

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

NGUYỄN NGỌC ĐIỆP

**NGHIÊN CỨU PHƯƠNG PHÁP HỌC MÁY
CHO NHẬN DẠNG HOẠT ĐỘNG
SỬ DỤNG CẢM BIẾN MANG TRÊN NGƯỜI**

LUẬN ÁN TIẾN SĨ KỸ THUẬT

HÀ NỘI – 2016

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

NGUYỄN NGỌC ĐIỆP

**NGHIÊN CỨU PHƯƠNG PHÁP HỌC MÁY
CHO NHẬN DẠNG HOẠT ĐỘNG
SỬ DỤNG CẢM BIẾN MANG TRÊN NGƯỜI**

**CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN
MÃ SỐ: 62.48.01.04**

LUẬN ÁN TIẾN SĨ KỸ THUẬT

NGƯỜI HƯỚNG DẪN KHOA HỌC:

- 1. PGS.TS. TỪ MINH PHƯƠNG**
- 2. TS. PHẠM VĂN CƯỜNG**

HÀ NỘI – 2016

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi. Các kết quả được viết chung với các tác giả khác đều được sự đồng ý của đồng tác giả trước khi đưa vào luận án. Các kết quả nêu trong luận án là trung thực và chưa từng được công bố trong các công trình nào khác.

Tác giả

Nguyễn Ngọc Điệp

LỜI CẢM ƠN

Thực hiện luận án tiến sĩ là một thử thách lớn, đòi hỏi sự kiên trì và tập trung cao độ. Tôi thực sự hạnh phúc với kết quả đạt được trong đề tài nghiên cứu của mình. Những kết quả đạt được không chỉ là nỗ lực cá nhân, mà còn có sự hỗ trợ và giúp đỡ của các thầy hướng dẫn, nhà trường, bộ môn, đồng nghiệp và gia đình. Tôi muốn bày tỏ tình cảm của mình đến với họ.

Trước hết, tôi xin gửi lời cảm ơn sâu sắc tới thầy PGS.TS. Từ Minh Phương và thầy TS. Phạm Văn Cường đã quan tâm hướng dẫn và giúp đỡ tôi trong suốt quá trình thực hiện và hoàn thành luận án.

Tôi xin trân trọng cảm ơn Khoa Công nghệ thông tin 1, Khoa Quốc tế và Đào tạo Sau Đại học và Lãnh đạo Học viện Công nghệ Bưu chính Viễn thông đã tạo điều kiện thuận lợi cho tôi trong suốt quá trình thực hiện luận án. Tôi xin cảm ơn tập thể cán bộ, giảng viên Khoa Công nghệ thông tin 1 – Học Viện Công nghệ Bưu chính Viễn thông đã cổ vũ động viên tôi trong quá trình nghiên cứu.

Tôi xin gửi lời cảm ơn chân thành tới tất cả những người bạn của tôi, những người luôn chia sẻ và cổ vũ tôi trong những lúc khó khăn. Cuối cùng, tôi xin bày tỏ lòng biết ơn vô hạn đối với cha mẹ và gia đình đã luôn bên cạnh ủng hộ, giúp đỡ tôi.

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN.....	ii
MỤC LỤC	iii
DANH MỤC CÁC BẢNG	vii
DANH MỤC CÁC HÌNH VẼ	viii
DANH MỤC CÁC TỪ VIẾT TẮT	x
PHẦN MỞ ĐẦU	1
1. GIỚI THIỆU	1
2. TÍNH CẤP THIẾT CỦA LUẬN ÁN	2
3. MỤC TIÊU CỦA LUẬN ÁN	5
4. CÁC ĐÓNG GÓP CỦA LUẬN ÁN	6
5. BỐ CỤC CỦA LUẬN ÁN	7
CHƯƠNG 1 TỔNG QUAN VỀ NHẬN DẠNG HOẠT ĐỘNG SỬ DỤNG CẢM BIẾN MANG TRÊN NGƯỜI	9
1.1. GIỚI THIỆU CHUNG VỀ NHẬN DẠNG HOẠT ĐỘNG SỬ DỤNG CẢM BIẾN MANG TRÊN NGƯỜI.....	9
1.1.1. <i>Khái niệm</i>	9
1.1.2. <i>Cách tiếp cận</i>	10
1.1.3. <i>Các ứng dụng</i>	13
1.1.4. <i>Một số khó khăn trong nghiên cứu</i>	14
1.2. CÁC CẢM BIẾN MANG TRÊN NGƯỜI.....	15
1.3. KIẾN TRÚC HỆ THỐNG NHẬN DẠNG HOẠT ĐỘNG NGƯỜI.....	17
1.4. CÁC PHƯƠNG PHÁP TRÍCH XUẤT ĐẶC TRƯNG	21

1.4.1. Trích xuất đặc trưng theo tri thức chuyên gia	22
1.4.2. Trích xuất đặc trưng tự động bằng học đặc trưng.....	30
1.5. KẾT LUẬN CHƯƠNG	36
CHƯƠNG 2 NHẬN DẠNG HOẠT ĐỘNG VỚI CÁC ĐẶC TRƯNG HALF	38
2.1. ĐẶT VẤN ĐỀ	39
2.2. THỐNG KÊ VỀ PHÂN PHỐI TẦN SUẤT GÓC CỦA CÁC HOẠT ĐỘNG	42
2.3. CÁC ĐẶC TRƯNG HALF	44
2.3.1. Thuật toán trích xuất đặc trưng.....	45
2.3.2. Biểu diễn đặc trưng đa mức.....	51
2.3.3. Độ phức tạp thuật toán	55
2.4. THỬ NGHIỆM	56
2.4.1. Giới thiệu một số nghiên cứu phát hiện ngã.....	57
2.4.2. Phương pháp phát hiện người ngã	59
2.4.3. Tập dữ liệu thử nghiệm	63
2.4.4. Kết quả thử nghiệm và đánh giá.....	65
2.5. KẾT LUẬN CHƯƠNG	71
CHƯƠNG 3 HỌC CÁC ĐẶC TRƯNG ĐA MỨC MPF.....	73
3.1. NHẬN DẠNG NHIỀU HOẠT ĐỘNG SỬ DỤNG ĐẶC TRƯNG ĐA MỨC.....	73
3.2. CÁC NGHIÊN CỨU CÓ LIÊN QUAN	75
3.3. PHƯƠNG PHÁP HỌC ĐẶC TRƯNG MPF	76
3.3.1. Sơ đồ hoạt động	76
3.3.2. Rừng ngẫu nhiên	78
3.3.3. MPF	81
3.3.4. Đặc trưng cục bộ	83
3.3.5. Bộ phân lớp.....	85

3.3.6. Độ phức tạp thuật toán	86
3.4. TẬP DỮ LIỆU THỬ NGHIỆM	86
3.4.1. Activity Prediction (AP)	87
3.4.2. Opportunity (OP)	87
3.4.3. Skoda (SK)	87
3.5. THỬ NGHIỆM VÀ KẾT QUẢ	88
3.5.1. Phương pháp thử nghiệm	88
3.5.2. Kết quả thử nghiệm và đánh giá	88
3.6. KẾT LUẬN CHƯƠNG	94
CHƯƠNG 4 XÂY DỰNG MỘT SỐ ỨNG DỤNG NHẬN DẠNG HOẠT ĐỘNG	96
4.1. PHÁT HIỆN NGƯỜI NGÃ TRONG THỜI GIAN THỰC	96
4.1.1. Các giải pháp phát hiện người ngã	96
4.1.2. Phần cứng và môi trường thu thập dữ liệu	98
4.1.3. Thuật toán phát hiện ngã	99
4.1.4. Tập dữ liệu thử nghiệm	101
4.1.5. Kết quả thử nghiệm	102
4.1.6. Phần mềm	102
4.2. XÁC THỰC NGƯỜI DÙNG VỚI CHỮ KÝ 3D – SIGVER3D	103
4.2.1. Vấn đề xác thực cho các hệ thống mang trên người	104
4.2.2. Phần cứng và môi trường thu thập dữ liệu	105
4.2.3. Thuật toán xác thực người dùng	107
4.2.4. Thu thập dữ liệu thử nghiệm	109
4.2.5. Kết quả thử nghiệm	110
4.2.6. Phần mềm	110
4.3. KẾT LUẬN CHƯƠNG	113

KẾT LUẬN.....	114
DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ.....	116
TÀI LIỆU THAM KHẢO	117

DANH MỤC CÁC BẢNG

Bảng 1.1. Tóm lược các cảm biến mang trên người sử dụng trong nhận dạng hoạt động ..	17
Bảng 1.2. Các phương pháp trích xuất đặc trưng cho tín hiệu gia tốc.....	24
Bảng 2.1. Thông tin tóm tắt các tập dữ liệu.....	65
Bảng 2.2. So sánh các kết quả phát hiện ngã thử nghiệm trên tập dữ liệu tFall.....	66
Bảng 2.3. Các kết quả phát hiện ngã trên 3 tập dữ liệu đánh giá.....	69
Bảng 2.4. Tốc độ phát hiện ngã trên tập dữ liệu tFall với 1000 mẫu	70
Bảng 3.1. Các đặc trưng cục bộ	84
Bảng 3.2. So sánh độ chính xác trong phân lớp.....	94

DANH MỤC CÁC HÌNH VẼ

Hình 1.1. Kiến trúc hệ thống nhận dạng hoạt động người.....	18
Hình 1.2. Các tín hiệu gia tốc trong các hoạt động khác nhau	23
Hình 1.3. (trái) tính đặc trưng thống kê; (giữa) mô hình PCA; (phải) mô hình DNN (tham khảo [148]).....	34
Hình 2.1. Ví dụ về các phân phối tần suất góc của 3 khung tín hiệu gia tốc cho 3 hoạt động <i>ngồi, đi bộ, chạy</i> như mô tả trong hình 1.2, với số ngăn/cụm (bin) là 9, đều nhau cho khoảng giá trị góc $(-90^0, 90^0)$	41
Hình 2.2. Phân phối xác suất góc với 180 cụm (góc trong khoảng $(-90^0, 90^0)$) cho sự kiện ngã trước, ngã sau và ngã sang bên cạnh.....	43
Hình 2.3. Phân phối xác suất góc với 180 cụm (góc trong khoảng $(-90^0, 90^0)$) cho hoạt động đứng, đi bộ và chạy bộ.....	43
Hình 2.4. Phân phối xác suất góc với 180 cụm (góc trong khoảng $(-90^0, 90^0)$) cho hoạt động nhảy, đi cầu thang và ngồi xuống	44
Hình 2.5. Ví dụ về mẫu tín hiệu ngã (a) và không phải ngã (b)	45
Hình 2.6. Ví dụ về một mảnh (quantum) và hướng của nó	46
Hình 2.7. Một mẫu rời rạc hóa 4 ngăn và véc-tơ đặc trưng phân phối tần suất.....	47
Hình 2.8. Thuật toán trích xuất đặc trưng HALF cho một khung tín hiệu	50
Hình 2.9. Thuật toán tự động xác định các ngăn sử dụng phương pháp phân cụm.....	51
Hình 2.10. Các đặc trưng phân phối tần suất giống nhau (các hình bên phải) được trích xuất từ 2 khung (các hình bên trái) không phân đoạn.....	52
Hình 2.11. Các đặc trưng khác nhau (các hình bên phải) được trích xuất từ 2 khung (các hình bên trái), mỗi khung được chia thành 2 đoạn (bởi đường gạch nối ở giữa)	53
Hình 2.12. Thuật toán biểu diễn đặc trưng đa mức cho một khung tín hiệu	54

Hình 2.13. Bốn pha của sự kiện ngã ([103]): trước khi ngã, tác động, sau khi ngã, hồi phục	59
Hình 2.14. Ảnh hưởng của số ngăn tới độ chính xác hệ thống.....	67
Hình 2.15. Ảnh hưởng của số đoạn trong khung cửa sổ thời gian tới độ chính xác hệ thống	68
Hình 3.1. (a) 3 lát tín hiệu (slice) được phân đoạn từ một khung hoạt động (frame) với độ chồng lấn là 50%, (b) các đặc trưng được trích xuất từ mỗi lát tín hiệu, (c) MPF với vai trò là từ vựng chuyển động, (d) phân phối tần suất của chuyển động gốc (motion primitive histogram)	77
Hình 3.2. Thuật toán rừng ngẫu nhiên cho phân lớp	80
Hình 3.3. Véc-tơ mã (code vector) tạo ra từ một rừng ngẫu nhiên với 3 cây cho một véc-tơ đặc trưng cục bộ. Đường dẫn đến các nút lá kết quả được hiển thị bằng màu vàng.	82
Hình 3.4. Thuật toán trích xuất đặc trưng MPF cho mảng véc-tơ đặc trưng cục bộ đầu vào thuộc một khung hoạt động	83
Hình 3.5. So sánh các thuật toán xây dựng từ vựng trên ba tập dữ liệu	90
Hình 3.6. Ảnh hưởng của các đặc trưng cục bộ tới độ chính xác.....	91
Hình 3.7. Ảnh hưởng của số lượng cây tới độ chính xác	92
Hình 3.8. Ảnh hưởng của kích thước từ vựng tới độ chính xác	93
Hình 4.1. Wii Remote (bên trái nhất), bo mạch Broadcom BCM2042 và vị trí đeo Wii Remote.....	99
Hình 4.2. (a) Giao diện chính của phần mềm phát hiện ngã tự động và (b) thông báo khi phát hiện sự kiện ngã	103
Hình 4.3. Samsung S3 (trái) và cách người dùng cầm điện thoại để ký trong không gian (phải).....	106
Hình 4.4. Các chữ ký trong tập dữ liệu.....	106
Hình 4.5. Các chữ ký gốc (hình phải) và biểu diễn gia tốc tương ứng (hình trái).....	107
Hình 4.6. Các hình ảnh minh họa cho phần mềm “Xác thực bằng chữ ký 3D”	112

DANH MỤC CÁC TỪ VIẾT TẮT

KÝ HIỆU	DIỄN GIẢI	
	TIẾNG ANH	TIẾNG VIỆT
ADL	Activities of Daily Living	Các hoạt động sống hàng ngày
AUC	Area Under the ROC Curve	Diện tích dưới đường cong ROC
BoW	Bag-of-Word	Mô hình túi từ/ tập các từ
CNN	Convolutional Neural Networks	Mạng nơ-ron nhân chập
DNN	Deep Neural Networks	Mạng nơ-ron sâu
ECDF	Empirical Cumulative Distribution Function	Hàm phân phối tích lũy thực nghiệm
EER	Equal Error Rate	Tỷ lệ lỗi cân bằng
FAR	False Acceptance Rate	Tỷ lệ chấp nhận sai
FFT	Fast Fourier Transform	Biến đổi Fourier nhanh
FN	False Negative	Âm tính giả (mẫu mang nhãn dương bị phân lớp sai vào lớp âm)
FP	False Positive	Dương tính giả (mẫu mang nhãn âm bị phân lớp sai vào lớp dương)
FRR	False Rejection Rate	Tỷ lệ từ chối sai
GMM	Gaussian Mixture Model	Mô hình Gauss hỗn hợp
GPS	Global Positioning System	Hệ thống định vị toàn cầu

HALF	Histograms of Angles of Line Fragments between Two Consecutive Points	Phân phối tần suất góc của các đoạn nối hai điểm dữ liệu kế
HAR	Human Activity Recognition	Nhận dạng hoạt động người
HMM	Hidden Markov Model	Mô hình Markov ẩn
k -NN	k Nearest Neighbor	k hàng xóm gần nhất
MCR	Mean Crossing Rate	Tỷ lệ vượt qua trung bình
MFCC	Mel-Frequency Cepstral Coefficients	Các hệ số cepstrum tần số Mel
MPF	Motion Primitive Forests	Rừng chuyển động gốc
PCA	Principal Component Analysis	Phân tích thành phần chính
RBF	Radial Basis Function	Hàm cơ sở bán kính
RBM	Restricted Boltzmann Machine	Máy Boltzmann hạn chế
ROC	Receiver Operator Characteristic	Đặc tính hoạt động của bộ thu nhận
SVM	Support Vector Machines	Máy véc-tơ tựa
TN	True Negative	Âm tính thật (mẫu mang nhãn âm được phân lớp đúng vào lớp âm)
TP	True Positive	Dương tính thật (mẫu mang nhãn dương được phân lớp đúng vào lớp dương)
ZCR	Zero Crossing Rate	Tỷ lệ vượt qua điểm không

PHẦN MỞ ĐẦU

1. GIỚI THIỆU

Nhận dạng hoạt động người (Human Activity Recognition - HAR) là quá trình giám sát và phân tích hành vi người dùng và trạng thái môi trường xung quanh nhằm suy diễn/nhận dạng các hoạt động đang xảy ra [33]. Nhận dạng hoạt động người không những là chủ đề nghiên cứu quan trọng trong tính toán nhận biết ngữ cảnh mà còn là chủ đề cho rất nhiều lĩnh vực khác như tính toán khắp nơi, tương tác người-máy hay tính toán di động. Một trong những mục tiêu của nhận dạng hoạt động là cung cấp thông tin về hành vi của người dùng, từ đó cho phép hệ thống tính toán chủ động hỗ trợ người dùng trong công việc [4].

