

Reconnaissance de l'activité humaine dans un environnement non contrôlé en utilisant l'internet des objets

Résumé—La reconnaissance de l'activité humaine, dans un environnement non contrôlé, reste une problématique de taille. Dans cet article, nous proposons une méthodologie permettant de reconnaître quatre types d'activité : debout, assis, allonger et marcher. Nous nous demandons si nous pouvons obtenir une bonne précision de reconnaissance de l'activité en utilisant des objets connectés. Cette étude exploite un smartphone et une smart Watch standards, portés par des participants durant des expériences non contrôlées. Nous proposons un pré-calcul en utilisant la transformée en cosinus discrète (DCT), nous identifions la meilleure taille de la fenêtre de découpage des signaux et le meilleur descripteur permettant d'avoir les meilleurs résultats de reconnaissance. Nous montrons que Support Vector Machines (SVM) permet d'avoir les meilleurs résultats par rapport aux autres algorithmes de classification (arbres de décision(DT)). Les résultats montrent que la précision de classification a atteint 91%, dans un environnement non contrôlé et en positions non contrôlées du smartphone. Nous définissons la notion de *transition* qui correspond à la transition entre deux activités et entre deux positions du smartphone. Le dernier résultat montre que la suppression des *transitions* permet d'obtenir de meilleurs résultats de classification : 98%.

Mots clés— Internet des objets, reconnaissance de l'activité, classification automatique, environnement non contrôlé.

Catégories et descripteurs du sujet :

[Reconnaissance de formes] : prétraitements, algorithmes de classification.

I. INTRODUCTION

L'émergence de l'Internet des objets (IdO) dans la vie quotidienne ouvre la porte à de nombreuses perspectives dans le domaine de l'interaction Homme-machine. Différents travaux de recherche ont étudié la reconnaissance de l'activité en utilisant des objets connectés [1, 2, 3, 6]. En effet, cette approche fournit des plates-formes bon marché et légères pour

un suivi transparent des activités avec différents types de capteurs embarqués (ex. centrales inertielles, capteur de proximité, capteur de pression, capteur de température,...). Ces capteurs peuvent être exploités avec des algorithmes de fusion de données [4,5].

Cependant, la recherche actuelle n'exploite pas suffisamment ce type de plates-formes. Des études ont été menées sur la reconnaissance de l'activité dans des environnements contrôlés, en demandant aux participants d'effectuer des tâches spécifiques (ex., se mettre debout, assis, monter l'escalier...) [9], ou en demandant aux participants d'annoter le placement du smartphone avant de commencer l'expérience (ex. dans la poche du pantalon, à la main, ...) [10]. En outre, de nombreuses études ont porté sur la reconnaissance de l'activité avec des capteurs dans des configurations spécifiques en fixant les dispositifs dans certains endroits sur le corps (ex. Smartphone dans la poche) [11] [13]. *Les transitions* de la position du smartphone d'un endroit à un autre ne sont pas abordées dans ces travaux. Enfin, les études qui tiennent compte de *la transition* entre les activités ont montré une diminution substantielle de la précision de reconnaissance [14]. Ainsi, la plupart des études suppriment les données de transition (ex. transition lors du passage de la position debout à la position assise, etc.) et ne gardent que les données correspondant à des configurations statiques (ex. assis, marché) [11].

En ce qui concerne les méthodes de classification déployées, plusieurs approches ont été utilisées avec des objets connectés: HMM [9], SVM [10,11,12,13,14], les arbres de décision, forêts aléatoires [10], KNN et NN [11,15]. Ces méthodes présentent des caractéristiques différentes et nécessitent des prétraitements spécifiques avant application de ces algorithmes (modèles multidimensionnels, etc.).

Le présent article se propose d'aller au-delà des méthodes actuelles de traitement afin de fournir une meilleure précision de reconnaissance des activités humaines en utilisant des objets connectés habituels. Sur la base de capteurs embarqués dans une smart Watch et un smartphone standards, nous allons aborder une série de questions pour concevoir un processus de traitement efficace pour la reconnaissance de quatre types

d'activité en utilisant des algorithmes de classification. Le Smartphone (un Iphone), comprend un accéléromètre et un capteur de proximité. La smartwatch (Apple watch), comprend un accéléromètre.

Premièrement, nous avons observé l'impact de l'utilisation de deux algorithmes de classification sur la précision de la reconnaissance de l'activité: SVM et les arbres de décision (DT). Ces algorithmes sont plus adaptés aux modèles multidimensionnels, car les autres méthodes exigent la réduction de la dimension des descripteurs.

Deuxièmement, l'étude a été effectuée dans un environnement non contrôlé. Les participants ont été observés au cours des activités quotidiennes sans instructions.

