems, where the positioning API may make unrealistic demands on the underlying measurement service, and unrealistic promises to the application.**

We show results for an implementation on a shopping street using cellular networks.

***Index Terms*— Context-Aware Computing, Location Fingerprinting, Machine Learning, Wireless Networks**

I. Introduction

It is well established that modelling the spatial environment is an essential part of developing a context-aware application [1], [2]. As such, multiple models have been developed over the past few years [3], [4], [5]. These spatial models can typically be classified as either topological (qualitative) or more commonly, as coordinate based (quantitative). Quantitative models generally take a geometric view of space with positional information supplied by location services using Euclidean or spherical coordinate systems. Coordinate tuples are processed by the application and behaviour is updated to reflect the new location information. In contrast, topological or symbolic models manage space in a qualitative manner with positional information mapped to human abstractions of physical places usually in the form of spatial zones. The relationships between zones forms a topology often expressed as a graph. Application behaviour varies depending upon the symbolic representation of space (zone) that the user is currently located in.

When constructing a symbolic model of the spatial environment developers must define spatial zones within the constraints of the underlying sources of positional information. For example, it is not possible to create zones with a physical coverage area that is finer than the granularity of the data produced by the positioning services.

With some positioning systems performance varies depending upon the physical environment where the system is deployed [6], [7]. For example, RF based systems suffer from multi-path fades, dead-spots, signal diffraction and reflection, creating inconsistent performance in different areas of the application environment. This poses problems for developers and forces a laborious offline calibration phase where positioning system performance is assessed and zones are created to reflect the limitations of the measurement service and environment.

In this paper we propose a strategy and set of algorithms for developing a positioning system that offers an appropriate service for the given spatial environment. In particular, our qualitative positioning service defines zones that are determined by the quality of the measurements. We use the term measurements to describe information that the positioning system can use to calculate location. We are agnostic about the type of measurements (cellular, 802.11, ultrasonics), as long as the measurements are position dependent. This qualitative approach differs from more common quantitative location systems, where the positioning API may, given the available measurements, make unrealistic demands and unrealistic promises to the application programs.

The rest of this paper is structured as follows: Section II provides an overview of pervasive positioning technologies, Section III discusses the underlying spatial model and introduces the concept of a logical path, Section IV demonstrates how a zone based representation of the spatial environment can be generated in an unsupervised manner via a simple calibration procedure, Section V demonstrates how zone topology can be inferred by applying Markov chain frequency analysis techniques and Section VI reports on a prototype implementation with cellular networks.

II. Background

In recent years there has been a huge increase in the popularity of context-aware applications. As such, there has been much research into developing positioning services to support the location sensing aspect of these applications. In this section we provide an overview of some of these technologies.

Location fingerprinting is a term used to describe a method of positioning a mobile device using signal strength levels from wireless beacons such as 802.11 access points and cellular base stations. Deploying a location fingerprinting positioning service is a two stage process. Firstly, an offline calibration phase is undertaken where the signal strength levels are recorded at fixed points in the application environment. The positions and associated signal strength levels (fingerprints) form a radio map of the environment. At runtime, users match their current signal strength levels from visible beacons against those in the radio map. Typically the associated position of the closest matching fingerprint (shortest Euclidean distance) is returned to the user as their current position.

I. Anderson and H. Muller are with the University of Bristol, Department of Computer Science, Merchant Venturers Building, Woodland Road, Bristol, BS8 1UB, U.K.

E-mail: {anderson},{henkm}@compsci.bris.ac.uk

This research was funded by the UK Engineering and Physical Science Research Council, Equator Interdisciplinary Research Collaboration EPSRC GR/N15986/01 (http://www.equator.ac.uk)

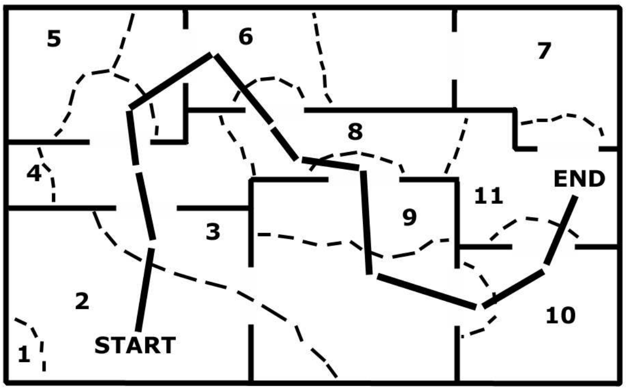

Fig. 1. A typical office floor plan - the spatial environment for a context-aware application. The spatial environment partitioned into zones that reflect the performance of the underlying positioning services.

