Indoor localisation based on fusing WLAN and image data

Milan Redžić, Conor Brennan and Noel E O'Connor
CLARITY: Centre for Sensor Web Technologies, Dublin City University, Ireland.
Email: {milan.redzic,brennanc,oconnorn}@eeng.dcu.ie

***Abstract*—In this paper we address the automatic identification of indoor locations using a combination of WLAN and image sensing. We demonstrate the effectiveness of combining the strengths of these two complementary modalities for very challenging data. We describe a fusion approach that allows localising to a specific office within a building to a high degree of precision or to a location within that office with reasonable precision. As it can be orientated towards the needs and capabilities of a user based on context the method becomes useful for ambient assisted living applications.**

Keywords: wireless communications, WLAN 802.11, Naive Bayes, SURF vocabulary tree, fusion

I. Introduction

Due to complex indoor environments, users should use more than one modality in order to improve localisation accuracy and precision [1], [2], [3]. This paper addresses the automatic identification of indoor locations using WLAN technology in addition to image sensing. By fusing these modalities we hope to obtain better performance than using them individually. Whilst GPS has become synonymous with user localisation, its robustness can be called into question. Outdoors, GPS signals can be affected by obstacles, multipath propagation and tall buildings causing serious errors in localisation. Indoors, GPS signals are weak or non-existent.

Using WLAN for indoor localisation has given promising results, but its performance is subject to change due to multipath propagation and changes in the environment [4], [5]. Recently researchers have investigated image-based localisation, for example in [6], but the limitations of this approach are occlusion, changes in lighting, noise and blur. Many localisation methods that have been proposed are based on fusion of UWB and WLAN, WLAN and RF tags (indoors) and GPS and WLAN (outdoors) [7], [8], [9], [10]. There are only a few techniques based on fusion of RF and image sensing methods. These fusion algorithms were used to build active tracking systems based on particle filtering models [11], [5]. For previously mentioned reasons, we propose combining WLAN and image data as complementary sensor modalities. Our motivation for this is that nowadays, any cellphone can be used as a WLAN and image data gathering hub e.g. see the Campaignr (http://www.campaignr.com) micropublishing platform [12]. As it can be orientated towards the needs and capabilities of a user based on context the method becomes useful for ambient assisted living applications. In this paper, we present the results for locating a user to within a specific office in a building or to a location within that office.

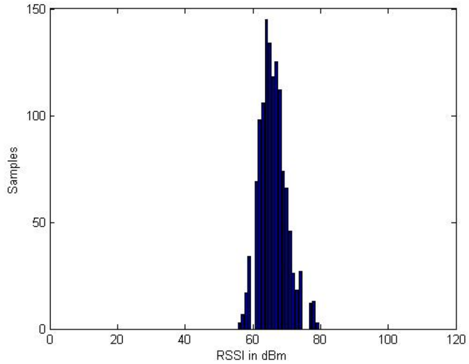

Fig. 1. RSSI histogram of one access point consisting of signal strengths of all orientations taken at one CP. RSSI is defined as the absolute value of RSS, given in dBm

II. Experimental Setup

For our experiments we use 20 offices of average size $8.9m^2$. Within each office we use 5 calibrations points (CP), A, B, C, D & E. Each orientation of a CP (N, S, W and E) has 8 (640×480 pixel) images and 300 associated received signal strength (RSS) observations taken with camera and laptop respectively. An observation consists of RSSs from up to 14 access points. An example of histogram is given in fig. 1. In total we gathered $5,000$ images, of which $3,200$ were used for training and $1,800$ for testing, and $125,000$ signal strengths observations of which $120,000$ were used for training and $5,000$ for testing. Offices are next to each other and look very similar inside thus resulting in very challenging data for both WLAN and image-based localisation methods (see examples in fig 2).

For testing we used one image and one signal strength observation per CP and tested how often we could localise to the correct CP. We also present results for localising to a given office whereby the office selected as the location is that associated with the 1^{st} ranked CP. Hence it is often possible to identify the correct office even if the identified CP is incorrect provided it is in the same office as the actual CP location. We examined localisation precision for 5 different combinations of 1, 2 and 3 CPs per office. We used two precision measures: normal precision and average precision. Average precision

Fig. 2. Some of the images used in the experiments

takes into account not only the top ranked guess but the entire location ranking and therefore gives more information. In case of N images (or signal strength observations) to evaluate, average precision is computed as $AVP = \sum_{k=1}^{N} P_k/N$ where P_k represents the position of the correct location in the k^{th} test.