Nhận dạng hoạt động sử dụng cảm biến mang trên người ứng dụng công nghệ cảm biến có thể mang theo người để giám sát hành vi của người dùng. Dữ liệu cảm biến được thu thập và sau đó được phân tích bằng cách sử dụng các kỹ thuật học máy và khai phá dữ liệu để xây dựng các mô hình hoạt động người cũng như thực hiện nhận dạng mẫu. Trong phương pháp này, các cảm biến có thể được gắn tại nhiều vị trí khác nhau trên cơ thể người. Cảm biến mang trên người cũng khá đa dạng về chủng loại, do đó có khả năng thu thập nhiều loại thông tin về hành vi người dùng. Ví dụ như cảm biến gia tốc, con quay hồi chuyển, cảm biến đo nhiệt độ, nhịp tim hay thậm chí cả cảm biến RFID (Radio-Frequency Identification).

Có rất nhiều ứng dụng hữu ích dựa trên nhận dạng hoạt động người, như các ứng dụng trong y tế, trong công nghiệp, trong thể thao, giải trí, v.v. Để các ứng dụng này có thể phục vụ cuộc sống con người trong thực tế, chúng phải hiện diện khắp mọi nơi và tại bất kỳ thời điểm nào người dùng cần. Điều này có nghĩa là hệ thống phải truy cập được thông tin của người dùng một cách liên tục. Hệ thống mang trên người có thể đáp ứng yêu cầu đó nhờ sử dụng các cảm biến có khả năng thu thập thông tin người dùng mà không bị gián đoạn. Một lợi ích quan trọng khác

của các hệ thống mang trên người là khả năng nhận thức thế giới từ góc nhìn nhân vật một cách liên tục, mà không cần hạ tầng hỗ trợ bên ngoài. Khả năng này giúp cải tiến các ứng dụng hiện tại, tạo ra nhiều ứng dụng nhận dạng hoạt động mới, như chăm sóc sức khỏe con người, hỗ trợ cuộc sống hàng ngày, các ứng dụng trong công nghiệp, giải trí và nghệ thuật.

Luận án này sẽ tập trung khai thác về *nhận dạng hoạt động sử dụng các cảm biến mang trên người*. Trước tiên, luận án sẽ khảo sát các loại cảm biến mang trên người, kiến trúc hệ thống nhận dạng hoạt động sử dụng cảm biến mang trên người và một số phương pháp nhận dạng. Sau đó luận án mô tả cách thức sử dụng cảm biến và kết hợp với các phương pháp học máy để thu thập dữ liệu, mô hình hóa, học và nhận dạng hoạt động người dùng.

Phần tiếp theo sẽ trình bày về tính cấp thiết, mục tiêu và khái quát về những đóng góp của luận án trong việc giải quyết bài toán này.

2. TÍNH CẤP THIẾT CỦA LUẬN ÁN

Nhờ vào sự tiên bộ của công nghệ và các nhu cầu ứng dụng trong những năm gần đây, nghiên cứu về nhận dạng hoạt động người đã có được sự phát triển mạnh mẽ. Trong một thập kỷ qua, các công nghệ cảm biến và kỹ thuật xử lý dữ liệu đã có những bước tiến lớn. Các cảm biến được thu nhỏ hơn, chính xác hơn, bộ nhớ lớn hơn, tiêu thụ năng lượng ít hơn và giá tiền cũng rẻ hơn, đồng thời có khả năng kết nối mạng. Sự tiên bộ của công nghệ đã thúc đẩy cộng đồng nghiên cứu chuyển từ truyền, thu nhận và xử lý dữ liệu mức thấp sang nghiên cứu tích hợp thông tin mức cao, xử lý ngữ cảnh, nhận dạng và suy diễn các hoạt động. Thêm vào đó, ngày càng có nhiều bài toán thực tế cần các giải pháp dựa vào nhận dạng hoạt động. Ví dụ như ứng dụng an ninh và theo dõi giám sát cần công nghệ nhận dạng hoạt động để xác định các mối đe dọa về khủng bố. Các ứng dụng trợ giúp con người trong cuộc sống hàng ngày cần công nghệ giám sát, nhận dạng hoạt động và hỗ trợ người sống một mình, đặc biệt là người già cô đơn. Hàng loạt ứng dụng mới khác như phòng họp thông minh, ngôi nhà thông minh, bệnh viện thông minh, v.v. cũng phụ thuộc vào

công nghệ nhận dạng hoạt động để cung cấp nhiều cách thức tương tác đa dạng, chủ động cung cấp các dịch vụ, trợ giúp người dùng hoàn thành công việc.

Hiện nay, có ba cách tiếp cận phổ biến để giải quyết bài toán nhận dạng hoạt động, bao gồm: nhận dạng hoạt động dựa trên thị giác máy tính, nhận dạng hoạt động dựa trên các cảm biến gắn trong môi trường xung quanh và nhận dạng hoạt động dựa trên cảm biến mang trên người.

Những nghiên cứu trong thị giác máy tính từ trước tới nay đã và đang được ứng dụng để giải quyết bài toán nhận dạng hoạt động [50,54,86,122]. Cách tiếp cận này dựa trên các thiết bị cảm biến thị giác như máy quay phim để quan sát hành vi người dùng và các thay đổi của môi trường xung quanh. Cách tiếp cận thứ hai là sử dụng cảm biến gắn trong môi trường xung quanh để theo dõi các hoạt động người dùng [53,63,116,129]. Cụ thể là các cảm biến được gắn trong các đối tượng hay vật thể cấu thành môi trường hoạt động của người dùng. Hai cách tiếp cận này đều có hạn chế là các hoạt động của người dùng bị giới hạn trong một môi trường cố định, đồng thời hệ thống cũng cần được triển khai, lắp đặt sẵn trong môi trường. Các hạn chế này là rào cản trong việc triển khai rộng rãi các ứng dụng nhận dạng hoạt động người trong thực tế. Các nỗ lực để nhận dạng hoạt động trong môi trường không hạn chế của cuộc sống hàng ngày đã tạo nên bước chuyển dịch trong nghiên cứu về nhận dạng hoạt động người, đó là sử dụng các *cảm biến có thể mang theo người* [5,11,47,99,121,136,138,144]. Đây chính là bước tiếp cận thứ ba trong nhận dạng hoạt động người. Cảm biến mang ngay trên cơ thể người đã mở ra nhiều ứng dụng tiềm năng trong nhận dạng hoạt động do con người không còn bị giới hạn trong những căn phòng được lắp sẵn các thiết bị. Chúng mang lại khả năng cung cấp những sự trợ giúp thông minh, các giao tiếp ảo tại bất kỳ nơi đâu và bất kỳ khi nào, thông qua việc quan sát các hoạt động từ góc nhìn nhân vật.

Bài toán nhận dạng hoạt động người nói chung và nhận dạng hoạt động dựa trên cảm biến mang theo người nói riêng cũng có thể tiếp cận theo hai hướng, đó là nhận dạng dựa trên tri thức và nhận dạng dựa trên dữ liệu [32]. Trong trường hợp

dựa trên tri thức, hệ thống nhận dạng sử dụng các luật suy diễn được xây dựng trước. Ưu điểm của cách tiếp cận này là rõ ràng về ngữ nghĩa, có tính logic cao và dễ thực hiện. Tuy vậy vẫn tồn tại những nhược điểm như cần nhiều chi phí về thời gian và kinh nghiệm của chuyên gia để xây dựng được tập luật suy diễn tốt, việc cập nhật tự động các luật là không khả thi do nguồn dữ liệu đầu vào thường không có cấu trúc và luôn biến động, đồng thời không có khả năng xử lý thông tin tạm thời và chưa rõ ràng. Chính vì vậy nhận dạng hoạt động dựa trên tri thức có xu hướng ít được sử dụng.

Cách tiếp cận thứ hai cho nhận dạng hoạt động người là dựa trên dữ liệu. Các quy tắc nhận dạng được xây dựng dựa trên các tập dữ liệu hành vi người dùng lớn có sẵn và các kỹ thuật học máy thống kê. Phương pháp này liên quan tới việc xây dựng nên các mô hình hoạt động dựa trên xác suất hoặc thống kê cùng với các tiến trình học và huấn luyện. Do dựa trên dữ liệu nên cách tiếp cận này đảm bảo được rằng hệ thống có thể cập nhật các quy tắc nhận dạng hoạt động một cách tự động mà không phụ thuộc vào tri thức chuyên gia cũng như có khả năng xử lý các thông tin tạm thời và chưa rõ ràng. Đồng thời, hệ thống cũng có khả năng thích nghi cao và tận dụng được nguồn dữ liệu có sẵn. Chính vì vậy, cách tiếp cận này được quan tâm nghiên cứu nhiều hơn so với phương pháp nhận dạng hoạt động dựa trên tri thức.

Dù có chung nhiều vấn đề về mặt phương pháp luận với các lĩnh vực nghiên cứu khác như thị giác máy, xử lý ngôn ngữ tự nhiên, hay nhận dạng tiếng nói, nhận dạng hoạt động dựa trên cảm biến mang theo người cũng vẫn phải đối mặt với nhiều thách thức riêng, đòi hỏi các phương pháp tính toán chuyên biệt kể cả khi đã sử dụng các phương pháp hiện đang được ứng dụng trong các lĩnh vực khác. Nguyên nhân là vì bên cạnh những vấn đề chung của bài toán nhận dạng mẫu, vẫn tồn tại một số vấn đề mang tính đặc thù đối với nhận dạng hoạt động sử dụng cảm biến mang trên người. Đó là sự chưa thống nhất về định nghĩa cũng như tính đa dạng của hoạt động thể chất, sự mất cân bằng của các loại hoạt động, khó khăn trong gán nhãn dữ liệu, khó khăn trong thu thập dữ liệu và thử nghiệm, và khó khăn

trong việc đảm bảo cân bằng giữa độ chính xác, độ trễ và khả năng xử lý. Mặc dù đã có nhiều nghiên cứu hướng tới các chủ đề này, nhưng đây vẫn là những vấn đề nghiên cứu mở, có tính thời sự và thu hút được sự quan tâm của cộng đồng nghiên cứu.

Đề tài “*Nghiên cứu phương pháp học máy cho nhận dạng hoạt động sử dụng cảm biến mang trên người*” được thực hiện trong khuôn khổ luận án tiến sĩ chuyên ngành hệ thống thông tin nhằm góp phần giải quyết một số vấn đề còn tồn tại trong phương pháp nhận dạng hoạt động sử dụng cảm biến mang trên người.

3. MỤC TIÊU CỦA LUẬN ÁN

Mục tiêu của luận án là nghiên cứu đề xuất một số phương pháp học máy nhằm tăng cường hiệu năng cho các hệ thống nhận dạng hoạt động sử dụng cảm biến mang trên người, cụ thể là đề xuất các phương pháp trích xuất đặc trưng mới và hiệu quả cho nhận dạng hoạt động. Đặc biệt, nghiên cứu tập trung vào hai mục tiêu:

- Nghiên cứu và đề xuất phương pháp trích xuất đặc trưng có *tốc độ nhanh* cho các ứng dụng nhận dạng một số *hoạt động riêng lẻ*, sử dụng cảm biến mang theo người với yêu cầu về *thời gian thực*. Các hệ thống nhận dạng hoạt động trên thiết bị nhúng như các hệ thống hỗ trợ người dùng thông minh mang trên người và có yêu cầu hoạt động theo thời gian thực cần các phương pháp trích xuất đặc trưng có khả năng tính toán nhanh nhưng vẫn phải đảm bảo độ chính xác cần thiết. Để có thể đáp ứng được yêu cầu này trong trường hợp tổng quát là vô cùng khó khăn. Phương pháp đề xuất sẽ giải quyết vấn đề đặt ra cho một lớp các ứng dụng nhận dạng hoạt động, đó là những ứng dụng nhận dạng một số hoạt động riêng lẻ.
- Nghiên cứu và đề xuất phương pháp trích xuất đặc trưng giúp nâng cao *độ chính xác* cho hệ thống cần nhận dạng cho *nhiều loại hoạt động*. Các hệ thống nhận dạng hoạt động phổ biến thường có một hạn chế dễ nhận

thấy là khi số lượng hoạt động cần nhận dạng càng nhiều thì độ chính xác của hệ thống càng giảm. Phương pháp đề xuất sẽ giải quyết được hạn chế này.

Các mục tiêu liệt kê trên đây cũng mô tả phạm vi và đối tượng nghiên cứu của luận án. Đó là sử dụng *các phương pháp học máy* để nghiên cứu và đề xuất các *phương pháp trích xuất đặc trưng* mới trong *các hệ thống nhận dạng hoạt động sử dụng cảm biến mang trên người*. Các phương pháp trích xuất đặc trưng này có thể phù hợp với các hệ thống trợ giúp cá nhân thông minh nhỏ gọn mang theo người, năng lực xử lý thấp và yêu cầu hoạt động theo thời gian thực, hoặc phù hợp với hệ thống cần nhận dạng nhiều hoạt động người phức tạp. Các hoạt động con người bao gồm các hoạt động trong cuộc sống hàng ngày, các hoạt động sản xuất, giải trí, thể thao, là các chuyển động của cơ thể có thể ghi nhận và phân biệt được bằng các cảm biến mang trên người.

4. CÁC ĐÓNG GÓP CỦA LUẬN ÁN

Đóng góp thứ nhất của luận án là đề xuất một phương pháp trích xuất đặc trưng mới đơn giản, hiệu quả dựa trên tín hiệu cảm biến mang trên người cho các bài toán nhận dạng hoạt động riêng lẻ. Cụ thể là:

- Đề xuất phương pháp giải quyết bài toán nhận dạng hoạt động người dựa trên phương pháp trích xuất đặc trưng mới, gọi là HALF. Các đặc trưng HALF có khả năng tính toán nhanh, phù hợp cho các hệ thống di động thông minh nhỏ gọn với năng lực tính toán thấp, đặc biệt là các hệ thống đòi hỏi thời gian thực. Các đặc trưng này cho kết quả cao khi ứng dụng trong việc phân biệt các hoạt động *ngã* và các hoạt động khác, sử dụng cảm biến gia tốc.
- Thử nghiệm và đánh giá kết quả của phương pháp đề xuất trên các tập dữ liệu đã được công bố rộng rãi. Kết quả thử nghiệm được so sánh với kết quả của một số nghiên cứu có độ chính xác cao và đáng tin cậy gần đây.

