Multimodal Identification of Journeys

Milan Redžić, Ciarán Ó Conaire, Noel E O'Connor
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Ireland
Email: {milan.redzic,oconaire,oconnorn}@eeng.dcu.ie

Conor Brennan[‡]
RF Modelling and Simulation Group
Dublin City University
Ireland
Email: brennanc@eeng.dcu.ie

***Abstract*—Due to the ubiquity of localisation technology, users now have the ability to keep a record of their own location, as a kind of 'location diary'. Such a large collection of data can become unmanageable without some way to structure that data to make it useful and searchable. We address this problem of structuring location data by proposing a framework for classifying the data into often-traversed routes. In this work, commonly traversed routes are identified with clusters based on sensed data. Our framework does not rely on any one source of location information, but can fuse data from multimodal localisation sources. We demonstrate the effectiveness of our algorithm by examining the combination of GPS, wireless signal strength readings and image-matching on very challenging data in a variety of environmental conditions. By fusing these three modalities we obtained better performance than any individual or combination of two modalities. As it can be orientated towards the needs and capabilities of the user based on context, this method becomes useful for some ambient assisted living applications.**

***Keywords*-Multimodal data-fusion; GPS; WLAN; image matching; MDTW; SURF**

I. Introduction

The ubiquity of localisation technology is allowing users to regularly collect location data. This large collection of data is potentially useless without some means of structuring the data to make it understandable and searchable. In this paper, we address this problem by examining the automatic identification of often-traversed routes for assisted living applications. Such applications of using large amounts of location data can be of benefit to a variety of users. For example, runners may wish to know how often they take a particular route whilst jogging. Localisation of people is considered to be a key task in ambient assisted living platforms. In caring for the elderly, allowing a mobile device to automatically determine whether they have deviated from their normal routine can trigger a notification to their carers [1]. Lifelogging describes recording different aspects of one's daily life, in digital form, for exclusive personal use. In the lifelogging community, route matching can add valuable structure to the months and years of recorded daily activities [2].

The problem of route-matching is complicated by a number of factors including the need to track users seamlessly in both indoor and outdoor environments, the need for robustness to slight deviations in the path and the user's speed taken along a route. We investigate the combined use of GPS, wireless signal strength readings (WLAN) and image-matching to provide reliable user route matching.

Whilst GPS has become synonymous with user localisation, its robustness is still questionable. Indoors, GPS signals are weak or non-existent. Outdoors, GPS signals can be affected by obstacles, multipath propagation and tall buildings causing serious errors in localisation [3], [4]; in the case of WLAN, localisation performance indoors is generally good. However, variations in the environment, such as temporary changes to building layout or weather conditions, can affect received signal strength (RSS) [5], [6]. Using image-matching to determine location is an alternative technique to radio-frequency based approaches. Occlusion and changes in lighting are the main problems for this approach [7]. By combining the strengths of these three complementary approaches, we hope to achieve high accuracy and robustness to the problems that affect individual modalities.

Our paper is structured as follows: Section II describes our experimental setup, including data capture, trip matching, classification and data fusion strategies. Section III presents results for each modality individually and verifies that fusion of the data outperforms any one modality. We give a summary and suggestions for future work in section IV.

II. Experimental Setup

In this section, we first describe the data we captured for use in our experiments. These consisted of a series of trips (walks along a *route*) taken at different times over a 6-week period. We next describe how we compare trips to each other in each modality using Multidimensional Dynamic Time Warping (MDTW), and how we classify trips into routes. Finally we detail how we fuse the data from multiple modalities.

A. Data Capture

A set of training data was collected simultaneously using a SenseCam [8], GiSTEQ GPS device, and Campaignr software [9] installed on a N95 Nokia phone (for collecting signal strengths data). Measurements were taken on 6 selected routes within and around the Dublin City University (DCU) campus, ranging from 330m to 615m in length. The devices were synchronized and the data recording was collected at regular time intervals (every 1, 15 and 30 seconds for GPS, SenseCam and Campaignr respectively). Each route was traversed many times over a period of 6 weeks, yielding 30 testing (6 routes × 5 trips) and 24 (6 routes × 4 trips) training sets of data overall. Signal strength information is considered to be 3-dimensional as the same 3 MAC addresses were discernible along each trip. GPS data is deemed to be 2-dimensional (consisting of longitude and latitude coordinates). On average, a trip consisted of approximately 30 images along the route.

