Hierarchical classification scheme for real-time recognition of physical activities and postural transitions using smartphone inertial sensors

Sid Ahmed Walid Talha[1], Anthony Fleury[1] *Member, IEEE* and Stéphane Lecoeuche[1]

***Abstract*— This paper introduces a novel approach for real-time classification of human activities using data from inertial sensors embedded in a smartphone. We propose a hierarchical classification scheme to recognize seven classes of activities including postural transitions. Its structure has three internal nodes composed of three Support Vector Machines $(SVMs)$ classifiers, each one is associated with a set of activities. Moreover, each $SVMs$ is fed with a feature vector from an adapted and optimal frequency band. Experimental results conducted on a challenging publicly available dataset named SBHAR show that our method is effective and outperforms various state-of-the-art approaches. We also show the suitability of our method to recognize postural transitions.**

I. Introduction

Human activity recognition (HAR) has become an important research area due to a high demand in various application domains, including healthcare monitoring. Computer vision has been the canonical way to develop HAR systems for decades, using several types of cameras as RGB [1] or more recently RGB-D cameras [2]. However, these approaches suffer from some limitations due to environmental restriction, such as illumination conditions or occlusions due to people and objects. Moreover, regarding the privacy context, cameras are more intrusive and disruptive for users.

Recent developments in wearable sensing technologies such as inertial sensors offer an interesting choice since it allows a continuous monitoring compatible with the user's privacy. However, wearable based approaches employ several sensors placed on different body locations. In addition to the cost, it can be uncomfortable to wear these sensors, especially for a long-term monitoring purpose.

More recently, sensors embedded in smartphones have gained the attention of researchers. Indeed, almost all smartphones include a complete Inertial Motion Unit (IMU) with a triaxial accelerometer, gyroscope and magnetometer. Nowadays, smartphones become more ubiquitous and widely used in our daily life, it can be a very good alternative to wearable devices. Moreover, smartphone-based approaches do not require any infrastructure to operate.

Several HAR approaches using inertial sensors have been proposed by researchers. The general architecture of a HAR system can be divided into the following modules: windowing, pre-processing, feature extraction and classification. The first main step is the windowing operation, it prepares the continuous data streaming for feature extraction process, i.e., inertial signals are divided into equal segments of size T_w, with a degree of overlap named O_v between consecutive segments. A wide range of window sizes has been used in HAR literature, from $T_w = 1s$ in [3], until $T_w = 30s$ in [4]. The size T_w is generally determined in an empirical way based on system accuracy. In [5], [6], after testing different window sizes, T is fixed to 5 seconds. The obtained segments can be pre-processed (optional) using filtering techniques in order to remove undesirable information. A Butterworth low-pass filter was employed in [7], the cutoff frequency was fixed to $20Hz$. In [8], a moving average filter was applied.

To describe each human activity, a vector of relevant features noted F is constructed. Features can be divided into three categories: time domain features (first and second order statistics, correlation coefficient, histogram, etc.), frequency domain features, extracted from Discrete Fourier Transform (DFT) using Fast Fourier Transform (FFT) algorithm, and time-frequency domain features, extracted from Discrete Wavelet Transform (DWT) or Wavelet Packet Decomposition (WPD). In order to obtain better accuracy, features are generally extracted from both frequency and time domains resulting in a high dimensional feature vector. In [7], [9] $F \in \mathbb{R}^{512}$. However, This method may require a high computational complexity and energy consumption to extract a large number of features or carry out the classification.

For the classification stage, various methods have been investigated in the state-of-the-art. Traditional classifiers such as k-Nearest Neighbors (kNN), Naive Bayes or decision trees are widely used. More recently, deep neural networks are employed [10], [11], mostly Convolutional Neural Networks.

In addition to physical activities, some researchers used wearable devices to detect transitions between postures [8], [12]. However, a few works fulfilled the classification of postural transitions using a smartphone. In [7], two different implementations are proposed: in the first one, only physical activities are learned by the classifier, transitions on the other hand, are considered as a switch of activities detected by the system. In the second approach, in addition of physical activities, transitions are learned by the classifier, they are considered as an additional class.

