# Fast features reduction of radio maps for real-time fingerprint-based wireless positioning systems

M.M. Atia, M.J. Korenberg and A. Noureldin

In fingerprint-based wireless positioning, a high number of wireless access points solicits feature reduction to obtain a compact radio map for accurate real-time positioning. Although principal component analysis (PCA) can be used to reduce dimensionality, PCA is computationally expensive. Additionally, PCA maps the data to a new space where physical meaning of the original features is lost. Presented is a faster features reduction approach using fast orthogonal search which selects the most informative features in the original space. The algorithm is applied to select the most informative access points in a radio map for accurate real-time wireless positioning. Experiments demonstrate the proposed method's superior performance to PCA in terms of speed and slightly better performance in terms of accuracy.

*Introduction:* The Global Positioning System (GPS) does not provide reliable positioning inside buildings due to satellite signal unavailability. The currently popular IEEE 802.11 Wireless Local Area Network (Wi-Fi) [1], on the other hand, provides a strong coverage through wireless access points (APs) distributed inside buildings and, hence, can be used as an alternative wireless positioning system in indoor areas. The Wi-Fi fingerprint positioning technique [1, 2] is one of the most accurate signal strength-based position estimation systems that can provide metre-level accuracy indoors. It utilises a radio map [3] which is a database consisting of known locations along with the power pattern measured in those locations. A power pattern is defined as a set of pairs of AP media access control ID (MAC)/signal strength measured by a Wi-Fi enabled device collected at known locations in an offline phase. In online phase, a Wi-Fi enabled device determines its position by classification based on its current power pattern and the saved power patterns in the radio map. The simplest algorithm for classification using power patterns is the K-nearest neighbours (KNN) algorithm [3].

*Problem description and research objectives:* The basic features in a radio map database are the MAC of APs. Experiments show that incorporating too large a number of APs may deteriorate the positioning accuracy and includes unnecessary computation. The ultimate objective of the presented work is to identify the minimal set of APs in a Wi-Fi area with the highest discrepancy power to be used for power patterns matching in a fingerprint-based Wi-Fi positioning system. This features reduction process must be fast enough to guarantee real-time performance. Principal component analysis (PCA) may be used to reduce features dimensionality as in [4]. However, PCA has two major drawbacks. The first is the expensive computation of the covariance matrix, eigenvectors, and data transformation computation. An observation table of $M$ rows by $N$ columns taking the first $C$ components, the complexity can be estimated as follows: $C_\Sigma = O(MN^2)$ for the covariance matrix computation, $C_E = O(N^3)$ for eigenvectors computation, and $C_T = O(MCN)$ for transformation. Another drawback of PCA is that the new features are combinations of the original features. Thus, the physical meaning of original features is lost.

*Methodology:* The canonical form of a radio map is a table of $M$ rows by $N$ columns. Each row contains a known location and $N$ signal strength measurements (power pattern) from $N$ APs. Our strategy to reduce the feature dimensionality of the radio map without the costly PCA and without transformations is to treat every data column as observations $Y_j[n]$ that need to be modelled using a small subset of the other $N - 1$ data columns. This can be achieved using the following model:

$$Y_j[n] = \sum_{m=0}^{C-1} a_{jm} P_m[n] + e_j[n] \qquad (1)$$

Here $j = 0, 1 \ldots N - 1$, $n = 1, 2 \ldots M$, $P_m[n]$ form a set of size $C$ of basis functions that will be selected from the other $N - 1$ columns set, and $a_{jm}$ are coefficients calculated by optimisation techniques such that the error $\|e_j^2[n]\|$ is minimised. The problem then is reduced to a search in the space of $N$ columns to find $C$ columns that if they are used as basis functions in (1) they would achieve the minimum total mean square error over all data columns ($\sum_{j=0}^{N-1} \frac{1}{M} \sum_{n=1}^{M} e_j^2[n]$) [5].

Finding such columns set is equivalent to finding the most informative 'true' APs in the radio map.

*Fast orthogonal search (FOS):* In orthogonal search techniques [5], the Gram-Schmidt procedure is used to replace the functions $P_m[n]$ in (1) by a set of orthogonal basis functions $W_m[n]$ where the model for a specific $j$ in (1) is represented by the following corresponding model:

$$Y[n] = \sum_{m=0}^{C-1} g_m W_m[n] + e[n] \qquad (2)$$

In orthogonal basis function space, the coefficients $g_m$ that minimise the mean square error over the observations is given by

$$g_m = \frac{\overline{Y[n] W_m[n]}}{\overline{W_m^2[n]}} \qquad (3)$$

The over-bar in (3) denotes the time average. The mean square error is given by:

$$\overline{e^2} = \overline{\left[ Y[n] - \sum_{m=0}^{C-1} g_m W_m[n] \right]^2} = \overline{Y^2[n]} - \sum_{m=0}^{C-1} Q_m \qquad (4)$$

where

$$Q_m = \frac{\overline{[Y[n] W_m[n]]^2}}{\overline{W_m^2[n]}} \qquad (5)$$

The reduction in mean square error resulting from adding a term $a_m P_m[n]$ is $Q_m$. The fast orthogonal search procedure [5] makes use of the fact that it is not necessary to create the orthogonal functions $W_m[n]$ explicitly. Only their correlations with $P_m[n]$, the data $Y[n]$, and with themselves are required.