B. Trip Matching

In order to find a similarity measure for data collected during different trips, the Multidimensional Dynamic Time Warping Algorithm [10], [11], [12] was employed. The classic DTW algorithm uses a local distance measure to determine the similarity between two sequences. These sequences may be discrete signals (time-series) or, more generally, feature sequences sampled at equidistant points in time [13]. In order to compare two different features from feature space F, a local distance measure is defined: $c : F \times F \rightarrow \Re \geq 0$. To measure the similarity between two sequences of data, the first C of length I and the second T of length J, an $I \times J$ distance table D is constructed, where each element of D, $d(i,j)$, represents the local distance between C_i, the i^{th} element of C and T_j, the j^{th} element of T. Warping paths W are then calculated from the distance table, each of which consists of a set of distance table elements that define a mapping and alignment between C and T:

$$W = \left\{ (i_w(q), j_w(q)) \,\middle|\, \begin{array}{l} q = 1, ..., Q, \\ \max(I,J) \leq Q \leq I + J - 1 \end{array} \right| \right\} \tag{1}$$

with $i_w(q) \in \{1, ..., I\}$ and $j_w(q) \in \{1, ..., J\}$. Given $(i_w(q), j_w(q))$ and $(i_w(q-1), j_w(q-1))$, the warping path is restricted by the following conditions [10]: continuity ($i_w(q) - i_w(q-1) \leq 1$ and $j_w(q) - j_w(q-1) \leq 1$), the endpoint ($i_w(1) = j_w(1) = 1$ and $i_w(Q) = I$ and $j_w(Q) = J$) and the monotonicity ($i_w(q-1) \leq i_w(q)$ and $j_w(q-1) \leq j_w(q)$). The similarity between the data sequences can be gauged by identifying the optimal warping path which minimises the overall distance. This minimised distance is given by

$$DTW(C,T) = \min_w \left(\sum_{q=1}^{Q} d(i_w(q), j_w(q)) \right) \tag{2}$$

DTW(C,T) is then normalised with the length of the optimal warping path (compensation due fact that warping paths may depend on the paths' lengths) [10]. Since the data in this paper was multidimensional, we switch to multidimensional sequences $C(I \times V)$ and $T(J \times V)$ (V is number of variables) and we use d_E, the extended *Euclidean distance* [10] as the local distance measure for two vectors of length V:

$$d_E(C_i^V, T_j^V) = \sqrt{\sum_{v=1}^{V} W(v)(C_{i,v} - T_{j,v})^2} \tag{3}$$

where W is a positive definite weight vector (gives more weight to certain variables but since the variables in our data are of equal importance, W is set to 1). The DTW distance between two multidimensional sequences $C(I \times V)$ and $T(J \times V)$ can be calculated recursively as [11]:

$$\begin{array}{l} DTW(C(I \times V), T(J \times V)) = d_E(C_I^V, T_J^V) + \\ min\{DTW(C((I-1) \times V), T(J \times V)), \\ DTW(C((I-1) \times V), T((J-1) \times V)), \\ DTW(C(I \times V), T((J-1) \times V))\} \end{array} \tag{4}$$