Đóng góp thứ hai của luận án là đề xuất một phương pháp trích xuất đặc trưng tự động có độ chính xác cao cho nhiều loại hoạt động người dựa trên học đặc trưng. Cụ thể là:

- Đề xuất phương pháp giải quyết bài toán nhận dạng hoạt động người sử dụng phương pháp trích xuất đặc trưng tự động dựa trên học đặc trưng nhằm tự động nắm giữ được sự phù hợp của các đặc trưng. Các đặc trưng đề xuất là các đặc trưng đa mức, gọi là motion primitive forests (MPF). Các đặc trưng đề xuất không những cải thiện được độ chính xác trong nhận dạng hoạt động người so với các đặc trưng đa mức kiểu cũ mà còn giúp loại trừ được các hạn chế về mặt tốc độ xử lý, đồng thời phù hợp với việc nhận dạng nhiều loại hoạt động người.
- Thử nghiệm và đánh giá kết quả của phương pháp đề xuất trên các tập dữ liệu đã công bố rộng rãi, so sánh kết quả thử nghiệm với một số phương pháp nhận dạng hoạt động người tiên tiến nhất hiện nay.

Đóng góp thứ ba của luận án là áp dụng phương pháp trích xuất đặc trưng đã đề xuất để xây dựng các ứng dụng nhận dạng hoạt động người, bao gồm: *ứng dụng phát hiện ngã trong thời gian thực* và *ứng dụng xác thực người dùng sử dụng chữ ký 3D*. Hệ thống phát hiện ngã sử dụng các bộ cảm biến đeo trên người rẻ tiền, và dễ triển khai, có thể phát hiện ngã trong thời gian thực. Hệ thống xác thực người dùng cho phép các ứng dụng cài đặt trên thiết bị di động thông minh có thể xác thực người dùng thông qua hành động “ký tên” trong không gian (gọi là chữ ký 3D) trong thời gian thực. Các hệ thống này cũng được thử nghiệm và đánh giá về hiệu năng cũng như tính khả thi một cách chặt chẽ trên tập dữ liệu thu thập về hoạt động người dùng.

5. BỐ CỤC CỦA LUẬN ÁN

Nội dung luận án được xây dựng thành bốn chương như sau.

Chương 1. Giới thiệu tổng quan về nhận dạng hoạt động sử dụng cảm biến mang trên người, bao gồm giới thiệu chung về nhận dạng hoạt động sử dụng cảm biến mang trên người và những khó khăn hiện tại trong lĩnh vực nghiên cứu này. Chương này cũng trình bày về các thành phần cấu thành hệ thống nhận dạng sử dụng cảm biến mang trên người bao gồm các loại cảm biến, các hoạt động người và kiến trúc hệ thống nhận dạng hoạt động. Cuối cùng là phân loại các phương pháp nhận dạng hoạt động người cùng những ưu nhược điểm trong từng phương pháp. Từ những cơ sở nghiên cứu này sẽ xác định rõ hướng nghiên cứu của luận án.

Chương 2. Trình bày một phương pháp trích xuất đặc trưng mới có tốc độ tính toán nhanh, phù hợp cho các hệ thống di động thông minh, nhỏ gọn với năng lực tính toán thấp, đặc biệt là các hệ thống đòi hỏi thời gian thực. Các đặc trưng này cho kết quả cao khi ứng dụng trong việc phân biệt các hoạt động ngã và các hoạt động khác. Nội dung trình bày trong chương này được tổng hợp dựa trên kết quả các công trình nghiên cứu số 1, số 2, số 4 và số 6 của tác giả.

Chương 3. Trình bày phương pháp trích xuất tự động mới nhờ học đặc trưng, phù hợp cho nhiều loại dữ liệu và nâng cao độ chính xác của hệ thống nhận dạng. Các đặc trưng đề xuất còn giúp loại trừ được các hạn chế về mặt tốc độ xử lý, đáp ứng được yêu cầu nhận dạng nhiều loại hoạt động người. Nội dung trình bày trong chương này được tổng hợp từ kết quả công trình nghiên cứu số 7 của tác giả.

Chương 4. Trình bày hai ứng dụng nhận dạng hoạt động người trong thời gian thực dựa trên các đặc trưng đã đề xuất trong chương 2, bao gồm ứng dụng phát hiện ngã trong thời gian thực sử dụng cảm biến gia tốc tích hợp trong thiết bị Wii Remote và ứng dụng xác thực người dùng sử dụng chữ ký 3D dựa trên chữ ký viết tay của người dùng. Nội dung trình bày trong chương này được tổng hợp dựa trên kết quả các công trình nghiên cứu số 4, số 5 và số 6 của tác giả.

Cuối cùng là một số kết luận về luận án.

CHƯƠNG 1

TỔNG QUAN VỀ NHẬN DẠNG HOẠT ĐỘNG SỬ DỤNG CẢM BIẾN MANG TRÊN NGƯỜI

Chương này trình bày những vấn đề tổng quan về nhận dạng hoạt động sử dụng cảm biến mang trên người, các khó khăn trong nghiên cứu, các loại cảm biến thường dùng và các hoạt động người. Phần tiếp theo trình bày về một kiến trúc chung thường dùng cho các hệ thống nhận dạng hoạt động sử dụng cảm biến mang trên người. Do phạm vi nghiên cứu của luận án tập trung chủ yếu vào trích xuất đặc trưng, nên để làm cơ sở trình bày các chương sau, phần tiếp theo sẽ trình bày các kết quả nghiên cứu đã có về nội dung này, đó là các phương pháp trích xuất đặc trưng cùng những vấn đề tồn tại trong từng phương pháp. Những cơ sở nghiên cứu này sẽ giúp xác định rõ hướng nghiên cứu cụ thể của luận án. Ở Việt Nam, nghiên cứu về nhận dạng hoạt động sử dụng cảm biến mang trên người là hầu như không có, đặc biệt là các nghiên cứu về cải tiến phương pháp nhận dạng. Do đó, các nghiên cứu trình bày ở đây đều là các nghiên cứu quốc tế. Cuối cùng, phần kết luận chương nêu ra một số vấn đề quan trọng mà luận án sẽ tập trung giải quyết.

1.1. GIỚI THIỆU CHUNG VỀ NHẬN DẠNG HOẠT ĐỘNG SỬ DỤNG CẢM BIẾN MANG TRÊN NGƯỜI

1.1.1. *Khái niệm*

Nhận dạng hoạt động người là quá trình giám sát và phân tích hành vi người dùng và trạng thái môi trường xung quanh nhằm suy diễn/nhận dạng các hoạt động đang xảy ra [33]. Một trong những mục tiêu của nhận dạng hoạt động là cung cấp thông tin về hành vi của người dùng, từ đó cho phép hệ thống tính toán chủ động hỗ trợ người dùng trong công việc [4].

Nhận dạng hoạt động sử dụng cảm biến mang trên người là một hướng tiếp cận hiệu quả với chi phí thấp cho thu thập dữ liệu và nhận dạng về các hoạt động của con người, nhờ vào công nghệ cảm biến gắn trên người. Hướng nghiên cứu này gần đây thu hút được nhiều sự quan tâm của cộng đồng nghiên cứu do những nhu cầu cấp thiết về nhận dạng các hoạt động người trong môi trường không hạn chế của cuộc sống hàng ngày. Thêm vào đó, sự tiến bộ không ngừng của công nghệ cảm biến đã tạo nên các cảm biến mang trên người mạnh mẽ, có tính ứng dụng cao, giá rẻ, cho phép khả năng hoạt động liên tục không hạn chế. Nhờ khả năng theo dõi các hoạt động từ góc nhìn của người dùng, cảm biến mang trên người cho phép các ứng dụng nhận dạng hiện tại khắc phục được những hạn chế do môi trường hoạt động, biến chúng trở thành các hệ thống trợ giúp thông minh cho con người mọi lúc mọi nơi.

Có nhiều loại hoạt động khác nhau tùy thuộc vào ứng dụng nhận dạng hoạt động người. Tuy nhiên, về cơ bản thì hoạt động người có thể được chia làm hai loại, đó là hoạt động mức thấp và hoạt động mức cao [55]. Các hoạt động mức thấp bao gồm các hoạt động như đi bộ, ngồi xuống, đứng lên, hút bụi, ăn, rửa bát, v.v. Đây là các hoạt động đặc trưng bởi chuyển động cơ thể, tư thế hoặc cách sử dụng vật dụng, thông thường kéo dài chỉ trong vài giây hay vài phút. Thậm chí chúng có thể chỉ tồn tại trong thời gian nhỏ hơn với các chuyển động cơ thể riêng biệt và ngắn gọn, ví dụ như một bước đi hoặc đong đưa một cái vợt hay một cử chỉ như gập cánh tay. Ngược lại, hoạt động mức cao thường gồm một tập hợp các hoạt động mức thấp và diễn ra trong thời gian dài hơn như lau nhà, đi ngắm cảnh hay làm việc tại văn phòng. Các hoạt động này có thể kéo dài vài phút hoặc nhiều giờ đồng hồ.

1.1.2. Cách tiếp cận

Nhận dạng hoạt động sử dụng cảm biến mang trên người là một quá trình phức tạp và có thể được chia thành bốn bước cơ bản (theo [32]), bao gồm:

- **Bước 1:** chọn và cài đặt các cảm biến thích hợp cho các đối tượng và môi trường để giám sát cũng như nắm giữ được thông tin về hành vi người dùng cùng với các thay đổi trạng thái môi trường.
- **Bước 2:** thu thập, lưu trữ và xử lý các thông tin nhận được nhờ sử dụng các kỹ thuật phân tích dữ liệu và biểu diễn tri thức ở các mức độ trừu tượng phù hợp.
- **Bước 3:** xây dựng các mô hình hoạt động để hệ thống có thể suy diễn và xử lý.
- **Bước 4:** chọn và xây dựng các giải thuật suy diễn để suy luận ra các hoạt động từ dữ liệu cảm biến.

Thông thường, phương pháp sử dụng trong một bước sẽ phụ thuộc vào phương pháp dùng trong một bước khác.

Phương pháp giải quyết bài toán này có thể được phân làm hai loại theo cách thức xây dựng các quy tắc nhận dạng (hay mô hình nhận hoạt động), là nhận dạng hoạt động dựa trên tri thức chuyên gia và nhận dạng dựa trên dữ liệu [32].

Nhận dạng hoạt động dựa trên tri thức chuyên gia. Trong trường hợp này, hệ thống nhận dạng sử dụng các luật suy diễn được xây dựng trước. Phương pháp này khai thác các biểu diễn tri thức cho việc mô hình hóa các hoạt động và dữ liệu cảm biến, sau đó sử dụng suy luận logic để thực hiện nhận dạng hoạt động. Cách thức chung đối với phương pháp nhận dạng hoạt động loại này bao gồm: (i) sử dụng hình thức logic để xác định và mô tả một tập các mô hình hoạt động rõ ràng cho tất cả các hoạt động có thể có trong lĩnh vực đang xét, (ii) tổng hợp và chuyển đổi dữ liệu cảm biến thành các công thức và hạng tử logic và (iii) thực hiện các suy diễn logic để trích xuất ra một tập tối thiểu các mô hình diễn tả từ tập mô hình hoạt động dựa trên các hoạt động đã theo dõi nhằm giải thích được các quan sát.

Ưu điểm của cách tiếp cận này là rõ ràng về ngữ nghĩa và có tính logic cao. Các luật suy diễn có thể được tính toán nhanh chóng với chi phí thấp. Thêm vào đó,

có thể dễ dàng tích hợp tri thức chuyên gia và các thuật giải heuristic cho các mô hình hoạt động và tổng hợp dữ liệu. Tuy nhiên, có một số nhược điểm ở đây là: cần nhiều chi phí về thời gian và kinh nghiệm của chuyên gia để xây dựng được tập luật suy diễn tốt; việc cập nhật tự động các luật là không khả thi do nguồn dữ liệu đầu vào thường không có cấu trúc và luôn biến động; không có khả năng xử lý thông tin tạm thời và không rõ ràng. Hầu hết các phương pháp đều không cung cấp cách thức để xác định xem một mô hình cụ thể nào đó có hiệu quả hơn các mô hình khác hay không. Một nhược điểm nữa là các phương pháp dựa trên tri thức và suy luận logic không có khả năng học. Chính vì vậy nhận dạng hoạt động dựa trên tri thức có xu hướng ít được sử dụng.

Nhận dạng hoạt động dựa trên dữ liệu. Trong cách tiếp cận này, các quy tắc nhận dạng được xây dựng dựa trên các tập dữ liệu hành vi người dùng lớn, có sẵn cùng với việc sử dụng các *kỹ thuật học máy*. Phương pháp này liên quan tới việc tạo ra các *mô hình hoạt động dựa trên xác suất hoặc thống kê cùng với các tiến trình học và huấn luyện*. Các phương pháp học máy có hai loại là có giám sát và không giám sát. Cách thức chung đối với nhận dạng hoạt động loại này, chẳng hạn như đối với phương pháp có giám sát, gồm các bước sau: (i) thu thập dữ liệu cảm biến, (ii) trích chọn đặc trưng dữ liệu đầu vào (feature extraction and selection), (iii) huấn luyện bộ phân lớp và (iv) phân loại hoạt động sử dụng bộ phân lớp đã huấn luyện.

Do dựa trên dữ liệu nên cách tiếp cận này đảm bảo hệ thống có thể cập nhật các quy tắc nhận dạng hoạt động một cách tự động mà không phụ thuộc vào tri thức chuyên gia, có khả năng xử lý các thông tin tạm thời và không rõ ràng, kể cả các thông tin không đầy đủ, đồng thời hệ thống có khả năng thích nghi cao và tận dụng được nguồn dữ liệu. Tuy chi phí tính toán để học hoạt động trong một mô hình xác suất cho các ứng dụng thực tế khá cao [32], độ phức tạp tính toán cần thiết để suy diễn các hoạt động là không cao nếu bước trích chọn đặc trưng và phân lớp không phức tạp. Với những ưu điểm như trên, cách tiếp cận này được quan tâm nghiên cứu nhiều hơn so với nhận dạng hoạt động dựa trên tri thức chuyên gia. Ngoài ra, nhờ

vào sự phát triển của hạ tầng công nghệ, việc thu thập dữ liệu cho hệ thống nhận dạng hoạt động ngày càng nhanh chóng và dễ dàng, làm cho nhận dạng hoạt động dựa trên dữ liệu trở thành cách tiếp cận chính. Chính vì vậy, luận án sẽ tập trung nghiên cứu nhận dạng hoạt động theo hướng này.

