Ambient Living Activity Recognition based on Feature-set Ranking Using Intelligent Systems

Oresti Baños, Héctor Pomares, Ignacio Rojas

***Abstract*— E-health and e-monitoring have become an increasingly important area during recent years, being the recognition of motion, postures and physical exercises one of the main topics. In this kind of problem is common to work with a huge training data set in a multidimensional space, so feature selection is absolutely necessary. Most works are based on knowledge extraction using features which permit to make decisions about the activity realized, being feature selection the most critical stage. Conventional feature selection procedures based on wrapper methods or 'branch and bound' are highly computationally expensive. In this work, we propose an alternative filter method using a feature-set ranking via a couple of two statistical criteria, which achieves remarkable accuracy rates in the classification process. We demonstrate the usefulness of our method on both laboratory and seminaturalistic activity ambient living datasets for real problems.**

I. Introduction

DAILY physical activity recognition is a very important task that is currently being applied in several fields such as e-health [11,12], robotics [7,9], sports [10] or videogames industry [4] among others. The primary difficulty consists in designing a system whose reliability is independent of who is carrying out the exercise and the particular activity execution style. Besides, the complexity is increased by distortion elements related to system monitoring and processing, along with the random character of the execution.

In the literature, most studies performed are based on supervised laboratory data. Nevertheless, the apparently good recognition results on supervised data that some works achieve cannot be extrapolated to unsupervised (semi-naturalistic) data [2,13]. In this paper we propose an automatic methodology to extract a set of the most important features to be used in activity recognition. One of the most important characteristic of the method proposed is that we do not provide a rank order for every individual feature but for every set of features, allowing for the synergical utility of several features when considered together at the same time.

The rest of the paper is organized as follows: In Section 2 a brief summary of the activity recognition process is presented. Next, the rank-based feature-set selection methodology is developed, describing the fundamentals of the proposed method and the algorithm's main steps. Finally the performance of the method for a specific example is evaluated and the accuracy results with related previous works are compared.

Manuscript sent February 11, 2010. This work was supported in part by the Spanish CICYT Project TIN2007-60587, Junta de Andalucia Projects P07-TIC-02768 and P07-TIC-02906, and the CENIT project AmIVital, of the "Centro para el Desarrollo Tecnológico Industrial" (CDTI- Spain).

O. B., H. P. and I. R. is with the Department of Computer Architecture and Computer Technology, University of Granada, Granada, 18071 SPAIN (e-mail: oresti@correo.ugr.es, {hector,irojas}@ugr.es).

II. Activity Recognition

Our experimental setup starts from a signal set corresponding to acceleration values measured by a group of sensors located in strategic different parts of the body (hip, wrist, arm, ankle, thigh), for several daily activities (see Figure 1). This philosophy of work can easily be generalized to other studies related to activity recognition from a set of features.

The initial information provided by the sensors has some artefacts and noise associated to the acquisition data process. A low pass and high pass filtering is normally used to remove these irregularities.

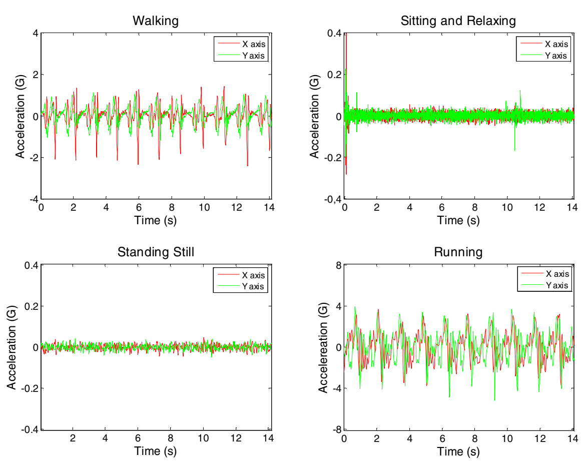

Fig. 1. Signals corresponding to 4 usual daily physical activities. It's relevant the signal pattern form similarity between "walking" and "running", and "sitting and relaxing" and "standing still" respectively.