For GPS and WLAN data $DTW(C,T)$ can be thus computed for each pair of trips, normalised and then used to populate a distance matrix. In the case of image data the elements of the distance table corresponded to the number of features [14], matched using Speeded Up Robust Features algorithm (SURF)[7], between every two images (one from each set). SURF is a scale and rotation invariant descriptor and detector. The detection process is based on the Hessian matrix. SURF descriptors are based on sums of $2D$ Haar wavelet responses, calculated in a 4×4 subregion around each interest point. The standard SURF descriptor has a dimension of 64 and the extended version (e-SURF) of 128. SURF features have been extensively compared against radial lines and SIFT features, showing SURF the best compromise between efficiency and accuracy in all the process, giving the most accurate results and allowing faster computations. A match between interest points was determined by using the distance ratio test, as described in [7]. Since this measure is asymmetrical, we also compute the matches in the reverse direction (from the target to the query) and we count the matches that occur in both directions (bi-directional matches). Such matches were found to be very stable and strong indicators of a good match [15]. A greater weight was put on these matches since the greater level of confidence ascribed to them (the measure is $d(i,j) = 10B + U_{ij} + U_{ji}$, where B stands for the number of bidirectional matches, U_{ij} the number of unidirectional matches from the i^{th} to the j^{th} image and vice versa). The distance table is then multiplied by -1 so that the optimal path corresponds to the path with most matches [13]. Resulting value was also normalised with the length of the optimal warping path. To transform the number of SURF matches between two trips into the distance matrix, a mapping process needed to be defined. It should be monotonically decreasing and produce non-negative values.

Sources	w_1	w_2	w_3	Acc (%)
SS	-	-	-	56.66
IMG	-	-	-	66.66
GPS	-	-	-	80.00
GPS,IMG	0.9720	0.0280	-	80.00
IMG,SS	-	0.8050	0.1950	70.00
GPS,SS	0.9840	-	0.0160	80.00
GPS,IMG,SS	0.8310	0.0090	0.1600	**83.33**

Table I
TRIP CLASSIFICATION PERFORMANCE: THE LEARNED WEIGHTS FOR EACH OF THE THREE SOURCES, AND THE ACCURACIES (ACC) OF THE CLASSIFIERS USING DIFFERENT SOURCES.

While there are many such functions, the reciprocal function was used for its simplicity [13].

C. Trip Classification

In order to classify a new trip into one of the known routes, we used a k-NN (nearest neighbor) classifier. This simple classifier can account for the large variability of the localisation sources, as well as being able to easily accommodate new trip examples for online training. In our case k was equal to 1,2,3 and 4.

D. Multimodal Fusion

To fuse the localisation data from our three sources, we computed a weighted linear combination of the distance matrices of the sources. These matrices were firstly normalized (using min-max normalisation), suitably weighted and then added. Using a training set of 24 trips (three normalized 24×24 matrices for GPS, image and signal strength data) we identified a set of optimal weights for each combination of sources using an exhaustive grid-search [16]. Grid-search consisted of using all possible combination of values w_1, w_2 and w_3 from the [0,1] domain with the step of 0.001. The weights were selected such that $\sum w_i = 1$ and the classification accuracy on the training set was maximised (for every k-NN this process was repeated). Table I shows the learned weights for which the highest accuracies were obtained. We evaluated the classification performance on 30 separate testing trips (three normalized 30×30 matrices for GPS, image and signal strength data) and gave the accuracies for the weights in the table I as well.

Table I clearly illustrates that GPS is the strongest individual modality. This is further emphasised by the high weight that is placed on this data source by the fusion process. As all our trips were outdoors, this was to be expected.

Figure 1 gives a visual representation of the similarities between trips in different modalities. We used the distance-matrix visualisation algorithm given in [16] to display in $2D$ a representation of the multidimensional trips and their similarities. This algorithm takes the difference between every two trips (distance matrix elements) and makes a chart (trips on the chart are presented as circles) in which the distances between them on the chart match those differences. This iterative algorithm first calculates the target distances between all the trips. Next all the trips were placed randomly on the two-dimensional chart. For every pair of trips the target distance is compared to the current distance and an error term is calculated. Then every trip is moved a small amount closer or further in proportoin to the error between the two trips. This procedure is repeated many times until the total amount of error cannot be reduced by moving the trips any more.