Troisièmement, nous avons proposé une nouvelle méthode pour le prétraitement des données sur la base de la Transformée en Cosinus Discrète (DCT) [7].

Quatrièmement, nous avons étudié l'effet des *transitions* à la fois pendant le changement des états d'activité du sujet (ex. Passer de la position assise à la position debout), et le changement de la localisation du Smartphone (ex. passage du smartphone de la main à la poche du pantalon). Nous avons appelé cela : *les transitions*.

Enfin, nous avons étudié l'effet de *la transition* sur le taux de reconnaissance de l'activité. La smartwatch étant fixée au niveau du poignet, le Smartphone passe d'une localisation à une autre, sur le corps (de la main à la poche du pantalon et vice versa).

II. COLLECTE DE DONNÉES

L'expérience a eu lieu dans une maison pendant une semaine. Trois participants, âgés entre 30 et 48 ans (une femme et deux hommes), ont été impliqués dans l'expérience au cours d'une semaine chacun. Trois caméras IP (Aquila vizion wifi 720p) ont été fixées à différents endroits de la pièce pour enregistrer l'activité des participants. Les vidéos ont été enregistrées dans un serveur local via une connexion wifi.

On a demandé aux participants de porter une smart Watch et un smartphone pendant l'expérience. Le smartphone (S1) contient deux capteurs : un accéléromètre à trois axes (Acc1) et un capteur de proximité (PS). La smartwatch embarque un accéléromètre à trois axes (Acc2). La fréquence d'échantillonnage des signaux de capteurs a été réglée à 120 Hz. La durée d'enregistrement est de 3 heures et 45 minutes, trois fois par jour pendant une semaine. Les données ont été envoyées via wifi à un serveur local, et a été stocké dans une base de données SQL SERVER. Un entrepôt de données a été créé pour intégrer automatiquement les données de la base de données à la fin de chaque enregistrement. Les enregistrements des signaux des capteurs et les vidéos ont été synchronisés et ont commencé simultanément. Les enregistrements vidéo et les signaux de capteurs ont été annotés avec le logiciel ELAN.

Les participants ont positionné le Smartphone soit dans la poche du pantalon ou soit le maintenir dans la main. La

smartwatch a été attaché au poignet (à gauche pour la femme et à droite pour les deux hommes). La figure 1 montre les différents états d'activité et la position des deux objets connectés.



Figure 1: Les quatre états d'activité (assis, debout, marcher et allonger) avec les différentes positions du smartphone (la smartwatch est fixée au niveau du poignet, le smartphone est posé soit dans la poche soit dans la main)

III. PREPARATION DES DONNEES

A. Traitement des signaux:

Dans cette section, nous nous concentrons sur la représentation parcimonieuse du signal. Nous avons utilisé la Transformée en Cosinus Discrète (DCT), plus précisément, la DCT-II [7]. La DCT est un très bon décorrélateur de signaux. Il permet également le regroupement de l'énergie dans les coefficients basses fréquences en raison de son approximation de la transformation de Karhunen Loève et de l'analyse en composantes principales (ACP).

La DCT applique une transformation au signal de départ et donc la majeure partie de son énergie est projetée dans une zone restreinte (un nombre réduit de coefficients) de l'espace transformé. La transformation utilisée est linéaire afin de fournir une solution analytique pour la reconstruction ultérieure du signal. La largeur de la fenêtre Δt est un paramètre important pour la méthode adoptée. Il est illustré sur la figure 2.

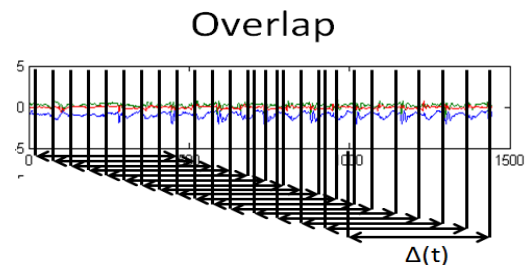


Figure 2: Le signal représentant le descripteur de la marche et la fenêtre de découpage (Δt).

IV. LES RESULTATS

A. Largeur de la fenetre de découpage:

Dans cette section, toutes les classifications ont été faites en utilisant Support Vector Machines (SVM) avec un noyau linéaire et entraînés avec *10-fold cross validation*. Dans le but d'identifier la taille optimale de la fenêtre de découpage de signaux, nous avons choisi de comparer deux tailles de fenêtres : $\Delta t = 256$ (2.13 s) et $\Delta t = 512$ (4.26 s), avec ou sans chevauchement entre les fenêtres :

(ΔT)	512		256	
	NON	OUI	NON	OUI
Standing/other	72.29%	78.97%	64.1%	68.3%
Sitting/other	71.19%	78.39%	66.3%	69.1%
Walking/having	71.26%	78.79%	66.2%	69.9%
AVG	71.24%	78.71%	65.53%	69.1%

Tableau 1: Performance de classification quand il y a (Oui) chevauchement ou pas (NON) entre les fenêtres de découpage, pour l'accéléromètre 1

Le tableau 1 montre que la performance de classification est meilleure quand la taille de la fenêtre de découpage est de 512 points avec un chevauchement de 256 points.