III. Localisation methods

A. WLAN-based localisation

Probabilistic WLAN-based localisation techniques presume a priori knowledge of the probability distribution of the user's location [4], [13]. A Naive Bayes method [4] was employed which takes into account the access points' (APs) signal strength values (RSS) and also the frequency of the appearance of these APs. A CP's signature is defined as a set of W distributions of signal strengths of W APs and a distribution representing the number of appearances of W APs received at this CP. We denote by C the CP random variable where K is the number of CPs, $X_m \in \{1, 2, ..., W\}$ represents the m^{th} AP random variable, $Y_m \in \{s_1, ..., s_V\}$ is the signal strength that corresponds to m^{th} AP where W is number of APs, M is number of APs of an observation and V is number of discrete values of signal strength. From a set of N training observations $D = \{\mathbf{o}_1, \mathbf{o}_2, ..., \mathbf{o}_n\}$ where $\mathbf{o}_n = (c^{(n)}, x_1^{(n)}, y_1^{(n)}, ..., x_M^{(n)}, y_M^{(n)})$, $n = 1, .., N$ we can then estimate the signature parameters. The joint distribution $P(C, X_1, Y_1, ..., X_M, Y_M)$ is given by:

$$P(C) \prod_{m=1}^{M} P(X_m|C) P(Y_m|C, X_m) \tag{1}$$

Let the identity function $I(a, b) = 1$ if $a = b$ else $= 0$, in the Naive Bayes estimation framework sufficient statistics are [4]:

$$n_c = \sum_{n=1}^{N} \sum_{m=1}^{M} I(c^{(n)}, c) \tag{2}$$

$$n_c^{(x)} = \sum_{n=1}^{N} \sum_{m=1}^{M} I(c^{(n)}, c) I(x_m^{(n)}, x) \tag{3}$$

$$n_{c,x}^{(y)} = \sum_{n=1}^{N} \sum_{m=1}^{M} I(c^{(n)}, c) I(x_m^{(n)}, x) I(y_m^{(n)}, y) \tag{4}$$

The parameters of $P(X_m|C)$ are estimated as

$$\hat{\pi}_c^{(x)} = \frac{n_c^{(x)} + 1}{n_c + W} \tag{5}$$

and the parameters of $P(Y_m|C, X_m)$ as

$$\hat{\gamma}_{c,x}^{(y)} = \frac{n_{c,x}^{(y)} + 1}{n_c^{(x)} + V} \tag{6}$$

Eventually at the prediction step we have:

$$l_i = P(c) \prod_{m=1}^{M} P(x_m|c) P(y_m|c, x_m) \tag{7}$$

The algorithm chooses the location which maximises l_i as being the user location. We rescaled these probabilities to sum to one and denoted their new values as the CP confidences, p_i.

B. Image-based localisation

For image-based localisation, we use an interest point based approach [14] using a variation of a hierarchical vocabulary tree [15] to efficiently match query images of a specific CP to the image training dataset of all CPs – see figure 3. 64-dimensional SURF descriptors were used [14].

SURF (Speeded Up Robust Features) is a robust image detector and descriptor that is used in computer vision tasks. It is inspired by SIFT descriptor [16] but it is several times faster and more robust against different image transformations than SIFT. It builds on the strengths of the best existing detectors and descriptors [17] which gives novel state-of-the art detection, description, and matching steps. Interest points must be selected at distinctive locations (T-junctions, corners, blobs). The main property of the detector is its robustness to changes. Thus it should be reliable in finding the same physical interest points under various viewing conditions. It uses a Haar wavelet approximation of the determinant of Hessian blob detector:

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}, \tag{8}$$

where $L_{xx}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point $\mathbf{x}$, and similarly for $L_{xy}(\mathbf{x}, \sigma)$ and $L_{yy}(\mathbf{x}, \sigma)$. The determinant approximation would be:

$$\det(\mathcal{H}_{\text{approx}}) = D_{xx} D_{yy} - (w D_{xy})^2 \tag{9}$$

Then every interest point's neighborhood is represented by a feature vector. In order to build the descriptor, one has to calculate orientation and descriptor vector for each interest point. Orientation of an interest point was calculated using Haar wavelet responses in x and y direction within circle of radius $6s$ around the same interest point (s is the scale at which the interest point was detected). The horizontal and vertical responses within the window are summed and yield a local orientation vector. The longest such vector among all windows gives the orientation of the interest point. The descriptor vector is calculated by splitting up the square region around interest point (the square size is $20s$) into 16 small sub-squares (4×4 within one square). Then one has to compute Haar wavelet responses at 5×5 regularly spaced sample points. There are four of them: $\sum d_x$ (the sum of Haar wavelet responses in horizontal direction) and $\sum d_y$ (the sum of Haar wavelet responses in vertical direction) and two sums of absolute values $\sum |d_x|$ and $\sum |d_y|$. This results in a descriptor vector for all 4×4 sub squares regions of length 64.

Eventually, the descriptor vectors are matched between different images. Every interest point in the test image is compared to every interest point in the training image by calculating the Euclidean distance between their descriptor vectors. The nearest neighbor ratio matching strategy gives pair (match) detected, if its distance is closer than e.g. $T = 0.7$ times the distance of the second nearest neighbor. Usually T is between 0.6 and 0.8. Since this measure is asymmetrical, we can also compute the matches in the reverse direction (from the training to the test image) and those that appear in both directions (bidirectional matches). Such matches are found to be very stable and strong indicators of a good match.

The SURF features from all $3,200$ database images were associated with the image and the CP to which they belonged. The features were split into two groups (denoted ± 1 respectively) based on the sign of the Laplacian which halves the search time. For each group, we created a hierarchical tree clustering the descriptors using the K-means algorithm repeatedly. This algorithm for partitioning (or clustering) N features into K disjoint subsets S_j containing N_j features minimizes the sum-of-squares criterion:

$$\mathbf{J} = \sum_{j=1}^{K} \sum_{n \in S_j} \|\mathbf{x}_n - \mu_j\|^2 \qquad (10)$$

where $\mathbf{x}_n$ is a vector representing the n^{th} data point and μ_j is the geometric centroid of the data points in S_j. The algorithm does not achieve a global minimum of $\mathbf{J}$ over the assignments since it uses discrete assignment rather than a set of continuous parameters. The algorithm consists of a simple re-estimation procedure as follows. Initially, the features are assigned at random to the sets. For step 1, the centroid is computed for each set. In step 2, every feature is assigned to the cluster whose centroid is closest to that feature. These two steps are alternated until a stopping criterion is met. This means that either there is no further change in the assignment or the algorithm reaches certain threshold value (for which we stop the iteration process). We also tried to use adaptive threshold values that depended on the previous iterations but this increased computational complexity. Nevertheless it might be taken into account in the future analysis.

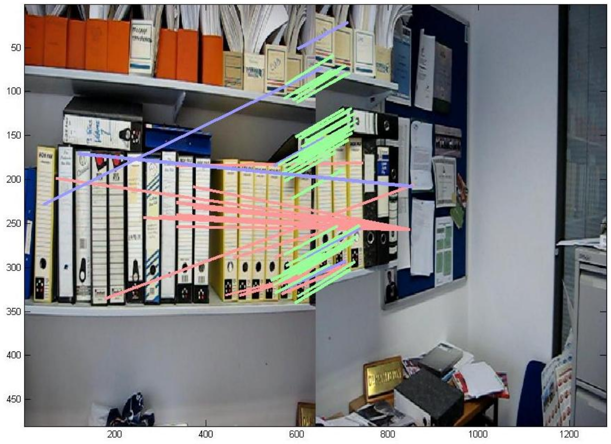

Fig. 3. SURF matching between two different images for the same CP. Unidirectional matches in right image that correspond to left image are represented with red lines (vice versa for blue lines) and bidirectional matches with the green lines

Initially, K clusters were created, then within each cluster, K more clusters, and so on until the last cluster contained less than K descriptor elements. In the case of query image, its SURF descriptors and the (corresponding) signs of the Laplacian were extracted and a match for each descriptor was found using $+1$ or -1 hierarchical tree. Since match was labelled with the image and location from which it was extracted, it therefore casted one vote for the location it belongs to. After each descriptor had voted for a location, we then had a ranked list of locations, from the most likely to the least likely one. We assigned a confidence for each CP (q_i) as the ratio of the number of votes associated with that CP and the total number of votes.