The trips noted as 1-5 belong to the first route (the green line in the figure 2), the trips 6-10 to the second (the yellow line), the trips 11-15 to the third (the pink line), the trips 16-20 to the fourth (the red line), the trips 21-25 to the fifth (the blue line) and the trips 26-30 to the sixth route (the purple line). It is noted that similar trips along the same route tend to cluster together and can be identified as such (these are explicitly circled in the figures). The clusters which contain less than three elements are discarded. The fusion algorithm was able to successfully identify each of the 6 routes, something not managed by any other combination of these modalities. Examining the GPS results in figure 1(a), it can be seen that the fourth route (trips $16 - 20$) and the sixth route (trips $26 - 30$) do not cluster well (red and purple routes shown in fig 2). They traversed environments where the GPS signal was degraded and attenuated (these areas are shown with the green circles), due to tall buildings (the sixth route) and to part of the path going into a tunnel (the fourth route), both of which are known to affect GPS signal quality [3], [4].

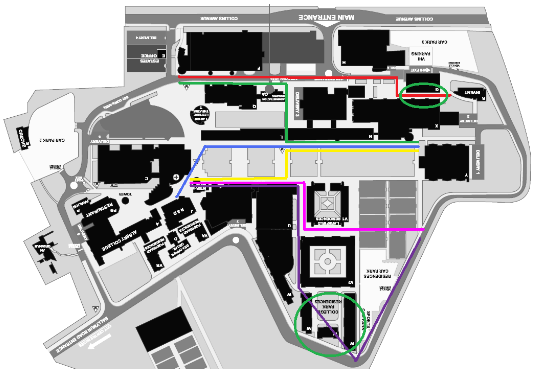
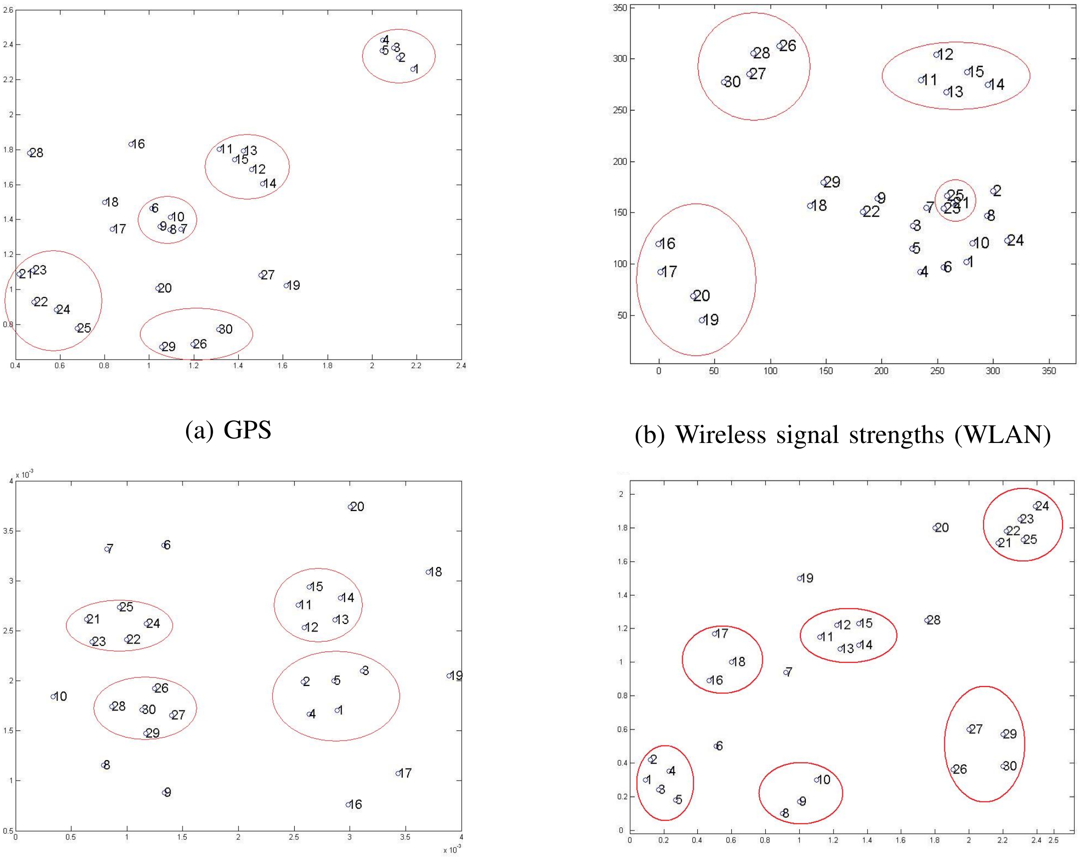


