Location by Scene Analysis of Wi-Fi Characteristics

B Cook[1, 2, 3], G Buckberry[2], I Scowcroft[2], J Mitchell[1], T Allen[3]

1 University College London, 2 Siemens Communications, 3 Nottingham Trent University

Abstract: This paper provides details of an experiment performed to investigate the use of Scene Analysis to provide location information. The method we used generates the characteristic signature using Position Estimates which are calculated by trilateration of Wi-Fi Received Signal Strength values. We provide details of the conclusions we have made regarding the resolution and error rates that the method provides.

1. Introduction

Over recent years, the potential of Location-based Services has been hotly debated, with several methods being put forward to provide location information. GPS has emerged as the leading method of providing position information but it has limited use indoors. There has been a lot of research into methods which are capable of providing indoor location information, however most of these require specialised equipment or have lengthy set-up procedures. We are therefore developing a system which has neither of these drawbacks. We are utilizing the standard Received Signal Strength (RSS) measurements available on Wi-Fi equipment because Wi-Fi networks are becoming increasingly available, meaning that extra specialised equipment is not required. In order to reduce the set-up time, we are not aiming to provide the location of every person at every location within a building – just at locations that are of interest to a user.

Location using Wi-Fi Signal Strength is generally thought to be too inaccurate indoors due to the variation caused by multipath. In order to reduce the effect of the variation, we calculate a Position Estimate by trilateration of RSS readings. Estimating position in this way is also inaccurate due to limitations of the radio model in cluttered environments. However, by using Scene Analysis to generate location information rather than using the Position Estimates as the actual position, we remove the requirement that the estimates be accurate. All we now require is that Position Estimates calculated at the same location are similar.

Having developed this technique and shown that Position Estimates are indeed repeatable [1], the next step was to determine how far apart locations need to be for the method to distinguish between them. This paper provides details of the experiment we performed to find this resolution, and so find out if it is usable in an office environment.

2. Using Scene Analysis for Location

2.1 Overview of Scene Analysis

Scene Analysis is a pattern recognition method which uses the characteristics of a scene from a particular viewpoint to match patterns. We are interested in static Scene Analysis, in which observed characteristics are compared to pre-stored characteristics for each pattern to determine a match.

The advantage of using Scene Analysis is that accurate physical quantities, such as distance, are not required. However, stored characteristics are needed for each pattern, which could take a long time to collect and require a lot of storage space. Also, changes to the environment may require the characteristics to be re-evaluated.

Scene Analysis has been used in many applications, such as image and speech recognition, as well as location. One of the most well-known indoor location systems, the RADAR system developed by Microsoft [2], uses Scene Analysis to match the signal strengths received from various Wi-Fi Access points to identify location.

2.2 Scene Analysis with Position Estimates

The RADAR system mentioned above requires characteristics, or signatures, to be stored covering the whole area of interest. This allows a 'best fit' method to be used to match observed characteristics, so that the set which produces the closest match is chosen. For our system, however, we only intend to provide locations covering a subset of the whole location area, which is determined by the user. This means we cannot use the 'best fit' method alone, as there is a case when the observed values do not match any of the signatures. Instead we will use a threshold value to determine whether there are any matching signatures. If more than one signature matches, we will use the 'best fit' method to choose between them.

3. Location Experiment

3.1 Purpose of the Experiment

We have performed an experiment to test the Scene Analysis technique in an open-plan office environment. The main aim was to determine the resolution of the method, that is the minimum distance at which the method can distinguish between locations, and the False Accept Rate at this distance. We also used the data to estimate the expected False Reject Rate.

3.2 Experiment Outline

An area of the office was divided into 32 regions corresponding to desk areas. The regions were between 2 and 3 metres apart in an open-plan arrangement with low dividers. A signature was generated at each location consisting of two sets of 150 position estimates. The sets were chosen to have a correlation value greater than or equal to 0.9 (see section 3.3) to reduce the possibility that unusual data was recorded, and statistical outliers were removed.

Verification datasets were also generated at one location, to take the part of observed data in the experiment. These datasets also contained 150 position estimates with statistical outliers removed. Three sets of Verification datasets were taken over one hour each.

The signatures were compared to each other to see how much they changed over distance, and the Verification datasets were compared to the signature for that location to see the variation at the same location over time.