1.1.3. Các ứng dụng

Ý tưởng nhận dạng hoạt động sử dụng cảm biến mang trên người đã có từ cuối thập niên 90 khi bắt đầu thời kỳ của tính toán di động và tính toán khắp nơi, khởi đầu với một số nghiên cứu khả thi về phân biệt một số hành động đơn giản của con người như chạy và đi bộ. Nhiều thành công nổi tiếp sau đó trong nhận dạng hoạt động đã tạo ra những động lực mạnh mẽ trong việc giải quyết các vấn đề khó khăn và thực tế hơn. Có thể dễ dàng thấy được, một số các lĩnh vực thực tế đã nhận được nhiều ích lợi từ nhận dạng hoạt động sử dụng cảm biến mang trên người, ví dụ như trong công nghiệp [95,126], an ninh [62], trong văn phòng, thể thao, giải trí [73,75,98], và đặc biệt là trong lĩnh vực chăm sóc sức khỏe. Cụ thể, các hoạt động sống hàng ngày (ADL) [67] đã thu hút được rất nhiều sự quan tâm của cộng đồng nghiên cứu [17,90,117,135]. Giám sát hoạt động hàng ngày đã cho thấy những ưu thế quan trọng so với các phương pháp y học truyền thống trong việc hỗ trợ chẩn đoán y tế, phục hồi chức năng hoặc giúp đỡ bệnh nhân suy giảm chức năng mãn tính [15,31,105,115,125,132,137]. Một mục tiêu quan trọng khác nữa đó là hỗ trợ sớm để khuyến khích con người sống một lối sống khỏe mạnh. Mục tiêu này dẫn đến nhiều nghiên cứu về các hoạt động có liên quan của con người, ví dụ như đánh răng [82] hoặc rửa tay, ăn uống [10,112] và uống thuốc [104,143], hay thói quen di chuyển [37,72] và tính toán tiêu thụ năng lượng [65]. Một vài nghiên cứu khác nữa cũng tập trung vào cảnh báo các nguy hiểm cho sức khỏe con người, ví dụ như phát hiện ra sự kiện người ngã [96,111], hoặc sự kiện có các dấu hiệu đe dọa sự sống con người [12,88,141].

Gần đây, nhận dạng hoạt động đã trở thành một thành phần quan trọng trong khá nhiều sản phẩm tiêu dùng. Ví dụ máy chơi trò chơi Wii của Nintendo hay

Kinect của Microsoft dựa vào nhận dạng các cử chỉ hoặc thậm chí cả chuyển động toàn thân để người chơi điều khiển trò chơi. Mặc dù ban đầu các hệ thống này được phát triển cho việc giải trí, chúng cũng được mở rộng sang các ứng dụng khác, như huấn luyện thể thao và phục hồi chức năng, tạo thêm động lực mới cho lĩnh vực nghiên cứu nhận dạng hoạt động [130]. Một số sản phẩm thể thao như DirectLife Philips hoặc giày chạy Nike+ cũng tích hợp cảm biến chuyển động dùng cho cả vận động viên chuyên nghiệp và không chuyên để hỗ trợ quá trình luyện tập thông qua những phản hồi về hiệu suất từng hoạt động của vận động viên.

1.1.4. Một số khó khăn trong nghiên cứu

Những mô tả trên về các ứng dụng đã thể hiện tầm quan trọng của nhận dạng hoạt động người trong cả nghiên cứu lẫn sản xuất hiện nay. Mặc dù đã có những tiến bộ đáng kể trong phương pháp suy diễn các hoạt động nhờ cảm biến mang trên người cũng như khả năng tạo bản mẫu và triển khai các hệ thống nhận dạng hoạt động [13,49] nhưng việc phát triển các hệ thống có thể đáp ứng được yêu cầu của người dùng và ứng dụng trong thực tế vẫn là một nhiệm vụ rất khó khăn. Thậm chí, các kỹ thuật nhận dạng hoạt động đã áp dụng thành công trong bài toán này nhưng chưa chắc đã thành công trong trường hợp khác.

Thêm vào đó, dù có chung nhiều vấn đề về mặt phương pháp luận với các lĩnh vực nghiên cứu khác như thị giác máy, xử lý ngôn ngữ tự nhiên hay nhận dạng tiếng nói, nhận dạng hoạt động cũng vẫn phải đối mặt với nhiều thách thức riêng, đòi hỏi một phương pháp tính toán chuyên biệt kể cả khi đã ứng dụng các phương pháp hiện đang sử dụng trong các lĩnh vực khác [28]. Chẳng hạn như, chưa có bất kỳ một phát biểu về bài toán chung một cách rõ ràng (hoạt động nào cần được nhận dạng, mô tả đặc trưng của một hoạt động cụ thể như thế nào, v.v). Đối với một số ứng dụng như giám sát hành vi trong thời gian dài, các hoạt động có liên quan thông thường thậm chí không được xác định rõ ràng từ trước. Ngoài ra, hoạt động của con người có mức độ đa dạng cao dẫn tới yêu cầu phải có sự lựa chọn kỹ càng giữa các loại cảm biến khác nhau về cả khả năng lẫn đặc tính. Khi đó cũng có thể phải thay

đòi hỏi các cảm biến tùy thuộc yêu cầu ứng dụng hiện tại. Thêm nữa, đối với các ứng dụng cụ thể, nhận dạng hoạt động thông thường đòi hỏi có thước đo đánh giá riêng, cụ thể để phản ánh chất lượng hệ thống. Sự khác biệt này đã tạo ra nhiều khó khăn trong nghiên cứu nhận dạng hoạt động sử dụng cảm biến mang trên người.

1.2. CÁC CẢM BIẾN MANG TRÊN NGƯỜI

Khái niệm cảm biến mang trên người là dùng để chỉ các cảm biến gắn trực tiếp hay gián tiếp trên cơ thể người. Cảm biến sinh ra tín hiệu khi người dùng thực hiện các hoạt động. Dựa vào những đặc điểm của tín hiệu đầu ra của cảm biến, hệ thống có thể phân biệt được các trạng thái sinh lý hoặc loại hoạt động đang thực hiện của con người. Cảm biến có thể được gắn trong quần áo, kính mắt, thắt lưng, giày, đồng hồ, các thiết bị di động hoặc đặt trực tiếp trong hoặc trên các bộ phận của cơ thể người. Các thông tin thu nhận được có thể là thông tin vị trí, các chuyển động cơ thể hay trạng thái sinh lý. Nhiều nghiên cứu cho thấy thông tin đầu ra của các cảm biến khác nhau có hiệu quả khác nhau trong việc phân loại các hoạt động.

Có thể phân loại các cảm biến theo cách thức sử dụng chúng trong phương pháp nhận dạng hoạt động người, bao gồm: các cảm biến chuyển động (có kết hợp các cảm biến khác), các cảm biến giúp xác định vị trí người dùng, các cảm biến được gắn vào đối tượng sử dụng và các cảm biến phát hiện dấu hiệu sống. Nhóm thứ nhất gồm các cảm biến mang trên người hay dùng nhất trong các nghiên cứu nhận dạng hoạt động. Đó là các cảm biến thu nhận được thông tin chuyển động của cơ thể, bao gồm cảm biến gia tốc, cảm biến con quay hồi chuyển, cảm biến âm thanh. Trong đó cảm biến gia tốc được dùng nhiều nhất [17,68,81,92] do khả năng đo được cả tần suất lẫn cường độ của chuyển động, và cũng có thể kết hợp với một số cảm biến khác như con quay hồi chuyển hay cảm biến âm thanh để tăng hiệu quả nhận dạng hoạt động trong nhiều điều kiện môi trường khác nhau (ví dụ [8,17,81,92]). Các cảm biến trong nhóm này có độ tin cậy cao, rẻ tiền và đỡ gây phiền hà cho người dùng so với các cảm biến khác như mi-crô hay máy quay phim. Nhóm thứ hai ít phổ biến hơn là các cảm biến dùng để xác định vị trí, mà phổ biến

nhất là GPS và con quay hồi chuyển [17,81]. Dữ liệu GPS có thể được sử dụng để xác định cụ thể hành vi mức cao của con người [108], học các địa điểm quan trọng và dự đoán di chuyển của nhiều người [14], hay học và suy diễn một mô hình di chuyển của người dùng để xác định các hành vi bất thường [85]. Còn dữ liệu con quay hồi chuyển giúp định vị người dùng và tăng độ chính xác nhận dạng hoạt động người cho một số dạng hoạt động cụ thể như ngồi, đứng và đi bộ [17,81]. Nhóm thứ ba là các cảm biến gắn vào đối tượng sử dụng, ví dụ như RFID, cảm biến phát hiện tia hồng ngoại, chuyển mạch cộng từ và cảm biến chuyển động, được ứng dụng để phát hiện sự kiện sử dụng các đối tượng (như mở cửa ra vào, cửa sổ, tủ, v.v.) và các điều kiện môi trường (như ánh sáng, nhiệt độ, độ ẩm) [90,127]. Nhóm các cảm biến này giúp tạo ra các cách tiếp cận linh hoạt cho nhận dạng hoạt động người nhưng gây ra những hạn chế do môi trường được thiết kế đóng. Cuối cùng là nhóm các cảm biến sinh học, giúp đo hiện các dấu hiệu sống (bao gồm huyết áp, nhịp tim, điện não, điện tim và thông tin hô hấp), từ đó ứng dụng trong nhận dạng hoạt động người như phát hiện hiện tượng hạ thân nhiệt [131], đo nhịp hô hấp [24], đo các cơn co thắt cơ bắp [91]. Bảng 1.1 liệt kê các nhóm cảm biến này.

Có thể thấy được rằng, các mô hình cảm biến đã được sử dụng và khai thác trong các nghiên cứu từ trước tới nay là rất đa dạng, từ các mô hình với chỉ một cảm biến đến các mô hình kết hợp nhiều loại cảm biến với nhau. Trong nghiên cứu [83] về nhận dạng hoạt động của mình, Lester thực hiện xếp hạng các cảm biến sử dụng một cơ chế lựa chọn đặc trưng và thấy rằng dữ liệu cảm biến gia tốc là dữ liệu đóng vai trò quan trọng nhất trong thực nghiệm. Đồng thời, cảm biến gia tốc cũng là cảm biến được trang bị và ứng dụng nhiều nhất trong các thiết bị trợ giúp cá nhân di động thông minh mang theo người. Do đó các phần sau của nghiên cứu này sẽ tập trung chủ yếu vào các phương pháp và hệ thống sử dụng cảm biến gia tốc.

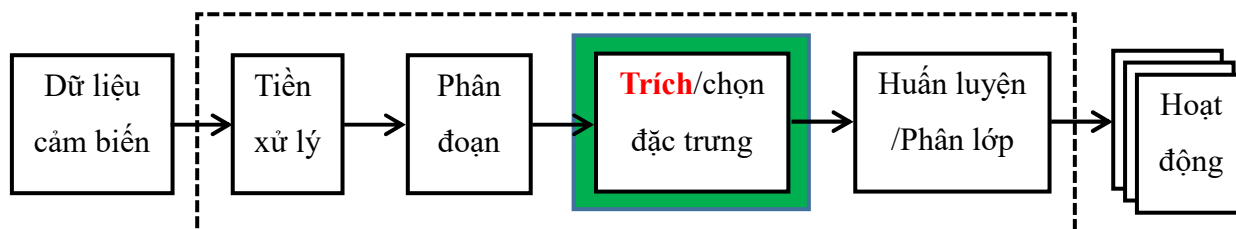
Bảng 1.1. Tóm lược các cảm biến mang trên người sử dụng trong nhận dạng hoạt động

Phương pháp nhận dạng hoạt động	Cảm biến
Dựa vào thông tin chuyển động của cơ thể (có thể kết hợp các cảm biến khác)	Cảm biến gia tốc. Có thể kết hợp cảm biến con quay hồi chuyển hoặc cảm biến âm thanh.
Dựa vào xác định vị trí người dùng	GPS, cảm biến âm thanh, cảm biến con quay hồi chuyển.
Dựa vào cảm biến gắn vào đối tượng sử dụng	Bộ đọc thẻ RFID, cảm biến phát hiện tia hồng ngoại, chuyển mạch cộng từ, các cảm biến đo nhiệt độ môi trường.
Dựa trên dấu hiệu sống	Huyết áp, nhịp tim, điện não, điện tim, hô hấp, cảm biến nhiệt độ cơ thể, cảm biến áp suất bọt, điện trở đo áp lực, cảm biến đo oxy, cảm biến độ dẫn điện của da, điện tim.

1.3. KIẾN TRÚC HỆ THỐNG NHẬN DẠNG HOẠT ĐỘNG NGƯỜI

Việc nhận dạng hoạt động có thể được thực hiện theo nhiều cách, với nhiều kiểu cảm biến hoặc hướng tới nhiều loại hoạt động khác nhau. Một hệ thống nhận dạng hoạt động người nói chung nhận dữ liệu đầu vào là dữ liệu thô lấy ra từ các cảm biến và thông thường gồm các bước như sau: tiền xử lý, phân đoạn, trích chọn đặc trưng, huấn luyện và phân lớp hoạt động (xem hình 1.1). Để có thể nhận dạng hoạt động, trước tiên hệ thống phải biết được cách thức các hoạt động được mô tả trong dữ liệu cảm biến. Sau đó, các mẫu dữ liệu sẽ được thu thập và gán nhãn để làm dữ liệu huấn luyện cho hệ thống. Kết quả nhận được như trong hình 1.1 sau khi

phân lớp là xác suất của từng hoạt động. Kiến trúc nhận dạng hoạt động này được sử dụng cho các hệ thống nhận dạng trong các chương sau (chương 2, 3, 4).



Hình 1.1. Kiến trúc hệ thống nhận dạng hoạt động người

Phần dưới đây sẽ mô tả đặc trưng của dữ liệu cảm biến và nhiệm vụ cụ thể của từng bước (như trong hình 1.1).

Dữ liệu cảm biến

Cần chú ý rằng, dữ liệu cảm biến đầu vào cho hệ thống là một trường hợp riêng của chuỗi dữ liệu thời gian, gồm một chuỗi các điểm dữ liệu được liệt kê theo trật tự về thời gian. Khi đó, chuỗi dữ liệu cảm biến đầu vào có trật tự về thời gian. Tính chất này khiến việc phân tích chuỗi thời gian khác biệt so với các loại nghiên cứu khác sử dụng các mẫu quan sát không tuân theo trật tự thời gian. Trong một chuỗi thời gian, các mẫu quan sát gần nhau về mặt thời gian sẽ có mối liên hệ chặt chẽ hơn các mẫu xa nhau. Thêm nữa, chuỗi dữ liệu thời gian có tính chất một chiều theo trật tự thời gian, do đó các giá trị trong một khoảng thời gian nhất định sẽ được thể hiện phát sinh từ các giá trị trong quá khứ, hơn là các giá trị trong tương lai [27]. Các tính chất này của chuỗi thời gian có ảnh hưởng đến các bước xử lý trong nhận dạng hoạt động người trình bày ở các phần tiếp theo.

Bước 1: Tiền xử lý

Tiền xử lý là giai đoạn đầu tiên của một hệ thống nhận dạng hoạt động điển hình. Dữ liệu đầu vào cho bước tiền xử lý (cũng chính là dữ liệu đầu vào cho hệ thống) là các luồng dữ liệu thu được từ các cảm biến mang trên người. Dữ liệu cảm biến đầu vào có thể là một chuỗi dữ liệu thời gian nhiều chiều do nhiều nguyên nhân, như là một số loại cảm biến có nhiều giá trị (như cảm biến gia tốc 3 chiều có

3 giá trị tương ứng với 3 trục x, y, z), hoặc sử dụng nhiều cảm biến khi lấy mẫu, hoặc các cảm biến được lấy mẫu sau các khoảng thời gian khác nhau.

Tuy vậy, tốc độ lấy mẫu của các loại cảm biến thường có thể khác nhau. Hoặc cảm biến có thể thay đổi tần số lấy mẫu tại một số tình huống, ví dụ như để tiết kiệm năng lượng hay do hệ thống yêu cầu. Hơn nữa, dữ liệu cảm biến chưa xử lý có thể bị ngắt quãng do nhiều nguyên nhân, có thể là do hoạt động của con người, lỗi cảm biến hoặc nhiễu sóng điện từ. Do đó, chức năng của tiền xử lý là đồng bộ, giảm nhiễu và loại bỏ các tín hiệu lạ, chuẩn bị các tín hiệu đã lấy được cho bước trích chọn đặc trưng. Tuy nhiên, tiền xử lý cần giữ lại được các đặc điểm tín hiệu có chứa thông tin quan trọng về các hoạt động đang quan tâm [28].

