



HAL
open science

Apport des techniques d'analyse et de traitement de données pour la reconnaissance des actions en vue d'un suivi du comportement humain

Sid Ahmed Walid Talha

► **To cite this version:**

Sid Ahmed Walid Talha. Apport des techniques d'analyse et de traitement de données pour la reconnaissance des actions en vue d'un suivi du comportement humain. Intelligence artificielle [cs.AI]. Ecole nationale supérieure Mines-Télécom Lille Douai, 2020. Français. NNT : 2020MTLD0006 . tel-02895440

HAL Id: tel-02895440

<https://theses.hal.science/tel-02895440>

Submitted on 9 Jul 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.

Apport des techniques d'analyse et de traitement de données pour la reconnaissance des actions en vue d'un suivi du comportement humain

par

Talha Sid Ahmed Walid

Pour obtenir le grade de

Docteur de l'Université de Lille délivré par IMT Lille Douai

Soutenue le 11 Février 2020

devant le jury composé de :

M. Eric CAMPO

Professeur, Université de Toulouse II Jean Jaurès, Rapporteur

M. Laurent NAVARRO

Enseignant-Chercheur HDR, Mines Saint-Etienne, Rapporteur

Mme. Sylvie CHARBONNIER

Maître de conférences, Université Grenoble Alpes, Examinatrice

M. Emmanuel CHAZARD

Professeur, Université de Lille, Examineur

M. Stéphane LECOEUICHE

Professeur, IMT Lille Douai, Directeur de thèse

M. Anthony FLEURY

Enseignant-Chercheur, HDR, IMT Lille Douai, Co-directeur de thèse

Résumé

Pour lutter contre la perte d'autonomie liée au vieillissement dû à des altérations physiques et/ou psychiques, les nouvelles technologies, œuvrent à retarder sa survenue, la détecter, l'évaluer et proposer des solutions modernes et innovantes. Dans ce contexte, notre projet de thèse vise à exploiter l'apport des techniques d'analyse et de traitement de données pour le suivi du comportement humain. Cette thèse cible deux parties importantes et complémentaires : la première réalise la synthèse journalière de l'ensemble des actions effectuées par la personne, afin de nous renseigner sur son degré d'autonomie. La deuxième partie propose une solution moderne basée sur l'exécution d'exercices physiques sous forme de mouvements contrôlés reconnus et corrigés. À partir d'une base de données de signaux acquis d'un accéléromètre et d'un gyroscope embarqués dans un smartphone, nous avons développé et mis en place un système intelligent pour la reconnaissance de l'action. Nous nous sommes tout d'abord intéressés à la construction d'un vecteur d'attributs pertinent et optimal suivant le problème de classification rencontré. Notre algorithme de sélection d'attributs est exécuté au niveau de chaque nœud interne de l'approche de classification nous permettant ainsi d'obtenir des performances supérieures aux différents travaux existants dans la littérature. Cette démarche a permis de classer trois catégories d'actions intimement corrélées avec l'autonomie et le bien-être : les actions sédentaires, les actions périodiques ou pseudo-périodiques et les actions non périodiques. Ce système reconnaît aussi six transitions posturales négligées dans la littérature mais importantes pour l'autonomie et le bien-être. Notre approche nous garantit la robustesse à l'emplacement des capteurs et permet de réduire considérablement le temps de calcul nécessaire pour reconnaître l'action.

À la base des actions effectuées par la personne durant la journée, un indicateur d'autonomie pourra être établi. Pour maintenir cette autonomie et pallier au risque de la perdre, il est important de s'entretenir physiquement. Dans ce contexte, nous proposons un second système intelligent de reconnaissance de l'action humaine basé sur des données squelette acquis à partir d'une caméra Kinect. Un nouvel algorithme d'ex-

traction d'attributs en temps réel appelé BDV (Body-part Directional Velocity) a été proposé. Dans ce cas le système intelligent de reconnaissance de l'action est basé sur les modèles de Markov cachés (HMMs) avec des distributions de sortie d'état représentées par des modèles de mélange gaussiens (GMMs). Les résultats expérimentaux sur deux jeux de données publiques ont démontré l'efficacité de notre approche et sa supériorité par rapport aux approches de l'état de l'art. L'invariance et la robustesse de l'approche à l'orientation ont été aussi abordées en positionnant ainsi notre technique parmi les meilleures approches testées sur deux jeux de données présentant ce challenge. La reconnaissance anticipée de l'action par notre système a aussi été considérée en montrant que la moitié des actions étaient prévisibles presque au milieu de l'ensemble de la séquence de données squelette, et que certaines classes ont été reconnues avec seulement 4% de l'ensemble des données de la séquence.

Remerciements

Je voudrais tout d'abord exprimer ma reconnaissance à la région Hauts-de-France d'avoir soutenu ce projet de thèse et d'avoir largement contribué à la finalisation de mes travaux de recherche.

Ma profonde gratitude va à mes deux directeurs, Professeur Stéphane LECOEUICHE et docteur Anthony FLEURY, qui m'ont fait confiance et m'ont donné la chance de réaliser cette thèse avec eux. Merci Professeur Stéphane pour votre soutien, vos conseils et votre disponibilité. Merci docteur Anthony de m'avoir intégré à votre équipe. Grâce à votre grande expérience dans la recherche, votre disponibilité, vos suggestions et vos conseils toujours pertinents j'ai pu venir à bout de cette thèse. Je profite de ces lignes pour vous témoigner mon respect.

Mes vifs remerciements vont aux membres du Jury. Je suis très honorée que mon travail soit jugé par un jury de qualité dont les membres sont les éminents : docteur Sylvie CHARBONNIER, docteur Laurent NAVARRO, Professeur Emmanuel CHAZARD, Professeur Eric CAMPO. Je vous remercie Madame et Messieurs les professeurs pour la spontanéité et l'amabilité avec lesquelles vous avez répondu et accepté d'examiner cette thèse.

Je n'oublierai pas de remercier tous les chercheurs, stagiaires, étudiants, et collègues de l'IMT Lille Douai qui ont contribué, de près ou de loin, à la finalisation de cette thèse. Ces quelques années ont été une expérience enrichissante et n'auraient pas été les mêmes sans votre présence.

Enfin, j'adresse mes plus sincères remerciements à ma famille : Mes parents pour leur persévérance, mes sœurs tous mes proches et amis, qui m'ont accompagné, aidé, soutenu et encouragé tout au long de la réalisation de cette thèse.

Publications de l'auteur

Revue internationale

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*, vol. 10, no. 4, pp. 894–902, 2018

Conférences internationales

S. A. W. Talha, A. Fleury, and S. Ambellouis, "Human action recognition from body-part directional velocity using hidden markov models," in 16th IEEE International Conference On Machine Learning And Applications (ICMLA). IEEE, 2017, pp. 1035–1040

S. A. W. Talha, A. Fleury, and S. Lecoeuche, "Hierarchical classification scheme for real-time recognition of physical activities and postural transitions using smartphone inertial sensors," in 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019

S. A. W. Talha, A. Fleury, and S. Lecoeuche, "Reconnaissance en temps réel d'activités physiques et de transitions posturales à l'aide de capteurs inertiels pour smartphone," in XXVIIème Colloque francophone de traitement du signal et des images (GRETSI)

Table des matières

I	Introduction	23
1	Introduction	25
1.1	Contexte sociétal	25
1.1.1	Vieillessement de la population	25
1.2	Dépendance et perte d'autonomie	27
1.3	Calcul de l'autonomie	29
1.3.1	L'échelle des activités de la vie quotidienne de Katz	29
1.3.2	L'échelle des activités instrumentales de la vie quotidienne	30
1.3.3	La grille AGGIR	30
2	Reconnaissance du comportement humain	33
2.1	Introduction	35
2.2	Problématique de la reconnaissance du comportement humain	35
2.3	Solutions existantes pour la reconnaissance des mouvements humains	37
2.3.1	Reconnaissance de mouvements dans un environnement intérieur	38
2.3.2	Reconnaissance de mouvements dans un environnement extérieur	43
2.4	Prévention de la perte d'autonomie	47
2.4.1	Nintendo Wii	47
2.4.2	Caméra EyeToy de Sony	50
2.4.3	Caméra Kinect de Microsoft	52
2.5	Objectifs des travaux de thèse	54
2.5.1	Reconnaissance des actions humaines et suivi sur le long terme	55
2.5.2	Prévention de la perte d'autonomie	56

2.5.3	Actions considérées dans notre travail	56
2.6	Contributions	59
2.6.1	Système de suivi des actions humaines	60
2.6.2	Système de prévention de la perte d'autonomie	61
2.7	Organisation	61
2.8	Conclusion	62
 II Reconnaissance des actions humaines à l'aide de capteurs inertiels		63
 3 État de l'art : Choisir les paramètres et analyser le signal		65
3.1	Introduction	66
3.2	Système HAR à l'aide de capteurs inertiels	66
3.3	Collecte de données	67
3.4	Emplacement du smartphone	68
3.5	Segmentation	70
3.6	Pré-traitement	71
3.7	Extraction des attributs	72
3.8	Réduction de la dimensionnalité des attributs	76
3.9	Classification des actions humaines	77
3.10	Classification des transitions posturales	80
3.11	Évaluation des performances	80
3.12	Conclusion	82
 4 Classification hiérarchique des actions humaines à l'aide d'un smartphone		85
4.1	Introduction	86
4.2	Conception de notre système HAR	86
4.2.1	Construction du vecteur d'attributs	89
4.2.2	Classification hiérarchique des actions humaines	98
4.2.3	Classification des transitions posturales	100

4.3	Validation expérimentale	103
4.3.1	Jeux de données	105
4.3.2	Estimation des performances	109
4.3.3	Résultats et comparaison avec les travaux de l'état de l'art . .	112
4.3.4	Temps de calcul	136
4.4	Suivi d'actions sur le long-terme	136
4.4.1	Résultats expérimentaux	140
4.5	Conclusion	143
 III Reconnaissance anticipée de l'action à l'aide de la Kinect		145
5	État de l'art	147
5.1	Introduction	148
5.2	Système HAR à l'aide de la Kinect	148
5.2.1	Approches basées sur les données profondeur	148
5.2.2	Approches basées sur les données squelette	152
5.2.3	Approches hybrides	155
5.2.4	Reconnaissance d'actions avec l'apprentissage profond	157
5.2.5	Reconnaissance anticipée de l'action	160
5.3	Conclusion	162
6	Reconnaissance anticipée des actions humaines	163
6.1	Introduction	164
6.2	Modèles de Markov cachés et les attributs Body-Part Directional Velocity	164
6.2.1	Body-part Directional Velocity (BDV)	165
6.2.2	Classification à l'aide d'un modèle GMM-HMM	167
6.3	Validation expérimentale	169
6.3.1	Jeux de données	170
6.3.2	Protocole expérimental	172
6.3.3	Expérimentation sur des jeux de données publiques	175

6.3.4	Reconnaissance anticipée de l'action humaine	178
6.3.5	Temps de calcul	182
6.4	Conclusion	183
IV	Conclusion et perspectives	185
7	Conclusion et perspectives	187
7.1	Conclusion générale	188
7.2	Perspectives	189
7.2.1	Proposition d'indicateurs d'autonomie	190
7.2.2	Transfert d'indicateurs d'autonomie aux professionnels de la santé	191
7.2.3	Conception d'un programme d'exercices suivant l'autonomie de la personne	191

Table des figures

1-1	Croissance du nombre de personnes âgées.	26
1-2	Population française.	26
1-3	Les personnes âgées ayant soit une maladie chronique, soit une limitation fonctionnelle, soit des difficultés pour se laver.	28
2-1	Taxonomie des mouvements humains utilisée dans notre travail. . . .	37
2-2	Architecture générique pour la reconnaissance du mouvement humain avec des capteurs portables.	43
2-3	Exemple de signaux inertiels d'une personne lors de la marche.	44
2-4	Dispositif contenant un accéléromètre de la marque Xsens (voir (XSSENS 2010) pour la documentation).	45
2-5	La Wii.	49
2-6	La caméra EyeToy.	51
2-7	La caméra Kinect (version 1).	53
2-8	Reconnaissance de l'action humaine à l'aide d'un smartphone.	55
2-9	Approche générale proposée dans notre travail.	57
2-10	Les catégories d'actions prises en considération dans notre travail. . .	58
3-1	Schéma fonctionnel d'un système HAR.	66
3-2	Signaux d'accélération issus d'un smartphone porté à différents emplacements du corps.	68
3-3	Visualisation de LOSO Cross-Validation vs k-fold cross-validation. . .	81
3-4	Exemples des performances obtenues à l'aide d'une validation dépendante et indépendante du sujet.	82

4-1	Aperçu de l’approche proposée pour la reconnaissance d’actions à l’aide de capteurs inertiels.	88
4-2	Méthode d’extraction des attributs suivie dans notre approche.	90
4-3	Moyennes des signaux d’accélération (A_x, A_y, A_z) pour les postures “assis” et “debout”.	96
4-4	Évolution du score <i>rappel</i> en fonction de Fr pour les actions périodiques.	97
4-5	Coefficients sélectionnés (en bleu) dans la bande fréquentielle $[0Hz - 5Hz]$ pour l’action “marcher”.	98
4-6	Exemple de correction des erreurs de classification.	100
4-7	Transitions posturales considérées dans notre travail.	101
4-8	Application de l’algorithme 4 afin d’estimer la transition posturale “debout vers assis”.	103
4-9	Scénario des actions réalisées dans le jeu de données SBHART.	107
4-10	Aperçu des positions des smartphones sur un participant dans le jeu de données “Multi-position” (SHOAIB, BOSCH, O. D. INCEL et al. 2014).	109
4-11	Score obtenu en fonction du degré de chevauchement avec un test 10-fold-subject-cross-validation.	117
4-12	Détection des transitions dans le jeu de données SBHART.	121
4-13	Manière de porter le smartphone en Bandoulière.	137
4-14	Durée de l’expérimentation.	137
4-15	Approche proposée pour la reconnaissance d’actions en mode non supervisé.	139
4-16	Résultats obtenus au premier niveau de classification pour 5 minutes de données.	140
4-17	Sortie du système pour une heure de données.	141
4-18	Actions reconnues pour une journée de données.	142
4-19	Actions reconnues pour un mois de données.	142
5-1	Le calcul des attributs dans (W. LI et al. 2010) à partir de des images de profondeur.	149

5-2	Le calcul des attributs dans (X. YANG, C. ZHANG et al. 2012) à partir des images de profondeur.	150
5-3	Le calcul de l'information DSTIPs dans (XIA et J. AGGARWAL 2013) à partir des images de profondeur.	151
5-4	Le calcul des attributs dans (OREIFEJ et al. 2013) à partir des images de profondeur.	152
5-5	Les données squelette estimées selon la version de la Kinect et du kit de développement utilisé.	153
5-6	Représentation d'une action (séquence de squelettes) sous forme de courbe dans le groupe de Lie.	154
5-7	Aperçu de l'approche proposée par DEVANNE et al. (2015).	155
5-8	Aperçu de l'approche proposée par OHN-BAR et al. (2013).	156
5-9	Aperçu de l'approche proposée par DU, FU et al. (2015).	158
5-10	Aperçu de l'approche proposée par DU, W. WANG et al. (2015).	159
5-11	Aperçu de l'approche Integral bag-of-words proposée par RYOO (2011).	160
5-12	Reconnaissance anticipée de l'action en utilisant des approches basées sur l'apprentissage profond.	161
6-1	Aperçu de la méthode proposée pour la reconnaissance d'actions à l'aide de la Kinect.	165
6-2	Corps humain (squelette) divisé en cinq parties différentes (B_1, B_2, B_3, B_4 et B_5).	166
6-3	Architecture du modèle de classification.	168
6-4	Aperçu de deux actions "dessiner une coche" et "service de tennis" issues du jeu de données MSRAction3D.	171
6-5	Action "regarder la montre" issue du jeu de données Florence 3D.	172
6-6	Aperçu de quelques images du jeu de données UTKinect-Action.	173
6-7	Aperçu du jeu de données Multiview3D.	174
6-8	Évolution de la vraisemblance lors de l'exécution d'actions de coups de pied sur le côté et marcher.	179

6-9	Boîtes à moustaches montrant le pourcentage des données squelette $([p_1, \dots, p_k])$ mises à l'entrée de notre système pour reconnaître les actions dans les trois sous-ensembles AS1, AS2, AS3 du jeu de données MSRAAction3D (de gauche à droite).	180
7-1	Travaux réalisés en thèse accompagnés des perspectives.	190

Liste des tableaux

1.1	Les activités instrumentales de la vie quotidienne.	32
3.1	Les attributs calculés dans l'état de l'art à partir du domaine temporel des signaux inertiels.	73
3.2	Les attributs calculés dans l'état de l'art à partir du domaine fréquentiel des signaux inertiels.	75
3.3	Les attributs calculés dans l'état de l'art à partir du domaine temps-fréquence des signaux inertiels.	76
3.4	Algorithmes de classification utilisés pour la reconnaissance de l'action humaine.	78
4.1	Attributs temporels calculés afin de construire le vecteur F_T	92
4.2	Vecteurs candidats pour chaque problème de classification.	96
4.3	Les actions prise en compte dans notre étude.	104
4.4	Les positions du capteur prise en compte dans notre étude.	105
4.5	Les jeux de données publiques utilisés dans la phase expérimentale.	110
4.6	Scores obtenus en comparaison avec l'état de l'art pour le jeu de données MobiAct.	113
4.7	Matrice de confusion (%) pour le jeu de données MobiAct.	114
4.8	Scores pour chaque classifieur et dimensions des attributs pour le jeu de données MobiAct (ici et dans la suite "Catégorie" signifie la séparation en méta-classes).	115
4.9	Scores obtenus en comparaison avec l'état de l'art pour le jeu de données Siirtola.	115

4.10	Matrice de confusion (%) pour le jeu de données Siirtola.	115
4.11	Scores pour chaque classifieur et dimensions des attributs pour le jeu de données Siirtola.	116
4.12	Scores obtenus en comparaison avec l'état de l'art pour le jeu de données WISDM.	117
4.13	Scores pour chaque classifieur et dimensions des attributs pour le jeu de données WISDM.	118
4.14	Scores obtenus en comparaison avec l'état de l'art pour le jeu de données SBHART.	120
4.15	Matrice de confusion (%) pour le jeu de données SBHART.	120
4.16	Scores de reconnaissance des transitions posturales sur le jeu de données SBHART.	121
4.17	Matrice de confusion des transitions posturales sur le jeu de données SBHART.	122
4.18	Scores pour chaque classifieur et dimensions des attributs pour le jeu de données SBHART.	123
4.19	Scores obtenus en comparaison avec l'état de l'art pour le jeu de données UniMiB SHAR.	124
4.20	Matrice de confusion (%) pour le jeu de données UniMiB SHAR.	124
4.21	Scores pour chaque classifieur et dimensions des attributs pour le jeu de données UniMiB SHAR.	125
4.22	Scores obtenus en comparaison avec l'état de l'art pour le jeu de données mHealth.	127
4.23	Scores obtenus pour le jeu de données mHealth.	129
4.24	Dimension des attributs pour le jeu de données mHealth.	131
4.25	Scores obtenus en comparaison avec l'état de l'art pour le jeu de données Multi-position.	132
4.26	Scores obtenus pour le jeu de données Multi-position.	134
4.27	Dimension des attributs pour le jeu de données Multi-position.	135

4.28	Heures déclaratif de réalisation de certaines activités de la vie quotidienne par le sujet 4.	138
6.1	Trois sous-ensembles d'actions construits dans le jeu de données MSRAction3D : AS1, AS2, AS3.	174
6.2	Précision de différentes méthodes sur le jeu de données MSRAction3D.	175
6.3	Précision de différentes méthodes sur le jeu de données Florence3D. . .	176
6.4	Précision de différentes méthodes sur le jeu de données UTKinect-Action.	176
6.5	Matrice de confusion (%) sur le jeu de données UTKinect-Action. . .	177
6.6	Précision (%) sur le jeu de données Multiview3D.	177
6.7	Temps de calcul dans chaque étape du système.	183

Liste des abréviations

ADL Activities of Daily Living

AGGIR Autonomie Gérontologique-Groupes Iso-Ressources

ANN Artificial Neural Network

AVC Accident Vasculaire Cérébral

BDV Body-part Directional Velocity

CNN Convolutional Neural Network

D-HMM Discrete Hidden Markov Model

DBN Deep Belief Network

DBSCAN Density-based spatial clustering of applications with noise

DFT Discrete Fourier transform

DMM Depth Motion Maps

DNN Deep Neural Network

DREES Direction de la recherche, des études, de l'évaluation et des statistiques

DSTIPs Depth Spatio-Temporal Interset Points

DTW Dynamic Time Warping

DWT Discrete Wavelet Transform

ESD Energy Spectral Density

FFT Fast Fourier transform

FTP Fourier temporal pyramid

GMM Gaussian Mixture Models

GPS Global Positioning System

H-HMM Hierarchical Hidden Markov Model

HAR Human Action Recognition

HMM Hidden Markov Model

HOF Histogram of Flow

HOG Histogram of Oriented Gradients

IMU Inertial Measurement Unit

IQR InterQuartile Range

kNN k Nearest Neighbor

LOSOVC Leave-One-Subject-Out-Cross-Validation

LSTM Long Short-Term Memory

MEMS Microelectromechanical systems

MET Metabolic Equivalent of Task

OMS Organisation Mondiale de la Santé

PCA Principal Components Analysis

RBF Radial Basis Function

RF Random Forest

RNN Recurrent Neural Network

SBS Sequential Backward Selection

SFFS Sequential Forward Floating Selection

SFS Sequential Forward Selection

STIPs Spatio Temporal Interest Points

SVM Support Vector Machine

TFD Transformée de Fourier Discrète

URIA Unité de Recherche en Informatique et Automatique

WPD Wavelet Packet Decomposition

Première partie

Introduction

Chapitre 1

Introduction

Contents

1.1	Contexte sociétal	25
1.1.1	Vieillessement de la population	25
1.2	Dépendance et perte d'autonomie	27
1.3	Calcul de l'autonomie	29
1.3.1	L'échelle des activités de la vie quotidienne de Katz	29
1.3.2	L'échelle des activités instrumentales de la vie quotidienne .	30
1.3.3	La grille AGGIR	30

1.1 Contexte sociétal

1.1.1 Vieillessement de la population

Selon les résultats d'un rapport réalisé par les Nations Unies, le nombre de personnes âgées va croître rapidement (MCNICOLL 2002). La figure 1-1a montre que le nombre de personnes de plus de soixante ans va plus que tripler d'ici 2050. Nous pouvons voir sur la figure 1-1b que les régions les plus concernées par cette hausse sont l'Europe et l'Amérique du Nord.

Au premier janvier 2019, la France comptait 66,993 millions d'habitants (MARTIAL et al. 2019). Les personnes âgées d'au moins 65 ans représentent 20% de la population.

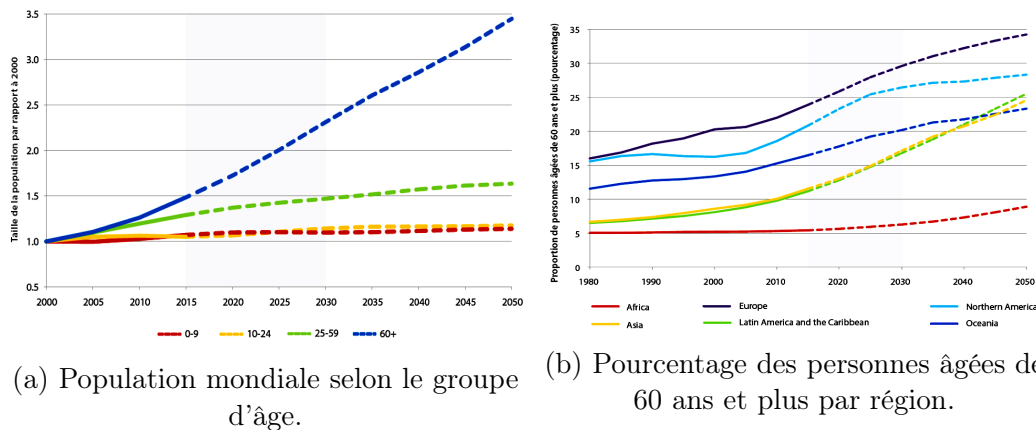


FIGURE 1-1 – Croissance du nombre de personnes âgées.

Cette tranche d'âge a augmenté en permanence depuis 1870. La figure 1-2a montre la pyramide des âges en 1870, 1920, 1970, 2018 et 2070. Nous pouvons voir que depuis 1870, la pyramide s'élargit au-delà du seuil de 65 ans. En effet, le nombre de seniors¹ n'arrête pas d'accroître depuis 1870 : la France comptait 3,5 millions de seniors en 1920, 6,5 millions en 1970 et 13,1 millions en 2018 (DEMAISON et al. 2018).

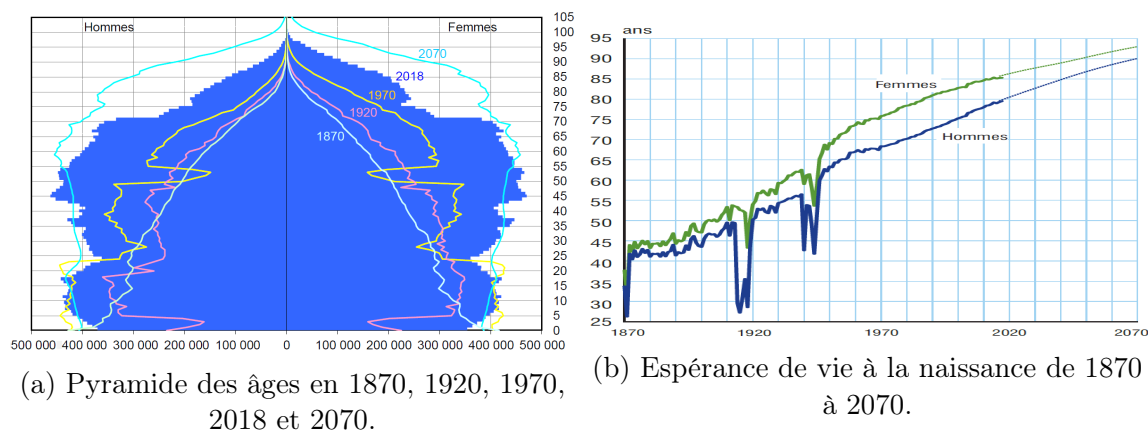


FIGURE 1-2 – Population française.

Cette progression des seniors devrait se poursuivre en raison de la hausse de l'espérance de vie à la naissance². La figure 1-2b montre que l'espérance de vie a quasiment doublé entre 1890 et 2017 passant de 44 à 85 ans pour les femmes et de

1. Personnes de 65 ans et plus pour les pays développés et 60 ans et plus pour les pays sous-développés selon l'OMS (KOWAL et al. 2001).

2. La durée de vie moyenne d'une génération fictive soumise aux conditions de mortalité de l'année.

41 à 79 ans pour les hommes, soit un gain moyen d'environ trois ans tous les 10 ans. Selon le scénario central des projections de population³, l'espérance de vie à la naissance pourrait atteindre 93 ans pour les femmes et 90 ans pour les hommes en 2070.

1.2 Dépendance et perte d'autonomie

Le vieillissement de la population implique une forte prévalence des maladies liées à l'âge. En effet, une population vieillissante est affectée par des maladies chroniques, des limitations fonctionnelles et des restrictions d'activités qui nécessitent des soins et un accompagnement de longue durée. Selon l'Organisation Mondiale de la Santé (OMS), une maladie chronique correspond à une maladie de longue durée et évolutive. Il peut s'agir de maladies cardio-vasculaires, de cancers, du diabète, ou de maladies neuro-dégénératives (Alzheimer, Parkinson), etc.

Les limitations fonctionnelles comprennent les difficultés à accomplir des actes mobilisant les fonctions physiques ou cognitives. Nous distinguons trois types de limitations :

- Limitations physiques : marcher sur 500 mètres, monter et descendre un escalier, lever le bras, se pencher et ramasser un objet.
- Limitations sensorielles : des difficultés à voir clairement le visage de quelqu'un à 4 mètres, voir clairement les caractères d'imprimerie d'un journal, des difficultés à entendre une conversation avec plusieurs personnes.
- Limitations cognitives : se concentrer plus de 10 minutes, se souvenir des choses importantes, résoudre les problèmes de la vie quotidienne.

Les restrictions d'activités incluent les problèmes qui peuvent se poser à un individu dans la réalisation des actes de la vie quotidienne. Il peut s'agir de faire sa toilette, se nourrir, s'habiller, se coucher/se lever, aller aux toilettes. Elles représentent

3. Que les tendances démographiques récentes se poursuivent et la mortalité continue à baisser au même rythme que celui observé au cours des vingt dernières années

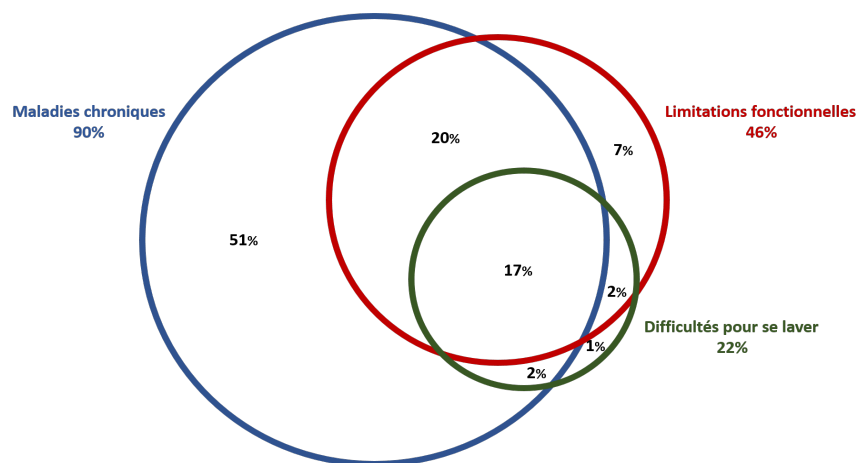


FIGURE 1-3 – Les personnes âgées ayant soit une maladie chronique, soit une limitation fonctionnelle, soit des difficultés pour se laver.

les formes les plus sévères d’incapacité chez les personnes âgées, qui nécessitent des soins et un accompagnement de longue durée.

Selon l’enquête Vie Quotidienne et Santé (VQS) du Ministère des Solidarités et de la Santé (BRUNEL et al. 2017), en 2014, en France, parmi les personnes de 60 ans et plus, vivant à domicile, 8 millions ont déclaré avoir une maladie chronique (51,3% des personnes âgées), 4,1 millions ont au moins un type de limitation fonctionnelle (26,3%) et 2,6 millions ont au moins une restriction d’activité (16,67%). Ces maladies liées à l’âge ne sont pas indépendantes entre elles. En effet, parmi les 8,9 millions de personnes âgées ayant soit une maladie chronique, soit une limitation fonctionnelle, soit des restrictions d’activité dans le domaine de la toilette, 17% cumulent les trois, comme le montre la figure 1-3. Nous pouvons voir aussi que très peu de personnes ont des difficultés à se laver sans limitations fonctionnelles ou maladies chroniques.

Ces maladies peuvent provoquer la perte d’autonomie chez les personnes âgées, augmentant ainsi le nombre d’entrées en institution. La perte d’autonomie se définit comme l’incapacité d’assurer seul certaines actions de la vie courante. En 2015, la France (hors Mayotte) comptait près de 2,5 millions de seniors en perte d’autonomie (15,3% des 60 ans ou plus) (LARBI et al. 2019), et si les tendances démographiques se poursuivent, 4 millions de seniors seraient touchés par la perte d’autonomie en 2050.

Pour la même année, la région des Hauts-de-France comptait 1,33 millions de

personnes de 60 ans et plus (BELHAKEM et al. 2018), 96% d’entre eux vivent à domicile. Cependant, plus de 30% de ces seniors à domicile déclarent une limitation fonctionnelle, contre 26% au niveau national. Les limitations fonctionnelles fréquemment déclarées sont de type physique, ce qui peut provoquer des situations de perte d’autonomie.

Pour maintenir constant le pourcentage de personnes en établissement par département, sexe, tranche d’âge et degré de perte d’autonomie, il faudrait que le nombre de places en hébergement permanent en établissements pour personnes âgées augmente de 20% d’ici à 2030 et de plus de 50% à l’horizon 2050 (LARBI et al. 2019).

Il est donc plus que nécessaire de promouvoir le maintien à domicile pour réduire ce nombre d’entrées (ou le retarder) et ainsi diminuer l’impact sur notre système de santé.

1.3 Calcul de l’autonomie

Afin d’évaluer la dépendance d’une personne, plusieurs échelles ont été proposées, permettant ainsi d’évaluer différentes facettes de la vie de la personne et utilisées dans différents contextes. Les plus utilisées dans la littérature sont détaillées ci-dessous.

1.3.1 L’échelle des activités de la vie quotidienne de Katz

L’indice de Katz est considéré comme une référence pour les professionnels de la santé (KATZ et al. 1970). Il a pour objectif de déterminer le niveau de dépendance d’une personne à travers l’évaluation de sa capacité à effectuer les activités élémentaires de la vie quotidienne. Six activités sont alors considérées : soins corporels, s’habiller, aller aux toilettes et les utiliser, se mettre au lit et se lever du lit et de la chaise sans aide (transitions posturales), continence et se nourrir sans aide. Pour calculer l’indice, un score “0” ou “1” est attribué au patient suivant la capacité ou non à exécuter chacune des six activités citées. Une personne âgée dont le score est inférieur à “3” est considérée comme dépendante. Les cliniciens utilisent souvent cet outil pour détecter les problèmes liés aux activités de la vie quotidienne et planifier les

soins en conséquence. À noter que plusieurs variantes ont été proposées par la suite (différents nombres d'activités ou scores attribués), certaines versions sont construites pour répondre à des besoins particuliers.

1.3.2 L'échelle des activités instrumentales de la vie quotidienne

L'indice de Katz permet d'évaluer les activités de base de la vie quotidienne. Toutefois, prendre en compte des activités impliquant les fonctions cognitives peut s'avérer pertinent pour évaluer l'autonomie d'une personne. Dans cette optique, LAWTON et al. (1969) ont proposé un outil pour évaluer la capacité d'une personne à exécuter des activités dites instrumentales, considérées plus complexes que celles proposées par Katz. Pour calculer l'indice de dépendance, huit activités sont considérées, le système de notation est le même que celui proposé par Katz, i.e., un score "0" ou "1" est attribué au patient suivant sa capacité ou non à exécuter les huit classes d'activités définies dans le tableau 1.1. Historiquement, les hommes étaient évalués sur cinq des classes d'activités proposées (les classes faire la cuisine, le ménage et le linge étaient exclues), par conséquent, un score global variant entre "0" (très dépendant) et "8" est attribué aux femmes et entre "0" et "5" pour les hommes.

1.3.3 La grille AGGIR

La grille Autonomie Gérontologique-Groupes Iso-Ressources (AGGIR) est l'outil actuellement utilisé en France pour mesurer le degré de perte d'autonomie des personnes âgées. Elle a été élaborée par des médecins de la Sécurité sociale et de la Société française de Gérontologie ainsi que par des informaticiens. Ils ont analysé les activités communs à 450 échelles existantes et repéré les plus pertinents. Ainsi, 10 critères, appelés "variables discriminantes" ont été sélectionnés pour déterminer le niveau de dépendance appelé GIR (Groupes Iso Ressources) : cohérence, orientation, toilette, habillage, alimentation, élimination, transferts (se lever, se coucher, s'asseoir), déplacement à l'intérieur, déplacement à l'extérieur et communication à

distance. À noter que 7 autres critères appelés “variables illustratives” font partie de la grille AGGIR mais n’entrent pas dans le calcul du GIR par contre elles sont exploitées dans l’élaboration du plan d’aide. Chacune des variables considérées a 3 modalités : (A) La personne fait seule et spontanément, habituellement, totalement et correctement ; (B) la personne peut faire seule, mais ne le fait pas spontanément ou/et correctement ou/et habituellement ou/et partiellement ; (C) la personne ne fait pas : il faut faire à sa place ou faire par quelqu’un d’autre.

À partir des critères cités, six groupes iso-ressources ont été définis :

(1) perte d’autonomie mentale, corporelle, locomotrice et sociale ; (2) fonctions mentales partiellement altérées ; (3) autonomie mentale mais besoin d’aide pour les soins corporels ; (4) autonomie mentale et capacités locomotrices mais nécessite une aide pour certaines tâches quotidiennes comme l’habillage et la toilette ; (5) autonomie mentale totale, besoin d’aide ponctuelle pour la toilette, la préparation des repas et le ménage ; (6) réalisation de tous les actes de la vie courante.

Ces grilles sont aussi la base des travaux qui sont présentés par la suite car les prendre en compte permet de savoir quelles activités sont d’intérêt à surveiller. Le chapitre suivant présentera les bases de la reconnaissance du comportement humain qui est le cœur des travaux de cette thèse.

Activités	Description des niveaux de dépendance	Score
Téléphone	-Utilise le téléphone de sa propre initiative	1
	-Compose quelques numéros connus	1
	-Décroche mais ne compose pas seul	1
	-N'utilise pas le téléphone	0
Courses	-Achète seul la majorité des produits nécessaires	1
	-Fait peu de courses	0
	-Nécessite un accompagnement lors des courses	0
	-Incapable de faire ses courses	0
Cuisine	-Prévoit et cuisine les repas seuls	1
	-Cuit les repas après préparation par une tierce personne	0
	-Fait la cuisine mais ne tient pas compte des régimes imposés	0
	-Nécessite des repas préparés et servis	0
Ménage	-S'occupe du ménage de façon autonome	1
	-Fait seule des tâches ménagères légères	1
	-Fait des travaux légers mais de façon insuffisante	1
	-Nécessite de l'aide pour les travaux ménagers	1
	-Nécessite de l'aide pour les travaux ménagers quotidiens	0
Linge	-Lave tout son linge seul	1
	-Lave le petit linge	1
	-Tout le linge doit être lavé à l'extérieur	0
Transport	-Utilise les moyens de transport de manière autonome	1
	-Commande et utilise seul un taxi	1
	-Utilise les transports publics avec une personne accompagnante	0
	-Parcours limités en voiture, en étant accompagné	0
	-Ne voyage pas	0
Médicaments	-Prend ses médicaments correctement et de façon responsable	1
	-Prend correctement les médicaments préparés	0
	-Ne peut pas prendre les médicaments correctement	0
Argent	-Règle ses affaires financières de façon autonome	1
	-Règle ses dépenses quotidiennes, aide pour les virements et dépôts	1
	-N'est plus capable de se servir de l'argent	0

TABLE 1.1 – Les activités instrumentales de la vie quotidienne.

Chapitre 2

Reconnaissance du comportement humain

Contents

2.1	Introduction	35
2.2	Problématique de la reconnaissance du comportement humain	35
2.3	Solutions existantes pour la reconnaissance des mouve- ments humains	37
2.3.1	Reconnaissance de mouvements dans un environnement in- térieur	38
2.3.2	Reconnaissance de mouvements dans un environnement ex- térieur	43
2.4	Prévention de la perte d'autonomie	47
2.4.1	Nintendo Wii	47
2.4.2	Caméra EyeToy de Sony	50
2.4.3	Caméra Kinect de Microsoft	52
2.5	Objectifs des travaux de thèse	54
2.5.1	Reconnaissance des actions humaines et suivi sur le long terme	55
2.5.2	Prévention de la perte d'autonomie	56
2.5.3	Actions considérées dans notre travail	56

2.6	Contributions	59
2.6.1	Système de suivi des actions humaines	60
2.6.2	Système de prévention de la perte d'autonomie	61
2.7	Organisation	61
2.8	Conclusion	62

2.1 Introduction

Le travail présenté dans cette thèse vise deux objectifs importants et complémentaires. Le premier est la surveillance du mouvement humain au quotidien et le second la proposition de mesures préventives afin de réduire la perte d'autonomie. Dans ce chapitre nous allons définir la problématique de la reconnaissance du mouvement humain et en présenter l'état de l'art, s'attarder sur les solutions existantes pour le reconnaître et identifier comment prévenir la perte d'autonomie.

2.2 Problématique de la reconnaissance du comportement humain

Dans les travaux de l'état de l'art pour la problématique de reconnaissance d'action, différents vocabulaires et terminologies ont été utilisés.

NAGEL (1988) a catégorisé le comportement en cinq niveaux : changement, événement, verbe, épisode et historique. Un changement est toute déviation dans un signal sensoriel qui diffère significativement du bruit. Un événement est tout changement qui a été défini a priori comme une primitive pour la construction de descriptions plus complexes. Un verbe décrit une activité ou l'absence explicite d'activité (se reposer par exemple). Une action est supposée être l'activité la plus élémentaire. Un épisode est une combinaison d'actions. Un historique est une série d'événements liés les uns aux autres par des relations de cause à effet.

Dans les travaux de BOBICK (1997), le comportement est classé en trois catégories : mouvement, activité et action. Afin de montrer la différence entre ces trois classes, un exemple autour du sport de Baseball est donné : la reconnaissance de trois différents comportements à savoir balancer la batte, lancer la balle et marquer un coureur. Balancer la batte consiste à déplacer la batte à une vitesse suffisante pour propulser la balle. Ce comportement est classé comme mouvement, car l'exécution est cohérente et facilement caractérisée par une trajectoire spatio-temporelle définie dans un espace de configuration (dans ce cas, la cinématique du corps humain). Lancer la balle implique

beaucoup plus d'étapes que de frapper : généralement, ce comportement consiste à (1) rapprocher les bras devant le corps pour atteindre l'équilibre sur les deux pieds (2) balancer les bras en arrière (3) donner un coup de pied dans la jambe avant tout en se penchant en arrière (4) livrer la balle. Le comportement est classé comme activité (séquence de mouvements). La reconnaissance d'un tel comportement nécessite de connaître à la fois l'apparition de chaque mouvement et les propriétés statistiques de la séquence temporelle. Pour marquer un coureur, un joueur avec la balle fait entrer son gant en contact avec un coureur qui ne touche pas la balle à temps. Ce comportement est classé comme action, car il inclut les primitives sémantiques relatives au contexte du mouvement.

MOESLUND et al. (2006) a utilisé la hiérarchie suivante pour définir un comportement : primitives d'action/moteur, actions et activités. La primitive action/moteur est un mouvement qui peut être décrit au niveau d'un membre, tel que le déplacement d'une jambe. Une action est une séquence de primitives d'action/moteur pouvant décrire un mouvement du corps entier, éventuellement cyclique, tel que la course. L'activité consiste en un ensemble d'actions qui donnent une interprétation de ce qui est exécuté (ce qui peut être décrit comme une compréhension de la situation), telle que jouer au football. Un exemple est donné sur le sport de tennis afin de montrer la différence entre les trois catégories de comportements citées. Dans ce sport, les mouvements "coup droit", "revers", "courir à gauche" et "courir à droite" pourraient être considérés comme primitives d'action. Le terme action est utilisé pour la séquence de primitives nécessaires pour retourner la balle. Une action particulière dépendra donc des primitives d'action requises (coup droit, un revers, un lob, une volée, etc.) pour pouvoir renvoyer la balle. Les activités sont des événements à plus grande échelle, dans cet exemple c'est "jouer au tennis".

Dans les travaux de la littérature pour la reconnaissance de mouvements avec des capteurs portables (accéléromètre par exemple), différents vocabulaires ont été employés par les chercheurs. Dans certains travaux, les mouvements de déambulation (marcher, courir, monter et descendre les escaliers, etc.) et les mouvements sédentaires (debout, assis et allongé) ont été considérés comme des "activités simples". Un

ensemble de ces activités forment une “activité complexe” (SHOAIB, BOSCH, O. INCEL et al. 2016 ; SAGUNA et al. 2013 ; L. LIU, PENG et al. 2016). Nous pouvons citer comme exemple : regarder la télévision, faire le ménage, manger, etc. Dans d’autres travaux, les activités sont groupées en différents niveaux. Dans (Jindong WANG et al. 2019 ; K.-J. KIM et al. 2009 ; M. ZHANG et al. 2012), les catégories de mouvements citées précédemment (activités simples et activités complexes) sont cette fois-ci classées en activités de bas niveau et activités de haut niveau.

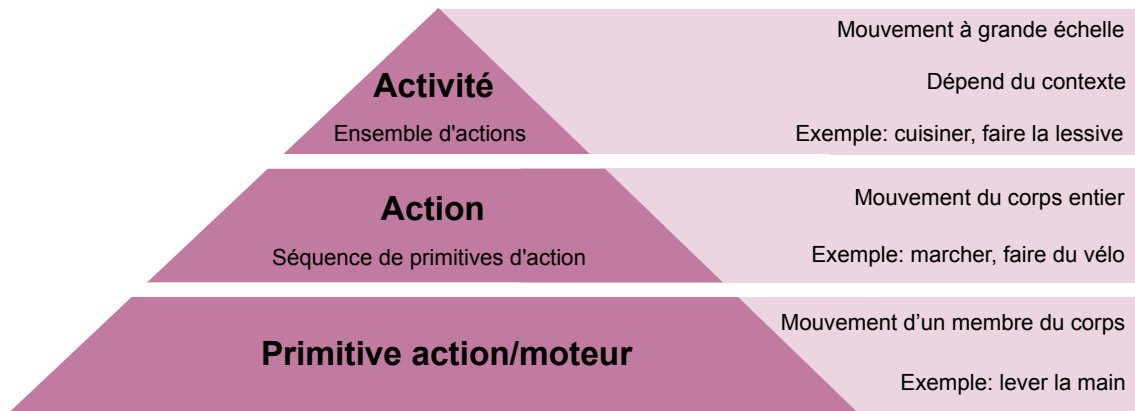


FIGURE 2-1 – Taxonomie des mouvements humains utilisée dans notre travail.

Dans notre travail, les mouvements humains seront décomposés suivant la taxonomie proposée par MOESLUND et al. (2006). Cette dernière est illustrée dans la figure 2-1. Les termes “activité”, “action” et “primitive action/moteur” seront donc utilisés afin de décrire un mouvement humain.

2.3 Solutions existantes pour la reconnaissance des mouvements humains

Afin de reconnaître les mouvements d’une personne, divers solutions existent utilisant différentes technologies de capteurs. Nous pouvons les regrouper en deux catégories suivant l’environnement visé : reconnaissance de mouvement dans un environnement intérieur, et sa reconnaissance dans un environnement extérieur.

2.3.1 Reconnaissance de mouvements dans un environnement intérieur

Les maisons intelligentes (smart homes en anglais) représentent une solution très utilisée pour la reconnaissance de mouvements humains ou d'activités humaines, plus particulièrement pour les personnes âgées. Ces maisons sont équipées de plusieurs capteurs afin de collecter des données sur le mouvement de la personne. Les principaux capteurs utilisés sont définis ci-dessous.

Caméras vidéo visibles

Les premiers systèmes de reconnaissance des mouvements humains utilisaient des caméras qui délivrent un flux d'images 2D couleur (*RGB*) ou en niveaux de gris. Des travaux qui regroupent les différentes approches proposées dans la littérature sont accessibles dans (POPPE 2010; Jake K AGGARWAL et al. 2011). Les caméras fournissent des images très riches en informations. Cependant, elles présentent certaines contraintes notamment sur les conditions d'éclairage qui font que les résultats obtenus au laboratoire dans les conditions contrôlées ne sont pas les mêmes que dans le cas pratique. Outre ceci, le plus grand obstacle des caméras vidéo visibles est qu'elles sont très intrusives dans la vie privée des utilisateurs. Certains travaux (RYOO et al. 2017; DAI et al. 2015; J. CHEN et al. 2017) utilisent des images de basse résolution afin de réduire l'impact de ce problème. Néanmoins, beaucoup de chercheurs se sont tournés vers d'autres types de capteurs moins intrusifs afin de préserver au maximum la vie privée des personnes.

Caméras de profondeur

Afin de pallier les limites observées dans les caméras vidéo visibles et de proposer des systèmes plus respectueux à l'intimité de la personne, d'autres chercheurs se sont intéressés à d'autres catégories de caméra. La plus populaire, est la caméra de profondeur. Celle-ci utilise le spectre infrarouge : un émetteur (projecteur) émet des faisceaux lasers infrarouges et un récepteur récupère l'information réfléchi par les

objets de la scène afin d'en calculer la profondeur.

Plusieurs chercheurs ont utilisé les caméras de profondeur pour la reconnaissance des mouvements humains. Dans (C. ZHANG et al. 2012), un système de reconnaissance de chute destiné aux personnes âgées est proposé. Cinq actions sont considérées : “debout”, “chute à partir de la position debout”, “chute à partir de la position assise”, “s’asseoir sur une chaise” et “s’asseoir sur le sol”. Les tests expérimentaux ont été menés suivant deux conditions : éclairage normal et sans éclairage. Les résultats montrent que les capteurs de profondeurs sont robustes aux conditions d'éclairage. Dans les travaux de (JALAL et al. 2014), un système de surveillance des activités des personnes âgées à l'intérieur d'un habitat est proposé. Chaque activité reconnue est stockée dans une base de données avec l'heure et la date. Pour la phase expérimentale, une caméra de profondeur a été fixée dans trois environnements différents : la maison, le bureau (espace de travail) et dans un hôpital. Six activités de la vie quotidienne sont considérées pour chaque environnement, comme, par exemple cuisiner, travailler sur un ordinateur ou s'allonger. Les résultats ont montré des performances satisfaisantes pour les trois environnements (des scores de classification supérieurs à 90%).

Capteurs audio

Les capteurs audio (microphones) sont parmi les capteurs les plus répandus. Ils permettent de capturer des informations à la fois sur une personne et sur son environnement ambiant, i.e., reconnaître ce que fait une personne (en conversation par exemple), où elle se trouve (signature audio d'une cafétéria en particulier). Dans les travaux de LU et al. (2009), un système appelé “SoundSense” a été proposé afin de modéliser les événements sonores. Il utilise une combinaison de techniques d'apprentissage supervisé et non supervisé pour classifier les types de son (musique ou voix par exemple) et découvrir de nouveaux événements sonores spécifiques à chaque utilisateur. La solution a été implémentée sur un Apple iPhone. Dans (FLEURY et al. 2008), huit microphones ont été installés dans une maison de santé intelligente à Grenoble. Les différents sons enregistrés sont traités et analysés afin de reconnaître différents types d'événements normaux tels que : frapper à la porte, appel téléphonique, faire

la vaisselle, etc., et anormaux, comme par exemple : bris de verre, chute d'un objet, crier. Dans (POPESCU et al. 2008), un ensemble de capteurs audio ont été exploités afin de détecter la chute d'une personne. Les auteurs se sont particulièrement intéressés à la réduction du taux de fausse alarme, ils exploitent pour cela la hauteur de la source sonore. Les résultats expérimentaux (sur un jeu de données très limités) ont montré un taux de détection de 100% avec cinq fausses alarmes.

Comme c'est le cas pour les caméras visibles, les signaux audio sont des données très sensibles, ils peuvent contenir des communications et discussions privées. Par conséquent, les capteurs audio sont considérés comme trop intrusifs pour notre travail.

Capteurs de vibration

Les capteurs de vibration sont généralement utilisés pour détecter la chute d'une personne. Pour ce faire, les signaux de vibration du sol sont exploités. Dans (ALWAN et al. 2006), un capteur piézoélectrique spécial couplé à la surface du sol est proposé. Ce dernier est associé à un système électronique alimenté par batterie afin d'évaluer les signaux de vibration du sol et générer une sortie de chute binaire. Un émetteur sans fil transmet par la suite l'alarme de la chute à travers une passerelle de communication. Dans les travaux de ZIGEL et al. (2009), l'objectif est de discriminer la chute des autres événements. Un système basé sur les signaux vibratoires et sonores est alors présenté. Un accéléromètre et un microphone sont fixés au sol par des rubans adhésifs. Les signaux collectés sont mis à l'entrée d'un dispositif d'acquisition de données portable qui échantillonne les signaux et les transmet par la suite au PC. D'autres chercheurs ont utilisé les capteurs de vibration pour détecter les flux d'eau utilisés par les personnes dans leur domicile. En effet, la consommation d'eau est corrélée avec certaines activités quotidiennes comme cuisiner ou faire la lessive par exemple. Dans (L. HU et al. 2013), un accéléromètre triaxial est accroché à la surface de la conduite d'eau principale de la maison. Le système proposé reçoit en entrée les signaux collectés afin de reconnaître quatre activités : prendre un bain, tirer la chasse d'eau, faire la lessive et cuisiner. Les résultats expérimentaux ont montré un score global de 70.37%.

Capteurs de pression

Afin de reconnaître des activités de la vie quotidienne, les chercheurs dans (LIM et al. 2008) ont installé des capteurs de pression sur les meubles et les sols de la maison. Les activités ciblées sont : manger, dormir, aller aux toilettes, sortir et repos (rester assis sur un canapé pendant un certain temps). Le système proposé par les auteurs a permis de repérer des comportements mauvais pour la santé comme le saut de repas par exemple. Dans (FOUBERT et al. 2012), trois tapis sensibles à la pression ont été placés côte à côte sur un cadre de lit d'hôpital et sous un matelas. L'objectif est de reconnaître trois actions : assis, allongé et transition posturale (allongé vers assis). Les tests expérimentaux ont été menés sur quatre catégories de personnes : jeunes en bonne santé, personnes âgées en bonne santé, personnes âgées en récupération post-AVC et personnes âgées en récupération d'une fracture de la hanche. Plusieurs algorithmes de classification ont été comparés. Les résultats ont montré des scores de précision élevés. Les capteurs de pression sont aussi exploités pour l'étude de la qualité du sommeil des personnes. Dans les travaux de ADAMI et al. (2006), afin de détecter un sommeil agité, les auteurs ont placé des capteurs de pression dans les coins du lit des sujets. Les mouvements effectués sont classés en deux catégories : mouvement faible ou élevé. La méthode proposée par les chercheurs a montré une faible erreur de classification (4%).

