35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Improving Activity Recognition using Temporal Coherence

Abbas Ataya[1], Pierre Jallon[1], Pascal Bianchi[2] and Maeva Doron[1]

***Abstract*— Assessment of daily physical activity using data from wearable sensors has recently become a prominent research area in the biomedical engineering field and a substantial application for pattern recognition. In this paper, we present an accelerometer-based activity recognition scheme on the basis of a hierarchical structured classifier. A first step consists of distinguishing static activities from dynamic ones in order to extract relevant features for each activity type. Next, a separate classifier is applied to detect more specific activities of the same type. On top of our activity recognition system, we introduce a novel approach to take into account the temporal coherence of activities. Inter-activity transition information is modeled by a directed graph Markov chain. Confidence measures in activity classes are then evaluated from conventional classifier's outputs and coupled with the graph to reinforce activity estimation. Accurate results and significant improvement of activity detection are obtained when applying our system for the recognition of 9 activities for 48 subjects.**

I. INTRODUCTION

Over recent years, there has been a significant research effort focusing on the assessment and classification of physical activities (PA) [1]. This has been driven by its importance for numerous health-related applications. The need for such applications has grown in response to studies that identified strong links between levels of PA and diseases as cardiovascular disease, hypertension, diabetes and certain cancer types and also the important role of PA in reducing obesity risk [2]. Although numerous studies employed computer vision-based systems for activity recognition, the use of wearable sensors as accelerometers showed promising advantages.

In previous research, machine learning approaches have been shown to be efficacious for recognizing a variety of activities from wearable sensors [1], [3]. The approaches principally utilize a two stage process [4]. A first step consists of evaluating representative features of the acceleration data over sliding windows. A classifier is then applied on the extracted features to associate each data window with an activity. Numerous classifiers have been used for this purpose. They included generative classifiers such as Gaussian mixture models (GMM) [5], a range of discriminative classifiers such as k-nearest neighbors (k-NN), support vector machines (SVM), and decision trees (DT) ([1], [4]), and ensembles of classifiers as AdaBoost (AdaB) and random forests (RF) ([6], [7]).

Discriminative classifiers distinguish between activities by directly learning decision boundaries in the feature space [8]. Whereas generative ones first create models to describe how features for each activity have been generated before inferring the boundaries to discriminate among activities. The authors in [8] have shown that discriminative classifiers usually outperform generative ones for classification tasks which justifies previous results obtained for activity classification [1], [6].

Nevertheless, the majority of discriminative classifiers are instance based that assume implicitly that the data measured from sensors comes from a time-independent sequence of activities. In real life, this assumption is not valid and activities performed in the same time range are sequentially correlated. Few approaches searched to consider these temporal dependencies of activities.

In [3], Mathie et al. performed classification of a predefined sequence of activities using decision trees and aposteriorily applied to the classification results a set of predefined rules to refine the estimated activity sequences. Taking advantage of the capacity of some generative classifiers as hidden Markov models (HMMs) to incorporate temporal information on how features and activities transition over time, recent work [6] applied a hybrid approach combining discriminative and generative classifiers. The authors reported 4% of increased accuracy compared to results of the discriminative classifier alone (AdaB with Decision Stumps). A big limitation of this approach is that its computational complexity can be very high. An HMM per activity is trained using posterior probabilities. Classification is then done over windows choosing the activity for which the HMM maximizes the likelihood over the windows. Computation complexity can also be increased if the number of states in HMMs and the window's length are to be optimized.

In this paper a novel less computationally expensive approach called graph method is presented to consider temporal coherence of activities. In this approach, the classifier is not used to make decisions but only to give confidence measures in the belonging of data to activity classes. These measures are then combined with the temporal dependencies of activities modeled by a directed graph Markov chain to refine recognition results. The instance based classification that we use is based on a hierarchical structure. The whole classification system will be evaluated using several classifiers and compared with state of the art methods.

