User tracking using a wearable camera

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*—This paper addresses automatic indoor user tracking based on fusion of WLAN and image sensing. Our motivation is the increasing prevalence of wearable cameras, some of which can also capture WLAN data. We propose a novel tracking method that can be employed when using image-based, WLAN-based and fusion-based approach only. The effectiveness of combining the strengths of these two complementary modalities is demonstrated for a very challenging data.**

Keywords: wearable cameras, WLAN 802.11, SURF vocabulary tree, tracking, Viterbi, HMM

I. INTRODUCTION

Wearable camera technology has evolved to the point whereby small unobtrusive cameras are now readily available, e.g. the Vicon Revue. This has allowed research effort to focus on analysis and interpretation of the data that such devices provide. Due to complex indoor environments and given the modalities that are currently available, using more than one improves accuracy [1], [2]. Using WLAN for indoor tracking has given promising results, but its performance is subject to change due to multipath propagation and changes in the environment [3]. Recently researchers have investigated image-based tracking but the limitations are occlusion, changes in lighting, noise and blur. There are only a few techniques based on fusion of RF and image sensing [2]. This paper addresses the automatic tracking of a user indoors using fusion of WLAN and image data. As it can be orientated towards the needs and capabilities of the user based on context the method is potentially useful for ambient assisted living applications.

II. EXPERIMENTAL SETUP

For our experiments we use 20 offices of average size $8.9m^2$. Within each office we have 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. 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 tracking methods. Humans were present in these offices which makes the tracking more difficult. We only use one CP per office: either A, B, C, D or E. Thus we have 5 different experimental scenarios. Times measured between consecutively visited locations were collected using a standard stopwatch. A user average speed of walking is approximately 1.1m/s and the user is able to pass a three meter distance in approximately 2.73 seconds. Using this approximation and the University building map together with its scale one is able to reconstruct real distances from the map and calculate all the times. The times are also checked in real world scenario thus showing robustness of this approach. The walking path is chosen to be fixed and along a line that halves the corridor next to the offices. In the testing phase one image and/or one signal strength observation per CP together with time interval ($t^*_{i,j}$) measured between two consecutively visited locations denoted by i and j ($i \neq j$, $1 \leq i, j \leq n$, $n \in N^+$) are used to track the user. n denotes the number of *visited* locations and can be greater than the total number of locations (20 in this work).

III. LOCALISATION METHODS

A. WLAN-based localisation

A Naive Bayes method [3] 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 RSSs 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 (locations), $X_m \in \{1, 2, ..., W\}$ represents the m^{th} AP random variable, $Y_m \in \{s_1, ..., s_V\}$ is the RSS that corresponds to m^{th} AP where W is number of APs, M is number of APs of an observation. At the prediction step we use eq. 1. 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 (location) confidences, p_i.

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

B. Image-based localisation

For image-based localisation, we use a hierarchical vocabulary tree [4] based on SURF to match query images of a specific CP to the image dataset of all CPs. The SURF features [5] 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. For each group, we created a hierarchical tree clustering the descriptors using the K-means algorithm repeatedly. Each match cast one vote for its associated location. After each descriptor had voted for a location, we then had a ranked list of locations, from the most to the least likely. 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.

C. Data fusion

To perform fusion, we take confidences p_i and q_i from both sensing modalities P (WLAN) and Q (image) into account. We built the fusion function described in [2]. The fusion confidence, denoted by f_i, represents a combination of confidences of the both sensing modalities, p_i and q_i.

IV. Novel tracking method

Let us denote by $t_{i,j}$ and $t^*_{i,j}$ time intervals measured in the training and the testing phase respectively between any two consecutively visited locations i and j. Here we refer to i and j as the location output by any of three possible methods used (WLAN-based, image-based and the fusion). Also let us denote by t^k, the k^{th} the nearest time interval to the $t^*_{i,j}$ in the training phase, such that it refers to locations i and j which are output by any of the three methods. If i or j is not obtained by the algorithm output we discard that t^k and do not include it (and its corresponding i and j) in the tracking process. Transitional probability, $T^k_{i,j}$, which says how likely the user passes by the a pair of locations i and j, $i \neq j$, is derived and given in equation 2.

$$T^k_{i,j} = 1 - \frac{\left|t^*_{i,j} - t^k\right|}{\max_k \left\{\left|t^*_{i,j} - t^k\right|, k \geq 1\right\}}, \quad (1 \leq i, j \leq n) \quad (2)$$

At every location the user can estimate position using either WLAN-based (p_i), image-based (q_i) or fusion-based approach (f_i) and obtain the ranking of possible locations from the most probable to the least probable. For a path consisting of several locations (e.g. $I-J-K-L-M-P$ where $1 \leq I, .., P \leq 20$ represent different locations with length equal to n) the total probability consists of the sum of probabilities of being at these locations and the sum of the transitional probabilities of visiting every two consecutive locations ($I-J$, $J-K$, $K-L$, $L-M$ and $M-P$). Equation 3 calculates the probability of visiting several locations.

$$P_{l_n} + \sum_{L_i=l_1}^{l_{n-1}} P_{L_i} + T^k_{L_i,L_{i+1}} \quad (3)$$

where P_{L_i} refers to the probability of being at location L_i and $T^k_{L_i,L_{i+1}}$ refers to the transitional probability $T^k_{i,j}$ as explained and given in equation 2. For each location we get a ranked list of possible locations from the most to the least probable. For a testing time stamp between two consecutively visited locations i and j we can find a ranked list of location pairs whose times (from the training phase) are very similar to the testing time-stamp (they are ranked as well from the most to the least probable). The top k ranked time stamps are chosen (as explained before), denoted by t^k where $k \in N^+$, and since it is known which location pair this particular time stamp belongs to, it can be connected to the *same* two location outputs given by any of the modalities used. Probabilities of being at specific locations and the corresponding transitional probabilities are normalised to $[0, 1]$ interval to reliably represent the influence of each of $(n-1)$ sections. Then these probabilities are added and the process is repeated (as given by equation 3) for all other locations until the last visited location is reached. Thus we have k different sequences each consisting of n locations. The one with the highest probability value gives the order of visited locations.

Dataset ID	P_W	P_I	P_F
A	73.41	61.66	82.66
B	76.18	62.74	84.71
C	66.39	67.19	76.25
D	71.83	57.24	79.02
E	65.42	59.82	82.83
Avg.	70.65	61.73	81.11

TABLE I
RESULTS: P_W, P_I, P_F REPRESENT PRECISION (IN %) WHEN USING WLAN, IMAGE AND FUSION METHOD RESPECTIVELY. ALSO THE LAST LINE OF THE TABLE SHOWS THE RESULTS ON AVERAGE THUS DEMONSTRATING THE EFFECTIVENESS OF THE APPROACH

V. Results

Table I shows the results comparing tracking performance when using either data source and the combination of both sources. Not only does the combination of both sources increase the performance, the difference between them is notably reduced.

VI. Conclusion

In this work, we presented results using two complementary data sources for indoor user tracking. By fusing them we achieve better precision than using them individually. Images are used in cases when WLAN breaks down or is unreliable. Moreover, they give contextual information about the user activities and do not bring extra costs in terms of additional capture.

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] M. Redzic, C. Brennan and N. O'Connor, *Dual-sensor fusion for indoor user localisation*, In Proceedings of International Conference on Multimedia, ACM Multimedia, 2011.

[3] 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.

[4] M. Redzic, C. Brennan and N. O'Connor, *Indoor localisation based on fusing WLAN and image data*, In IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN), 21-23 Sept 2011, Guimaraes, Portugal.

[5] 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.