Subsequently we generate a parameter set made up of 861 features corresponding to a combination of statistical functions such as mean, kurtosis, mode, variance, etc., and magnitudes obtained from a domain transformation of the original data such as energy spectral density, spectral coherence or wavelet decomposition, among others (the completely set is shown in table 1). These features are evaluated over the complete signal, although other alternatives based on windowing and sub-segmentation signal feature extraction could also been tested.

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

TABLE I
FEATURE SET GENERATION FUNCTIONS

Magnitudes	Statistical functions
Amplitude	4^{th} and 5^{th} central statistical moments
Autocorrelation	Energy
Cepstrum	Arithmetic/Harmonic/Geometric/ Trimmed mean
Correlation lags	Entropy
Cross correlation	Fisher asymmetry coefficient
Energy Spectral Density	Maximum / Position of
Spectral coherence	Median
Spectrum amplitude/phase	Minimum / Position of
Histogram	Mode
Historical data lags	Kurtosis
Minimum phase reconstruction	Data range
Wavelet decomposition	Standard deviation
	Total harmonic deviation
	Variance
	Zero crossing counts

Every magnitude is combined with each statistical function respectively generating the set of 861 features.

In this stage, we now must rely on a feature selection process that has the responsibility of deciding which features or magnitudes are the most important ones to decide the kind of activity the person is carrying out. In the next section, the method we have designed to accomplish this task is described.

III. Proposed Ranking Technique Based on Discrimination and Robustness

Obtaining a specific group of variables from a big initial set is not a trivial task, because the number of possible feature combinations is huge. In our experimental setup the sample space is represented by n = 861 features, so brute force techniques like 'branch and bound' ($O(2^n)$ convergence implies $2^{861} \approx 1.5 \times 10^{259}$ possible permutations) or wrapper methods are impractical. In this section, we present an alternative method based on the concepts of discrimination and robustness for a complete set of features.

Let us define the sample range of a class as the set of values included between the maximum and the minimum value (both inclusive) that a feature or variable takes for this class. Every circle showed in the example of figure 2 represents a sample corresponding to the feature value calculated over the data from a concrete subject and class. We represent for every class the sample range using a double arrow with the same class tonality.

Given a group of samples (associated to every class) we rank its discriminant capability with respect to that class through the overlapping probability between this class and the others. This is calculated computing the number of samples from the analyzed class which are inside of the sample range defined by the others. For N classes and M samples for each class (let us suppose that this number is independent of the class), we define the overlapping probability of a set of samples as follows:

$$p(k) = \frac{1}{N-1}\sum_{n \neq k}^{N} \frac{m(k,n)}{M} \qquad (1)$$

with *m(k,n)* being the number of samples from the class *k* inside the sample range of class *n*.

In order to make this more understandable, for the example given in figure 2 (N = 4, M = 8), the overlapping probability for the class "running" (identified as k = 1 for simplicity) is p(k=1) = ⅓(3/8+0/8+0/8) = 0.125, since there are 3 samples from the class "running" in the data range defined for class "walking", and 0 in the rest of classes. Accordingly, this feature permits to discriminate a priori the activity *running* from the activity *standing still* or *sitting and relaxing*, but it could be mistaken with an approximately 12.5% probability with one of all the others activities.

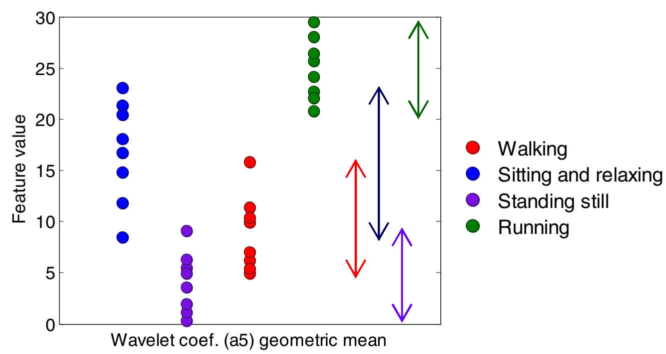

Fig. 2. Feature values extracted for the ankle accelerometer. Every circle represents the value (sample) of the feature "decomposition wavelet coefficients a5 geometric mean" for a specific subject (8 in all), identifying with different tonalities the belonging class.