Capteurs infrarouges

Les capteurs infrarouges sont utilisés soit pour détecter la présence humaine dans une pièce (ou un mouvement dans une zone spécifique) soit pour localiser un humain dans une maison. YAMAGUCHI et al. (1998) se sont fixés pour objectif de surveiller le comportement humain au quotidien pour améliorer sa qualité de vie. Quatre capteurs infrarouges de type pyroélectrique¹ ont été installés dans la cuisine, la chambre, les toilettes et le bureau. Les résultats ont montré que le système proposé est capable d'estimer le temps passé dans différentes pièces et le nombre de passages par

1. Matériau dans lequel un changement de température entraîne une variation de polarisation électrique

les portes. Cependant, les mouvements humains ne sont pas reconnus dans le cas où l'air est chaud (lié aux propriétés des capteurs pyroélectriques). Le problème s'accroît quand il y a plus d'une personne dans la pièce (incapacité de distinguer les individus). CHAN et al. (2005) visaient à mettre en place un système de diagnostic des événements anormaux afin d'aider les personnes âgées vivant seules. 10 capteurs infrarouges étaient alors installés au plafond d'une résidence pour personnes âgées. Le système proposé était capable de détecter des événements comme se lever, aller aux toilettes et sortir. Dans les travaux de VIRONE et al. (2008), le rythme d'activité de 22 personnes âgées est suivi grâce à un ensemble de capteurs, principalement des capteurs infrarouges installés dans chaque pièce des appartements des participants. Une douzaine de modèles de comportements ont été définis. Les résultats présentés par les auteurs ont montré que le système proposé a permis de détecter des déviations sur ces comportements, qui pourraient être liées à un changement de l'état de santé de la personne. Les soignants sont donc avertis dans ce cas afin d'établir un diagnostic rapide.

Capteurs à ultrasons

Les capteurs à ultrasons sont généralement utilisés pour la détection et la localisation des personnes dans une maison. Dans (HELAL et al. 2003), un système de localisation d'une personne à l'intérieur d'un habitat a été proposé. Pour la phase expérimentale, deux récepteurs ultrasons ont été placés sur l'épaule droite et gauche d'un mannequin de laboratoire. Pour les émetteurs ultrasons, ils sont placés dans chaque coin de l'appartement. Les résultats ont montré une erreur de précision très faible (22 cm maximum). Dans les travaux de PHAM et al. (2007), une approche est proposée pour reconnaître les trajectoires de mouvement des personnes. Deux classes sont considérées dans les expérimentations : "aller en direction des capteurs", et "à travers les capteurs". Plusieurs techniques de reconnaissance de trajectoires ont été testées. La plus faible erreur de classification obtenue est de 10%. Dans (KALGAONKAR et al. 2009), un dispositif à base de capteurs à ultrasons a été proposé pour la reconnaissance de gestes. Cela consiste à utiliser un émetteur ultrason central et

trois récepteurs. Les gestes sont caractérisés par l'effet Doppler (décalage de fréquence d'une onde observée entre les mesures à l'émission et à la réception). Pour la phase expérimentale, les auteurs ont choisi d'évaluer leur système sur huit gestes d'une seule main (comme le mouvement de main de droite à gauche). Un score de précision de 88.42% est alors obtenu.

2.3.2 Reconnaissance de mouvements dans un environnement extérieur

Dans un environnement extérieur, et dans un contexte de suivi de la santé d'une personne à travers une application de suivi de mouvements, les capteurs utilisés dans un environnement intérieur ne sont plus exploitables. Nous nous orientons alors vers une autre catégorie de capteurs, nommés "capteurs portables" (wearable sensors en anglais). La figure 2-2 montre l'architecture générique pour la reconnaissance du mouvement humain avec des capteurs portables.

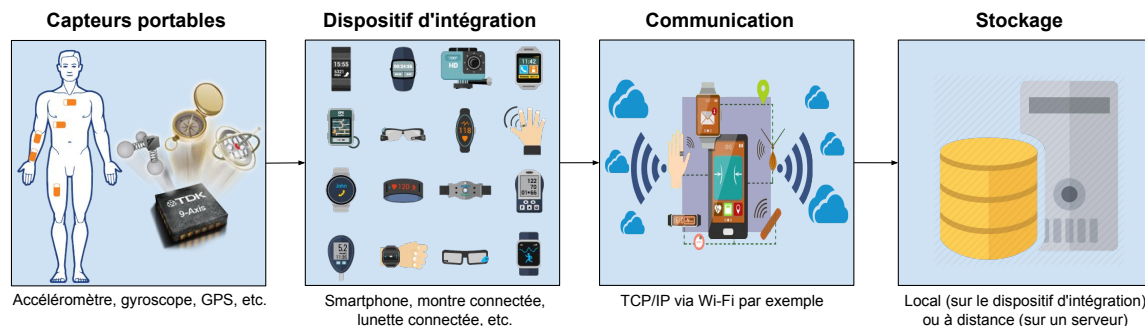


FIGURE 2-2 – Architecture générique pour la reconnaissance du mouvement humain avec des capteurs portables.

Microsystèmes électromécaniques inertiels

Les microsystèmes électromécaniques (Microelectromechanical systems - MEMS) sont des dispositifs miniaturisés comprenant un ou plusieurs éléments mécaniques utilisant l'électricité comme source d'énergie, en vue de réaliser une fonction de capteur ou d'actionneur. Les premiers MEMS ont été mis au point dans les années 70 et ils ont été mis sur le marché dans les années 80. Ils sont aujourd'hui très présents dans

notre vie quotidienne : dans la téléphonie, l'automobile, les appareils médicaux, etc. Les progrès dans le domaine de la miniaturisation font que ce genre de capteurs est peu coûteux, consomme moins d'énergie comparés aux capteurs précédemment cités, et en plus, ils sont moins intrusifs comparés aux caméras. Il existe plusieurs types de MEMS chacun destiné à un domaine bien spécifique. Ici, nous nous intéressons aux MEMS inertiels, nous parlons alors de centrale à inertie (Inertial Measurement Unit - IMU). Elle contient trois capteurs de mouvements : l'accéléromètre qui permet de calculer les accélérations statiques et dynamiques d'un dispositif, le gyroscope qui calcule la vitesse angulaire, et le magnétomètre qui détermine la force du champ magnétique local. La figure 2-3 illustre un exemple des signaux inertiels collectés par ces trois capteurs, l'action exécutée est la marche.

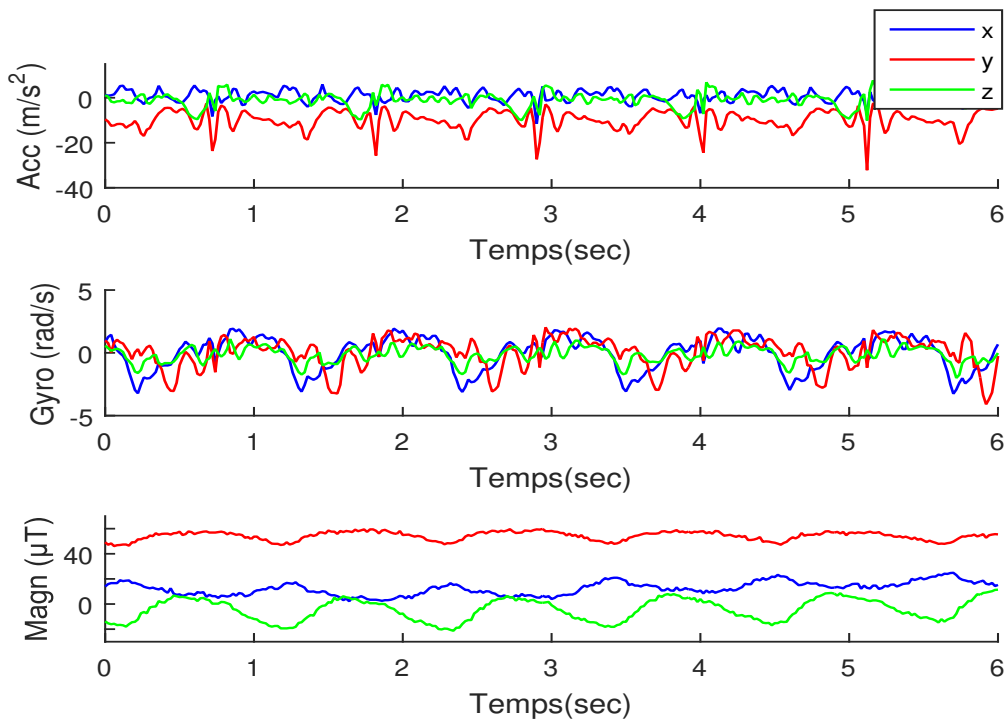


FIGURE 2-3 – Exemple de signaux inertiels d'une personne lors de la marche.

L'accéléromètre est le capteur le plus utilisé pour la reconnaissance des actions de déambulation (marcher ou courir par exemple). Dans les travaux de ATTAL et al. (2015), un système de reconnaissance d'actions chez les personnes âgées est proposé, trois capteurs accéléromètre triaxial de technologie Xsens sont utilisés (voir figure 2-

4). Les capteurs sont fixés sur trois emplacements à savoir la poitrine, la jambe droite et la cheville gauche du sujet. Plusieurs méthodes ont été testées pour reconnaître 12 actions (actions de type dynamiques, postures et transitions posturales), des scores de classification autour de 99% ont été obtenus pour certaines méthodes.

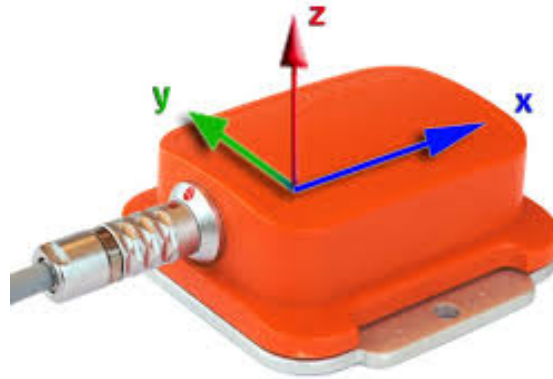


FIGURE 2-4 – Dispositif contenant un accéléromètre de la marque Xsens (voir (XSENS 2010) pour la documentation).

Dans (GUPTA et al. 2014), un accéléromètre triaxial de la marque Freescale est placé à la taille à l'aide d'une ceinture personnalisée. Les données d'accélération collectées sont envoyées à une passerelle de communication en utilisant le protocole ZigBee². Les données reçues par la passerelle sont par la suite envoyées à un serveur pour la phase de traitement. Six actions sont reconnues avec un score global de 98%.

Les capteurs portables de type IMU sont très populaires dans le domaine de la reconnaissance d'actions. Ils permettent d'obtenir des scores de performance très élevés comme nous venons de le voir ci-dessus. Cependant, ils présentent quelques inconvénients relatifs à l'utilisateur. En effet, porter plusieurs capteurs dans diverses positions peut s'avérer très inconfortable pour la personne, surtout à des fins de surveillance sur du long terme. En plus de cela, le coût d'achat de plusieurs capteurs par personne est aussi à prendre en considération.

2. Protocole de haut niveau permettant la communication d'équipements personnels ou domestiques qui sont dotés de petits émetteurs radios à faible consommation

Smartphone

De nos jours, les smartphones deviennent de plus en plus omniprésents et largement utilisés dans notre vie quotidienne. Ils peuvent constituer une très bonne alternative aux autres appareils portables (comme Xsens par exemple). De plus, les approches basées sur les smartphones ne nécessitent aucune infrastructure pour fonctionner. La contrepartie de ces faits est que le smartphone est utilisé non seulement à des fins de surveillance, mais également par la personne. De plus, contrairement aux appareils portables, les smartphones sont transportés dans divers emplacements (par exemple, une poche de pantalon, dans la main ou dans une pochette de ceinture). Par conséquent, les signaux collectés pour une action vont être fortement dépendants de la localisation des capteurs. Malgré cela, les capteurs intégrés dans les smartphones ont retenu l'attention des chercheurs. En effet, les smartphones d'aujourd'hui contiennent plusieurs capteurs qui fournissent des informations discriminantes pour la reconnaissance d'actions. En plus des capteurs inertiels (accéléromètre, gyroscope et magnétomètre), les smartphones incluent d'autres capteurs qui peuvent être utilisés pour la reconnaissance d'actions tels qu'un microphone, un GPS, ou des connectivités diverses (Bluetooth ou Wi-fi). Cependant, ces derniers consomment beaucoup plus d'énergie que les IMU. Dans (KJAERGAARD 2010), la consommation d'énergie des capteurs embarqués dans un smartphone Nokia N95 est étudiée. Des applications basées sur la localisation sont considérées dans ce cas. Le scénario consiste à estimer la position chaque 30 secondes. Les résultats ont montré que l'accéléromètre consomme en moyenne 0,05 watts, beaucoup moins que le Wi-fi (1,37 watts), le GPS (0,32 watts), le Bluetooth (0,28 watts) ou le microphone (0,26 watts). Dans (Jihe WANG et al. 2015), la métrique "puissance par Hz " est utilisée pour représenter la consommation d'énergie d'un échantillon par seconde. Les capteurs ont été paramétrés sur plusieurs modes : éteint, lent, normal, jeu et rapide. Les résultats ont montré que l'accéléromètre a consommé $0,35mW/Hz$, le gyroscope $15,25mW/Hz$ et le magnétomètre $10,00mW/Hz$. Ces capteurs IMU cités consomment beaucoup moins que le GPS ($166,10mW/Hz$), le Bluetooth ($44,90mW/Hz$) ou le microphone

(289, 50mW/Hz).

Étant donné la faible consommation d'énergie que présentent les capteurs inertiels, beaucoup de chercheurs ont choisi d'exploiter ces capteurs embarqués dans un smartphone pour la reconnaissance d'actions. Que se soit pour des personnes en bonne santé (ANGUITA et al. 2012; RONA O et al. 2016), âgées (CONCEPCIÓN et al. 2017; OUCHI et al. 2013), ou en récupération post-AVC (CAPELA et al. 2016; O'BRIEN et al. 2017), le smartphone a montré sa capacité à discriminer différentes classes d'actions.

2.4 Prévention de la perte d'autonomie

Afin de réduire les risques de perte d'autonomie, différentes mesures préventives ont été proposées. Une des mesures populaires dans la littérature consiste à faire des exercices basés sur les jeux vidéo, ce concept est appelé "exergaming". Ces exercices permettent de pratiquer plus de mouvements physiques et ainsi améliorer la santé des personnes âgées (PRIMACK et al. 2012). Des études ont montré que l'exergaming améliore la mobilité de la personne (MAILLOT et al. 2012), le contrôle de l'équilibre (RENDON et al. 2012), la force musculaire des membres inférieurs (JORGENSEN et al. 2012) et la cognition (O'LEARY et al. 2011) chez les personnes âgées.

Les exercices basés sur les jeux vidéo sont très avantageux par rapport aux exercices traditionnels. Ils permettent la réalisation d'un mouvement particulier avec différents niveaux de difficulté, ce qui permet aux utilisateurs de commencer par le niveau de difficulté approprié, puis d'évoluer progressivement en difficulté, en fonction des performances en temps réel.

Afin d'étudier l'apport des exercices basés sur les jeux vidéo sur la santé de la personne, différentes technologies commerciales ont été utilisées par les auteurs, les plus populaires sont détaillées ci-dessous.

2.4.1 Nintendo Wii

La console Nintendo Wii (figure 2-5) est la technologie la plus utilisée dans la littérature sur les exercices basés sur les jeux vidéo. Dans (YOUNG et al. 2011), la

“Wii Balance Board” (figure 2-5c) a été utilisée pour calculer le centre des pression des personnes âgées qui présentent des risques de chute. Les valeurs de centre des pression calculées sont ensuite récupérées et affichées en temps réel sur une interface pour réaliser un diagnostic. Afin d’atteindre cet objectif, deux jeux d’exercices d’équilibre ont été créés, chacun comprenant des niveaux de difficulté croissants (les jeux existants étaient jugés trop difficiles pour les personnes âgées). Le premier jeu invite les participants à contrôler la position médio-latéral d’un panier en utilisant leur centre des pression pour attraper des pommes qui tombent d’un arbre. Dans le deuxième jeu, le centre des pression est exploité pour contrôler la position médio-latéral et antérieur-postérieur d’un personnage dans le but d’éclater des bulles. Afin d’évaluer l’approche proposée, six personnes âgées en bonne santé ont participé à 10 sessions de jeux de 20 minutes chacune, étendues sur une période de 4 semaines. Les résultats ont montré la diminution de la variabilité du balancement médio-latéral et antérieur-postérieur. Dans (AGMON et al. 2011), la sécurité et la faisabilité de l’utilisation du jeu pour Nintendo “Wii Fit”³ sont étudiées afin d’améliorer l’équilibre chez les personnes âgées. Pour la phase expérimentale, les chercheurs ont recruté sept personnes âgées et atteintes de troubles de l’équilibre pour participer à l’évaluation du jeu “Wii Fit”. Les participants ont joué à 4 jeux différents pour en moins 30 minutes, trois fois par semaine, et cela pendant 3 mois. Les résultats ont montré qu’un entraînement à la maison avec le jeu “Wii Fit” est sûr et faisable pour l’échantillon d’adultes âgés sélectionnés. En outre, les participants ont enregistré une amélioration de l’équilibre, et ils ont noté qu’ils ont apprécié l’activité. Dans (RENDON et al. 2012), 40 personnes âgées ont été réparties aléatoirement dans deux groupes. Le premier groupe a pratiqué le jeu “Wii Fit” 3 fois par semaine durant 6 semaines. La durée d’une séance de pratique était d’environ 35 à 45 minutes. Le second groupe (groupe de contrôle) n’a reçu aucune intervention. Les auteurs ont conclu que le jeu “Wii Fit” a permis au premier groupe d’améliorer l’équilibre et la stabilité posturale.

D’autres chercheurs ont exploité la télécommande Wii (figure 2-5b) qui permet

3. Un ensemble de 40 épreuves physiques utilisant la balance et comprenant par exemple le yoga, la gymnastique, l’aérobic, des jeux d’équilibre etc.



FIGURE 2-5 – La Wii.

de contrôler la console. Cette dernière est équipée d'un accéléromètre permettant de reconnaître des actions effectuées par les utilisateurs. Dans les travaux de [KEOGH et al. \(2014\)](#), l'impact du jeu "Wii sports" sur l'état physique des personnes âgées est analysé. Afin de mener cette étude, 34 résidents d'une maison de retraite ont participé à l'expérience suivante : les sujets ont d'abord été répartis dans deux groupes. Le premier groupe (groupe d'intervention) comptait 19 personnes qui ont joué à 5 jeux de sport inclus dans le jeu "Wii Sports" (baseball, boxe, golf, tennis et bowling). L'expérience a duré 8 semaines, avec une moyenne de 30 minutes par semaine pour chaque participant. Le second groupe est considéré comme groupe témoin à titre comparatif. Afin d'analyser l'impact du jeu "Wii sports" sur les participants, trois mesures étaient calculées une semaine avant et après l'expérience : la capacité fonctionnelle, qui est calculée par le biais d'un test de force musculaire et d'endurance ([RIKLI et al. 1999](#)), une mesure des activités physiques, calculée à l'aide d'un questionnaire (voir ([TOPOLSKI et al. 2006](#)) pour plus de détails) et une mesure de la qualité de vie, calculée avec un questionnaire aussi (voir ([GROUP et al. 1998](#))). Les résultats obtenus ont montré que le groupe d'intervention présentait une amélioration significative de la mobilité et l'intensité des activités physiques, une augmentation de la force des bras est aussi observée. En outre, un entretien avec le groupe d'intervention a révélé qu'ils ont apprécié les sessions de jeu.

En plus de lutter contre les limitations physiques, l'exergaming est utilisé pour lutter contre les maladies psychologiques comme la dépression par exemple. Cette maladie est très répandue chez les personnes âgées. Selon une enquête de la Direction

de la recherche, des études, des évaluations et des statistiques (DREES), parmi les résidents en établissements d’hébergement pour personnes âgées, 34% se trouvent dans un état dépressif (PERRIN-HAYNES et al. 2011). Dans cette thématique, les auteurs de (JUNG et al. 2009) ont étudié l’impact de jouer à la console Wii sur le bien-être psychologique et physique des personnes âgées résidentes dans un établissement de soins. Pour atteindre leur objectif, ils ont collaboré avec un établissement de soins de longue durée à Singapour nommé SASCO. Quarante-cinq adultes âgées entre 56 et 92 ans issus de cet établissement ont participé à cette expérience. Ces derniers ont été divisés en deux groupes de 30 et de 15 personnes. En utilisant la télécommande Wii, le premier groupe a joué au jeu “Wii sports” (tennis, bowling, baseball et boxe) plus un jeu de cuisine. Dans le deuxième groupe (groupe de contrôle), les participants jouaient à des jeux traditionnels comme les jeux de mémoires ou UNO (jeu de cartes). Les séances d’activités duraient une heure et demie et elles ont eu lieu trois fois par semaine. Afin d’estimer le bien-être des personnes, quatre mesures ont été calculées : un score de solitude nommé “UCLA Loneliness Scale” (voir (RUSSELL 1996) pour plus de détails), un score d’estime de soi⁴, un score d’affectivité (niveau affect des participants, positif ou négatif, voir (BRADBURN et al. 1965) pour plus de détails), et le dernier score a été calculé suivant un questionnaire sur les activités physiques pratiquées durant la semaine (YASUNAGA et al. 2007). Les résultats obtenus ont montré que les personnes qui ont joué à la console Wii présentent un bien-être psychologique meilleur que les personnes qui ont joué aux jeux traditionnels.

2.4.2 Caméra EyeToy de Sony

EyeToy est une caméra commercialisée par Sony en 2003 pour être utilisée avec la console PlayStation 2 (figure 2-6). Elle permet aux utilisateurs d’interagir avec certains jeux de la console en utilisant le mouvement.

La caméra EyeToy a été exploitée par des chercheurs afin d’étudier l’apport des exercices physiques sur la santé de la personne. Dans (Sunwoo LEE et al. 2013), une

4. Proposé par ROSENBERG (2015) pour mesurer le sentiment de confiance des adolescents. Depuis, il a été étendu aux personnes âgées



FIGURE 2-6 – La caméra EyeToy.

étude est menée sur des personnes âgées atteintes de diabète de type 2. L’objectif est de déterminer si un programme d’exercice utilisant la caméra EyeToy est bénéfique pour l’état physique de cette catégorie de personnes. 55 sujets âgés de plus de 65 ans ont participé à la phase expérimentale. Ils ont été divisés aléatoirement en deux groupes : le groupe d’intervention, qui a participé à un programme d’entraînement comptant des sessions de 50 minutes, répétées trois fois par semaine, et cela pendant 10 semaines. La caméra EyeToy était utilisée pour l’entraînement, 6 jeux ont été sélectionnés suivant des critères moteurs, comme par exemple le mouvement dans plusieurs directions des membres supérieurs et inférieurs du corps, ou encore l’inclinaison du tronc. Le second groupe (groupe de contrôle) n’a reçu qu’une éducation sanitaire sur la gestion du diabète. Des mesures ont été prises avant et après l’expérience pour chaque groupe à l’aide de tests cliques : test “debout sur une jambe” (utilisé pour évaluer la stabilité posturale), test d’équilibre de Berg (utilisé pour évaluer l’équilibre des personnes, voir les travaux de BERG et al. (1992) pour plus de détails), le test de portée fonctionnelle (DUNCAN et al. 1990) (Functional Reach test en anglais), qui permet de mesurer l’équilibre dynamique et la flexibilité lorsque les sujets effectuent des tâches fonctionnelles, le test “timed up-and-go”, qui consiste à mesurer le temps qu’il faut à un sujet pour se lever d’un fauteuil (46 cm de hauteur), marcher sur une distance de 3 mètres, puis se retourner, regagner son fauteuil et se rasseoir. Et le

dernier test “sit-to-stand”, où il est demandé aux sujets de se lever d’une chaise sans accoudoirs, les bras croisés, puis de revenir s’asseoir sur la chaise le plus rapidement possible cinq fois de suite. Les résultats obtenus dans le groupe d’intervention ont montré une amélioration significative de l’équilibre, la diminution de la durée nécessaire pour effectuer le test “sit-to-stand”, l’augmentation de la vitesse et de la cadence de la marche. Les auteurs de (FLYNN et al. 2007) ont exploré la faisabilité de l’utilisation de la caméra EyeToy pour un individu en phase chronique de rétablissement d’un AVC. Une femme âgée de 76 ans a participé à cette étude, elle était victime d’un AVC hémorragique 17 mois avant le début de l’étude. La patiente a pratiqué 20 séances d’entraînement en utilisant la caméra EyeToy, chaque séance durait une heure et la période totale du programme était de quatre semaines et demie. L’équipe de recherche n’a pas supervisé les séances d’entraînement, cependant, la patiente était surveillée et encouragée par son entourage pendant ces séances dans la pièce ou à la maison. Afin d’évaluer la faisabilité de l’utilisation du dispositif, la perception du participant a été obtenue à l’aide d’un journal quotidien. De plus, les chercheurs ont procédé à un entretien approfondi après un mois d’entraînement. En outre, des mesures ont été effectuées sur la patiente avant le début de l’entraînement, après 10 séances, à la fin, et enfin 6 mois après. 11 mesures étaient considérées, parmi elles, la mesure de la récupération motrice (FUGL-MEYER et al. 1975), l’indice de la démarche (à travers l’évaluation de l’équilibre, le contrôle postural, la mobilité et les capacités fonctionnelles de la démarche), score de dépression, etc. Les résultats obtenus ont montré la faisabilité de l’utilisation d’un programme d’entraînement à base de caméra EyeToy chez les patients post-AVC. Les résultats des tests post-entraînement de la patiente ont indiqué des améliorations sur l’indice de la démarche, la récupération motrice, l’équilibre et une diminution de la dépression.

2.4.3 Caméra Kinect de Microsoft

La caméra Kinect (figure 2-7) est un périphérique fabriqué par Microsoft, elle est utilisée comme accessoire de jeu pour la console Xbox. Deux mois après sa sortie, Microsoft annonce avoir vendu 8 millions d’exemplaires dans le monde, et inscrit la

Kinect dans le livre Guinness des records comme étant l'accessoire high-tech le plus vendu dans un court laps de temps.

La caméra Kinect permet de collecter les données de profondeur, elle est de la même famille que les capteurs mentionnés dans la partie 2.3.1. Le système contient une caméra couleur (données *RGB*), un émetteur (projecteur) et un récepteur infrarouge. Il existe deux types d'approches utilisées par ces capteurs pour le calcul des données de profondeur : la première, exploitée dans la version 1 de la Kinect est nommée lumière structurée, elle consiste à projeter un motif connu sur une scène. Suivant la manière dont le motif se déforme en heurtant les surfaces, le système calcul les données de profondeur relatives aux objets de la scène. Le deuxième procédé de calcul des données de profondeur est appelé temps-de-vol (time of flight en anglais), on le trouve dans la version 2 de la Kinect. Cela consiste à émettre un signal lumineux et calcule le temps qu'il met pour effectuer le trajet entre l'objet et la caméra.



FIGURE 2-7 – La caméra Kinect (version 1).

Le succès de la Kinect dépasse largement sa simple utilisation comme périphérique d'une console de jeux. Dans (J. KIM et al. 2013), la Kinect est exploitée pour évaluer les effets d'un programme d'exercices physiques sur la force musculaire de la hanche et le contrôle de l'équilibre chez les personnes âgées. Trente-six adultes résidents en établissement d'hébergement, âgés entre 65 et 75 ans ont participé à cette étude. Les sujets ont été divisés aléatoirement en deux groupes : le premier groupe (groupe d'intervention) a suivi un programme d'entraînement de 8 semaines en exploitant la Kinect. Le programme consiste à pratiquer des exercices de Tai Chi et yoga issu du jeu "Your Shape : Fitness Evolved". Le second groupe (groupe de contrôle) n'a reçu aucune intervention. Afin d'évaluer l'impact du programme sur les participants, la

force musculaire est mesurée avant et après l'expérimentation à l'aide d'un dynamomètre conçu à cet effet. Les résultats obtenus dans le groupe d'intervention ont montré une amélioration significative de la force musculaire de la hanche, plus précisément les muscles extenseurs, fléchisseurs, adducteurs et abducteurs. En ce qui concerne le groupe témoin, aucune amélioration significative n'a été observée. Dans (POMPEU et al. 2014), la faisabilité et la sécurité de pratiquer le jeu "Kinect Adventures" pour les personnes atteintes de la maladie de Parkinson sont analysées. L'étude a été menée dans un centre de rééducation de l'université de São Camilo, au Brésil. Sept adultes âgés entre 60 et 85 ans ont participé à l'expérience, tous atteints de la maladie de Parkinson idiopathique de stade 2 ou 3 sur l'échelle de Hoehn et Yahr⁵. Les sujets ont participé à 3 sessions d'entraînement par semaine, une heure chacune (14 sessions au total). Le programme consiste à pratiquer 4 mini-jeux inclus dans le jeu "Kinect Adventures". Les jeux étaient sélectionnés suivant des critères moteurs, comme par exemple le transfert de poids entre les membres inférieurs ou l'inclinaison du tronc. Les critères cognitifs sont aussi considérés, comme par exemple l'attention visuo-spatiale ou le temps de réaction. Les auteurs ont conclu que l'entraînement avec la Kinect était sans danger et réalisable par des patients atteints de la maladie de Parkinson de stade 2 et 3. En effet, les participants ont fini les 14 sessions d'entraînement avec une bonne adhésion. Malgré la stimulation visuelle intense et le mouvement rapide de la tête requis par les jeux, aucun des participants n'a connu de vertige, de nausée ni aucun effet indésirable. En outre, l'entraînement avec la Kinect a permis aux patients d'améliorer leur endurance cardio-pulmonaire, leur équilibre et leur démarche.

2.5 Objectifs des travaux de thèse

Ce projet de thèse vise à étudier l'apport des techniques d'analyse de données pour le maintien de l'autonomie et le bien-être des personnes adultes. Ce travail peut être divisé en deux volets : reconnaissance des actions humaines ainsi que leur suivi

5. Système couramment utilisé pour décrire l'évolution des symptômes de la maladie de Parkinson. Voir les travaux de GOETZ et al. (2004) pour plus de détails

sur le long-terme et la prévention de la perte d'autonomie.

2.5.1 Reconnaissance des actions humaines et suivi sur le long terme

Cela consiste à construire un système de reconnaissance des actions humaines effectuées durant la journée. Afin d'atteindre notre objectif, quatre aspects doivent être pris en compte dans le choix du capteur :

(1) précision : performances élevées du système à reconnaître les actions considérées. (2) robustesse : la préservation des caractéristiques du système, telles que sa performance en présence des perturbations inconnues et du bruit. (3) respect de la vie privée : système non intrusif qui préserve au maximum la vie privée des utilisateurs. (4) coût : proposer une solution à faible coût accessible à une grande partie de la population.

Comme nous l'avons expliqué dans la partie 2.3, diverses solutions de capteurs existent dans la littérature, cependant une seule solution réunit les quatre aspects cités ci-dessus à savoir l'utilisation des capteurs inertiels embarqués dans un smartphone.

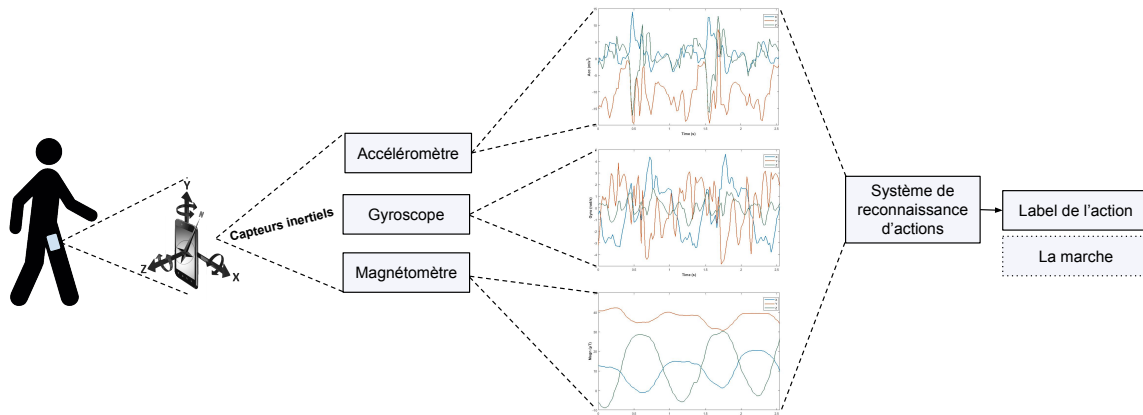


FIGURE 2-8 – Reconnaissance de l'action humaine à l'aide d'un smartphone.

La figure 2-8 schématise la reconnaissance d'actions avec un smartphone. Les capteurs inertiels embarqués à l'intérieur du smartphone porté par la personne à savoir l'accéléromètre (au minimum), le gyroscope (fréquemment présent) et le magnétomètre (selon les modèles) permettent de collecter les signaux inertiels, ces derniers

seront ensuite mis à l'entrée d'un système de reconnaissance d'actions afin d'identifier l'action exécutée en temps réel. Dans l'exemple de la figure 2-8, le système est censé émettre l'étiquette de l'action "la marche", autrement cela sera considéré comme une erreur de classification.

2.5.2 Prévention de la perte d'autonomie

En fonction des actions effectuées par la personne durant la journée, un programme d'exercices pourra lui être recommandé afin de promouvoir le maintien de son autonomie et de son bien-être. Le choix du capteur à utiliser étant primordial pour notre application, en plus des aspects pris en compte précédemment (précision, robustesse, respect de la vie privée et coût), la facilité d'utilisation est aussi à considérer. Diverses solutions autour de la problématique "prévention de la perte d'autonomie" ont été exposées dans la partie 2.4, en prenant en compte les avantages et les inconvénients de chaque solution, nous avons opté pour l'utilisation de la Kinect.

Un programme d'exercices comme ceux proposés dans le jeu "Kinect Adventures" passe par la reconnaissance des actions considérées dans le jeu. Dans le même objectif, ce volet de notre travail vise à reconnaître les actions humaines qui correspondent à des mouvements susceptibles d'être utilisés dans les jeux. Pour répondre à cette problématique, nous avons développé un système robuste, précis et peu coûteux en temps de calcul, montrant des résultats supérieurs à plusieurs travaux de l'état de l'art. Afin de diminuer au mieux la latence, notre système est capable de reconnaître une action avant la fin de l'exécution de celle-ci (reconnaissance anticipée).

Pour résumer, l'approche générale est illustrée dans la figure 2-9.

2.5.3 Actions considérées dans notre travail

Dans notre étude, les actions humaines effectuées par la personne sont suivies au cours de la journée à l'aide de son smartphone. En fonction de la catégorie d'actions effectuées, un programme d'exercice basé sur la Kinect pourra être recommandé à l'utilisateur. Il consiste à l'exécution d'actions favorables au maintien de son auto-

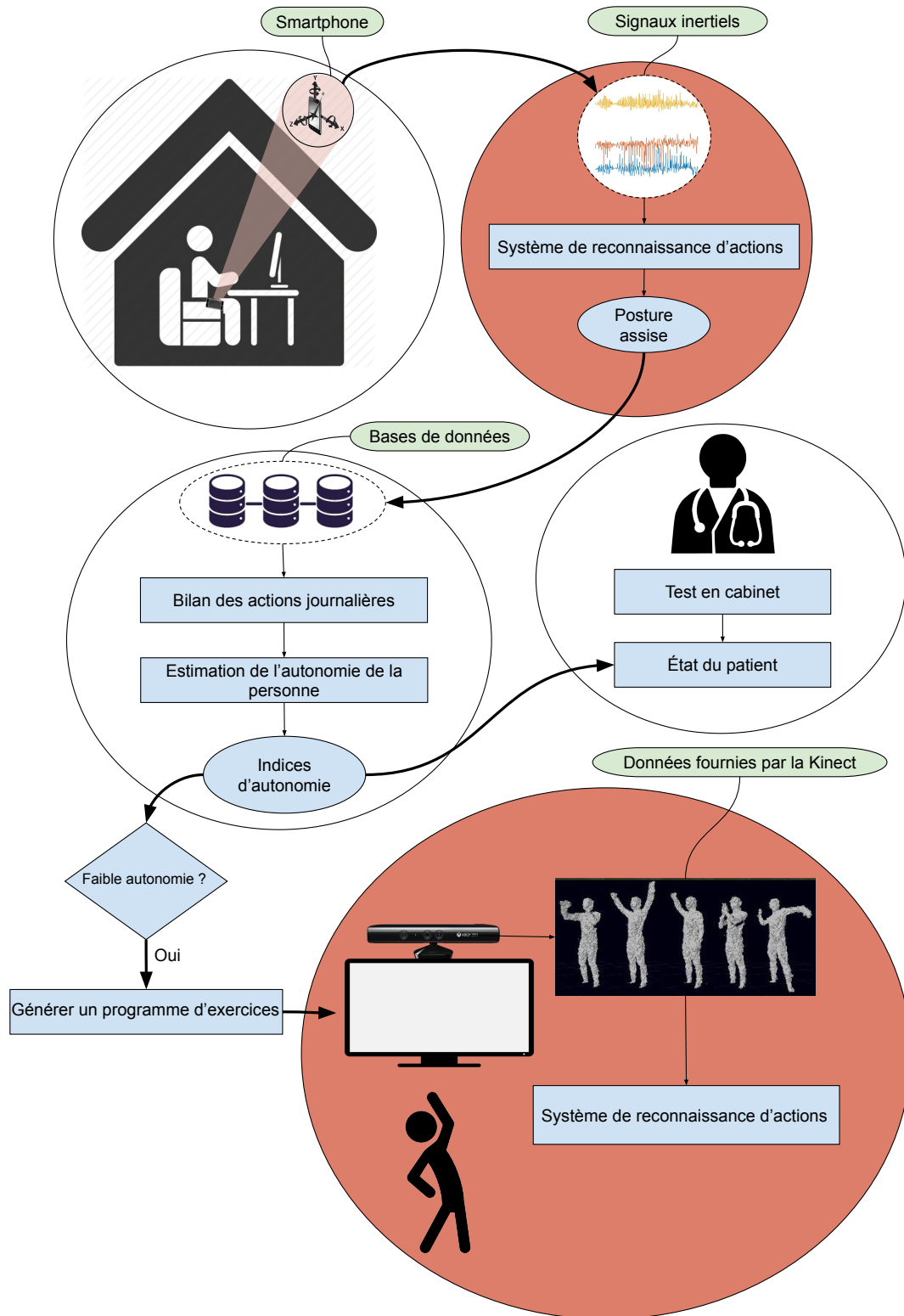


FIGURE 2-9 – Approche générale proposée dans notre travail.

nomie et son bien-être. Pour mener à bien notre travail, les actions corrélées avec l'autonomie et le bien-être doivent être suivies. Par conséquent, trois catégories d'actions sont prises en considération dans notre approche globale : actions sédentaires, périodiques (ou pseudo-périodique) et non périodiques (figure 2-10).

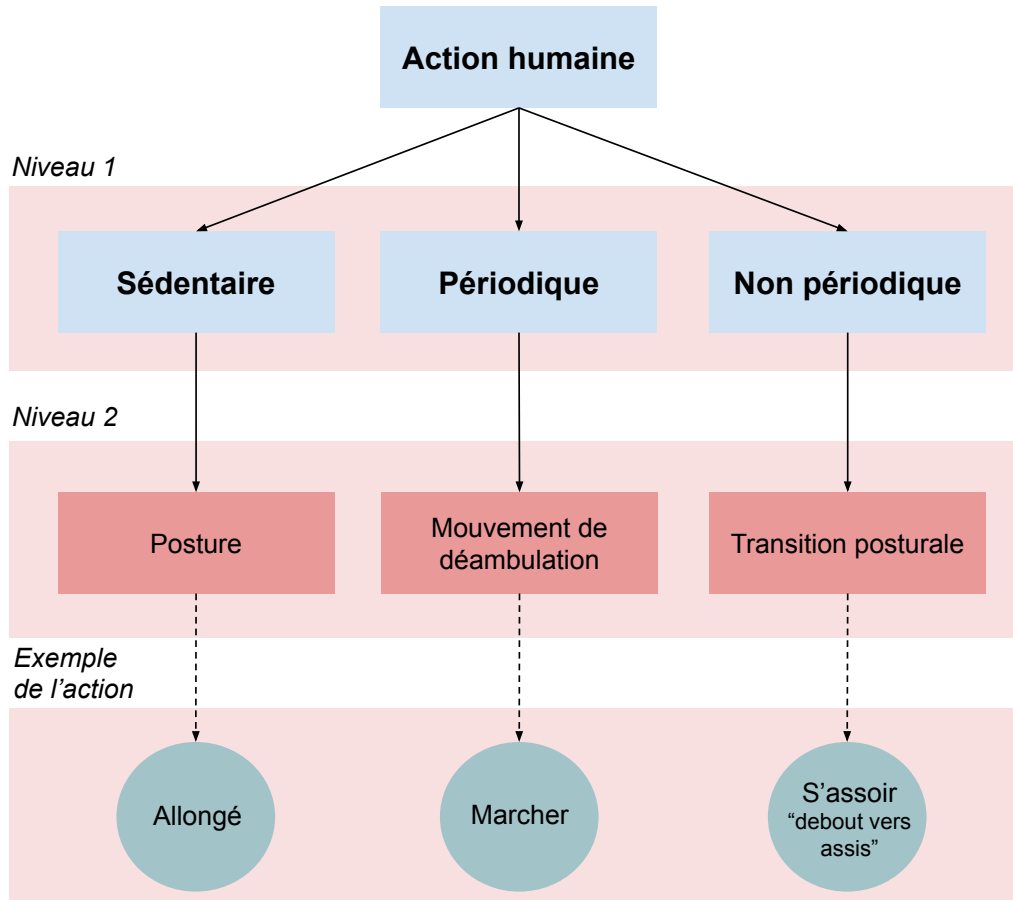


FIGURE 2-10 – Les catégories d'actions prises en considération dans notre travail.

Les actions journalières d'une personne provoquent un réel impact sur son autonomie, elles peuvent être le signe d'apparition de facteurs de risque ou bénéfiques pour la santé. Les comportements sédentaires sont à surveiller pour les personnes adultes. Ce terme désigne les comportements caractérisés par une faible dépense d'énergie, comme dormir, s'allonger, conduire, regarder la télévision. Les comportements sédentaires impliquent une dépense d'énergie entre 1,0 et 1,5 MET (équivalent métabolique) (AINSWORTH et al. 2000). Dans la société actuelle, ces comportements sont de plus en plus présents dans la vie quotidienne. Le mode de vie sédentaire expose

la personne aux risques d'obésité (F. B. HU et al. 2003), de diabète et de maladies cardio-vasculaires (WILMOT et al. 2012; F. B. HU et al. 2003; KATZMARZYK et al. 2009).

Les actions de nature périodique correspondent aux mouvements de déambulation, comme par exemple : marcher, jogging, monter et descendre les escaliers. Ces dernières sont recommandées pour le maintien d'une bonne santé. Dans (HASKELL et al. 2007; LACROIX et al. 1996; MANSON et al. 1999), le manque de marche est associé aux maladies cardio-vasculaires. Par conséquent, pour un adulte, il est recommandé de marcher pendant 30 minutes en deux jours par semaine, et de faire un jogging pendant 20 minutes en deux autres jours.

La troisième catégorie d'action considérée dans notre étude est de nature non périodique et cela correspond à des actions courtes telles que les transitions posturales. En effet, les transitions entre postures sont des actions très importantes et possèdent un impact sur la santé d'un adulte (KRALJ et al. 1990; RILEY et al. 1991). Les transitions "assis vers debout" et "debout vers assis" sont par exemple recommandées dans les travaux de (ATLAS et al. 2001). L'objectif est d'éviter les positions assis et debout d'une façon prolongée. Au travail ou dans un véhicule, une transition toute les 30 minutes au minimum est recommandée. En outre, des études ont montré que des paramètres liés aux transitions debout-assis et assis-debout sont corrélés au risque de chute chez les patients âgés (NAJAFI et al. 2002), aux accidents vasculaires cérébraux (CHENG et al. 1998) et à la maladie de Parkinson (ASHBURN et al. 2001), de même pour le niveau fonctionnel d'une personne (JANSSEN et al. 2002), l'évaluation du contrôle postural (LORD et al. 2002) et la mesure du handicap global (GURALNIK et al. 1995).

2.6 Contributions

Les contributions sont divisées suivant l'application : le suivi des actions humaines à l'aide d'un smartphone ou la prévention de la perte d'autonomie avec la Kinect.

2.6.1 Système de suivi des actions humaines

Un système de classification en temps réel des actions humaines est proposé dans notre travail. Les contributions sont citées ci-dessous.

Approche de classification hiérarchique des actions humaines

Dans notre système, une approche de classification hiérarchique est proposée, i.e., les actions sont d'abord regroupées en méta-classes (postures ou actions périodiques par exemple), le premier niveau de classification permet de reconnaître ces méta-classes, ensuite selon la catégorie de l'action reconnue, un classifieur adapté est sélectionné afin de reconnaître l'action proprement dite.

Construction d'un vecteur d'attributs adapté au problème de classification

Dans notre système, selon le problème de classification rencontré, un vecteur d'attributs optimal et pertinent est construit, i.e., au niveau de chaque nœud interne de la méthode de classification, un algorithme de sélection d'attributs SFSS (Sequential Forward Floating Search) est appliqué afin d'optimiser au maximum les entrées des classifieurs.

Robustesse de l'approche proposée à l'emplacement des capteurs

Dans les approches basées sur les données inertielles, l'emplacement des capteurs est considéré comme l'un des paramètres les plus importants. Afin de répondre à la problématique de sa méconnaissance, une méthode indépendante à l'emplacement des capteurs est proposée, i.e., l'algorithme de sélection des attributs est appliqué pour le choix d'attributs indépendants à la position des capteurs.

Proposition d'une méthode de reconnaissance des transitions posturales

Afin de reconnaître la classe des transitions posturales (souvent négligé dans la littérature), nous proposons une méthode originale exécutée en deux parties : la première consiste à la détection de la transition sans aucune information sur son type. La

seconde partie vise à identifier la transition parmi six catégories de transitions posturales, à savoir : “debout vers assis”, “assis vers debout”, “allongé vers assis”, “allongé vers debout”, “debout vers allongé”, “assis vers allongé”.

2.6.2 Système de prévention de la perte d’autonomie

Un système de reconnaissance d’action en temps réel utilisant la Kinect est proposé dans notre travail. Dans notre recherche, nous nous sommes focalisés sur les aspects suivants : la reconnaissance rapide et anticipée des actions et la robustesse du système.

Reconnaissance anticipée des actions humaines

Pour reconnaître une action avant la fin de son exécution par le sujet, un système qui repose sur les modèles de Markov cachés (HMMs) avec des distributions de sorties d’état représentées par des modèles de mélange gaussiens (GMMs) est proposé dans nos travaux. Afin de classifier une action en cours d’exécution, chaque classe considérée par notre système fournit en sortie de notre approche un score de vraisemblance. La classe qui procure la plus grande vraisemblance à un instant donné est représentative de l’action exécutée.

Construction d’un vecteur d’attributs adaptés pour la reconnaissance anticipée

Afin de réaliser efficacement la reconnaissance anticipée de l’action, nous proposons un nouvel algorithme d’extraction d’attributs nommé Body-part Directional Velocity (BDV). Cet algorithme est rapide à calculer et permet de modéliser la vitesse et la direction du mouvement des membres du corps.

2.7 Organisation

Le manuscrit est organisé en trois parties qui suivent l’introduction. La première partie est dédiée à la reconnaissance des actions humaines à l’aide de capteurs inertiels,

la seconde partie est consacrée à la reconnaissance anticipée des actions avec la Kinect. La conclusion générale et les perspectives sont présentées dans la troisième partie.

La première partie contient deux chapitres définis comme suite :

Le chapitre 3 approfondi et critique l'état de l'art autour de la problématique de reconnaissance d'actions à l'aide de capteurs inertiels.

Le chapitre 4 présente et décrit notre approche proposée afin de répondre à la problématique et aux défis concernant la reconnaissance d'actions.

La seconde partie contient elle aussi deux chapitres répartis de la même façon que la première partie :

Le chapitre 5 présente les différentes méthodologies utilisées pour la reconnaissance d'actions à l'aide de données de profondeur.

Le chapitre 6 décrit le système de reconnaissance anticipé d'actions proposé et compare les résultats expérimentaux avec d'autres approches de l'état de l'art.

La troisième partie clôt ce manuscrit par une conclusion générale et les perspectives.

2.8 Conclusion

L'aspect matériel est essentiel à notre application, la qualité et les résultats de cette dernière en dépendent. Ainsi, deux volets incontournables ont été abordés dans ce chapitre. Le premier volet présente la panoplie de capteurs et d'appareils exploités pour reconnaître les mouvements d'une personne. En faisant référence aux travaux de l'état de l'art et en développant l'apport de chacun d'eux, nous avons justifié notre choix du smartphone. Pour son aspect non intrusif, sa disponibilité, sa capacité d'intégrer des capteurs de détection de mouvement et même d'embarquer notre application, ce matériel se présente comme le plus approprié. Le deuxième volet de ce chapitre a été consacré au matériel préconisé à titre préventif de la perte d'autonomie. Dans ce cas, nous avons opté pour la Kinect pour sa facilité et sa simplicité d'utilisation.

Deuxième partie

Reconnaissance des actions humaines à l'aide de capteurs inertiels

Chapitre 3

État de l'art : Choisir les paramètres et analyser le signal

Contents

3.1	Introduction	66
3.2	Système HAR à l'aide de capteurs inertiels	66
3.3	Collecte de données	67
3.4	Emplacement du smartphone	68
3.5	Segmentation	70
3.6	Pré-traitement	71
3.7	Extraction des attributs	72
3.8	Réduction de la dimensionnalité des attributs	76
3.9	Classification des actions humaines	77
3.10	Classification des transitions posturales	80
3.11	Évaluation des performances	80
3.12	Conclusion	82

3.1 Introduction

La reconnaissance d'actions humaines (Human Action Recognition - HAR) à l'aide de capteurs inertiels est un sujet très actif et populaire. Les premiers travaux remontent aux années 90 (FOERSTER et al. 1999). Cependant, il y a encore beaucoup de défis à résoudre, et l'utilisation d'un smartphone comme dispositif pour les capteurs rajoute encore plus de contraintes. Dans ce chapitre, nous examinons l'état de l'art autour de la reconnaissance d'actions humaines à l'aide de capteurs inertiels. Nous nous focalisons en particulier sur les systèmes utilisant des capteurs inertiels embarqués sur un smartphone.

3.2 Système HAR à l'aide de capteurs inertiels

La figure 3-1 présente le schéma de base suivi dans la littérature pour la reconnaissance d'actions à l'aide d'un smartphone.

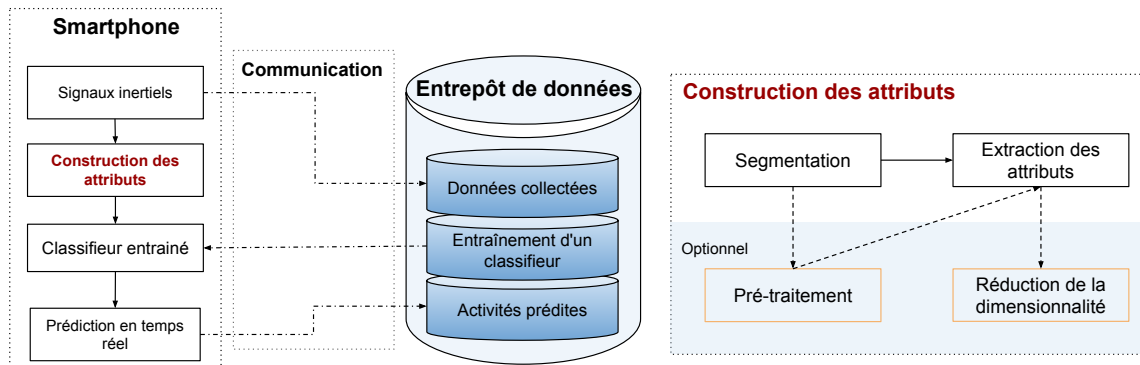


FIGURE 3-1 – Schéma fonctionnel d'un système HAR.

Les données collectées sont d'abord segmentées à l'aide d'une fenêtre glissante rectangulaire avec un degré de chevauchement entre les fenêtres adjacentes. N segments égaux de taille T sont alors obtenus : $W = \{W_{T_1}, W_{T_2}, \dots, W_{T_N}\}$. Les segments sont pré-traités à l'aide des techniques de traitement du signal afin d'éliminer le bruit et les informations indésirables. Une liste d'attributs est ensuite extraite des signaux bruts ou pré-traités pour décrire chaque action humaine. Cette étape peut être suivie par des techniques de réduction de la dimensionnalité afin d'optimiser le nombre

d'attributs. Enfin, les attributs sélectionnés sont utilisés pour entraîner un classifieur choisi.

Dans le reste du chapitre chaque phase suivie dans le processus de reconnaissance d'actions est examinée. Ce travail va nous permettre par la suite de justifier le modèle et les traitements retenus.