In this paper, a novel HAR method using inertial signals from a smartphone is proposed. It aims to carry out a real-time classification of physical activities and postural transitions. The main contribution of this work is the development of a hierarchical classification scheme to classify three groups of activities: postures, dynamic activities and postural transitions. Each internal node of the model is

*This work is supported by DRESS, Région Haut-de-France, and Institut Mines-Télécom Lille Douai

[1] IMT Lille Douai, Univ. Lille, Unité de Recherche Informatique Automatique, F-59000 Lille, France. E-mails: {firstname.lastname}@imt-lille-douai.fr

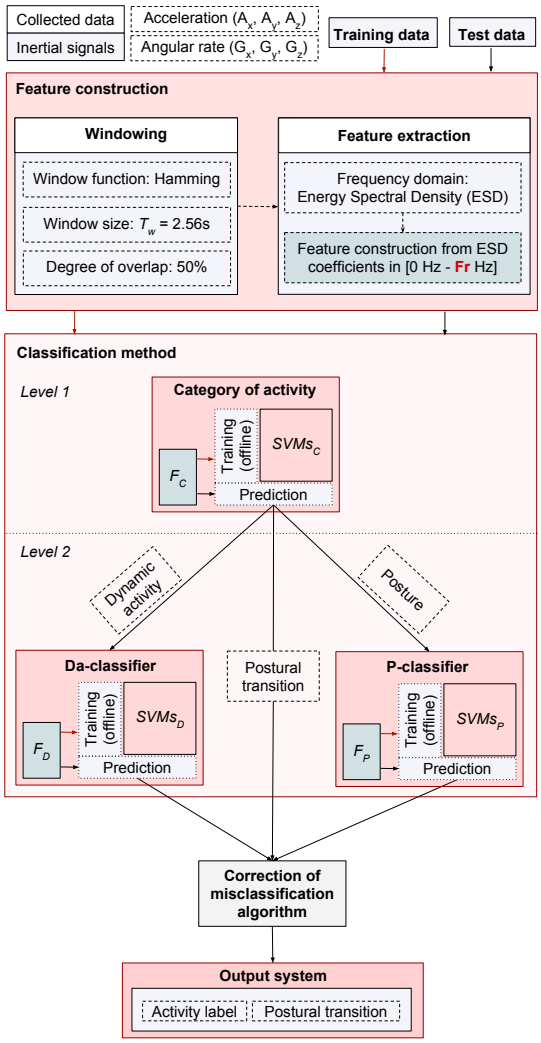

Fig. 1: Overview of the proposed system.

composed of *SVMs* classifier fed with an adapted feature set. We demonstrate experimentally that the proposed method is effective and outperforms various state-of-the-art approaches.

The remaining of the paper is organized as follows: Section II introduces the proposed methodology. Section III describes the experimental setups, then presents the results with a comparison over existing approaches. Finally, section IV concludes this work.

II. Method

The overview of our approach is shown in Figure 1. Inertial signals collected from a triaxial accelerometer and gyroscope sensors are employed in our methodology, it involves: acceleration (A_x, A_y, A_z), angular rate (G_x, G_y, G_z), and their magnitudes A_m and G_m with $A_m = \sqrt{A_x^2 + A_y^2 + A_z^2}$ and $G_m = \sqrt{G_x^2 + G_y^2 + G_z^2}$.

A. Windowing

In the state-of-the-art approaches, a rectangular window function is generally employed in windowing operation. However, in our methodology, features are extracted from frequency domain of inertial signals, hence, a rectangular window is not an advantageous choice, it presents a high leakage factor (ratio of power in the side lobes to the total window power), therefore, to minimize this effect, a Hamming window is employed instead. The size of the window is fixed at $2.56s$, and the degree of overlap O_v at 50% of the window size, these parameters are set equal to HAR approaches experimented on SBHAR dataset, so a fair comparison can be carried out.

B. Feature extraction and classification algorithms

In our approach, features are extracted from frequency domain following the algorithm 1: for each signal, after windowing operation, DFT is obtained by the FFT, then the Energy Spectral Density (ESD) containing the energy distribution at each frequency bin is calculated. Afterwards, feature vector is constructed from coefficients included in frequency band $[0Hz - \boldsymbol{Fr}Hz]$ (Fr denote the cutoff frequency). The size of the frequency band depends on the classification problem. As displayed in Figure 1, a hierarchical classification method composed of three *SVMs* classifiers is fed with three vectors of features named: F_C, F_D and F_P ($\{$C,D,P$\}$ denotes "category of activity", "dynamic activity" and "posture"). Each couple $(F_i, SVMs_i)$ with $i \in \{C, D, P\}$ is designed to solve a specific classification problem. In the first level of the scheme, activities are separated into three categories: dynamic activities, postures and postural