Bước 2: Phân đoạn

Bước phân đoạn (còn gọi là phát hiện hoạt động, hay là đánh dấu hoạt động) lấy dữ liệu được xử lý (từ luồng dữ liệu sau khi tiền xử lý), sau đó chia thành các phần nhỏ hơn có khả năng chứa thông tin về hoạt động hay cử chỉ. Các phần dữ liệu này được gọi là các đoạn (segment).

Phân đoạn một luồng cảm biến liên tục là một việc khó. Có 3 phương pháp khác nhau để thực hiện phân đoạn [28], bao gồm phân đoạn sử dụng cửa sổ trượt, phân đoạn dựa vào năng lượng và phân đoạn dựa vào các vị trí nghỉ (bằng cách sử dụng một loại cảm biến để phân đoạn cho dữ liệu của một loại cảm biến khác hay sử dụng các thông tin ngữ cảnh bên ngoài). Trong nhận dạng hoạt động sử dụng cảm biến mang trên người, phân đoạn sử dụng cửa sổ trượt thường được sử dụng nhiều nhất [28]. Do đó, cửa sổ trượt được sử dụng để phân đoạn dữ liệu trong các hệ thống nhận dạng ở các chương tiếp theo trong luận án này.

Bước 3: Trích chọn đặc trưng

Trích chọn đặc trưng bao gồm hai phần: tách/trích xuất đặc trưng (feature extraction) và lựa chọn đặc trưng (feature selection). Trích chọn đặc trưng nhằm rút gọn các tín hiệu thành các đặc trưng để phân biệt các hoạt động đang có và sau đó

được sử dụng làm dữ liệu đầu vào cho bước phân lớp. Tùy thuộc vào từng hệ thống cụ thể mà lựa chọn đặc trưng có thể được thực hiện hoặc không.

Các đặc trưng có thể được trích xuất tự động (ví dụ như trong [114]) hoặc dựa trên tri thức chuyên gia. Tập các đặc trưng có được từ dữ liệu được gọi là không gian đặc trưng. Nói chung, khi các hoạt động được phân tách càng rõ ràng trong không gian đặc trưng thì hiệu suất nhận dạng của hệ thống càng cao. Lý tưởng nhất là các đặc trưng của cùng hoạt động có thể nhóm thành một nhóm trong không gian đặc trưng và ngược lại các đặc trưng của các hoạt động khác nhau cần phân biệt càng xa càng tốt.

Không gian đặc trưng có số chiều càng lớn thì cần càng nhiều dữ liệu huấn luyện để tính toán các tham số cho mô hình cũng như cần càng nhiều lượng tính toán khi thực hiện phân lớp. Đặc biệt là trong các hệ thống nhúng cần xử lý theo thời gian thực, mục tiêu cần đạt là tối thiểu hóa yêu cầu về bộ nhớ, lượng tính toán và băng thông. Vì vậy, điều quan trọng là cần sử dụng số lượng ít nhất các đặc trưng trong khi vẫn đảm bảo hiệu năng cho hệ thống nhận dạng. Ngoài ra, việc lựa chọn tập đặc trưng tối thiểu như vậy bằng thủ công là rất khó. Do đó, nhiều phương pháp phân hạng (ranking) và lựa chọn đặc trưng tự động đã được đề xuất (xem [45]). Các phương pháp này có thể được phân loại thành khung bao (wrapper) [70], bộ lọc [109], hoặc phương pháp lai [124]. Thêm nữa, cần chú ý rằng các phương pháp học máy hiện đại như SVM hay AdaBoost cũng có cơ chế chọn đặc trưng được tích hợp sẵn.

Như đã đề cập đến trong mục tiêu và phạm vi của luận án, bước trích chọn đặc trưng trong các hệ thống nhận dạng hoạt động đề xuất ở các chương sau sẽ chỉ tập trung vào việc trích xuất đặc trưng mà không có lựa chọn đặc trưng (hình 1.1).

Bước 4: Huấn luyện và phân lớp

Ở chế độ huấn luyện, các đặc trưng được tách ra và cùng với các nhãn tương ứng, các đặc trưng sẽ được sử dụng làm dữ liệu đầu vào để huấn luyện mô hình hoạt động. Ở chế độ phân lớp, các đặc trưng và mô hình đã huấn luyện trước đó được sử

dụng để tính ra giá trị của mỗi lớp hoạt động và ánh xạ các giá trị này với một nhãn riêng trong lúc phân lớp. Khi sử dụng nhiều cảm biến hay nhiều bộ phân lớp, đầu ra của các bộ phân lớp này được kết hợp với nhau để đưa ra một quyết định duy nhất.

Việc chọn lựa một phương pháp suy diễn tùy thuộc vào sự cân bằng giữa độ phức tạp tính toán và hiệu năng nhận dạng. Khi phân lớp trong hệ thống nhúng có tài nguyên hạn chế, cần tối thiểu hóa độ phức tạp tính toán và bộ nhớ sao cho vẫn đảm bảo hệ thống có hiệu năng cao. Các phương pháp suy diễn thông thường được chọn phụ thuộc vào loại hình hoạt động và độ phức tạp của không gian đặc trưng, hoặc có thể dựa trên các nhân tố như độ trễ, hoạt động trực tuyến (online) hay độ thích nghi. Đồng thời, tùy thuộc vào chế độ hoạt động của hệ thống nhận dạng hoạt động hiện tại là huấn luyện hay phân lớp mà hệ thống tiếp tục xử lý các đặc trưng đã tính được.

1.4. CÁC PHƯƠNG PHÁP TRÍCH XUẤT ĐẶC TRƯNG

Như mô tả ở phần 1.3, có thể thấy rằng, để một hệ thống HAR nhận dạng được các hoạt động, tín hiệu từ các cảm biến phải được xử lý qua mô-đun trích chọn đặc trưng (gồm trích xuất và lựa chọn đặc trưng). Sau đó, thông qua sử dụng các kỹ thuật học máy để xây dựng mô hình nhận dạng hoạt động với dữ liệu đầu vào là tập các đặc trưng có được. Sau khi huấn luyện xong mô hình, các mẫu hoạt động đầu vào (ví dụ như các cửa sổ thời gian) chưa biết sẽ được xác định và nhận dạng thông qua mô hình này. Như đã đề cập đến trong mục tiêu và phạm vi của luận án, phần này sẽ chỉ tập trung vào các phương pháp trích xuất đặc trưng. Các phương pháp lựa chọn đặc trưng sẽ nằm ngoài phạm vi của luận án này.

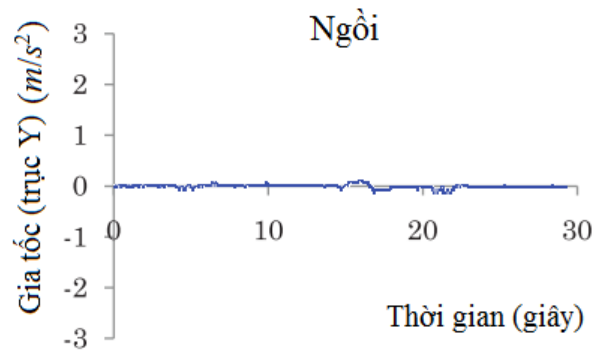
So với tần suất lấy mẫu của các cảm biến thì hoạt động người được thực hiện trong một khoảng thời gian tương đối dài. Hoạt động có thể kéo dài vài giây hay vài phút trong khi tần số lấy mẫu có thể tới 250Hz. Bên cạnh đó thì một mẫu đơn lẻ tại một thời điểm cụ thể không đủ để cung cấp đủ thông tin cho hoạt động. Vì vậy, hoạt động cần được ghi nhận trong một cửa sổ thời gian hơn là trong một mẫu. Vấn đề là làm thế nào để so sánh hai cửa sổ thời gian? Rất khó để có thể so sánh được các tín

hiệu này, ngay cả khi chúng được thực hiện bởi cùng một người, cho cùng một hoạt động lặp lại. Do đó, cần sử dụng các phương pháp trích xuất đặc trưng đối với mỗi cửa sổ thời gian để lọc ra các thông tin cần, từ đó có được các số đo định lượng dùng cho việc so sánh các tín hiệu.

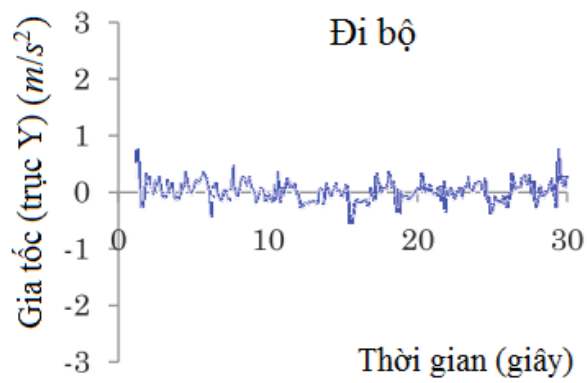
Tiêu chí cho việc lựa chọn phương pháp trích xuất đặc trưng phù hợp sẽ phụ thuộc vào bản chất của tín hiệu. Phần tiếp theo sẽ trình bày các kỹ thuật trích xuất đặc trưng phổ biến nhất, tập trung vào tín hiệu gia tốc được sử dụng chủ yếu trong các chương tiếp theo của luận án. Đồng thời phân loại kỹ thuật trích xuất đặc trưng thành hai loại: một là dựa trên tri thức chuyên gia để xác định các đặc trưng và loại thứ hai là học tự động từ tập dữ liệu huấn luyện.

1.4.1. Trích xuất đặc trưng theo tri thức chuyên gia

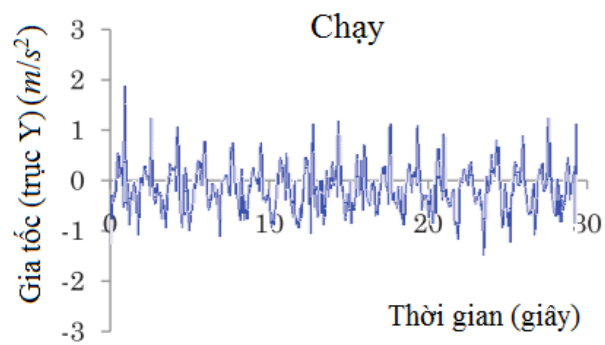
Tín hiệu gia tốc (xem ví dụ trong hình 1.2) là tín hiệu có mức độ dao động lớn, do đó rất khó có thể nhận dạng những mẫu cơ bản khi chỉ sử dụng các giá trị thô của tín hiệu chưa xử lý. Hầu hết các hệ thống HAR hiện thời sử dụng các đặc trưng theo miền thời gian hoặc tần số [40]. Phương pháp trích chọn đặc trưng thường được sử dụng nhất là tính toán các đại lượng thống kê trực tiếp trên dữ liệu thô đầu vào của cảm biến, độc lập trên từng khung dữ liệu được phân đoạn bởi cửa sổ trượt. Các đại lượng thống kê phổ biến nhất bao gồm trung bình, độ lệch chuẩn, phương sai, năng lượng, trung bình đạo hàm, khoảng tứ phân vị, entropy, tương quan giữa các trục, skewness, độ nhọn (kurtosis). Trong miền tần số, các phương pháp như biến đổi Cô-sin rời rạc (Discrete Cosine Transform) và biến đổi Fourier (Fourier Transform) được áp dụng [52,106] và mang lại những kết quả khả quan. Các phương pháp này được hình thành để giải quyết tính biến đổi cao vốn có của tín hiệu loại này. Tóm tắt về các phương pháp trích xuất đặc trưng được mô tả trong bảng 1.2.



a) Tín hiệu cho hoạt động ngồi



b) Tín hiệu cho hoạt động đi bộ



c) Tín hiệu cho hoạt động chạy

Hình 1.2. Các tín hiệu gia tốc trong các hoạt động khác nhau

Bảng 1.2. Các phương pháp trích xuất đặc trưng cho tín hiệu gia tốc

Nhóm	Phương pháp
Miền thời gian	Trung bình, RMS, độ lệch chuẩn, phương sai, MAD, trung bình đạo hàm, MCR, ZCR, khoảng tứ phân vị, entropy, tương quan giữa các trục, skewness, độ nhọn (kurtosis) [17,94,107,134,150]. Các đặc trưng vật lý của chuyển động [150].
Miền tần số	Chuyển đổi Fourier [17,34,106], biến đổi Cô-sin rời rạc [9].

Các đặc trưng thống kê được sử dụng nhiều nhất trong các nghiên cứu do tính đơn giản và hiệu năng cao trong rất nhiều bài toán nhận dạng hoạt động người [28]. Một số đặc trưng đã được khảo sát kỹ lưỡng và được chứng minh là hiệu quả cho nhận dạng hoạt động, đặc biệt là các trường hợp cần phân biệt một số hoạt động cụ thể, đơn giản. Ví dụ như trung bình [17] chỉ ra mức trung bình của giá trị giá tốc, từ đó giúp phân biệt các tư thế như ngồi, đứng im, và nằm xuống (thông qua tính toán các véc-tơ trọng lực), còn phương sai được chứng minh là giúp phân biệt tốt cho các hoạt động đi bộ, chạy bộ và nhảy [58] do cho thấy mức độ chuyển động có trong tín hiệu. Tương quan giữa các cặp trục gia tốc giúp phân biệt các hoạt động liên quan tới các chuyển đổi theo một hướng như đi bộ, chạy và các hoạt động chuyển đổi nhiều hướng như leo cầu thang [117]. Một số đặc trưng hiệu quả khác, như tỷ lệ vượt qua điểm không ZCR (zero crossing rate), tỷ lệ vượt qua trung bình MCR (mean crossing rate), và first order derivative, là các đặc trưng thống kê quan trọng hay được sử dụng trong các bài toán xử lý dữ liệu chuỗi thời gian, ví dụ như nhận dạng tiếng nói [149].

Một tập đặc trưng khác được giới thiệu trong [150] được chứng minh hiệu quả hơn các đặc trưng thống kê khác trong phân biệt các chuyển động đi lại cơ bản gọi là *đặc trưng vật lý*. Tên gọi này dựa trên các giải thích vật lý về chuyển động

của con người. Các đặc trưng vật lý được tính toán khác với cách tính đặc trưng thống kê. Đối với đặc trưng thống kê, mỗi đặc trưng được trích ra từ một trục cảm biến (kênh) riêng. Trong khi đó, hầu hết các đặc trưng vật lý được trích ra từ nhiều kênh cảm biến. Nói cách khác, sự kết hợp cảm biến được thực hiện ở mức đặc trưng đối với các đặc trưng vật lý. Trong nghiên cứu của mình, tác giả đã chỉ ra rằng tập các đặc trưng vật lý có thể phân biệt tốt hơn tập các đặc trưng thống kê rất phức tạp cho các chuyển động như *đứng, ngồi, đi các hướng, đi lên cầu thang, xuống cầu thang, chạy và nhảy*.

Các đặc trưng trong miền tần số cũng được sử dụng cho nhận dạng hoạt động người và thường dựa vào chuyển đổi Fourier [17,34,106] hay Cô-sin rời rạc tính toán trên các khung hoạt động [9]. Trước đó, Huynh [58] cũng đã thực hiện đánh giá các mô tả đặc trưng dựa trên thống kê và đặc trưng trong miền tần số bằng thực nghiệm và cho rằng các đặc trưng sử dụng hệ số Fourier thậm chí còn tốt hơn các đặc trưng thống kê khi phân biệt các hoạt động có các mẫu chuyển động lặp lại như chạy hay đạp xe.

