Direct Localisation using Ray-tracing and Least-Squares Support Vector Machines

Benny Chitambira, Simon Armour, Stephen Wales, Mark Beach
Communication Systems & Networks Group
University of Bristol
Bristol, BS8 1UB, UK
Email: {b.chitambira, simon.armour, m.a.beach}@bristol.ac.uk; stephen.wales@roke.co.uk

***Abstract*— This paper evaluates a novel scheme for direct 2D localization that employs least-squares support vector machines, using ray-tracing data. The scheme does not require non-line-of-sight identification or mitigation, which means it can be applied under any conditions. The approach requires perfect knowledge of base-station positions and the ray-tracing data is location specific. This approach shows that when an outage probability of twenty percent is considered, the mobile's location can be determined to within 15m of accuracy in a dense urban environment. Usage and application contexts for this approach are also provided.**

Keywords—localisation; positioning; support vector machines

I. INTRODUCTION

The problem of geolocation in dense urban areas is widely researched, with the aim of providing better accuracy in areas where Global Navigation Satellite Systems (GNSS) cannot provide adequate coverage due to urban canyon. The European GNSS Agency (GSA) reported in 2014 [1] that their tests had confirmed that Galileo will improves location accuracy in challenging environments (urban canyons and indoors) when used in addition to GPS and GLONASS. However, their tests still produced a horizontal accuracy of just 38m in a typical urban canyon environment.

Common mobile signals-based positioning systems, determine the location by first, determining the range (distance from the base-station (BS)) of the mobile. Measurements like the time-delay or the power level, of the received signal, are used in trilateration, and Angle-of-Arrival (AOA) measurements are used in triangulation. In triangulation, the challenge is to determine the true line-of-sight (LOS) angle of arrival at the BS, which is a difficult task in urban multipath environments, when all signals arriving at the BS are in non-line-of-sight (NLOS). In trilateration, the time delay measurements generally consist of a positive bias due to multipath. Both above cases give rise to the need for NLOS identification and mitigation.

There are several ways to achieve NLOS identification and mitigation, one of which is the use of Least Squares Support Vector Machines (LSSVM) [3]. In this study, we take further, the ideas in [3] and apply a similar methodology for LSSVM function estimation, to directly estimate the *x*-coordinates and the *y*-coordinates of the mobile station (MS) positions. Use of support vector machines in these contexts may be thought of as applying artificial intelligence or machine learning concepts, to the problem of localisation. The approach in this study achieves localisation of the mobile without first having to go through NLOS identification or mitigation.

Other related approaches like fingerprinting [2] involve matching the measured/observed received signal quantities, commonly the received signal strength, to the values that are pre-recorded in the fingerprinting database for a particular area. Fingerprinting database is built by collecting measurements per grid, of the area of interest. Positioning accuracy therefore depends on the grid size. Also fingerprinting requires cell matching before correlation with grids around that cell, whereas LSSVM handles BS matching and location estimation within the same model. Fingerprinting employs, either probabilistic algorithms like maximum likelihood, to estimate the position, or deterministic algorithms that calculate the similarity between the UE measurement and the database grid-based measurements. Fingerprinting is commonly talked of as an augmenting scheme to other approaches, to improve accuracy .

II. EXPERIMENTAL SETUP

A. Ray-tracing

A ray-tracing setup similar to the one described in [3] was used to generate data for 3 different areas of the greater city of Bristol, UK. These areas were chosen to represent 3 different environments which are; a dense urban area, an urban peripheral area and farm land. MS positions are randomly placed within the base-stations' coverage area and ray-tracing is run for each BS-MS link. MS positions falling onto obscure areas like court yards do not produce any ray data, so they are excluded from the study.

The key outputs from the ray-tracer, that are used in this study are; the BS and MS locations (*x* and *y* coordinates), the azimuth AOA at BS, the received power and the time delay, for each ray. No noise modeling is incorporated, and the ray-tracer output values are considered to be accurate. See [3] for more detail on the ray-tracing setup and data processing.