[1]A. Ataya, P. Jallon, and M. Doron are with CEA LETI, Minatec Campus, DTBS, 17 rue des Martyrs, 38054 Grenoble, France `firstname.lastname at cea.fr`

[2]P. Bianchi is with the Department of signal and image processing, TÃľlÃľcom ParisTech, 37 rue Dareau, F-75014 Paris `pascal.bianchi at telecom-paristech.fr`

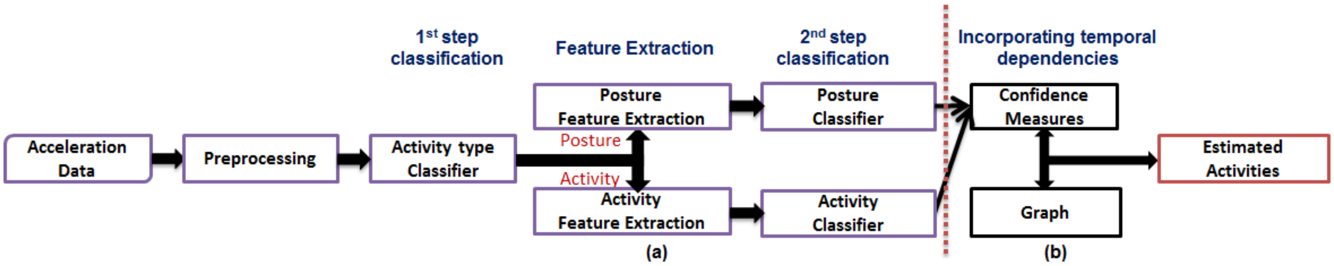

Fig. 1. Block diagram describing the classification system presented in this paper. (a) Instance based classification. (b) Graph method incorporating temporal dependencies to classification results.

II. Framework for activity classification

Our classification system aims at classifying 9 activities (lying down, slouching, standing, stamping, cycling, running, slow/fast walking, and using stairs) using data from hip-worn accelerometer. The classification system is characterized by 2 main parts (Fig. 1) : (a) Instance based classification and (b) Incorporation of inter-activity temporal dependencies in the recognition. Instance based classification part has the hierarchical configuration of a one-node binary tree where leaves themselves are classifiers. At the first level, a binary classification of activities is applied to differentiate between two activity types where an activity is classified as static (posture) or dynamic. The next stage consists of applying a separate classifier for each type to perform more detailed activity classification. The split between static and dynamic activities permits the extraction of more relevant set of features for each activity type. It also leads to less classification confusion since reduced number of activities is presented to each of the second stage classifiers.

A. *Preprocessing techniques*

Before classifying the acceleration data, some preprocessing methods were applied. The acceleration signals were first median filtered to remove noisy spikes. The resulting signals were then high and low pass filtered at a cutoff frequency of 0.1 Hz. Applying the high pass filter attempts to remove the static component of the acceleration signal which basically captures postural information concerning the body inclination with respect to the ground. Low pass filtering has an opposite effect which is eliminating the information related to dynamic motion to conserve only information about static activities.

B. *Feature computation and classification of data*

Time and frequency domain features were evaluated over 2 seconds sequential sliding windows with an overlap of 1 second. This kind of features has already shown advantage over time-frequency features such as wavelets for the recognition of activities [9]. The use of 2 seconds windows was inspired from previous research that found them sufficient for recognizing a wide range of activities [10]. The 50% overlap between consecutive windows has also been proven efficient for activity recognition [1].

The distinction between posture and activity was done using a normalized Signal Magnitude Area (SMA) value over windows of the high-pass filtered acceleration data. The SMA value is the sum of the moduli of the 3 acceleration values. If this value is higher than a certain threshold then the subject is considered involved in dynamic activity else a static phase is detected. The threshold is estimated from the annotated database using 1-fold cross validation and chosen to be the value maximizing classification accuracy between dynamic and static activities. For static activities, 14 features were evaluated from the low-pass filtered acceleration data. These features included the average mean values, temporal energy, average of L-1 norm of acceleration vector, sensor's tilt angle from ground, area under the curve and mean distances between axis. As for dynamic activities, 18 features were extracted from high-pass filtered acceleration data. These features comprised median frequencies of the 3 axis, entropy, mean-cross rate, peak-to-peak distances, mutual correlation between axis, and spectral energy. Most of these features have already demonstrated success in activity recognition [1], [9].

