

Reconnaissance en temps réel d'activités physiques et de transitions posturales à l'aide de capteurs inertiels pour smartphone

Sid Ahmed Walid Talha, Anthony Fleury, Stéphane Lecoeuche

▶ **To cite this version:**

Sid Ahmed Walid Talha, Anthony Fleury, Stéphane Lecoeuche. Reconnaissance en temps réel d'activités physiques et de transitions posturales à l'aide de capteurs inertiels pour smartphone. XXVIIème Colloque francophone de traitement du signal et des images, Aug 2019, Lille, France. hal-02523809

HAL Id: hal-02523809
https://hal.science/hal-02523809

Submitted on 29 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reconnaissance en temps réel d'activités physiques et de transitions posturales à l'aide de capteurs inertiels pour smartphone

Sid Ahmed Walid TALHA, Anthony FLEURY, Stéphane LECOEUCHE

Unité de Recherche Informatique et Automatique
IMT Lille Douai, Univ. Lille, F-59000 Lille, France.
{prénom.nom}@imt-lille-douai.fr

Résumé – Cet article présente une nouvelle approche pour la classification en temps réel des activités humaines à l'aide de données provenant de capteurs inertiels embarqués dans un smartphone. Nous proposons un système de classification hiérarchique permettant de reconnaître sept classes d'activités y compris les transitions posturales (PTs). Sa structure contient trois nœuds internes composés de trois Support vector machines (*SVM*) classifieurs où chacun est associé à un ensemble d'activités. De plus, chaque *SVM* reçoit un vecteur d'attributs provenant d'une bande de fréquence adaptée et optimale. Les résultats expérimentaux réalisés sur une base de données publique, montrent que notre méthode est efficace et surpasse les approches de l'état de l'art. Nous montrons également la pertinence de notre méthode pour reconnaître les transitions posturales.

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 *SVM* classifiers, each one is associated with a set of activities. Moreover, each *SVM* 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.

1 Introduction

La reconnaissance de l'activité humaine (HAR) est devenue un domaine de recherche important en raison d'une forte demande dans divers domaines d'application notamment la santé. Depuis des décennies, la vision par ordinateur s'avère le moyen le plus utilisé pour développer des systèmes HAR [1]. Cependant, ces approches souffrent de certaines limitations dues aux contraintes de l'environnement, telles que les conditions d'éclairage ou les occlusions dues aux personnes et aux objets. De plus, en ce qui concerne le contexte de la confidentialité, les caméras sont plus intrusives et perturbent les utilisateurs. Les développements des capteurs inertiels offrent un choix intéressant et leur intégration dans les smartphones a retenu l'attention car cet élément de la vie quotidienne est courant.

Dans ces systèmes, pour décrire chaque activité, un vecteur d'attributs est construit avec des attributs : temporels (statistiques du premier et du second ordre, coefficient de corrélation, histogramme, etc.), fréquentiels, avec la transformée de Fourier discrète (DFT), et temps-fréquence, avec la transformée en ondelettes discrète (DWT) ou la transformée par paquets d'ondelettes (WPD). Pour l'étape de classification, diverses méthodes ont été expérimentées dans la littérature. Récemment, des réseaux de neurones profonds ont été utilisés [2, 3, 4].

En plus des activités physiques, certains travaux détectent les transitions entre les postures [5]. Cependant, peu ont réalisé la classification des transitions posturales à l'aide d'un smartphone. Dans [6], deux approches sont proposées pour détecter les transitions posturales : dans la première, seules les activités de base sont apprises par le classifieur, les transitions, en revanche, sont considérées comme un changement d'activités détecté par le système. Dans la deuxième approche, en plus des activités de base, les transitions sont apprises par le classifieur et sont considérées comme une classe supplémentaire.

Par la suite, une nouvelle approche HAR basée sur les données smartphone est proposée, pour une classification en temps réel des activités physiques et des transitions posturales. La principale contribution de ce travail est l'élaboration d'un système de classification hiérarchique permettant de classer trois groupes d'activités : les postures, les activités dynamiques et les transitions posturales. Chaque nœud interne du modèle est composé d'un *SVM* alimenté par un ensemble d'attributs adaptés et optimaux. Le reste de l'article est organisé comme suit : La section 2 présente la méthodologie proposée. La section 3 décrit les configurations expérimentales, puis présente les résultats en les comparant aux approches existantes. Enfin, la section 4 conclut ce travail.

2 Approche proposée

Les différentes étapes de notre approche sont représentées sur la figure 1. Notre méthodologie utilise des signaux inertiels recueillis à partir des capteurs accéléromètre et gyroscope.