B. Localisation perfomance

Performance of this scheme is defined by the MS location error *e*, which is the distance between the true MS position and the estimated position. Location error is calculated as shown in equation 1 below.

$$e_i = \sqrt{[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2} \tag{1}$$

(x_i, y_i) are the actual coordinates for the i^{th} MS position as obtained during ray-tracing, and $(\hat{x}_i, \hat{y}_i)$ are the corresponding LSSVM estimated MS coordinates. Location error cumulative distribution functions (CDF) are plotted to compare performance for different scenarios.

III. Direct Localisation with Support Vector Machines

A. Methodology

LSSVMs can be a robust and effective technique for solving non-linear function or density estimation in linear, kernel-based systems. Typically used for classification and regression as in [3] and [4], they can produce very good results for function estimation of which this study is concerned. Function estimation methodology, using Support Vector Machines (SVMs), seeks to construct a regressor, which is a function $\Re^n \rightarrow \Re$, of the form

$$y(x) = \sum_{i=1}^{N} \alpha_i \psi(x.x_i) + b \tag{2}$$

given a training data set of N data points $\{x_i\, y_i\}_{i=1}^{N}$ where $x_i \in \Re^n$ is the i^{th} input and $y_i \in \Re$ is the corresponding i^{th} "output". α_i are positive real constants and b is a real constant, both which form the parameters of the regressor. $\psi(x, x_i)$ is the kernel. We choose the Radial Basis Function (RBF) kernel in this study because it gives the best validation and test set performance [5]. The LSSVM formulation leads to a linear system that incorporates a hyper-parameter γ which is used to tune the trade-off between the level of tolerable training errors and model complexity [6].

The inputs $x_i \in \Re^n$ form a (N x 5) matrix whose 5 columns are; the BS *x*-coordinate, BS *y*-coordinate, the signal/ray's AOA at BS, logarithm of its time delay, and logarithm of its received power. The output sequences used for training, $y_i \in \Re$ forms a column vector with the *x*-coordinates or the *y*-coordinates of the MS depending on the coordinates being estimated at that point. A data-set size, N, of 10 000 points was used for training, of which half were LOS and half were NLOS. Training yields the regressor tuning parameters and constants, which are then used to estimate the coordinates of the MS for any new given data set. Training data was generated per each considered area and it is that training dataset, that is used for the LSSVM location estimation within that area. Training data from different areas or databases may be merged into a single dataset to allow one-time training, and then re-using of parameters for the localisation stage.

Training is done separately for the *x* and *y* coordinates using the appropriate output sequences. This approach means estimation of the MS position is O(2) as compared to the traditional regression for NLOS mitigation. It is however possible to just estimate, say the *y*-coordinate, and use it together with LOS information where available (via NLOS identification or otherwise), to calculate the *x*-coordinates for those positions that are determined to be in LOS as shown in Fig 1. below. Once the estimate of the *y*-coordinate is obtained, the *x*-coordinate can then be calculated as follows

$$x_i = y_i.tan(\theta_i) \tag{3}$$

Where x_i is the *x*-coordinate corresponding to the *y*-coordinate y_i and θ_i is the AOA. This approach is only suitable for LOS positions. After obtaining the estimates for both the x and y coordinates of the MS, the location error is calculated in the same way as in equation 1.