III. Graph method for temporal coherence

Physical activities naturally present coherence in time. For instance, it seems unlikely that a person lies down straight after running, but is more likely to be engaged in a walking phase or standing. Similarly, people do not abruptly alternate between activities but rather tend to perform the same activity over long periods of time. Despite these correlations, classifications are usually done over sliding windows that are supposed independent of each other which leads to many spurious classifications. The graph method attempts to remedy these time independence assumptions by coupling, in an overall decision technique, information about the inter-activity transitional behavior with confidence measures in the belonging of data windows to each of the activities. The activities are modeled using a directed graph described by a Markov chain and are therefore completely characterized by the initial probabilities $p(A_1 = i)$ $(i = 1, 2, \ldots, N)$, and the transition probability distribution $p(A_m = j|A_{m-1} = i)$. Here, N is the total number of activities and A_m is the performed activity at instant m indexed from 1 to N.

The distribution of the transition probabilities between activities can be directly estimated from training data.

$p\left(A_m = j | A_{m-1} = i\right)$ is thus taken to be the number of transitions from activity i to activity j divided by the total number of transitions from activity i. This leads to high probability values for most likely transitions and lower probabilities for those that are less likely to happen. In real life, the activity by which people start is unpredictable. For this reason we choose an equiprobable distribution for initial probabilities. Given a sequence of feature vectors to classify $\{\mathcal{F}_m\}_{1\leq m\leq M}$, incorporating temporal dependencies into classifier's decisions is done by estimating the performed activities as follows :

$$\hat{A}_{1:M} = \underset{A_{1:M}}{\arg\max}\, p(A_1) \prod_{m=2}^{M} p(A_m|A_{m-1}) \prod_{m=1}^{M} \phi(\mathcal{F}_m|A_m)$$

This optimization problem is solved using Viterbi algorithm. $\phi(\mathcal{F}_m|A_m)$ represents the confidence measures for the belonging of $\mathcal{F}_m$ to the different activities. In other words, $\phi(\mathcal{F}_m|A_m = i)$ stands for the level of confidence that the classifier has in associating $\mathcal{F}_m$ with activity i. In the above equation we can see that the temporal dependencies modeled in the graph regularize the decisions taken by the classifier. Estimated activities represent thus the most likely sequence of activities taking into account, at a time, the temporal coherence of activities and the classifier's confidence in its decisions. We evaluate the confidence measures from the soft outputs of the classifier using appropriate functional mappings. Due to lack of space, the evaluation of these measures for the applied classifiers will be briefly explained in Section V.

IV. Data Collection

Acceleration data was collected using a Motion POD™ (MOVEA) with a built-in triaxial accelerometer sensor. The monitor sensor was placed at the hip level and used to collect, at a rate of 100 Hz, acceleration data of 48 subjects free of any chronic diseases and including 26 men and 22 women of age range varying between 19 and 55 (mean ± SD = 36 ± 11 y). About 55 minutes of acceleration data were collected for each subject and during which subjects performed activities of various intensities as naturally as they do in their everyday life. During data collection, a supervising medical team took charge of labeling the principal performed activities by each subject. Nine activities that mainly constitute the most practiced activities in everyday life were searched to be recognized. Among these activities, 3 are postures or static activities (lying down, slouching/sitting, and standing) and the other 6 activities correspond to motion or dynamic activities (stamping, cycling, running, slow walking, fast walking, and using stairs).