Cela implique la mise en jeu des signaux suivants : l'accélération (A_x, A_y, A_z), vitesse angulaire (G_x, G_y, G_z) et les signaux d'amplitudes A_m et G_m avec $A_m = \sqrt{A_x^2 + A_y^2 + A_z^2}$ et $G_m = \sqrt{G_x^2 + G_y^2 + G_z^2}$.

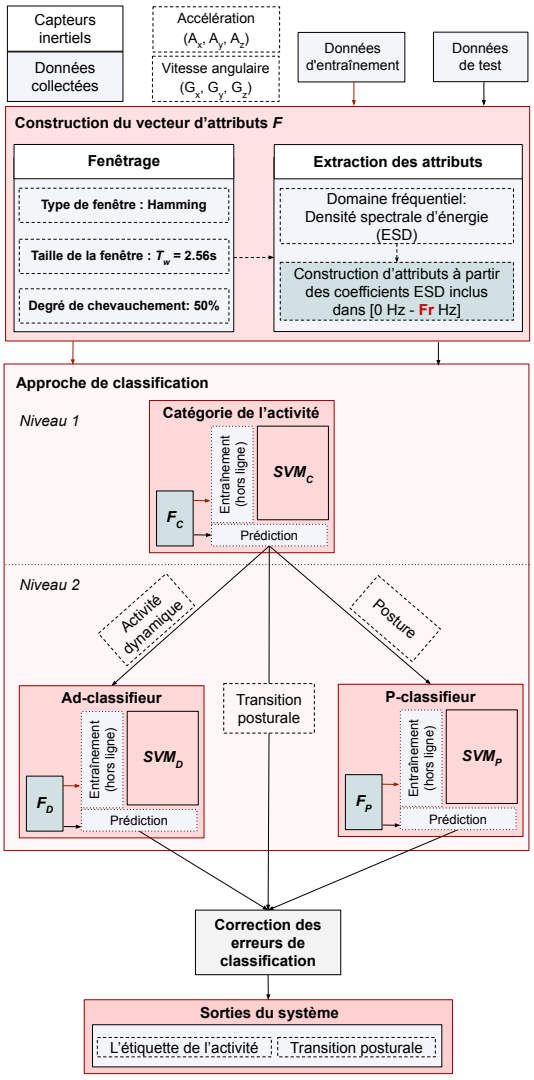


FIGURE 1 – Aperçu du système proposé.

2.1 Fenêtrage

Dans la littérature, une fenêtre rectangulaire est généralement utilisée dans les opérations de fenêtrage. Cependant, dans notre approche, les attributs sont calculés à partir du domaine fréquentiel des signaux inertiels. Par conséquent, une fenêtre rectangulaire n'est pas un choix avantageux, elle présente un facteur de fuite élevé (leakage). Afin de minimiser cet effet, une fenêtre de Hamming est utilisée. La taille de la fenêtre est fixée à $2.56s$ et le degré de chevauchement nommé O_v à 50%. Ces paramètres sont réglés de la même manière que ceux des approches expérimentées sur la base de données SBHAR, afin de nous permettre une comparaison équitable.

2.2 Extraction d'attributs et classification

Dans notre approche, les attributs sont extraits à partir du domaine fréquentiel comme suit : pour chaque signal, après l'opération de fenêtrage, la DFT est obtenue par la FFT, puis la densité spectrale d'énergie (ESD) est calculée. Le vecteur d'attributs est construit à partir des coefficients inclus dans la bande de fréquence $[0Hz - \boldsymbol{Fr}Hz]$ (Fr indique la limite de fréquence supérieure). La taille de la bande de fréquence dépend du problème de classification. Comme le montre la figure 1, une méthode de classification hiérarchique composée de trois classifieurs *SVM* est proposée, chacun est alimenté par un vecteur d'attributs nommés : F_C, F_D et F_P ({C,D,P} dénote "catégorie d'activité", "activité dynamique" et "posture"). Chaque couple (F_i, SVM_i) avec $i \in \{C, D, P\}$ est construit pour résoudre un problème de classification spécifique. Dans le premier niveau du schéma, les activités sont divisées en trois catégories : activités dynamiques, postures et transitions posturales. Pour ce faire, le classifieur SVM_C est alimenté par le vecteur d'attributs $F_C = [C_1, .., C_7]$ contenant les coefficients ESD dans $[0Hz - 2.3Hz]$. Au deuxième niveau, les activités dynamiques sont classées à l'aide du classifieur SVM_D. Le vecteur $F_D = [C_1, .., C_{10}]$ est utilisé en tant qu'entrée ($Fr_2 = 3.5Hz$). De manière similaire, les postures sont reconnues en employant le classifieur SVM_P alimenté par $F_P = [C_1, .., C_{10}]$ avec $Fr_3 = 3.5Hz$. Il est à noter que les bandes de fréquences présentées sont déterminées de manière empirique en fonction de la précision du système, c'est-à-dire que le modèle a été évalué en faisant varier Fr pour chaque vecteur d'attributs de $0Hz$ (composante continue) à $25Hz$ (tous les coefficients). Les résultats de cette expérience sont décrits dans la section 3.2.