Các đặc trưng đã nêu trên không phải lúc nào cũng có thể sử dụng tốt để phân biệt các hoạt động đơn giản. Trong trường hợp một bài toán đã được quan tâm từ trước đó rất lâu là phân biệt *ngã* và các hoạt động khác thì các đặc trưng đơn giản trên lại không phải rất hiệu quả [38]. Lý do là *ngã* và một số hoạt động tương tự *ngã* khác như *đứng ngồi, ngồi nằm, nhảy* có tín hiệu có tính chất tương tự nhau và dễ gây nhầm lẫn [1,84]. Cách tiếp cận truyền thống dựa trên *ngưỡng* có khả năng phát hiện *ngã* nhanh, đơn giản và tương đối tốt, tuy vậy thường bị tỉ lệ cảnh báo giả cao và trong nhiều tình huống lại không hiệu quả, dẫn tới khả năng khái quát hóa thấp [60]. Để tăng tính khái quát hóa và hiệu quả phân biệt *ngã* thì các cách tiếp cận dựa trên học máy gần đây như [7,69,106] đề xuất sử dụng các tập gồm nhiều đặc trưng thống kê phức tạp hơn, bao gồm cả skewness, kurtosis, tự tương quan, hệ số Fourier. Tuy vậy, độ phức tạp của tập đặc trưng lớn dẫn tới tính khả thi của các phương pháp này bị hạn chế do khó có thể tính toán nhanh trên hệ thống có năng lực xử lý thấp và thời lượng pin không nhiều như đa số các thiết bị trợ giúp cá nhân

đi động thông minh mang theo người, đặc biệt là các hệ thống đòi hỏi thời gian thực. Do đó, chúng khó có thể ứng dụng được trong thực tế hiện nay, khi người dùng cần sự thuận tiện cao với các thiết bị di động trợ giúp cá nhân thông minh, nhỏ gọn. Đây chính là một vấn đề tồn tại trong nghiên cứu nhận dạng hoạt động sử dụng cảm biến mang trên người, đó là cần phải có các *phương pháp biểu diễn đặc trưng mới, hiệu quả có thể phân biệt tốt ngã và các hoạt động khác có đặc tính dữ liệu tương tự, có khả năng tính toán nhanh để chạy được trên các hệ thống yêu cầu xử lý theo thời gian thực và bị hạn chế về tài nguyên*. Vấn đề tồn tại này sẽ được bàn luận và giải quyết trong chương 2.

Phần dưới đây liệt kê một số đặc trưng thường được sử dụng nhất đối với một tín hiệu $S = \{s_1, \dots, s_n\}$, với s_i là một giá trị đầu ra cảm biến tại một thời điểm ($i = 1, \dots, n$). Các đặc trưng này cũng được sử dụng trong các chương tiếp theo như trong phần so sánh các phương pháp trích chọn đặc trưng trong chương 2 (trong phương pháp của Pham [110], các đặc trưng cục bộ trong chương 3 (tập 5 đặc trưng thống kê và tập các đặc trưng vật lý).

- Độ đo hướng tâm như trung bình số học \bar{s} và trung bình bình phương RMS.

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i \quad (1.1)$$

$$RMS(S) = \sqrt{\frac{1}{n} \sum_{i=1}^n s_i^2} \quad (1.2)$$

- Độ đo phân tán như độ lệch chuẩn σ_s , phương sai σ_s^2 , và độ lệch tuyệt đối trung bình (MAD).

$$\sigma_s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (s_i - \bar{s})^2} \quad (1.3)$$

$$\sigma_s^2 = \frac{1}{n-1} \sum_{i=1}^n (s_i - \bar{s})^2 \quad (1.4)$$

$$MAD(S) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n |s_i - \bar{s}|} \quad (1.5)$$

- Các độ đo chuyển đổi miền, chẳng hạn như năng lượng theo FFT, trong đó nơi F_i là thành phần thứ i của biến đổi Fourier của S .

$$Energy(S) = \frac{\sum_{i=1}^n F_i^2}{n} \quad (1.6)$$

- Các độ đo thống kê khác như:
 - Trung bình đạo hàm (Averaged derivatives): Là giá trị trung bình của các đạo hàm đầu tiên của tín hiệu trong cửa sổ thời gian.
 - MCR (Mean crossing rate): Số lần tín hiệu vượt qua đường trung bình.
 - ZCR (Zero crossing rate): Số lần tín hiệu chuyển từ âm sang dương và ngược lại.
 - Khoảng tứ phân vị (Interquartile range): độ khác biệt giữa khoảng 75% và khoảng 25%.
 - Entropy: $H_y = -\sum_i^N p(s_i) \log(p(s_i))$, với $p(s_i)$ là phân bố xác suất của s_i trong cửa sổ thời gian, tính bằng số các s_i trong cửa sổ thời gian chia cho n .
 - Tương quan giữa mỗi cặp trục:

$$\delta_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{s}_x)(y_i - \bar{s}_y)}{N\sigma_x\sigma_y} \quad (1.7)$$

- Skewness: đo độ bất đối xứng của phân phối xác suất tín hiệu.

$$Sk(S) = \sqrt{T} \frac{\sum_{i=1}^r (s_i - \bar{s})^3}{\left(\sum_{i=1}^r (s_i - \bar{s})^2\right)^{3/2}} \quad (1.8)$$

- Độ nhọn (Kurtosis): độ nhọn của phân phối xác suất tín hiệu.

$$Kur(S) = T \frac{\sum_{i=1}^T (s_i - \bar{s})^4}{\left(\sum_{i=1}^T (s_i - \bar{s})^2\right)^2} - 3 \quad (1.9)$$

- Các đặc trưng vật lý:

- Cường độ chuyển động (Movement Intensity - MI):

$$MI(t) = s_x(t)^2 + s_y(t)^2 + s_z(t)^2 \quad (1.10)$$

Đây là khoảng cách O-clit của tổng các véc-tơ gia tốc sau khi loại bỏ gia tốc trọng trường tĩnh. Trong đó, $s_x(t)$, $s_y(t)$, và $s_z(t)$ đại diện cho các mẫu gia tốc thứ t của các trục x , y , z tương ứng trong mỗi cửa sổ thời gian. Đặc trưng này độc lập với hướng của các thiết bị cảm biến và dùng để đo cường độ tức thời của chuyển động con người tại thời điểm t . Cường độ chuyển động không được sử dụng trực tiếp. Thay vào đó, hai đặc trưng gồm giá trị trung bình (Average of MI - AI) và phương sai (Variance of MI) của cường độ chuyển động qua cửa sổ (độ dài T) được sử dụng và được tính như sau:

$$AI = \frac{1}{T} \left(\sum_{t=1}^T MI(t) \right) \quad (1.11)$$

$$VI = \frac{1}{T} \left(\sum_{t=1}^T (MI(t) - AI)^2 \right) \quad (1.12)$$

- Diện tích của độ lớn tín hiệu chuẩn hóa (Normalized Signal Magnitude Area - SMA):

$$SMA = \frac{1}{T} \left(\sum_{t=1}^T |s_x(t)| + \sum_{t=1}^T |s_y(t)| + \sum_{t=1}^T |s_z(t)| \right) \quad (1.13)$$

Đây là tổng của độ lớn gia tốc trên ba trục trong mỗi cửa sổ, được chuẩn hóa theo độ dài cửa sổ. Đặc trưng này được coi là ước lượng gián tiếp của tiêu hao năng lượng.

- Trị riêng của các hướng chính (Eigenvalues of Dominant Directions): Khi một đối tượng nhảy, dự kiến sẽ có một thành phần gia tốc tăng lên theo hướng thẳng đứng. Tương tự như vậy, khi một đối tượng chạy về phía trước thì một thành phần gia tốc tăng lên theo hướng chạy và một thành phần gia tốc cũng tương đối lớn theo hướng thẳng đứng. Để có được những hiệu ứng này cần tính ma trận hiệp phương sai của dữ liệu gia tốc dọc theo 3 trục x , y , z trong mỗi cửa sổ để từ đó tính được các véc-tơ riêng của ma trận hiệp phương sai tương ứng với các hướng chính chuyển động của con người. Các giá trị riêng đo được cường độ chuyển động tương đối tương ứng theo các hướng. Hai giá trị riêng đầu được đề xuất sử dụng làm đặc trưng, tương ứng với độ lớn chuyển động tương đối theo hướng thẳng đứng và hướng di chuyển.
- Tương quan giữa gia tốc theo lực hấp dẫn và hướng di chuyển (Correlation between Acceleration along Gravity and Heading Directions): Dựa vào các vị trí và hướng của các thiết bị cảm biến được mô tả, hướng trọng lực sẽ gần song song với trục x , và hướng di chuyển của đối tượng khi đi bộ là kết hợp các trục y và z . Đầu tiên lấy được khoảng cách O-clit của tổng các véc-tơ gia tốc theo hướng di chuyển, và sau đó tính toán các hệ số tương quan giữa các gia tốc theo hướng trọng lực và gia tốc có được theo hướng di chuyển để tạo ra đặc trưng.
- Vận tốc trung bình theo hướng di chuyển (Averaged Velocity along Heading Direction): được xấp xỉ bằng cách sau: ban đầu tính toán vận tốc trung bình theo trục y và z trong cửa sổ, sau đó tính toán khoảng cách O-clit của hai vận tốc.
- Vận tốc trung bình theo hướng trọng lực (Averaged Velocity along Gravity Direction): được tính bằng trung bình cộng vận tốc tức thời theo hướng trọng lực tại mỗi thời điểm t trong cửa sổ. Vận tốc tức thời tại mỗi thời điểm t được tính bằng tích phân của gia tốc theo hướng trọng lực.

- Trung bình góc quay theo hướng trọng lực (Averaged Rotation Angles related to Gravity Direction): tính các góc quay tích lũy xung quanh hướng trọng lực. Tổng tích lũy sau đó được chuẩn hóa theo độ dài cửa sổ. Nếu cảm biến gắn ở phía trước hông phải của đối tượng, đặc trưng này mô tả góc quay của thân người xung quanh hướng trọng lực.
- Tần số chính (Dominant Frequency): Tần số chính được định nghĩa là tần số tương ứng với cực đại của căn bậc hai các độ lớn thành phần FFT rời rạc của tín hiệu từ mỗi trục cảm biến.
- Năng lượng (Energy): Năng lượng được tính bằng tổng của căn bậc hai độ lớn các thành phần FFT rời rạc của tín hiệu từ mỗi trục cảm biến. Tổng này sau đó được chuẩn hóa theo độ dài cửa sổ.
- Trung bình năng lượng gia tốc (Averaged Acceleration Energy): được định nghĩa là giá trị trung bình của năng lượng trên ba trục gia tốc.
- Trung bình năng lượng quay (Averaged Rotation Energy): được định nghĩa là giá trị trung bình của năng lượng trên ba trục của con quay hồi chuyển.

1.4.2. Trích xuất đặc trưng tự động bằng học đặc trưng

Các đặc trưng dựa trên tri thức chuyên gia thường được sử dụng trong các hệ thống nhận dạng hoạt động do tính đơn giản và hiệu quả. Các đặc trưng dạng này thường được thiết kế dựa trên kết quả của một quá trình phân tích kỹ lưỡng về bản chất của dữ liệu để từ đó xây dựng nên một biểu diễn đặc trưng có khả năng nắm bắt được các đặc tính dữ liệu của ứng dụng. Tuy nhiên, chúng chỉ hoạt động hiệu quả khi ứng dụng để phân biệt một số hoạt động đơn giản, cụ thể mà không đáp ứng được khi cần phân biệt tập các hoạt động ADL nhiều hơn và phức tạp hơn. Ví dụ như các đặc trưng thống kê đơn giản như trung bình, độ lệch chuẩn hay phương sai có thể phân biệt tốt các hoạt động *đi bộ*, *chạy* và *đứng yên* nhưng khi tích hợp thêm các hoạt động phức tạp như *xem ti vi*, *ăn*, *đánh răng*, *đi lên cầu thang*, *đi xuống cầu*

thang thì tập đặc trưng trên không có khả năng nhận dạng tốt. Một cách tiếp cận cho vấn đề này là sử dụng một tổ hợp các đặc trưng thống kê phức tạp hơn. Chẳng hạn như trong [150], Zhang sử dụng các đặc trưng bao gồm 13 đặc trưng thống kê, trong đó có cả năng lượng, entropy, các hệ số tương quan, skewness, kurtosis để phân biệt 7 hoạt động. Pham [110] sử dụng tập gồm 7 đặc trưng, bao gồm năng lượng các hệ số tương quan, skewness, kurtosis để phân biệt 14 hoạt động ADL phức tạp và đạt kết quả tương đối tốt.

Một điểm yếu dễ nhận thấy của cách tiếp cận sử dụng các đặc trưng dựa trên tri thức chuyên gia như trên là, mỗi một tập các hoạt động khác nhau cần phân biệt trong mỗi ứng dụng cụ thể lại yêu cầu một tập các đặc trưng riêng được lựa chọn, thiết kế kỹ lưỡng để sao cho phù hợp, hiệu quả đối với tập các hoạt động đang xem xét [78]. Yêu cầu này làm tăng chi phí và gây khó khăn cho việc xây dựng hệ thống nhận dạng hoạt động người. Nhiều nhà nghiên cứu đã tìm cách cải tiến các phương pháp nhận dạng để có được độ chính xác cao hơn, hiệu quả với nhiều loại hoạt động người và hỗ trợ được nhiều ứng dụng nhận dạng hoạt động. Một cách tiếp cận hiệu quả là sử dụng học đặc trưng, tức là tìm cách tối ưu hóa một hàm mục tiêu để “bắt được” sự phù hợp của các đặc trưng cho mọi dữ liệu, đồng thời có thể tự động khám phá các biểu diễn có ý nghĩa của dữ liệu được phân tích. Cách tiếp cận này có thể tìm được một biểu diễn đặc trưng chung cho mọi dữ liệu, đồng thời rất khả quan trong việc cải thiện độ chính xác của nhận dạng. Một số phương pháp có thể kể đến bao gồm tối thiểu hóa năng lượng [80] và học sâu [79]. Các phương pháp khác được thử nghiệm và có kết quả khả quan trong nhận dạng hoạt động người như: PCA [51], ECDF dựa trên PCA [114], học đặc trưng sử dụng các kỹ thuật học sâu với RBM [114], CNN [148] và học đặc trưng dựa vào phương pháp túi từ [55,71,151]. Hiện tại, các nghiên cứu nhận dạng hoạt động cho kết quả tốt nhất sử dụng phương pháp học đặc trưng ECDF dựa trên PCA [114] và học đặc trưng dựa trên học sâu như CNN [148]. Phần tiếp theo đây sẽ mô tả về các kỹ thuật này.

a) Học đặc trưng với PCA và ECDF

Phương pháp phân tích thành phần chính PCA là một kỹ thuật được tạo ra nhằm khử tương quan và giảm chiều dữ liệu. PCA là một hình thức cơ bản của học đặc trưng với khả năng tự động phát hiện các biểu diễn nhỏ gọn và có ý nghĩa của dữ liệu thô mà không dựa vào tri thức chuyên gia hay tri thức về kiểu ứng dụng cụ thể. Các véc-tơ riêng của hiệp phương sai cho một tập mẫu (tương ứng với giá trị riêng lớn nhất) được sử dụng để mở rộng thành một không gian phụ có số chiều thấp hơn, tập trung vào phương sai của dữ liệu gốc. Việc chiếu dữ liệu gốc lên không gian con với phương sai cực đại được sử dụng làm biểu diễn đặc trưng.