We now carry out a thresholding process which allows us to define the feature analyzed as discriminative or not. This *overlapping* threshold (o_{th}) takes values from 0 (the most restrictive, for cases with no overlapping between classes) to 1 (the most relaxed, when every sample from a class is inside the others). In general for a specific feature, if the analyzed class exceeds the threshold, the feature will be considered as no discriminant for this class. Therefore, being $p_f(k)$ the overlapping probability for a feature *f* and class *k*, we have:

$$f \mapsto \begin{cases} discrim.\ class\ k & if \quad p_f(k) \leq o_{th} \\ no\ discrim.\ class\ k & if \quad p_f(k) > o_{th} \end{cases} \qquad (2)$$

In figure 3 we can see the evolution of the discriminant feature set as overlapping probability is relaxed. The number of features which permit discriminate the current activity from the rest increases as we expect, due to samples from other classes are positively accepted within the sample range. Moreover, we want to stress the importance of feature extraction method in this context. In figure 4 we show for zero overlapping threshold no discriminant features are available for most of the classes (only a few features overcome in "running" activity). This is related to the feature extraction method used to obtain the measures. In fact, features evaluated over the complete signal offer better performance (in terms of discriminability) than windowing methods. Consequently, the feature selection method allows us to have a general measure of generalization capability.

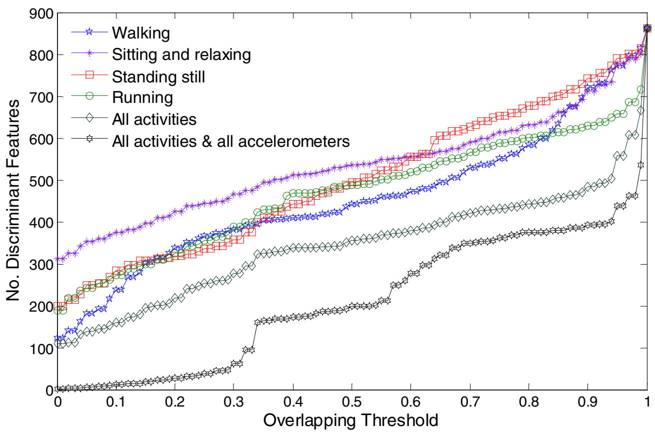

Fig. 3. Evolution of discriminant features for several overlapping probabilities (features extracted over the complete signal from data corresponding to hip accelerometer). "All activities" label is referred to common discriminant features for all activities respectively. "All activities & all accelerometers" label is for activities that reproduce their discriminability for all accelerometers.

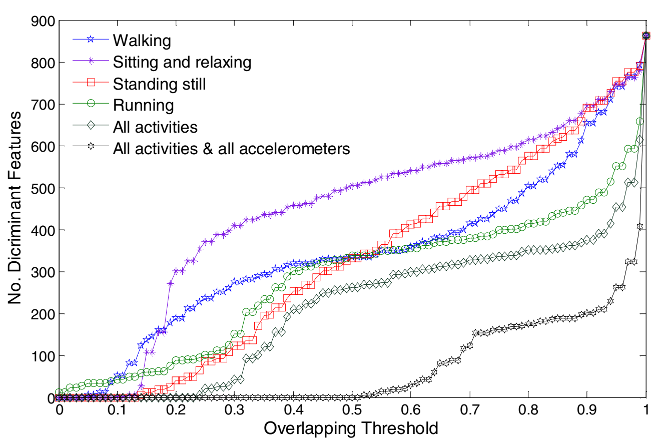

Fig. 4. Evolution of discriminant features for several overlapping probabilities (features extraction based on a windowing method; data corresponding to hip accelerometer).