Algorithm 1: Feature construction algorithm

Data:
A: Segments of acceleration signals // (A_x, A_y, A_z)
A_m: Segments of magnitude of acceleration signals
G: Segments of angular rate signals // (G_x, G_y, G_z)
G_m: Segments of magnitude of angular rate signals
T_w: Window size
$Fr_1 = 2.3$: Frequency range for SVM_C // $[0\ Hz - 2.3\ Hz]$
$Fr_2 = 3.5$: Frequency range for SVM_D // $[0\ Hz - 3.5\ Hz]$
$Fr_3 = 3.5$: Frequency range for SVM_P // $[0\ Hz - 3.5\ Hz]$

Function `Feature_construction` (A, A_m, G, G_m)
  $w = 0.54 - 0.46\cos\left(\frac{2\pi n}{M-1}\right)$; // Hamming window
  $S_1 \leftarrow$Spectrum$(A.w)$; // Spectrum of A_x, A_y and A_z
  $S_2 \leftarrow$Spectrum$(A_m.w)$;
  $S_3 \leftarrow$Spectrum$(G.w)$;
  $S_4 \leftarrow$Spectrum$(G_m.w)$;
  $k_1 = round(T_w.Fr_1)$;
  $k_2 = round(T_w.Fr_2)$;
  $k_3 = round(T_w.Fr_3)$;
  $F_C \leftarrow [S_1(1:k_1)\ \ S_2(1:k_1)\ \ S_3(1:k_1)\ \ S_4(1:k_1)]$;
  $F_D \leftarrow [S_1(1:k_2)\ \ S_2(1:k_2)\ \ S_3(1:k_2)\ \ S_4(1:k_2)]$;
  $F_P \leftarrow [S_1(1:k_3)\ \ S_2(1:k_3)\ \ S_3(1:k_3)\ \ S_4(1:k_3)]$;
end
Function `Spectrum` (S)
  $S_f \leftarrow$`fft` (S); // Fast Fourier transform calculation
  $S_{ESD} \leftarrow S_f.S_f^*$; // Energy spectral density calculation (S_f^* is the complex conjugate of S_f)
  return S_{ESD};
end

transitions. Here, $SVMs_C$ classifier is fed by feature vector $F_C = [C_1, .., C_7]$ containing ESD coefficients in $[0Hz - 2.3Hz]$. In the second level, dynamic activities are classified using $SVMs_D$ classifier, the vector $F_D = [C_1, .., C_{10}]$ is employed as input ($Fr_2 = 3.5Hz$). In a similar way, postures are recognized employing $SVMs_P$ classifier fed by $F_P = [C_1, .., C_{10}]$ with $Fr_3 = 3.5Hz$. Noted that the frequency bands presented here are determined empirically based on system accuracy, i.e., the model is evaluated by varying Fr for each feature vector from $0Hz$ (DC components) to $25Hz$ (all coefficients). The results of this experiment are detailed in section III-B.

C. Correction of misclassification

The output label given by *SVMs* classifiers is put in a misclassification correction process.The impulsive prediction error is rectified based on the following hypothesis: in our experiments, as $O_v = 50\%$, then a segmented window at an instant t named W_t shares 50% of data with W_{t-1}, and the remaining 50% with W_{t+1}. It implies that in the case of W_{t-1} and W_{t+1} outputs the same activity label, it should be equal to W_t output label. The implementation of the proposed solution is reported in the algorithm 2.

Algorithm 2: Correction of misclassification

Data:
$F(t)$: Feature extracted from inertial signals at an instant t
$PR(t-1) = [pred(t-N), .., pred(t-2), pred(t-1)]$: Vector containing labels of predicted activities
$pr \leftarrow SVM\_Predict(F(t))$;
if $[pr \neq PR(end)]$ *AND* $[pr == PR(end-1)]$ **then**
  $PR(end) \leftarrow pr$;
end
$PR = [PR \quad pr]$;

III. Experiments

A. Experimental setup

The experiments are conducted on SBHAR dataset [7], it contains acceleration and angular rate signals sampled at $F_s = 50Hz$. Six activities are performed by thirty subjects, each one wore a Samsung Galaxy S2 smartphone on the waist. Three of collected activities are static (standing, sitting and lying) and three are dynamic (walking, up and down stairs). In addition, the dataset also contains transitions between static activities named postural transitions (stand to sit, sit to stand, sit to lie, lie to sit, stand to lie, lie to stand). Each subject performed a scenario of sequential activities twice.