3.3 Collecte de données

Avant d'alimenter un système HAR, une collecte de données est réalisée en suivant un protocole bien particulier : un groupe de personnes porte un smartphone dans un emplacement fixe et effectue un ensemble d'actions. Les signaux de plusieurs capteurs sont alors collectés à une fréquence d'échantillonnage notée F_e . Cela inclut des données inertielles : signaux d'accélération à l'aide d'un accéléromètre triaxial, signaux de vitesse angulaire en utilisant un gyroscope triaxial et mesure du champ magnétique avec un magnétomètre triaxial. La fréquence d'échantillonnage est fixée en fonction de l'ensemble des actions considérées par le système HAR. Dans (KARANTONIS et al. 2006), les actions de marche (marche faible, normale et lente), les transitions posturales et la chute ont été étudiées à l'aide d'un accéléromètre triaxial accroché à la taille. Selon les auteurs, les signaux générés par ces actions présentent des fréquences en dessous de $20Hz$. Par conséquent, une fréquence d'échantillonnage de $40Hz$ pourrait être suffisante. Dans (ANGUITA et al. 2012; SHOAI, BOSCH, O. D. INCEL et al. 2014; ATALLAH et al. 2011), F_e est fixée à $50Hz$. D'autres chercheurs ont choisi une fréquence plus importante, dans (J.-Y. YANG et al. 2008; REISS et al. 2012; HE et al. 2009) $F_e = 100Hz$. Une fréquence d'échantillonnage élevée permet de surmonter les effets de repliements du spectre. Cependant, cela augmente le temps de calcul du système HAR, les exigences en énergie (plus d'échantillons à traiter) et le bruit présent dans le signal final.

3.4 Emplacement du smartphone

La position d'un smartphone est un paramètre important à prendre en considération, i.e., les signaux de sortie des capteurs inertiels présentent des caractéristiques différentes selon la position du smartphone. La figure 3-2 montre un exemple de signaux d'accélération recueillis à partir d'un smartphone placé dans quatre positions : poche du pantalon, position de la ceinture côté jambe droite, bras droit et poignet droit. L'action exécutée par la personne dans ce cas est la marche.

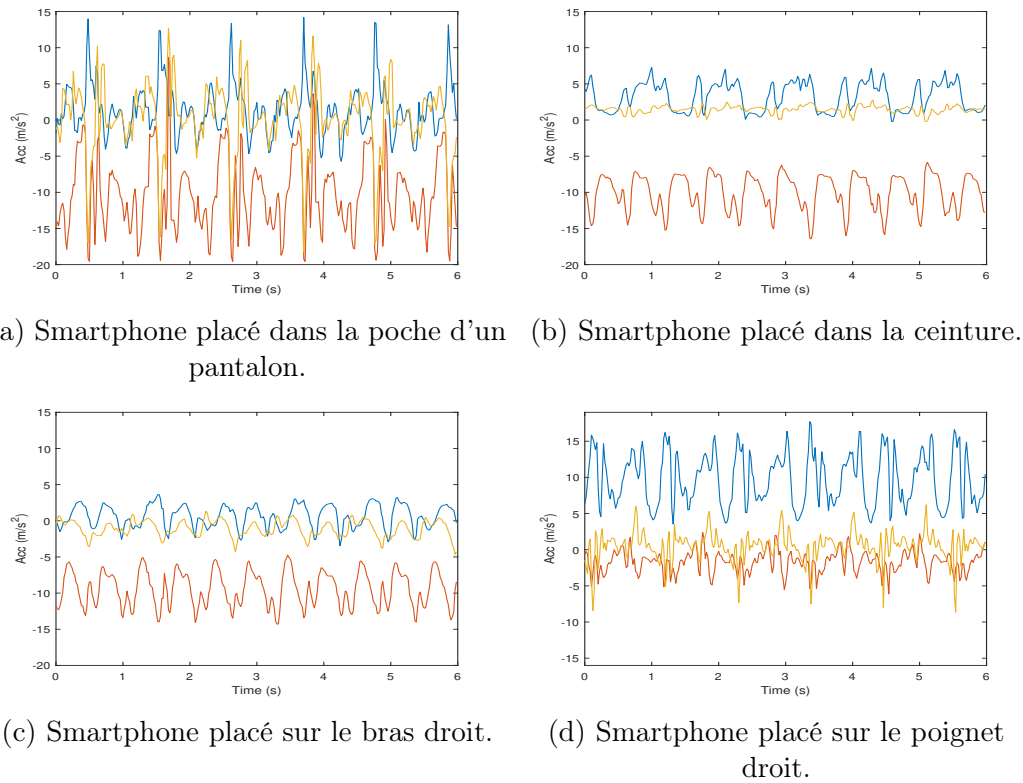


FIGURE 3-2 – Signaux d'accélération issus d'un smartphone porté à différents emplacements du corps.

Différents emplacements de smartphones ont été choisis dans la littérature, tels que la poche du pantalon (WU et al. 2012; KWAPISZ et al. 2011; SIIRTOLA et al. 2012a), la taille (ANGUITA et al. 2012; REYES-ORTIZ et al. 2016) ou la main (Y.-S. LEE et al. 2011; BAYAT et al. 2014). Cependant, pour certaines applications impliquant une surveillance des actions de la vie quotidienne (Activities of Daily Living - ADL) à long terme (étant donné que l'objectif est de mesurer en permanence le

comportement de la personne), les smartphones sont portés dans divers endroits. Par conséquent, un système basé sur les données d'un seul emplacement de smartphone n'est pas adapté. Quelques chercheurs ont essayé de répondre à cette problématique. Les travaux proposés peuvent être divisés en deux catégories :

- Approches sensibles à la position : la reconnaissance de l'action suivant K emplacements de smartphone s'effectue en deux phases : la reconnaissance de l'emplacement du smartphone dans un premier temps, suivie par la sélection d'un classifieur entraîné sur des données associées à l'emplacement reconnu.
- Approches indépendantes de la position : un seul classifieur est construit pour effectuer la classification des actions humaines. Les données de tous les emplacements de smartphone sont exploitées pour entraîner le classifieur.

Dans (SZTYLER et al. 2017), une approche sensible à la position est proposée pour reconnaître huit actions collectées à partir de sept emplacements du smartphone. Un classifieur est alors entraîné pour la reconnaissance de la position du smartphone. En fonction de la position détectée, un des sept classifieurs entraînés séparément sur des données issues de chaque emplacement est choisi. Dans les méthodes de classification, les forêts d'arbres décisionnels (Random Forest - RF) sont utilisées. Dans le travail de (MARTÍN et al. 2013), en plus des signaux inertiels, des données de capteurs de lumière et de proximité sont utilisées pour reconnaître la position du smartphone. Dans (Rong YANG et al. 2016), les attributs calculés sont ajustés en fonction de l'emplacement du téléphone. Plus récemment, les auteurs de (ALMASLUKH et al. 2018) ont proposé une approche sensible à la position et une autre indépendante à la position. Sept emplacements de smartphone sont considérés dans cette dernière étude. Dans les deux approches, les réseaux de neurones à convolution (Convolutional Neural Network - CNN) sont utilisés pour l'étape de classification. En utilisant un test indépendant du sujet, un score F -*measure* de 78% est obtenu pour la reconnaissance des actions et de 90% pour la position du smartphone.

Les approches sensibles à la position reposent sur deux niveaux de classification. Le premier (reconnaissance de la position du smartphone) est pertinent, i.e., l'erreur

de prédiction dans la localisation du smartphone implique la sélection du mauvais classifieur pour la reconnaissance des actions (classifieur pour un autre emplacement de smartphone), ce qui implique une probabilité plus élevée d'obtenir une erreur de classification de l'action. En outre, le suivi de la position d'un smartphone est une tâche difficile. Cette méthode vise à extraire des attributs discriminants à partir des signaux inertiels afin d'entraîner un classifieur de position. Alors que les actions dynamiques contiennent des attributs pertinents sur la position du smartphone, les postures ne donnent que des informations sur l'orientation du smartphone, ce qui rend la tâche difficile lors de l'exécution de ces derniers.

Dans les approches indépendantes de la position, une seule phase de classification est effectuée pour reconnaître les actions. Cette catégorie de méthodes repose sur le calcul d'attributs robustes à la position de smartphone dans le but d'obtenir des résultats satisfaisants.

3.5 Segmentation

La phase de segmentation prépare les données pour le processus d'extraction d'attributs. Elle permet de diviser les signaux inertiels en segments de données plus petits. Dans le domaine de la reconnaissance d'actions humaines avec des capteurs inertiels, les techniques de segmentation peuvent être regroupées en trois catégories : détection d'évènements, détection des changements d'actions et l'utilisation d'une fenêtre glissante. Cette dernière est la plus populaire et largement utilisée pour la reconnaissance des actions. Elle permet le découpage des signaux inertiels en temps réel. Les segments obtenus possèdent une taille identique correspondant à un temps T , avec un degré de chevauchement O_v entre deux segments adjacents. Une autre technique de segmentation est la détection d'évènements. Cette méthode est plus utilisée dans l'analyse de démarche. Elle consiste à identifier des évènements spécifiques dans les signaux tels que le contact d'un pied (coup de talon) avec le sol ou le décollage d'un pied. La dernière technique de segmentation a pour objectif de détecter le changement d'actions. Les échantillons de début et de fin des actions sont détectés en analysant

la variation des signaux dans le domaine fréquentiel en particulier.

Dans l'état de l'art, les méthodes proposées appliquent généralement une fenêtre glissante de type rectangulaire. La taille de la fenêtre est un paramètre important à fixer. Elle affecte directement les performances d'un système HAR. Dans la littérature, des fenêtres de tailles très variables sont utilisées, allant de $T = 1s$ dans (SZTYLER et al. 2017; ATTAL et al. 2015), jusqu'à $T = 30s$ dans (S. LIU et al. 2012). La taille T est généralement déterminée de manière empirique en fonction des performances du système et de ce que nous voulons identifier (la durée "caractéristique" des actions considérées). La plage dans laquelle cette taille sera recherchée pour être optimisée dépendra de l'application ciblée, en déterminant la granularité attendue. Dans (CONCEPCIÓN et al. 2017; ATALLAH et al. 2011), après avoir testé différentes tailles de fenêtres, T est fixée à 5 secondes. Les auteurs de (G. WANG et al. 2018) ont examiné l'impact de la taille de la fenêtre sur les performances de leur système. Les résultats ont montré qu'une fenêtre entre 2,5s et 3,5s offrait un compromis optimal entre la précision et la latence.

En utilisant une fenêtre courte, le système s'enclenche plus souvent avec une fenêtre large (même degré de chevauchement). Ainsi, la latence de prédiction est réduite. Cependant, une courte fenêtre ne peut fournir suffisamment d'informations pour décrire l'action exécutée, ce qui diminue la précision du système.

3.6 Pré-traitement

Après l'étape de segmentation, des filtres numériques peuvent être appliqués pour enlever les informations indésirables. Dans (REYES-ORTIZ et al. 2016; SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016; SAN-SEGUNDO, MONTERO et al. 2016; NI et al. 2016; MOHAMMAD et al. 2018), un filtre passe-bas de Butterworth a été utilisé afin d'atténuer les composantes hautes fréquences. La fréquence de coupure a été fixée à 20Hz. Dans (NOOR et al. 2017; GARCIA-CEJA et al. 2014; KHAN et al. 2010; NAM et al. 2013) une moyenne mobile a été appliquée. Notez que la moyenne est un filtre passe-bas classique. D'autres ont choisi d'utiliser le filtre médian (REYES-ORTIZ et al.

2016; NI et al. 2016; HASSAN et al. 2018; ARIF et al. 2014; SCHEURER et al. 2017).

Les techniques de pré-traitement peuvent également être utilisées pour générer des signaux discriminants à partir des signaux bruts. Les données brutes d'accélération par exemple, sont souvent séparées en composantes d'accélération du corps et la gravité. Dans ce sens, un filtre passe-bas est appliqué dans (NOOR et al. 2017; SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016; SAN-SEGUNDO, MONTERO et al. 2016; SZTYLER et al. 2017; HASSAN et al. 2018; LEUTHEUSER et al. 2013; SUAREZ et al. 2015a; WELHENGE et al. 2018), un filtre passe-haut dans (REYES-ORTIZ et al. 2016; NI et al. 2016; J.-Y. YANG et al. 2008; MACHADO et al. 2015; MIAO et al. 2015) et la transformée en ondelettes dans (N. WANG et al. 2007). Ces techniques de filtrage pourraient ajouter des informations utiles pour le système HAR, par contre, cela nécessite un temps de calcul supplémentaire pour exécuter et stocker ou transférer les données.

3.7 Extraction des attributs

Pour décrire chaque action humaine, une liste d'attributs est extraite des signaux bruts ou pré-traités. Les attributs peuvent être divisés en trois catégories : attributs du domaine temporel, du domaine fréquentiel et attributs temps-fréquence. Les tableaux 3.1, 3.2 et 3.3 présentent les attributs calculés dans l'état de l'art en fonction du domaine considéré. Les plus populaires, attributs du domaine temporel, sont calculés à partir de signaux dans leurs représentations temporelles ; aucune transformation n'est nécessaire au préalable. Comme indiqué dans le tableau, dans le domaine temporel, deux sortes d'attributs sont calculés : attributs statistiques, (nous citons par exemple des moments d'ordre $k \in \mathbb{N}$), ou encore des attributs heuristiques, ces derniers sont moins utilisés que leurs prédécesseurs. Ils sont généralement calculés dans le but de détecter des actions particulières comme par exemple la chute.

La deuxième famille d'attributs, les fréquentiels, sont extraits à partir d'une représentation fréquentielle des signaux. Pour ce faire, les données temporelles sont d'abord transformées à l'aide d'un outil mathématique de transformation. Diverses

Domaine des attributs			Références
Temporel	Statistique	Moyenne, moyenne quadratique, écart type, médiane, quartiles, asymétrie, Kurtosis, entropie, énergie, maxima et minima, coefficient de corrélation, histogramme, coefficients d'autorégression, passages par zéro, temps entre pics, nombre d'occurrences de pics	(CATAL et al. 2015 ; KWAPISZ et al. 2011 ; REYES-ORTIZ et al. 2016 ; NOOR et al. 2017 ; SZTYLER et al. 2017 ; SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016 ; ATTAL et al. 2015 ; Rong YANG et al. 2016 ; ANTOS et al. 2014 ; ALTUN et al. 2010 ; WU et al. 2012 ; SHOAIB, BOSCH, O. D. INCEL et al. 2014 ; BAO et al. 2004 ; HASSAN et al. 2018 ; NI et al. 2016 ; ZDRAVEVSKI et al. 2017 ; S. LIU et al. 2012 ; LEUTHEUSER et al. 2013 ; CASALE et al. 2015 ; MANNINI et SABATINI 2010 ; CLELAND et al. 2013 ; C. HU et al. 2018 ; KHAN et al. 2010 ; BAYAT et al. 2014 ; SIIRTOLA et al. 2012a ; N. RAVI et al. 2005 ; LIANG et al. 2014 ; GUPTA et al. 2014 ; DENG et al. 2014 ; MACHADO et al. 2015 ; SUAREZ et al. 2015b ; NAM et al. 2013 ; SUN et al. 2010 ; DERNBACH et al. 2012 ; MIAO et al. 2015 ; MARTÍN et al. 2013 ; Y. CHEN et al. 2017 ; GU et al. 2015 ; PEI et al. 2013 ; ARIF et al. 2014 ; D. RAVI et al. 2017 ; K. LEE et al. 2018 ; SHEN et al. 2016 ; O. D. INCEL 2015 ; A. WANG et al. 2016 ; ANGUITA et al. 2012 ; RONAO et al. 2017 ; LIN et al. 2012 ; LONG et al. 2009 ; Z. ZHAO et al. 2011 ; SCHEURER et al. 2017 ; JANIDARMIAN et al. 2017 ; BANOS, VILLALONGA et al. 2015)
	Heuristique	Angle d'inclinaison	(NI et al. 2016 ; KHAN et al. 2010 ; LIANG et al. 2014)

TABLE 3.1 – Les attributs calculés dans l'état de l'art à partir du domaine temporel des signaux inertiels.

transformées existent. La plus populaire est la transformée de Fourier \mathcal{F} . Elle permet de décrire le spectre fréquentiel d'une fonction intégrable sur \mathbb{R} suivant la formule :

$$\mathcal{F}(x) : f \mapsto X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} dt \quad (3.1)$$

Dans notre cas, puisque nous exploitons des signaux échantillonnés (domaine discret), nous devons considérer alors la transformée de Fourier discrète (TFD) définie

comme suit :

$$\mathbf{X}(k) = \sum_{n=0}^{N-1} x(n)e^{-2\pi jk \frac{n}{N}} \quad \text{avec } 0 \leq k < N - 1 \quad (3.2)$$

L'implémentation directe de la formule 3.2 peut s'avérer très coûteuse en temps de calcul. En pratique, un algorithme appelé transformation de Fourier rapide (FFT) (COOLEY et al. 1965) est utilisé pour le calcul de la DFT. Ce dernier est doté d'une complexité qui varie en $O(n \log n)$ plutôt que $O(n^2)$ en utilisant la formule 3.2. À partir de la DFT obtenue, la densité spectrale d'énergie (Energy Spectral Density - ESD) des signaux inertiels est calculée comme suit :

$$\text{ESD} \triangleq |X(f)|^2 = X(f)X^*(f) \quad (3.3)$$

Ainsi, les attributs fréquentiels peuvent être extraits de trois manières. La plus fréquente est le calcul de mesures statistiques sur le spectre obtenu, de la même façon que l'extraction de certains attributs temporels. La deuxième approche est le calcul de mesures utilisées en traitement du signal pour caractériser le spectre comme l'énergie ou la puissance d'un signal. Notez qu'il est possible de calculer ces dernières directement à partir du domaine temporel, et cela ne nécessite pas forcément le passage vers le domaine fréquentiel (méthode plus coûteuse en temps de calcul). Enfin, la troisième approche utilisée pour extraire les attributs fréquentiels est l'exploitation directe du spectre du signal, en gardant les distributions d'énergie à certaines fréquences. Dans les travaux de (D. RAVI et al. 2017), les spectres des signaux inertiels sont calculés sur des fenêtres temporelles de 4s pour certaines bases de données et 10s pour d'autres. Le vecteur d'attributs est construit à partir de la concaténation des spectres obtenus combinés avec des attributs temporels statistiques. Dans (SHEN et al. 2016; Y. CHEN et al. 2017), les huit premiers coefficients du spectre après avoir retiré la composante continue, sont considérés comme étant le vecteur d'attributs. Dans (C. HU et al. 2018; O. D. INCEL 2015) cinq coefficients seulement sont pris en compte.

La dernière famille d'attributs est extraite d'une représentation temps-fréquence du signal. La transformée en ondelettes est l'outil mathématique le plus utilisé pour

Domaine des attributs			Références
Fréquentiel (ESD)	Brute	Spectre	(C. HU et al. 2018; Y. CHEN et al. 2017; D. RAVI et al. 2017; JIANG et al. 2015; SHEN et al. 2016; O. D. INCEL 2015)
	Traitement du signal	Énergie, puissance, centroïde spectral	(LEUTHEUSER et al. 2013; CASALE et al. 2015; CLELAND et al. 2013; N. RAVI et al. 2005; BAO et al. 2004; GUPTA et al. 2014; MACHADO et al. 2015; NAM et al. 2013; SUN et al. 2010; MARTÍN et al. 2013; K. LEE et al. 2018; O. D. INCEL 2015; A. WANG et al. 2016; RONAO et al. 2017; Z. ZHAO et al. 2011; SCHEURER et al. 2017)
	Statistique	Mode, entropie, maximum, asymétrie, kurtosis	(SAN-SEGUNDO, MONTERO et al. 2016; REYES-ORTIZ et al. 2016; NOOR et al. 2017; SZTYLER et al. 2017; SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016; ATTAL et al. 2015; Rong YANG et al. 2016; ANTOS et al. 2014; ALTUN et al. 2010; WU et al. 2012; SHOAIB, BOSCH, O. D. INCEL et al. 2014; BAO et al. 2004; HASSAN et al. 2018; S. ZHANG et al. 2012; NI et al. 2016; ZDRAVEVSKI et al. 2017; S. LIU et al. 2012; MANNINI et SABATINI 2010; C. HU et al. 2018; DENG et al. 2014; MACHADO et al. 2015; SUN et al. 2010; MARTÍN et al. 2013; PEI et al. 2013; K. LEE et al. 2018; A. WANG et al. 2016; RONAO et al. 2017; LONG et al. 2009; SAEZ et al. 2016; PEI et al. 2013)

TABLE 3.2 – Les attributs calculés dans l’état de l’art à partir du domaine fréquentiel des signaux inertiels.

transformer le signal tout en conservant son information temporelle et fréquentielle simultanément. Ce dernier est projeté sur un ensemble de fonctions de base qui varient en fréquence et en espace. La transformée en ondelettes présente une bonne résolution fréquentielle pour les basses fréquences et une bonne résolution temporelle pour les hautes fréquences (principe d’incertitude d’Heisenberg).

En traitement numérique du signal la transformée en ondelettes discrète (Discrete Wavelet Transform - DWT) est utilisée. Dans ce cas, le signal passe par une série de

filtres numériques. Des filtres passe-haut pour analyser les hautes fréquences et une série de filtres passe-bas pour analyser les basses fréquences. Il est à noter qu’une variante de la DWT, la transformation par Paquets d’ondelettes (Wavelet Packet Decomposition - WPD) est aussi exploitée. Ces transformations permettent d’obtenir deux ensembles de coefficients : les coefficients d’approximation et les coefficients de détail. Des attributs comme l’énergie ou l’écart type sont ensuite calculés à partir de ces coefficients.

Domaine des attributs			Références
Temps-fréquence	WPD	Énergie des coefficients de détails, écart type des coefficients de détails, écart type des coefficients d’approximation	(ATTAL et al. 2015 ; N. WANG et al. 2007)
	DWT	Énergie des coefficients de détail, énergie des coefficients d’approximation	(S. ZHANG et al. 2012 ; MANNINI, ROSENBERGER et al. 2017 ; MITCHELL et al. 2013 ; Y. CHEN et al. 2017 ; SHEN et al. 2016 ; MANNINI, INTILLE et al. 2013)

TABLE 3.3 – Les attributs calculés dans l’état de l’art à partir du domaine temps-fréquence des signaux inertiels.

Afin d’obtenir une meilleure précision, les attributs extraits du domaine fréquentiel sont généralement combinés aux attributs temporels, menant généralement à un vecteur d’attributs de grande dimension. Cette technique peut s’avérer coûteuse en terme de temps de calcul et de consommation d’énergie.

3.8 Réduction de la dimensionnalité des attributs

L’étape d’extraction d’attributs est très coûteuse en temps de calcul, certains d’entre eux sont redondants ou non discriminatoires. Pour cela, des techniques de réduction de la dimensionnalité sont utilisées pour réduire le nombre d’attributs à

calculer. Nous pouvons en citer deux types de techniques : la sélection des attributs ou la projection des attributs. La première technique permet d’identifier et de supprimer les attributs inutiles et non pertinents de l’ensemble d’origine. Elles sont divisées en trois catégories : les méthodes de filtre, les méthodes d’enveloppe et les méthodes embarquées. Dans (NI et al. 2016), trois méthodes de sélection d’attributs de type filtres ont été considérées : le gain d’information, la sélection basée sur la corrélation et la méthode nommée Relief. Avec ces techniques, 21 attributs ont été sélectionnés sur les 77 à l’origine. Dans (Y. CHEN et al. 2017), un test Kolmogorov-Smirnov est effectué pour sélectionner 68 attributs sur un ensemble de 89. En projection d’attributs, la création de nouveaux attributs non corrélés est examinée, une cartographie des entités originales dans un espace de dimension inférieure est ensuite effectuée. Dans (HE et al. 2009 ; LONG et al. 2009 ; ALTUN et al. 2010 ; JANIDARMIAN et al. 2017), l’analyse en composantes principales (ACP) est utilisée à cette fin.

3.9 Classification des actions humaines

Pour le processus de classification, diverses méthodes de classification ont été utilisées dans l’état de l’art. Le tableau 3.4 présente les classifieurs les plus exploités. Des méthodes traditionnelles telles que les k-plus proches voisins (k Nearest Neighbors - kNN) ou arbre de décision ont été utilisées dans la littérature. Plus récemment, des réseaux de neurones profonds (Deep Neural Network - DNN) ont été appliqués, principalement des réseaux de neurones à convolution (CNN).

Plusieurs facteurs peuvent influencer le choix d’un classifieur (type de données, précision, complexité du classifieur, réponse en temps réel, etc.). Dans la littérature HAR, une évaluation comparative de divers algorithmes de classifieurs est largement effectuée. L’environnement Weka (WITTEN et al. 2016) est l’un des logiciels permettant d’explorer et de comparer les algorithmes traditionnels et de les évaluer facilement pour un ensemble de données. Le score de précision est généralement le critère utilisé pour la comparaison des classifieurs. Dans ces études comparatives (voir le tableau 3.4), quelques travaux ont considéré la stratégie de validation indépendante du su-

Méthode de classification	Études pertinentes	
k-plus proches voisins	(ATALLAH et al. 2011; GUPTA et al. 2014; A. WANG et al. 2016)	
Machine à vecteurs de support	Noyau linéaire	(REYES-ORTIZ et al. 2016)
	Noyau RBF	(MANNINI, ROSENBERGER et al. 2017; SUN et al. 2010; Rong YANG et al. 2016; ANGUIA et al. 2012)
Arbres de décision	(S. ZHANG et al. 2012; LIN et al. 2012)	
Naïve bayésienne	(A. WANG et al. 2016; LONG et al. 2009)	
Modèles de Markov cachés	H-HMM	(Y.-S. LEE et al. 2011; RONAO et al. 2017)
	HMMs- GMMs ¹	(SAN-SEGUNDO, MONTERO et al. 2016; SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016)
	D-HMMs	(Y.-J. KIM et al. 2015)
Réseau de neurones	(J.-Y. YANG et al. 2008; KHAN et al. 2010; N. WANG et al. 2007)	
Forêts d'arbres décisionnels	(K. LEE et al. 2018)	
Apprentissage profond	CNN ²	(RONAO et al. 2016; IGNATOV 2018; JIANG et al. 2015; MOHAMMAD et al. 2018; JORDAO et al. 2018; J. YANG et al. 2015; ALMASLUKH et al. 2018; CHO et al. 2018; HA et al. 2016)
	DBN ³	(HASSAN et al. 2018)
	RNN ⁴	(MURAD et al. 2017; WELHENGE et al. 2018)
Évaluation individuelle de plusieurs classifieurs	Weka (WITTEN et al. 2016)	(WU et al. 2012; BAYAT et al. 2014; N. RAVI et al. 2005; BAO et al. 2004; KWAPISZ et al. 2011; DERNBACH et al. 2012; ALTUN et al. 2010; MIAO et al. 2015; SIIRTOLA et al. 2012a; SHOAI B, BOSCH, O. D. INCEL et al. 2014; MARTÍN et al. 2013; GU et al. 2015; NAM et al. 2013; ARIF et al. 2014; VAVOULAS et al. 2016; O. D. INCEL 2015; TAPIA et al. 2007; BANOS, VILLALONGA et al. 2015)
	Scikit-learn (PEDREGOSA et al. 2011)	(ZDRAVEVSKI et al. 2017; SAEZ et al. 2016)
Fusion de décisions de plusieurs classifieurs	(ANTOS et al. 2014; CATAL et al. 2015)	

¹ Gaussian mixture models.

² Convolutional Neural Network.

³ Deep Belief Network.

⁴ Recurrent Neural Network.

TABLE 3.4 – Algorithmes de classification utilisés pour la reconnaissance de l'action humaine.

jet. Dans (ALTUN et al. 2010), cinq capteurs portables ont été utilisés pour collecter des signaux inertiels. Les machines à vecteurs de support (Support vector machines - SVMs), les réseaux de neurones artificiels (Artificial neural network - ANN), la déformation temporelle dynamique (Dynamic time warping - DTW), la prise de décision bayésienne, la méthode des moindres carrés, kNN et les arbres de décision ont été comparés à l'aide de différentes techniques de validation, notamment une stratégie de validation indépendante du sujet. Les résultats obtenus par (ALTUN et al. 2010) indiquent que le classifieur SVM a atteint le taux de classification le plus élevé par rapport aux autres techniques de classification. Afin de choisir un algorithme de classification, les auteurs de (CHOWDHURY et al. 2018) ont expérimenté les classifieurs suivants : SVM, RF, arbres de décision, Adaboost et DNN. Deux ensembles de données publiques, PAMAP2 (REISS et al. 2012) et mHealth (BANOS, GARCIA et al. 2014), contiennent les données d'un accéléromètre triaxial recueillies à trois endroits différents (cheville, poitrine et poignet). Une comparaison des classifieurs est effectuée pour chaque emplacement en fonction de la précision, une stratégie de validation indépendante du sujet est utilisée à cette fin. Les résultats de (CHOWDHURY et al. 2018) indiquent que les SVMs et RF offrent de meilleures performances pour tous les emplacements des deux jeux de données. Cependant, les SVMs fonctionnent mieux que les RF en faisant la moyenne des performances sur tous les emplacements et les deux jeux de données. Par conséquent, le classifieur SVM est sélectionné. Dans (N. RAVI et al. 2005), les classifieurs Bagging et Boosting ont été comparés. Différentes stratégies de validation ont été testées, notamment le test suivant : les données collectées d'un sujet pendant une journée ont été utilisées pour l'entraînement et les données collectées d'un autre sujet pendant une autre journée ont été utilisées pour les tests. Les résultats de (N. RAVI et al. 2005) montrent que les SVMs boostés offrent les meilleures performances.

3.10 Classification des transitions posturales

Outre les actions humaines classiques, certains chercheurs ont utilisé des capteurs inertiels pour détecter les transitions entre les actions (KHAN et al. 2010 ; MATHIE et al. 2004 ; GUPTA et al. 2014 ; KARANTONIS et al. 2006). Cependant, peu de travaux ont utilisé des smartphones pour la détection des transitions. Dans (REYES-ORTIZ et al. 2016), deux approches différentes sont proposées pour détecter les transitions posturales : dans la première approche, seules les actions de base sont apprises par le classifieur, les transitions, en revanche, sont considérées comme un changement d’actions détectées par le système. Dans la seconde approche, en plus des actions de base, les transitions sont apprises par le classifieur. Elles sont considérées comme une classe supplémentaire. Dans (NOOR et al. 2017), le premier bloc du système proposé comprend un détecteur de transition. Il distingue les transitions posturales des actions de base. La transition détectée est ensuite mise à l’entrée d’un classifieur de transition afin de reconnaître la classe de la transition détectée. Dans les deux processus, un arbre de décision est utilisé pour la classification. Dans (HASSAN et al. 2018), le même processus est appliqué aux actions de base et aux transitions. Un réseau de croyances profond (Deep Belief Network -DBN) est ensuite utilisé pour la classification.

3.11 Évaluation des performances

Afin de démontrer l’efficacité de l’approche proposée ou de la comparer aux méthodes existantes, une évaluation est effectuée sur un jeu de données. Cette tâche peut être divisée en deux étapes : les techniques de validation du modèle et le calcul des métriques. La validation du modèle vise à reproduire le comportement du système avec suffisamment de fidélité. Par conséquent, le jeu de données est divisé en jeu d’apprentissage pour former le modèle et le jeu de test pour la validation. Dans la littérature, “ k -fold cross-validation” (validation croisée à k partitions) est une technique populaire pour la validation du modèle. Le jeu de données est partitionné de manière aléatoire

en k sous-ensembles de taille égale. $k - 1$ sous-ensembles sont utilisés pour l'apprentissage et le sous-ensemble restant est utilisé pour le test. Le processus est ensuite répété k fois pour prendre chaque sous-ensemble une fois pour le test. Les résultats obtenus peuvent être moyennés pour produire un score. La méthode "Hold-out" est également populaire dans la littérature. Elle consiste à découper le jeu de données aléatoirement en deux sous-ensembles, un pour l'apprentissage et l'autre pour le test. Cette méthode est généralement appliquée lorsqu'un système nécessite beaucoup de calcul pour construire le modèle. Les algorithmes d'apprentissage profond sont les plus concernés par cette technique.

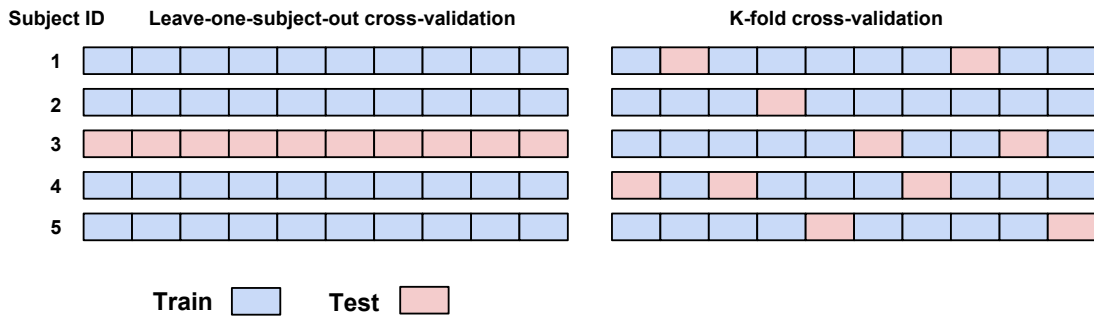


FIGURE 3-3 – Visualisation de LOSO Cross-Validation vs k-fold cross-validation.

Selon la méthode de partitionnement du jeu de données, une validation dépendante ou indépendante du sujet peut être réalisée. Dans la stratégie de validation indépendante du sujet, les sous-ensembles d'apprentissage et de test ne doivent pas contenir des données issues du même sujet. Cette validation montre comment le modèle répond lorsqu'il est testé sur un sujet différent. Il donne des résultats plus proches d'une situation pratique et permet une analyse plus objective que la validation dépendant du sujet. La validation croisée simple (k -fold cross-validation) est alors remplacée par la validation croisée suivant le sujet (k -fold-subject-cross-validation), ou par leave-one-subject-out-cross-validation (LOSOCV) quand k est égal au nombre de sujets. La validation indépendante du sujet la plus simple est le découpage du jeu de données en deux sous-ensembles avec différents sujets chacun. Cette méthode est nommée Hold-out subjects dans la phase expérimentale.

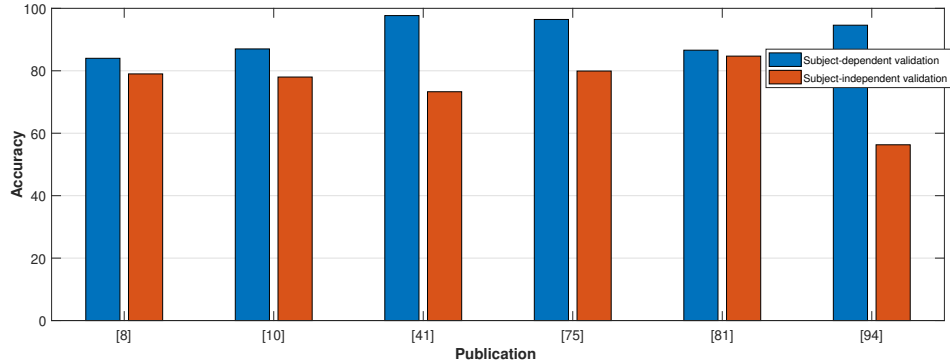


FIGURE 3-4 – Exemples des performances obtenues à l’aide d’une validation dépendante et indépendante du sujet.

La figure 3-3 illustre un exemple de fractionnement d’un jeu de données à l’aide de LOSOCV et 10-fold cross-validation. Chaque rectangle de l’illustration représente une instance. Nous pouvons voir que dans la validation croisée, il est très courant d’avoir une instance utilisée pour le test et des instances adjacentes pour l’apprentissage. Comme indiqué ci-dessus, dans la plupart des méthodes HAR proposées dans l’état de l’art, un chevauchement entre des fenêtres adjacentes est effectué. Un degré de chevauchement de 50% est généralement choisi. Par conséquent, pour la validation croisée simple, les sous-ensembles apprentissage et de test partagent des échantillons, ce qui rend les sous-ensembles non indépendants. La figure 3-4 présente des performances obtenues dans la littérature où des stratégies indépendantes et dépendantes du sujet sont appliquées dans les mêmes conditions (même approche et même jeu de données). Elle montre que les scores de précision pour les stratégies dépendantes au sujet surpassent les scores pour les stratégies indépendantes au sujet.

3.12 Conclusion

Dans ce chapitre consacré à l’état de l’art, nous avons présenté les principaux blocs de traitement formant un système HAR. En se basant sur les différentes méthodes proposées par les chercheurs, nous avons identifié six blocs principaux qui sont : la collecte de données, la segmentation, le pré-traitement, l’extraction des attributs, les algorithmes de sélection des attributs et les méthodes de classification. Nous avons

décrit et critiqué les différents choix de la littérature à l'intérieur de chaque bloc. Ces critiques vont servir par la suite à justifier les décisions prises dans la construction de notre approche et notamment dans le choix des paramètres de réglage des algorithmes ou des éléments de traitement du signal.

Chapitre 4

Classification hiérarchique des actions humaines à l'aide d'un smartphone

Contents

4.1	Introduction	86
4.2	Conception de notre système HAR	86
4.2.1	Construction du vecteur d'attributs	89
4.2.2	Classification hiérarchique des actions humaines	98
4.2.3	Classification des transitions posturales	100
4.3	Validation expérimentale	103
4.3.1	Jeux de données	105
4.3.2	Estimation des performances	109
4.3.3	Résultats et comparaison avec les travaux de l'état de l'art	112
4.3.4	Temps de calcul	136
4.4	Suivi d'actions sur le long-terme	136
4.4.1	Résultats expérimentaux	140
4.5	Conclusion	143

4.1 Introduction

Dans ce chapitre, nous allons présenter notre système de reconnaissance d’actions humaines (HAR pour Human Activity Recognition) à l’aide d’un smartphone. Ce chapitre est organisé en trois sections qui suivent cette introduction. La première section décrit en détail notre approche proposée. Afin de faire le parallèle avec le chapitre de l’état de l’art, la section est décomposée en trois grandes étapes : la construction du vecteur d’attributs, la classification hiérarchique des actions humaines et la classification des transitions posturales. La seconde section est consacrée à la validation expérimentale. Elle comprend quatre étapes : les jeux de données utilisés dans l’expérimentation, les méthodes d’estimation des performances utilisées pour notre système, les résultats obtenus ainsi qu’une comparaison avec les travaux de l’état de l’art, et pour finir les performances de notre approche en terme de temps de calcul. La troisième section est consacrée à la conclusion.

4.2 Conception de notre système HAR

Dans notre approche, l’objectif est de construire un système de reconnaissance des actions journalières qui répond aux différents challenges cités dans le chapitre 2 : une précision élevée, un temps de calcul faible et une invariance à la position du smartphone. Suivant l’architecture d’un système HAR, nous nous sommes focalisés principalement sur deux aspects : la construction du vecteur d’attributs et la méthode de classification.

Dans notre démarche, trois catégories d’actions humaines sont étudiées : postures, actions périodiques (ou quasi-périodiques) et actions non périodiques. Les postures représentent les comportements sédentaires de la personne durant la journée. Trois classes d’actions sont prises en considération : “debout”, “assis” et “allongé”. Les actions quasi-périodiques correspondent aux mouvements dynamiques affichant une périodicité irrégulière. Les actions suivantes sont considérées dans notre étude : marcher, monter et descendre les escaliers, courir et cyclisme. Les actions non périodiques sont

des mouvements très courts par rapport aux autres catégories d'actions. Trois actions non périodiques sont considérées dans notre travail : transitions posturales, sauter et la chute de la personne.

Les postures, les actions périodiques et non périodiques sont des groupes d'actions homogènes de natures différentes, i.e., les classes à l'intérieur de chaque groupe présentent des caractéristiques similaires et peuvent être difficiles à discriminer. C'est le cas des actions "marcher" et "monter les escaliers" par exemple. En revanche, la problématique de distinguer deux classes issues de groupes différents est plus facile à résoudre ("marcher" et "assis" par exemple).

Par conséquent, dans notre travail, les actions sont d'abord décomposées en trois méta-classes dans un premier niveau de classification. Cela correspond aux trois catégories d'actions citées ci-dessus. Par la suite, un vecteur d'attributs adaptés est calculé suivant le groupe d'actions détectées, suivi par une classification à l'aide d'un modèle entraîné sur les attributs sélectionnés comme le montre la figure 4-1. Cette approche est motivée par le fait que les trois problèmes de classifications qui découlent des trois groupes d'actions sont différents les uns des autres en terme de complexité, i.e., un vecteur d'attributs de dimension moins importante est demandé en général afin de classifier les postures contrairement aux actions quasi-périodiques par exemple. Dans les travaux de l'état de l'art (partie 3.7), nous avons montré que pour toutes les actions, un vecteur d'attributs de nature identique et de dimension très élevée est habituellement calculé. Par conséquent, le temps de calcul peut s'avérer très élevé et contraignant pour une application mobile. De ce fait, dans notre approche, les trois problèmes de classifications relatifs aux trois catégories d'actions sont traités séparément. Des attributs optimaux et adaptés sont identifiés et calculés suivant chaque groupe d'actions, pour cela, des techniques de sélections d'attributs sont appliquées afin de déterminer les vecteurs d'attributs adéquats.

Comme il est mentionné ci-dessus, trois classes d'actions sont considérées dans notre étude, y compris les transitions posturales. Ces dernières sont labellisées comme telles à la sortie du classifieur en charge des actions. Dans notre système, en plus de détecter les transitions, un bloc de classification est proposé afin de les reconnaître.

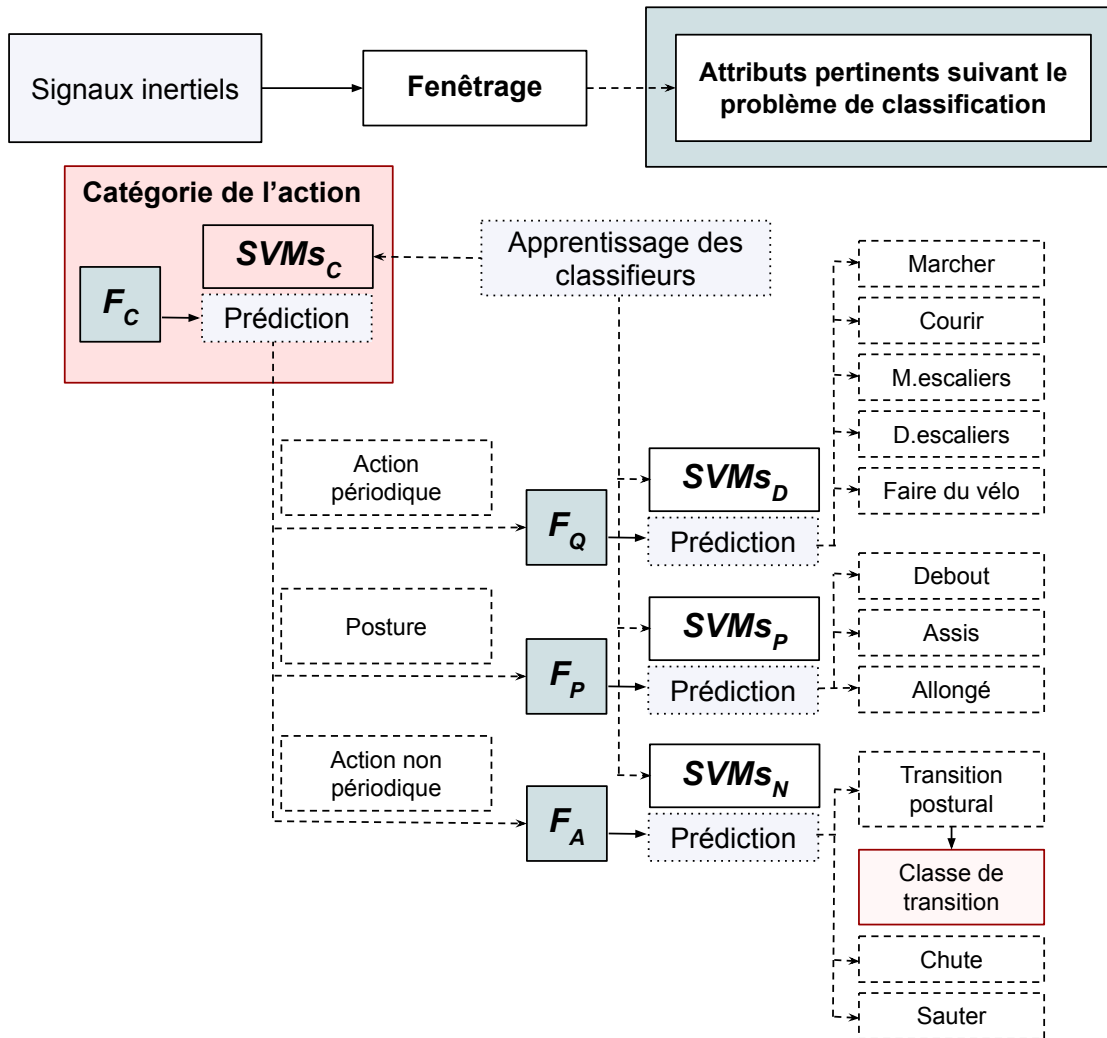


FIGURE 4-1 – Aperçu de l’approche proposée pour la reconnaissance d’actions à l’aide de capteurs inertiels.

Les classes suivantes sont considérées dans notre étude : “debout vers assis”, “assis vers debout”, “allongé vers assis”, “allongé vers debout”, “debout vers allongé”, “assis vers allongé”.

Dans ce qui suit et afin de faire le parallèle avec les travaux de l’état de l’art évoqués dans la section 3.2, chaque étape de notre approche est présentée individuellement : construction du vecteur d’attributs, classification hiérarchique des actions humaines et classification des transitions posturales.

4.2.1 Construction du vecteur d'attributs

Dans l'état de l'art, nous avons constaté que l'étape d'extraction des attributs se fait d'une façon arbitraire. Dans notre méthode, nous nous sommes focalisés sur le design d'un vecteur d'attributs optimal et pertinent selon le problème de classification rencontré. Pour décrire chaque action humaine, les attributs sont extraits de signaux inertiels provenant d'un accéléromètre triaxial et d'un gyroscope triaxial. Les mesures du champ magnétique issues d'un magnétomètre triaxial sont écartées dans notre système car elles ont une forte probabilité d'être affectées par le bruit de l'environnement. La magnitude des signaux définis comme suit est également prise en compte dans cette étude.

$$\mathbf{S}_m = \sqrt{S_x^2 + S_y^2 + S_z^2} \quad (4.1)$$

Tous les smartphones ne contiennent pas de gyroscope (les premiers modèles contiennent seulement des accéléromètres). Par conséquent, selon la disponibilité du capteur, quatre ou huit signaux sont impliqués dans le processus de construction des attributs.

Comme indiqué dans la partie état de l'art, les attributs peuvent appartenir à trois domaines : temporel, fréquentiel et temps-fréquence. Les attributs temps-fréquence, extraits d'une représentation temps-fréquence du signal sont peu utilisés dans l'état de l'art comme indiquée dans la partie 3.7, et ils n'ont pas vraiment montré un réel pouvoir discriminant comparés aux autres familles d'attributs pour le problème de reconnaissance d'actions/d'activités. De ce fait, dans notre travail, les attributs temps-fréquence sont écartés, ce qui nous laisse les attributs temporels, extraits directement du signal original, et les attributs fréquentiels, extraits à partir d'une représentation fréquentielle du signal.

Dans notre démarche, comme il apparaît dans la figure 4-2, afin de reconnaître la catégorie de l'action, un premier vecteur d'attributs F_C est d'abord construit dans le premier niveau de classification. Il est suivi par F_Q si l'action identifiée est de type quasi-périodique, F_P si l'action est une posture, et F_A si c'est une action non péri-

dique. En ce qui concerne la nature des attributs pour chaque vecteur, nous pouvons suivre le schéma illustré dans figure 4-2 : le vecteur F_Q est construit à partir d'attributs fréquentiels. Ce choix est justifié par le fait que les actions quasi-périodiques sont caractérisées par des signaux contenant des propriétés ressemblant à celles des fonctions périodiques. Par conséquent, l'extraction des attributs à partir d'une représentation fréquentielle peut s'avérer très pertinente. Pour les autres vecteurs F_C , F_P et F_A , les attributs sont calculés directement à partir du signal temporel sous forme de mesures statistiques.

Pour les deux familles d'attributs, un large panel de mesures est utilisé dans la littérature (revoir les tableaux 3.1 et 3.2). Afin d'identifier les attributs les plus pertinents, des algorithmes de sélection d'attributs sont appliqués. Ceci nous permet d'écartier certains attributs et ainsi diminuer la complexité du modèle, réduire le temps de calcul, réduire la consommation des ressources (batterie du smartphone), et enfin dans certains cas, augmenter la performance du système.

Dans les parties qui suivent, la construction des vecteurs d'attributs pour chaque catégorie d'actions est développée.

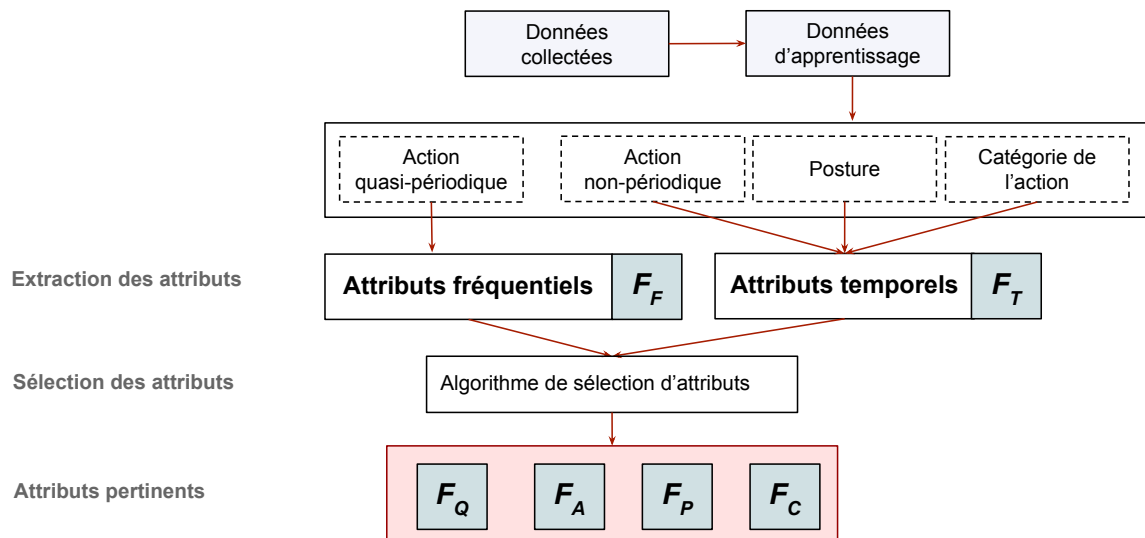


FIGURE 4-2 – Méthode d'extraction des attributs suivie dans notre approche.

Fenêtrage

La phase de segmentation prépare les données pour le processus d'extraction d'attributs. Par conséquent, une fenêtre glissante est utilisée afin de découper les signaux inertiels en segments de taille T , avec un degré de chevauchement O_v entre deux fenêtres adjacentes. En plus de ces deux paramètres cités, un autre paramètre souvent négligé dans la littérature doit être bien choisi : il s'agit de la forme de la fenêtre. Dans les systèmes HAR proposés dans l'état de l'art, une fenêtre de type rectangulaire est généralement appliquée. Il est à noter que la forme de la fenêtre utilisée a une incidence sur la représentation fréquentielle.

La forme rectangulaire présente un facteur de fuite élevé (rapport de la puissance dans les lobes latéraux à la puissance totale de la fenêtre). Dans le système que nous proposons, la fenêtre de Hamming (équation 4.2), très utilisée dans le traitement du signal de la parole, est exploitée.

$$w(n) = 0,54 - 0,46 \cos\left(\frac{2\pi n}{M-1}\right) \quad 0 \leq n \leq M-1 \quad (4.2)$$

Extraction des attributs temporels

Comme il a été indiqué dans la partie 3.7 de l'état de l'art, les attributs extraits à partir des signaux temporels sont les plus utilisés dans la littérature. Nous pouvons les diviser en deux catégories : les attributs statistiques et les attributs heuristiques. Traditionnellement, les attributs heuristiques sont exploités dans des applications particulières comme par exemple la détection de chute. Comme nous ne nous intéressons pas à une action en particulier, et puisque toutes les actions ont le même degré d'importance dans notre système, les attributs heuristiques sont donc écartés de notre système. Ce qui nous laisse les attributs statistiques pour la construction des vecteurs F_C , F_P et F_A .