Apart from the discriminant capacity of a feature or a set of features, a second characteristic is now defined which takes into account the usability of this set of features in different information contexts or sources. For instance, a specific measure taken from, let's say, the ankle accelerometer can be very discriminative to distinguish between the activities *walking* and *standing still*, but this very same measure may not be that reliable when taken from the thigh accelerometer. There may be some measures with the same discriminant capability between those activities which are not so dependent of the exact location of the sensor or, at least, which are still reliable when taken from a bigger number of sensors. We will denote this measure as the robustness criterion of a set of features. In short, discriminant capacity says how useful a motion feature is in general, and robustness is how this depends on where the sensor is.

Combining both criteria we obtain a quality ranking procedure capable of grouping features in different stages. For the sake of simplicity, let us suppose a recognition system with 4 classes and 5 sources; features will be classified in groups defining a ranking (see table 2). For instance, features that discriminate 4 classes in every source will be added to group #1 (the best). Group #14 will be completed with features that classify 2 classes (the same) in 3 sources at least. This example is extensible to any classes and sources.

TABLE II
FEATURES RANKING

Discriminant capacity	Robustness	Quality group
4	5	#1
4	4	#2
4	3	#3
4	2	#4
4	1	#5
3	5	#6
3	4	#7
3	3	#8
3	2	#9
3	1	#10
2	5	#11
2	4	#12
2	3	#13
2	2	#14
2	1	#15
1	5	#16
1	4	#17
1	3	#18
1	2	#19
1	1	#20
0	-	#21

Example for 4 classes and 5 sources of quality feature set (ranking) based on discriminant (number of activities discriminated, first column) and robustness (number of motion sensor where the feature is discriminant, second column) criterions.

IV. RESULTS

To evaluate the effectiveness of the ranking method developed, we use a signal database[1] corresponding to the

[1] Database facilitated in [2] by Prof. Stephen Intille (Massachussets Institute of Technology).

data monitored by 5 biaxial accelerometers (hip, wrist, arm, ankle and thigh) for 4 activities (introduced in figure 1) in laboratory (supervised) and semi-naturalistic (unsupervised) environments.

Most remarkable features (set #1 and #2 primarily) for supervised and unsupervised data are geometric mean for amplitude signal, autocorrelation and some wavelets coefficients obtained through a 5-level Daubechies decomposition. Two classification strategies have been tested: support vector machines (SVM) and decision trees (DT). First, we achieve an accuracy rate about 96% (96.37% ± 3.58%) for laboratory data and 78% (78.20% ± 2.14%) for semi-naturalistic data using SVM (one against all process). Results were improved using a C4.5 DT implementation, which permits to achieve an accuracy rate close to 99% (98.92% ± 1.08%) for laboratory data and 95% (95.05% ± 1.20%) for semi-naturalistic data. Both cases we used a cross validation method for training and testing. The results are quite good working on laboratory data, but DT performance is particularly better in unsupervised situations, due to the relative parallelism between probability model used in feature selection defined and the entropy-based model used in DT. In any case, these results improves the mean scores for medium ranked features (sets #3 to #8) and low ranked features (sets #9 to end) more than 30% and 70% respectively.

Although a strict comparison with other studies cannot be made since the data and the number of classes may differ, in [9] a 83-90% classification accuracy was reached for laboratory conditions, 92.85%-95.91% in [8] (also for lab conditions), 89% in [2] for supervised and unsupervised, or 93% and 89% on recent works ([3] and [5] respectively) for semi-naturalistic data.

V. Conclusion

In this work we have very briefly shown a direct application of ranking selection methods used on daily physical activity automatic recognition. An efficient classification method requires a productive and limited feature set, being necessary a selection process since the initial set is quite huge. We have defined a feature selector based on statistical discrimination and robustness criteria, focused on low computational time and resources, defining a real alternative to other selection processes.