*FOS features reduction of radio maps:* In an $M$ by $N$ radio map, the aim is to reduce columns from $N$ to $C$ most informative columns where $C < N$. Thus, we will have $N$ observations set and the model that needs to be optimised is given by (1). The significance of a data column is evaluated by adding it to the model in (1) and the total mean square error (RMSE) reduction over all data columns is calculated using (5). In FOS iteration, the column with the greatest RMSE reduction is added to model (1) until $C$ columns are obtained. The complexity of the cross-correlations between all pairs of data columns is $C_{corr} = O(MN^2)$. The complexity of applying FOS mean square error reduction $N$ times is $C_{FOS} = O(MN^2 + N^2 C)$. Owing to the fact that $C$ is much smaller than $M$, the overall complexity is dominant by $O(MN^2)$. By comparing this complexity with that of PCA, we note that the term of $N^3$ resulting from the eigenvectors computations is eliminated and the term of $MCN$ resulting from transformation is also eliminated.

*Experiments and results:* A radio map was built in an indoor $30 \times 30$m area in Queen's University Kingston, Canada (Fig. 1) and a total of 67 reference positions were recorded in a radio map table of 67 points by 132 unique MAC addresses of APs. Radio map features reduction using PCA and FOS was performed separately and the feature reduction processing times of both were recorded. In the FOS-based method, execution stops after adding four data columns to the model in (1). Fig. 1 shows the testing trajectory (in white) in which online power fingerprints measurements were recorded. The KNN positioning algorithm was applied on the FOS-reduced radio map and the PCA-reduced radio map to test the performance of both techniques.
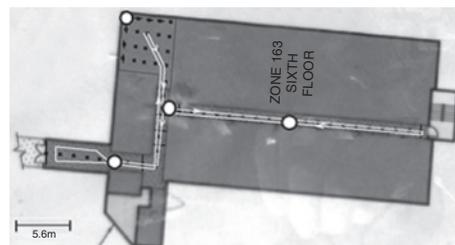


**Fig. 1** *Radio map area and testing trajectory with four heuristically distributed APs*

The following different tests were performed. In each test, the positioning error in terms of RMSE in metres was calculated. In the full radio map test, the 132 MACs were used for positioning obtaining a RMSE of 3.40m. In the PCA-reduced radio map test, a RMSE of 3.9645m was obtained using the first four components in the transformed radio map. In the FOS-reduced radio map test, a slightly better RMSE of 3.3072m was obtained using the four APs suggested by FOS. To obtain a reference solution, four APs were carefully chosen heuristically in predefined places (see Fig. 1) to give the best coverage during the whole trajectory achieving a RMSE of 1.7449m. Fig. 2 shows the error cumulative percentage and Table 1 shows RMSE for all the tests. Table 2 shows the number of multiplications and the feature reduction processing times of PCA and FOS on an Intel Core i5 2.53 CPU 2GB RAM laptop using Matlab 2009 on Windows 7 OS. Table 2 shows that FOS took 0.07 seconds and PCA took 0.1521 seconds including transformations.
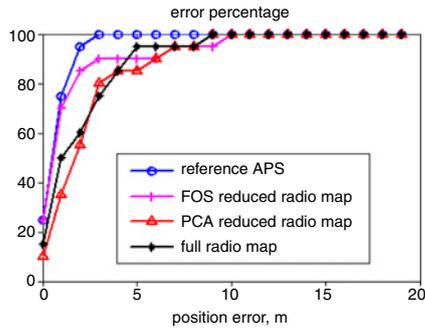


**Fig. 2** *Cumulative positioning error percentage*

**Table 1:** Positioning root mean square error in four test cases

|  | Full radio map | PCA-reduced | FOS-reduced | Reference |
|---|---|---|---|---|
| RMSE | 3.4074m | 3.9645m | 3.3072m | 1.7449m |

**Table 2:** Radio map features reduction processing times

|  | PCA | FOS |
|---|---|---|
| Number of multiplications | 3 502 752 | 1 167 540 |
| Processing time | 0.1521 s | 0.07 s |

*Conclusion:* An alternative features reduction technique based on FOS is presented. FOS eliminates the need for eigenvectors and transformation computation overhead. Additionally, the reduced features are a subset of the original features which gives meaning to the output. Experiments on a 132 features radio map on a Wi-Fi fingerprint positioning system showed that the FOS-based feature reduction approach is 54% faster than PCA. In addition, using only four features suggested by FOS achieved a slightly better accuracy than that of the full radio map with 132 MACs. It also achieved 16.57% less RMSE than that of PCA.

M.M. Atia and M.J. Korenberg (*Department of Electrical & Computer Engineering, Queens University, 99 University Ave, Kingston, ON, K7L 3N6, Canada*)

E-mail: mohamad.m.atia@queensu.ca

A. Noureldin (*Royal Military College of Canada, Electrical and Computer Engineering*)

## References

1 Farivar, R., Wiczer, D., Gutierrez, A., and Campbell, R.H.: 'A statistical study on the impact of wireless signals' behavior on location estimation accuracy in 802.11 fingerprinting systems'. IEEE Int. Symp. on Parallel & Distributed Processing, Rome, Italy, July 2009, pp. 1–8

2 Bahl, P., and Padmanabhan, V.: 'RADAR: an in-building RF-based user location and tracking system'. Proc. IEEE INFOCOM 2000, Tel Aviv, Isrel, 2000, pp. 775–784

3 Roxin, A., Gaber, J., Wack, M., and Nait-Sidi-Moh, A.: 'Survey of wireless geolocation techniques'. GLOBECOM Workshops, Washington, DC, USA, 2007, pp. 1 – 9

4 Xu, Y., Zhou, M., and Ma, L.: 'WiFi indoor location determination via ANFIS with PCA methods'. IEEE Int. Conf. on Network Infrastructure and Digital Content, (IC-NIDC 2009)

5 Korenberg, M.J., and Adeney, K.M.: 'Iterative fast orthogonal search for modeling by a sum of exponentials or sinusoids', *Ann. Biomed. Eng.*, 1998, **26**, pp. 315–327