Figure 2. DCU map with the routes overlaid

III. RESULTS

The reason why the second and the fourth route failed as the image data on these routes were collected randomly during a variety of different conditions (rain/sun, morning/evening/nighttime, obstacles). While SURF features are somewhat robust to changes in lighting [7], large changes cause problems, as shown in figure 3. Consequently the results show that 4 of the 6 routes could be properly clustered



(a) GPS

(b) Wireless signal strengths (WLAN)

(c) Image-based matching

(d) Fusion of all three sources

Figure 1. Visualisation of trip similarity using different localisation sources: We project the distances between trips into 2-dimensions for visualisation. Circles are drawn to show trips from the same route that tightly cluster together. Route 1 contains trips {1,..,5}, Route 2 contains trips {6,..,10}, etc.

using image data alone. For the other routes image-matching performed quite well, considering the low sampling rate it used (1/15Hz). The format above the image describes matching process and is given as $[U_{12}$:$U_{21}]$:$[B]$, where U_{12} is represented with red lines which connect matches on the two images, U_{21} with blue lines and B with the green lines.

On its own, WLAN signal strength readings perform worst for trip classification, which was expected due to the many environmental factors that can influence signal strengths outdoors and the fact that only 3 MACs were discernible. Figure 4 shows signal readings for 3 trips from the same route, illustrating the degree of variability inherent in the readings.

IV. Conclusions

In this work, we presented preliminary results of combining three complementary sources of data for classifying trips from localisation data. By fusing GPS, wireless signal strength readings and image-based matching, we achieve better performance than any individual/combined modality. Future work will investigate other fusion methods, such as adaptive confidence-based weighting, more sophisticated classifiers (such as SVMs) and evaluating performance on indoor routes where GPS is expected to perform poorly and

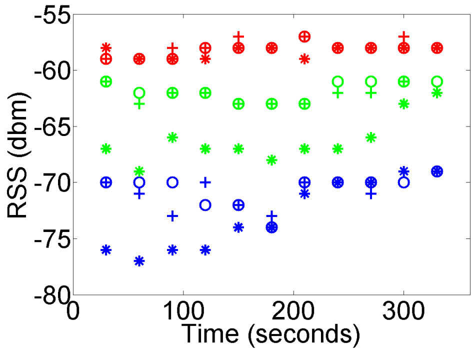
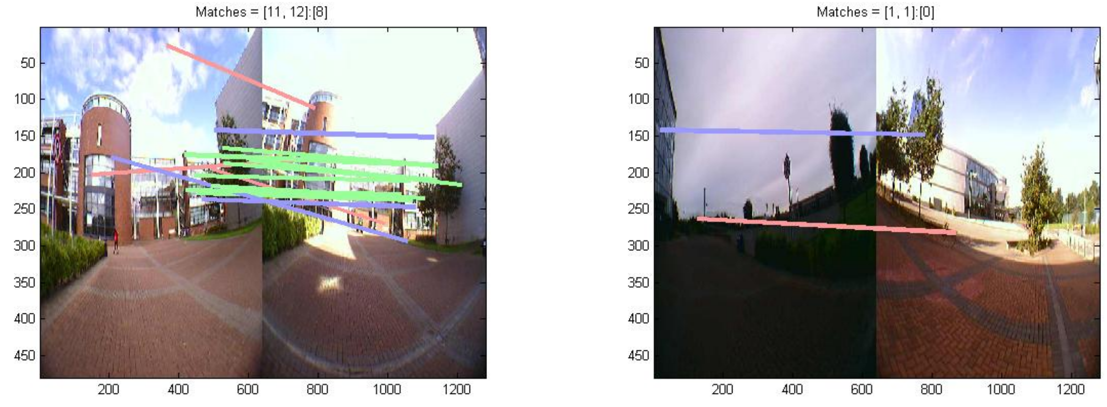