En considérant les mesures statistiques calculées dans les travaux de l'état de l'art, une liste d'attributs est sélectionnée et listée dans le tableau 4.1. Remarquez que certains attributs sont écartés comme les coefficients d'autorégression car leur

Attribut	Formule et description
Moyenne	$\bar{s} = \frac{1}{N} \sum_{i=1}^N s_i$
Moyenne quadratique	$s_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N s_i^2}$
Écart type	$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (s_i - \bar{s})^2}$
Asymétrie	$\gamma_1 = E \left[\left(\frac{s-\bar{s}}{\sigma} \right)^3 \right]$
Kurtosis	$\beta_2 = E \left[\left(\frac{s-\bar{s}}{\sigma} \right)^4 \right]$
Amplitude maximum	$max(s)$
Amplitude minimum	$min(s)$
Peak-to-Peak	$max(s) - min(s)$
Énergie	$E_s = \sum_{i=1}^N s_i ^2$
Médiane	Valeur médiane du signal
Variance	$\sigma = \frac{1}{N-1} \sum_{i=1}^N (s_i - \bar{s})^2$
Premier quartile	$Q_1(s)$
Troisième quartile	$Q_3(s)$
Écart inter-quartile	$iqr = Q_3(s) - Q_1(s)$
Peak-magnitude-to-RMS	Le rapport de la plus grande valeur absolue du signal à la moyenne quadratique
Coefficient de corrélation	$r = \frac{Cov(s_j, s_k)}{\sigma_{s_j} \sigma_{s_k}}$
Nombre de passage par zéro	$zcr = \frac{1}{N} \sum_{i=1}^{N-1} \mathbb{I} \{s_i s_{i-1} < 0\}$
Centiles	$10^e, 20^e, 30^e, 40^e, 60^e, 70^e, 80^e, 90^e$

TABLE 4.1 – Attributs temporels calculés afin de construire le vecteur F_T .

temps de calcul est assez important, comparé à celui des autres attributs.

Nous obtenons alors un vecteur F_T de dimension $dim(F_T) = (24 * 4 + 3) * N_{cap}$,

cela correspond à 24 attributs extraits par signal, 4 signaux fournis par capteurs (signaux suivant les trois axes x , y , z et leur magnitude), 3 coefficients de corrélations entre les signaux, et N_{cap} le nombre de capteurs utilisés. En utilisant par exemple un accéléromètre et un gyroscope, nous obtenons $dim(F_T) = 198$. Le vecteur F_T construit sera par la suite mis à l'entrée d'un algorithme de sélection d'attributs afin de choisir les attributs les plus pertinents.

Extraction des attributs fréquentiels

Dans notre approche, les attributs sont extraits du domaine fréquentiel pour les actions périodiques. L'algorithme 1 est donc appliqué dans cette finalité : pour chaque capteur, les signaux inertiels suivant les trois axes (S_x, S_y, S_z) et la magnitude du signal (S_m) sont d'abord segmentés par la fenêtre de Hamming. La transformée DFT est ensuite obtenue pour chaque signal (nommée S_f) en utilisant l'algorithme de la FFT. Enfin, la densité spectrale d'énergie (ESD) contenant la distribution d'énergie de chaque fréquence est calculée.

Notre vecteur d'attributs est construit à partir des coefficients de l'ESD pertinents pour la reconnaissance d'actions. Pour ce faire un algorithme de sélection d'attributs est appliqué. La démarche est expliquée en détail ci-dessous.

Sélection des attributs pertinents

Comme il est présenté ci-dessus, quatre vecteurs d'attributs sont construits afin de répondre à quatre problématiques de classification différentes, à savoir, la détermination de la catégorie de l'action dans un premier niveau de classification, suivie par la classification séparée des actions quasi-périodiques, postures et actions non périodiques. Dans la littérature, afin de reconnaître ces mêmes types d'actions, un seul vecteur d'attributs est calculé dans la plupart des cas. Ce vecteur possède généralement une taille très importante (par exemple \mathbb{R}^{561} dans les travaux de (REYES-ORTIZ et al. 2016; HASSAN et al. 2018; SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016; DENG et al. 2014)). La construction d'un tel vecteur peut s'avérer coûteuse en terme de temps de calcul et de consommation d'énergie. Pour y remédier, d'autres chercheurs

Algorithm 1: Algorithme de construction du vecteur d'attributs F_F .

Data:

S : Signaux issus d'un capteur // (S_x, S_y, S_z)

$S_m \leftarrow \sqrt{S_x^2 + S_y^2 + S_z^2}$; // Calcul de la magnitude du signal

$F_F \leftarrow \text{Feature_construction}(S, S_m)$;

Function $\text{Feature_construction}(S, S_m)$

$w = 0,54 - 0,46 \cos\left(\frac{2\pi n}{M-1}\right)$; // Fenêtre de Hamming

$F_k \leftarrow \text{Spectrum}(S_k \cdot w)$; // $k \in \{x, y, z\}$ (Spectres de S_x, S_y et S_z)

$F_m \leftarrow \text{Spectrum}(S_m \cdot w)$;

$F_F \leftarrow [F_k \ F_m]$;

end

Function $\text{Spectrum}(S)$

$S_f \leftarrow \text{fft}(S)$; // Calcul de la Transformée de Fourier

$S_{ESD} \leftarrow S_f \cdot S_f^*$; // Calcul de la ESD (S_f^* est le conjugué de S_f)

return S_{ESD} ;

end

ont appliqué des algorithmes de sélection d'attributs afin de garder que les plus discriminants (voir la partie 3.8 pour plus de détails). Étant donné qu'une approche de classification hiérarchique est proposée dans notre système, un algorithme de sélection d'attributs est appliqué au niveau de chaque nœud interne de sa structure. Nous démontrerons dans la partie expérimentale (section 4.3) que cette approche est plus pertinente et avantageuse que l'approche traditionnelle. Concernant la méthode de sélection d'attributs utilisée, nous exploitons l'algorithme "Sequential Forward Floating Search" (SFFS) proposé dans les travaux de (PUDIL et al. 1994). Cette méthode est considérée comme l'une des plus efficaces de la littérature (JAIN et al. 1997; KUDO et al. 2000). Dans le but d'améliorer nos résultats, nous proposons une version améliorée de cette dernière en apportant certaines modifications des paramètres. La sélection des attributs est présentée dans l'algorithme 2 : un vecteur F est d'abord initialisé par un ensemble vide, ce dernier va contenir par la suite les attributs sélectionnés. L'approche commence par l'application de l'algorithme "Sequential Forward Selection" (SFS) jusqu'à ce que le vecteur F atteigne une cardinalité égale à 2. Ensuite l'étape 1, nommée inclusion est appliquée. Elle consiste à réaliser une seule itération de l'algorithme SFS. Cela implique l'ajout au vecteur F l'attribut qui maximise

la fonction de critère J . Cette démarche est suivie par l'étape 2 appelée exclusion, relative à l'application de l'algorithme "Sequential Backward Selection" SBS. À ce niveau, à chaque itération, l'attribut le moins significatif (selon la fonction de critère) est écarté du vecteur F . L'étape 2 est répétée tant que la condition $J(F_k \setminus x^-) > J(F_k)$ est vérifiée, sinon retour à l'étape 1. L'algorithme continu jusqu'à ce que la fonction critère ne s'accroît plus.

Algorithm 2: Sequential Forward Floating Selection (SFFS)

Data:

$F_a = \{f_i, 1 \leq i \leq d\}$: Vecteur contenant d attributs calculés (F_T ou F_F)

F : vecteur candidat //Dans l'approche de base : $F = \{\emptyset\}$

Étape 1 (Inclusion) : //One step of Sequential Forward Selection (SFS)

$x^+ \leftarrow \arg \max_{x \in F_a \setminus F_k} J(F_k \cup x);$

$F_{k+1} \leftarrow F_k \cup x^+;$

$k = k + 1;$

Étape 2 (Exclusion) : //Sequential Backward Selection (SBS)

$x^- \leftarrow \arg \max_{x \in F_k} J(F_k \setminus x);$

if $J(F_k \setminus x^-) > J(F_k)$ **then**

$F_{k-1} \leftarrow F_k \setminus x^-;$

$k = k - 1;$

Go to Étape 2

else

Go to Étape 1

end

Dans notre méthode une modification est apportée au niveau du vecteur de l'attribut initial F , qui est à la base un ensemble vide. Afin d'obtenir un résultat plus proche de l'optimum global, l'algorithme débute par un vecteur candidat F considéré comme discriminant en se servant des informations a priori sur la nature des classes d'actions. Le tableau 4.2 montre les vecteurs candidats définis pour chaque problème de classification. Nous pouvons voir que pour classer les actions en catégories dans le premier niveau de classification, l'écart type des signaux inertiels suivant les trois axes (x, y, z) est choisi. Nous savons que cette information est discriminante pour discerner les postures des actions périodiques. Le même vecteur candidat est aussi choisi pour les actions non périodiques. Pour les postures, ce sont les moyennes des signaux

d'accélération qui sont sélectionnées. La figure 4-3 montre un exemple de cette information extraite pour les postures "assis" et "debout", trois positions de smartphone sont considérées : poche droite, ceinture et poignet droit. Nous pouvons voir que pour les trois positions, les données des deux classes sont linéairement séparables (ou presque, hormis quelques points pour la position poignet droit).

Problème de classification	Vecteur candidat
Catégorie de l'action	$F = [std(S_x); std(S_y); std(S_z)]$
Périodique	$F = [F_k(1 : r); F_m(1 : r)]$
Posture	$F = [mean(A_x); mean(A_y); mean(A_z)]$
Non périodique	$F = [std(S_x); std(S_y); std(S_z)]$

TABLE 4.2 – Vecteurs candidats pour chaque problème de classification.

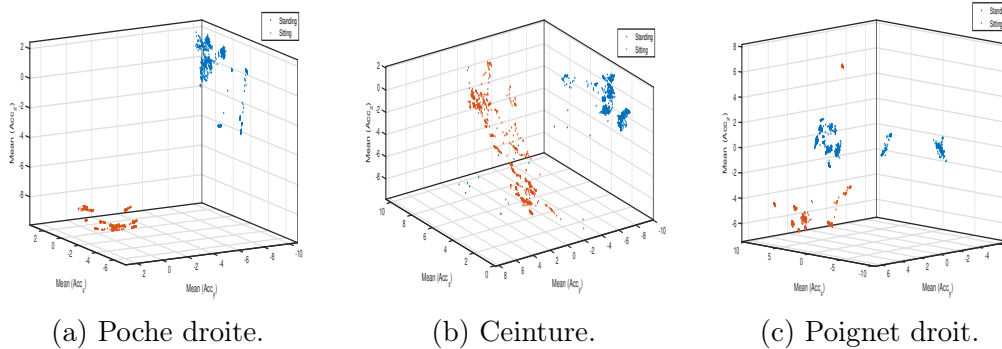


FIGURE 4-3 – Moyennes des signaux d'accélération (A_x, A_y, A_z) pour les postures "assis" et "debout".

Pour les actions périodiques, le vecteur candidat F contient pour chaque signal (S_x, S_y, S_z, S_m) les r premiers coefficients de sa densité spectrale d'énergie. Pour déterminer r , nous nous sommes basés sur les résultats expérimentaux de nos travaux antérieurs (TALHA et al. 2019) : dans un test mené sur un jeu de données publique, un classifieur SVM a été alimenté par un vecteur d'attributs calculé de la même manière que le vecteur candidat pour les actions périodiques. Les performances du classifieur ont été évalué en faisant varier r de $0Hz$ (composantes continues) à $25Hz$ (tous les coefficients). Les résultats obtenus sont représentés dans la figure 4-4. Nous pouvons

voir clairement une croissance du score de performance jusqu'à atteindre le maximum à la fréquence $Fr \simeq 3,5Hz$, une légère baisse est observée au-delà de cette fréquence. À noter que pour des raisons de lisibilité, seules les fréquences se situant entre $0Hz$ et $\approx 10Hz$ sont affichées, au-delà, le score converge. Nous concluons de cette expérience que seuls quelques coefficients de la densité spectrale d'énergie sont pertinents pour la reconnaissance d'actions périodiques. Nous choisissons donc de sélectionner r afin de construire un vecteur candidat F contenant les coefficients ESD inclus dans la bande fréquentielle $[0Hz - 5Hz]$. Un exemple des coefficients sélectionnés est illustré dans la figure 4-5, elle montre les signaux d'accélération de l'action "marcher" accompagnés de leurs densités spectrales d'énergie. Nous pouvons voir en bleu les coefficients sélectionnés pour chaque signal.

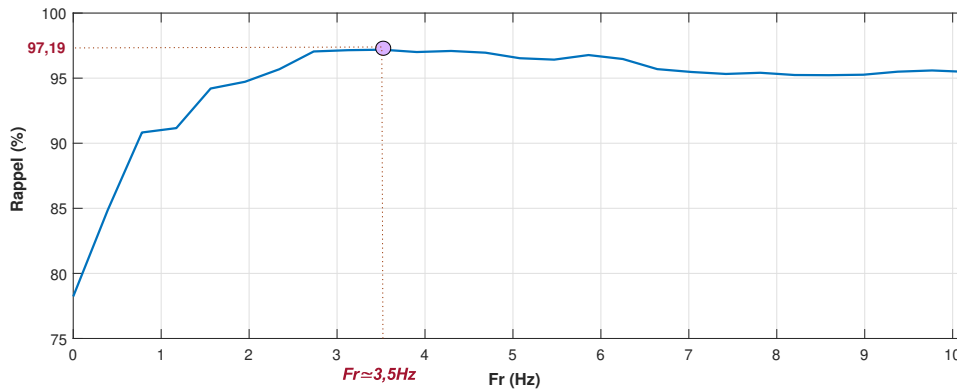


FIGURE 4-4 – Évolution du score *rappel* en fonction de Fr pour les actions périodiques.

Comme toute phase de sélection d'attributs, l'algorithme est appliqué sur un ensemble de données indépendantes. Dans notre approche, des données issues de k sujets sont choisies aléatoirement pour former l'ensemble de données qui va permettre la sélection. Afin de trier des attributs robustes au sujet, la stratégie de validation LOSOCV est appliquée (voir 3.11 pour plus de détails). En ce qui concerne la fonction critère, nous choisissons de maximiser la mesure de la précision $F - score$.

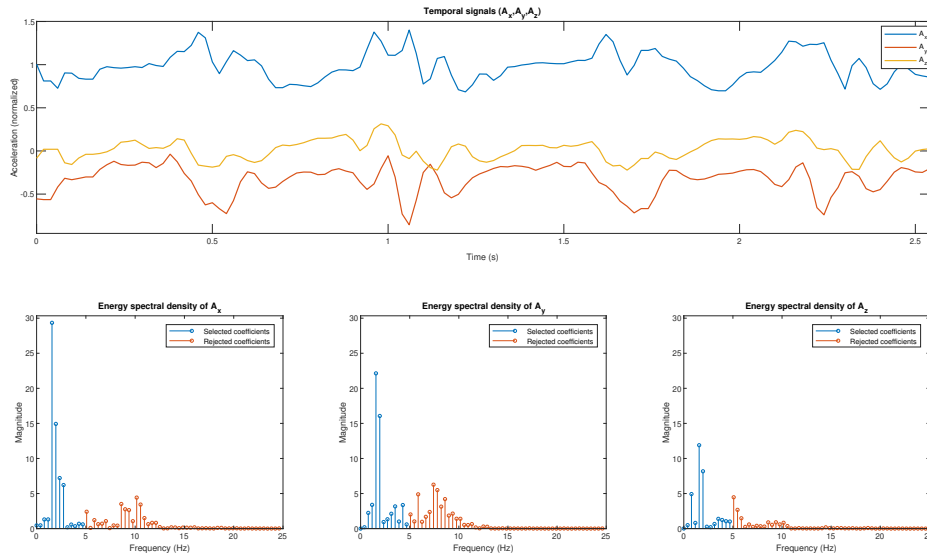


FIGURE 4-5 – Coefficients sélectionnés (en bleu) dans la bande fréquentielle $[0Hz - 5Hz]$ pour l’action “marcher”.

4.2.2 Classification hiérarchique des actions humaines

Une méthode de classification hiérarchique composée de quatre classifieurs est proposée. Chacun est alimenté par un vecteur d’attributs nommés : F_C , F_Q , F_P et F_A ($\{C, Q, P, A\}$ dénote “catégorie d’action”, “action quasi-périodique”, “posture” et “action non périodique”).

Pour le choix du classifieur, nous avons montré dans la section 3.9 du chapitre 3 que l’approche commune adoptée consiste à expérimenter le système avec différentes techniques de classification (méthodes traditionnelles) et de choisir celui qui donne les meilleures performances (score de précision par exemple). Les résultats rapportés indiquent que les machines à vecteurs de support (Support vector machines - SVMs) présentent les meilleures performances. Par conséquent, pour la phase de classification, nous avons opté pour les SVMs implémentés dans LIBSVM (CHANG et al. 2011).

Les SVMs ont été conçus à l’origine pour la classification binaire. Afin de l’étendre efficacement pour la classification à plusieurs classes, différentes méthodes ont été proposées, un-contre-un (one-against-one en anglais) et un-contre-tous (one-against-all en anglais) sont les deux stratégies les plus populaires selon la littérature. Les au-

teurs de (REYES-ORTIZ et al. 2016; ANGUIA et al. 2012) ont utilisé la stratégie un-contre-tous qui construit un SVM pour chaque classe, tandis que dans (SUN et al. 2010; Rong YANG et al. 2016; MANNINI, ROSENBERGER et al. 2017), la stratégie un-contre-un est appliquée, dans laquelle un SVM est construit pour chaque paire de classes. Dans (HSU et al. 2002), une comparaison des méthodes pour les SVMs multi-classes (y compris un-contre-un et un-contre-tous) est effectuée. Les résultats montrent que l’approche un-contre-un est plus appropriée pour une utilisation pratique que l’approche un-contre-tous. Par conséquent, dans notre système proposé, une approche un-contre-un est utilisée pour obtenir un support multi-classe. Pour n nombre de classes d’action, $\frac{n(n-1)}{2}$ classifieurs sont construits, chacun entraînant les données de deux classes. Une stratégie de vote est appliquée sur les scores fournis par les classifieurs binaires. Une instance est indiquée comme appartenant à la classe avec le nombre maximal de votes.

Chaque couple $(F_i, SVMs_i)$ avec $i \in \{C, Q, P, A\}$ est construit pour résoudre un problème de classification spécifique. Dans le premier niveau de notre système, les actions sont divisées en trois catégories : actions périodiques, actions non périodiques et postures. Pour ce faire, le classifieur SVM_C est alimenté par le vecteur d’attributs F_C contenant des attributs temporels sélectionnés par l’algorithme de sélection d’attributs. Au deuxième niveau, les actions non périodiques et les postures utilisent respectivement les classifieurs SVM_A et SVM_P alimentés par les vecteurs d’attributs temporels F_A et F_P déterminés de la même manière que F_C . Enfin pour les actions périodiques, le vecteur F_Q contenant les coefficients ESD discriminants et optimaux est utilisé en tant qu’entrée pour le classifieur SVM_Q .

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. La figure 4-6 montre les erreurs pouvant être rencontrées. Nous pouvons constater certains événements impulsifs qui représentent des erreurs de classification. Dans notre approche, l’erreur impulsive de prédiction est rectifiée sur la base de l’hypothèse suivante : puisque $O_v = 50\%$ dans nos expériences, alors une

fenêtre segmentée à un instant t nommée W_t partage 50% de données avec W_{t-1} , et les 50% restantes avec W_{t+1} . Cela implique que dans le cas où W_{t-1} et W_{t+1} génèrent la même étiquette, elle est supposée être égale à l'étiquette de W_t . L'implémentation de la solution proposée est décrite par l'algorithme 3.

Algorithm 3: Correction des erreurs de classification

Data:

$F(t)$: Vecteur d'attributs à l'instant t

$PR(t-1) = [pred(t-N), \dots, pred(t-2), pred(t-1)]$: Vecteur contenant les étiquettes d'actions prédites

$pr \leftarrow SVM_Predict(F(t));$

if $[pr \neq PR(end)]$ **AND** $[pr == PR(end-1)]$ **then**

$PR(end) \leftarrow pr;$

end

$PR = [PR \ pr];$

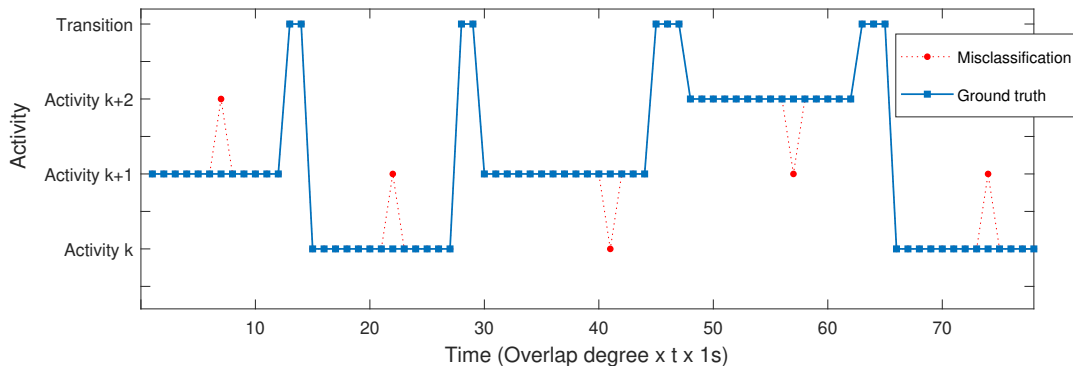


FIGURE 4-6 – Exemple de correction des erreurs de classification.

4.2.3 Classification des transitions posturales

Dans notre travail, les transitions entre les différentes postures (assis, debout et couché) sont étudiées. Des tests expérimentaux sur le jeu de données SBHART (REYES-ORTIZ et al. 2016) sont effectués à cette fin. SBHART est une version mise à jour du jeu de données SBHAR (ANGUITA et al. 2012). Elle comprend, en plus des actions de base, six classes de transition posturales. La figure 4-7 illustre les signaux d'accélération de différentes transitions posturales prises en compte dans notre étude.

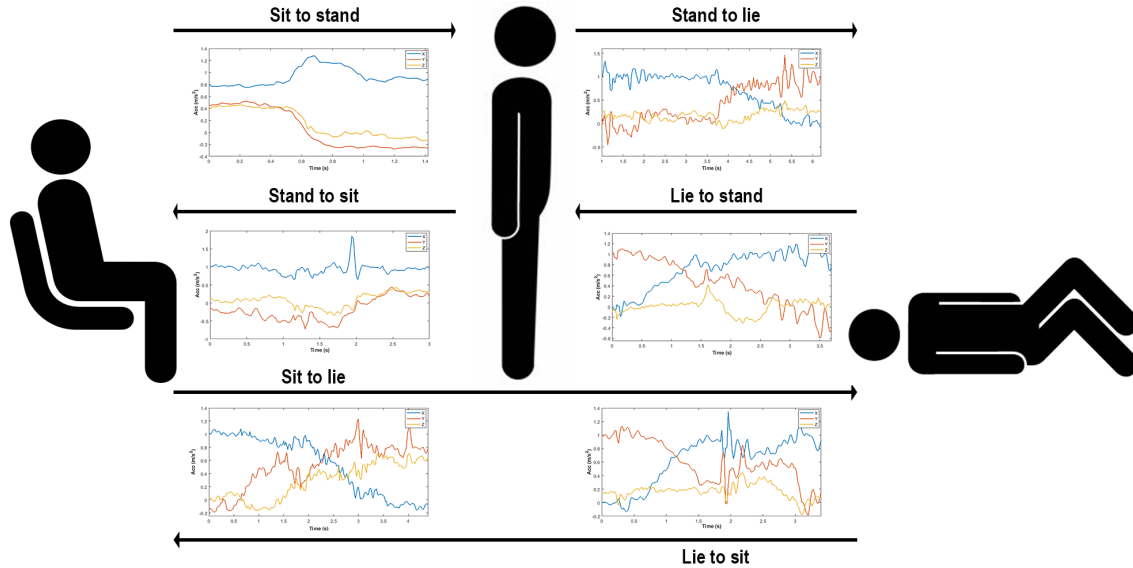


FIGURE 4-7 – Transitions posturales considérées dans notre travail.

Afin de reconnaître efficacement les transitions posturales, nous proposons de diviser le problème en deux parties : détection de la transition et son identification. Le processus de détection de la transition vise à détecter cette dernière, sans aucune information sur son type. Par conséquent, à l'étape de classification, toutes les transitions posturales sont considérées comme une classe supplémentaire ; l'étiquette de sortie du classifieur est notée pr avec $pr \in \{1, \dots, N_{ha}, Tr_lab\}$ (N_{ha} correspond au nombre d'actions humaines considérées par le système et Tr_lab l'étiquette de détection de la transition posturale, la sortie est donc soit l'une des actions soit une transition).

Une fois la transition détectée, une approche d'estimation de transition est exécutée afin de reconnaître sa classe parmi six catégories de transitions posturales, à savoir : “debout vers assis”, “assis vers debout”, “allongé vers assis”, “allongé vers debout”, “debout vers allongé”, “assis vers allongé”. L'idée est d'inspecter les étiquettes prédites adjacentes à la transition (voir l'algorithme 4). Pour expliquer l'algorithme, son exécution sur un exemple du jeu de données SBHART est présentée dans la figure 4-8. Elle montre les valeurs affectées aux variables principales de l'algorithme pour estimer la transition posturale “debout vers assis”.

Dans l'exemple, $Tr_lab = 7$, pour une bonne visibilité de l'illustration, seuls

Algorithm 4: Estimation d'une transition posturale

Data: $F(t)$: Attributs extraits des signaux inertiels à l'instant t
 $PR(t-1) = [pred(t-N), \dots, pred(t-2), pred(t-1)]$: Vecteur contenant des étiquettes des actions humaines reconnues antérieurement
 D_labs : Dictionnaire contenant une paire d'étiquettes de posture accompagnée de l'étiquette de la transition correspondante
 Tr_lab : L'étiquette d'une détection de transition posturale
 $Flag_tr$: Déclencheur pour enregistrer l'étiquette de l'action située avant la transition
 $Save_tr$: Déclencheur pour identifier l'étiquette de la transition détectée
 $Size_tr$: Taille de la transition posturale détectée

Initialisation :
 $Flag_tr \leftarrow 0$;
 $Save_t \leftarrow 0$;
 $Size_tr \leftarrow 0$;
 $pr \leftarrow SVM_Predict(F(t))$;

if $pr == Tr_lab$ **then**
 $Size_tr \leftarrow Size_tr + 1$;
 if $Flag_tr == 0$ **then**
 $Tr_1 \leftarrow PR(end)$;
 $Flag_tr \leftarrow 1$;
 end

else if $Flag_tr == 1$ **then**
 $Tr_2 \leftarrow pr$;
 $Flag_tr \leftarrow 0$;
 $Save_tr \leftarrow 1$;
end

if $Save_tr == 1$ **then**
 $Tr \leftarrow Find(D_labs == [Tr_1 \ Tr_2])$;
 $PR(end - Size_tr : end) \leftarrow Tr$;
 $Save_tr \leftarrow 0$;
 $Size_tr \leftarrow 0$;
end

$PR = [PR \ pr]$;

les signaux d'accélération sont tracés et aucun degré de chevauchement n'est effectué. Dans la figure 4-8, chaque signal est segmenté en trois fenêtres : $W = \{W_{T_1}, W_{T_2}, W_{T_3}\}$ avec $T = 1s$.

Dans W_{T_1} , le classifieur SVM prédit l'action *stand*. Cela correspond à $pr = 5$. Dans W_{T_2} , une transition posturale est détectée ($pr = 7$). Par conséquent, l'action prédite

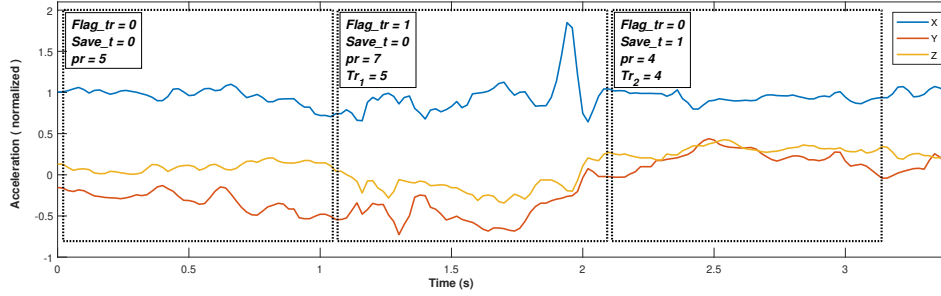


FIGURE 4-8 – Application de l’algorithme 4 afin d’estimer la transition posturale “debout vers assis”.

antérieurement est enregistrée dans la variable Tr_1 . Enfin, dans W_{T_3} , la fenêtre est en dehors de la transition posturale. Ainsi, l’action prédite à ce moment-là est enregistrée dans la variable Tr_2 . À partir du dictionnaire “ D_labs ” cité dans l’algorithme 4, l’étiquette d’une transition posturale correspondant à $[Tr_1, Tr_2]$ est récupérée.

4.3 Validation expérimentale

Afin de montrer les performances et la robustesse de notre système, une validation expérimentale est menée sur plusieurs jeux de données publiques, chacun possédant un challenge particulier. Comme indiqué dans la partie 2.5.3 du chapitre 2, dans notre travail, nous nous sommes intéressés à trois catégories d’actions intimement corrélées avec l’autonomie et le bien-être, à savoir : les actions sédentaires (postures), les actions périodiques ou quasi-périodiques et les actions non périodiques. Le tableau 4.3 décrit les classes d’actions considérées dans la phase expérimentale et leurs labels. Nous pouvons voir que dans certains jeux de données, des actions sont regroupées pour former une nouvelle classe comme “monter et descendre les escaliers” par exemple (label 4). Nous notons aussi que certaines classes sont décrites par le lieu d’acquisition comme “assis dans une voiture”. Pour les actions non périodiques, trois classes sont testées : transitions posturales, la chute et sauter. Pour rappel, après la détection des transitions posturales, un algorithme est appliqué par la suite afin de reconnaître le type de transition.

En ce qui concerne les positions des capteurs, elles sont définies dans le tableau 4.4.

Catégorie de l'action	Label	Action
Action périodique (quasi-périodique)	1	Marcher
	2	Monter les escaliers
	3	Descendre les escaliers
	4	MD. les escaliers (act. 2 et 3)
	5	Marcher (act. 4 inclus)
	6	Courir (ou jogging)
	7	Cyclisme
Posture	8	Debout
	9	Assis
	10	Assis dans une voiture
	11	Assis et debout
	12	Allongé
Action non périodique	13	Transitions posturales
	14	La chute
	15	Sauter

TABLE 4.3 – Les actions prise en compte dans notre étude.

Les positions sont étiquetées de P_1 à P_{10} . Cette notation sera utilisée par la suite dans la phase expérimentale. À noter que dans certains jeux de données, les poches droite et gauche sont considérées comme des positions différentes, dans d'autres jeux de données, ils forment une seule position. En outre, dans certains protocoles d'acquisition, les auteurs demandent aux sujets de mettre le smartphone dans une orientation fixe, d'autres laissent le choix aux sujets de mettre le smartphone dans n'importe quelle orientation. Par conséquent, nous avons attribué un label différent pour chacun des cas énoncés.

Label	Position du capteur
1	Poche du pantalon (orientation libre)
2	Poche droite du pantalon (orientation fixe)
3	Poche gauche du pantalon (orientation fixe)
4	Poche droite du pantalon (orientation libre)
5	Poche gauche du pantalon (orientation libre)
6	Hanche (ceinture et taille)
7	Poitrine
8	Poignet (ou avant-bras)
9	Haut du bras
10	Cheville

TABLE 4.4 – Les positions du capteur prise en compte dans notre étude.

4.3.1 Jeux de données

Les tests expérimentaux sont effectués sur les jeux de données détaillés dans le tableau 4.5. Dans notre étude, nous exploitons non seulement les données issues du smartphone mais aussi les IMU.

Le jeu de données “MobiAct” a été collecté par (VAVOULAS et al. 2016). Il contient des données échantillonnées à $200Hz$, issues d’accéléromètre et de gyroscope embarqués dans un smartphone Samsung Galaxy S3. Il est à noter que 57 sujets ont participé à cette expérimentation dont 42 hommes et 15 femmes âgés de 20 à 47 ans. Le smartphone est localisé dans la poche droite ou gauche du pantalon choisie librement par les sujets et dans une orientation arbitraire. Le jeu de données contient les actions suivantes : “marcher”, “monter les escaliers”, “descendre les escaliers”, “courir”, “debout”, “assis sur une chaise”, “allongé”, “sauter”, “monter dans une voiture” et “descendre d’une voiture”. Quatre types de chutes sont aussi collectées : chute vers l’avant avec utilisation des mains pour amortir la chute, chute vers l’avant avec premier impact sur les genoux, chute sur le côté (jambes pliées), et chute vers l’arrière en essayant de s’asseoir sur une chaise. Dans notre travail, les actions “monter dans une

voiture” et ”descendre d’une voiture” sont écartées du jeu de données dans la phase expérimentale, et les différentes catégories de chutes sont regroupées dans une seule classe.

Le jeu de données “Siirtola” (SIIRTOLA et al. 2012b) contient trois actions quasi-périodiques : “marcher”, “courir”, “cyclisme”, et deux classes de postures où la première regroupe deux postures “assis sur une chaise au travail et debout” et la seconde classe “assis dans une voiture”. 8 personnes âgées entre 25 et 34 ans ont participé à cette expérimentation, chacun portait un smartphone Nokia N8 dans la poche droite ou gauche (choix libre aux sujets). Seuls les signaux d’accélération sont collectés. La fréquence d’échantillonnage est fixée à $40Hz$.

Le jeu de données “WISDM” (KWAPISZ et al. 2011) contient les signaux d’accélération de six actions : “marcher”, “monter les escaliers”, “descendre les escaliers”, “courir”, “debout” et “assis”. 36 personnes ont porté des smartphones de différentes marques dans la poche droite ou gauche du pantalon. La fréquence d’échantillonnage, réglée à $20Hz$, est relativement faible comparée à celle des autres jeux de données utilisés dans notre travail.

Le jeu de données “SBHART” (REYES-ORTIZ et al. 2016) contient six actions exécutées par trente sujets, chacun portant un smartphone Samsung Galaxy S2 à la taille. Trois des actions rassemblées sont les postures (debout, assise et couchée) et les trois autres des actions quasi-périodiques (marcher, monter et descendre les escaliers). En outre, le jeu de données contient également des transitions posturales : “debout vers assis”, “assise vers debout”, “allongée vers assise”, “allongée vers debout”, “debout vers allongée”, “assise vers allongée”. Chaque sujet a réalisé deux répétitions d’un scénario d’actions séquentielles décrit dans la figure 4-9. Dans chaque scénario, toutes les actions sont effectuées par les sujets dans un ordre spécifique comme il est indiqué sur la figure. Dans les extrémités de chaque action, la fenêtre glissante est susceptible d’inclure la classe “transition” en plus de l’action physique. Deux situations peuvent être rencontrées : action suivie ou précédée d’une transition . Dans notre système, ces situations sont gérées comme suit : dans le premier cas, la prédiction à un instant t appelée $pred(t)$ est considérée comme correcte si elle est étiquetée comme

TR ou $GT(t - 1)$ correspondant au label de transition ou vérité de terrain à l’instant précédent. Dans le second cas, la prédiction est correcte si $pred(t)$ est étiquetée comme TR ou $GT(t + 1)$. Sinon, $pred(t)$ est évaluée en tant qu’erreur de classification.

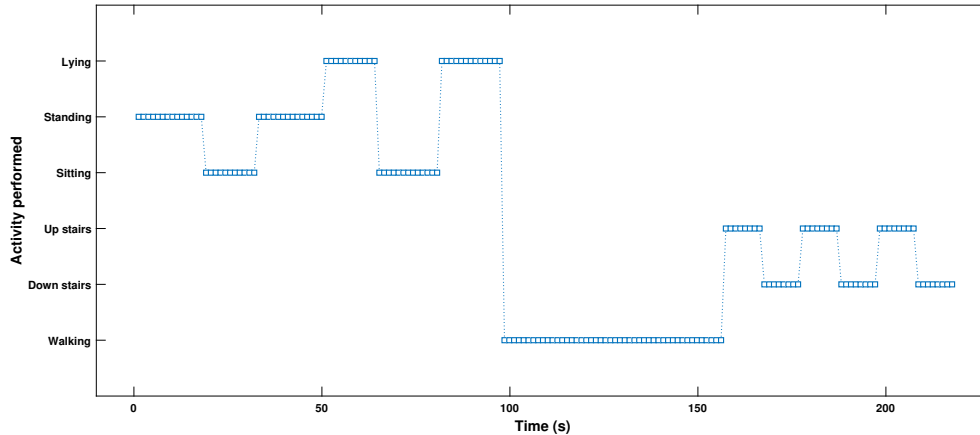


FIGURE 4-9 – Scénario des actions réalisées dans le jeu de données SBHART.

Le jeu de données “UniMiB SHAR” (MICUCCI et al. 2017) contient les signaux d’accélération de quatre actions quasi-périodiques : “marcher”, “monter les escaliers”, “descendre les escaliers”, “courir”. Le jeu de données contient aussi quatre transitions posturales : “debout vers assis sur une chaise”, “assis sur une chaise vers debout”, “debout vers allongée sur un lit”, “allongée sur un lit vers debout”, et 8 types de chutes : chute vers l’arrière (cas général), chute vers l’arrière en essayant de s’asseoir sur une chaise, chute vers l’avant avec utilisation des mains pour amortir la chute, chute vers l’avant (avec des protections), chute sur le côté droit à partir de la position debout, chute sur le côté gauche à partir de la position debout, tomber en ayant un contact avec un obstacle avant de heurter le sol, et perte de connaissance. 30 sujets ont participé à cette collecte de données, dont 24 femmes âgées entre 18 et 55 ans et 6 hommes âgés entre 20 et 60 ans. Les sujets portaient un smartphone de marque Samsung Galaxy Nexus, la fréquence d’échantillonnage est réglée à $50Hz$. Deux protocoles d’actions ont été définis par les auteurs du jeu de données. Chaque protocole était répété deux fois. Dans la première, le smartphone était positionné à droite et dans la seconde à gauche. Au final, nous obtenons deux sous-ensembles de données avec deux positions différentes de smartphone. Dans notre travail, les

différentes catégories de chutes sont regroupées dans une seule classe (semblable au jeu de données “MobiAct”). En outre, dans ce jeu de données, la chronologie des actions est absente (les actions sont regroupées par classe). De ce fait, notre approche de reconnaissance de la nature des transitions posturales (détaillée dans la partie 4.2.3) n’est pas applicable. Par conséquent, les transitions posturales sont regroupées elles aussi en une seule classe.

Afin d’étudier la robustesse de notre approche aux changements de position du smartphone, un jeu de données collectées suivant plusieurs positions est nécessaire. Ainsi, le jeu de données “Multi-position” (SHOAIB, BOSCH, O. D. INCEL et al. 2014) est exploité à cette fin. Il contient des signaux d’accélération et vitesse angulaires de sept actions. Cinq d’entre elles sont quasi-périodiques : “marcher”, “courir”, “monter les escaliers”, “descendre les escaliers”, “cyclisme”, et le reste des postures : “debout” et “assis”. Les actions ont été exécutées par dix participants. Chacun portait cinq smartphones Samsung Galaxy S2 dans cinq positions différentes : (1) poche droite du pantalon, (2) poche gauche du pantalon, (3) la hanche (ceinture) du côté de la jambe droite, (4) bras droit et (5) poignet droit. Les positions sont illustrées dans la figure 4-10.

Comme il est mentionné dans la partie 3.4, quelques chercheurs seulement ont essayé de répondre à la problématique de la reconnaissance d’actions suivant plusieurs positions de smartphone. Par conséquent, un manque est observé au niveau des jeux de données publiques. Ainsi pour pallier à ce problème, en plus du jeu de “Multi-position”, un ensemble de données collectées avec des dispositifs IMU nommé “mHealth” est aussi exploité. Il contient des signaux inertiels issus de trois dispositifs IMU placés sur la poitrine, cheville gauche et l’avant-bras droit. Les dispositifs contiennent un accéléromètre triaxial, deux d’entre eux (ceux localisés à la cheville et l’avant-bras) incluent en plus un gyroscope et un magnétomètre. Dans nos tests expérimentaux nous étudions l’impact des changements de position du dispositif IMU sur les performances de la classification. Nous exploitons ainsi que les signaux communs sur les trois positions, à savoir les accélérations. Dix sujets ont porté les capteurs afin d’exécuter les 12 actions suivantes : “marcher”, “monter et descendre les escaliers”, “cy-

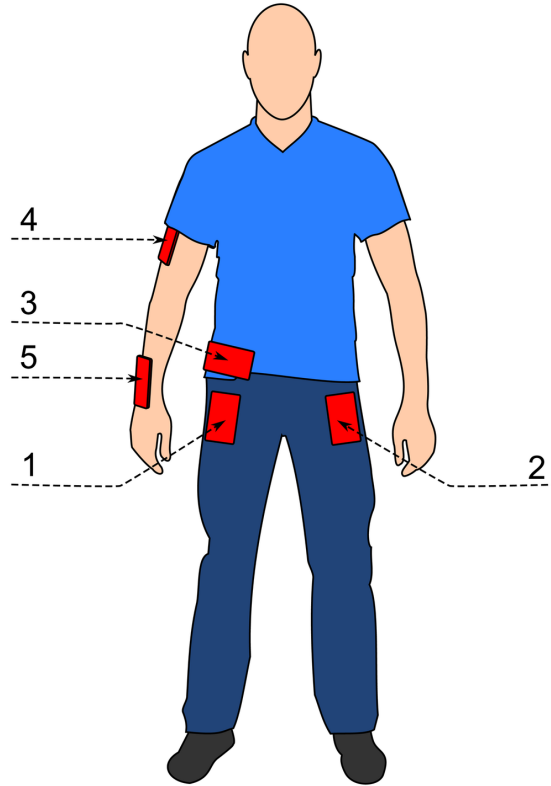


FIGURE 4-10 – Aperçu des positions des smartphones sur un participant dans le jeu de données “Multi-position” (SHOAIB, BOSCH, O. D. INCEL et al. 2014).

clisme”, “courir”, “jogging”, “debout”, “assis”, “allongé”, “saut vers l’avant et l’arrière”, “se pencher en avant”, “élévation frontale des bras”, “fléchir les genoux”. Dans notre travail, les deux dernières actions sont écartées car elles ne correspondent pas à notre objectif. En outre, les actions “courir” et “jogging” sont regroupées en une seule classe. Au final 8 classes d’actions sont considérées dans nos tests expérimentaux.

4.3.2 Estimation des performances

Afin d’estimer les performances de reconnaissances globales de notre système et de les comparer aux méthodes de la littérature, quatre mesures sont calculées : *précision*, *rappel*, *F-score* et *exactitude* (*accuracy* en anglais). Les formules suivantes sont utilisées pour le calcul :

$$Rappel = \frac{V_P}{V_P + F_N} \quad (4.3)$$

Jeu de données	Actions ¹	N_{cap} ²	Position ³	$F_e(Hz)$	N_i ⁴
MobiAct (VAVOULAS et al. 2016)	$A_1, A_2, A_3, A_6, A_8, A_9, A_{12}, A_{14}, A_{15}$	2	P_1	200	57
Siirtola (SIIRTOLA et al. 2012b)	$A_5, A_6, A_7, A_{10}, A_{11}$	1	P_1	40	8
WISDM (KWAPISZ et al. 2011)	$A_1, A_2, A_3, A_6, A_8, A_9$	1	P_1	20	36
SBHART (REYES-ORTIZ et al. 2016)	$A_1, A_2, A_3, A_8, A_9, A_{12}, A_{13}$	2	P_6	50	30
UniMiB SHAR (MICUCCI et al. 2017)	$A_1, A_2, A_3, A_6, A_{13}, A_{14}, A_{15}$	1	P_4, P_5	50	30
mHealth (BANOS, GARCIA et al. 2014)	$A_1, A_4, A_6, A_7, A_8, A_9, A_{12}, A_{15}$	1	P_7, P_8, P_{10}	50	10
Multi-position (SHOAIB, BOSCH, O. D. INCEL et al. 2014)	$A_1, A_2, A_3, A_6, A_7, A_8, A_9$	2	P_2, P_3, P_6, P_8, P_9	50	10

¹ La notation pour les actions est définie dans le tableau 4.3

² Nombre de capteurs inertiels utilisés

³ La notation pour les positions est définie dans le tableau 4.4

⁴ Nombre de sujets qui ont participé à l'expérimentation

TABLE 4.5 – Les jeux de données publiques utilisés dans la phase expérimentale.

$$Précision = \frac{V_P}{V_P + F_P} \quad (4.4)$$

$$Exactitude = \frac{V_P + V_N}{V_P + V_N + F_P + F_N} \quad (4.5)$$

$$F - score = 2 \cdot \frac{Précision \cdot Rappel}{Précision + Rappel} \quad (4.6)$$

Dans les formules ci-dessus, V_P , V_N , F_P et F_N correspondent respectivement à vrais positifs, vrais négatifs, faux positifs et faux négatifs.

À noter que $F-score$ est appelé “score” par abus de langage dans les tests expérimentaux.

Dans les approches de l'état de l'art évaluées sur les jeux de données sélectionnées,

différents scores ont été calculés, principalement *l'exactitude*. Par conséquent, dans le but 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 : la première est le Leave-One-Subject-Out-Cross-Validation (LOSOVCV) (cf. partie 3.11). Dans notre travail, une technique de sélection d'attributs est appliquée. Par conséquent, pour ne pas biaiser les scores calculés, un tiers des sujets sont choisis aléatoirement et mis à l'écart pour le processus de sélection d'attributs. Ces derniers sont ensuite utilisés dans la phase d'apprentissage mais pas dans la phase de test. Dans ce travail, ce protocole est choisi afin de détailler les résultats (matrice de confusion, scores par catégories d'actions, les tests avec différents capteurs, etc.). La seconde technique est la validation simple et indépendante du sujet "hold-out subjects". Dans ce cas, les données de certains sujets sont choisies aléatoirement pour la phase d'apprentissage et la sélection d'attributs, et les données des sujets restants pour la phase de test. Cette technique de validation est utilisée dans l'éventualité où un découpage apprentissage/test particulier est considéré par les chercheurs pour un des jeux de données choisis. C'est le cas par exemple du découpage du jeu de données "SBHART" proposé par (ANGUITA et al. 2012), qui a été repris par la suite par d'autres chercheurs, où les données de 21 sujets ont été sélectionnées pour la phase d'apprentissage et les 9 sujets restants pour la phase de test.

En ce qui concerne les tests effectués sur des jeux de données collectés suivant plusieurs emplacements du dispositif (IMU ou smartphone), trois types de tests expérimentaux sont menés :

- Tests dépendants de la position : les données de la phase d'apprentissage et de test sont collectées depuis le même emplacement de dispositif.
- Tests positions croisées : les données de la phase d'apprentissage et de test sont collectées suivant les positions P_{k1} et P_{k2} respectivement, avec $k1 \neq k2$.

- Tests indépendants de la position : les données de la phase d'apprentissage et de test sont collectées suivant les positions P_k , avec k allant de 1 jusqu'à N_p (nombre de positions du dispositif dans un jeu de données). Ce test est dénommé P_i dans ce qui suit.

À noter que dans les tests indépendants de la position, des données issues d'une seule position sont utilisées à la fois. Cependant, dans certains travaux de l'état de l'art (travaux avec des dispositifs IMU généralement), des données issues de plusieurs emplacements de dispositifs sont utilisées simultanément à l'entrée du système. Ce test est dénommé P_a dans ce qui suit afin de ne pas le confondre avec le test P_i .

4.3.3 Résultats et comparaison avec les travaux de l'état de l'art

La comparaison de notre approche avec les travaux de l'état de l'art est présentée en trois phases. Dans la première, trois jeux de données "MobiAct", "Siirtola" et "WISDM" sont exploités. Ils contiennent des actions quasi-périodiques, postures et les actions "sauter", "la chute" pour le jeu de données "MobiAct". Dans la deuxième partie des tests, en plus des actions citées dans la partie 1, les transitions posturales sont aussi prises en considération. Les jeux de données "SBHART" et "UniMiB SHAR" sont alors exploités dans ce cas. Dans ces deux parties des expérimentations citées, le smartphone est mis en position fixe. Afin d'étudier la robustesse de notre approche face au changement de position du smartphone, des expérimentations sont effectuées dans la dernière partie en exploitant deux jeux de données additionnelles "mHealth" et "Multi-position smartphone".

Actions humaines

Le tableau 4.6, présente les résultats obtenus dans le jeu de données "MobiAct". La technique de validation LOSOCV a donné un score de 92,97%. Dans (ZHENG et al. 2018), la validation "Hold-out subjects" a été retenue. 18 des 57 sujets du jeu de données sont d'abord pré-sélectionnés. Les données de 15 de ces sujets ont

Validation	Approche	Rap (%)	Pr (%)	F-s (%)	Ex (%)
LOSOCV ¹	Nous	93,44	92,51	92,97	96,55
HS ²	(ZHENG et al. 2018)	86,44	84,25	85,33	90,09
	(WELHENGE et al. 2018)	-	-	-	90,00
CV ³	(VAVOULAS et al. 2016)	-	-	-	99,88
	(HUR et al. 2018)	100	100	100	100

¹ Leave-one-subject-out cross-validation.

² Hold-out subjects.

³ k -fold cross-validation.

TABLE 4.6 – Scores obtenus en comparaison avec l’état de l’art pour le jeu de données MobiAct.

été utilisées pour la phase d’apprentissage, et celles des trois restants pour le test. Un score de 85,33% est obtenu, bien inférieur au nôtre. En outre, dans l’approche proposée, une fenêtre de 4 secondes est utilisée, que nous jugeons trop large pour certaines actions comme la chute ou le saut par exemple. Dans (WELHENGE et al. 2018), le modèle LSTM proposé est entraîné sur 20 sujets et testé sur les 20 autres. Seul le score “exactitude” est indiqué dans ce travail. Il est égal à 90,00%, inférieur à celui que nous avons obtenu (96,55%). De plus, seules cinq actions sont considérées par (WELHENGE et al. 2018), à savoir : marcher, monter les escaliers, courir, sauter et debout. Il est à remarquer que, si dans notre approche nous avons considéré ces mêmes actions en écartant les autres, un score bien supérieur aurait été obtenu (en se référant à la matrice de confusion 4.7 qui va être détaillée par la suite). Dans (VAVOULAS et al. 2016 ; HUR et al. 2018), une validation croisée simple (k -fold cross-validation) est appliquée. La même taille de fenêtre, égale à 5 secondes est utilisée dans les deux travaux, avec un degré de chevauchement de 80%. Un score “exactitude” de 99,88% est obtenu dans (VAVOULAS et al. 2016) et 100% dans (HUR et al. 2018). Comme nous l’avons expliqué dans la partie 3.11, cette technique de validation est dépendante du sujet. Elle présente des résultats ne correspondant pas au cas pratique. De plus, le degré de chevauchement élevé ($O_v = 80\%$), rend les résultats encore plus biaisés.

La matrice de confusion de nos résultats sur le jeu de données MobiAct est présentée dans le tableau 4.7. Les résultats montrent que les actions dynamiques sont

bien reconnues, entre 93,24% pour “monter les escaliers” et 99,62% pour la marche. Cependant, nous observons des confusions importantes entre les postures “assise” et “allongée”. Le smartphone est localisé dans la poche droite ou gauche du pantalon du sujet. Suivant la conception de la poche, des signaux similaires peuvent être observés à la sortie des capteurs pour les postures “assise” et “allongée”. De plus l’orientation du smartphone, qui est arbitraire, rend la classification encore plus difficile.

	M	ME	DE	Co	D	As	Al	S	Ch
Marcher	99,62	0,06	0,16	0,04	0	0	0	0	0,12
M. escaliers	4,10	93,24	0,23	2,43	0	0	0	0	0
D. escaliers	3,56	0,79	93,98	0,24	0	0	0	0	1,43
Courir	0,07	0	0,63	98,53	0	0	0	0,77	0
Debout	0	0	0,02	0	99,88	0	0,10	0	0
Assis	0	0	0	0	1,00	74,18	24,82	0	0
Allongé	0	0	0	0	0,45	13,08	86,47	0	0
Sauter	0	0	0,08	2,13	0	0	0	97,79	0
Chute	0	0	0	2,74	0	0	0	0	97,26

TABLE 4.7 – Matrice de confusion (%) pour le jeu de données MobiAct.

Le tableau 4.8 présente les scores obtenus pour chaque classifieur SVM_i avec $i \in \{C, D, P, A\}$ (C , la catégorie, est la première classification faite pour séparer en méta-classes), et les dimensions des vecteurs d’attributs mis à l’entrée de ces classifieurs. Les résultats montrent que dans le premier niveau de classification (répartition des actions en catégories), en employant 11 attributs statistiques, un score de 99,18% est atteint. Pour les actions quasi-périodiques, nous atteignons un score de 97% avec 91 attributs. Les postures sont reconnues à 87,03% avec seulement 9 attributs, et enfin nous descendons jusqu’à 2 attributs pour classifier les actions “sauter” et “la chute”. À noter que dans notre approche, les attributs sélectionnés constituent des mesures statistiques basiques, simple et peu coûteuse à calculer (voir le tableau 4.1 de la partie 4.2.1). À titre de comparaison, dans (ZHENG et al. 2018), 162 attributs sont calculés afin d’atteindre un score global de 85,33%.

Le tableau 4.9 présente les résultats obtenus dans le jeu de données “Siirtola”. La validation par LOSOCV a donné un score “*rappel*” de 91,20%, inférieur à celui de (SIIRTOLA et al. 2012b) qui est de 94,53% quoiqu’une fenêtre de taille importante est utilisée ($T = 7,5s$).

	Catégorie	Quasi-périodique	Posture	Non périodique
F-s (%)	99,18	97,00	87,03	93,69
$Dim(F_i)_{i \in \{C,D,P,A\}}$	11	91	9	2

TABLE 4.8 – Scores pour chaque classifieur et dimensions des attributs pour le jeu de données MobiAct (ici et dans la suite “Catégorie” signifie la séparation en méta-classes).