Biểu diễn dữ liệu cảm biến dựa trên ECDF: PCA thường kém hiệu quả khi các dữ liệu đầu vào không được chuẩn hóa. Tuy nhiên, việc chuẩn hóa trên một dải giá trị chưa biết thường gây ra nhiều vấn đề khi các thành phần liên quan đến các khía cạnh hoàn toàn khác nhau của một hiện tượng; cụ thể là việc sử dụng khung có kích thước lớn sẽ không hiệu quả trong trường hợp nhận dạng hoạt động người. Một cách biểu diễn dữ liệu thô dựa trên hàm phân phối tích lũy thực nghiệm (ECDF) của dữ liệu mẫu được đưa ra với ý tưởng là xây dựng một biểu diễn cho dữ liệu đầu vào sao cho độc lập với các phạm vi tuyệt đối mà vẫn giữ được thông tin về cấu trúc. Phương pháp này được mô tả như dưới đây.

Đối với mỗi khung hoạt động (frame) $\vec{f} = (\vec{x}_i)^T$, lấy được các hàm phân phối tích lũy thực nghiệm ECDF ε_i của các mẫu \vec{x}_i (đã thực hiện whitened) dọc theo mỗi trục $i = \{1,2,3\}$ sử dụng ước lượng chuẩn Kaplan-Meier [36]. Những hàm ECDF này đơn điệu tăng trong phạm vi $[0,1]$, và mô tả xác suất dữ liệu cảm biến nhỏ hơn hoặc bằng một giá trị cụ thể. Nhờ phép nội suy bậc ba $C^{\bar{p}}$, tính giá trị của các nghịch đảo ECDF ε_i^{-1} trên một tập cố định N điểm $\vec{p} = \{p_1, \dots, p_N\}$. Các giá trị này được coi là biểu diễn của dữ liệu cảm biến \vec{x}_i . Trong phương pháp này, dữ liệu mẫu được chuẩn hóa thành một phạm vi chung mà không phá hủy cấu trúc phụ thuộc vốn có. Với $i = \{1,2,3\}$, có được:

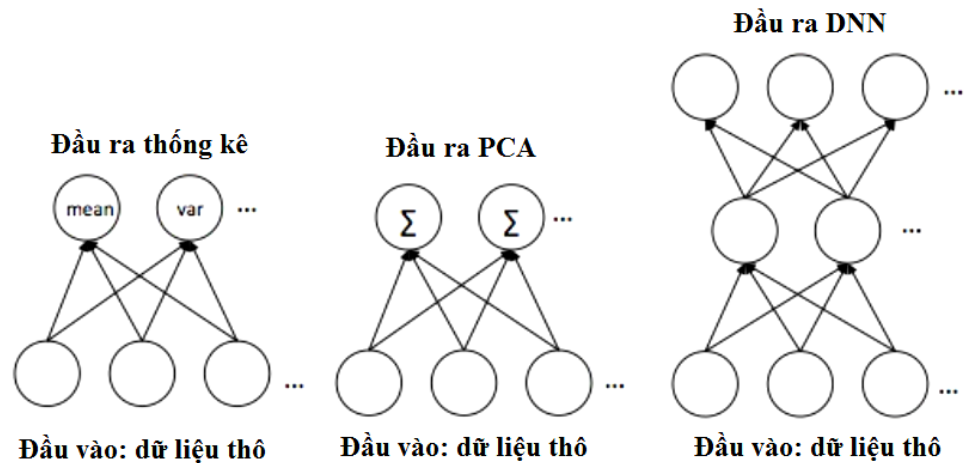
$$\vec{x}_i = C^{\bar{p}}(\varepsilon_i^{-1}(x_i)) \in \mathbb{R}^N \quad (1.14)$$

với $\varepsilon_i^{-1} = x_i \mapsto F_{X_i}(x_i) = P(X_i \leq x_i)$ và $C^{\bar{p}}$ là phép nội suy bậc ba sử dụng $\bar{p} = \{p_1, \dots, p_N\}$

b) Học đặc trưng dựa trên kỹ thuật học sâu

Mặc dù PCA có thể học đặc trưng theo cách không giám sát nhưng nó chỉ là kết hợp tuyến tính của các đặc trưng chưa xử lý. Do đó, PCA không đủ khả năng mô hình hóa cho các phụ thuộc phi tuyến phức tạp [18]. Plotz và cộng sự đã đề xuất sử dụng mạng nơ-ron sâu (DNN) [114] để tách được các đặc trưng có ý nghĩa hơn. Một trong những khác biệt chính giữa mạng nơ-ron truyền thống và DNN là DNN có thể có nhiều lớp, trong khi mạng nơ-ron truyền thống thường sử dụng nhiều nhất 3 lớp. Một lợi thế quan trọng khác nữa của DNN là khả năng biểu diễn các dữ liệu đầu vào. DNN có thể mô hình hóa nhiều hoạt động đa dạng mà không cần nhiều dữ liệu huấn luyện. Nó có thể chia sẻ các phần không gian đầu vào tương tự chỉ với một vài nút ẩn, trong khi giữ cho các nút khác “cảm nhận” một tập con các đặc trưng đầu vào quan trọng cho việc nhận dạng [114].

Gần đây, DNN đạt được các đột phá trong nhiều lĩnh vực nghiên cứu. Những cấu trúc sâu có thể diễn tả các hàm phức tạp một cách ngắn gọn, thể hiện những kết quả vượt trội so với các thuật toán tiên tiến nhất trong nhiều ứng dụng như nhận dạng mặt người, nhận dạng tiếng nói [18]. Hình 1.3 dưới đây so sánh DNN với các kỹ thuật khác.



Hình 1.3. (trái) tính đặc trưng thống kê; (giữa) mô hình PCA; (phải) mô hình DNN (tham khảo [148])

Trong nghiên cứu [114], Plotz đưa ra một cấu trúc mạng DNN dựa trên RBM (Restricted Boltzmann Machine). Việc huấn luyện mạng DNN này được thực hiện theo cách có giám sát. Cách huấn luyện này khác với cách huấn luyện mô hình học sâu khác trong [19,142] là Shift-Invariant Sparse Coding, sử dụng học không giám sát để huấn luyện. Hạn chế của RBM [114] và Shift-Invariant Sparse Coding [19,142] là các mô hình DNN kết nối đầy đủ, do vậy không bắt giữ được các phụ thuộc cục bộ của các tín hiệu chuỗi thời gian [3] nên kết quả đạt được còn hạn chế, kém hơn phương pháp sử dụng đặc trưng ECDF kết hợp PCA [114]. Zeng và cộng sự [148] đã đề xuất sử dụng CNN để khắc phục vấn đề này. CNN bao gồm một hoặc nhiều các cặp thuộc lớp convolution và pooling. Các bộ phận nhỏ cục bộ của đầu vào được bắt giữ bởi các lớp convolution với một tập các bộ lọc cục bộ. Lớp pooling có thể lưu giữ được các đặc trưng bất biến. Cuối cùng, lớp kết nối đầy đủ đứng đầu sẽ kết hợp các đầu vào từ tất cả các đặc trưng để thực hiện phân loại toàn bộ đầu vào. Kết quả đạt được là rất khả quan do Zeng đã cải tiến cấu trúc của CNN dựa vào một kỹ thuật gọi là *partial weight sharing* nhằm cải thiện hiệu năng nhận dạng hoạt động cho các đầu vào là chuỗi gia tốc.

Một nghiên cứu khác sử dụng DNN của Nils [102] thử nghiệm một số tiếp cận học sâu gọi là “deep, convolutional, and recurrent network” và cũng đạt được

kết quả tốt trên các tập dữ liệu khó. Nghiên cứu này đã khảo sát sự phù hợp của mỗi mô hình cho từng ứng dụng nhận dạng hoạt động, thử nghiệm ảnh hưởng của các tham số sử dụng khung làm việc fANOVA và đưa ra được các khuyến nghị cho việc ứng dụng học sâu vào nhận dạng hoạt động sử dụng cảm biến mang trên người.

Mặc dù học đặc trưng dựa trên học sâu (deep learning) đã mang lại được những kết quả rất đáng kể trong việc cải thiện hiệu năng nhận dạng hoạt động người, việc đưa học sâu vào các ứng dụng nhận dạng hoạt động thực tế trên các thiết bị di động thông minh để trợ giúp người dùng vẫn rất khó khăn do các hạn chế về năng lực xử lý của các thiết bị này. Hầu hết các ứng dụng đều hoạt động dưới dạng dịch vụ đám mây chứ chưa thực sự chạy trực tiếp trên thiết bị. Hiện nay, mới chỉ có một số rất ít các thử nghiệm chạy hoạt động trên các thiết bị di động trang bị phần cứng tương đối tốt, ví dụ như thử nghiệm của Lane trong [77].

c) Học đặc trưng dựa trên phương pháp túi từ

Một phương pháp học đặc trưng khác trong nhận dạng hoạt động là kỹ thuật dựa trên mô hình túi từ BoW (bag-of-word) hay túi đặc trưng BoF (bag-of-feature). Phương pháp này được sử dụng phổ biến trong lĩnh vực phân loại văn bản và phân loại ảnh. Trong nhận dạng hoạt động, học đặc trưng dựa trên phương pháp túi từ tạo ra các đặc trưng phức tạp gọi là đặc trưng đa mức. Có thể hiểu rằng các đặc trưng này được xác định bằng tổ hợp của các đặc trưng cơ sở của chuyển động cơ thể người. Đặc trưng đa mức không quá đòi hỏi năng lực xử lý cao như các đặc trưng dựa trên học sâu.

Đặc trưng đa mức được tính thông qua các bước: ban đầu dữ liệu được phân cụm sử dụng một phương pháp phân cụm (như k -means) để tính các thống kê sự xuất hiện của nhóm trên một cửa sổ trượt, sau đó các đặc trưng có ý nghĩa được tạo nên thông qua các tính toán về quãng thời gian, tần số và sự kiện đồng xảy ra từ dữ liệu. Các mô hình BoW này cho nhận dạng hoạt động thường được xây dựng theo ba bước sau: trích xuất đặc trưng, học từ vựng (vocabulary hay codebook) và cuối cùng là xây dựng phân phối tần suất (histogram) của các từ (chuyển động). Các đặc

trung được trích xuất thường là các đặc trưng cục bộ, được tính toán từ các phân đoạn nhỏ của cửa sổ thời gian. Từ vựng có thể được học bằng cách sử dụng thuật toán phân cụm không giám sát như k -means [55,71,151] hoặc Gaussian Mixture Model (GMM) [44,151] để nhóm các đặc trưng thành các cụm (cluster). Trung tâm của cụm tạo thành một từ (hay véc-tơ mã/code vector), và một tập các từ tạo ra từ tập dữ liệu tạo thành từ vựng. Tần suất xuất hiện của các từ có trong từ vựng tạo nên biểu đồ tần suất của các từ.

Các đặc trưng đa mức có được thông qua học đặc trưng dựa trên mô hình BoW như trên cũng đã đạt được những kết quả khả quan trong việc cải thiện độ chính xác nhận dạng. Tuy nhiên việc sử dụng các phương pháp phân cụm không giám sát như k -means hoặc GMM trong các nghiên cứu về nhận dạng hoạt động sử dụng cảm biến mang trên người [55,151] để xây dựng từ vựng *có các hạn chế về độ phức tạp tốc độ tính toán khi dữ liệu đầu vào có số chiều lớn*, gây ra khó khăn cho việc áp dụng trong thực tế. Vấn đề này sẽ được phân tích và giải quyết trong chương 3, với đề xuất của tác giả về sử dụng các đặc trưng MPF để tăng tốc độ tính toán. Kết quả thử nghiệm, đánh giá trong chương 3 cho thấy các đặc trưng đề xuất này không những có tốc độ tính toán nhanh mà còn giúp hệ thống nhận dạng chính xác hơn các phương pháp khác có kết quả nhận dạng hoạt động tốt nhất hiện nay dựa trên các đặc trưng ECDF [114] và CNN [148].

1.5. KẾT LUẬN CHƯƠNG

Chương này đã trình bày giới thiệu tổng quan về các khía cạnh của nghiên cứu nhận dạng hoạt động sử dụng cảm biến mang trên người, bao gồm các khái niệm chung, các cách tiếp cận và ứng dụng của nhận dạng hoạt động, một số khó khăn trong nghiên cứu, các loại cảm biến sử dụng, kiến trúc và các bước thực hiện trong một hệ thống nhận dạng hoạt động điển hình. Ngoài ra, các phương pháp trích xuất đặc trưng cho nhận dạng hoạt động người cũng được khảo sát kỹ lưỡng với một loạt các công trình nghiên cứu gần đây.

Nội dung đã trình bày trong chương cho thấy còn rất nhiều vấn đề khó khăn trong nghiên cứu nhận dạng hoạt động người. Tuy nhiên, luận án sẽ tập trung vào việc đề xuất các phương pháp trích xuất đặc trưng để giải quyết hai vấn đề quan trọng, cụ thể như sau:

- Nghiên cứu và đề xuất phương pháp trích xuất đặc trưng để giải quyết vấn đề *đảm bảo nhận dạng hoạt động chính xác cho hệ thống yêu cầu xử lý theo thời gian thực, bị hạn chế về tài nguyên*. Phương pháp này có tốc độ tính toán nhanh, ứng dụng để phân biệt ngã và các hoạt động thường ngày khác, sử dụng cảm biến mang theo người với yêu cầu xử lý theo thời gian thực. Cụ thể, phương pháp đề xuất được trình bày trong Chương 2.
- Nghiên cứu và đề xuất phương pháp trích xuất đặc trưng để giải quyết vấn đề *xây dựng phương pháp nhận dạng hiệu quả cho nhiều loại hoạt động người*. Nghĩa là, phương pháp trích xuất đặc trưng đề xuất tạo ra các đặc trưng phù hợp với nhiều kiểu dữ liệu, giúp nâng cao độ chính xác của hệ thống nhận dạng hoạt động kể cả khi số hoạt động nhiều. Phương pháp đề xuất được trình bày trong Chương 3.

CHƯƠNG 2

NHẬN DẠNG HOẠT ĐỘNG VỚI CÁC ĐẶC TRƯNG HALF

Chương này trình bày các đặc trưng mới và hiệu quả cho nhận dạng hoạt động người, gọi là HALF (Histograms of Angles of Line Fragments between Two Consecutive Points – Phân phối tần suất góc của các đoạn nối hai điểm dữ liệu kề), với mục tiêu giải quyết vấn đề là *đảm bảo nhận dạng hoạt động chính xác cho các hệ thống yêu cầu xử lý theo thời gian thực và bị hạn chế về tài nguyên*. Phương pháp trích xuất đặc trưng hoạt động nhanh và hiệu quả trong các ứng dụng nhận dạng một số hoạt động riêng lẻ, đặc biệt là phân biệt các hoạt động ngã trong hệ thống sử dụng cảm biến mang theo người.