2.3 Correction des erreurs de classification

L'étiquette de sortie donnée par le modèle de classification est mise à l'entrée d'une méthode de correction. L'erreur impulsive de prédiction est rectifiée sur la base de l'hypothèse suivante : dans nos expériences, puisque $O_v = 50\%$, alors une fenêtre segmentée à un instant t nommée W_t partage 50% de données avec W_{t-1}, et les 50% restants avec W_{t+1}. Cela implique que dans le cas où W_{t-1} et W_{t+1} génèrent la même étiquette, elle doit être égale à l'étiquette de W_t. L'implémentation de la solution proposée est décrite dans l'algorithme 1.

3 Expérimentations et résultats

3.1 Protocole expérimental

Les tests expérimentaux sont effectués sur la base de données SBHAR [6]. Elle contient les signaux d'accélération et de vitesse angulaire échantillonnés à $F_s = 50Hz$. Six activités sont effectuées par trente sujets, chacun portant un smart-

Algorithm 1: Correction des erreurs de classification

Data:
$F(t)$: Vecteur d'attributs à l'instant t
$PR(t-1) = [pred(t-N), .., pred(t-2), pred(t-1)]$: Vecteur contenant les étiquettes d'activités prédites
$pr \leftarrow SVM_Predict(F(t))$;
if $[pr \neq PR(end)] \;\; AND \;\; [pr == PR(end-1)]$ **then**
  $PR(end) \leftarrow pr$;
end
$PR = [PR \quad pr]$;

phone Samsung Galaxy S2 à la taille. Trois de ces activités sont des postures (debout, assis et allongé) et les trois autres dynamiques (marcher, monter et descendre les escaliers). En outre, la base de données contient également des transitions entre les postures appelées transitions posturales (les passages : assis-debout, debout-assis, allongé-assis, assis-allongé, debout-allongé, allongé-debout).

Pour estimer la performance de reconnaissance globale de notre système et la comparer à des méthodes de la littérature, quatre mesures sont calculées : *précision*, *rappel*, *F-score* et *exactitude* (*accuracy*). Dans les approches de l'état de l'art évaluées sur la base données SBHAR, différents scores ont été calculés, principalement *l'exactitude*. Par conséquent, afin de réaliser une comparaison objective avec ces approches, pour chaque méthode, les quatre scores considérés sont soit récupérés s'ils ont déjà été estimés, soit recalculés à l'aide de la matrice de confusion fournie dans les articles. Pour les travaux où deux approches sont proposées, une comparaison avec chaque approche notée m_1 et m_2 est effectuée.

Une validation croisée indépendante du sujet est réalisée en appliquant les techniques de validation suivantes : leave-one-subject-out-cross-validation (LOSOCV) et une validation simple (hold-out) mais indépendante du sujet (proposée par [7]), où les données de 21 sujets choisis aléatoirement sont utilisées pour l'apprentissage, et les 9 autres sujets pour le test.

3.2 Bandes fréquentielles optimales

Avant d'entamer la comparaison avec les approches de l'état de l'art, une étude expérimentale est réalisé afin de déterminer les bandes fréquentielles optimales pour construire les vecteurs F_D et F_P. Dans cette expérience, la stratégie de validation $LOSOCV$ est appliqué. Le score $rappel$ est calculé en fonction de la fréquence de coupure Fr. Les résultats obtenus sont représentés sur la figure 2. On peut voir clairement une croissance du score pour les deux catégories d'activités, le maximum est atteint à la fréquence $Fr \simeq 3.5Hz$, une légère baisse est observée au-delà de cette fréquence. Il est à noter que dans cette expérience, le vecteur F_c est construit à partir des coefficients inclus dans la bande $[0Hz - 2.3Hz]$, une séparation parfaite entre les postures et activités dynamiques est alors réalisée (score de 100%).