Validation	Approche	Rap (%)	Pr (%)	F-s (%)	Ex (%)
LOSOCV ¹	(SIIRTOLA et al. 2012b)	94,53	-	-	-
	Nous	91,20	94,65	92,89	92,78

¹ Leave-one-subject-out cross-validation.

² Hold-out subjects.

³ k -fold cross-validation.

TABLE 4.9 – Scores obtenus en comparaison avec l’état de l’art pour le jeu de données Siirtola.

La matrice de confusion est présentée dans le tableau 4.10. Elle montre des scores “*rappel*” très élevés pour les actions quasi-périodiques, supérieur à 99%. Nous remarquons aussi que la classe qui a fait baisser le score global est l’action “assis dans une voiture”, avec un score de 61,52%. Celle-ci est mélangée avec la classe “postures” contenant à la fois les postures “assise” et “debout”. De ce fait, nous pouvons dire que cette confusion est tout à fait logique vu que les signaux de sortie des capteurs pour les postures “assis dans une chaise” ou “assis dans une voiture” sont quasi identiques.

	M	Co	Cy	P	A
Marcher ¹	99,19	0	0,25	0,56	0
Courir	0,29	99,71	0	0	0
Cyclisme	0	0	99,21	0,46	0,33
Postures ²	0,63	0	0,69	96,39	2,29
Assis ³	0	0	1,64	36,84	61,52

¹ Marcher, monter et descendre les escaliers.

² Assis et debout.

³ Assis dans une voiture.

TABLE 4.10 – Matrice de confusion (%) pour le jeu de données Siirtola.

Le tableau 4.11 présente les scores obtenus pour chaque classifieur de notre système et les dimensions des vecteurs d’attributs. Les résultats montrent que 9 attributs sont calculés afin de reconnaître la catégorie de l’action. Il faut compter 24 attributs pour

les actions quasi-périodiques et 5 attributs pour les postures.

	Catégorie	Quasi-périodique	Posture
F-s (%)	98,99	98,97	83,57
$Dim(F_i)_{i \in \{C,D,P\}}$	9	24	5

TABLE 4.11 – Scores pour chaque classifieur et dimensions des attributs pour le jeu de données Siirtola.

Le tableau 4.12 présente les résultats obtenus dans le jeu de données “WISDM”. Pour notre approche, un score de 89,38% est obtenu avec technique de validation LOSOCV, bien supérieur au score obtenu par (KOLOSNAJI et al. 2015), qui est de 81,64%. Dans (MOHAMMAD et al. 2018), la technique de validation 10-fold-subject-cross-validation est appliquée. La taille de la fenêtre utilisée est $T = 10s$, quatre fois supérieure à la nôtre. Un score “exactitude” de 86,03% est alors obtenu. Dans (IGNATOV 2018), les 26 premiers sujets sont exploités pour la phase d’apprentissage et le reste des sujets pour la validation. Un score “rappel” de 88,59% est obtenu, inférieur au score que nous avons obtenu, en revanche, le score “Exactitude” est plus élevé, 93,32%, cela est dû au non équilibrage du nombre d’exemples dans chaque classe (la classe “marche” est majoritaire à 38,66% d’exemples au total). Il est à noter aussi qu’une taille de fenêtre de 4 secondes est employée dans les tests de (IGNATOV 2018).

Dans (KWAPISZ et al. 2011 ; CATAL et al. 2015 ; VAVOULAS et al. 2016 ; MOHAMMAD et al. 2018 ; ARIF et al. 2014), la technique de validation k -fold-subject-cross-validation est utilisée, des scores entre 81,00% et 99,19% sont alors obtenus. Comme nous l’avons expliqué précédemment, ce test est très biaisé, et nous pouvons le voir à travers les tests expérimentaux réalisés dans (MOHAMMAD et al. 2018), i.e., en plus d’une validation indépendante du sujet, une validation croisée simple est aussi appliquée. Les résultats ont montré que le score est passé de 86,03% à 98,40%. De plus, si une validation dépendante du sujet est appliquée en choisissant un degré de chevauchement O_v très important comme dans (ARIF et al. 2014) ($O_v = 75%$) ou (VAVOULAS et al. 2016) ($O_v = 80%$), des scores avoisinant les 99% peuvent être obtenus.

nus. Pour confirmer ces prévisions, des tests expérimentaux sont réalisés en fonction du degré de chevauchement O_v . Le résultat illustré dans la figure 4-11 montre une courbe monotone croissante, qui commence à 94% quand $O_v = 0\%$, et monte jusqu'à 99,26% quand $O_v = 90\%$.

Validation	Approche	Rap (%)	Pr (%)	F-s (%)	Ex (%)
LOSOCV ¹	(KOLOSNAJI et al. 2015)	83,31	80,03	81,64	85,36
	Nous	90,55	88,24	89,38	90,91
SCV ²	(MOHAMMAD et al. 2018)	-	-	-	86,03
HS ³	(IGNATOV 2018)	88,59	-	-	93,32
CV ⁴	(KWAPISZ et al. 2011)	80,46	81,56	81,00	85,33
	(CATAL et al. 2015)	91,61	92,08	91,85	94,35
	(MOHAMMAD et al. 2018)	-	-	-	98,40
	(VAVOULAS et al. 2016)	99,67	-	-	99,79
	(ARIF et al. 2014)	99,08	99,29	99,19	99,43

¹ Leave-one-subject-out-cross-validation.

² k -fold-subject-cross-validation.

³ Hold-out subjects.

⁴ k -fold cross-validation.

TABLE 4.12 – Scores obtenus en comparaison avec l'état de l'art pour le jeu de données WISDM.

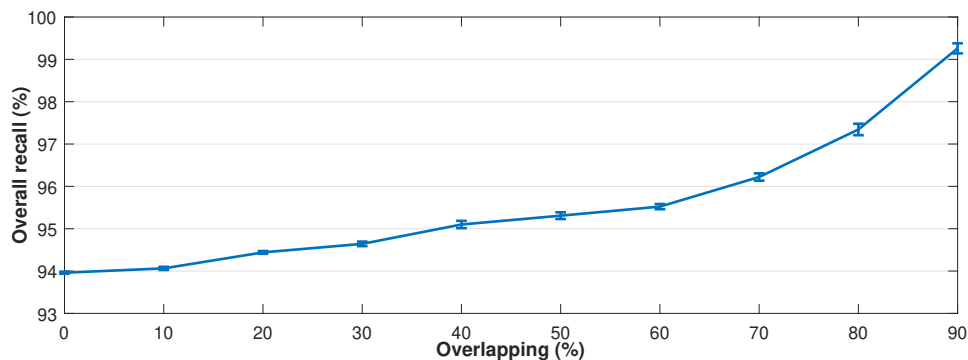


FIGURE 4-11 – Score obtenu en fonction du degré de chevauchement avec un test 10-fold-subject-cross-validation.

Le tableau 4.13 montre les scores obtenus pour chaque classifieur et les dimensions des vecteurs d'attributs correspondants. Nous pouvons voir que la classification des actions en méta-classes "posture" et "quasi-périodique" nécessite seulement 4 attributs. Un score de 98,28% est alors obtenu. Pour les trois actions quasi-périodiques :

“marcher”, “monter les escaliers” et “descendre les escaliers”, nous atteignons un score de 93,79% avec 31 attributs. En ce qui concerne les postures “debout” et “assis”, un score de 85,26% est obtenu. Ceci est probablement dû à la position du smartphone (poche du pantalon avec orientation libre), ajouté à cela l’exploitation d’un seul capteur et la faible quantité de données pour les classes de type posture (5,46% pour la classe “assis” et 4,41% pour la classe “debout”).

	Catégorie	Quasi-périodique	Posture
F-s (%)	98,28	93,79	85,26
$Dim(F_i)_{i \in \{C,D,P\}}$	4	31	5

TABLE 4.13 – Scores pour chaque classifieur et dimensions des attributs pour le jeu de données WISDM.

Actions humaines et transitions posturales

Dans cette partie, nous exploitons les jeux de données “SBHART” et “UniMiB SHAR”, contenant à la fois des actions humaines et des transitions posturales.

Pour le jeu de données “SBHART”, la plupart des travaux existants testés sur ce jeu ne prennent pas en considération les transitions posturales. Pour une comparaison équitable, dans un premier temps, le score obtenu pour les transitions posturales n’est pas inclus dans le calcul du score global, et cela bien que les transitions sont considérées dans les tests expérimentaux. Il est à noter que dans ce cas nous sommes moins avantagés que les autres approches car certaines confusions peuvent apparaître entre la classe “transitions posturales” et les autres classes. Le tableau 4.14 présente les scores que nous avons obtenus suivant ce calcul, plus une comparaison avec les méthodes de la littérature. Un score de 96,94% est alors obtenu avec la technique de validation LOSOCV, supérieur à celui des approches existantes. Afin de se comparer à d’autres approches, la technique “Hold-out subjects” est aussi appliquée comme nous l’avons expliqué dans la partie 4.3.2, nous obtenons alors un score de 96,75%, au-dessus de celui des approches déjà proposées.

Comme nous l’avons constaté précédemment, des tests dépendants du sujet ont aussi été effectués avec le jeu de données “SBHART”. Dans les travaux de SAN-SEGUNDO, ECHEVERRY-CORREA et al. (2016), 6-fold cross-validation est appliquée. Le jeu de données est alors divisé aléatoirement en six ensembles de données ; quatre ensembles sont exploités pour la phase d’apprentissage, un autre pour la validation (réglage des paramètres du système) et le dernier ensemble pour le test. L’expérience est répétée six fois en prenant à chaque fois un sous-ensemble différent pour le test. Un score “exactitude” moyen de 94,70% est obtenu. Dans (SUAREZ et al. 2015a; HASSAN et al. 2018), le jeu de données a été divisé aléatoirement en deux ensembles d’apprentissage (70% des données) et de test (30% des données). Un score “exactitude” de 94,23% est alors obtenu dans (SUAREZ et al. 2015a), et un score “rappel” de 97,05% dans (HASSAN et al. 2018).

En ce qui concerne les transitions posturales, la figure 4-12 montre un exemple de performance permettant de reconnaître les postures et les transitions posturales en utilisant notre approche. Notez que dans les extrémités des actions, les étiquettes prédites ne correspondent pas toujours à la vérité du terrain. Ces résultats correspondent à la situation d’une fenêtre glissante incluant une classe de transition en plus de la posture. En considérant les conditions d’évaluation de ce cas particulier, mentionnées dans la partie 4.3.1, les étiquettes prédites dans les extrémités des actions comme le montre la figure 4-12 sont bien classées.

La matrice de confusion pour toutes les actions (incluant les transitions posturales) est décrite dans le tableau 4.15. Elle montre que toutes les actions quasi-périodiques et les transitions posturales atteignent de très bons scores de performances, entre 96,08% pour l’action “descendre les escaliers” et 100% pour les transitions posturales. À titre de comparaison REYES-ORTIZ et al. (2016) ont obtenu un score de 92,18% pour les transitions posturales avec la première méthode proposée et 99,76% avec la deuxième. La matrice montre aussi quelques confusions pour les postures “assis” et “debout” qui sont dues à la position du smartphone (la taille) rendant les signaux de sortie similaires.

Jusqu’à présent les transitions posturales étaient considérées comme étant une

Validation	Approche	Rap (%)	Pr (%)	F-s (%)	Ex (%)
LOSOCV ¹	(REYES-ORTIZ et al. 2016) (m_1)	-	-	-	96,74
	(REYES-ORTIZ et al. 2016) (m_2)	-	-	-	96,50
	(KOLOSNAJI et al. 2015)	78,23	78,45	78,34	77,81
	Nous	96,92	96,96	96,94	96,92
HS ²	(ANGUITA et al. 2012)	89,35	89,23	89,29	88,97
	(Y.-J. KIM et al. 2015)	-	-	-	83,51
	(A. WANG et al. 2016) (m_1)	88,12	87,30	87,71	87,03
	(A. WANG et al. 2016) (m_2)	90,35	89,75	90,05	90,13
	(RONAO et al. 2017)	92,67	92,95	92,81	92,92
	(JIANG et al. 2015)	-	-	-	95,18
	(RONAO et al. 2016) (m_1)	94,79	94,78	94,79	94,50
	(RONAO et al. 2016) (m_2)	95,75	-	-	-
	(MOHAMMAD et al. 2018)	-	-	-	95,46
	(ZDRAVEVSKI et al. 2017)	95,90	95,97	95,93	-
	(SEERA et al. 2014)	96,59	96,86	96,72	96,64
Nous	96,67	96,83	96,75	96,69	
CV ³	(SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016)	-	-	-	94,70 \pm 2, 1
H ⁴	(SUAREZ et al. 2015a)	-	-	-	94,23
	(HASSAN et al. 2018)	97,05	-	-	-

¹ Leave-one-subject-out cross-validation.

² Hold-out subjects.

³ k -fold cross-validation.

⁴ Hold-out (découpage des données aléatoire).

TABLE 4.14 – Scores obtenus en comparaison avec l'état de l'art pour le jeu de données SBHART.

	M	ME	DE	As	D	AI	TP
Marcher	99,00	1,00	0	0	0	0	0
M. escaliers	0,46	99,08	0,46	0	0	0	0
D. escaliers	1,57	2,35	96,08	0	0	0	0
Assis	0	0	0	91,85	8,15	0	0
Debout	0	0	0	4,48	95,52	0	0
Allongé	0	0	0	0	0	100	0
T. posturales	0	0	0	0	0	0	100

TABLE 4.15 – Matrice de confusion (%) pour le jeu de données SBHART.

seule classe. Afin de reconnaître la nature de la transition, l'algorithme défini dans la partie 4.2.3 est appliqué. Les scores obtenus sont présentés dans le tableau 4.16 avec quelques approches de l'état de l'art testées sur le jeu de données "SBHART". Un

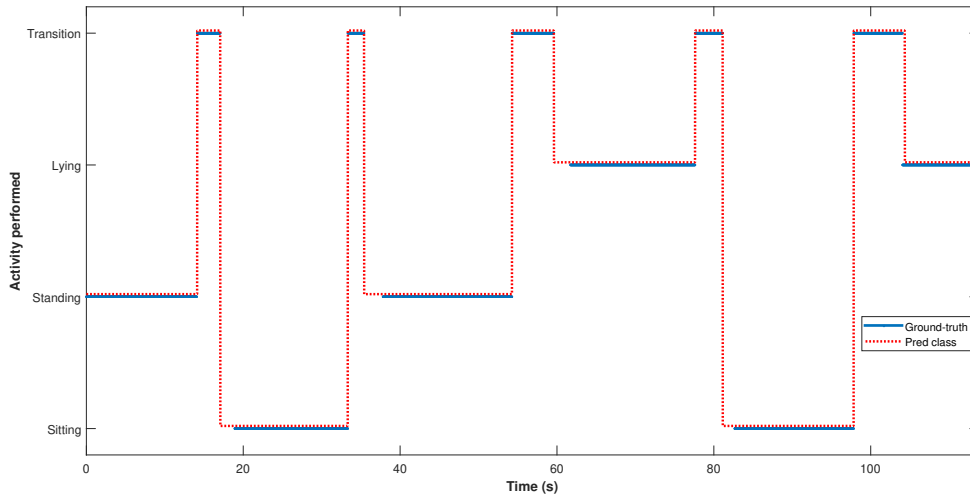


FIGURE 4-12 – Détection des transitions dans le jeu de données SBHART.

score global de 92,73% est obtenu avec notre approche. Dans (HASSAN et al. 2018), l’approche utilisant un réseau de croyances profond a donné un score “rappel” de 82,18%, largement inférieur au nôtre. Dans (NOOR et al. 2017), un score de 95,10% est obtenu, cependant, seules quatre transitions posturales sont considérées dans les tests expérimentaux, les transitions “debout vers allongée” et “allongée vers debout” (les transitions présentant d’ailleurs le plus d’erreurs) sont écartées.

Validation	Approche	Rap (%)	Pr(%)	F-s (%)	Ex (%)
LOSOVCV ¹	Nous	92,68	92,78	92,73	92,64
H ²	(HASSAN et al. 2018)	82,18	-	-	-
	(NOOR et al. 2017) (4PT ³)	95,10	-	-	-

¹ Leave-one-subject-out cross-validation.

² Hold-out (découpage des données aléatoire).

³ Quatre transitions posturales considérées dans les tests expérimentaux.

TABLE 4.16 – Scores de reconnaissance des transitions posturales sur le jeu de données SBHART.

Afin de montrer les résultats en détail, la matrice de confusion est présentée dans le tableau 4.17. Nous pouvons remarquer que toutes les transitions sont bien reconnues avec un score supérieur ou égal à 90%. Nous observons cependant de légères confusions entre la transition “debout vers allongée” et les transitions “debout vers assise”, “assise vers allongée”. Elle est due au fait que pour certains sujets, la transition “debout vers

allongée” est effectuée en deux étapes : “debout vers assise” suivie de “assise vers allongée ”. Ce qui peut créer des confusions dans le cas où une fenêtre glissante intercepte une partie des signaux “debout vers allongé” au lieu de l’ensemble des informations de l’action. La même remarque peut être faite pour la transition “allongée vers debout”.

	De-As	As-De	As-Al	Al-As	De-Al	Al-De	Rappel (%)
Debout vers assis	35	0	0	0	3	0	92,11
Assis vers debout	0	39	0	0	0	1	97,50
Assis vers allongé	0	0	34	0	3	0	91,89
Allongé vers assis	0	0	0	32	0	2	94,12
Debout vers allongé	2	0	2	0	36	0	90,00
Allongé vers debout	0	2	0	2	0	38	90,48

TABLE 4.17 – Matrice de confusion des transitions posturales sur le jeu de données SBHART.

Le tableau 4.18 montre les scores obtenus pour chaque classifieur de notre système et les dimensions des vecteurs d’attributs. Nous constatons que les actions sont séparées parfaitement en trois catégories (périodique, posture et transitions posturales) avec seulement deux attributs. Pour les actions périodiques, 46 attributs ont été sélectionnés pour atteindre un score de 98,07%, et seulement 7 attributs pour les postures. En général le nombre d’attributs sélectionnés est relativement faible comparé à celui des autres approches de l’état de l’art testées sur le jeu de données “SBHART”. Dans (REYES-ORTIZ et al. 2016 ; SAN-SEGUNDO, ECHEVERRY-CORREA et al. 2016 ; HASSAN et al. 2018) par exemple, 561 attributs sont calculés à partir du domaine temporel et fréquentiel de neuf signaux (signaux générés par filtrage, voir la partie 3.6 pour plus de détails). Dans (NOOR et al. 2017), 169 attributs sont calculés. Ce nombre important d’attributs calculés peut avoir un impact non négligeable dans la consommation d’énergie de l’appareil sur lequel est embarqué le système.

Dans le jeu de données “UniMiB SHAR”, les données sont collectées suivant deux positions de smartphone, poche droite et gauche. À noter que les tests relatifs à ce jeu de données sont présentés dans cette partie plutôt que dans la dernière (tests sur des jeux de données avec des capteurs dans diverses positions). Nous estimons que les

	Catégorie	Quasi-périodique	Posture	TP ¹
F-s (%)	100	95,81	98,07	92,73
$Dim(F_i)_{i \in \{C,D,P\}}$	2	46	7	-

¹ Transitions posturales.

TABLE 4.18 – Scores pour chaque classifieur et dimensions des attributs pour le jeu de données SBHART.

données issues des positions poche droite et gauche du smartphone ne sont pas assez éloignées dans l’espace des attributs, en comparaison avec les données issues d’autres positions comme le poignet par exemple. Cependant des tests sont effectués avec une seule position à la fois puis en combinant les deux. Les résultats sont alors présentés dans le tableau 4.19. Ils montrent un score de 93,47% obtenu avec un smartphone dans la poche droite, et 92,79% quand le smartphone est à gauche. En combinant les deux positions dans la phase d’apprentissage et de test, nous obtenons un score légèrement supérieur qui est de 94,36%. Nous remarquons que le score “exactitude” est plus élevé que les autres scores car le jeu de données est déséquilibré (les classes “marcher” et “courir” sont majoritaires). À noter que la comparaison avec les travaux de l’état de l’art n’est pas réalisée car les conditions de test ne sont pas les mêmes (le nombre de classes est différent). Nous indiquons par exemple que dans les travaux de (MICUCCI et al. 2017), différents classifieurs sont exploités. Quatre problèmes de classification sont définis. Dans le premier toutes les classes sont considérées, le meilleur score “rappel” obtenu est de 56,58%. Dans le deuxième problème, les chutes sont écartées, un score de 72,67% est obtenu. Inversement pour le troisième problème, où seulement les chutes sont considérées, un score de 49,35% est obtenu. Enfin dans le dernier problème, une classification binaire est effectuée sur deux groupes, le premier inclut les actions quasi-périodiques et les transitions posturales, le second groupe contient les types de chutes. Un score de 97,57% est alors obtenu dans ce cas.

La matrice de confusion obtenue avec les deux positions de smartphone (poches droite et gauche) est présentée dans le tableau 4.20. Elle montre que les classes “marcher” et “courir” (classes majoritaires) sont bien reconnues avec un score de 99,08% et 94,36% respectivement. Quelques erreurs sont observées entre la classe “descendre

Validation	Approches	Pos	Rap (%)	Pr (%)	F-s (%)	Ex (%)
LOSOCV ¹	Nous	P_4	92,89	94,05	93,47	96,14
		P_5	91,15	94,49	92,79	94,51
		P_i	93,60	95,13	94,36	96,43

¹ Leave-one-subject-out cross-validation.

TABLE 4.19 – Scores obtenus en comparaison avec l’état de l’art pour le jeu de données UniMiB SHAR.

les escaliers” et les classes similaires à savoir : “marcher” et “monter les escaliers”. Les actions montrent globalement un bon score, nous observons cependant certaines confusions entre les classes “sauter” et “courir”. Ce résultat était prévisible puisque le mouvement courir se caractérise par une suite d’élan. Des confusions sont aussi observées entre “la chute” et “les transitions posturales”. Les résultats détaillés ont montré que les classes “debout vers allongé sur un lit” et “debout vers assis sur une chaise” sont les plus concernées par ces erreurs. Cela est dû au fait que ces classes sont très similaires à certains types de chutes comme par exemple : chute vers l’arrière en essayant de s’asseoir sur une chaise.

	M	ME	DE	Co	S	Ch	TP
Marcher	99,08	0	0,09	0,18	0	0,28	0,37
M. escaliers	2,26	90,68	3,67	0	0,28	0,57	2,54
D. escaliers	4,97	5,23	85,40	1,10	0	1,10	2,20
Courir	1,68	0	1,09	94,36	2,57	0,20	0,1
Sauter	0	0	0,62	4,95	91,02	3,41	0
Chute	0,03	0	0	0,03	0	99,56	0,38
T. posturales	0,05	0,05	0	0	0	4,82	95,08

TABLE 4.20 – Matrice de confusion (%) pour le jeu de données UniMiB SHAR.

Les tableaux 4.21a, 4.21b et 4.21c montrent les scores obtenus pour chaque classifieur de notre système et les dimensions des vecteurs d’attributs. Ces résultats sont séparés suivant les positions de smartphone considérées dans la phase d’apprentissage et de test, à savoir : poche droite, poche gauche, et la combinaison des deux. Les résultats montrent que dans le premier niveau de classification (la séparation des actions en catégories quasi-périodiques et actions), un score autour de 98% est atteint et entre 10 et 14 attributs seulement sont sélectionnés à cet égard. Dans le deuxième niveau

de classification, 42 attributs sont sélectionnés quand le smartphone est mis dans la poche gauche et 50 attributs pour le cas des deux poches. Un score de moyen autour de 92% est obtenu. Et enfin, les actions ont donné un score relativement élevé comparé à celui des actions quasi-périodiques (autour de 95%), et cela en sélectionnant 11 jusqu'à 16 attributs seulement.

(a) Position : poche droite.			
	Catégorie	Quasi-périodique	Non périodique
F-s (%)	98,74	91,68	95,80
$Dim(F_i)_{i \in \{C,D,A\}}$	11	46	14
(b) Position : poche gauche.			
	Catégorie	Quasi-périodique	Non périodique
F-s (%)	97,87	90,88	95,32
$Dim(F_i)_{i \in \{C,D,A\}}$	14	42	11
(c) Position : poches droite et gauche.			
	Catégorie	Quasi-périodique	Non périodique
F-s (%)	98,89	93,64	95,30
$Dim(F_i)_{i \in \{C,D,A\}}$	10	50	16

TABLE 4.21 – Scores pour chaque classifieur et dimensions des attributs pour le jeu de données UniMiB SHAR.

Capteurs dans diverses positions

Dans cette partie, des données inertielles collectées suivant plusieurs positions de capteurs sont exploitées. Les jeux de données “mHealth” et “Multi-position smartphone” sont donc utilisés pour ces tests.

Les résultats expérimentaux obtenus pour le jeu de données *mHealth* sont présentés dans le tableau 4.22. Pour notre approche, deux types de tests sont présentés ici : test dépendant de la position et test indépendant de la position (P_i). Dans le premier test, nous pouvons voir que globalement la position “avant-bras droit” (P_8) réalise le meilleur score à 92,31%. Un score de 91,82% est obtenu avec la position “poitrine” et le score le plus bas est de 87,49% correspondant à la position “cheville gauche”. Le

test indépendant de la position donne un score de 91,45%. Comme nous l’avons expliqué dans la partie 3.4, deux axes de travail existent pour la reconnaissance d’actions suivant plusieurs positions de capteurs : le premier, celui adopté par notre système, consiste à entraîner le modèle sur les données d’actions collectées suivant plusieurs positions de capteurs. La deuxième catégorie d’approche repose quant à elle sur deux phases : la reconnaissance de l’emplacement des capteurs dans un premier temps, suivie par la sélection d’un classifieur entraîné sur des données associées à l’emplacement détecté. Dans l’hypothèse où cette deuxième approche serait sélectionnée pour notre système, dans le meilleur des cas (score de 100% pour la reconnaissance de l’emplacement des capteurs), nous obtiendrons un score de 90,54% (moyenne des scores obtenus pour chaque position). Nous pouvons déduire alors que notre approche vient d’atteindre de meilleur score de performance (91,82%) que dans le meilleur des cas d’un système identique basé sur le suivi de l’emplacement des capteurs. Dans les travaux de (HA et al. 2016), l’approche basée sur les réseaux de neurones à convolution est entraînée sur des données issues des trois emplacements simultanément (test P_a). De plus, les gyroscopes sont aussi utilisés ce qui n’est pas le cas dans nos travaux où seulement les données d’accélération sont sélectionnées (revoir la partie 4.3.1 pour plus d’explications). Cependant les 12 actions sont considérées. Un score de 91,94% est alors obtenu. Dans la méthode de (BANOS, VILLALONGA et al. 2015), la technique 10-fold cross-validation est appliquée. Trois capteurs sont utilisés (accéléromètre, gyroscope et magnétomètre). Le test P_a a donné un score de 97,58%. Dans (SAID et al. 2018), le jeu de données a été divisé aléatoirement en deux ensembles d’apprentissage (80% des données) et de test (20% des données). En plus des signaux issues des trois capteurs inertiels, des signaux électrocardiogramme (ECG) sont utilisés. Un score “exactitude” de 98,1% est obtenu avec le test P_a . Les auteurs de (MEHMOOD et al. 2016) se sont restreints à quatre actions seulement : “debout”, “sauter”, “cyclisme” et “se pencher en avant”. L’ensemble des signaux sont utilisés dans leur système. Un découpage aléatoire en deux ensembles est choisi pour la partie expérimentale, avec 70% des données pour l’apprentissage et le reste pour la validation. Le test P_a a montré un score de 99,79%.

Validation	Approches	N_{cap}	N_{act}	Pos	Rap (%)	Pr (%)	F-s (%)	Ex (%)
LOSOCV ¹	(HA et al. 2016)	2	12	P_a	-	-	-	91,94
	Nous	1	8	P_i	90,82	92,08	91,45	91,67
				P_7	90,78	92,88	91,82	91,33
				P_8	91,13	93,52	92,31	92,09
				P_{10}	87,36	87,63	87,49	87,81
CV ²	(BANOS, VILLALONGA et al. 2015)	3	12	P_a	97,5	97,75	97,58	-
H ³	(SAID et al. 2018)	4	12	P_a	-	-	-	98,1
	(MEHMOOD et al. 2016)	4	4	P_a	-	-	-	99,79

¹ Leave-one-subject-out cross-validation.

² k -fold cross-validation.

³ Hold-out.

TABLE 4.22 – Scores obtenus en comparaison avec l’état de l’art pour le jeu de données mHealth.

Outre les tests dépendants et indépendants de la position, des tests “positions croisées” sont appliqués comme il est expliqué dans la partie 4.3.2. Les résultats obtenus sont séparés suivant le problème de classification. Les tableaux 4.23a, 4.23b, 4.23c, 4.23d et 4.23e montrent respectivement les résultats obtenus pour la reconnaissance de l’ensemble des actions, la catégorie d’action, les actions quasi-périodiques, les postures et les actions. Les tests dépendants de la position sont indiqués dans la diagonale de la partie des tableaux délimités par des pointillés (en gras). Les résultats hors diagonale correspondent aux tests “positions croisées”. Le test indépendant de la position est quant à lui décrit dans les dernières lignes des tableaux.

Comme il a été indiqué précédemment, la position “poignet” a montré le meilleur score global dans les tests dépendants de la position. Les résultats obtenus dans chaque catégorie d’action montrent que cela est intégralement dû aux scores obtenus pour la reconnaissance des postures : comme le montre le tableau 4.23d, la position “poignet” a réalisé un écart de +7,33% par rapport à la position “poitrine” et +13,33% par rapport à la position “cheville”. Cependant, pour les autres catégories actions, c’est la position “poitrine” qui réalise le meilleur score, un écart par rapport à la position “poignet” de +3,74% pour les actions quasi-périodiques et +2,81% pour les actions.

En ce qui concerne le test indépendant de la position, les résultats observés révèlent que les actions quasi-périodiques et les actions atteignent de bons scores, 96,68% et

95,91% respectivement. Les postures quant à elles montrent un score plus bas à 82,96%.

Les tests “positions croisées” montrent des résultats intéressants avec les scores globaux entre 74,25% et 85,54%. La meilleure paire apprentissage/test correspond à la combinaison poignet/poitrine. Elle présente un score important de 96,91% pour les actions quasi-périodiques et des scores relativement faibles pour les postures et les actions (73,27% et 70,09% respectivement). La paire poignet/poitrine est talonnée par cheville/poignet avec un score global de 84,95%. Cette combinaison présente des résultats plutôt équilibrés pour les actions quasi-périodiques et postures avec des scores respectifs de 87,98% et 87,30%. Néanmoins les actions montrent un score faible égal à 64,04%. Les deux paires apprentissage/test restantes atteignant un bon score (au dessus de 80%) correspondent aux combinaisons poitrine/poignet et poignet/cheville avec respectivement des scores de 82,66% et 82,25%. Pour finir, les scores les plus faibles correspondent au croisement des positions poitrine et cheville avec un score de 74,25% pour poitrine/cheville et 76,23% pour cheville/poitrine.

À présent, nous nous intéressons aux dimensions des vecteurs d’attributs de notre système. Cette information est présentée dans les tableaux 4.24a, 4.24b et 4.24c. Les tests dépendants de la position montrent que pour séparer les actions en catégories (quasi-périodique, posture et actions) entre 4 et 12 attributs seulement sont utilisés. Par la suite au deuxième niveau de classification, la position “poitrine” a montré que les actions quasi-périodiques sont reconnues en exploitant 11 attributs. Ces derniers sont prélevés à partir des spectres des signaux. Plus précisément, 10 attributs sont extraits du spectre de l’accélération suivant l’axe x (A_x), et 1 attribut du spectre suivant l’axe y (A_y). Ce constat indique que les signaux d’accélération A_z et A_m (magnitude des signaux) ne sont pas exploités pour la reconnaissance des actions quasi-périodiques. Par conséquent, il n’est pas nécessaire de calculer le spectre de ces signaux d’où réduction du temps de calcul pour notre système. Pour rappel, la position “poitrine” a montré le meilleur score de reconnaissance des actions quasi-périodiques (98,59%) comparé à celui des autres positions. Pour la position “cheville”, 10 attributs sont utilisés seulement pour la reconnaissance des actions quasi-périodiques. Ils sont

(a) F-score global.

		Test et sélection d'attributs			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	91,82	82,66	74,25	-
	Poignet	85,54	92,31	82,25	-
	Cheville	76,23	84,95	87,49	-
	Ensemble (P_i)	93,87	90,72	90,86	91,45

(b) F-score du premier niveau de classification (catégorie de l'action).

		Test et sélection d'attributs			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	98,72	90,95	81,43	-
	Poignet	90,92	98,08	96,61	-
	Cheville	79,14	87,43	97,16	-
	Ensemble (P_i)	99,54	96,58	99,53	98,56

(c) F-score des actions quasi-périodiques.

		Test et sélection d'attributs			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	98,59	88,95	78,44	-
	Poignet	96,91	94,85	91,18	-
	Cheville	83,45	87,98	95,81	-
	Ensemble (P_i)	98,70	93,62	97,90	96,68

(d) F-score des postures.

		Test et sélection d'attributs			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	80,83	77,13	74,58	-
	Poignet	73,27	88,16	67,58	-
	Cheville	71,36	87,30	74,83	-
	Ensemble (P_i)	85,81	86,95	78,76	82,96

(e) F-score des actions non périodiques.

		Test et sélection d'attributs			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	97,33	72,44	40,00	-
	Poignet	70,09	94,52	89,71	-
	Cheville	26,46	64,04	91,82	-
	Ensemble (P_i)	98,68	90,00	98,65	95,91

TABLE 4.23 – Scores obtenus pour le jeu de données mHealth.

répartis comme suit : 3 attributs sont extraits du spectre du signal suivant l'axe x , les 4 suivants sont extraits selon l'axe y et 3 derniers selon l'axe z . Dans ce cas-là, le signal magnitude A_m peut être écarté du système réduisant ainsi le temps de calcul.

Pour la troisième position le “poignet” plus d’attributs sont sélectionnés, 38 au total répartis comme suit : 9 attributs extraits du spectre de A_x , 11 extraits du spectre de A_y , 10 extraits du spectre de A_z et 8 extraits du spectre de A_m . Ici contrairement aux positions “poitrine” et “cheville”, tous les signaux sont nécessaires pour le calcul du vecteur d’attributs. Les résultats obtenus pour la reconnaissance des postures sont présentés dans le tableau 4.24c. Ils montrent que 6 attributs sont calculés pour la position “poitrine”, 4 attributs pour la position “poignet” et 2 attributs seulement pour la position “cheville”.

Dans le test indépendant de la position, 9 attributs sont sélectionnés afin de séparer les actions en catégories alors que la position “poitrine” nécessite 12 attributs. De plus, comme l’indique le tableau 4.23b, le modèle construit avec le test indépendant de la position reconnaît mieux les catégories d’actions que le modèle construit en utilisant seulement les données issues de la position “poitrine”, et cela avec moins d’attributs en entrée du classifieur concerné. Cela montre que des données issues d’une position différente de celle qui est considérée peuvent contribuer à mieux reconnaître celle-ci. Dans le but de reconnaître les actions quasi-périodiques, 22 attributs sont extraits des spectres des signaux, A_x , A_y et A_z comptent 6 attributs chacun et 4 attributs pour A_m . La même remarque peut être faite que précédemment. Le modèle construit avec le test indépendant de la position présente de meilleures performances que les modèles construits avec des données d’une seule position à savoir “poitrine” et “cheville”. Pour reconnaître les postures, 11 attributs sont calculés donnant un résultat meilleur suivant les positions “poitrine” et “cheville” que les tests dépendants de la position.

En ce qui concerne les tests “positions croisées”, entre 5 et 11 attributs sont calculés pour reconnaître la catégorie de l’action. Pour les actions quasi-périodiques, entre 10 et 39 attributs sont extraits des spectres. Pour la combinaison cheville/poignet par exemple, 11 attributs sont extraits au total. 3 attributs extraits des spectres de chacun des signaux A_x et A_m , 4 attributs extraits du spectre de A_y et 1 attribut du spectre de A_z . Ce dernier correspond à la composante continue du spectre. Par conséquent, il peut être calculé sans le passage vers le domaine fréquentiel (valeur moyenne du

signal), réduisant ainsi le temps de calcul. Pour reconnaître les postures, moins de 6 attributs sont extraits des signaux. La combinaison poignet/cheville montre par exemple qu’un seul attribut est calculé (correspondant à la moyenne du signal A_y). Cependant, un score relativement faible est obtenu (67,58%), cela est dû à la difficulté de séparer les données issues de la position “cheville” dans l’espace des attributs pour les postures “debout” et “assis”.

(a) mHealth : Dimension des attributs pour la catégorie de l’action (F_C).

		Test			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	12	9	6	-
	Poignet	10	8	11	-
	Cheville	5	9	4	-
	Ensemble (\bar{P}_i)	-	-	-	9

(b) mHealth : Dimension des attributs pour les actions quasi-périodique (F_Q).

		Test			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	11	20	10	-
	Poignet	22	38	39	-
	Cheville	12	11	10	-
	Ensemble (\bar{P}_i)	-	-	-	22

(c) mHealth : Dimension des attributs pour postures (F_P).

		Test			
		Poitrine	Poignet	Cheville	Ensemble (P_i)
Train	Poitrine	6	4	5	-
	Poignet	3	4	1	-
	Cheville	3	5	2	-
	Ensemble (\bar{P}_i)	-	-	-	11

TABLE 4.24 – Dimension des attributs pour le jeu de données mHealth.

Nous passons maintenant au deuxième jeu de données nommé “Multi-position smartphone” dans lequel les signaux sont collectés suivant plusieurs positions de smartphone. Les résultats expérimentaux obtenus sont indiqués dans le tableau 4.25. Comme pour le jeu de données “mHealth”, deux types de tests sont appliqués dans notre approche : tests dépendants et indépendants à la position. Les tests dépendants de la position montrent que les poches droite et gauche (P_2 et P_3) obtiennent les meilleurs scores, respectivement 99,22% et 99,24% , talonnés par la position “han-

che” (P_6), qui atteint un score de 98,92%. Les positions “poignet” et “haut du bras” (P_8 et P_9) enregistrent des scores relativement bas, respectivement 90,10% et 88,79%. En ce qui concerne le test indépendant de la position (P_i), un score “rappel” de 94,41% est obtenu. Dans les travaux de SAN BUENAVENTURA et al. (2018), deux méthodes sont proposées, la première basée sur un modèle LSTM a donné un score “rappel” de 80,02%, la seconde, utilisant des réseaux de neurones à convolution a atteint un score “rappel” meilleur qui est de 90,61%. De plus, les deux méthodes utilisent les signaux issus d’un magnétomètre ce qui n’est pas le cas dans notre approche qui donne un score bien supérieur.

Validation	Approches	Pos	Rap (%)	Pr (%)	F-s (%)	Ex (%)
LOSOCV ¹	(SAN BUENAVENTURA et al. 2018) (m1)	P_i	80,02	-	-	80,02
	(SAN BUENAVENTURA et al. 2018) (m2)	P_i	90,61	-	-	90,61
	Nous	P_i	94,32	94,50	94,41	94,32
		P_2	99,24	99,24	99,24	99,24
		P_3	99,22	99,23	99,22	99,22
		P_6	98,91	98,93	98,92	98,91
		P_8	89,74	90,46	90,10	89,74
P_9	88,63	88,95	88,79	88,63		

¹ Leave-one-subject-out cross-validation.

TABLE 4.25 – Scores obtenus en comparaison avec l’état de l’art pour le jeu de données Multi-position.

De la même manière que les expérimentations réalisées sur le jeu de données “mHealth”, des tests “positions croisées” sont aussi effectués. Les tableaux 4.26a, 4.26b, 4.26c et 4.26d montrent respectivement les résultats obtenus pour la reconnaissance de l’ensemble des actions, la catégorie d’action, les actions quasi-périodiques et les postures. Dans les tests dépendant à la position, nous avons indiqué précédemment que les positions “poche droite”, “poche gauche” et “hanche” ont présenté les meilleurs scores. Les résultats obtenus pour les actions quasi-périodiques et postures révèlent des scores autour de 99% pour ces positions. Les emplacements “poignet” et “haut du bras” donnent quant à eux des scores relativement bas pour les actions quasi-périodiques, respectivement 86,98% et 91,50%. Pour les postures, la position “haut

du bras” est la seule à atteindre un score bas (81, 89%).

Le test indépendant de la position indique que les actions sont bien séparées en catégories avec un score de 98, 99%. Dans le deuxième niveau de classification, les actions quasi-périodiques et les postures atteignent des scores satisfaisants, 94, 41% et 94, 19% respectivement. Nous remarquons aussi que les scores obtenus pour chaque position (la dernière ligne dans le tableau) sont très proches des scores obtenus dans les tests dépendants de la position.

En ce qui concerne les tests “positions croisées”, ils présentent des résultats intéressants. Le croisement des positions poche droite et poche gauche montre des scores très élevés, ce qui est cohérent vu que les données des deux emplacements sont très similaires. C’est pour cette raison d’ailleurs que des tests étaient effectués en combinant les données des deux positions dans le jeu de données “UniMiB SHAR”. Néanmoins, une baisse est constatée pour la reconnaissance des actions quasi-périodiques. Les autres croisements de positions montrent que les scores globaux varient entre 62, 01% et 80, 59%, la meilleure performance correspond au croisement poignet/hanche et la plus faible au croisement bras/poche gauche. Les résultats obtenus sur chaque catégorie d’action indiquent des scores disproportionnés, en effet les postures présentent des scores largement meilleurs que les actions quasi-périodiques. Nous notons des scores allant jusqu’à 100% pour les postures.

Pour compléter cette partie, nous présentons la dimension des vecteurs d’attributs de notre système de la même manière que celle effectuée sur le jeu de données “mHealth”. Cette information est indiquée dans les tableaux 4.27a, 4.27b et 4.27c. Les tests dépendants de la position indiquent qu’entre 4 et 12 attributs seulement sont utilisés pour séparer les actions en catégories (même constat que dans le jeu de données “mHealth”). Afin de reconnaître les actions périodiques, entre 41 et 76 attributs sont sélectionnés parmi les spectres des signaux d’accélération et rotation angulaire. Les postures quant à elles, nécessitent peu d’attributs (moins de 13 attributs pour toutes les positions). Les emplacements “poche droite” et “poignet” par exemple utilisent uniquement la moyenne du signal d’accélération A_y afin de construire le vecteur d’attributs et cela a permis d’obtenir des scores respectifs de 99, 43% et 97, 90%.

(a) F-score global.

		Test et sélection d'attributs					
		Poche. D	Poche. G	Hanche	Poignet	Bras	Ensemble (P_i)
Train	Poche. D	99,24	94,96	77,41	74,88	80,50	-
	Poche. G	91,63	99,22	74,75	75,48	73,80	-
	Hanche	70,81	66,94	98,92	71,15	76,58	-
	Poignet	66,92	74,95	80,59	90,10	80,32	-
	Bras	67,94	62,01	75,83	79,53	88,79	-
	Ensemble (P_i)	99,15	98,84	98,82	89,57	86,51	94,41

(b) F-score du premier niveau de classification (catégorie de l'action).

		Test et sélection d'attributs					
		Poche. D	Poche. G	Hanche	Poignet	Bras	Ensemble (P_i)
Train	Poche. D	99,60	100	93,04	99,45	99,42	-
	Poche. G	99,87	99,75	94,86	98,61	99,50	-
	Hanche	99,95	99,70	99,82	98,64	98,47	-
	Poignet	99,95	100	96,71	99,27	98,91	-
	Bras	100	100	99,55	98,85	96,87	-
	Ensemble (P_i)	99,92	100	99,75	99,22	96,23	98,99

(c) F-score des actions quasi-périodiques.

		Test et sélection d'attributs					
		Poche. D	Poche. G	Hanche	Poignet	Bras	Ensemble (P_i)
Train	Poche. D	99,17	92,94	72,26	65,10	79,32	-
	Poche. G	88,33	99,05	75,44	66,46	77,78	-
	Hanche	59,48	57,44	98,65	63,69	72,59	-
	Poignet	53,22	64,93	74,28	86,98	78,33	-
	Bras	54,82	47,89	70,19	73,95	91,50	-
	Ensemble (P_i)	98,85	99,03	98,90	86,74	89,34	94,49

(d) F-score des postures.

		Test et sélection d'attributs					
		Poche. D	Poche. G	Hanche	Poignet	Bras	Ensemble (P_i)
Train	Poche. D	99,43	100	90,26	99,06	83,43	-
	Poche. G	99,82	99,64	72,92	98,02	62,78	-
	Hanche	99,07	90,62	99,61	89,70	86,51	-
	Poignet	99,93	100	95,35	97,90	85,17	-
	Bras	100	96,93	89,80	93,47	81,89	-
	Ensemble (P_i)	99,89	98,37	98,64	96,64	79,31	94,19

TABLE 4.26 – Scores obtenus pour le jeu de données Multi-position.

Dans le test indépendant de la position, 8 attributs sont sélectionnés afin de séparer les actions en catégories, 64 attributs pour reconnaître les actions quasi-périodiques, et 5 attributs pour les postures.

(a) mHealth : Dimension des attributs pour la catégorie de l'action (F_C).

		Test					Ensemble (P_i)
		Poche. D	Poche. G	Hanche	Poignet	Bras	
Train	Poche. D	4	2	7	3	4	-
	Poche. G	2	6	5	5	3	-
	Hanche	1	2	5	2	5	-
	Poignet	1	1	3	8	3	-
	Bras	1	1	3	6	12	-
	Ensemble (P_i)	-	-	-	-	-	8

(b) mHealth : Dimension des attributs pour les actions quasi-périodiques (F_Q).

		Test					Ensemble (P_i)
		Poche. D	Poche. G	Hanche	Poignet	Bras	
Train	Poche. D	58	67	64	56	53	-
	Poche. G	39	76	68	63	60	-
	Hanche	36	35	74	33	49	-
	Poignet	50	34	36	41	47	-
	Bras	47	47	36	63	48	-
	Ensemble (P_i)	-	-	-	-	-	64

(c) mHealth : Dimension des attributs pour postures (F_P).

		Test					Ensemble (P_i)
		Poche. D	Poche. G	Hanche	Poignet	Bras	
Train	Poche. D	1	1	4	1	8	-
	Poche. G	1	3	1	1	8	-
	Hanche	1	2	12	3	6	-
	Poignet	1	1	3	1	5	-
	Bras	4	6	4	3	9	-
	Ensemble (P_i)	-	-	-	-	-	5

TABLE 4.27 – Dimension des attributs pour le jeu de données Multi-position.

Pour finir, les tests “positions croisées” montrent que moins de 8 attributs sont sélectionnés pour reconnaître la catégorie de l'action. Pour cinq croisements de positions, 1 attribut seulement est calculé. Dans les paires bras/poche gauche, bras/poche droite et poignet/poche gauche l'attribut “écart interquartile” du signal rotation angulaire suivant l'axe z (G_z) est sélectionné, cela a conduit à un score de 100%. Pour reconnaître les actions quasi-périodiques entre 36 et 68 attributs sont sélectionnés. Les postures quant à elles, nécessitent moins de 9 attributs. Pour la paire poignet/poche gauche par exemple, le premier quartile du signal A_z est sélectionné, cela a permis d'obtenir un score de 100%.

4.3.4 Temps de calcul

Avant de conclure, le temps de calcul de l’approche proposée est évalué sur le jeu de données SBHART. 8 signaux issus de capteurs accéléromètre et gyroscope sont traités dans cette expérience. La fréquence d’échantillonnage est de $50Hz$, et la taille de la fenêtre est de $T = 2,56s$ ce qui nous donne 128 échantillons à traiter. Trois catégories d’actions sont contenues dans ce jeu de données : périodique, postures et transitions posturales. Par conséquent notre approche est définie par le couple $(F_i, SVMs_i)$ avec $i \in \{C, Q, P\}$. Trois classifieurs sont alors construits et ils sont alimentés par les vecteurs d’attributs F_C, F_Q et F_P de dimension 2, 7 et 46 respectivement (tableau 4.18). À noter qu’aucun classifieur n’est construit pour reconnaître les transitions posturales. L’algorithme 4 est exécuté à cette fin.

Tous les tests sont effectués sur un ordinateur ordinaire doté d’un processeur Intel Core i5 à 2,60 GHz et de 4 Go de RAM. Le temps de calcul est évalué par la fonction intégrée tic-toc de MATLAB qui fournit une résolution de $1\mu s$ (KNAPP-CORDES et al. 2011a). L’expérience a été répétée 100 fois pour obtenir des résultats significatifs. Seule la phase de prédiction est concernée dans cette expérience puisque la phase d’apprentissage est réalisée en mode hors-ligne. Les résultats obtenus ont montré un temps de $2,4ms \pm 0,05$ pour chaque fenêtre.

À cette étape, toutes les expériences ont été menées sur un ordinateur personnel, car nous avons utilisé des bases de données publiques pour tester notre approche et pour pouvoir comparer nos résultats. La prochaine étape (en perspective) est l’implémentation de notre méthode sur smartphone.

4.4 Suivi d’actions sur le long-terme

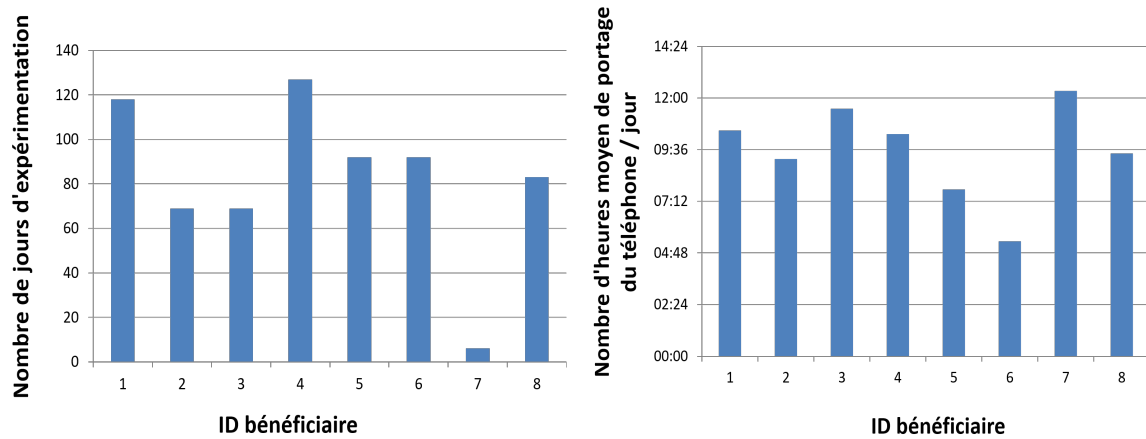
Dans la phase expérimentale de nos travaux, nous avons exploité des jeux de données contenant moins d’une dizaine de jours d’actions. Dans cette partie, nous allons exploiter un jeu de données acquises par Orange et contenant plusieurs mois d’activité. Cette base de données a été collectée à l’aide de Samsung Nexus S2 portés en bandoulière par 8 personnes (5 femmes et 3 hommes) (figure 4-13). 2 d’entre-elles

vivent seules et 6 en couple ou avec l'aide d'un proche. Ces personnes ont un âge moyen de 77 ans (entre 64 et 90), habitent dans six maisons et deux studios.



FIGURE 4-13 – Manière de porter le smartphone en Bandoulière.

La durée d'expérimentation est en moyenne de 92 jours avec une activité moyenne globale de 9 heures 32 minutes comme le montre la figure 4-14. Il est à remarquer que 4 personnes ont arrêté prématurément l'expérimentation : 3 pour des raisons de santé et 1 pour vol de portable. Seules 4 personnes sont arrivées au terme de l'expérimentation.



(a) Temps resté en expérimentation.

(b) Activité moyenne globale.

FIGURE 4-14 – Durée de l'expérimentation.

Dans cette expérience, nous allons considérer trois classes de postures qui sont “debout”, “assis”, “allongé” et deux classes dynamiques¹ : “la marche” et “autres actions”. Les autres actions représentent toutes les actions dynamiques détectées différentes de la marche. Enfin, une classe supplémentaire nommée “hors service” est considérée, elle définit le non-port du smartphone par l’individu (smartphone posé sur la table ou en charge par exemple).

Jours	Lever	Petit déjeuner	Sortie		Déjeuner		Sortie		Souper		Se coucher
		début	début	fin	début	fin	début	fin	début	fin	
Lun	07 :45	08 :15	10 :15	11 :30	12 :30	-	15 :15	17 :30	19 :50	-	22 :45
Mar	07 :40	08 :15	10 :15	11 :50	12 :15	-	13 :15	21 :00	21 :30	-	22 :30
Mer	07 :30	08 :00	-	-	12 :15	-	14 :15	17 :30	-	19 :45	22 :30
Jeu	08 :15	08 :45	-	-	13 :15	-	-	-	-	19 :45	23 :00
Ven	07 :45	08 :15	09 :30	-	12 :40	13 :45	14 :00	16 :00	-	19 :45	22 :30
Sam	07 :45	08 :15	09 :15	11 :45	12 :45	-	-	-	-	20 :00	23 :00
Dim	08 :00	08 :30	10 :00	11 :30	13 :00	-	16 :00	17 :00	-	19 :45	22 :45
Moy	07 :48	08 :19	09 :51	11 :38	12 :40	13 :45	14 :33	17 :48	-	20 :02	22 :42
Écart	00 :14	00 :14	00 :27	00 :10	00 :22	-	01 :04	01 :53	-	00 :38	00 :13

TABLE 4.28 – Heures déclaratif de réalisation de certaines activités de la vie quotidienne par le sujet 4.