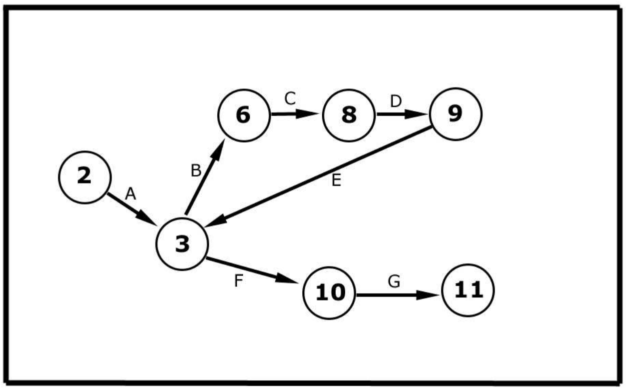

Fig. 2. The physical path illustrated in Figure 1 can be represented as a directed graph. The nodes in this graph correspond to qualitative locations and the arcs indicate order.

The RADAR system [8] was one of the first to apply this location technique and achieved a positional accuracy of 1.5 meters using a network of 802.11 (WiFi) wireless access points. Since then there has been much work trying to improve on these initial results by building statistical models, applying complex RF signal propagation models and other tracking/filtering techniques [9], [10], [11]. As such location fingerprinting is now a very popular positioning technique. Other favourable traits include: user privacy, low cost, operates in environments where the Global Positioning System (GPS) would fail (indoors and in dense urban environments) and the number of wireless beacons available in our cities and towns has increased dramatically over the last few years. For example, in 2005 during a war driving survey it was shown that downtown Seattle has a WiFi access point density of 1200 per km^2 [12].

The Place Lab project [13] uses the known position of approximately 2.2 million radio beacons to position mobile devices such as cell phones, PDAs and laptops. By applying Bayesian filtering techniques such as a particle filter, a median position accuracy of 20-30 meters has been achieved with almost a 100% environmental coverage where coverage is by assessed by the availability of location information in peoples daily lives [14].

These quantitative approaches provide a position and an associated accuracy error that reflects the limits of the underlying measurements. Our work differs from this in that we contain the error within the positioning system and return a qualitative location that reflects the best achievable performance given the available measurements and constraints of the spatial environment.

III. Logical Management of Space

In this section we introduce the concept of a spatial zone and illustrate how transitions between zones can be expressed as directed graphs.

We use the term *spatial zone* to describe a portion of space that, when using a measurement of, for example, signal strength of a wireless beacon, can be distinguished from other areas of space. The area of physical space that a zone symbolizes, reflects both the quality of the positional measurements and the spatial environment. Thus zones represent the finest, reliable position that the measurement service can offer, i.e. if it is possible to reliably determine position within different areas of a zone then the zone should be split into smaller, child zones. Consequently, zones do not necessarily cover the same amount of physical space and hence are assumed to be of unequal size.

Our positioning service returns the zone the user is currently located in as their qualitative location. The way that zone membership is determined depends on the type of positional measurements available. For example, Figure 1 shows the layout of a typical office environment. Positional measurements are obtained from ultrasonic transmitters distributed throughout the environment. Zone membership (a users qualitative location) is determined by looking at the strongest transmitter signal. As such, the floor has been partitioned into zones that reflect the areas that can be reliably distinguished from one another using the measurements obtained from the transmitters.

Figure 1 also shows the physical path an application user took when walking through the spatial environment. In terms of qualitative location, this path simply represents a series of zone transitions in the form of a directed graph as shown in Figure 2. We use the term *logical path* to describe the series of zone transitions equivalent to the physical path.

By constructing logical paths based on users interactions with the application environment it is possible to infer the relationships between zones. This has the advantage that once sufficient data has been collected it is possible to identify popular paths and invalid zone transitions in an unsupervised manner, improving positioning service performance.

IV. Automatic Zone Creation

In this section we demonstrate how it is possible to construct a zone based representation of the application environment in an unsupervised manner that reflects the accuracy and reliability of the available positional measurements.

Constructing the zone representation is a simple offline calibration process. Firstly, the deployer collects samples of positional measurements throughout the application environment. Unlike traditional location fingerprinting calibration, the associated physical positions do not need to be stored with these measurements. Once this training data has been collected it is partitioned into sets of similar measurements. These sets contain the data that will be used to determine zone membership. Hence a set, or cluster of training data defines the boundaries of a spatial zone.

At runtime, the qualitative location of a user is determined by finding the cluster most similar to a position dependent measurement taken at the users current, physical location.

With many partitioning algorithms the developer must specify in advance, the number of clusters (zones) to create. Therefore a range of values should be tested and the performance of each assessed in order to select the optimum solution. Performance can be evaluated by generating several logical paths recorded over the same physical path. As they all represent the same physical path in theory all the logical paths should be identical. This however is not realistic as typically the source of positional measurements are inherently noisy. Therefore the optimum solution is a trade off between two factors: number of clusters and similarity of logical paths. The higher the number of clusters the greater the positional granularity since more zones represent the spatial environment. But the lower the number of clusters the greater the similarity between logical paths that represent the same physical path. It is important to note that the quality of the generated zones largely depends on the amount of training data collected and whether the data was gathered throughout the spatial environment.