V. Recognition results

Instance based classification and classifications taking into account the temporal coherence of activities were applied using 7 different classifiers (DT, RF, SVM, k-NN, GMM, AdaBoost with decision stumps, and Adaboost with decision trees). To evaluate the 9-activity classification accuracy for each of the classifiers, a leave-one-subject-out validation technique was applied. The method applying HMMs to take into account the temporal continuity was extended and adapted to each of the employed classifiers for comparison. One HMM is learned for each activity using probabilistic outputs of instance based classifiers. The same activity is then assumed to be practiced over fixed length windows or sub-sequences of the final estimated activity sequence. Therefore, the activity for which the HMM maximizes the likelihood over these windows of pre-specified length is estimated. Using Gaussian mixture models with 2 components as observation probabilities for HMM hidden states, we test this method for different number of states per HMM (2, 3, and 4 states) and also for windows of variable lengths (5, 10, 15, and 20 s). For each classifier, the combination giving best results is considered.

For the AdaBoost classifiers, the classification margins which correspond to the weighted ratio of votes given by the weak classifiers to each of the activities are used to estimate the confidence measures in these activities. This is done by applying parametrized sigmoid functions to these margins. The parameters of the sigmoid functions are learned from the training data. Similar sigmoid functions were used to map the SVM classifier's output to confidence measures [11]. Concerning random forest classifier, the fractions of votes given by the learned trees to each of the activities are used as the confidence measures (50 trees). For the k-NN classifier, the proportion of nearest neighbors belonging to each of the considered activities are evaluated. These proportions are further weighted with the inverse of the distance from nearest neighbors to the considered data point to obtain the confidence measures. For decision trees, in order to estimate the confidence in the belonging of a data point to the different activities, the classifying node is considered. Confidence measures are obtained by evaluating the fraction of training data arriving to this classifying node and belonging to each of the activities. Finally, applying Bayes rule for the results of the GMM classifier permits us to obtain for this classifier confidence measures in each of the activities.

Table I reports the classification results of the different applied methods for the classification of activities. From

TABLE I

Summary of classifiers' results (Mean Accuracy ± SD)

		Applied Method		
		Instance based	with HMM	with graph
Classifier	DT	77.8 ± 7.3	83.9 ± 8	86.7 ± 6.6
Classifier	RF	81 ± 7.1	86.5 ± 6.5	**89 ± 5.8**
Classifier	SVM	76 ± 7.1	83 ± 7.4	87.1 ± 7.5
Classifier	*k-NN*	76.7 ± 6.9	83.8 ± 7.3	86 ± 7.2
Classifier	GMM	73.9 ± 8.7	76.6 ± 9.3	80.2 ± 9
Classifier	AdaB(DS)	77.4 ± 8.2	81.1 ± 7	83.7 ± 6.8
Classifier	AdaB(DT)	74.1 ± 7.9	81.6 ± 9.6	84.5 ± 10

TABLE II

Confusion matrix : Random Forest classifier with graph using leave-one-subject-out validation for 48 subjects.

	Precision = 89 % Recall = 88.4 %	Detected Activity								
		Lying	Slouching	Standing	Stamping	Cycling	Running	Slow Walking	Fast Walking	Stairs
Labeled Activity	Lying	**92.4**	3.3	0	4.2	0.1	0	0	0	0
	Slouching	3.5	**88.5**	4.5	3.3	0.2	0	0	0	0
	Standing	0.2	2.6	**93.2**	3.8	0	0	0	0	0
	Stamping	0	0	1	**80.1**	7.9	0	9.8	0.4	0.8
	Cycling	0	0	0.3	0.9	**98.5**	0	0.3	0	0
	Running	0	0	0	0.1	0.3	**94.2**	0.1	0.1	5.2
	Slow Walking	0	0	0.2	6.2	3.1	0	**76**	12.5	2
	Fast Walking	0	0	0	0.2	1.3	0	14.2	**77.8**	6.5
	Stairs	0	0	0.8	1.2	0	0.8	2.2	2.4	**92.6**

this table, we can notice the superiority in performance of discriminative classifiers with respect to the generative one (GMM). As can also be seen from this table, taking into account the temporal dependencies between activities improves significantly the classification accuracy results. This improvement ranges between 4 to 11% depending on the employed classifier and the used method for considering these temporal dependencies. Even though simpler for implementation and much less computationally expensive, the graph method outperforms the method that uses HMMs for all tested classifiers. The better performance of the graph method is basically due to the fact that the method using HMMs fails very frequently to correctly classify windows over which the conventional classifiers gave multiple erroneous activity estimations. Overall, the best recognition results were reported for the random forest classifier combined with the graph method where an average recognition accuracy of 89% was obtained for the nine activities.