Afin d’obtenir les informations relatives aux activités réalisées par les sujets, un entretien individuel sur les heures de pratiques des activités de base est effectué. Les résultats de l’entretien du sujet 4 sont présentés dans le tableau 4.28. Ils montrent les heures d’exécution de six activités de la vie quotidienne suivant les jours de la semaine. Les activités concernées sont : se lever, prise du petit déjeuner, sortie (se promener ou faire les courses), déjeuner, dîner, se coucher.

Le jeu de données d’Orange n’est pas étiqueté, un apprentissage non supervisé est donc nécessaire. Afin d’appliquer notre système sur ce jeu de données, nous avons proposé une variante de notre approche pour l’apprentissage non supervisé. L’aperçu de notre méthode est illustré dans la figure 4-15. La révision principale effectuée dans notre système réside dans le remplacement des algorithmes d’apprentissage su-

1. Toutes les actions qui impliquent un mouvement considérable du corps (actions non sédentaires).

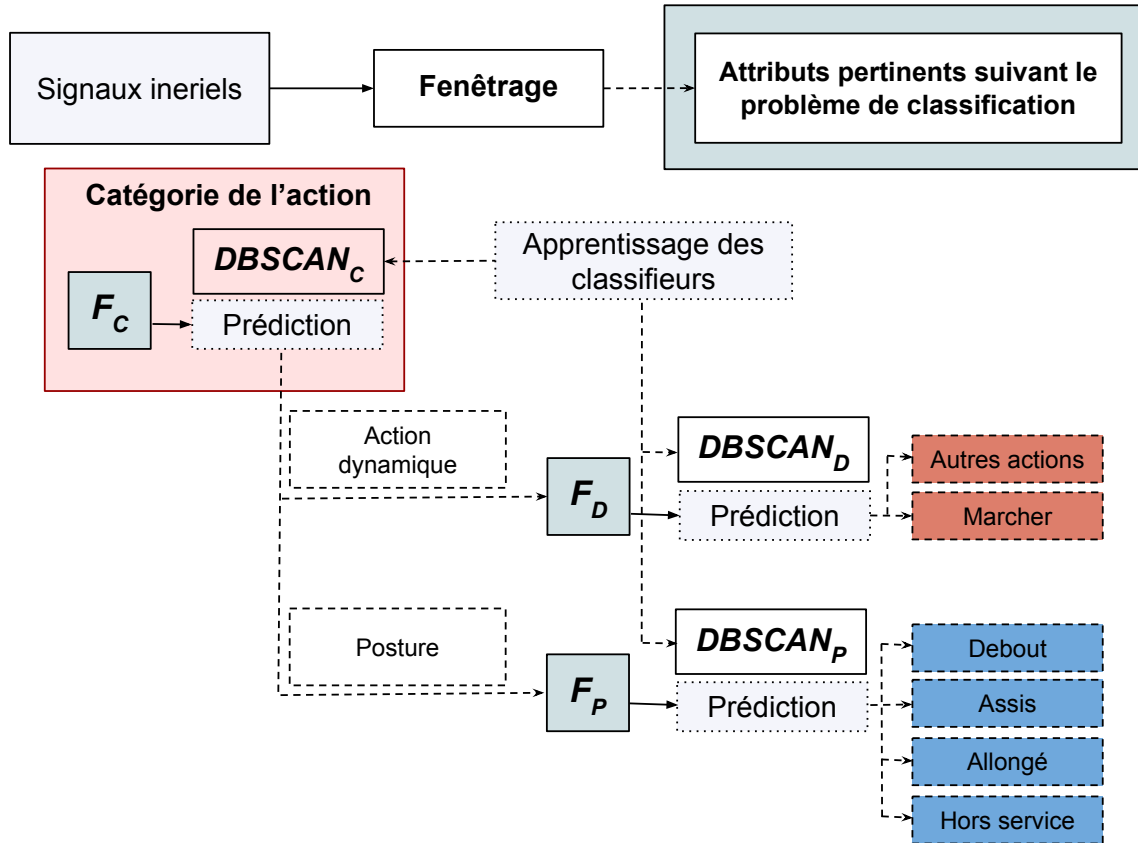


FIGURE 4-15 – Approche proposée pour la reconnaissance d’actions en mode non supervisé.

pervisé, dans notre cas les classifieurs SVMs par des méthodes d’apprentissage non supervisé. Nous avons pour cela exploité l’algorithme DBSCAN (ESTER et al. 1996). L’algorithme reçoit en entrée deux paramètres : la distance ϵ et le nombre minimum de points nommé *MinPts* se trouvant dans un rayon ϵ afin que ces points soient comptés comme un cluster.

Conformément à la figure 4-15, nous pouvons voir que dans le premier niveau de classification les actions sont d’abord regroupées en méta-classes : actions dynamiques et postures. Pour atteindre cet objectif, l’algorithme $DBSCAN_C$ est alimenté avec le vecteur F_C contenant des attributs discriminants pour ce problème. Même principe dans le deuxième niveau de classification de notre système : les classes des actions dynamiques sont reconnues par le couple $(F_D, DBSCAN_D)$ et les postures par le couple $(F_P, DBSCAN_P)$.

4.4.1 Résultats expérimentaux

Afin de montrer la pertinence de notre approche, des résultats d’expérimentations sur cette base de données sont présentés. Des données inertielles du sujet 4 ont été utilisées pour l’apprentissage (sujet qui est resté le plus longtemps dans l’expérimentation comme indiqué dans la figure 4-14). Cependant, seuls 10 jours de données collectées durant le mois de mai (jours sélectionnés aléatoirement) ont été utilisées pour entraîner les modèles, cette quantité de données est jugée suffisante pour cette expérimentation. Pour les données de test, nous avons utilisé les données du même sujet collectées durant tout le mois de juin.

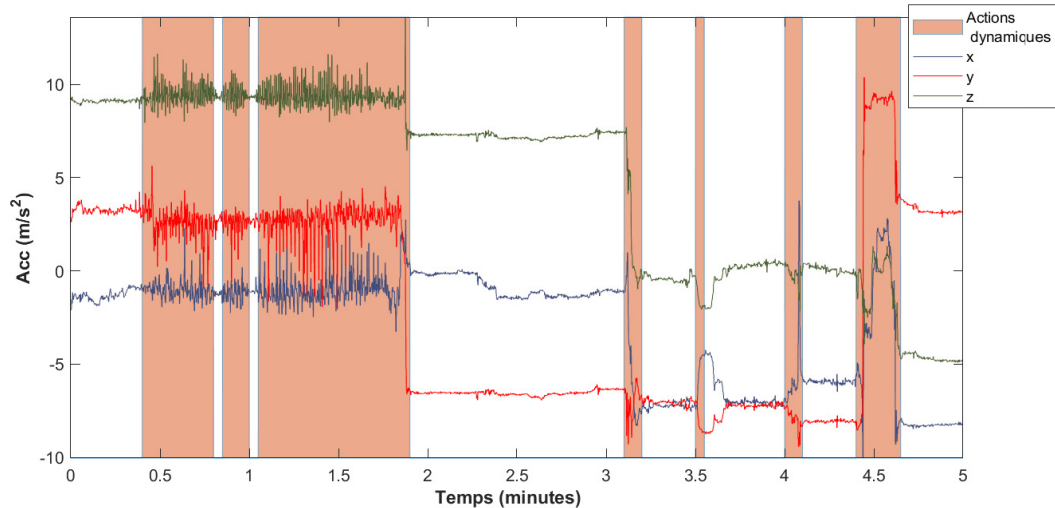


FIGURE 4-16 – Résultats obtenus au premier niveau de classification pour 5 minutes de données.

La figure 4-16 montre 5 minutes de données d’accélération, accompagnées des résultats de séparation obtenus entre les postures et les actions dynamiques (sortie de l’algorithme $DBSCAN_C$). la partie colorée de la figure correspond aux actions dynamiques, nous pouvons voir qu’elles sont parfaitement séparées des postures.

La figure 4-17 montre une heure de signaux inertiels (accélération, rotations angulaires, mesures de champ magnétique), ces données sont accompagnées du résultat de prédiction final de notre système. Nous pouvons voir que la classe “marcher” et la posture “debout” sont très présentes et proches l’une de l’autre temporellement. Ces

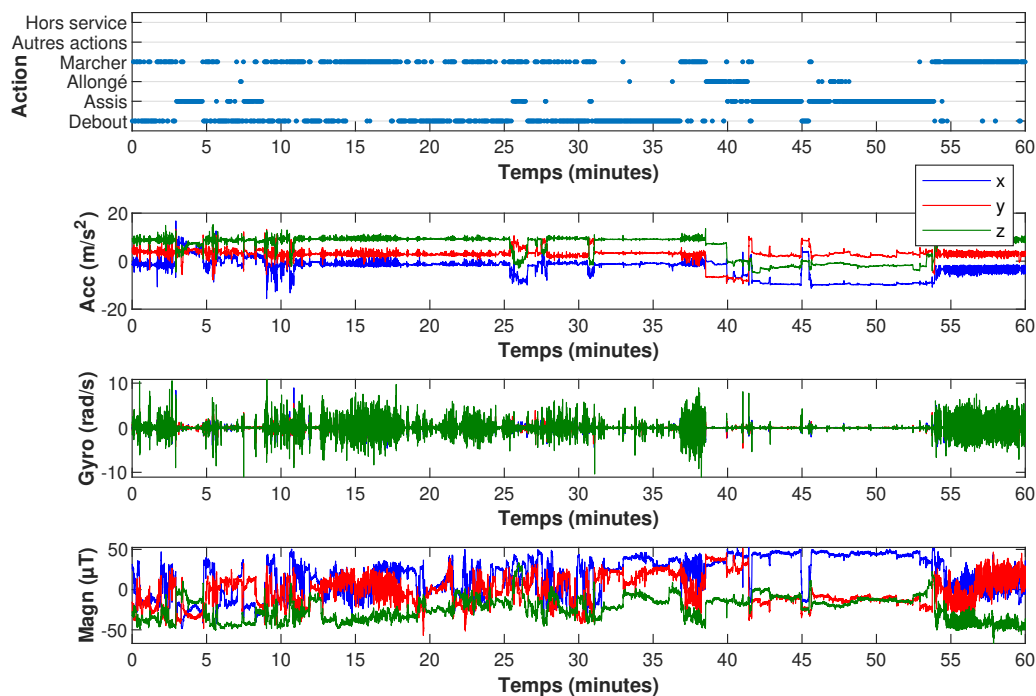


FIGURE 4-17 – Sortie du système pour une heure de données.

résultats semblent corrects, car dans un environnement intérieur, alterner les actions “marcher” et “debout” est un comportement commun, contrairement à un environnement extérieur où les actions ont plus tendance à durer (marcher pour atteindre une destination particulière, faire du vélo, etc.).

L’action “assis” est aussi très présente dans cette partie des résultats, principalement dans la tranche $[40min, 53min]$. Nous remarquons que cette posture s’étend sur une longue durée comparée à la posture “debout”, car les activités faisant appel à cette posture telles-que regarder la télévision, lecture, utiliser un ordinateur durent plus d’une dizaine de minutes et d’une façon continue. À noter aussi que les signaux issus du magnétomètre varient moins pour la posture “assis” que “debout” car les personnes ont plus tendance à s’orienter dans diverses directions quand elles sont debout qu’assis.

La figure 4-18 montre les résultats de reconnaissance obtenus durant une journée entière (journée choisie aléatoirement). Nous pouvons voir que la posture “allongé” est très présente entre 12 heures et 18 heures. Dans la tranche $[14h, 15h]$ en particulier,

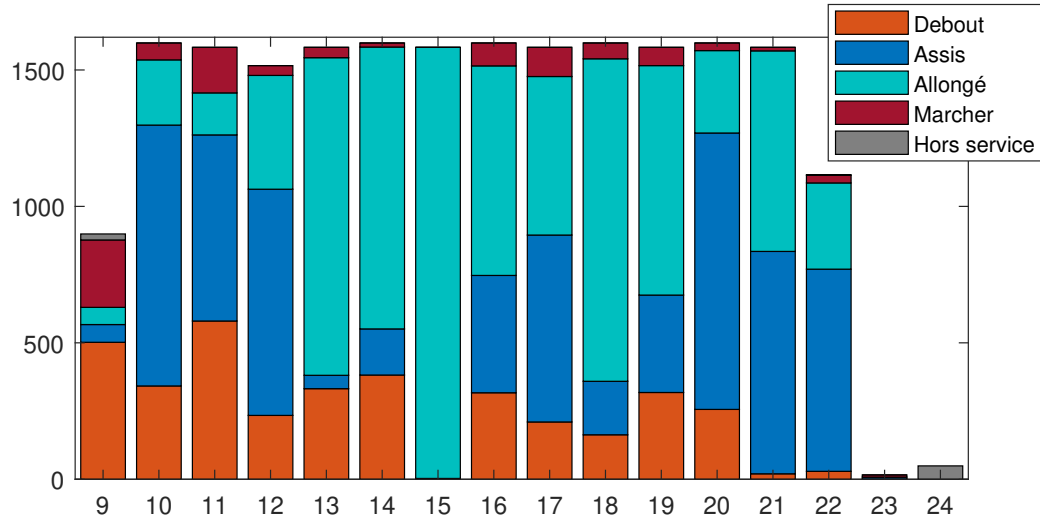


FIGURE 4-18 – Actions reconnues pour une journée de données.

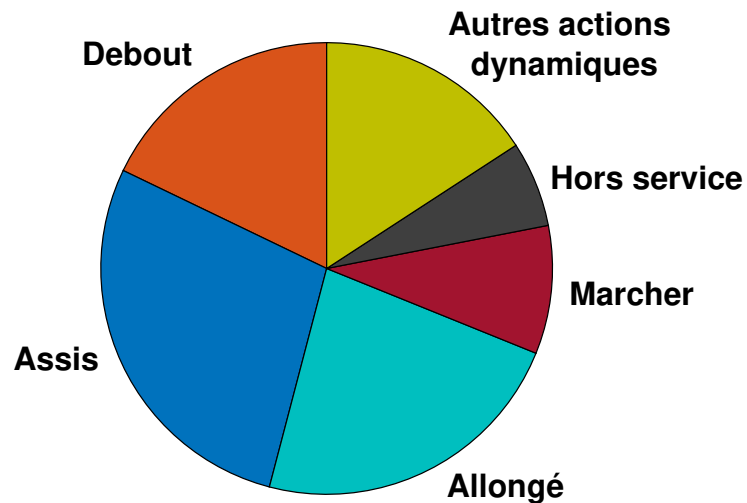


FIGURE 4-19 – Actions reconnues pour un mois de données.

la personne est allongée durant toute cette durée, probablement faisant une sieste. La posture “assis” est elle aussi très courante, principalement avant midi et le soir (après 19 heures), cela correspond peut-être à l’activité “regarder la télévision”. Nous remarquons aussi que l’action “debout” est reconnue pendant les heures de préparation des repas (petit-déjeuner, déjeuner et dîner). Et enfin pour “la marche”, elle est en

grande partie présente le matin (dans les tranches $[8h, 9h]$ et $[10h, 11h]$). À noter aussi que nous trouvons l'évènement que nous avons appelé "hors-service" au début et à la fin de la journée, correspondant peut être au moment où le téléphone est posé sur la table avant que le sujet le porte au réveil et au moment où il dépose le téléphone avant d'aller se coucher. Enfin pour finir, le diagramme circulaire dans la figure 4-19 présente les actions reconnues durant tout le mois de juin pour le sujet 4. Nous pouvons voir que les postures représentent presque trois quarts de l'activité de la personne comparées aux actions dynamiques. La posture "assis" est la plus répandue avec plus d'un quart des actions en totale, suivie par la posture "allongé" en dernier la posture "debout". Pour les actions dynamiques la proportion de l'action "marche" est faible comparée aux postures. Enfin d'autres types d'actions, considérés comme dynamiques sont aussi présents à une échelle égale que la posture "debout". Ces actions peuvent correspondre à différents mouvements effectués par la personne comme par exemple : faire le ménage, faire la vaisselle, se brosser les dents, etc.

4.5 Conclusion

Dans ce chapitre, nous avons présenté notre système pour reconnaître en temps réel les actions humaines à l'aide de capteurs inertiels (accéléromètre et gyroscope) intégrés dans un smartphone. L'approche que nous avons proposée permet de classer d'une manière hiérarchique les signaux collectés, i.e., les actions sont d'abord regroupées en méta-classes (postures ou actions périodiques par exemple). Le classifieur adapté entraîné sur les actions incluses dans la méta-classe est ensuite sélectionné afin de reconnaître l'action proprement dite. Dans notre approche, au lieu de calculer un vecteur d'attributs de taille importante (qui est généralement le cas dans les travaux de littérature), nous construisons un vecteur d'attributs pertinents et optimaux suivant le problème de classification rencontré, i.e., un algorithme de sélection d'attributs nommé SFFS est exécuté au niveau de chaque nœud interne de l'approche de classification. Cette méthode nous a permis d'obtenir des performances supérieures aux différents travaux existants dans la littérature. En outre, l'aspect robustesse à

l'emplacement des capteurs a été pris en compte dans notre démarche. Également, cette méthode nous a permis de réduire considérablement le temps de calcul nécessaire pour reconnaître l'action. En plus des actions communes avec les travaux de l'état de l'art, dans notre approche, nous avons aussi mis l'accent sur la reconnaissance des transitions posturales, considérées comme importantes pour l'autonomie et le bien-être.

Cette partie présentait une première série de travaux sur l'analyse des signaux inertiels. La partie suivante va s'attarder elle sur la reconnaissance d'actions avec des caméras de profondeur.

Troisième partie

Reconnaissance anticipée de l'action à l'aide de la Kinect

Chapitre 5

État de l'art

Contents

5.1	Introduction	148
5.2	Système HAR à l'aide de la Kinect	148
5.2.1	Approches basées sur les données profondeur	148
5.2.2	Approches basées sur les données squelette	152
5.2.3	Approches hybrides	155
5.2.4	Reconnaissance d'actions avec l'apprentissage profond	157
5.2.5	Reconnaissance anticipée de l'action	160
5.3	Conclusion	162

5.1 Introduction

La reconnaissance de l'action humaine (Human Action Recognition - HAR) est une problématique importante et populaire en recherche sur la vision par ordinateur. Cela est dû à son utilisation dans diverses applications telles que la vidéo-surveillance, la robotique ou le divertissement. Les premiers travaux se sont focalisés sur l'exploitation des séquences d'images *RGB*, en raison de l'omniprésence des caméras vidéo classiques. Récemment, des capteurs *RGB-D* à faible coût et disponibles dans le commerce, tels que la Kinect, ont contribué à la popularité des capteurs de profondeur. Les données (profondeur ou squelette) acquises en temps réel par ce genre de capteurs ont permis de proposer diverses méthodologies pour reconnaître les actions. Dans ce chapitre, nous allons donner un aperçu des différents systèmes HAR proposés dans la littérature.

5.2 Système HAR à l'aide de la Kinect

Les systèmes de reconnaissance des actions humaines utilisant la Kinect peuvent être divisés en trois catégories suivant les types de données exploitées : approches basées sur la profondeur, sur les données squelette ou approches hybrides. Ces trois catégories d'approches sont abordées plus en détails ci-dessous.

5.2.1 Approches basées sur les données profondeur

Les données profondeur représentent la distance euclidienne entre le capteur *RGB-D* (Kinect dans notre cas) et les points de la scène. Ce type d'information rend la séparation du nuage de points relatif au corps humain et l'arrière-plan plus accessible comparé aux données *RGB*.

Une fois que les points 3D relatifs au corps humain sont définis, des attributs peuvent être calculés à partir de ces données pour discriminer les actions humaines. Parmi les approches de calcul proposées dans la littérature, le passage préalable par une projection des points 3D sur trois plans 2D orthogonaux : vue de face, de côté

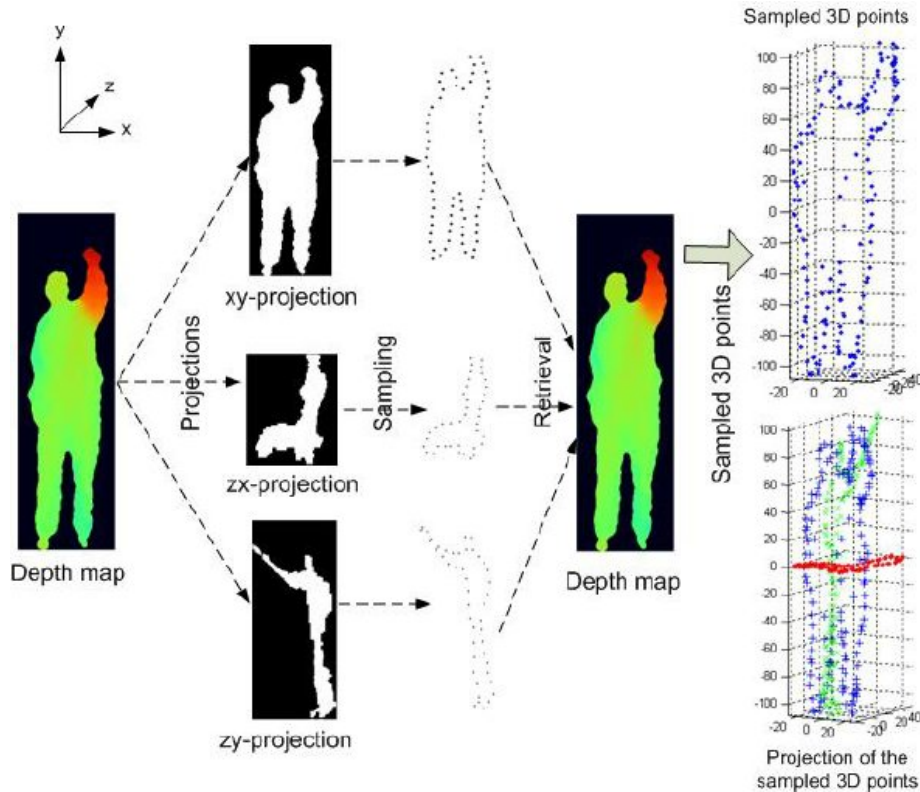


FIGURE 5-1 – Le calcul des attributs dans (W. LI et al. 2010) à partir de des images de profondeur.

et de dessus. Dans (W. LI et al. 2010), le contour 2D est calculé et ensuite sous-échantillonné pour chaque projection. Les points 3D avoisinants le résultat obtenu sont alors récupérés comme le montre la figure 5-1. Dans (X. YANG, C. ZHANG et al. 2012), après la phase de projection du nuage de points 3D, la différence entre deux images de profondeur adjacentes est calculée pour chaque plan, suivie par un seuillage. Une image binaire indiquant les régions de mouvement pour chaque intervalle de temps est ainsi obtenue. Cette représentation des données est nommée “Depth Motion Maps” (DMM). Ce processus de calcul est ensuite suivi par l’extraction des histogrammes de gradients orientés (HOG), ce dernier est très utilisé en vision par ordinateur pour la détection d’objet ou de personnes (DALAL et al. 2005). Les résultats obtenus pour chaque plan sont ainsi concaténés pour enfin obtenir le vecteur d’attributs *DMM-HOG*. Le schéma de calcul complet est présenté dans la figure 5-2. Plus tard dans les travaux de C. CHEN et al. (2016), les données DMM sont également

calculées, à une différence près, le seuillage n'est pas appliqué, et les résultats obtenus pour chaque plan sont directement concaténés sans l'extraction des données HOG.

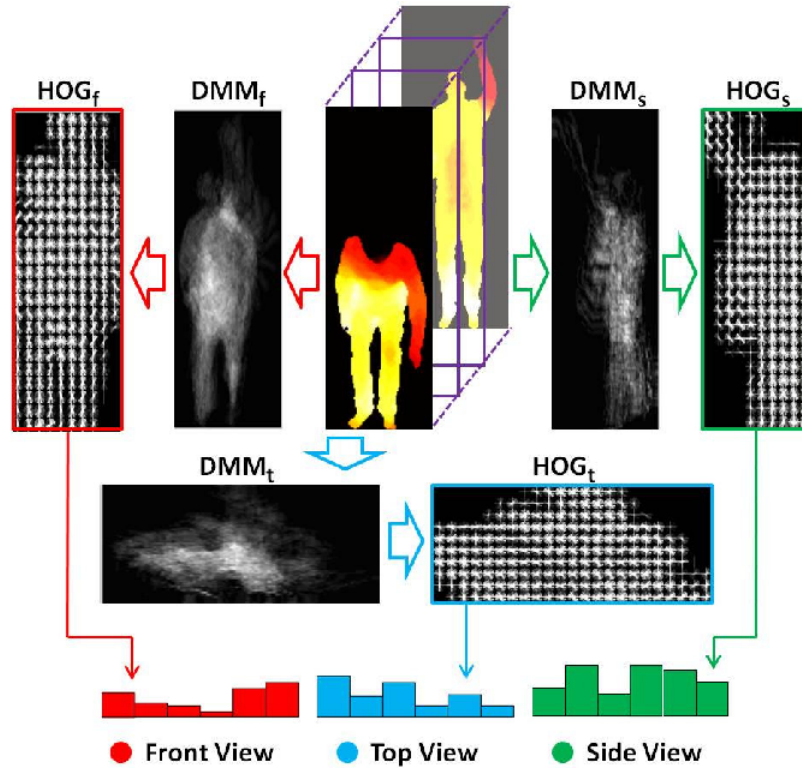


FIGURE 5-2 – Le calcul des attributs dans (X. YANG, C. ZHANG et al. 2012) à partir des images de profondeur.

D'autres chercheurs se sont inspirés des approches utilisées dans la reconnaissance d'actions à partir des images 2D pour les étendre aux données profondeur. La plus populaire est la segmentation des données en régions d'intérêts sur lesquelles des attributs pertinents sont ensuite calculés. Dans cette logique, des points d'intérêt spatio-temporel (Spatio-Temporal Interest Points - STIPs) sont définis pour les approches basées sur des images 2D (DOLLÁR et al. 2005 ; LAPTEV 2005). Pour les données profondeur, le terme DSTIPs est utilisé dans la littérature. Dans les travaux de XIA et J. AGGARWAL (2013), l'information DSTIPs est extraite en employant des méthodes de filtrage de données profondeur. Un exemple des résultats obtenus sur deux actions "boire en étant assis" et "boire en étant debout" est illustré dans la figure 5-3. Afin de construire le vecteur d'attributs, des cubes 3D sont centrés autour des données DSTIPs. Chaque cube est divisé en blocs à partir desquels des histogrammes

sont calculés. Des mesures de similarité sont ensuite calculées en réalisant la combinaison de tous les blocs. La concaténation du résultat forme le vecteur d’attributs. Dans (M. LIU et al. 2016), l’information DSTIPs est calculée en s’inspirant des travaux de BELONGIE et al. (2002) concernant la représentation de la forme des images 2D. Les points riches en informations de mouvement et de forme sont alors extraits des images profondeurs. Un nouveau groupe d’attributs nommé “contexte de profondeur” est proposé afin d’encoder l’information DSTIPs. La principale limitation de ces approches est l’exigence d’un temps de calcul supplémentaire pour la détection des régions d’intérêts.

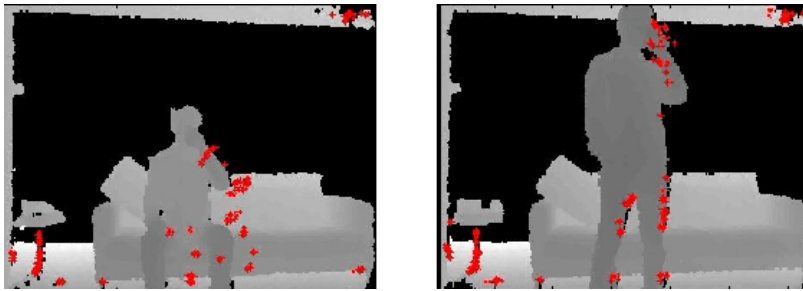


FIGURE 5-3 – Le calcul de l’information DSTIPs dans (XIA et J. AGGARWAL 2013) à partir des images de profondeur.

Une autre représentation qui a montré son efficacité dans la littérature est le calcul des vecteurs normaux à la surface à partir du nuage de points 3D. Cette idée a été introduite par OREIFEJ et al. (2013) : les vecteurs normaux à la surface sont calculés sur une dimension \mathbb{R}^4 (\mathbb{R}^3 pour l’espace et \mathbb{R} pour le temps). Le résultat est ensuite quantifié en utilisant des objets appelés polychorons (extensions 4D d’un polygone 2D). Ces étapes de calcul sont représentées dans la figure 5-4. La même représentation est utilisée dans (X. YANG et TIAN 2014). Des vecteurs normaux à une hypersurface sont d’abord regroupés pour obtenir les “polynormaux”. Un codage parcimonieux est ensuite appliqué afin d’apprendre un dictionnaire et de coder les polynormaux. Une agrégation des polynormaux de bas niveau est adoptée pour obtenir des attributs nommés “super normal vector”.

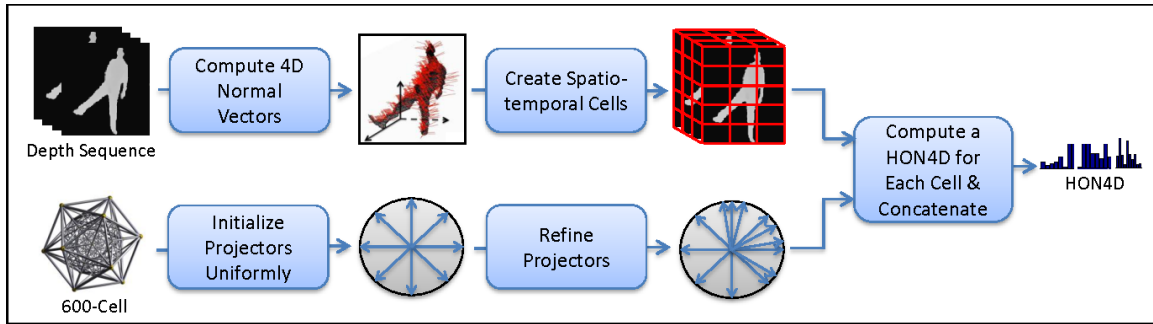


FIGURE 5-4 – Le calcul des attributs dans (OREIFEJ et al. 2013) à partir des images de profondeur.

5.2.2 Approches basées sur les données squelette

Les approches basées sur la profondeur doivent traiter une quantité importante de données impliquant un calcul coûteux dans la phase d'extraction des attributs. Par conséquent, des chercheurs se sont tournés vers d'autres types d'approches dont l'approche basée sur les données squelette. Cette dernière est devenue populaire grâce aux travaux de SHOTTON et al. (2013). Son algorithme permet d'extraire en temps réel des données squelette à partir des images profondeur. Ces données correspondent à l'estimation des positions 3D des articulations du corps. Selon la version de la Kinect et du kit de développement utilisé, un nombre d'articulations différent est estimé comme le montre la figure 5-5. La librairie OpenNI fournit 15 positions d'articulation, la Microsoft Kinect SDK version 1 fournit 20 articulations, et pour la version 2, elle est capable d'estimer 25 articulations.

En se basant sur ces données de faible dimension (comparées aux images de profondeur), plusieurs approches récentes ont montré leur capacité à reconnaître des actions. Suivant les approches utilisées dans la littérature pour l'extraction des attributs, nous proposons de diviser les travaux en deux catégories : méthodes basées sur les postures clés, méthodes basées sur la trajectoire.

Méthodes basées sur les postures clés

Dans cette catégorie d'approches, une action est représentée par une séquence de postures clés. Dans les travaux de XIA, C.-C. CHEN et al. (2012), un histogramme des

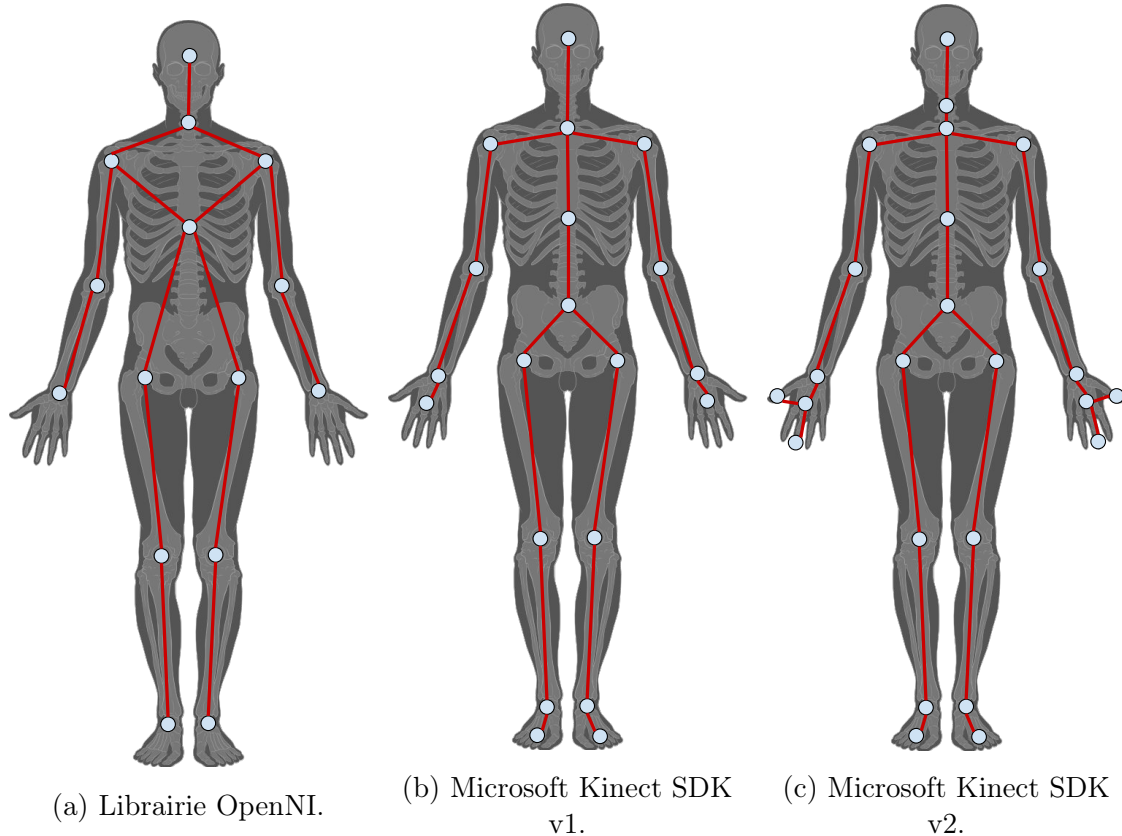


FIGURE 5-5 – Les données squelette estimées selon la version de la Kinect et du kit de développement utilisé.

emplacements des articulations sur un système de coordonnées sphériques a été proposé. Une analyse discriminante linéaire est ensuite appliquée suivie par regroupement des données obtenues en clusters à l'aide de l'algorithme k -means. Un vocabulaire de postures est ainsi obtenu. L'évolution temporelle de ces dernières a été modélisée à l'aide de modèles de Markov cachés discrets (D-HMMs). Dans (CHAARAOUI et al. 2014), un sous-ensemble optimal des articulations du squelette est d'abord sélectionné par un algorithme évolutif proposé. Afin d'obtenir les postures clés pour chaque action, un regroupement des données sélectionnées en clusters est appliqué en utilisant l'algorithme k -means. Les centres des clusters sont concaténés pour former le vecteur d'attributs final. Une classification basée sur la déformation temporelle dynamique (DTW) est ensuite appliquée. Dans les travaux de VEMULAPALLI et al. (2014), chaque squelette est représenté dans un groupe de Lie. La séquence de squelettes constituant l'action est modélisée par une courbe comme le montre la figure 5-6. Une classification

est ensuite effectuée en utilisant une combinaison de DTW, Fourier temporel pyramid (FTP) et SVM linéaire. Dans (CIPPITELLI et al. 2016), l'extraction d'attributs débute en normalisant les données squelette par la distance euclidienne entre l'articulation du torse et du cou. L'algorithme k -means a été appliqué par la suite pour regrouper des postures similaires dans des clusters. Enfin, un classifieur SVM a été appliqué pour identifier les actions humaines.

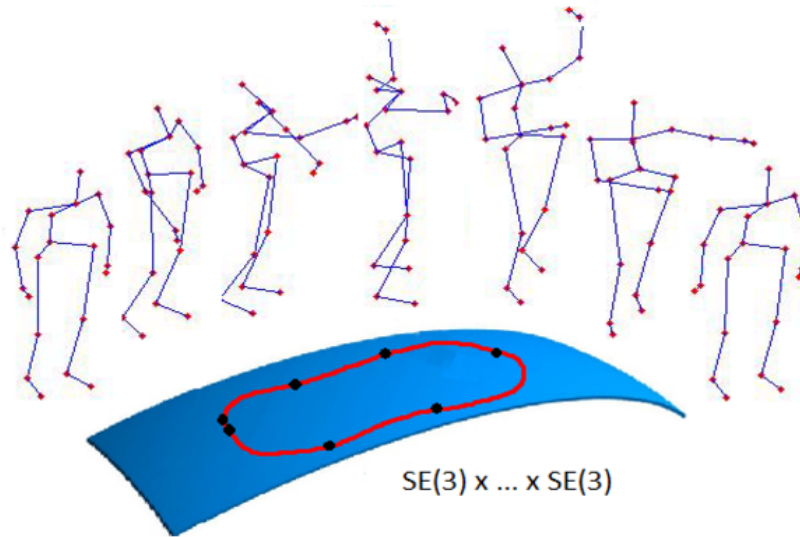


FIGURE 5-6 – Représentation d’une action (séquence de squelettes) sous forme de courbe dans le groupe de Lie.

Méthodes basées sur la trajectoire

Dans cette catégorie d’approches, les attributs sont extraits des squelettes en exploitant les trajectoires des différentes articulations. DEVANNE et al. (2015) représentent les trajectoires des articulations du squelette dans la variété riemannienne. La classification a été effectuée à l’aide du classifieur les k -plus proches voisins (k Nearest Neighbors - k NN) en utilisant une métrique d’élasticité dans le calcul de la distance. La figure 5-7 résume l’approche proposée. Dans les travaux de GOWAYYED et al. (2013), les trajectoires des articulations sont projetées sur trois plans 2D orthogonaux (xy, yz, xz) . FTP est ensuite construite pour chaque trajectoire 2D afin de capturer l’information temporelle. Enfin, un nouveau descripteur est proposé, nommé “Histogram of Oriented Displacements”, il repose sur le calcul d’un histogramme des

directions entre deux points consécutifs. Z. SHAO et al. (2015) proposent de nouveaux invariants intégraux pour les trajectoires de mouvement spatial. Cette dernière à un effet de lissage et sont donc moins sensibles au bruit (pas besoin de pré-traiter la trajectoire du mouvement). Pour la classification, les méthodes kNN et DTW sont exploitées.

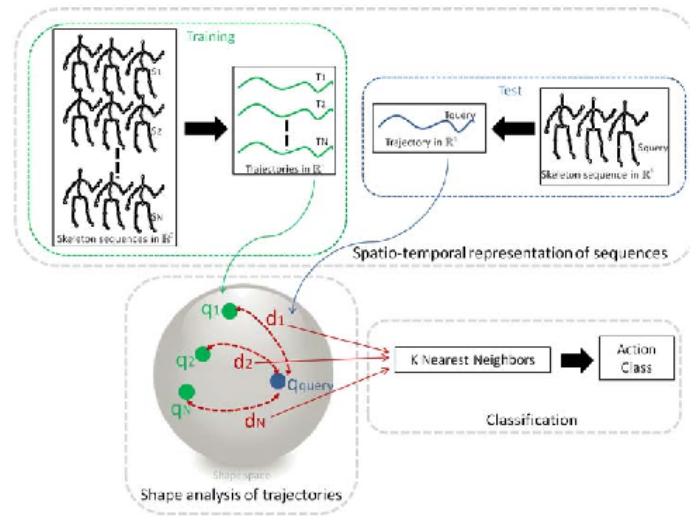


FIGURE 5-7 – Aperçu de l’approche proposée par DEVANNE et al. (2015).

5.2.3 Approches hybrides

Au lieu d’utiliser un seul type de données, les approches hybrides considèrent la combinaison d’attributs calculés à partir de données de nature différentes. Ici on parle de données profondeur, squelette et images *RGB*. Dans (L. LIU et L. SHAO 2013), les attributs sont extraits à partir des images *RGB* et des données de profondeur. Ces derniers sont d’abord mis à l’entrée d’un système de filtrage contenant plusieurs filtres numériques (filtres gaussiens, laplacien, ondelettes, etc.). Des techniques d’apprentissage profond (“max pooling” et “average pooling” en anglais) sont ensuite utilisées pour la phase d’extraction des attributs. La même combinaison de données (profondeur, images *RGB*) est utilisée dans (Y. ZHAO et al. 2012), les algorithmes de calcul d’attributs HOG et Histogramme de flux (Histogram of Flow - HOF) (LAPTEV et al. 2008) sont appliqués sur les données profondeur et images *RGB*. Un algorithme supplémentaire nommé “Local Depth Pattern” est proposé par les auteurs, ce dernier est

appliqué uniquement sur les données profondeur, il est défini par la différence des valeurs de profondeur moyenne sur des régions d'intérêts. Les résultats expérimentaux ont montré que la combinaison des attributs profondeur et *RGB* donne de meilleurs résultats que les attributs extraits d'une seule source.

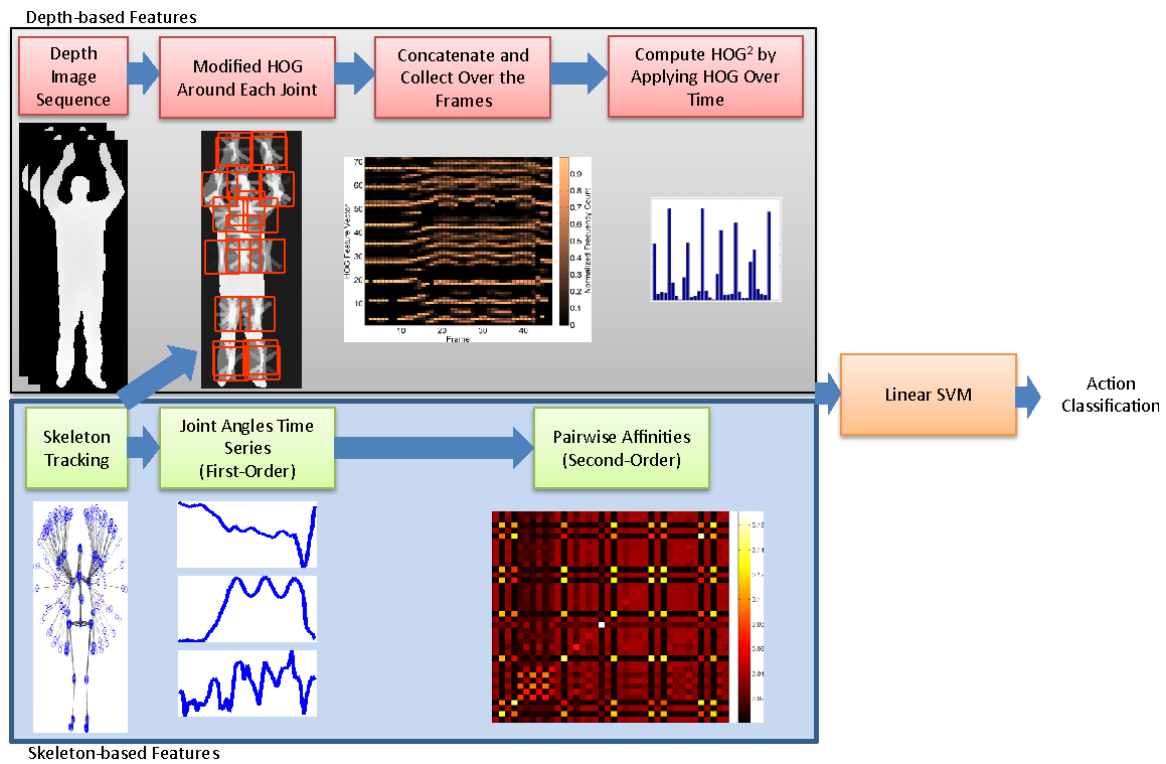


FIGURE 5-8 – Aperçu de l’approche proposée par OHN-BAR et al. (2013).

Dans (OHN-BAR et al. 2013), les attributs sont calculés à la fois des données profondeur et squelette comme indiqué dans la figure 5-8. Concernant les données profondeur, une version modifiée de la méthode HOG nommée HOG^2 est proposée pour extraire les attributs. L’approche est appelée ainsi car l’algorithme HOG est appliqué deux fois (une fois dans le domaine spatial puis à nouveau dans le domaine temporel). Pour les données squelette, des attributs nommés “Joint Angles Similarities” sont calculés à partir des distances de similarités entre les angles des articulations. Pour l’étape de classification, les deux vecteurs d’attributs sont concaténés et mis à l’entrée d’un classifieur SVM à noyau linéaire.

5.2.4 Reconnaissance d’actions avec l’apprentissage profond

Dans ce qui précède, l’état de l’art a été divisé suivant les différentes approches utilisées pour l’extraction des attributs. Cela est dû au fait que dans les travaux existants, la contribution des chercheurs est plus accentuée sur la construction du vecteur d’attributs. Plus récemment, les chercheurs se sont focalisés sur l’apprentissage des classifieurs, les techniques dites d’apprentissage profond sont de plus en plus populaires. Dans le domaine de la reconnaissance d’actions à l’aide des images *RGB*, les méthodes d’apprentissage profond ont entièrement remplacé les méthodes traditionnelles (dites “handcrafted” en anglais). Cependant, il y a beaucoup moins de travaux sur la reconnaissance d’actions à l’aide des données de profondeur utilisant les techniques d’apprentissage profond. La première raison est l’absence de couleur et de texture dans les données profondeur, cela peut atténuer le pouvoir discriminant des réseaux de neurones à convolution (Convolutional Neural Network - CNN). La seconde raison est que les jeux de données contenant des images de profondeur sont limités en taille comparés aux jeux de données avec des images *RGB*.

Apprentissage profond avec les données profondeur

L’apprentissage profond d’un modèle à partir des données profondeur brutes est rarement employé dans la littérature pour les causes citées ci-dessus. Cependant, quelques chercheurs appliquent au préalable des transformations sur les données comme le calcul de DMM (citée dans la partie 5.2.1) afin d’obtenir une représentation texturée des données. C’est le cas dans (P. WANG, W. LI, GAO, TANG, J. ZHANG et al. 2015), les données DMM sont d’abord calculées sur les images de profondeur. Les résultats obtenus pour chaque plan sont mis aux entrées de trois modèles CNN d’une manière indépendante. Rui YANG et al. (2014) propose une extension des données DMM appelée DMM-Pyramid : l’action est d’abord segmentée récursivement en plusieurs parties le long de l’axe du temps, le calcul DMM est ensuite appliqué sur chaque partie. Pour la phase d’apprentissage, un modèle CNN est entraîné sur les données calculées. P. WANG, W. LI, GAO, TANG et OGUNBONA (2018) a pro-

posé d'utiliser trois représentations des données profondeur nommée "Dynamic Depth Images", "Dynamic Depth Normal Images" et "Dynamic Depth Motion Normal Images". Ces représentations sont construites à l'aide d'une technique de "pooling" dites "hierarchical bidirectional rank pooling". Un modèle CNN est ensuite utilisé pour la reconnaissance d'actions.

Apprentissage profond avec les données squelette

Les approches d'apprentissage profond utilisant les données squelette peuvent être divisées en deux catégories suivant le modèle utilisé : approches basées sur les modèles CNN et approches basées sur les modèles réseaux de neurones récurrents (Recurrent neural network - RNN).

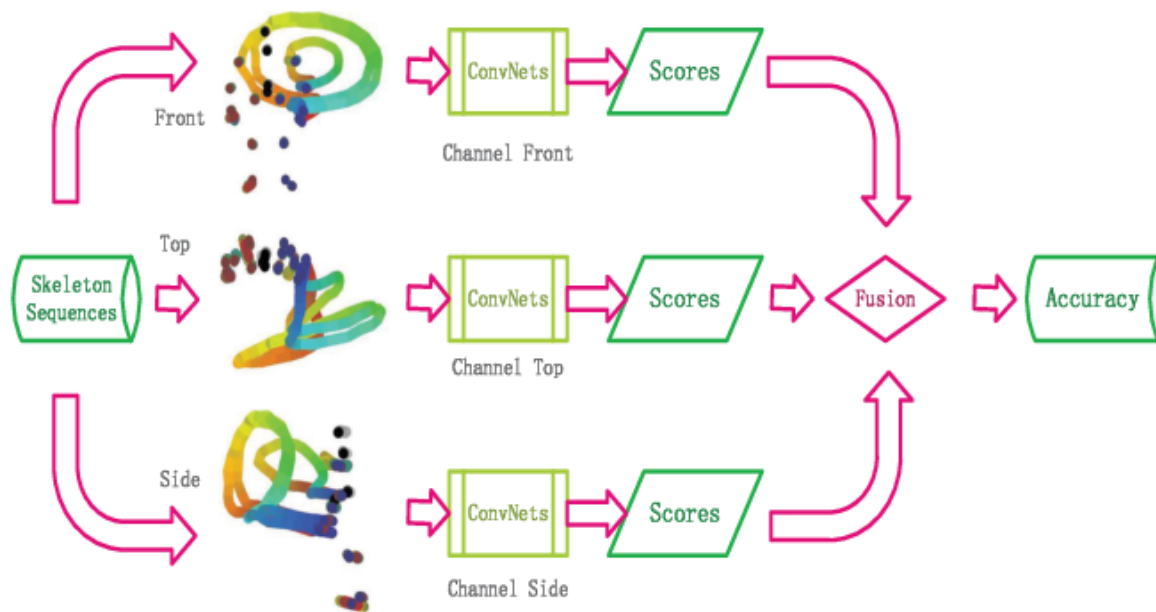


FIGURE 5-9 – Aperçu de l'approche proposée par DU, FU et al. (2015).

Dans les approches basées sur les modèles CNN, l'information spatio-temporelle des données squelette est d'abord transformée en images 2D. Dans cette démarche, DU, FU et al. (2015) proposent de transformer les données squelette en trois images de texture contenant les trajectoires des articulations comme le montre la figure 5-9. Les images obtenues sont encodées en teinte, saturation et luminosité (système de description des couleurs). Enfin, trois modèles CNN sont utilisés pour la classification,

et les scores en sorties de ces trois modèles sont fusionnés afin d’obtenir un score global. Une approche similaire est adoptée par C. LI et al. (2017) où les distances entre les articulations sont codées en images de texture nommées “Joint Distance Maps”. Le résultat est mis aux entrées des modèles CNN pour la phase d’apprentissage.

Concernant les approches basées sur les modèles RNN, les données squelette sont directement mises aux entrées des modèles. Dans (DU, W. WANG et al. 2015), un modèle RNN hiérarchique est proposé pour la reconnaissance d’actions. Le squelette a été d’abord divisé en cinq groupes d’articulations représentant deux bras, deux jambes et un tronc. Chaque groupe a été utilisé pour nourrir cinq RNN bidirectionnels comme le montre la figure 5-10. À mesure que le nombre de couches augmente, les représentations extraites par les sous-réseaux sont fusionnées de manière hiérarchique pour constituer les entrées des couches supérieures. Les représentations finales des données squelette sont mises à l’entrée d’un perceptron monocouche, et la sortie accumulée du perceptron constitue la décision finale.

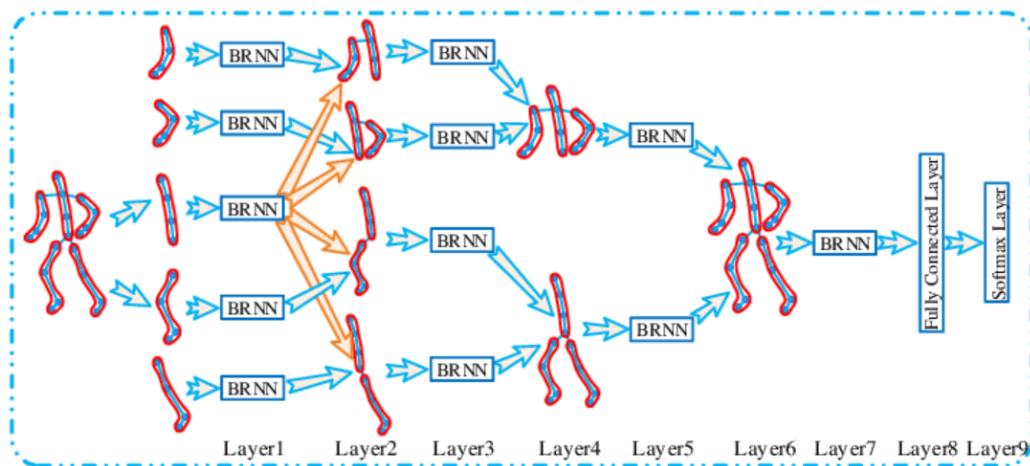


FIGURE 5-10 – Aperçu de l’approche proposée par DU, W. WANG et al. (2015).

W. ZHU et al. (2016) proposent d’utiliser un modèle “Long short-term memory” (LSTM) pour reconnaître les actions. Une technique de régularisation a été introduite afin de permettre au modèle d’apprendre les attributs de co-occurrence des articulations au niveau des différentes couches. De plus, une deuxième nouvelle technique de régularisation (“dropout” en anglais) a été proposée afin d’aider le réseau à apprendre des dynamiques complexes du mouvement.