In order to estimate the overall recognition performance of our system and to compare it to state-of-the-art methods, four metrics are calculated: *recall*, *precision*, F_1 score and *accuracy*. In the state-of-the-art methods evaluated on SBHAR dataset, different scores are calculated, mostly accuracy. Hence, in order to make a fair comparison with these approaches, for each method, *recall*, *precision*, *accuracy* and F_1 scores are either recovered if it was already estimated like in our paper or recomputed using the confusion matrix that is given in the article. For papers where two methods are proposed, a comparison with each method noted m_1 and m_2 is carried out.

A subject-independent validation is carried out using leave-one-subject-out-cross-validation (LOSOCV) and hold-out subjects (proposed by [13]), where data from 21 randomly selected subjects are used for training, and the remaining 9 subjects for the test.

B. Optimal frequency band

Before presenting the results, an experimental study is carried out to determine the optimal frequency bands to construct F_D and F_P vectors. In this experiment, *LOSOCV* strategy is applied. The *recall* score is calculated for each value of cutoff frequency Fr. The obtained results are shown in the Figure 2. We can see clearly a growth of the score for postures and dynamic activities, the maximum is reached at the frequency $Fr \simeq 3.5Hz$, a slight decrease is observed beyond this frequency. Noted that in this experiment, the vector F_C is constructed from coefficients included in the band $[0Hz - 2.3Hz]$, a perfect separation between the postures and dynamic activities is then achieved (score of 100%).

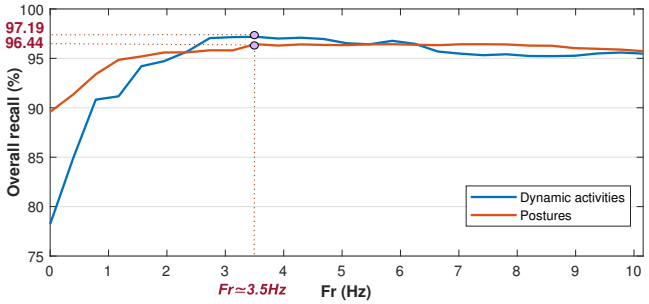

Fig. 2: *Recall* score depending on Fr.

C. Experimental results

In the state-of-the-art, a few works investigated the recognition of postural transitions using a smartphone, hence, in our experiments, the performances of physical activities and postural transitions are presented separately.

1) Physical activities: Table I reports performance scores of our system to recognize the physical activities and a comparison with state-of-the-art methods. It shows scores around 97% for both LOSOCV and hold-out subjects strategies, outperforming the existing approaches tested on SBHAR dataset. The confusion matrix is shown in Table II. All activities achieve a good performance score. However, some confusions appear for "sitting" and "standing" postures due to the position of the smartphone (waist), making the output signals similar for these postures.

2) Postural transitions: Table III reports performance scores of our method to recognize postural transitions and comparison with two methods proposed by [7] (the only existing method performing recognition of PTs on SBHAR dataset). It shows that the proposed approach achieves a *recall* score of 99.81% using LOSOCV strategy, outperforming both methods proposed by [7]. With hold-out subjects strategy, an overall *recall* of 99.63% is obtained.

Validation strategy	Relevant Studies	Recall (%)	Precision (%)	F-score (%)	Accuracy (%)
LOSOCV[1]	[7] (m_1)				96.74
	[7] (m_2)				96.50
	[14]	78.23	78.45	78.34	77.81
	Ours	**97.09**	**97.10**	**97.10**	**97.09**
HS[2]	[13]	89.35	89.23	89.29	88.97
	[15]				83.51
	[16] (m_1)	88.12	87.30	87.71	87.03
	[16] (m_2)	90.35	89.75	90.05	90.13
	[17]	92.67	92.95	92.81	92.92
	[18]				95.18
	[10] (m_1)	94.79	94.78	94.79	94.50
	[10] (m_2)	95.75			
	[11]				95.46
	[19]	95.90	95.97	95.93	
	[20]	96.59	96.86	96.72	96.64
	Ours	**97.02**	**97.17**	**97.10**	**97.04**

[1] Leave-one-subject-out cross-validation.
[2] Hold-out subjects.

TABLE I: Scores of relevant studies on SBHAR dataset.