For future work, we aim to make a time-based comparison to traditional and recent features selectors [6, 14, 15].

Acknowledgment

This work has been supported in part by the Spanish CICYT Project TIN2007-60587, Junta de Andalucia Projects P07-TIC-02768 and P07-TIC-02906, and the CENIT project AmIVital, of the "Centro para el Desarrollo Tecnológico Industrial" (CDTI- Spain). We want to express our gratitude to Prof. Stephen S. Intille, Technology Director of the House_n Consortium in the MIT Department of Architecture for the experimental data provided.

References

[1] J. Baek, G. Lee, W. Park, and B. J. Yun, "Accelerometer Signal Processing for User Activity Detection", in *Techniques of Computational Intelligence for Affective Computing (Knowledge-Based Intelligent Information and Engineering Systems)*, LNCS, vol. 3215, pp. 1611–3349. Springer, Heidelberg, 2004.

[2] L. Bao, and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data", in *Activity Recognition (Pervasive Computing)*. LNCS, vol. 3001, pp. 1–17, Springer, Heidelberg, 2004.

[3] A. G. Bonomi, A. H. C. Goris, B. Yin, and K. R. Westerterp, "Detection of Type, Duration, and Intensity of Physical Activity Using an Accelerometer". *Medicine & Science in Sports & Exercise*, vol. 41, 2009, pp. 1770–1777.

[4] N. Crampton, K. Fox, H. Johnston, and A. Whitehead, "Dance, Dance Evolution: Accelerometer Sensor Networks as Input to Video Games", *IEEE International Workshop on Haptic, Audio and Visual Environments and Games*, 2007, pp. 107–112.

[5] M. Ermes, J. Pärkka, J. Mantyjarvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions". *IEEE Trans. Inf. Technol. Biomed*. vol. 12, 2008, pp. 20–26.

[6] R. Kohavi and D. Sommereld, "Feature Subset Selection Using the Wrapper Method: Overtting and Dynamic Search Space Topology", presented at the *First International Conference on Knowledge Discovery and Data Mining*, 1995.

[7] H. Koskimaki, V. Huikari, P. Siirtola, P. Laurinen, and J. Roning, "Activity recognition using a wrist-worn inertial measurement unit: A case study for industrial assembly lines", in *17th Mediterranean Conference on Control and Automation*, 2009, pp. 401–405.

[8] S. W. Lee and K. Mase, "Activity and location recognition using wearable sensors", in *IEEE Pervasive Computing,* 2002, vol. 1, pp. 24–32.

[9] J. Mantyjarvi, J. Himberg, and T. Seppanen, "Recognizing human motion with multiple acceleration sensors", in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2001, pp. 747–752.

[10] G. McCormack, B. Giles-Corti, A. Lange, T. Smith, K. Martin and T. J. Pikora, "An update of recent evidence of the relationship between objective and self-report measures of the physical environment and physical activity behaviours", *Journal of Science and Medicine in Sport*, 2004, pp. 81–92.

[11] M. McCue1, J. Hodgins, and A. Bargteil, "Telerehabilitation in Employment/Community Supports Using Videobased", in *Activity Recognition. RERC on Telereahabilitation*, 2008.

[12] E. Munguia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King, and, R. Friedman, "Real-Time Recognition of Physical Activities and Their Intensities Using Wireless Accelerometers and a Heart Rate Monitor", in *Proceedings of the 11th IEEE International Symposium on Wearable Computers*, 2007, pp. 1–4.

[13] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity Recognition from Accelerometer Data", in *Proceedings of the 17th conference on Innovative applications of artificial intelligence*, 2005, pp. 1541–1546.

[14] L. Song, A. Smola, A. Gretton, K. M. Borgwardt, and J. Bedo, "Supervised feature selection via dependence estimation", in *Proceedings of the 24th international conference on Machine learning*, 2007, pp. 823–830.

[15] Z. Xu, R. Jin, J. Ye, M. R. Lyu, and I. King, "Non-monotonic feature selection", in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 1145–1152.