5.2.5 Reconnaissance anticipée de l'action

La reconnaissance anticipée vise à reconnaître l'action avant la fin de l'exécution de celle-ci. Quelques chercheurs ont proposé des approches adaptées pour la reconnaissance anticipée de l'action. Dans (RYOO 2011), la reconnaissance anticipée de six classes d'actions à partir d'une séquence vidéo (données *RGB*) est étudiée. Deux méthodes basées sur la représentation par sac de mots (bag of words en anglais) ont été introduites. La première est nommée "Integral bag-of-words", où des attributs locaux spatio-temporels sont calculés comme suit : pour chaque image, des régions d'intérêts basées sur le mouvement sont d'abord identifiées. Des gradients sont ensuite calculés sur ces régions suivis par un regroupement des données résultantes en clusters à l'aide de l'algorithme *k*-means. Les attributs obtenus sont appelés mots visuels (visual words en anglais). Afin de reconnaître l'action d'une manière anticipée, les attributs calculés sont représentés sous forme d'histogrammes (figure 5-11), ces derniers sont comparés avec les histogrammes du modèle au fur et à mesure que l'action évolue. Cette approche proposée ne prend pas en considération la relation temporelle entre les attributs calculés. Par conséquent, une deuxième méthode est proposée, appelée "dynamic bag-of-words", ici, le modèle de l'action et la séquence observée sont divisés en segments de taille Δd définie d'une manière dynamique, la similarité entre les seg-

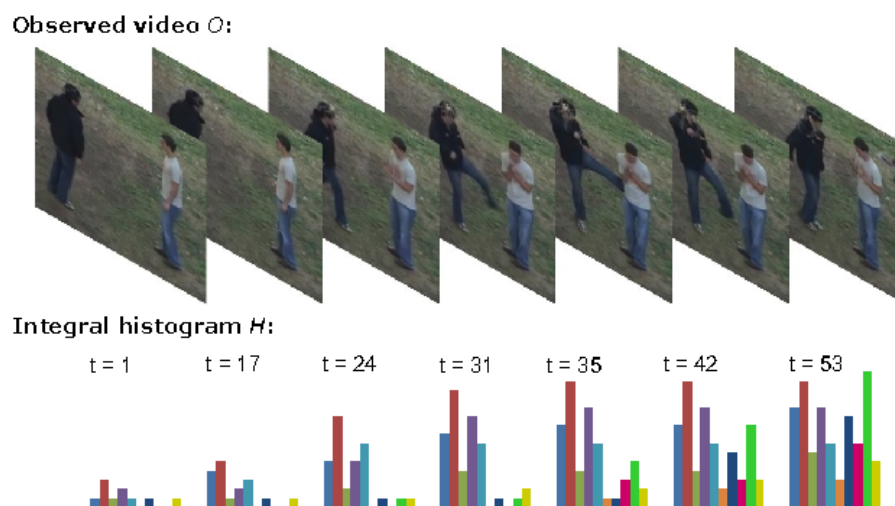
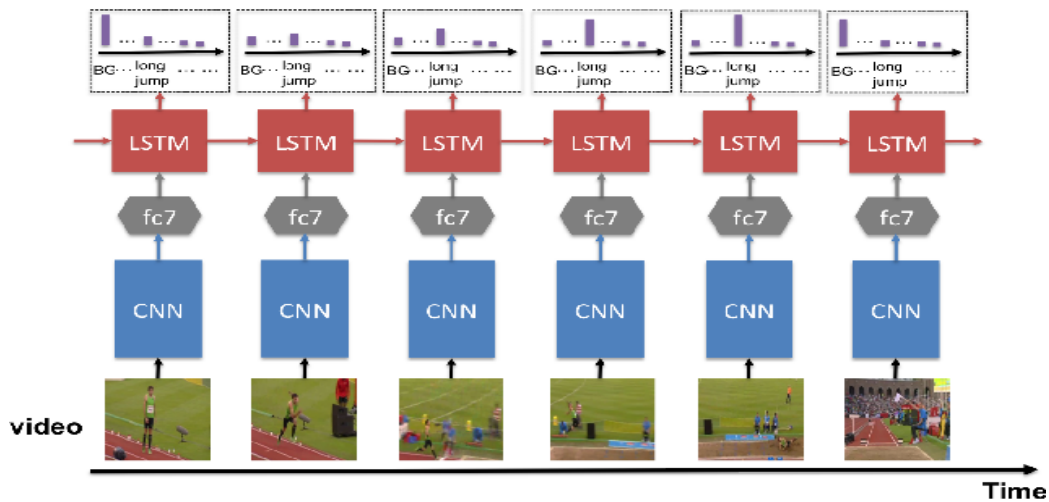
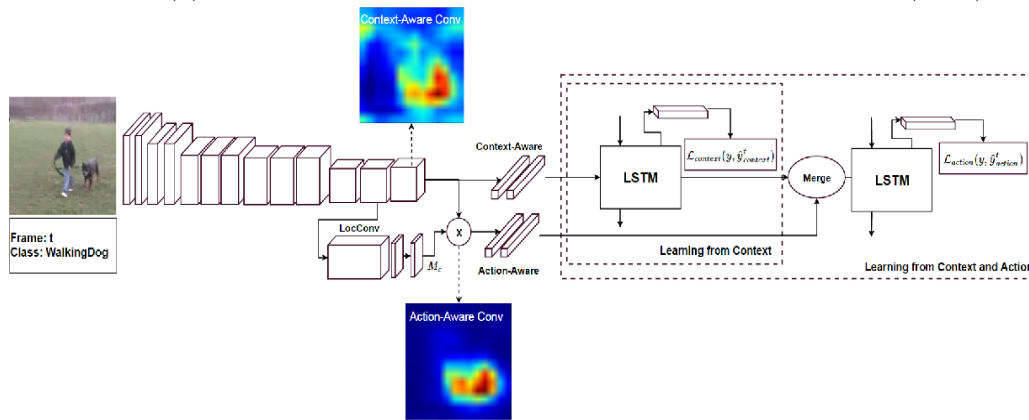


FIGURE 5-11 – Aperçu de l'approche Integral bag-of-words proposée par RYOO (2011).

ments du modèle et ceux observés est ensuite calculée. Les résultats expérimentaux sur un jeu de données publique ont montré qu'en exploitant la séquence entière de l'action, des scores de précisions de 81,7% et 85,0% sont atteints respectivement pour la première et deuxième méthode. En exploitant la moitié de la séquence seulement, ces scores chutent à 65,0% et 70,0%.



(a) Aperçu de l'approche générale proposée par MA et al. (2016).



(b) Aperçu de l'approche générale proposée par SADEGH ALIAKBARIAN et al. (2017).

FIGURE 5-12 – Reconnaissance anticipée de l'action en utilisant des approches basées sur l'apprentissage profond.

Les récents systèmes qui traitent le problème d'anticipation d'actions utilisent des approches basées sur l'apprentissage profond. Dans cette optique, MA et al. (2016) proposent de combiner deux modèles d'apprentissage profond. Dans un premier temps, un réseau de neurones à convolution (CNN) est appliqué afin de calculer les attributs, ces derniers sont ensuite mis à l'entrée d'un modèle LSTM (figure 5-12a),

l'objectif est de modéliser à la fois l'information temporelle et spatiale des actions. En outre, les auteurs estiment que l'évolution du score de reconnaissance de l'action en cours d'exécution, ou l'évolution de la marge entre ce score et les scores des autres actions doit être croissante. Pour atteindre cet objectif, le modèle est pénalisé dans le cas du non-respect de cette condition. Une approche un peu similaire est utilisée par SADEGH ALIAKBARIAN et al. (2017) où un modèle CNN est exploité afin d'extraire à la fois des attributs contextuels et des attributs pour caractériser l'action (figure 5-12b). Les attributs contextuels extraits sont ensuite mis à l'entrée d'un modèle LSTM afin de calculer un score pour chaque action considérée, cette sortie est ensuite combinée avec les attributs d'actions afin d'alimenter un deuxième modèle LSTM.

5.3 Conclusion

Dans ce chapitre, nous avons donné un aperçu de l'état de l'art sur les systèmes HAR à base de capteurs *RGB-D*. Suivant la nature des données exploitées, les approches proposées ont été divisées en trois catégories : approches basées sur les données profondeur, approches basées sur les données squelette, et approches hybrides. En outre, les travaux exploitants les techniques d'apprentissage profond pour la reconnaissance d'actions sont aussi rapportés comme une catégorie à part. Pour finir, nous avons aussi donné un aperçu des techniques de reconnaissance anticipée des actions.

Chapitre 6

Reconnaissance anticipée des actions humaines

Contents

6.1	Introduction	164
6.2	Modèles de Markov cachés et les attributs Body-Part	
	Directional Velocity	164
6.2.1	Body-part Directional Velocity (BDV)	165
6.2.2	Classification à l'aide d'un modèle GMM-HMM	167
6.3	Validation expérimentale	169
6.3.1	Jeux de données	170
6.3.2	Protocole expérimental	172
6.3.3	Expérimentation sur des jeux de données publiques	175
6.3.4	Reconnaissance anticipée de l'action humaine	178
6.3.5	Temps de calcul	182
6.4	Conclusion	183

6.1 Introduction

Dans ce chapitre, nous présentons une nouvelle approche pour la reconnaissance anticipée des actions humaines utilisant des données squelette extraites à partir des données profondeurs. Nous proposons une nouvelle méthode afin de construire image par image et en temps réel le vecteur d'attributs. Nommé Body-Part Directional Velocity (BDV), ce vecteur est calculé en considérant la vitesse algébrique produite par différentes parties du corps. Les attributs sont ensuite mis en temps réel à l'entrée des modèles de Markov cachés HMMs avec des distributions de sortie d'états représentées par des modèles de mélange gaussiens (GMMs). Les résultats expérimentaux sur quatre jeux de données montrent que notre méthode surpasse plusieurs approches de la littérature. Nous montrons également la pertinence de notre méthode pour la reconnaissance anticipée des actions humaines.

6.2 Modèles de Markov cachés et les attributs Body-Part Directional Velocity

Dans cette partie, la méthode proposée est décrite en détail dans la figure 6-1. Nous pouvons distinguer deux blocs principaux : un apprentissage hors ligne et une reconnaissance en temps réel. Dans la partie apprentissage, l'extraction des données squelette à partir de l'image de profondeur est effectuée grâce à l'algorithme de Shotton et al. (SHOTTON et al. 2013). Les données squelette sont ensuite filtrées pour réduire le bruit d'acquisition. Pour décrire les actions, un nouvel algorithme d'extraction des attributs nommé Body-part Directional Velocity (BDV) est appliqué sur les images une à une. Enfin, un modèle de classification de type modèle de mélange gaussien avec modèle de Markov Caché (*GMM-HMM*) est entraîné sur les attributs BDV. Le déroulement du processus de la partie apprentissage est effectué en temps réel. Par conséquent, dans la partie reconnaissance de l'action, le même processus est appliqué pour l'extraction des attributs. Afin de calculer la probabilité d'appartenance à une classe donnée, la vraisemblance est calculée pour chaque modèle HMM. Le modèle

présentant la vraisemblance la plus élevée correspond à l'action réalisée. Les étapes d'extraction des attributs et la construction du classifieur sont décrites plus en détail ci-dessous.

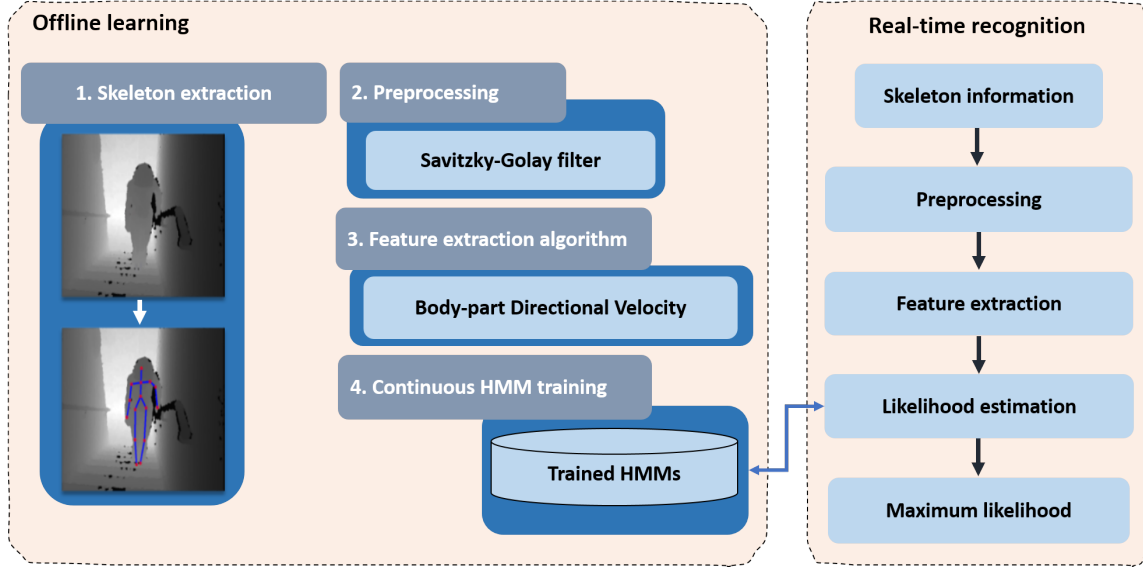


FIGURE 6-1 – Aperçu de la méthode proposée pour la reconnaissance d’actions à l’aide de la Kinect.

6.2.1 Body-part Directional Velocity (BDV)

Dans notre méthode, les attributs BDV sont extraits image par image afin d’accomplir une reconnaissance en temps réel de l’action. Dans cette section, nous détaillons le calcul des attributs.

Une séquence des données squelette \mathbf{p} représente une série de N poses ordonnées et décrites par l’équation (6.1).

$$\mathbf{p} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_t, \dots, \mathbf{p}_N] \quad (6.1)$$

La pose du squelette à chaque instant t , nommée \mathbf{p}_t , est composée d’un ensemble de n positions d’articulations, tel qu’il est décrit par l’équation (6.2) (où p_t^i est la position de la i^{th} articulation à un instant t).

$$\mathbf{p}_t = [p_t^1, p_t^2, \dots, p_t^i, \dots, p_t^n] \quad (6.2)$$

Premièrement, les données du squelette 3D n'étant pas toujours précises en raison du bruit d'acquisition et des occlusions, un pré-traitement est alors effectué. Ainsi, un filtre Savitzky-Golay est appliqué pour toutes les positions d'articulations comme suit : $\forall(i, t) \in \llbracket 1, n \rrbracket \times \llbracket 1, N \rrbracket$,

$$\mathbf{P}_t^i = \frac{1}{35}(-3\mathbf{p}_{t-2}^i + 12\mathbf{p}_{t-1}^i + 17\mathbf{p}_t^i + 12\mathbf{p}_{t+1}^i - 3\mathbf{p}_{t+2}^i) \quad (6.3)$$

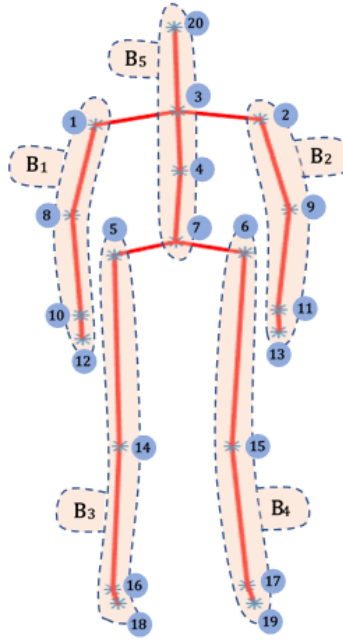


FIGURE 6-2 – Corps humain (squelette) divisé en cinq parties différentes (B_1 , B_2 , B_3 , B_4 et B_5).

où \mathbf{P}_t^i fait référence à la position de l'articulation i à un instant t après le filtrage.

Ensuite, la vitesse à un instant t de chaque articulation i , considérée comme une caractéristique très discriminante, est calculée comme dans (ZANFIR et al. 2013) en utilisant l'équation (6.4).

$$\mathbf{V}_t^i = \mathbf{P}_{t+1}^i - \mathbf{P}_{t-1}^i \quad (6.4)$$

Afin de calculer le vecteur d'attributs, nous proposons de diviser le squelette humain en cinq parties à savoir bras gauche (B_1), bras droit (B_2), jambe gauche (B_3), jambe droite (B_4) et la colonne vertébrale (B_5), comme illustré dans la figure 6-2.

L'ensemble des parties du corps est donc noté $B = \{B_1, B_2, B_3, B_4, B_5\}$.

Ensuite, pour chaque partie du corps, les vitesses négatives et positives des articulations associées sont respectivement additionnées et notées $D_{B_l}^+$ et $D_{B_l}^-$ dans l'équation (6.5) et l'équation (6.6).

$$\mathbf{D}_{B_l}^+(t) = \sum_{i \in B_l} (\mathbf{V}_t^i \geq 0) \quad (6.5)$$

$$\mathbf{D}_{B_l}^-(t) = \sum_{i \in B_l} (\mathbf{V}_t^i < 0) \quad (6.6)$$

Le vecteur final d'attributs à un instant t , noté $\mathbf{D}(t)$, est obtenu suivant l'équation (6.7).

$$\mathbf{D}(t) = \bigcup_{l=1}^5 [\mathbf{D}_{B_l}^+(t), \mathbf{D}_{B_l}^-(t)] \quad (6.7)$$

Par conséquent, la dimension du vecteur d'attributs proposé nommé BDV est de 30 ($\dim(BDV) = 30$).

6.2.2 Classification à l'aide d'un modèle GMM-HMM

Comme nous l'avons expliqué précédemment, les attributs BDV proposés dans la partie 6.2.1 sont générés graduellement. Parallèlement, une reconnaissance d'actions est effectuée en temps réel et d'une manière anticipée quand c'est possible. Pour atteindre cet objectif, les attributs calculés sont modélisés à l'aide de plusieurs HMMs dans des espaces continus (voir figure 6-3).

Un modèle de Markov Caché (Hidden Markov Model - HMM) (RABINER et al. 1986) est un modèle statistique utilisé pour décrire l'évolution des événements observables. Il est principalement utilisé pour modéliser des données temporelles dans la reconnaissance de la parole, du geste et de l'activité. Un HMM repose sur deux processus stochastiques. Le premier processus est observable, il représente la séquence des symboles observés. Le second n'est pas observable (caché) et peut être déduit indirectement en analysant la séquence des symboles observés.

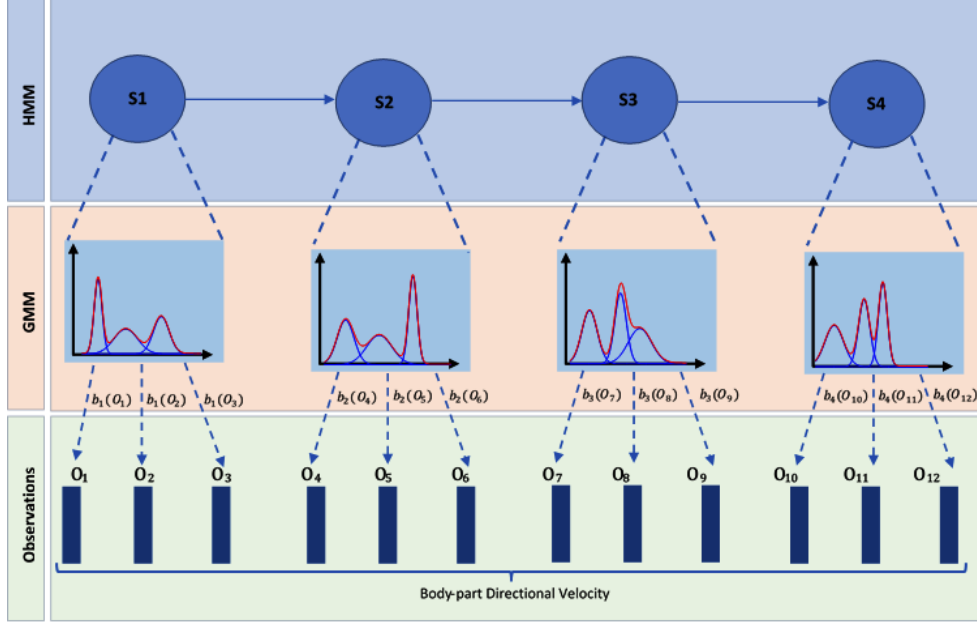


FIGURE 6-3 – Architecture du modèle de classification.

Pour chaque HMM^c entraîné pour une action c , N^c et M indiquent respectivement le nombre d'états dans le modèle et le nombre de symboles observés. $S^c = \{s_1^c, s_2^c, \dots, s_N^c\}$ se réfère à l'ensemble des états distincts, $V = \{v_1, v_2, \dots, v_M\}$ l'alphabet d'observation, $Q^c = \{q_1^c, q_2^c, \dots, q_T^c\}$ T états de S^c et $O = \{o_1, o_2, \dots, o_T\}$ T observations de l'alphabet V . Chaque HMM^c peut être écrit sous une forme compacte notée par $\lambda^c = (\pi^c, \mathbf{A}^c, \mathbf{B}^c)$.

π^c est la loi de probabilité de l'état initial :

$$\pi^c = \{\pi_i\}, \pi_i = P(q_1 = s_i)_{1 \leq i \leq N^c} \quad (6.8)$$

\mathbf{A}^c la matrice des probabilités de transition d'un état vers un autre :

$$\mathbf{A}^c = \{a_{ij}\}, a_{i,j} = P(q_{t+1} = s_j | q_t = s_i)_{1 \leq i, j \leq N^c} \quad (6.9)$$

\mathbf{B}^c la matrice des probabilités d'émission des symboles de V pour chaque état :

$$\mathbf{B}^c = \{b_{ik}\}, b_{ik} = P(o_t = v_k | q_t = s_i)_{1 \leq i \leq N^c, 1 \leq k \leq M} \quad (6.10)$$

Un HMM discret (D-HMM) considère les observations comme des symboles discrets issus d'un alphabet fini. Par conséquent, les attributs calculés sont quantifiés en utilisant un algorithme de classification non supervisé comme le k -means par exemple. Une quantification implique généralement une dégradation du modèle. Pour surmonter ce problème, des fonctions avec des lois de probabilités continues sont utilisées pour modéliser les attributs comme le montre l'équation 6.11.

$$b_i(o_t) = \sum_{r=1}^{N_g} w_{ir} g_r(o_t, \mu_{ir}, C_{ir})_{1 \leq i \leq N^c, 1 \leq k \leq M} \quad (6.11)$$

Où w_{ir} , μ_{ir} et C_{ir} représentent respectivement les poids, le vecteur de moyenne et la matrice de covariance r^{eme} d'un modèle gaussien dans l'état i . N_g est le nombre de composantes du mélange, dans nos tests expérimentaux, il est fixé empiriquement à $N_g = 3$. Nous rappelons que d_D est la dimension du vecteur d'attributs BDV.

Les fonctions de densité de probabilité employées sont les mélanges gaussiens multivariés (GMMs) où chacune est définie comme suit :

$$g_r(o_t, \mu_{ir}, C_{ir}) = \frac{1}{(2\pi)^{d_D/2} |C_{ir}|^{1/2}} e^{-\frac{1}{2}(o_t - \mu_{ir})^T C_{ir}^{-1} (o_t - \mu_{ir})} \quad (6.12)$$

Afin de classifier une action donnée, les HMMs entraînés reçoivent en entrée les observations. Une vraisemblance est ensuite estimée pour chaque HMM^c suivant l'algorithme 5. La classe qui fournit la plus grande vraisemblance est sélectionnée comme le montre l'équation 6.13.

$$c^*(t) = \arg \max(P(o_t | \lambda_i)) \quad (6.13)$$

6.3 Validation expérimentale

Afin de démontrer l'efficacité de notre approche, des tests expérimentaux sont menés sur des jeux de données publiques. Cette section comprend quatre parties : la première est la présentation des jeux de données sélectionnées pour tester notre ap-

Algorithm 5: Classification en temps réel des actions humaines

Data:

$C = \{action_1, action_2, \dots, action_n\}$: Listes des actions (HMMs entraînés) considérés par le système

p_t : Vecteur contenant les données squelette à l'instant t

π^c : Probabilité de l'état initial de l'action c

$A^c = \{a_{ij}\}$: Matrice des probabilités de transition de l'action c

$D(t) \leftarrow BDV(p_t);$ // Calcul du vecteur d'attributs

$p \leftarrow b_i(D(t));$ // Probabilités d'émission de $D(t)$ dans chaque état (équation 6.11)

foreach $c \in C$ **do**

 // $t=1$ correspond à la première image de l'action en cours d'exécution

if $t == 1$ **then**

for $j \leftarrow 1$ **to** N^c **do**

$\alpha_1(j) \leftarrow \pi^c(j) \cdot p(j);$

$loglik_1^c \leftarrow sum(log(\alpha_1));$

end

else

for $j \leftarrow 1$ **to** N^c **do**

for $i \leftarrow 1$ **to** N^c **do**

$m(i) \leftarrow a_{ij} \cdot \alpha_{t-1}(i);$

end

$\alpha_t(j) \leftarrow sum(m) \cdot p(j);$

end

$loglik_t^c \leftarrow loglik_{t-1}^c + sum(log(\alpha_t));$

end

end

$label_t \leftarrow \arg \max(loglik_t);$

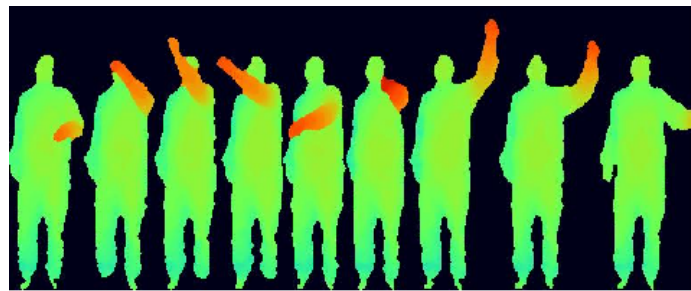
proche. La seconde est le protocole expérimental adopté pour nos expérimentations. Dans la troisième partie, une comparaison avec les travaux de l'état de l'art est réalisée. Enfin, dans la dernière partie, les résultats obtenus sont présentés et analysés pour la reconnaissance anticipée de l'action.

6.3.1 Jeux de données

Afin de démontrer l'efficacité de notre approche, et de se comparer avec les travaux de l'état de l'art, quatre jeux de données publiques sont exploités : MSRAAction3D (W. LI et al. 2010), UTKinect-Action (XIA, C.-C. CHEN et al. 2012), Florence 3D

(SEIDENARI et al. 2013) et Multiview3D (HAMMOUCHE et al. 2015).

Le jeu de données MSRAction3D est considéré comme une référence dans la littérature pour la reconnaissance des actions humaines basée sur des données *RGB-D*. Il contient à la fois les données de profondeur et de squelette. Ce jeu de données a été collecté par Microsoft Research et comprend 20 actions : *mouvement du bras vers le haut, mouvement du bras vers l'horizontale, mouvement du bras en marteau, prise de main, coup de poing en avant, lancer vers le haut, dessiner un "x", dessiner une coche, dessiner un cercle, applaudissement, mouvement des deux mains vers le haut, boxe sur le côté, se pencher, coup de pied en avant, coup de pied sur le côté, jogging, un tir au tennis, service de tennis, un tir de golf, ramasser et jeter*. Chaque action est effectuée par 10 sujets 2 ou 3 fois, pour un total de 567 séquences. La figure 6-4 montre deux exemples des séquences de profondeurs représentant les actions "dessiner une coche" et "service de tennis".



(a) *Draw tick*



(b) *Tennis serve*

FIGURE 6-4 – Aperçu de deux actions “dessiner une coche” et “service de tennis” issues du jeu de données MSRAction3D.

Le jeu de données Florence 3D a été collecté à l’université de Florence. Comme MSRAction3D, des données de profondeur et squelette sont disponibles. L’ensemble de données comprend 9 actions différentes : *mouvement du bras, boire d’une bouteille,*

répondre au téléphone, applaudir, lacer ses chaussures, s’asseoir, se lever, regarder la montre, s’incliner. Chaque action est effectuée par 10 sujets, 2 ou 3 fois, pour un total de 215 séquences. Le principal défi de ce jeu de données est que la même action est effectuée de diverses manières. Pour les actions qui sollicitent la main par exemple, elles sont exécutées une fois avec la main droite et une autre avec la main gauche (forte variance intra-classe). La figure 6-5 illustre un exemple de quelques images *RGB* représentant l’action “regarder la montre”.



FIGURE 6-5 – Action “regarder la montre” issue du jeu de données Florence 3D.

Le jeu de données UTKinect-Action contient 10 sujets réalisant chacun deux exécutions de 10 actions : *marcher, s’asseoir, se lever, ramasser, porter, lancer, pousser, tirer, mouvement des deux mains et applaudir.* Le jeu de données contient un défi supplémentaire par rapport à MSRAction3D et Florence 3D : il est collecté en utilisant différentes orientations du corps humain par rapport à la caméra (vue de droite, vue de face et vue de dos). La figure 6-6 montre quelques exemples d’images *RGB* et données profondeurs issues du jeu de données UTKinect-Action.

Multiview3D est un nouveau jeu de données collecté à l’Institut Mines Télécom Lille Douai (HAMMOUCHE et al. 2015). Il inclut 8 sujets effectuant 12 actions : *mouvement d’une main, boxe, s’asseoir, mouvement des deux mains, tenir la tête avec les mains, répondre au téléphone, ramasser, coup de pied, retenir, regarder la montre, sauter et jeter par-dessus la tête.* Chaque sujet effectue la même action deux fois pour trois orientations différentes ($30^\circ, 0^\circ, -30^\circ$) par rapport à la caméra. La structure de la base de données est présentée dans la figure 6-7.

6.3.2 Protocole expérimental

Afin de démontrer la robustesse de notre méthode et son indépendance aux sujets, deux protocoles expérimentaux sont appliqués suivant le jeu de données. Le premier,



FIGURE 6-6 – Aperçu de quelques images du jeu de données UTKinect-Action.

leave-one-subject-out-cross-validation (LOSOCV) : l'expérimentation est réalisée en N itérations (N dénote le nombre de sujets dans le jeu de données) en prenant à chaque fois les données d'un sujet pour la phase de test et les données des autres sujets pour la phase d'apprentissage. Les scores obtenus sont moyennés pour obtenir une mesure globale caractérisant la robustesse de notre système. Dans les expérimentations avec les jeux de données Florence 3D, UTKinect-Action et Multiview3D, la méthode de validation LOSOCV est utilisée afin de se comparer aux travaux existants. Le deuxième protocole expérimental dans notre travail, nommé "Hold-out subjects", est le découpage du jeu de données en deux sous-ensembles avec chacun, différents sujets. Cette méthode est généralement utilisée dans le jeu de données MSRAction3D. Nous l'appliquons également, suivant le découpage proposé dans les travaux de (W. LI et al. 2010) afin de réaliser une comparaison équitable. Outre la méthode de validation, le jeu de données MSRAction3D a été divisé en trois sous-ensembles : AS1, AS2, AS3 comme le montre la figure 6.1. Les étapes d'apprentissage et de test sont effectuées pour chaque sous-ensemble séparément. Les scores obtenus sont ensuite moyennés.

Dans le jeu de données Multiview3D, les actions sont observées à partir de trois

Actions	Orientation -30°	Orientation 0°	Orientation 30°
1 One Hand waving			
2 Box with 2 hands			
3 Sitting(chair)			
4 Two Hand waving			
5 Holding head			
6 Phone answering			
7 Picking up			
8 kicking			
9 Holding back			
10 Check watch			
11 jumping			
12 Throw over r head			

FIGURE 6-7 – Aperçu du jeu de données Multiview3D.

AS1	AS2	AS3
[a02] Bras vers l'horizontale	[a01] Bras vers le haut	[a06] lancer vers le haut
[a03] Bras en marteau	[a04] Prise de main	[a14] Coup de pied en avant
[a05] Coup de poing	[a07] Dessiner un "x"	[a15] Coup de pied sur le côté
[a06] Lancer vers le haut	[a08] Dessiner une coche	[a16] Jogging
[a10] Applaudissement	[a09] Dessiner un cercle	[a17] un tir au tennis (swing)
[a13] Se pencher	[a11] Mouvement des deux mains	[a18] Service de tennis
[a18] Service de tennis	[a12] Boxe sur le côté	[a19] un tir de golf (swing)
[a20] Ramasser et jeter	[a14] Coup de pied en avant	[a20] Ramasser et jeter

TABLE 6.1 – Trois sous-ensembles d'actions construits dans le jeu de données MSRAAction3D : AS1, AS2, AS3.

orientations différentes ($30^\circ, 0^\circ, -30^\circ$). Afin de démontrer l'impact de l'orientation (*personne, caméra*), trois techniques de tests sont appliquées : test "same-orientation" (les données d'apprentissage et test sont issues de la même orientation), le test "cross-orientation" (les données d'apprentissage et de test sont issues d'orientations dif-

férentes), et le dernier, test “all-orientation” (le modèle est appris suivant les trois orientations).

6.3.3 Expérimentation sur des jeux de données publiques

Reconnaissance de l’action humaine

Le tableau 6.2 présente les scores de reconnaissance comparés aux approches de l’état de l’art testées sur le jeu de données MSRAction3D. Les résultats présentés ci-dessous montrent que notre méthode atteint un score moyen de 92,9% de précision surpassant la plupart des méthodes existantes.

Algorithme	AS1 (%)	AS2 (%)	AS3 (%)	Moyenne globale (%)
(W. LI et al. 2010)	72,90	71,90	79,20	74,70
(VENKATARAMAN et al. 2013)	77,50	63,10	87,00	75,90
(C. CHEN et al. 2016)	96,20	83,20	92,00	90,50
(MIRANDA et al. 2014)	96,00	57,10	97,30	83,50
(CHAARAOUI et al. 2014)	91,59	90,83	97,28	93,23
(VEMULAPALLI et al. 2014)	95,29	83,87	98,22	92,46
(DU, W. WANG et al. 2015)	93,33	94,64	95,50	94,49
(CIPPITELLI et al. 2016)	79,50	71,90	92,30	81,50
(Z. LIU et al. 2016)	86,79	76,11	89,29	84,07
Nous	91,40	91,07	96,23	92,90

TABLE 6.2 – Précision de différentes méthodes sur le jeu de données MSRAction3D.

De plus, le tableau 6.3 compare notre méthode aux approches de l’état de l’art testées sur le jeu de données Florence3D. Notre approche présente l’un des meilleurs scores de précision par rapport aux méthodes de la littérature. Il enregistre 90,32% de précision (légèrement moins que l’approche de Vamulapalli et al. (VEMULAPALLI et al. 2014) qui présente un score de 90,88% de précision).

De cette manière, les résultats obtenus sur deux benchmarks prouvent que notre approche présente des résultats très satisfaisants comparés aux approches de la littérature.

Algorithme	Précision (%)
(SEIDENARI et al. 2013)	82,00
(ANIRUDH et al. 2015)	89,67
(DEVANNE et al. 2015)	87,04
(CIPPITELLI et al. 2016)	76,10
(VEMULAPALLI et al. 2014)	90,88
Nous	90,32

TABLE 6.3 – Précision de différentes méthodes sur le jeu de données Florence3D.

Invariance à l’orientation

Dans les expérimentations effectuées jusqu’à présent, les sujets effectuant les actions se trouvent face à la caméra, ce qui constitue une condition idéale pour un système de reconnaissance d’actions. Cependant, dans un contexte réel, le sujet n’est pas forcément face à la caméra, ce qui rend la tâche de reconnaissance plus complexe. Ainsi, pour tester la robustesse de notre méthode aux changements de l’orientation de la personne par rapport à la caméra, des expériences sont menées sur deux jeux de données UTKinect-Action et Multiview3D collectés en utilisant différents points de vue.

Algorithme	Précision (%)
(Y. ZHU et al. 2013)	87,90
(SLAMA et al. 2015)	88,50
(XIA, C.-C. CHEN et al. 2012)	90,92
(VEMULAPALLI et al. 2014)	97,08
Nous	91,10

TABLE 6.4 – Précision de différentes méthodes sur le jeu de données UTKinect-Action.

Le tableau 6.4 indique les scores de reconnaissance obtenus dans le jeu de données UTKinect-Action comparés aux méthodes de l’état de l’art. Les résultats montrent que notre approche surpasse diverses méthodes existantes atteignant un score de précision de 91,1%. Nous notons que l’approche proposée par VEMULAPALLI et al. (2014) présente une meilleure précision, sauf que cette méthode est très lente dans

	M	S	Le	R	Por	La	Pous	T	MM	App
Marcher	61	0	0	0	39	0	0	0	0	0
S’asseoir	0	95	0	5	0	0	0	0	0	0
Se lever	0	0	100	0	0	0	0	0	0	0
Ramasser	0	0	0	100	0	0	0	0	0	0
Porter	5	0	0	0	95	0	0	0	0	0
Lancer	0	0	0	0	5	80	0	0	5	10
Pousser	0	0	0	0	0	5	95	0	0	0
Tirer	0	0	0	0	0	5	0	95	0	0
Mouv. des mains	0	0	0	0	0	0	0	0	95	5
Applaudir	0	0	0	0	0	0	0	0	5	95

TABLE 6.5 – Matrice de confusion (%) sur le jeu de données UTKinect-Action.

la prise de décision. Dans (GHORBEL et al. 2016), le temps de calcul du système de (VEMULAPALLI et al. 2014) a été évalué sur le jeu de données MSRAction3D. Les résultats ont indiqué que l’approche prend en moyenne 17,61 secondes pour le calcul des attributs (temps moyen sur une séquence complète), sachant que MSRAction3D contient des vidéos segmentées composées de 12 à 54 images. L’algorithme a été testé sur un ordinateur portable doté d’un processeur Intel Core i7 et d’une mémoire vive de 4 Go (similaire à la nôtre). Nous montrerons en détail à la fin de cette section que notre approche est beaucoup plus intéressante en termes de temps de calcul.

La matrice de confusion est présentée dans le tableau 6.5. Nous pouvons remarquer que presque toutes les actions sont bien reconnues. Cependant, nous observons une certaine confusion entre les actions *marcher* et *porter*. La raison principale est que dans certaines séquences du jeu de données, les sujets marchent en effectuant l’action “porter”. Ce verrou peut être levé en exploitant les informations liées aux interactions d’objets, qui ne sont pas disponibles dans les données squelette.

		Test		
		0°	30°	−30°
Apprentissage	0°	92,71	90,63	90,10
	30°	90,63	90,70	87,10
	−30°	92,19	87,00	92,23
	(0° 30° −30°)	92,71	91,15	91,67

TABLE 6.6 – Précision (%) sur le jeu de données Multiview3D.

Table 6.6 rapporte la précision obtenue suivant les trois techniques de tests énoncées dans la partie 6.3.2. Les résultats montrent que lorsque la même orientation est utilisée pour l'apprentissage et le test, nous atteignons une précision moyenne de 91,88%, score assez prometteur et proche des résultats obtenus dans les autres jeux de données. En appliquant le test “cross-orientation”, un score moyen de 89,61% est alors obtenu. Nous remarquons que les résultats obtenus suivant les deux techniques de tests “same-orientation” et “cross-orientation” sont assez proches (différence de 2,27%). Ces résultats démontrent la robustesse de notre méthode à l'orientation de la personne par rapport à la caméra. De plus, la dernière technique de test, “all-orientation” montre une précision moyenne de 91,84%, presque similaire aux résultats du test “same-orientation”.

6.3.4 Reconnaissance anticipée de l'action humaine

Dans notre travail, pour réaliser une reconnaissance anticipée de l'action humaine, les données squelette sont générées image par image. À chaque nouvelle donnée reçue, notre système est capable de générer un label de l'action en cours d'exécution.

Afin de révéler l'efficacité de notre démarche, et démontrer que les actions sont prédictibles bien avant la fin de l'exécution de celles-ci, nous proposons un test expérimental décrit comme suit : soit un jeu de données contenant x séquences de données squelette sous format $p = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_t, \dots, \mathbf{p}_{N_x}]$. Chaque séquence de données est mise à l'entrée de notre système d'une manière incrémentale. Les labels en sortie sont analysés afin de déterminer K telle que :

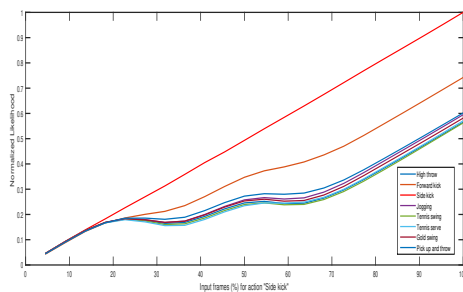
$$\text{label}_K = \text{label}_i \quad \forall i \in \mathbb{N} \quad \text{avec} \quad K \leq i \leq N_x \quad (6.14)$$

L'indice K correspond au squelette p_K 0 à partir duquel la décision du système sur la classe de l'action est stable. Étant donné que les actions ne contiennent pas le même nombre de données squelette, l'indice est divisé sur la taille de la séquence complète N_x afin de représenter cette information en pourcentage.

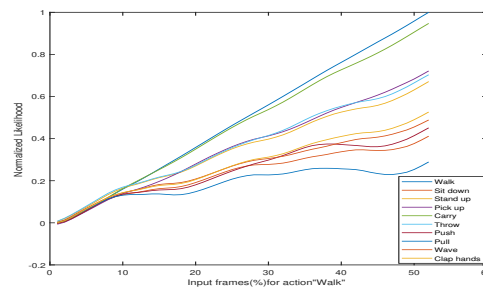
Le jeu de données MSRAction3D est sélectionné pour cette expérience. La figure 6-

8a montre un exemple des l'évolution des vraisemblances obtenues lors de l'exécution de l'action "coups de pied sur le côté" en fonction de la taille de la séquence. Les résultats montrent qu'au début de l'exécution de l'action, les vraisemblances obtenues pour chaque HMM^c sont presque égales. À partir de 20% de la séquence globale, la vraisemblance issue du HMM correspondant à l'action "coups de pied sur le côté", surpasse les vraisemblances des autres classes et cela jusqu'à la fin de l'exécution de l'action.

Un autre exemple intéressant appartenant au jeu de données UTKinect-Action est illustré dans la figure 6-8b. Il montre les vraisemblances obtenues lors de l'exécution de l'action "Marcher". Après la réception de seulement 10% des données de la séquence globale, les vraisemblances correspondant aux actions "marcher" et "porter" deviennent nettement supérieures aux vraisemblances des autres classes. La raison est que ces actions sont les seules dans le jeu de données UTKinect-Action impliquant fortement le mouvement des jambes (les sujets marchaient tout en effectuant l'action "porter"), ce qui a permis leur discrimination dès les premières données squelette. Nous remarquons aussi que les vraisemblances de ces actions prédominantes sont très proches au-delà de 10% des données réceptionnées, et cela jusqu'à la fin de l'exécution de l'action. Rappelons que les résultats de la matrice de confusion présentés dans le tableau 6.5 ont montré la présence de confusion entre l'action "marcher" et "porter". En effet, il est très difficile de discriminer ces deux actions, même après avoir observé la séquence complète.



(a) Coup de pied sur le côté.



(b) Marcher.

FIGURE 6-8 – Évolution de la vraisemblance lors de l'exécution d'actions de coups de pied sur le côté et marcher.

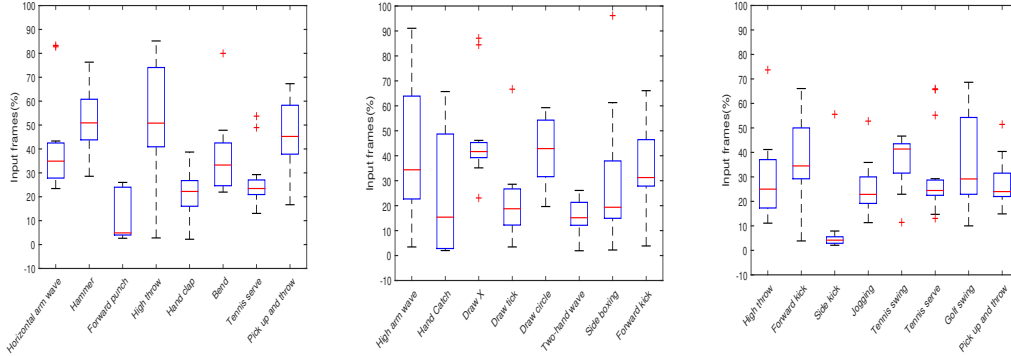


FIGURE 6-9 – Boîtes à moustaches montrant le pourcentage des données squelette ($[p_1, \dots, p_k]$) mises à l’entrée de notre système pour reconnaître les actions dans les trois sous-ensembles AS1, AS2, AS3 du jeu de données MSRAction3D (de gauche à droite).

Afin de visualiser les résultats obtenus d’une manière globale, nous utilisons des boîtes à moustaches. Elles permettent de décrire graphiquement la distribution des données numériques à travers un rectangle allant du premier quartile au troisième quartile et couper par la médiane. des segments sont ajoutés aux extrémités menant jusqu’aux valeurs extrêmes (échantillon minimum, échantillon maximum). L’écart interquartile (IQR) représente la différence entre le troisième et le premier quartile, il est illustré par la longueur de la boîte à moustache. Contrairement à la moyenne et l’écart-type, la médiane et l’écart interquartile sont robustes aux valeurs aberrantes.

Les boîtes à moustaches obtenues sur les sous-ensembles AS1, AS2 et AS3 du jeu de données MSRAction3D sont illustrées dans la figure 6-9. Elles montrent pour chaque action appartenant à un sous-ensemble, la distribution du nombre de données squelette ($[p_1, \dots, p_k]$) nécessaire pour reconnaître celle-ci. D’une manière générale, nous observons des distributions différentes pour chaque action. La valeur médiane (Med) est représentée par un segment rouge à l’intérieur du rectangle. Suivant les résultats obtenus dans les trois sous-ensembles, $Med \in [4\%, 52\%]$. La borne supérieure (52%), correspond à la classe “mouvement du bras en marteau”. Cela démontre que la moitié des actions présentes dans le jeu de données sont identifiables presque au milieu de l’exécution de celles-ci, ce qui est très prometteur.

La valeur minimale de la médiane (4%) est obtenue pour les deux classes “coup

de poing en avant” et “coup de pied sur le côté”. La moitié de ces actions sont reconnues très rapidement. Cela est probablement dû aux propriétés des attributs BDV qui prennent en compte la direction du mouvement des membres du corps. En effet l’action “coup de poing en avant” est la seule classe du sous-ensemble AS1 dont un seul bras se déplace vers l’avant. Pour la classe “Coup de pied sur le côté”, une seule jambe se déplace dans la direction latérale.

Dans les boîtes à moustaches, L’écart interquartile (IQR) indique la répartition en pourcentage de la taille des données squelette nécessaires à la reconnaissance d’une action. Dans la figure 6-9, $IQR \in [3\%, 45\%]$. Ces résultats pourraient être dûs à la variabilité d’exécution de l’action dans chaque classe. Les actions “mouvement du bras vers le haut” et “prise de main” présentent la plus grande dispersion enregistrée, i.e., que le nombre de données squelette pour reconnaître l’action peut considérablement varier suivant la personne ou la manière d’exécuter l’action. À l’inverse, la classe “coup de pied sur le côté” montre la plus faible dispersion.

Les valeurs aberrantes sont également indiquées par les boîtes à moustaches. Elles sont tracées individuellement à l’aide du symbole "+". Elles représentent les valeurs d’observation éloignées des autres observations. Notons q_1 et q_3 le premier et le troisième quartile, les valeurs aberrantes sont supérieures à $q_3 + 1,5 \times (q_3 - q_1)$ ou inférieures à $q_1 - 1,5 \times (q_3 - q_1)$. La valeur la plus élevée observée est $K = 98\%$ appartenant à la classe “boxe sur le côté”. Nous notons que ces valeurs aberrantes représentent des cas très rares (quelques exemples sur un total 567 séquences d’actions).

Les trois sous-ensembles AS1, AS2, AS3 partagent les actions suivantes : “lancer vers le haut”, “ramasser et jeter”, “coup de pied en avant” et “service de tennis”. Les boîtes à moustaches des deux premières classes montrent que, selon le sous-ensemble auquel ces actions appartiennent (AS1 ou AS3), la distribution de la taille des données pour reconnaître les actions est différente. En revanche, les classes “coup de pied en avant” (AS2, AS3) et “service de tennis” (AS1, AS3) montrent des distributions presque similaires. Cette variabilité peut être expliquée par la complexité du problème de classification dans chaque sous-ensemble. Comme nous l’avons montré dans les exemples illustrés dans la figure 6-8, une action qui appartient à un sous-ensemble

contenant des actions similaires à celle-ci est plus difficile à discriminer et elle nécessite plus de données squelette (probablement toute la séquence dans certains cas). Par contre si la même action est déplacée dans un sous-ensemble présentant des actions plus éloignées à celle-ci, elle est rapidement reconnue et nécessite moins de données en entrée du système.

6.3.5 Temps de calcul

Afin de prouver que notre approche convient à une reconnaissance anticipée, nous rapportons le temps de traitement nécessaire pour chaque image. Auparavant, il convient de souligner que la fréquence d'échantillonnage d'un capteur Kinect est de $30Hz$. Cela signifie qu'une image de profondeur est traitée toutes les $33,33ms$. Dans (SHOTTON et al. 2013), Shotton et al. indiquent que la mise en œuvre optimisée de l'algorithme proposé permet l'extraction du squelette en moins de $5ms$. Par conséquent, pour effectuer une reconnaissance anticipée de l'action en temps réel, le temps de calcul pour chaque donnée squelette doit être inférieur à $33,33ms$ pour une approche basée sur la profondeur et à $28,33ms$ pour une approche basée sur le squelette (en soustrayant le temps nécessaire pour le calcul des données squelette). Dans (C. CHEN et al. 2016), le calcul des données DMM pour chaque donnée profondeur (partie 5.2.1) est réalisé en $5,3ms$ en moyenne. Après avoir calculé les attributs pour toute la séquence de données profondeur (de taille 30 en moyenne), la méthode PCA suivie par la classification de l'action sont appliquées en $2,5ms$ et $1,5ms$ respectivement. Dans (DEVANNE et al. 2015), la reconnaissance d'une séquence complète de données squelette est réalisée en $50ms$ (38 images par séquence en moyenne). La méthode proposée par CHAARAoui et al. (2014) présente un temps de calcul de $1,89ms$ par image squelette en moyenne.

Le temps de calcul de notre approche proposée est évalué sur le jeu de données MSRAAction3D. Tous les tests expérimentaux sont exécutés en temps réel à l'aide de Matlab sur un processeur Intel Core i5 2,60 GHz et 4 Go de RAM. Le temps de calcul est évalué par la fonction tic-toc de MATLAB qui fournit une résolution de $1\mu s$ (KNAPP-CORDES et al. 2011b). L'expérience a été répétée 100 fois pour obtenir

des résultats significatifs. Seule la phase de prédiction est concernée dans cette expérience puisque la phase d'apprentissage est réalisée en mode hors-ligne. Pour chaque étape de notre approche, nous avons calculé la moyenne et l'écart type du temps de calcul estimé. Les résultats sont indiqués dans le tableau 6.7. Avec une moyenne de $0,015ms$ pour chaque image, nous montrons que non seulement notre approche est adaptée pour la reconnaissance anticipée de l'action (temps de calcul bien inférieur à $28,33ms$), mais aussi qu'elle est plus rapide comparée aux approches de l'état de l'art citées.

Étape	Temps de traitement (ms)
Pre-traitement (filtrage)	$0,003 \pm 0,001$ /image
Calcul des attributs BDV	$0,005 \pm 0,001$ /image
Likelihood par HMM	$0,007 \pm 0,002$ /image

TABLE 6.7 – Temps de calcul dans chaque étape du système.

6.4 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle approche permettant de reconnaître l'action humaine à l'aide de capteurs *RGB-D*. Nous avons concentré nos travaux sur deux défis majeurs pour les applications de divertissements contenant des mouvements physiques : la robustesse face aux changements de points de vue et la propriété de reconnaissance anticipée de l'action.

Un nouvel algorithme d'extraction d'attributs en temps réel appelé BDV (Body-part Directional Velocity) a été proposé. Une classification basée sur les modèles de Markov cachés (HMMs) avec des distributions de sortie d'états représentées par des modèles de mélange gaussiens (GMMs) a été proposée pour reconnaître les actions humaines. Afin de démontrer la pertinence de notre approche, nous avons dans un premier temps testé notre système en utilisant l'ensemble de la séquence de données squelette pour reconnaître les actions. Les résultats expérimentaux sur deux jeux de données publiques ont démontré l'efficacité de notre approche et sa supériorité par

rapport aux approches de l'état de l'art en atteignant une précision moyenne de 92,9% dans le jeu de données MSRAction3D et de 90,32% dans Florence3D. La deuxième partie des expérimentations concerne la validation de l'invariance de l'approche à l'orientation (changements de l'orientation de la personne par rapport à la caméra). Pour ce faire, des expérimentations ont été effectuées sur deux autres jeux de données présentant ce challenge. Dans le premier, UTKinect-Action, un score de 91,1% a été obtenu, se positionnant ainsi parmi les meilleures approches testées sur ce jeu de données. Dans le deuxième test, un jeu de données collecté au sein de l'équipe de recherche URJA est exploité. Différentes techniques de tests ont été considérées démontrant ainsi la robustesse de notre approche à l'orientation. La troisième partie de notre expérimentation a impliqué une reconnaissance anticipée des actions humaines. Dans notre étude, nous nous sommes intéressés à la proportion des données squelette nécessaires pour reconnaître chaque action. Pour ce faire un test expérimental a été proposé et une validation a été menée sur le jeu de données MSRAction3D. Les boîtes à moustaches ont été exploitées afin de visualiser les résultats obtenus. Notre solution a démontré que toutes les actions appartenant au jeu de données sont identifiables avant la fin de leurs exécutions. La moitié des actions sont prévisibles presque au milieu de l'ensemble de la séquence de données squelette, ce qui est très prometteur. Certaines classes ont été reconnues avec seulement 4% de l'ensemble des données de la séquence.