Nous avons établi les mêmes tests avec la smart Watch (acc2). La performance de classification est meilleure (91.03%) quand la taille de la fenêtre de découpage est de 512 points avec chevauchement de 256 points.

Ces mesures ont été prises avec une taille de descripteur égale à 144 (48 pour chaque axe). La taille optimale de chaque fenêtre de découpage qui permet d'avoir la meilleure performance est $\Delta t=512$ avec chevauchement entre les coupes, et ce quel que soit la position de l'accéléromètre.

B. Taille de descripteur :

Le but de cette section est de déterminer la meilleure taille de descripteur qui permet d'avoir les meilleures performances de classification. La taille du descripteur a été variée entre 2 et 80 et le calcul de la performance a été fait par rapport à ces dimensions en utilisant le classifieur SVM avec un noyau linéaire.

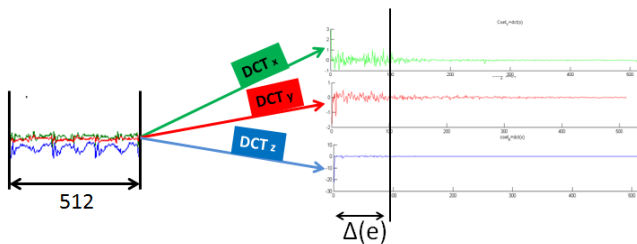


Figure 3: Détermination de la taille des descripteurs

Les performances de classification sont meilleures pour le smartphone et la smart Watch quand la dimension des descripteurs est de 48. Nous gardons donc les premiers 48 facteurs de la DCT, pour la suite de cet article.

Pour extraire les descripteurs du capteur de proximité, nous avons appliqué un filtre de fenêtre de moyennage avec une taille de 256 points pour lisser les données. Nous gardons la dérivée première des valeurs pour éliminer l'effet de décalage sur les données. Pour récupérer les quatre types d'activités à partir du capteur de proximité, nous avons appliqué une ACP sur les données prétraitées et utilisé la première composante principale comme premier descripteur. Nous avons ensuite appliqué Short Time Fourier Transform sur une fenêtre glissante de 256 échantillons sur le premier descripteur et

utilisé l'énergie moyenne de la bande entre 7.9 Hz et 14.7 Hz comme deuxième descripteur.

La section suivante décrit le processus d'apprentissage basé sur la DCT comme méthode de sélection de descripteurs:

C. Le descripteur DCT:

La DCT des axes des données des capteurs ont été concaténés et stockés dans une base pour apprentissage, et données en entrées de SVM et DT. La précision de classification a atteint 92.06% (SVM) contre 89.9% (DT)

D. Les transitions des localisations du smartphone:

Cette partie étudie l'effet des transitions du smartphone sur la reconnaissance de l'activité. Nous avons exploré l'impact de la suppression des patterns correspondant aux transitions entre deux positions du smartphone (de la poche du pantalon à la main et de la main vers la poche du pantalon). Nous avons collecté les positions suivantes : smartphone dans la main, smartphone dans la poche du pantalon. Nous avons comparé le temps correspondant aux transitions et le temps correspondant aux configurations fixes. Nous avons observé que le temps cumulatif de transition du smartphone depuis la poche du pantalon vers la main et de la main vers la poche du pantalon est très petit ou insignifiant comparé au temps où le smartphone est dans une configuration stable (fixe). Nous avons, ensuite, appliqué les algorithmes d'apprentissage pour classier les quatre types d'activités par rapport au cas suivant : sans et avec les données de transition.

Signals	S1	Acc2	S1, acc2
1- En prenant en considération les données de transition			
SVM	90.83%	93.4%	95.1%
DT	89.27%	91.2%	93.2%
2- Sans prendre en considération les données de transition			
SVM	92.06%	94.1%	96.12%
DT	89.2%	90.21%	93.87%

Tableau 2 : précisions de classification dans le cas de la transition des localisations du smartphone

En prenant en compte les données de transition, la performance de classification a atteint 95.1%, en fusionnant toutes les données de tous les capteurs. En revanche, en supprimant ces données de transition, la performance de classification à grimper vers 96.12%.