	W	WU	WD	Si	St	L	PT
Walking	**97.83**	2.07	0	0	0	0	0.1
W. upstairs	0.19	**95.49**	3.5	0	0	0	0.82
W. downstairs	0.23	1.37	**98.25**	0	0	0	0.15
Sitting	0	0	0	**95.05**	4.78	0	0.17
Standing	0	0	0	5.26	**94.69**	0	0.05
Laying	0	0	0	0	0	**99.58**	0.42
P. transition	0	0	0	0.14	0.05	0	**99.81**

TABLE II: Confusion matrix on SBHAR dataset (%).

Validation strategy	Relevant Studies	Recall (%)	Precision (%)	F-score (%)
LOSOCV[1]	[7] (m_1)	92.18		
	[7] (m_2)	99.76		
	Ours	**99.81**	**98.29**	**99.04**
HS [2]	**Ours**	**99.63**	**97.89**	**98.75**

[1] Leave-one-subject-out cross-validation.
[2] Hold-out subjects.

TABLE III: Scores of postural transitions.

IV. Conclusions

In this paper, we introduced a novel method to carry out a real-time recognition of physical activities and postural transitions. Three meta-classes representing the different categories of activity (postures, dynamic activities and postural transitions) have been created. A hierarchical classification scheme has been proposed to classify efficiently each group of activity. Each internal node is composed of *SVMs* classifier and made of frequency domain features extracted from an optimal frequency band determined empirically. The experimental results on a challenging publicly available dataset named SBHAR have demonstrated that our method is effective and outperforms various state-of-the-art approaches, reaching overall scores around 97% for physical activities and 99.81% for postural transitions.

References

[1] D. Tran and A. Sorokin, "Human activity recognition with metric learning," in *European conference on computer vision*. Springer, 2008, pp. 548–561.

[2] S. A. W. Talha, M. Hammouche, E. Ghorbel, A. Fleury, and S. Ambellouis, "Features and classification schemes for view-invariant and real-time human action recognition," *IEEE Transactions on Cognitive and Developmental Systems*, 2018.

[3] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31 314–31 338, 2015.

[4] S. Liu, R. X. Gao, D. John, J. W. Staudenmayer, and P. S. Freedson, "Multisensor data fusion for physical activity assessment," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 687–696, 2012.

[5] M. Á. Á. de la Concepción, L. M. S. Morillo, J. A. Á. García, and L. González-Abril, "Mobile activity recognition and fall detection system for elderly people using ameva algorithm," *Pervasive and Mobile Computing*, vol. 34, pp. 3–13, 2017.

[6] L. Atallah, B. Lo, R. King, and G.-Z. Yang, "Sensor positioning for activity recognition using wearable accelerometers," *IEEE transactions on biomedical circuits and systems*, vol. 5, no. 4, pp. 320–329, 2011.

[7] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754–767, 2016.

[8] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 5, pp. 1166–1172, 2010.

[9] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307–313, 2018.

[10] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.

[11] Y. Mohammad, K. Matsumoto, and K. Hoashi, "Primitive activity recognition from short sequences of sensory data," *Applied Intelligence*, pp. 1–14, 2018.

[12] P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single triaxial accelerometer," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1780–1786, 2014.

[13] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *International workshop on ambient assisted living*. Springer, 2012, pp. 216–223.

[14] B. Kolosnjaji and C. Eckert, "Neural network-based user-independent physical activity recognition for mobile devices," in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2015, pp. 378–386.

[15] Y.-J. Kim, B.-N. Kang, and D. Kim, "Hidden markov model ensemble for activity recognition using tri-axis accelerometer," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3036–3041.

[16] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4566–4578, 2016.

[17] C. A. Ronao and S.-B. Cho, "Recognizing human activities from smartphone sensors using hierarchical continuous hidden markov models," *International Journal of Distributed Sensor Networks*, vol. 13, no. 1, p. 1550147716683687, 2017.

[18] W. Jiang and Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 2015, pp. 1307–1310.

[19] E. Zdravevski, P. Lameski, V. Trajkovik, A. Kulakov, I. Chorbev, R. Goleva, N. Pombo, and N. Garcia, "Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering," *IEEE Access*, vol. 5, pp. 5262–5280, 2017.

[20] M. Seera, C. K. Loo, and C. P. Lim, "A hybrid fmm-cart model for human activity recognition," in *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 182–187.