Quatrième partie

Conclusion et perspectives

Chapitre 7

Conclusion et perspectives

Contents

7.1	Conclusion générale	188
7.2	Perspectives	189
7.2.1	Proposition d'indicateurs d'autonomie	190
7.2.2	Transfert d'indicateurs d'autonomie aux professionnels de la santé	191
7.2.3	Conception d'un programme d'exercices suivant l'autonomie de la personne	191

7.1 Conclusion générale

Dans cette thèse nous avons ciblé une catégorie de notre société qui est la personne âgée. Notre étude oeuvre au maintien de leur bien-être et de leur autonomie. Le premier verrou de notre travail consistait à faire un choix judicieux sur le matériel adéquat susceptible de nous permettre la mise en place d'une base de données adaptée à notre application. Cette base devait être en mesure de nous permettre un traitement et une analyse pertinente afin d'évaluer le degré d'autonomie de la personne âgée dans une première phase. Nous avons choisi le smartphone comme matériel central à notre application, pour son respect de la vie privée (du moins dans notre application), son faible coût et l'ensemble des capteurs et moyens de communication disponibles (donnant de bonnes performances en présence du bruit et une bonne précision à reconnaître les actions). Grâce à la base de données contenant les valeurs de capteurs embarqués dans le smartphone, i.e l'accéléromètre et le gyroscope, nous avons entamé la phase de mise en place du système de reconnaissance de l'action. Dans notre analyse, nous avons distingué trois catégories d'actions intimement corrélées avec l'autonomie et le bien-être : les actions sédentaires, les actions périodiques ou pseudo-périodiques et les actions non périodiques. Dans cette phase de notre étude, il s'agissait de développer un système intelligent de suivi des trois types d'actions humaines que nous avons déterminé. Dans ce contexte, un système de classification en temps-réel est mis au point et une approche de classification hiérarchique des actions humaines est proposée. Notre système commence dans un premier niveau de classification par trouver la catégorie de l'action puis ensuite sélectionner le classifieur adapté à reconnaître l'action proprement dite. La qualité et la performance de cette classification a été atteinte grâce à la sélection d'attributs optimaux et pertinents, adapté à notre classification, au niveau de chaque nœud interne de notre méthode de classification. L'aspect robustesse à l'emplacement des capteurs de la méthode proposée a été pris en compte dans notre démarche et nous avons justifié l'indépendance à la position des capteurs obtenue par le biais de l'algorithme de sélection des attributs que nous avons proposé. Il est à noter que nous avons aussi mis l'accent dans cette étude

sur la reconnaissance des transitions posturales par la mise en place d'un système de détection et de reconnaissance de six types de postures. La synthèse journalière de l'ensemble des actions effectuées par la personne âgée, va nous renseigner sur son degré d'autonomie et nous permettra d'établir un programme d'exercices physiques susceptibles de promouvoir son autonomie et son bien-être. C'est dans cette optique que nous avons proposé un début de solution à cette problématique et de premiers tests sur une base de données adaptée. Dans un second temps et dans une optique de proposer par la suite des moyens d'améliorer leur bien-être aux personnes suivies, nous avons choisi la Kinect pour la réalisation des exercices physiques sous forme de mouvements contrôlés reconnus et corrigés à partir de la mise en place d'un deuxième système intelligent qui va analyser et traiter les données délivrées par la caméra en question. La simplicité d'utilisation de cette caméra a motivé à la choisir pour cette application. Ce volet de notre étude cible la reconnaissance des actions humaines qui correspondent aux mouvements physiques qui seront proposés après l'analyse des résultats obtenus à partir de notre premier système. Dans ce contexte nous avons développé le deuxième système intelligent de reconnaissance de l'action, opérant en temps réel, robuste et précis avec la capacité de reconnaître une action avant qu'elle ne soit achevée.

7.2 Perspectives

Dans ces travaux de thèse, nous avons proposé deux approches pertinentes pour la reconnaissance d'actions, chacune destinée à une application particulière dans le domaine de la santé en général : la première consiste à exploiter le smartphone des utilisateurs afin de réaliser un bilan des actions et activités journalières, la seconde approche utilise la Kinect pour la prévention de la perte d'autonomie. Afin de montrer les perspectives et leurs liens avec les travaux achevés jusque-là, nous proposons de suivre le schéma illustré dans la figure 7-1 (version moins détaillée de l'approche générale montrée dans figure 2-9 du chapitre 2). Sur cette illustration, nous pouvons voir, en bleu, le travail effectué dans cette thèse et en rouge les perspectives. Ces

derniers seront détaillées ci-dessous.

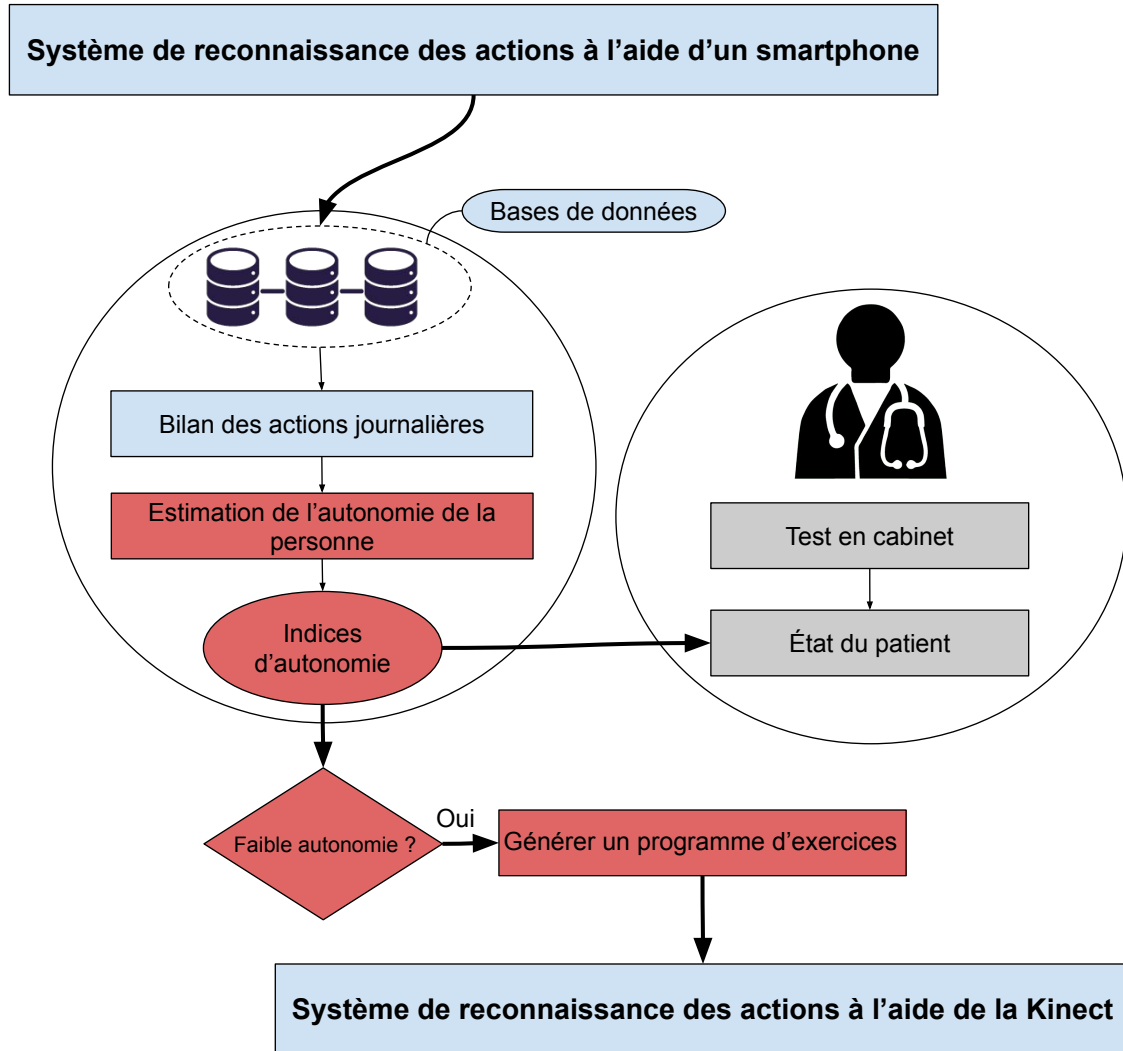


FIGURE 7-1 – Travaux réalisés en thèse accompagnés des perspectives.

7.2.1 Proposition d'indicateurs d'autonomie

Le système que nous avons proposé pour la surveillance des mouvements humains a la capacité de reconnaître les actions humaines avec une grande précision. Afin de profiter des résultats obtenus à travers notre système et de l'étendre encore plus au domaine de la santé, un indicateur d'autonomie de la personne (ou plusieurs indicateurs) sera(ont) proposé(s) en perspective. L'objectif visé est le même que les travaux cités dans la partie 1.2 du chapitre 1 : calcul de l'autonomie d'une personne. Dans les

travaux mentionnés (l'indice de Katz, la grille, etc.), l'évaluation de l'autonomie ou le degré de dépendance d'un individu est calculé manuellement au travers d'une série de questions. L'objectif recherché dans les perspectives est de réaliser d'une manière automatique cette opération : en exploitant les bases de données contenant les actions et leurs moments d'exécution, des indicateurs d'autonomie peuvent être calculés suivant l'occurrence et la nature des actions (sédentaire, mouvements de déambulation, etc.). De cette manière, des résultats plus détaillés et précis peuvent être obtenus en comparaison de l'utilisation d'un questionnaire. De plus, l'opération est beaucoup moins coûteuse que l'envoi d'une équipe médicale au domicile des personnes. C'est le cas de l'évaluation de l'autonomie avec la grille AGGIR, où une équipe médico-sociale du département est envoyée aux domiciles des personnes demandeuses de l'allocation personnalisée d'autonomie.

7.2.2 Transfert d'indicateurs d'autonomie aux professionnels de la santé

Cette perspective s'inscrit dans la même continuité que la proposition d'indicateurs d'autonomie. Ces derniers, une fois calculés, peuvent servir aux professionnels de la santé afin d'évaluer l'état du patient : les indicateurs d'autonomie d'un patient peuvent être combinés avec les résultats de tests en cabinets effectués par un médecin. On cite par exemple le test d'équilibre ou le test "sit-to-stand" (revoir la partie 2.4 du chapitre 2 pour plus de détails).

7.2.3 Conception d'un programme d'exercices suivant l'autonomie de la personne

Les indices d'autonomie ont pour objectif de résumer l'activité physique de la personne durant la journée. Dans le cas où cette activité est faible (ratio des actions sédentaires très élevé comparé aux mouvements de déambulation par exemple), nous avons proposé dans notre travail de thèse de pratiquer des exercices physiques à l'aide de la Kinect. Afin de profiter au mieux de ce système, la conception d'un programme

d'exercices particulier et spécifique à chaque utilisateur s'inscrit en perspective : en exploitant les différents indicateurs d'autonomie, nous pouvons proposer (d'une manière automatique) un programme d'exercices physiques qui vise à améliorer certains paramètres liés à l'autonomie comme la force musculaire des membres inférieurs.

Bibliographie

- ADAMI, AM et al. (2006). "Detection and classification of movements in bed using load cells". In : *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, p. 589-592.
- AGGARWAL, Jake K et Michael S RYOO (2011). "Human activity analysis: A review". In : *ACM Computing Surveys (CSUR)* 43.3, p. 16.
- AGMON, Maayan et al. (2011). "A pilot study of Wii Fit exergames to improve balance in older adults". In : *Journal of geriatric physical therapy* 34.4, p. 161-167.
- AINSWORTH, Barbara E et al. (2000). "Compendium of physical activities: an update of activity codes and MET intensities". In : *Medicine and science in sports and exercise* 32.9; SUPP/1, S498-S504.
- ALMASLUKH, Bandar, Abdel ARTOLI et Jalal AL-MUHTADI (2018). "A Robust Deep Learning Approach for Position-Independent Smartphone-Based Human Activity Recognition". In : *Sensors* 18.11, p. 3726.
- ALTUN, Kerem, Billur BARSHAN et Orkun TUNÇEL (2010). "Comparative study on classifying human activities with miniature inertial and magnetic sensors". In : *Pattern Recognition* 43.10, p. 3605-3620.
- ALWAN, Majd et al. (2006). "A smart and passive floor-vibration based fall detector for elderly". In : *Information and Communication Technologies*. T. 1. 2, p. 1003-1007.
- ANGUITA, Davide et al. (2012). "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine". In : *International workshop on ambient assisted living*. Springer, p. 216-223.

- ANIRUDH, Rushil et al. (2015). “Elastic functional coding of human actions: From vector-fields to latent variables”. In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 3147-3155.
- ANTOS, Stephen A, Mark V ALBERT et Konrad P KORDING (2014). “Hand, belt, pocket or bag: Practical activity tracking with mobile phones”. In : *Journal of neuroscience methods* 231, p. 22-30.
- ARIF, Muhammad et al. (2014). “Better physical activity classification using smart-phone acceleration sensor”. In : *Journal of medical systems* 38.9, p. 95.
- ASHBURN, Ann et al. (2001). “A community-dwelling sample of people with Parkinson’s disease: characteristics of fallers and non-fallers”. In : *Age and ageing* 30.1, p. 47-52.
- ATALLAH, Louis et al. (2011). “Sensor positioning for activity recognition using wearable accelerometers”. In : *IEEE transactions on biomedical circuits and systems* 5.4, p. 320-329.
- ATLAS, Steven J et Richard A DEYO (2001). “Evaluating and managing acute low back pain in the primary care setting”. In : *Journal of general internal medicine* 16.2, p. 120-131.
- ATTAL, Ferhat et al. (2015). “Physical human activity recognition using wearable sensors”. In : *Sensors* 15.12, p. 31314-31338.
- BANOS, Oresti, Rafael GARCIA et al. (2014). “mHealthDroid: a novel framework for agile development of mobile health application”. In : *International workshop on ambient assisted living*. Springer, p. 91-98.
- BANOS, Oresti, Claudia VILLALONGA et al. (2015). “Design, implementation and validation of a novel open framework for agile development of mobile health applications”. In : *Biomedical engineering online* 14.2, S6.
- BAO, Ling et Stephen S INTILLE (2004). “Activity recognition from user-annotated acceleration data”. In : *International Conference on Pervasive Computing*. Springer, p. 1-17.

- BAYAT, Akram, Marc POMPLUN et Duc A TRAN (2014). “A study on human activity recognition using accelerometer data from smartphones”. In : *Procedia Computer Science* 34, p. 450-457.
- BELHAKEM, Nadia et Cyril GICQUIAUX (2018). “Une perte d’autonomie chez les seniors plus fréquente qu’au niveau national”. In :
- BELONGIE, Serge, Jitendra MALIK et Jan PUZICHA (2002). “Shape matching and object recognition using shape contexts”. In : *IEEE Transactions on Pattern Analysis & Machine Intelligence* 4, p. 509-522.
- BERG, Katherine O et al. (1992). “Measuring balance in the elderly: validation of an instrument”. In : *Canadian journal of public health= Revue canadienne de sante publique* 83, S7-11.
- BOBICK, Aaron F (1997). “Movement, activity and action: the role of knowledge in the perception of motion”. In : *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 352.1358, p. 1257-1265.
- BRADBURN, Norman M et David CAPLOVITZ (1965). *Reports on happiness: A pilot study of behavior related to mental health*. 3. Aldine Pub. Co.
- BRUNEL, Mathieu et Amélie CARRÈRE (2017). *Incapacités et perte d’autonomie des personnes âgées en France: une évolution favorable entre 2007 et 2014*. La Direction de la recherche, des études, de l’évaluation et des statistiques (Drees).
- CAPELA, NA et al. (2016). “Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants”. In : *Journal of neuroengineering and rehabilitation* 13.1, p. 5.
- CASALE, Pierluigi, Marco ALTINI et Oliver AMFT (2015). “Transfer learning in body sensor networks using ensembles of randomized trees”. In : *IEEE Internet of Things Journal* 2.1, p. 33-40.
- CATAL, Cagatay et al. (2015). “On the use of ensemble of classifiers for accelerometer-based activity recognition”. In : *Applied Soft Computing* 37, p. 1018-1022.
- CHAARAOU, Alexandros Andre et al. (2014). “Evolutionary joint selection to improve human action recognition with RGB-D devices”. In : *Expert systems with applications* 41.3, p. 786-794.

- CHAN, Marie, Eric CAMPO et Daniel ESTÈVE (2005). "Assessment of activity of elderly people using a home monitoring system". In : *International Journal of Rehabilitation Research* 28.1, p. 69-76.
- CHANG, Chih-Chung et Chih-Jen LIN (2011). "LIBSVM: a library for support vector machines". In : *ACM transactions on intelligent systems and technology (TIST)* 2.3, p. 27.
- CHEN, Chen, Kui LIU et Nasser KEHTARNAVAZ (2016). "Real-time human action recognition based on depth motion maps". In : *J. Real-Time Image Processing* 12.1, p. 155-163.
- CHEN, Jiawei et al. (2017). "Semi-coupled two-stream fusion convnets for action recognition at extremely low resolutions". In : *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, p. 139-147.
- CHEN, Yufei et Chao SHEN (2017). "Performance analysis of smartphone-sensor behavior for human activity recognition". In : *Ieee Access* 5, p. 3095-3110.
- CHENG, Pao-Tsai et al. (1998). "The sit-to-stand movement in stroke patients and its correlation with falling". In : *Archives of physical medicine and rehabilitation* 79.9, p. 1043-1046.
- CHO, Heeryon et Sang Min YOON (2018). "Divide and Conquer-Based 1D CNN Human Activity Recognition Using Test Data Sharpening". In : *Sensors* 18.4, p. 1055.
- CHOWDHURY, Alok Kumar et al. (2018). "Physical Activity Recognition Using Posterior-Adapted Class-Based Fusion of Multiaccelerometer Data". In : *IEEE journal of biomedical and health informatics* 22.3, p. 678-685.
- CIPPITELLI, Enea et al. (2016). "A human activity recognition system using skeleton data from rgbd sensors". In : *Computational intelligence and neuroscience 2016*, p. 21.
- CLELAND, Ian et al. (2013). "Optimal placement of accelerometers for the detection of everyday activities". In : *Sensors* 13.7, p. 9183-9200.

- CONCEPCIÓN, Miguel Ángel Álvarez de la et al. (2017). “Mobile activity recognition and fall detection system for elderly people using Ameva algorithm”. In : *Pervasive and Mobile Computing* 34, p. 3-13.
- COOLEY, James W et John W TUKEY (1965). “An algorithm for the machine calculation of complex Fourier series”. In : *Mathematics of computation* 19.90, p. 297-301.
- DAI, Ji et al. (2015). “Towards privacy-preserving recognition of human activities”. In : *2015 IEEE international conference on image processing (ICIP)*. IEEE, p. 4238-4242.
- DALAL, Navneet et Bill TRIGGS (2005). “Histograms of oriented gradients for human detection”. In :
- DEMAISON, Catherine et al. (2018). “France, portrait social”. In :
- DENG, Wan-Yu, Qing-Hua ZHENG et Zhong-Min WANG (2014). “Cross-person activity recognition using reduced kernel extreme learning machine”. In : *Neural Networks* 53, p. 1-7.
- DERNBACH, Stefan et al. (2012). “Simple and complex activity recognition through smart phones”. In : *Intelligent Environments (IE), 2012 8th International Conference on*. IEEE, p. 214-221.
- DEVANNE, Maxime et al. (2015). “3-d human action recognition by shape analysis of motion trajectories on riemannian manifold”. In : *IEEE transactions on cybernetics* 45.7, p. 1340-1352.
- DOLLÁR, Piotr et al. (2005). “Behavior recognition via sparse spatio-temporal features”. In : VS-PETS Beijing, China.
- DU, Yong, Yun FU et Liang WANG (2015). “Skeleton based action recognition with convolutional neural network”. In : *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. IEEE, p. 579-583.
- DU, Yong, Wei WANG et Liang WANG (2015). “Hierarchical recurrent neural network for skeleton based action recognition”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 1110-1118.
- DUNCAN, Pamela W et al. (1990). “Functional reach: a new clinical measure of balance”. In : *Journal of gerontology* 45.6, p. M192-M197.

- ESTER, Martin et al. (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise”. In : *Kdd*. T. 96. 34, p. 226-231.
- FLEURY, Anthony et al. (2008). “Sound and speech detection and classification in a health smart home”. In : *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, p. 4644-4647.
- FLYNN, Sheryl, Phyllis PALMA et Anneke BENDER (2007). “Feasibility of using the Sony PlayStation 2 gaming platform for an individual poststroke: a case report”. In : *Journal of neurologic physical therapy* 31.4, p. 180-189.
- FOERSTER, F, M SMEJA et J FAHRENBERG (1999). “Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring”. In : *Computers in Human Behavior* 15.5, p. 571-583.
- FOUBERT, Nicholas et al. (2012). “Lying and sitting posture recognition and transition detection using a pressure sensor array”. In : *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*. IEEE, p. 1-6.
- FUGL-MEYER, Axel R et al. (1975). “The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance”. In : *Scandinavian journal of rehabilitation medicine* 7.1, p. 13-31.
- GARCIA-CEJA, Enrique et al. (2014). “Long-term activity recognition from wrist-watch accelerometer data”. In : *Sensors* 14.12, p. 22500-22524.
- GHORBEL, Enjie et al. (2016). “A fast and accurate motion descriptor for human action recognition applications”. In : *Pattern Recognition (ICPR), 2016 23rd International Conference on*. IEEE, p. 919-924.
- GOETZ, Christopher G et al. (2004). “Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations the Movement Disorder Society Task Force on rating scales for Parkinson’s disease”. In : *Movement disorders* 19.9, p. 1020-1028.
- GOWAYYED, Mohammad Abdelaziz et al. (2013). “Histogram of oriented displacements (HOD): Describing trajectories of human joints for action recognition”. In : *Twenty-Third International Joint Conference on Artificial Intelligence*.

- GROUP, Whoqol et al. (1998). “Development of the World Health Organization WHOQOL-BREF quality of life assessment”. In : *Psychological medicine* 28.3, p. 551-558.
- GU, Fuqiang et al. (2015). “User-independent motion state recognition using smart-phone sensors”. In : *Sensors* 15.12, p. 30636-30652.
- GUPTA, Piyush et Tim DALLAS (2014). “Feature selection and activity recognition system using a single triaxial accelerometer”. In : *IEEE Transactions on Biomedical Engineering* 61.6, p. 1780-1786.
- GURALNIK, Jack M et al. (1995). “Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability”. In : *New England Journal of Medicine* 332.9, p. 556-562.
- HA, Sojeong et Seungjin CHOI (2016). “Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors”. In : *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, p. 381-388.
- HAMMOUCHE, Mounir, Sebastien AMBELLOUIS et Anthony FLEURY (2015). *URIA Kinect Dataset*. Online:
<http://ia.ur.mines-douai.fr/en/datasets/>.
- HASKELL, William L et al. (2007). “Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association”. In : *Circulation* 116.9, p. 1081.
- HASSAN, Mohammed Mehedi et al. (2018). “A robust human activity recognition system using smartphone sensors and deep learning”. In : *Future Generation Computer Systems* 81, p. 307-313.
- HE, Zhenyu et Lianwen JIN (2009). “Activity recognition from acceleration data based on discrete cosine transform and SVM”. In : *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*. IEEE, p. 5041-5044.
- HELAL, Sumi et al. (2003). “Enabling location-aware pervasive computing applications for the elderly”. In : *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003)*. IEEE, p. 531-536.

- HSU, Chih-Wei et Chih-Jen LIN (2002). “A comparison of methods for multiclass support vector machines”. In : *IEEE transactions on Neural Networks* 13.2, p. 415-425.
- HU, Chunyu et al. (2018). “A novel random forests based class incremental learning method for activity recognition”. In : *Pattern Recognition* 78, p. 277-290.
- HU, Frank B et al. (2003). “Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women”. In : *Jama* 289.14, p. 1785-1791.
- HU, Lisha et al. (2013). “A nonintrusive and single-point infrastructure-mediated sensing approach for water-use activity recognition”. In : *2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing*. IEEE, p. 2120-2126.
- HUR, Taeho et al. (2018). “Iss2Image: A Novel Signal-Encoding Technique for CNN-Based Human Activity Recognition”. In : *Sensors* 18.11, p. 3910.
- IGNATOV, Andrey (2018). “Real-time human activity recognition from accelerometer data using Convolutional Neural Networks”. In : *Applied Soft Computing* 62, p. 915-922.
- INCEL, Ozlem Durmaz (2015). “Analysis of movement, orientation and rotation-based sensing for phone placement recognition”. In : *Sensors* 15.10, p. 25474-25506.
- JAIN, Anil et Douglas ZONGKER (1997). “Feature selection: Evaluation, application, and small sample performance”. In : *IEEE transactions on pattern analysis and machine intelligence* 19.2, p. 153-158.
- JALAL, Ahmad, Shaharyar KAMAL et Daijin KIM (2014). “A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments”. In : *Sensors* 14.7, p. 11735-11759.
- JANIDARMIAN, Majid et al. (2017). “A comprehensive analysis on wearable acceleration sensors in human activity recognition”. In : *Sensors* 17.3, p. 529.
- JANSSEN, Wim GM, Hans BJ BUSSMANN et Henk J STAM (2002). “Determinants of the sit-to-stand movement: a review”. In : *Physical therapy* 82.9, p. 866-879.

- JIANG, Wenchao et Zhaozheng YIN (2015). “Human activity recognition using wearable sensors by deep convolutional neural networks”. In : *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, p. 1307-1310.
- JORDAO, Artur, Leonardo Antônio Borges TORRES et William Robson SCHWARTZ (2018). “Novel approaches to human activity recognition based on accelerometer data”. In : *Signal, Image and Video Processing*, p. 1-8.
- JORGENSEN, Martin G et al. (2012). “Efficacy of Nintendo Wii training on mechanical leg muscle function and postural balance in community-dwelling older adults: a randomized controlled trial”. In : *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 68.7, p. 845-852.
- JUNG, Younbo et al. (2009). “Games for a better life: effects of playing Wii games on the well-being of seniors in a long-term care facility”. In : *Proceedings of the Sixth Australasian Conference on Interactive Entertainment*. ACM, p. 5.
- KALGAONKAR, Kaustubh et Bhiksha RAJ (2009). “One-handed gesture recognition using ultrasonic Doppler sonar”. In : *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, p. 1889-1892.
- KARANTONIS, Dean M et al. (2006). “Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring”. In : *IEEE transactions on information technology in biomedicine* 10.1, p. 156-167.
- KATZ, Sidney et al. (1970). “Progress in development of the index of ADL”. In : *The gerontologist* 10.1_Part_1, p. 20-30.
- KATZMARZYK, Peter T et al. (2009). “Sitting time and mortality from all causes, cardiovascular disease, and cancer”. In : *Medicine & Science in Sports & Exercise* 41.5, p. 998-1005.
- KEOGH, Justin WL et al. (2014). “Physical and psychosocial function in residential aged-care elders: effect of Nintendo Wii Sports games”. In : *Journal of aging and physical activity* 22.2, p. 235-244.
- KHAN, Adil Mehmood et al. (2010). “A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer”. In : *IEEE transactions on information technology in biomedicine* 14.5, p. 1166-1172.

- KIM, Jungjin et al. (2013). “Unsupervised virtual reality-based exercise program improves hip muscle strength and balance control in older adults: a pilot study”. In : *Archives of physical medicine and rehabilitation* 94.5, p. 937-943.
- KIM, Kyu-Jin et al. (2009). “Dementia wandering detection and activity recognition algorithm using tri-axial accelerometer sensors”. In : *Proceedings of the 4th International Conference on Ubiquitous Information Technologies & Applications*. IEEE, p. 1-5.
- KIM, Yong-Joong, Bong-Nam KANG et Daijin KIM (2015). “Hidden markov model ensemble for activity recognition using tri-axis accelerometer”. In : *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, p. 3036-3041.
- KJAERGAARD, Mikkel (2010). “Location-based services on mobile phones: minimizing power consumption”. In : *IEEE Pervasive Computing* 11.1, p. 67-73.
- KNAPP-CORDES, Martin et Bill MCKEEMAN (2011a). “Improvements to tic and toc functions for measuring absolute elapsed time performance in MATLAB”. In : *Matlab Digest. mathworks. com*.
- (2011b). “Improvements to tic and toc functions for measuring absolute elapsed time performance in MATLAB”. In : *Matlab Technical Articles and Newsletters*. The MathWorks Inc.
- KOLOSNAJI, Bojan et Claudia ECKERT (2015). “Neural network-based user-independent physical activity recognition for mobile devices”. In : *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, p. 378-386.
- KOWAL, Paul et J Edward DOWD (2001). “Definition of an older person. Proposed working definition of an older person in Africa for the MDS Project”. In : *World Health Organization, Geneva, doi 10.2.1*, p. 5188-9286.
- KRALJ, Alojz, Robert J JAEGER et Marko MUNIH (1990). “Analysis of standing up and sitting down in humans: definitions and normative data presentation”. In : *Journal of biomechanics* 23.11, p. 1123-1138.
- KUDO, Mineichi et Jack SKLANSKY (2000). “Comparison of algorithms that select features for pattern classifiers”. In : *Pattern recognition* 33.1, p. 25-41.

- KWAPISZ, Jennifer R, Gary M WEISS et Samuel A MOORE (2011). “Activity recognition using cell phone accelerometers”. In : *ACM SigKDD Explorations Newsletter* 12.2, p. 74-82.
- LACROIX, Andrea Z et al. (1996). “Does walking decrease the risk of cardiovascular disease hospitalizations and death in older adults?” In : *Journal of the American Geriatrics Society* 44.2, p. 113-120.
- LAPTEV, Ivan (2005). “On space-time interest points”. In : *International journal of computer vision* 64.2-3, p. 107-123.
- LAPTEV, Ivan et al. (2008). “Learning realistic human actions from movies”. In :
- LARBI, Khaled et Delphine ROY (2019). “4 millions de seniors seraient en perte d’autonomie en 2050”. In :
- LAWTON, M Powell et Elaine M BRODY (1969). “Assessment of older people: self-maintaining and instrumental activities of daily living”. In : *The gerontologist* 9.3_Part_1, p. 179-186.
- LEE, Kangjae et Mei-Po KWAN (2018). “Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results”. In : *Computers, Environment and Urban Systems* 67, p. 124-131.
- LEE, Sunwoo et Sungrae SHIN (2013). “Effectiveness of virtual reality using video gaming technology in elderly adults with diabetes mellitus”. In : *Diabetes technology & therapeutics* 15.6, p. 489-496.
- LEE, Young-Seol et Sung-Bae CHO (2011). “Activity recognition using hierarchical hidden markov models on a smartphone with 3D accelerometer”. In : *International Conference on Hybrid Artificial Intelligence Systems*. Springer, p. 460-467.
- LEUTHEUSER, Heike, Dominik SCHULDHAUS et Bjoern M ESKOFIER (2013). “Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset”. In : *PloS one* 8.10, e75196.
- LI, Chuankun et al. (2017). “Joint distance maps based action recognition with convolutional neural networks”. In : *IEEE Signal Processing Letters* 24.5, p. 624-628.

- LI, Wanqing, Zhengyou ZHANG et Zicheng LIU (2010). “Action recognition based on a bag of 3D points”. In : *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, p. 9-14.
- LIANG, Yunji et al. (2014). “Energy-efficient motion related activity recognition on mobile devices for pervasive healthcare”. In : *Mobile Networks and Applications* 19.3, p. 303-317.
- LIM, Joon-Ho et al. (2008). “Daily activity recognition system for the elderly using pressure sensors”. In : *2008 30th annual international conference of the IEEE engineering in medicine and biology society*. IEEE, p. 5188-5191.
- LIN, Che-Wei et al. (2012). “A wearable sensor module with a neural-network-based activity classification algorithm for daily energy expenditure estimation”. In : *IEEE Transactions on Information Technology in Biomedicine* 16.5, p. 991-998.
- LIU, Li, Yuxin PENG et al. (2016). “Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors”. In : *Information Sciences* 340, p. 41-57.
- LIU, Li et Ling SHAO (2013). “Learning discriminative representations from RGB-D video data”. In : *Twenty-Third International Joint Conference on Artificial Intelligence*.
- LIU, Mengyuan et Hong LIU (2016). “Depth Context: A new descriptor for human activity recognition by using sole depth sequences”. In : *Neurocomputing* 175, p. 747-758.
- LIU, Shaopeng et al. (2012). “Multisensor data fusion for physical activity assessment”. In : *IEEE Transactions on Biomedical Engineering* 59.3, p. 687-696.
- LIU, Zhi, Chenyang ZHANG et Yingli TIAN (2016). “3d-based deep convolutional neural network for action recognition with depth sequences”. In : *Image and Vision Computing* 55, p. 93-100.
- LONG, Xi, Bin YIN et Ronald M AARTS (2009). “Single-accelerometer-based daily physical activity classification”. In : *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, p. 6107-6110.

- LORD, Stephen R et al. (2002). “Sit-to-stand performance depends on sensation, speed, balance, and psychological status in addition to strength in older people”. In : *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 57.8, p. M539-M543.
- LU, Hong et al. (2009). “SoundSense: scalable sound sensing for people-centric applications on mobile phones”. In : *Proceedings of the 7th international conference on Mobile systems, applications, and services*. ACM, p. 165-178.
- MA, Shugao, Leonid SIGAL et Stan SCLAROFF (2016). “Learning activity progression in lstms for activity detection and early detection”. In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 1942-1950.
- MACHADO, Inês P et al. (2015). “Human activity data discovery from triaxial accelerometer sensor: Non-supervised learning sensitivity to feature extraction parametrization”. In : *Information Processing & Management* 51.2, p. 204-214.
- MAILLOT, Pauline, Alexandra PERROT et Alan HARTLEY (2012). “Effects of interactive physical-activity video-game training on physical and cognitive function in older adults”. In : *Psychology and aging* 27.3, p. 589.
- MANNINI, Andrea, Stephen S INTILLE et al. (2013). “Activity recognition using a single accelerometer placed at the wrist or ankle”. In : *Medicine and science in sports and exercise* 45.11, p. 2193.
- MANNINI, Andrea, Mary ROSENBERGER et al. (2017). “Activity Recognition in Youth Using Single Accelerometer Placed at Wrist or Ankle”. In : *Medicine and science in sports and exercise* 49.4, p. 801-812.
- MANNINI, Andrea et Angelo Maria SABATINI (2010). “Machine learning methods for classifying human physical activity from on-body accelerometers”. In : *Sensors* 10.2, p. 1154-1175.
- MANSON, JoAnn E et al. (1999). “A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women”. In : *New England Journal of Medicine* 341.9, p. 650-658.
- MARTIAL, Françoise et al. (2019). *Tableaux de l'économie française*. Institut national de la statistique et des études économiques (INSEE).

- MARTÍN, Henar et al. (2013). “Activity logging using lightweight classification techniques in mobile devices”. In : *Personal and ubiquitous computing* 17.4, p. 675-695.
- MATHIE, MJ et al. (2004). “Classification of basic daily movements using a triaxial accelerometer”. In : *Medical and Biological Engineering and Computing* 42.5, p. 679-687.
- MCNICOLL, Geoffrey (2002). “World Population Ageing 1950-2050”. In : *Population and Development Review* 28.4, p. 814-816.
- MEHMOOD, Amir et al. (2016). “Study of multi-classification of advanced daily life activities on shimmer sensor dataset”. In : *International Journal of Communication Networks and Information Security* 8.2, p. 86.
- MIAO, Fen et al. (2015). “Identifying typical physical activity on smartphone with varying positions and orientations”. In : *Biomedical engineering online* 14.1, p. 32.
- MICUCCI, Daniela, Marco MOBILIO et Paolo NAPOLETANO (2017). “Unimib shar: A dataset for human activity recognition using acceleration data from smartphones”. In : *Applied Sciences* 7.10, p. 1101.
- MIRANDA, Leandro et al. (2014). “Online gesture recognition from pose kernel learning and decision forests”. In : *Pattern Recognition Letters* 39, p. 65-73.
- MITCHELL, Edmond, David MONAGHAN et Noel E O’CONNOR (2013). “Classification of sporting activities using smartphone accelerometers”. In : *Sensors* 13.4, p. 5317-5337.
- MOESLUND, Thomas B, Adrian HILTON et Volker KRÜGER (2006). “A survey of advances in vision-based human motion capture and analysis”. In : *Computer vision and image understanding* 104.2-3, p. 90-126.
- MOHAMMAD, Yasser, Kazunori MATSUMOTO et Keiichiro HOASHI (2018). “Primitive activity recognition from short sequences of sensory data”. In : *Applied Intelligence*, p. 1-14.
- MURAD, Abdulmajid et Jae-Young PYUN (2017). “Deep recurrent neural networks for human activity recognition”. In : *Sensors* 17.11, p. 2556.
- NAGEL, H-H (1988). “From image sequences towards conceptual descriptions”. In : *Image and vision computing* 6.2, p. 59-74.

- NAJAFI, Bijan et al. (2002). "Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly". In : *IEEE Transactions on biomedical Engineering* 49.8, p. 843-851.
- NAM, Yunyoung et Jung Wook PARK (2013). "Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor". In : *IEEE journal of biomedical and health informatics* 17.2, p. 420-426.
- NI, Qin et al. (2016). "Dynamic detection of window starting positions and its implementation within an activity recognition framework". In : *Journal of biomedical informatics* 62, p. 171-180.
- NOOR, Mohd Halim Mohd et al. (2017). "Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer". In : *Pervasive and Mobile Computing* 38, p. 41-59.
- O'BRIEN, Megan K et al. (2017). "Activity recognition for persons with stroke using mobile phone technology: toward improved performance in a home setting". In : *Journal of medical Internet research* 19.5, e184.
- O'LEARY, Kevin C et al. (2011). "The effects of single bouts of aerobic exercise, exergaming, and videogame play on cognitive control". In : *Clinical Neurophysiology* 122.8, p. 1518-1525.
- OHN-BAR, Eshed et Mohan Manubhai TRIVEDI (2013). "Joint angles similarities and HOG2 for action recognition". In : *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on. IEEE*, p. 465-470.
- OREIFEJ, Omar et Zicheng LIU (2013). "Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences". In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 716-723.
- OUCHI, Kazushige et Miwako DOI (2013). "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. ACM*, p. 103-106.
- PEDREGOSA, Fabian et al. (2011). "Scikit-learn: Machine learning in Python". In : *Journal of machine learning research* 12.Oct, p. 2825-2830.

- PEI, Ling et al. (2013). “Human behavior cognition using smartphone sensors”. In : *Sensors* 13.2, p. 1402-1424.
- PERRIN-HAYNES, Jacqueline et al. (2011). *Les personnes âgées en institution*. La Direction de la recherche, des études, de l'évaluation et des statistiques (Drees).
- PHAM, Viet Thang et al. (2007). “Application of ultrasonic sensors in a smart environment”. In : *Pervasive and Mobile computing* 3.2, p. 180-207.
- POMPEU, JE et al. (2014). “Feasibility, safety and outcomes of playing Kinect Adventures!TM for people with Parkinson’s disease: a pilot study”. In : *Physiotherapy* 100.2, p. 162-168.
- POPESCU, Mihail et al. (2008). “An acoustic fall detector system that uses sound height information to reduce the false alarm rate”. In : *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, p. 4628-4631.
- POPPE, Ronald (2010). “A survey on vision-based human action recognition”. In : *Image and vision computing* 28.6, p. 976-990.
- PRIMACK, Brian A et al. (2012). “Role of video games in improving health-related outcomes: a systematic review”. In : *American journal of preventive medicine* 42.6, p. 630-638.
- PUDIL, Pavel, Jana NOVOVIČOVÁ et Josef KITTLER (1994). “Floating search methods in feature selection”. In : *Pattern recognition letters* 15.11, p. 1119-1125.
- RABINER, Lawrence et B JUANG (1986). “An introduction to hidden Markov models”. In : *ieee assp magazine* 3.1, p. 4-16.
- RAVI, Daniele et al. (2017). “A deep learning approach to on-node sensor data analytics for mobile or wearable devices”. In : *IEEE journal of biomedical and health informatics* 21.1, p. 56-64.
- RAVI, Nishkam et al. (2005). “Activity recognition from accelerometer data”. In :
- REISS, Attila et Didier STRICKER (2012). “Introducing a new benchmarked dataset for activity monitoring”. In : *Wearable Computers (ISWC), 2012 16th International Symposium on*. IEEE, p. 108-109.

- RENDON, Abel Angel et al. (2012). “The effect of virtual reality gaming on dynamic balance in older adults”. In : *Age and ageing* 41.4, p. 549-552.
- REYES-ORTIZ, Jorge-L et al. (2016). “Transition-aware human activity recognition using smartphones”. In : *Neurocomputing* 171, p. 754-767.
- RIKLI, Roberta E et C Jessie JONES (1999). “Development and validation of a functional fitness test for community-residing older adults”. In : *Journal of aging and physical activity* 7.2, p. 129-161.
- RILEY, Patrick O et al. (1991). “Mechanics of a constrained chair-rise”. In : *Journal of biomechanics* 24.1, p. 77-85.
- RONAO, Charissa Ann et Sung-Bae CHO (2016). “Human activity recognition with smartphone sensors using deep learning neural networks”. In : *Expert Systems with Applications* 59, p. 235-244.
- (2017). “Recognizing human activities from smartphone sensors using hierarchical continuous hidden Markov models”. In : *International Journal of Distributed Sensor Networks* 13.1, p. 1550147716683687.
- ROSENBERG, Morris (2015). *Society and the adolescent self-image*. Princeton university press.
- RUSSELL, Daniel W (1996). “UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure”. In : *Journal of personality assessment* 66.1, p. 20-40.
- RYOO, Michael S (2011). “Human activity prediction: Early recognition of ongoing activities from streaming videos”. In : *2011 International Conference on Computer Vision*. IEEE, p. 1036-1043.
- RYOO, Michael S et al. (2017). “Privacy-preserving human activity recognition from extreme low resolution”. In : *Thirty-First AAAI Conference on Artificial Intelligence*.
- SADEGH ALIAKBARIAN, Mohammad et al. (2017). “Encouraging lstms to anticipate actions very early”. In : *Proceedings of the IEEE International Conference on Computer Vision*, p. 280-289.

- SAEZ, Yago, Alejandro BALDOMINOS et Pedro ISASI (2016). “A comparison study of classifier algorithms for cross-person physical activity recognition”. In : *Sensors* 17.1, p. 66.
- SAGUNA, Saguna, Arkady ZASLAVSKY et Dipanjan CHAKRABORTY (2013). “Complex activity recognition using context-driven activity theory and activity signatures”. In : *ACM Transactions on Computer-Human Interaction (TOCHI)* 20.6, p. 32.
- SAID, Ahmed Ben et al. (2018). “Deep learning and low rank dictionary model for mHealth data classification”. In : *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, p. 358-363.
- SAN BUENAVENTURA, Charlene V, Nestor Michael C TIGLAO et Rowel O ATIENZA (2018). “Deep Learning for Smartphone-Based Human Activity Recognition Using Multi-sensor Fusion”. In : *International Wireless Internet Conference*. Springer, p. 65-75.
- SAN-SEGUNDO, Rubén, Julián David ECHEVERRY-CORREA et al. (2016). “Human activity monitoring based on hidden Markov models using a smartphone”. In : *IEEE Instrumentation & Measurement Magazine* 19.6, p. 27-31.
- SAN-SEGUNDO, Rubén, Juan Manuel MONTERO et al. (2016). “Feature extraction from smartphone inertial signals for human activity segmentation”. In : *Signal Processing* 120, p. 359-372.
- SCHEURER, Sebastian et al. (2017). “Human activity recognition for emergency first responders via body-worn inertial sensors”. In : *Wearable and Implantable Body Sensor Networks (BSN), 2017 IEEE 14th International Conference on*. IEEE, p. 5-8.
- SEERA, Manjeevan, Chu Kiong LOO et Chee Peng LIM (2014). “A hybrid FMM-CART model for human activity recognition”. In : *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*. IEEE, p. 182-187.
- SEIDENARI, Lorenzo et al. (2013). “Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses”. In : *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, p. 479-485.

- SHAO, Zhanpeng et Youfu LI (2015). “Integral invariants for space motion trajectory matching and recognition”. In : *Pattern Recognition* 48.8, p. 2418-2432.
- SHEN, Chao, Yufei CHEN et Gengshan YANG (2016). “On motion-sensor behavior analysis for human-activity recognition via smartphones”. In : *Identity, Security and Behavior Analysis (ISBA), 2016 IEEE International Conference on*. IEEE, p. 1-6.
- SHOAIB, Muhammad, Stephan BOSCH, Ozlem Durmaz INCEL et al. (2014). “Fusion of smartphone motion sensors for physical activity recognition”. In : *Sensors* 14.6, p. 10146-10176.
- SHOAIB, Muhammad, Stephan BOSCH, Ozlem INCEL et al. (2016). “Complex human activity recognition using smartphone and wrist-worn motion sensors”. In : *Sensors* 16.4, p. 426.
- SHOTTON, Jamie et al. (2013). “Real-time human pose recognition in parts from single depth images”. In : *Communications of the ACM* 56.1, p. 116-124.
- SIIRTOLA, Pekka et Juha RÖNING (2012a). “Recognizing human activities user-independently on smartphones based on accelerometer data”. In : *IJIMAI* 1.5, p. 38-45.
- (2012b). “Recognizing human activities user-independently on smartphones based on accelerometer data”. In : *IJIMAI* 1.5, p. 38-45.
- SLAMA, Rim et al. (2015). “Accurate 3D action recognition using learning on the Grassmann manifold”. In : *Pattern Recognition* 48.2, p. 556-567.
- SUAREZ, Isabel et al. (2015a). “Improved activity recognition by using enriched acceleration data”. In : *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, p. 1011-1015.
- (2015b). “Improved activity recognition by using enriched acceleration data”. In : *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, p. 1011-1015.
- SUN, Lin et al. (2010). “Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations”. In : *International conference on ubiquitous intelligence and computing*. Springer, p. 548-562.

- SZTYLER, Timo, Heiner STUCKENSCHMIDT et Wolfgang PETRICH (2017). “Position-aware activity recognition with wearable devices”. In : *Pervasive and mobile computing* 38, p. 281-295.
- TALHA, Sid Ahmed Walid, Anthony FLEURY et Stéphane LECOEUCE (2019). “Hierarchical classification scheme for real-time recognition of physical activities and postural transitions using smartphone inertial sensors”. In : *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, p. 1243-1246.
- TAPIA, Emmanuel Munguia et al. (2007). “Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor”. In : *Wearable Computers, 2007 11th IEEE International Symposium on*. IEEE, p. 37-40.
- TOPOLSKI, Tari D et al. (2006). “Peer reviewed: the Rapid Assessment of Physical Activity (RAPA) among older adults”. In : *Preventing chronic disease* 3.4.
- VAVOULAS, George et al. (2016). “The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones”. In : *ICT4AgeingWell*, p. 143-151.
- VEMULAPALLI, Raviteja, Felipe ARRATE et Rama CHELLAPPA (2014). “Human action recognition by representing 3d skeletons as points in a lie group”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 588-595.
- VENKATARAMAN, Vinay et al. (2013). “Attractor-shape for dynamical analysis of human movement: Applications in stroke rehabilitation and action recognition”. In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, p. 514-520.
- VIRONE, Gilles et al. (2008). “Behavioral patterns of older adults in assisted living”. In : *IEEE Transactions on Information Technology in Biomedicine* 12.3, p. 387-398.
- WANG, Aiguo et al. (2016). “A comparative study on human activity recognition using inertial sensors in a smartphone”. In : *IEEE Sensors Journal* 16.11, p. 4566-4578.

- WANG, Gaojing et al. (2018). “Impact of Sliding Window Length in Indoor Human Motion Modes and Pose Pattern Recognition Based on Smartphone Sensors”. In : *Sensors* 18.6, p. 1965.
- WANG, Jihe et al. (2015). “Low-power sensor polling for context-aware services on smartphones”. In : *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*. IEEE, p. 617-622.
- WANG, Jindong et al. (2019). “Deep learning for sensor-based activity recognition: A survey”. In : *Pattern Recognition Letters* 119, p. 3-11.
- WANG, Ning et al. (2007). “Accelerometry based classification of walking patterns using time-frequency analysis”. In : *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, p. 4899-4902.
- WANG, Pichao, Wanqing LI, Zhimin GAO, Chang TANG et Philip O OGUNBONA (2018). “Depth pooling based large-scale 3-d action recognition with convolutional neural networks”. In : *IEEE Transactions on Multimedia* 20.5, p. 1051-1061.
- WANG, Pichao, Wanqing LI, Zhimin GAO, Chang TANG, Jing ZHANG et al. (2015). “Convnets-based action recognition from depth maps through virtual cameras and pseudocoloring”. In : *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, p. 1119-1122.
- WELHENGE, Anuradhi Malshika et Attaphongse TAPARUGSSANAGORN (2018). “Human activity classification using long short-term memory network”. In : *Signal, Image and Video Processing*, p. 1-6.
- WILMOT, Emma G et al. (2012). *Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis*.
- WITTEN, Ian H et al. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- WU, Wanmin et al. (2012). “Classification accuracies of physical activities using smartphone motion sensors”. In : *Journal of medical Internet research* 14.5.

- XIA, Lu et JK AGGARWAL (2013). “Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 2834-2841.
- XIA, Lu, Chia-Chih CHEN et JK AGGARWAL (2012). “View invariant human action recognition using histograms of 3d joints”. In : *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, p. 20-27.
- XSENS (2010). “Technical Documentation”. In : *Xsens Technologies 27*.
- YAMAGUCHI, Akifumi et al. (1998). “Monitoring behavior in the home using positioning sensors”. In : *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*. T. 4. IEEE, p. 1977-1979.
- YANG, Jhun-Ying, Jeen-Shing WANG et Yen-Ping CHEN (2008). “Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers”. In : *Pattern recognition letters* 29.16, p. 2213-2220.
- YANG, Jianbo et al. (2015). “Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition”. In : *Ijcai*. T. 15, p. 3995-4001.
- YANG, Rong et Baowei WANG (2016). “PACP: A Position-Independent Activity Recognition Method Using Smartphone Sensors”. In : *Information* 7.4, p. 72.
- YANG, Rui et Ruoyu YANG (2014). “DMM-pyramid based deep architectures for action recognition with depth cameras”. In : *Asian Conference on Computer Vision*. Springer, p. 37-49.
- YANG, Xiaodong et YingLi TIAN (2014). “Super normal vector for activity recognition using depth sequences”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 804-811.
- YANG, Xiaodong, Chenyang ZHANG et YingLi TIAN (2012). “Recognizing actions using depth motion maps-based histograms of oriented gradients”. In : *Proceedings of the 20th ACM international conference on Multimedia*. ACM, p. 1057-1060.

- YASUNAGA, Akitomo et al. (2007). “Development and evaluation of the physical activity questionnaire for elderly Japanese: the Nakanojo study”. In : *Journal of aging and physical activity* 15.4, p. 398-411.
- YOUNG, William et al. (2011). “Assessing and training standing balance in older adults: a novel approach using the ‘Nintendo Wii’Balance Board”. In : *Gait & posture* 33.2, p. 303-305.
- ZANFIR, Mihai, Marius LEORDEANU et Cristian SMINCHISESCU (2013). “The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection”. In : *Proceedings of the IEEE International Conference on Computer Vision*, p. 2752-2759.
- ZDRAVEVSKI, Eftim et al. (2017). “Improving Activity Recognition Accuracy in Ambient-Assisted Living Systems by Automated Feature Engineering”. In : *IEEE Access* 5, p. 5262-5280.
- ZHANG, Chenyang et Yingli TIAN (2012). “RGB-D camera-based daily living activity recognition”. In : *Journal of computer vision and image processing* 2.4, p. 12.
- ZHANG, Mi et Alexander A SAWCHUK (2012). “USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors”. In : *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, p. 1036-1043.
- ZHANG, Shaoyan et al. (2012). “Physical activity classification using the GENE A wrist-worn accelerometer”. In :
- ZHAO, Yang et al. (2012). “Combing rgb and depth map features for human activity recognition”. In : *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, p. 1-4.
- ZHAO, Zhongtang et al. (2011). “Cross-people mobile-phone based activity recognition”. In : *IJCAI*. T. 11. 3, p. 2545-2250.
- ZHENG, Zengwei et al. (2018). “TASG: An Augmented Classification Method for Impersonal HAR”. In : *Mobile Information Systems* 2018.
- ZHU, Wentao et al. (2016). “Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks”. In : *Thirtieth AAAI Conference on Artificial Intelligence*.

- ZHU, Yu, Wenbin CHEN et Guodong GUO (2013). “Fusing spatiotemporal features and joints for 3d action recognition”. In : *Computer vision and pattern recognition workshops (CVPRW), 2013 IEEE conference on*. IEEE, p. 486-491.
- ZIGEL, Yaniv, Dima LITVAK et Israel GANNOT (2009). “A method for automatic fall detection of elderly people using floor vibrations and sound—Proof of concept on human mimicking doll falls”. In : *IEEE Transactions on Biomedical Engineering* 56.12, p. 2858-2867.