E. Temps de transition pour l'activité de transition:

Signals	S1	Acc2	S1, acc2
En prenant en compte les données de transition			
SVM	92.13%	92.8%	95.02%
Decision trees	89.2%	91.2%	92.32%
En supprimant les données de transition			
SVM	92.8%	93%	98.03%
Decision trees	89.29%	91.87%	92.21%
SVM	92.8%	93%	98.03%

Tableau 3 : précision de classification dans le cas où on prend ou pas les données de transition

V. DISCUSSION & CONCLUSION

Dans cet article, nous avons montré que la DCT, avec une largeur et taille appropriée de la fenêtre de découpage des signaux, peut conduire à des résultats très précis de reconnaissance de l'activité de classification. Dans cette étude, les sujets tiennent un smartphone équipé d'un accéléromètre et un capteur de proximité et portent une smart Watch, équipée d'un accéléromètre. L'étude a porté sur la reconnaissance des quatre types d'activités (debout, assis, marcher et allongé) dans un environnement non contrôlé. Nous avons particulièrement mis l'accent sur la notion de *transition*, qui correspondent aux transitions entre deux activités ainsi entre deux configuration des capteurs.

Les résultats de classification montrent une meilleure performance 98% en supprimant les données de transition entre deux activités et 95% de précision en supprimant les transitions entre deux localisations du smartphone. Ceci est dû à ce que les descripteurs des transitions peuvent biaiser la classification en fournissant des entrées similaires à une classe et une autre, mais différente. De plus, il est difficile de distinguer le moment exact de transition entre deux classes. Les descripteurs sur les frontières entre deux classes peuvent être considérés comme un bruit pour le classifieur. Le même raisonnement s'applique aux transitions entre deux localisations du smartphone, mais dans ce cas, le résultat est moins significatif.

Nous pouvons conclure que les données transitoires peuvent être supprimées sans perte significative d'information et peuvent même conduire à de meilleurs résultats de classification en supprimant les signaux transitoires qui peuvent ajouter du bruit au classifieur.

REFERENCES

- [1] Geroch, M. S. (2004). Motion capture for the rest of us. *J.Comput.Small Coll.*, 19(3), 157-164.
- [2] Fontana, J. M., Farooq, M., and Sazonov, E. Automatic Ingestion Monitor: A Novel Wearable Device for Monitoring of Ingestive Behavior. *IEEE Transactions on Biomedical Engineering* 61, 6 (2014), 1772–1779.
- [3] T. Isoda, Y. Nohara, S. Inoue, M. Shirouzu, Y. Sugiyama, M. Hirata, K. Machida, and N. Nakashima, « Experiment for nursing activity analysis using mobile sensors and proximity sensors », in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, 2015, p. 153–156.

- [4] Braud, T. and Ouarti, N. "Constrained Sigma Points for Attitude Estimation", *ICARCV* 2016, to appear
- [5] Braud, T. and Ouarti, N. "Comparison of Nonlinear Attitude Fusion Filter ", *Information Fusion (FUSION)*, 2016, 07/2016.
- [6] S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling, « In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass », in *Proceedings of the 5th augmented human international conference*, 2014, p. 15.
- [7] F. M. de S Matos and L. V. Batista, « Face recognition using DCT coefficients selection », in *Proceedings of the 2008 ACM symposium on Applied computing*, 2008, p. 1753–1757.
- [8] D. Meyer, « Support vector machines », *Porting R to Darwin/X11 and Mac OS X*, p. 23, 2001.
- [9] C. Zhu and W. Sheng, « Multi-sensor fusion for human daily activity recognition in robot-assisted living », in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 2009, p. 303–304.
- [10] A. Ghosh and G. Riccardi, « Recognizing human activities from smartphone sensor signals », in *Proceedings of the ACM International Conference on Multimedia*, 2014, p. 865–868.
- [11] L. Gao, A. K. Bourke, and J. Nelson, « A comparison of classifiers for activity recognition using multiple accelerometer-based sensors », in *Cybernetic Intelligent Systems (CIS)*, 2012 IEEE 11th International Conference on, 2012, p. 149–153.
- [12] K. Murao, H. Mogari, T. Terada, and M. Tsukamoto, « Evaluation function of sensor position for activity recognition considering wearability », in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, 2013, p. 623–632.
- [13] K. Ouchi and M. Doi, « Smartphone-based monitoring system for activities of daily living for elderly people and their relatives etc. », in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, 2013, p. 103–106.
- [14] N. Li, Y. Hou, and Z. Huang, « A real-time algorithm based on triaxial accelerometer for the detection of human activity state », in *Proceedings of the 6th International Conference on Body Area Networks*, 2011, p. 103–106.
- [15] S. A. Rahman, C. Merck, Y. Huang, and S. Kleinberg, « Unintrusive eating recognition using Google Glass », in *Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare*, 2015, p. 108–111.