3.3 Matching Datasets

3.3.1 The Correlation Value

In order to match datasets we have used the square of the correlation ratio, η^2 (1), a measure of non-linear correlation.

$$\eta^2 = \frac{\sum_x n_x (\bar{y}_x - \bar{y})^2}{\sum_{xi} (y_{xi} - \bar{y})^2}$$

where y_{xi} is observation i in category x, n_x is the number of observations in category x, $\bar{y}_x = \frac{\sum_i y_{xi}}{n_x}$ and $\bar{y} = \frac{\sum_x n_x \bar{y}_x}{\sum_x n_x}$

(1)

We combined the value of η^2 for the x- and y- co-ordinates of the position estimates to give a single correlation value in the range [0, 1], where 1 is correlated and 0 is uncorrelated.

3.3.2 Correlation Value vs. Distance

In order to evaluate the resolution of the location method, we calculated the correlation value between the signatures of each location and plotted them against the distance between the locations (Figure 1). This allowed us to show that there is a significant relationship between correlation value and distance. The distribution follows a negative exponential distribution, with an R^2 value of around 0.4 showing that 40% of variation in the correlation value can be explained by distance. Although there is a lot of variation in the correlation value, there are clear boundaries above which it does not exceed a particular level after a given distance. From this we have concluded that the correlation value changes enough over the distance tested to be able to distinguish between locations within the test area.

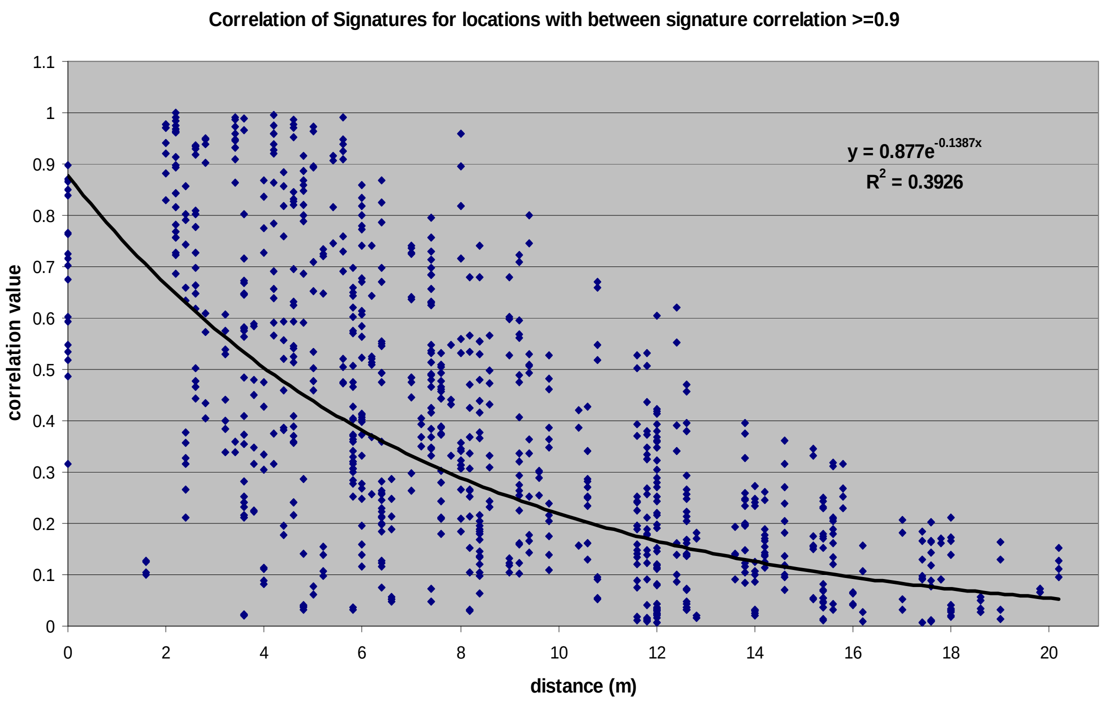

Figure 1 Correlation value vs Distance

3.4 Resolution and Error Rates

The resolution and error rates achievable by the Scene Analysis method are dependent on the correlation threshold used to indicate matching datasets. This value must be chosen to provide reasonable values for the resolution, the False Accept Rate (FAR) at this distance, and the False Reject Rate (FRR). To do this we calculated the resolution and error rates for different thresholds between 0 and 1, at increments of 0.1. We then chose the threshold which provides a suitable trade-off between resolution and False Reject Rate.

3.4.1 False Reject Rate

To find the False Reject Rate (FRR) at different thresholds we calculated the correlation value of each Verification dataset with both datasets of the signature. If either of the signature datasets matched we concluded that the location matched.