IV. Data fusion

To perform fusion, we take confidences p_i and q_i from both sensing modalities P (WLAN) and Q (image) into account. Here, i refers to a given CP. The first ranked, the second ranked, the third ranked, sorted confidences are denoted by p_{max1}, p_{max2}, p_{max3}, respectively (similarly for Q). We decided to use a large training dataset of confidences in building a robust fusion function which would be reliably used on (unknown) testing data. First, let us define $P_{ij} = p_{maxi} - p_{maxj}$. Observing P_{12} and Q_{12} in many training confidence pairs we concluded that for values P_{12} and/or Q_{12} beyond some reliably large thresholds, denoted T_1 and T_2, we were sure that the correct CP (location) was the 1^{st} ranked one, based either on P or Q (or both).

Moreover we deduced that introducing multiplication ($p_i q_i$) and/or addition ($p_i + q_i$) functions under some conditions can improve precision (and/or average precision) even more. This improvement is small though and the main improvement comes from the previous steps. Also we found that the ranking of the correct location did not fall below some positions in both sets of rankings (m^{th} position for P and the n^{th} for Q modality). If none of conditions are satisfied we decided

No of CPs	P_W	P_I	P_F	P_{WO}	P_{IO}	P_{FO}
1	70.00	57.83	76.09	70.00	57.83	76.09
2	64.78	51.74	72.39	72.61	60.87	76.96
3	59.42	46.67	68.41	75.94	65.37	81.45

TABLE I

LOCALISATION RESULTS: P_W, P_I, P_F DECREASE WHILE P_{WO}, P_{IO}, P_{FO} INCREASE WHEN THE NUMBER OF CPS PER OFFICE INCREASES

to take the ranking of the one with $min(n, m)$. Eventually, the main steps in the fusion process are given in the eq. 11. Here f_i represents fusion confidence and k_i confidence of the method to which $min(n, m)$ corresponds. The location output by the algorithm is the one with the maximum value of the fusion confidence.

$$f_i = \begin{cases} p_i, & P_{12} \geq Q_{12} \wedge P_{12} \geq T_1 \wedge Q_{12} \geq T_2 \\ q_i, & Q_{12} \geq P_{12} \wedge P_{12} \geq T_1 \wedge Q_{12} \geq T_2 \\ p_i, & P_{12} \geq T_1 \wedge Q_{12} < T_2 \\ q_i, & Q_{12} \geq T_2 \wedge P_{12} < T_1 \\ k_i, & \text{else} \end{cases} \qquad (11)$$

V. RESULTS

The fusion function can show the behaviour of precision considering the top N ranked results, thus illustrating how often each modality returned the correct location as the top ranked result, 2^{nd} ranked results, and so on and also how precision increases if the top N ranked results are considered. Overall results are presented in the table I. The left hand side of the table shows results for the precision *on average* when using 1, 2 and 3 CPs per office using WLAN data only (P_W), image data only (P_I) and the fusion of both modalities (P_F). Every CP represents a different location. The right hand side of the table shows results where we only try to localise to the correct office, rather than the correct location within the office. The office chosen is that associated with the 1^{st} ranked CP. This gives the precision to a particular office. P_{WO}, P_{IO} and P_{FO} are WLAN, image and the fusion precision *on average* respectively.

From the table it is clear that fusion of WLAN and images significantly improves the performance over using either approach on its own. The performance variation for the localisation to within an office obtained by using a variable number of CPs also gives an interesting conclusion. The quality of the results improve as we increase the number of CPs used in each office, but acceptable performance is obtained using a single CP: 76.09% *on average*. Thus, the manual data collection stage for model creation outlined in section II is viable as it only needs to be performed once (i.e. at one CP) to obtain reasonably accurate performance.