We use the K-means clustering algorithm to create zones. K-means partitions N data points (training vectors) into K clusters Sj. The objective function:

$$J = \sum_{j=1}^{K} \sum_{n \in S_j}^{n} |x_n - \mu_j|^2 \tag{1}$$

where x_n is a vector representing a positional measurement and μ_j is the centroid of the data points in S_j and $|x_n - \mu_j|^2$ represents the distance between the sample and the cluster centre is used to partition the training data. If for example, the positional measurements were signal strength levels on a cellular network then x_n would represent a snapshot of these levels for all visible cells. K-means can be initialized with vectors selected at random from the training data. The Euclidean distance for each subsequent sample x_n to the centre of each centroid μ_j is then calculated. This sample x_n is then added to the centroid that it is closest too. The centroids are then recalculated and the membership of each of the points S_j for each centroid μ_j is then re-evaluated until there are no further changes in membership.

V. Zone Topology

In this section we demonstrate how logical paths, expressed as directed graphs, can be used to infer zone topology and hence improve positioning service performance.

As the physical coverage area of a zone is decreased (number of zones covering the entire spatial environment is increased) support for richer location services can be provided. But decreasing the physical coverage area of a zone (reducing the training data) reduces the reliability of accurately matching the same physical position to the same spatial zone. This problem can be minimized by identifying the zone topology and hence distinguishing between valid and invalid zone transitions. An invalid transition is one where the user is reported to have moved from a zone to another that is not a neighbour of the first zone. The zone topology is not directly observable with both valid and invalid transitions appearing to have equal legitimacy. We can however determine a zones neighbours by applying frequency analysis techniques such as Markov chains.

In terms of logical paths, a Markov chain is a sequence of zone transitions where the current zone that the user is located in is conditionally dependent of the previous zone. That is:

$$P(X_n+1 = x|X_0, X_1, X_2, ..., X_n) = P(X_n+1 = x|X_n) \tag{2}$$

where X is a spatial zone and X_n is the current zone a user is located in. The one-step transitional probability:

$$P(X_n + 1|X_n) \tag{3}$$

is implemented as a transition matrix containing the probabilities of moving from one spatial zone to any other spatial zone in the environment. This matrix is populated by processing the zone transitions contained in logical paths. Once trained one can distinguish between valid and invalid transitions by looking at the transition probabilities. Invalid zone transitions should be associated with lower probabilities than the more frequently occurring valid transitions.

VI. Shopping Street

In this section we discuss the results for an implementation of this work on a busy shopping street using cellular networks.

The aim of our experiments was to identify if people had taken the same physical path by comparing the equivalent logical paths. To implement this users were equipped with Orange SPV C500 cell phones running software to record the signal strengths levels for up to 7 of the nearest base stations on a cellular network. We used a 500 meter stretch of road in the centre of Bristol in the UK as our test environment. Users walked from one end of the road to the other collecting measurement samples.

In total we collected 8457 measurements during 25 passes and encountered 24 different Cell-IDs. We used 15 passes as training data (data used to create the clusters) and 10 passes as test data (data used to assess performance). Zones were created using K-means, with K values (2-28). Once the clusters had been generated, the test data was processed, producing logical paths of the users transitions through the zones representing the spatial environment.

Logical paths were compared using the following equation:

$$s = \frac{\sum_{z=1}^{r} (\frac{x_z}{n})}{r} \tag{4}$$

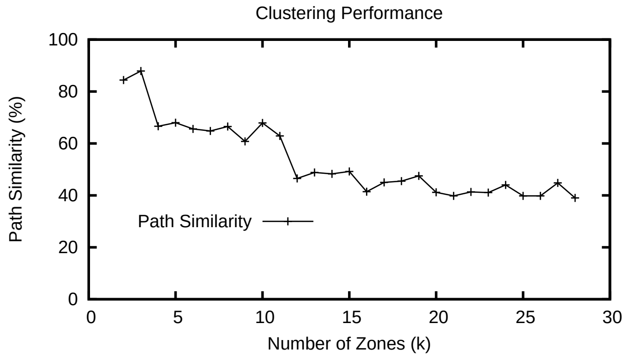

Fig. 3. Clustering performance is assessed by comparing multiple logical paths generated over the same physical path.

where n is the number of logical paths being compared, X_z is the number of matching zones at logical path position z with a path size of r and s is a real number between 1 and 0 representing path similarity with 1 indicating paths are identical and 0 meaning paths share no similarity. The performance of this function is shown in Figure 3.