Figure 4. Signal strengths distribution example: Data from 3 MAC addresses shown in red, green and blue, corresponding to trip 16, 17 and 18 (plotted with circles, crosses and asterisks) respectively in figure 1(b). Note the discrepancy in signal strengths for trip 18 compared to others.



(a)

(b)

Figure 3. Image matching examples for trips taken in different light conditions: (a) in similar lighting conditions, many matches are found, (b) matching is more difficult due to lighting changes.

WLAN to improve.

Acknowledgment

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

References

[1] H. Aghajan, J. C. Augusto, P. Mccullagh, and J. A. Walkden, "Distributed visionbased accident management for assisted living," in *ICOST 2007*.

[2] A. Doherty, C. Ó Conaire, M. Blighe, A. F. Smeaton, and N. O'Connor., "Combining image descriptors to effectively retrieve events from visual lifelogs," in *MIR 2008 - ACM International Conference on Multimedia Information Retrieval 2008*, 2008.

[3] R. Klukas, G. Lachapelle, C. Ma, and G. Jee, "Gps signal fading model for urban centres," *IEE Proc.-Microw. Antennas Propag*, August 2003.

[4] G. MacGougan, G. Lachapelle, R. Klukas, and K. Siu, "Degraded gps signal measurements with a stand-alone high sensitivity receiver," in *Proceedings of National Technical Meeting*. The Institute of Navigation, January 2002.

[5] K. Whitehouse, C. Karlof, and D. Culler, "A practical evaluation of radio signal strength for ranging-based localization," *ACM Mobile Computing and Communications Review (MC2R), Special Issue on Localization Technologies and Algorithms*, pp. 41–52, January 2007.

[6] M. Klepal, M. Weyn, W. Najib, I. Bylemans, S. Wibowo, W. Widyawan, and B. Hantono, "Ols: Opportunistic localization system for smart phones devices," in *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds*, August 2009.

[7] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346–359, August 2008.

[8] D. Byrne, A. R. Doherty, G. J. F. Jones, A. F. Smeaton, and K. Jarvelin, "The sensecam as a tool for task observation," *HCI 2008 - 22nd BCS HCI Group Conference (HCI 2008)*, September 2008.

[9] S. Reddy, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "A framework for data quality and feedback in participatory sensing," *SenSys '07: Proceedings of the 5th international conference on Embedded networked sensor systems*, November 2007.

[10] M. H. Ko, G. West, S. Venkatesh, and M. Kumar, "Using dynamic time warping for online temporal fusion in multisensor systems," *Information Fusion*, vol. 9, no. 3, pp. 370–388, January 2008.

[11] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh, "Indexing multidimensional time-series," *The VLDB Journal*, vol. 15, July 2006.

[12] G. ten Holt, M. Reinders, and E. Hendriks, "Multidimensional dynamic time warping for gesture recognition," *Annual Conference on the Advanced School for Computing and Imaging*, June 2007.

[13] M. Muller, *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007.

[14] T. Rath and R. Manmatha, "Word image matching using dynamic time warping," *in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. II: 521–527, 2003.

[15] C. O. Connaire, M. Blighe, and N. O'Connor., "Sensecam image localisation using hierarchical surf trees," in *MMM 2009 - 15th international Multimedia Modeling Conference*, 2009.

[16] T. Segaran, *Programming Collective Intelligence:Building Smart Web 2.0 Applications*. Cambridge, Massachusetts: O'Reilly Media, 2007.