VI. CONCLUSION

In this work, we presented results combining two complementary data sources for classifying indoor locations. By fusing them we achieve better performance than any individual modality. Thus we demonstrated the effectiveness of the method for very challenging data. Use of images is justified as there were situations where WLAN broke down. Moreover, we have to collect the images as they give contextual information about user's activities, so it does not bring extra costs in terms of additional capture. Future work will investigate the possibility of seamlessly tracking a user indoors, using dynamic confidence-based weighting between these modalities, and more sophisticated classifiers such as neural networks.

ACKNOWLEDGMENT

This work is supported by Science Foundation Ireland under grant 07/CE/I1147.

REFERENCES

[1] M. Redzic, C. O'Conaire, C. Brennan and N. O'Connor, *A Hybrid Method for Indoor User Localisation*, In EuroSSC 2009 - 4th European Conference on Smart Sensing and Context, 2009.

[2] C. O'Conaire, K. Fogarty, C. Brennan and N. O'Connor, *User Localization using Visual Sensing and RF signal strength*, In Sixth ACM Conference on Embedded Networked Sensor Systems (SenSys), 2008.

[3] M. Redzic, C. O'Conaire, C. Brennan and N. O'Connor, *Multimodal Identification of Journeys*, In Proceedings of International Conference on Database and Expert Systems Applications (DEXA), 2010.

[4] K. Tran, D. Phung, B. Adams and S. Venkatesh, *Indoor location prediction using multiple wireless received signal strengths*, In Proceedings of Australasian Data Mining Conference (AusDM), 2008.

[5] S. Mazuelas, A. Bahillo, R.M. Lorenzo, P. Fernandez, F.A. Lago, E. Garcia, J. Blas and E.J. Abril, *Robust Indoor Positioning Provided by Real-Time RSSI Values in Unmodified WLAN Networks*, IEEE Journal of Selected Topics in Signal Processing, vol.3, no.5, pp.821-831, Oct. 2009.

[6] L. Ledwich and S. Williams, *Reduced SIFT features for image retrieval and indoor localisation*, In Australian Conference on Robotics and Automation, 2004.

[7] S. Zirari, P. Canalda and F. Spies, *WiFi GPS based combined positioning algorithm*, In IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS), 2010.

[8] R. Hansen, R. Wind, C.S. Jensen and B. Thomsen, *Seamless Indoor/Outdoor Positioning Handover for Location-Based Services in Streamspin*, In Tenth International Conference on Mobile Data Management: Systems, Services and Middleware (MDM), 2009.

[9] M. Kourogi, N. Sakata, T. Okuma and T. Kurata, *Indoor/Outdoor Pedestrian Navigation with an Embedded GPS/RFID/Self-contained Sensor System*, In 16th International conference on Artificial Reality and Telexistence (ICAT), 2006.

[10] N. Viandier, F.D. Nahimana, J. Marais and E. Duflos, *MRERA (Minimum Range Error Algorithm): RFID - GPS Integration for vehicle navigation in urban canyons*, In IEEE Position, Location and Navigation Symposium, 2008

[11] T. Germa, F. Lerasle, N. Ouadah and V. Cadenat, *Vision and RFID data fusion for tracking people in crowds by a mobile robot*, Computer Vision and Image Understanding (CVIU), 114(6):641–651, June 2010.

[12] A. Joki, A.J. Burke and D. Estrin, *Campaignr: A Framework for Participatory Data Collection on Mobile Phones*, Papers, Center for Embedded Network Sensing, UC Los Angeles, 2007.

[13] A. Paul and E. Wan, *RSSI-based indoor localization and tracking using sigma-point Kalman smoothers*, IEEE Journal of Selected Topics in Signal Processing, 3(5):860-873, 2009.

[14] H. Bay, T. Tuytelaars and L.V. Gool, *SURF: Speeded Up Robust Features*, Computer Vision and Image Understanding (CVIU), 110(3):346-359, August 2008.

[15] D. Nister and H. Stewenius, *Scalable recognition with a vocabulary tree*, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.

[16] D.G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, In International Journal of Computer Vision (IJCV), 2003.

[17] K. Mikolajczyk and C. Schmid, *A Performance Evaluation of Local Descriptors*, IEEE Trans. on PAMI, Vol.20(10):1615-1630, 2005.