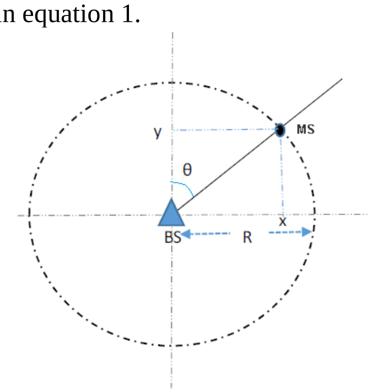

Fig 1. Obtaining the second coordinate for LOS scenarios

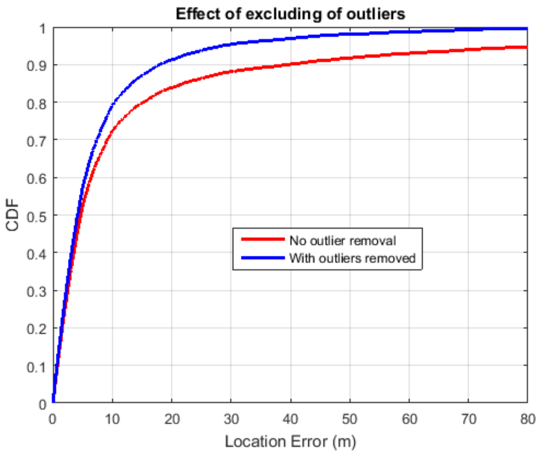

Fig 2. Outlier removal

B. Post-processing and outlier removal

The ray-tracing setup has BS-BS distance of 300m so a coverage radius for each BS, of 150m is considered, for determining outliers. The BS deployment seeks to approximate envisaged 5G deployments, where a dense deployment of small cells is expected. The process of determining and excluding outliers involves calculating the distance d_i between the known BS position and the estimated MS location, as follows

$$d_i = \sqrt{\left[\left(BSx_i - \widehat{MSx}_i\right)^2 + \left(BSy_i - \widehat{MSy}_i\right)^2\right]} \qquad (4)$$

where BSx_i is the i^{th} BS *x*-coordinate and $\widehat{MSx}_i$ is the estimated i^{th} MS *x*-coordinate. The other symbols' meaning follow.

A BS receives multiple rays from an MS and each ray is used to estimate the MS position. For a single MS position, some rays will estimate the position better than others, so those rays that result in the BS-MS distance greater than 150m are discarded. Empirical tests show that more regressor errors start increasing for MS positions beyond 100m from the BS. Outlier removal criteria may be tightened to any distance, but that will create more coverage black spots, hence we settled on 150m. On average, the total number of data points that were excluded because of this criteria were around 10%. Fig 2., shows the effect of excluding those rays that are resulting in outlying MS positions.

IV. Results and Discussion

Ray-tracing data for different areas of the city (Figs 4, 5 and 6) are used and the results for the location error CDF are shown in Fig 3. below.

The LSSVM performance produced best accuracy for the dense urban environment. This is because, for a given MS position, the fingerprint (set of the parameters for rays that are received from that position) are generally unique in a multipath environment. This is demonstrated in Figs 4, 5 and 6 taken from the ray-tracer databases used for the dense urban area, an urban peripheral area and farm land area, respectively.

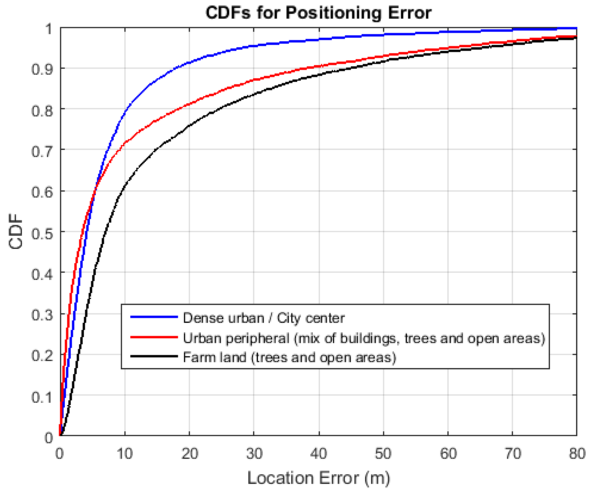

Fig 3. Location error for different environments

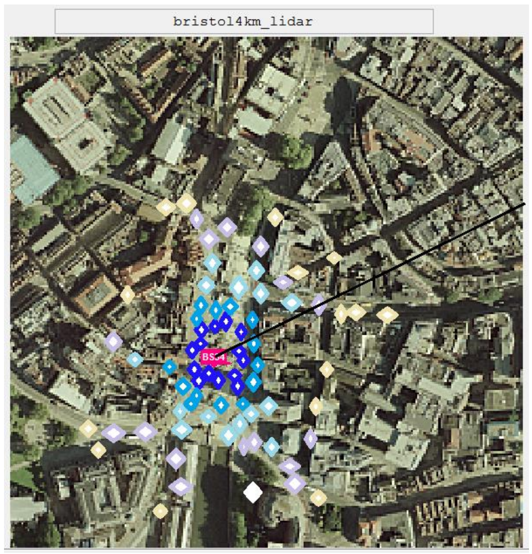

Fig 4. Dense urban area / City center (sampled color-coded positions: same color means positions with same received signal power.)