As can be seen from Figure 2, there is a large difference in the FRR value between using a correlation threshold of 0.4 and 0.5. We therefore conclude that a correlation value no greater than 0.4 will be needed to provide a reasonable False Reject Rate.

Threshold	FRR
1	100%
0.9	95%
0.8	79%
0.7	63%
0.6	42%
0.5	37%
0.4	16%
0.3	8%
0.2	5%
0.1	5%
0	0%

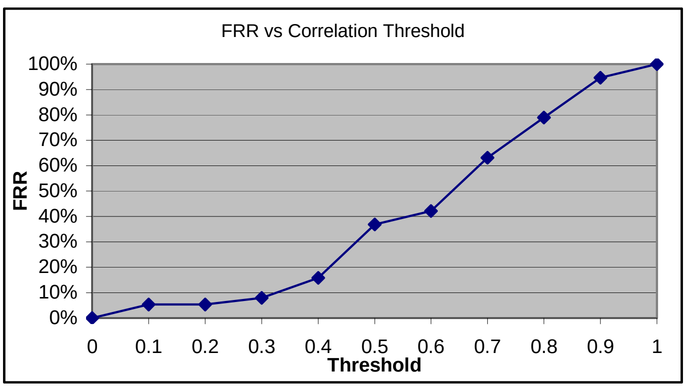

Figure 2 False Reject Rate using different correlation thresholds

3.4.2 Resolution

To determine the possible resolution, we used the correlation value vs. distance data (Section 3.3.2). We calculated the percentage of values greater than or equal to the different correlation thresholds, at increments of 1m from the actual location.

From this data we see that the as the correlation value falls, the resolution increases (Figure 3). Using a correlation threshold of 0.4, the maximum allowable for a reasonable False Reject Rate, we get a resolution of 10m with a False Accept Rate of 6% at this distance, which we believe will be acceptable.

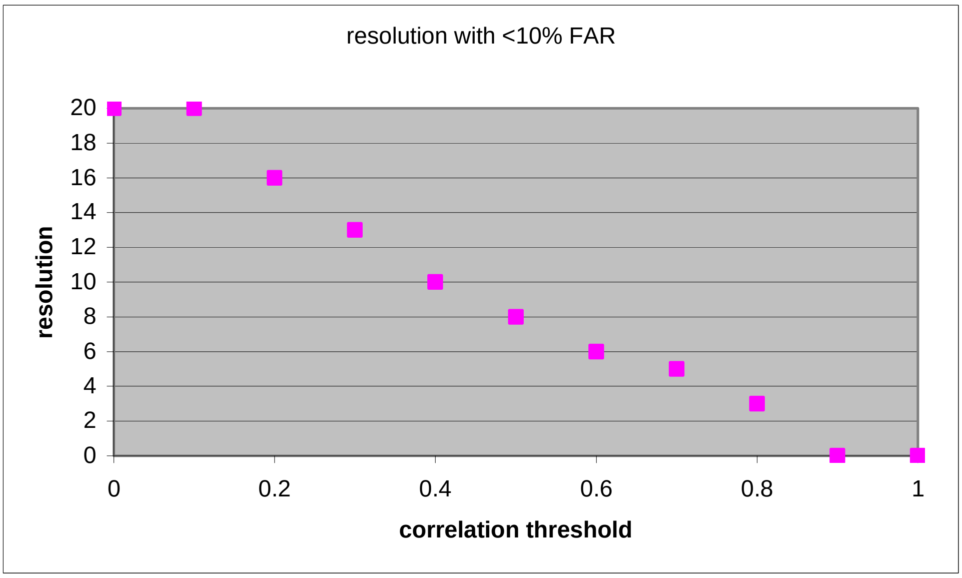

Figure 3 The possible resolution increases as lower correlation thresholds are used.

4. Conclusions

The experiment has shown us that the Scene Analysis method is feasible. Using a correlation threshold of 0.4 provides a reasonable trade-off between having a small resolution and a low False Reject Rate. The resolution we can achieve is 10m with a 6% False Accept rate at this distance, and a False Reject rate of 16%. We believe that these values are suitable for our uses.

We now intend to test the suitability of these limitations by performing a Proof of Concept trial. This trial will test whether the resolution and error rates are acceptable, and that the method is stable over time. We will also use the trial to verify that location information is of use in an office environment.

References

[1] Cook B, Buckberry G, Scowcroft I, Mitchell J, Allen T, "Indoor Location Using Trilateration Characteristics", London Communications Symposium 2005.

[2] Bahl P, Padmanabhan VN, "RADAR: An In-Building RF-Based User Location and Tracking System", Proc IEEE Infocom 2000.