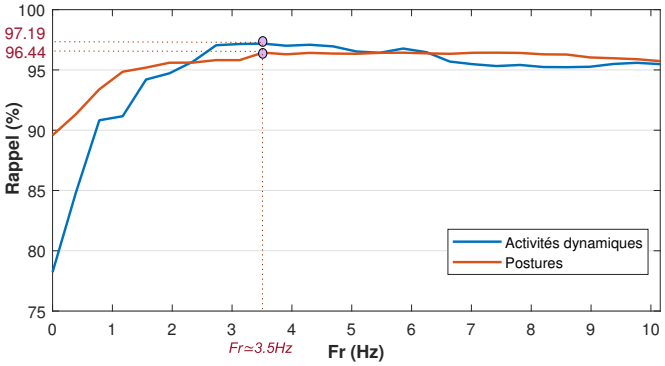


FIGURE 2 – Évolution du score $rappel$ en fonction de Fr.

Validation	Études pertinentes	Rappel (%)	Précision (%)	F-score (%)	Exactitude (%)
LOSOCV[1]	[6] (m_1)				96.74
	[6] (m_2)				96.50
	[8]	78.23	78.45	78.34	77.81
	Nous	**97.24**	**97.25**	**97.24**	**97.24**
Hold-out[2]	[7]	89.35	89.23	89.29	88.97
	[9]				83.51
	[10] (m_1)	88.12	87.30	87.71	87.03
	[10] (m_2)	90.35	89.75	90.05	90.13
	[11]	92.67	92.95	92.81	92.92
	[3]				95.18
	[2] (m_1)	94.79	94.78	94.79	94.50
	[2] (m_2)	95.75			
	[4]				95.46
	[12]	95.90	95.97	95.93	
	[13]	96.59	96.86	96.72	96.64
	Nous	**97.02**	**97.17**	**97.10**	**97.04**

[1] Leave-one-subject-out cross-validation.
[2] Validation simple indépendante du sujet.

TABLE 1 – Scores de reconnaissance des activités physiques.

3.3 Comparaison avec l'état de l'art

Dans l'état de l'art, peu de travaux ont porté sur la reconnaissance des transitions posturales à l'aide d'un smartphone ; par conséquent, les résultats de reconnaissance des activités physiques et des transitions posturales sont présentées séparément.

3.3.1 Activités physiques

La table 1 présente les scores de reconnaissance des activités physiques avec notre approche, plus une comparaison avec les méthodes de la littérature. Des scores autour de 97% sont alors obtenus pour les deux stratégies de validations, surpassant les approches existantes.

La matrice de confusion est décrite dans le tableau 2. Elle montre que toutes les activités atteignent un bon score de performance. Cependant quelques confusions apparaissent pour les postures "assis" et "debout", elles sont dues à la position du smartphone (la taille) rendant les signaux de sorties similaires.

	M	ME	DE	As	D	Al	TP
Marcher	**97.83**	2.07	0	0	0	0	0.1
M. escaliers	0.19	**95.49**	3.5	0	0	0	0.82
D. escaliers	0.23	1.37	**98.25**	0	0	0	0.15
Assis	0	0	0	**95.05**	4.78	0	0.17
Debout	0	0	0	5.26	**94.69**	0	0.05
Allongé	0	0	0	0	0	**99.58**	0.42
T. posturales	0	0	0	0.14	0.05	0	**99.81**

TABLE 2 – Matrice de confusion (%).

3.3.2 Transitions posturales

La table 3 présente les scores de reconnaissance des transitions posturales avec notre approche, plus une comparaison avec les deux méthodes proposées par [6] (les seules méthodes existantes permettant la reconnaissance des PTs sur la base de données SBHAR). Il en ressort que l'approche proposée a atteint un score $rappel$ de $99,81\%$ en utilisant la stratégie LOSOCV, surpassant les deux méthodes proposées par [6]. Avec la validation simple, un rappel de $99,63\%$ est obtenu.

Validation	Études pertinentes	Rappel (%)	Précision (%)	F-score (%)
LOSOCV	[6] (m_1)	92.18		
	[6] (m_2)	99.76		
	Nous	**99.81**	**98.29**	**99.04**
Hold-out	**Nous**	**99.63**	**97.89**	**98.75**

TABLE 3 – Scores de reconnaissance des transitions posturales.

4 Conclusion

Dans cet article, nous avons introduit une nouvelle méthode pour réaliser une reconnaissance en temps réel des activités physiques et transitions posturales. Une méta-classe nommée catégorie d'activité comprenant postures, activités dynamiques et transitions posturale a été créé. Un système de classification hiérarchique a été proposé afin de classer efficacement chaque groupe d'activités. Chaque nœud interne est composé de classifieur *SVM* et des attributs du domaine fréquentiel extraits à partir d'une bande de fréquence optimale déterminée empiriquement. Les résultats expérimentaux obtenus à partir d'une base de données publique ont démontré l'efficacité et la robustesse de notre méthode par rapport aux diverses approches de l'état de l'art en atteignant un score d'environ 97% pour les activités physiques et 99.81% pour les transitions posturales.

Références

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.