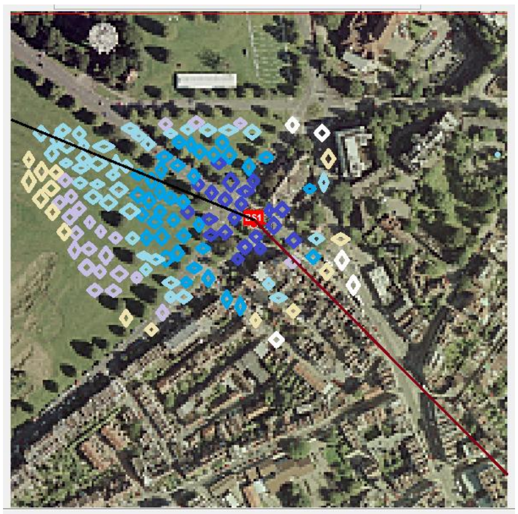

Fig 5. Urban peripheral area

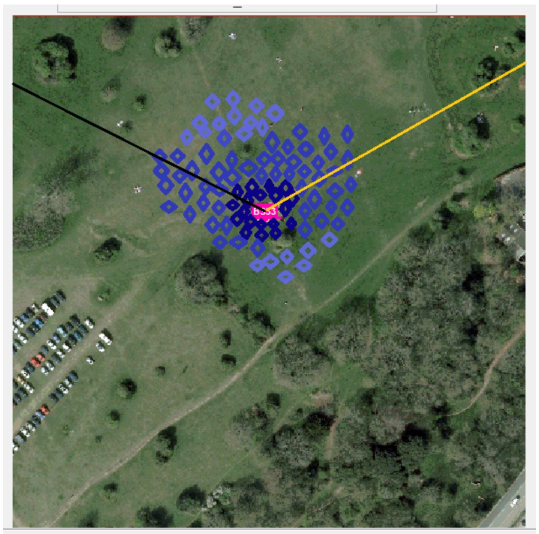

Fig 6. Farmland / trees and open areas

Sampled MS positions are indicated on the figures and they are color-coded according to quantized received signal power. To show why the location accuracy performance is better in dense multipath environments, consider a straight line extending from the BS (line of bearing) outward in any direction. It is easy to see that the line will cross multiple MS positions of the same given received power, in farmland type (LOS) area than in dense urban (NLOS) areas. This is the same reason why AOA is a key metric, as demonstrated later in Fig. 7, because it resolves ambiguities when multiple measurements have same TOA and received power. Peri-urban environment performed better for low resultant location errors because the BS was closer to the buildings, so more MS positions within 50m of the BS, were within the built-up area.

In practical systems, both measurement and model errors (discrepancies between ray-tracing model and the actual environment) would be present.

The sensitivity of the localisation estimate to measurement errors in the three quantities (AOA, TOA and received signal strength) used to determine location, has been investigated by introducing a Gaussian distributed error in each. This also gives an indication of which parameters/quantities that are more sensitive to errors, hence more critical for the performance of the LSSVM regressor. A 5% standard error was introduced in each parameter/quantity, each time performing location estimation, and also in all quantities at once, and performing location estimation, following our methodology.

The results in Fig 7. show that our LSSVM approach is most sensitive to AOA errors. In practice, next-generation wireless systems that employ antenna arrays at the BS, like Massive MIMO [7], will likely provide very accurate AOA estimates, so this approach sits well with envisaged fifth generation (5G) systems.