The aggregate confusion matrix showing individual recognition rates for activities is given in Table II. The slight confusions between the static and dynamic activities is due to classification errors done by the first step classifier. The majority of static activities that were confused with dynamic ones were classified by the dynamic classifier as stamping. This makes sense since stamping is much closer to static activities than any other studied dynamic activity. The most significant confusion happens to be between slow walking and fast walking. This confusion is quietly expected due to similarities in motion patterns that these 2 activities share when the sensor is placed on the hip. Moreover, these activities are subject-dependent in the sense that the execution of these activities can significantly differ from one person to another. Using stairs is confused in almost 2.5% of cases with each of slow walking and fast walking activities which is due to similarities in the acceleration data that correspond to these activities. It is important to note that more important confusion between these 2 activities was obtained with classifiers other than random forest.

VI. Conclusion

In this paper, we presented a systematic approach for recognizing physical activities from a hip-mounted acceleration data. Activity recognition was refined by combining confidence measures evaluated from the instance based classifier soft outputs with the temporal coherence of activities modeled by a graph. The instance based classifier was designed hierarchically so as to preliminarily distinguish between postures and motion. These latter were subsequently processed to separate classifiers for more specific classification of the performed activities. For the recognition of 9 physical activities, the whole classification system was applied using several classifiers over a data set containing 48 subjects. Obtained results showed the proposed classification approach to outperform other methods, confirming its effectiveness in detecting and discriminating among the different activities.

References

[1] L. Bao, S. S. Intille, Activity recognition from user-annotated acceleration data, Pervasive computing, 2004.

[2] World Health Organization Global Strategy on Diet, Physical Activity and Health. Geneva Switzerland ; WHO, 2004.

[3] M.J. Mathie, B.G. Celler, N.H. Lovell, A. C. F. Coster, Classification of basic daily movements using triaxial accelerometer. Medical and Biological Engineering Computing, 2004.

[4] S. Preece, J. Y. Goulermas, L. P. J. Kenney, D. Howard, K. Meijer, R. Crompton, Activity identification using body-mounted sensors : a review of classification techniques, Physiological Measurement, vol. 30, no. 4, pp. R1, 2009.

[5] F. R. Allen, E. Ambikairajah, N. H. Lovell, and B. G. Celler, Classification of a known sequence of motions and postures from accelerometry data using adapted gaussian mixture models. Phys. Measurement, vol. 27, pp. 935-951(17), Oct. 2006.

[6] J. Lester, T. Choudhury, N. Kern, G. Borriello, B. Hannaford, A hybrid discriminative/generative approach for modeling human activities. In proceedings of the Nineteenth Internation Joint Conference on Artificial Intelligence, 2005, pp. 766-772.

[7] C. Pierluigi, P. Oriol, R. Petia, Human activity recognition from accelerometer data using wearable device, Proceedings of the 5th Iberian conf. on Pattern recognition and image analysis, 2011.

[8] A. Y. Ng, M. I. Jordan, On discriminative vs. generative classifiers : A comparison of logistic regression and naive bayes, 2001.

[9] S. Preece, J. Goulermas, L. Kenney, D. Howard, A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. In Biomedical Engineering, IEEE Transactions, vol. 56, no. 3, pp. 871-879, march 2009.

[10] W. Wu, S. Dasgupta, E. E. Ramirez, C. Peterson, G. J. Norman, Classification accuracies of physical activities using smartphone motion sensors. J Med Internet Res, 2012.

[11] J. C. Platt, Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In Advances in large margin classifiers, pp. 61-74, 1999, MIT Press.