In our experiments all logical paths represent the same physical path so, in theory, all logical paths should be identical. This is not realistic though as cellular signal strength levels are inherently noisy and the mobile station can only track up to 7 levels at any one time. Instead, our results showed that for this environment performance was optimal with 3 zones, producing a path similarity of 87%. With K values less than 11, a path similarity of 60% or above was achieved, adequate for many applications.

At most, only 10 zones were used, even when K was 28. This served as a good indication that with higher K values the number of zones is not reflective of the environment and available training data. Illustrating it is not possible to force a granularity beyond the limits of the positional measurements.

Using the generated zone representation would not provide a positional granularity precise enough to differentiate between two adjacent shops on the street. But by augmenting the cellular data with WiFi data this level of performance could easily be achieved. It is important to note that zone size is typically unequal. Hence, in places, multiple zones may provide the coverage area for the front of a single shop but at other shop fronts coverage may be provided by a single zone. This is dependent upon the quality of the positional measurements, training data and the spatial environment.

VII. Conclusions and Future Work

We have presented a strategy for developing positioning services that provide users with a qualitative location. The supplied location reflects the quality and reliability of positional measurements obtainable in a particular spatial environment. We have demonstrated that, via a simple calibration phase, the spatial environment can be automatically partitioned into a series of distinguishable zones.

Instead of offering a location and accuracy which may not be met, we offer a zone and a confidence, where the zone depends on the environment. If it is not possible to distinguish between two places then a single, qualitative location will represent both areas.

At present, we have prototyped this work using cellular signal strength levels on a busy shopping street. We achieved promising results, accurately matching logical paths with low numbers of spatial zones. But, as expected, we found that as we increased the number of zones in the environment the performance of the location service decreased. Although performance can be improved by knowledge of the zone topology, ultimately a trade off must be made between positional granularity and measurement reliability. In our experiments, we found cellular signal strength information to be very noisy. The calibration phase gave us an accurate picture of performance in this environment enabling us to build location aware applications that operated within these constraints.

For the future, we wish to assess the viability of using this approach to fuse heterogeneous positional measurements. In particular, WiFi and cellular signal strength levels.

References

[1] C. Jiang and P. Steenkiste, "A hybrid location model with a computable location identifier for ubiquitous computing," in *UbiComp '02: Proceedings of the 4th international conference on Ubiquitous Computing*. London, UK: Springer-Verlag, 2002, pp. 246–263.

[2] M. Bauer, C. Becker, and K. Rothermel, "Location models from the perspective of context-aware applications and mobile ad hoc networks," *Personal Ubiquitous Computing*, vol. 6, no. 5-6, pp. 322–328, 2002.

[3] A. K. Narayanan, "Realms and states: a framework for location aware mobile computing," in *WMC '01: Proceedings of the 1st international workshop on Mobile commerce*. New York, NY, USA: ACM Press, 2001, pp. 48–54.

[4] S. Volz, D. Fritsch, D. Klinec, A. Leonhardi, and J. Schtzner, "Nexus - spatial model servers for location-aware applications on the basis of arcview," in *Proceedings of ESRI European User Conference*, 1999.

[5] M. Beigl, "Using spatial co-location for coordination in ubiquitous computing environments." in *HUC '99: Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, vol. 1707, 1999, pp. 259–273.

[6] E. Trevisani and A. Vitaletti, "Cell-id location technique, limits and benefits: An experimental study," in *WMCSA '04: Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'04)*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 51–60.

[7] D.-J. Shyy and B. Rohani, "Indoor location technique for 2g and 3g cellular/pcs networks." in *LCN '00: Proceedings of the 25th IEEE Conference on Local Computer Networks*, 2000, pp. 264–271.

[8] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *INFOCOM (2)*, 2000, pp. 775–784.

[9] J. Krumm and J. Platt, "Minimizing calibration effort for an indoor 802.11 device location measurement system."

[10] M. Brunato and C. K. Kall, "Transparent location fingerprinting for wireless services," in *Proceedings of Med-Hoc-Net, Mediterranean Workshop on Ad-hoc Networks*, Baia Chia, Cagliari, September 2002.

[11] W. H. Wong, J. K. Ng, and W. M. Yeung, "Wireless lan positioning with mobile devices in a library environment," in *Proceedings of the 3rd International Workshop on Mobile Distributed Computing (MDC)*, vol. 6, no. 6, 2005, pp. 633–636.

[12] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit, "Place lab: Device positioning using radio beacons in the wild," in *Proceedings of PERVASIVE 2005, Third International Conference on Pervasive Computing*, Munich, Germany, 2005.

[13] The Place Lab project, http://www.placelab.org/.

[14] J. Hightower and G. Borriello, "Particle filters for location estimation in ubiquitous computing: A case study." in *Ubicomp*, 2004, pp. 88–106.