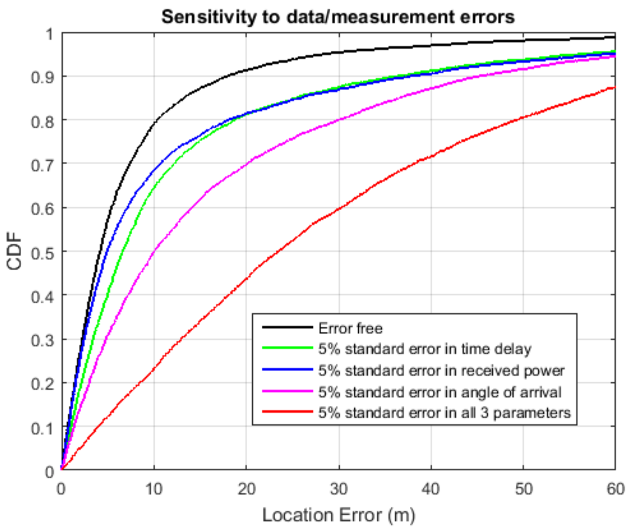

Fig 7. Sensitivity to measurement errors

V. CONCLUSIONS

This study has demonstrated an approach that is especially relevant for dense urban environments. Although these results may be location specific, it is easy to see how they can be relevant for similar urban environments. Extension of this approach, to 3D positioning increases computation, but is straight-forward. Employing this approach requires availability of ray-tracing data for any particular environment, and this is becoming more and more common for most cities as the ray-tracing equipment becomes more portable, and the general availability of such equipment is increasing. Granularity and performance of the scheme can be further controlled by the training data size. A larger training data set improves the tuning parameters. Training can be done per BS, with the tuning parameters stored and referenced per each BS. This further simplifies the primary question "if the BS *m*, received *n* rays, each which records a set of measurements for user *k*, where is the user likely to be located?". Since LSSVM training is done once per given coverage area, this approach can be used for cities, with the ray-tracing database getting updated regularly. The whole process of localising a MS position can then be thought of as a "measure and look-up" process. Artificial intelligence or machine learning, and big data, are popular and promising technologies for the future which are envisaged to become common place in the next decade. This study sits well with these topics, so whilst data availability may be a constraint today, we believe that will not be the case in the future.

ACKNOWLEDGEMENT

The authors acknowledge the financial support of the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training (CDT) in Communications (EP/I028153/1) and Roke Manor Research.

REFERENCES

[1] European GNSS Agency (GSA) [Online]. Available: https://www.gsa.europa.eu/newsroom/news/results-are-galileo-increases-accuracy-location-based-services

[2] Y. Wang, H. Li, X. Luo, Q. Sun, and J. Liu, 'A 3D Fingerprinting Positioning Method Based on Cellular Networks', International Journal of Distributed Sensor Networks, vol. 10, no. 7, p. 248981, 2014.

[3] B. Chitambira, S. Armour, S. Wales, and M. Beach, 'NLOS Identification and Mitigation for Geolocation Using Least-squares Support Vector Machines', in 2017 IEEE Wireless Communications and Networking Conference (WCNC) (IEEE WCNC 2017), San Francisco, USA, 2017.

[4] S. Marano, W. M. Gifford, H. Wymeersch, and M. Z. Win, 'NLOS identification and mitigation for localization based on UWB experimental data', IEEE Journal on Selected Areas in Communications, vol. 28, no. 7, pp. 1026–1035, Sep. 2010.

[5] T. van Gestel et al., 'Benchmarking Least Squares Support Vector Machine Classifiers', Machine Learning, vol. 54, no. 1, pp. 5–32, 2004

[6] J. A. K. Suykens and J. Vandewalle, 'Least Squares Support Vector Machine Classifiers', Neural Processing Letters, vol. 9, no. 3, pp. 293–300, 1999.

[7] P. Harris *et al.*, 'An Overview of Massive MIMO Research at the University of Bristol', *CoRR*, vol. abs/1705.07540, 2017.