Các đặc trưng này được trích xuất thông qua việc tính toán các phân phối tần suất góc của các mảnh nhỏ giữa hai điểm dữ liệu liên tiếp, trên các đoạn (segment) nằm trong cửa sổ trượt, trên luồng dữ liệu gia tốc liên tục của cảm biến khi biểu diễn dưới dạng hình học theo biểu đồ đường (line chart). Các *ngăn* (bin) của phân phối tần suất được tính toán tự động hoặc dựa trên cách tiếp cận heuristic. Việc tính toán *ngăn* tự động có thể được thực hiện dựa trên các cụm (cluster) góc của các *mảnh* (quantum), nhờ vậy sẽ giảm bớt sự ảnh hưởng của các tham số tới việc lựa chọn *ngăn*. Đặc trưng của một cửa sổ hoạt động/khung được tính toán bằng việc gộp các đặc trưng cục bộ thành véc-tơ đặc trưng tổng. Cuối cùng, các đặc trưng được sử dụng với máy véc-tơ tựa (SVM) đã huấn luyện để phân loại các khung của các luồng dữ liệu thành các loại hoạt động. Trong chương này, các đặc trưng đề xuất được áp dụng cho một ứng dụng cụ thể là phát hiện ngã. Việc đánh giá hiệu năng của phương pháp trích xuất đặc trưng được thực hiện dựa trên 3 tập dữ liệu đã được công bố, chứa các dữ liệu gia tốc cho hoạt động ngã và các hoạt động sống hàng ngày (ADL) khác. Kết quả cho thấy, hiệu năng của phương pháp đề xuất cao hơn đáng kể so với các phương pháp phát hiện ngã hiệu quả khác.

2.1. ĐẶT VẤN ĐỀ

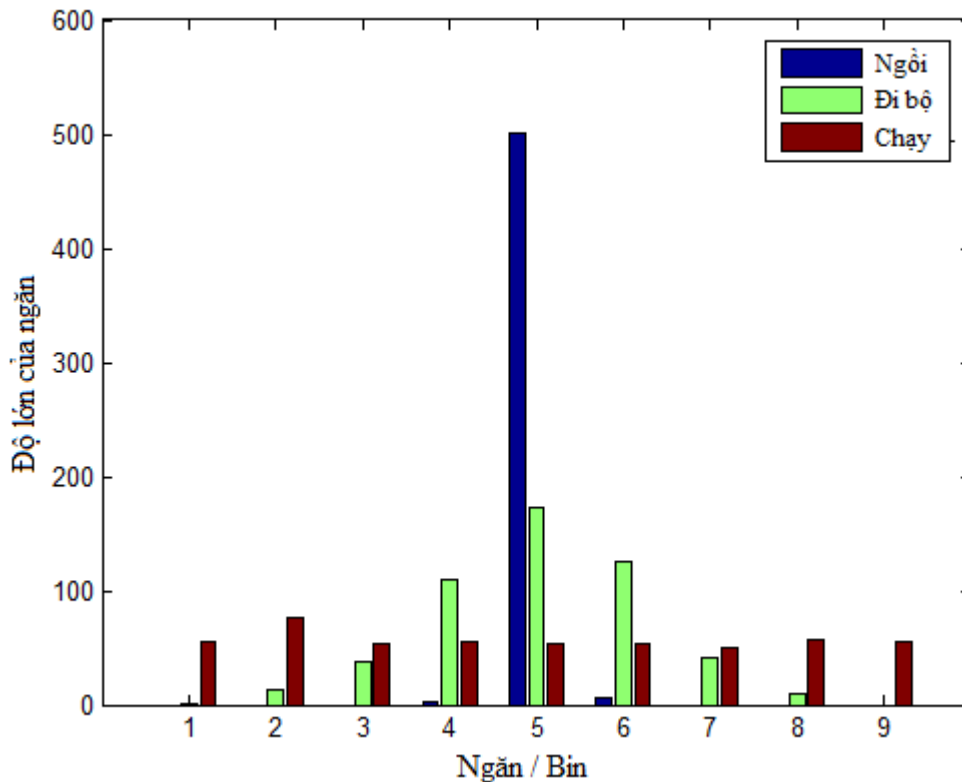
Nhiều nghiên cứu về nhận dạng hoạt động của người đã sử dụng các đặc trưng thống kê (như thấy trong chương 1) do tính đơn giản và dễ sử dụng của chúng, nhất là trong các hoạt động đơn giản và phổ biến trong cuộc sống như ngồi-đứng, đứng-ngồi, leo cầu thang, đi bộ, v.v. Nhiều đặc trưng đã thể hiện được ảnh hưởng quan trọng trong việc phân biệt các hoạt động. Chẳng hạn như phương sai, được chứng minh là có độ chính xác cao khi phân biệt các hoạt động đi bộ, chạy bộ, nhảy [58], hay độ tương quan giữa các cặp trục gia tốc giúp phân biệt tốt các hoạt động liên quan tới các chuyển đổi theo một hướng như đi bộ, chạy và các hoạt động chuyển đổi nhiều hướng như leo cầu thang [117].

Tuy nhiên, các đặc trưng đã nêu trên không phải lúc nào cũng phù hợp để sử dụng và phân biệt tốt các hoạt động đơn giản, ví dụ như *ngã* và các hoạt động khác. Lý do là các hoạt động với các tín hiệu có tính chất tương tự nhau như *ngã*, *đứng ngồi*, *ngồi nằm*, *nhảy*, v.v., rất dễ nhầm lẫn với nhau và các đặc trưng thống kê không có hiệu quả trong việc phân biệt các hoạt động này [1,84]. Trong khi đó cách tiếp cận truyền thống dựa trên *ngưỡng* có khả năng phát hiện ngã nhanh, đơn giản, chính xác thì lại thường có tỉ lệ cảnh báo giả cao và trong nhiều tình huống lại không hiệu quả, dẫn tới khả năng khái quát hóa thấp [60].

Để giải quyết hạn chế trên trong khi vẫn đạt được mục tiêu về độ chính xác cần thiết, các cách tiếp cận dựa trên học máy gần đây như [7,69,106] đề xuất sử dụng các đặc trưng được thiết kế phức tạp hơn (ví dụ như sử dụng skewness, kurtosis, tự tương quan, hệ số Fourier), kết hợp chúng với nhau nhằm đạt được mức độ cải thiện tốt hơn. Việc thiết kế các đặc trưng như vậy thực sự đã mang lại những cải thiện rõ rệt trong việc nhận dạng các hoạt động. Tuy nhiên vấn đề nảy sinh là các đặc trưng phức tạp hay tập hợp nhiều các đặc trưng phức tạp dẫn tới lượng tính toán tăng, không phù hợp khi thuật toán nhận dạng hoạt động phải chạy trên các thiết bị trợ giúp cá nhân di động nhỏ gọn, bộ nhớ và năng lực tính toán thấp, nhất là khi cần phải thực thi cho các ứng dụng cần xử lý thời gian thực. Vấn đề dẫn tới hạn

chế về khả năng ứng dụng trong thực tế hiện nay, khi mà đa phần người dùng đòi hỏi sự thuận tiện cao nhờ các thiết bị di động thông minh nhỏ gọn mang theo người, nhưng đi kèm theo là các giới hạn về pin, bộ nhớ, năng lực xử lý, chẳng hạn như điện thoại thông minh, đồng hồ thông minh hay máy tính bảng.

Để giải quyết các vấn đề trên, chương này trình bày một phương pháp trích xuất đặc trưng mới, đơn giản (không đòi hỏi tính toán phức tạp), đồng thời hiệu quả trong nhận dạng hoạt động người, gọi là HALF. Ý tưởng cơ bản tạo nên đặc trưng HALF được bắt nguồn từ việc quan sát hình dạng của tín hiệu cảm biến gia tốc của các hoạt động trên biểu đồ đường (line chart). Với các loại chuyển động khác nhau, biểu đồ tín hiệu gia tốc sẽ có hình dạng thay đổi khác nhau, trên từng đoạn ngắn nối hai điểm dữ liệu liên tiếp. Vấn đề là làm sao bắt được thông tin về sự khác nhau trong thay đổi gia tốc của từng hoạt động nhanh và hiệu quả. Có thể thấy sự thay đổi gia tốc giữa hai điểm dữ liệu liên tiếp có thể được thể hiện thông qua góc của véc-tơ nối hai điểm dữ liệu này và trục thời gian (xem hình 2.3b). Đặc trưng về sự thay đổi gia tốc trong một cửa sổ thời gian của tín hiệu gia tốc có thể được nắm bắt thông qua thống kê, phân phối về tần suất góc của các đoạn nối hai điểm dữ liệu liên tiếp này. Đặc trưng này có thể được tính toán rất nhanh chóng, hiệu quả và phân biệt được các loại hoạt động khác nhau. Như ví dụ trong hình 2.1 về đặc trưng dựa phân phối tần suất góc của các đoạn dữ liệu gia tốc của các khung tín hiệu cho 3 hoạt động ngồi, đi bộ và chạy như mô tả trong hình 1.2, dựa vào các đặc trưng này có thể dễ dàng phân biệt 3 loại hoạt động.



Hình 2.1. Ví dụ về các phân phối tần suất góc của 3 khung tín hiệu gia tốc cho 3 hoạt động *ngồi, đi bộ, chạy* như mô tả trong hình 1.2, với số ngăn/cụm (bin) là 9, đều nhau cho khoảng giá trị góc $(-90^0, 90^0)$.

Tính hiệu quả của đặc trưng HALF không chỉ nhờ vào cách trích xuất đặc trưng thông qua tính toán phân phối tần suất góc của các mảnh nhỏ trong khung tín hiệu, mà còn bắt nguồn từ khả năng tự động học để tính các ngăn cho phân phối tần suất cùng khả năng nắm giữ các đặc trưng cục bộ của khung tín hiệu đầu vào. Có thể thấy rằng phương pháp trích xuất đặc trưng HALF không bị hạn chế vào phương pháp nhận dạng và học hoạt động hay phụ thuộc vào kiểu cảm biến cụ thể nào. Thậm chí, có thể áp dụng một cách tổng quát cho nhiều loại cảm biến, nhiều kiểu ứng dụng và các ứng dụng thời gian thực trên các hệ thống bị hạn chế về bộ nhớ cũng như năng lực tính toán. Trong chương này, các đặc trưng HALF được áp dụng thử nghiệm cho ứng dụng phân biệt hoạt động ngã sử dụng cảm biến gia tốc đeo trên người và đã cho kết quả tốt trên cả 3 các tập dữ liệu ngã thử nghiệm. Kết

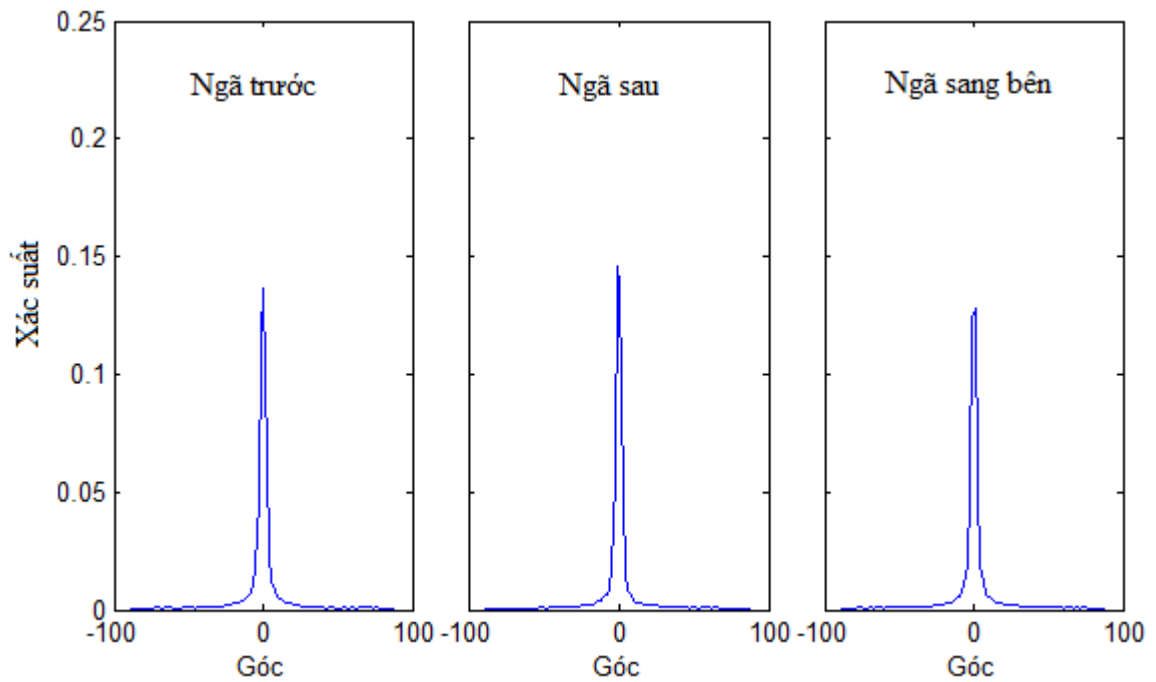
quả có được thể hiện sự cải thiện rõ ràng về hiệu suất so với các phương pháp phát hiện ngã khác có độ chính xác tốt nhất hiện nay.

Các phần tiếp theo trong chương này trước tiên sẽ mô tả về các thống kê phân phối tần suất góc của các hoạt động, sau đó trình bày phương pháp trích xuất đặc trưng, phương pháp phát hiện ngã sử dụng đặc trưng HALF và thực nghiệm.

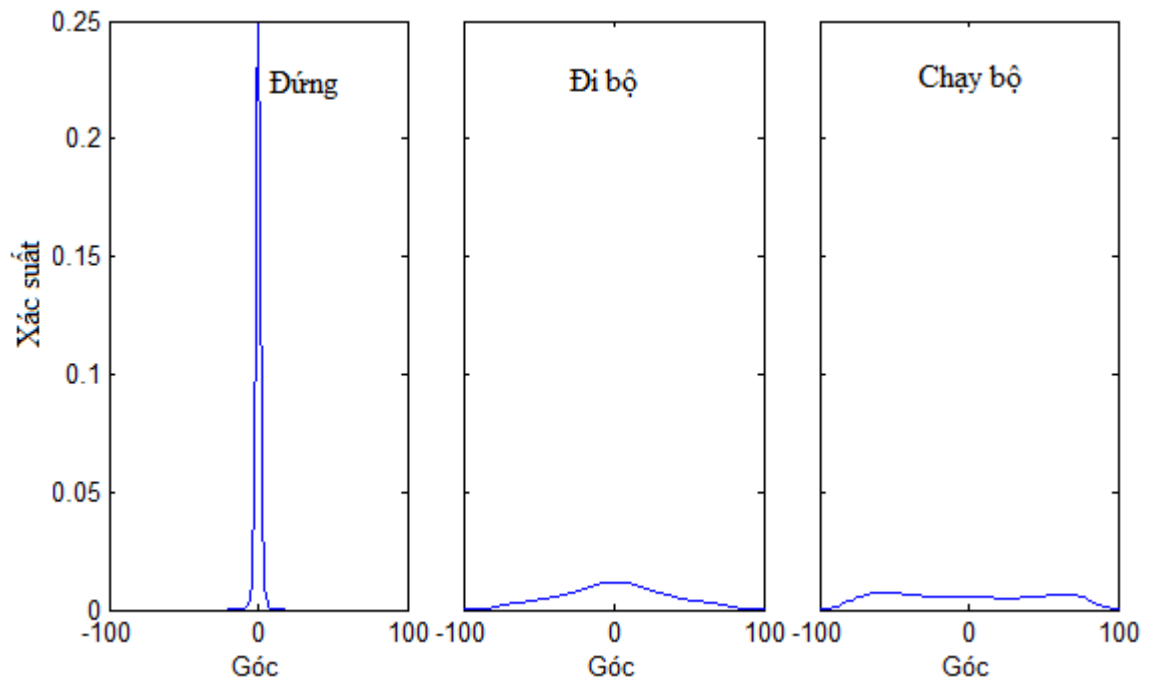
2.2. THỐNG KÊ VỀ PHÂN PHỐI TẦN SUẤT GÓC CỦA CÁC HOẠT ĐỘNG

Trước khi mô tả về phương pháp trích xuất đặc trưng HALF, phần này sẽ xem xét các phân phối xác suất góc cho nhiều loại hoạt động với nguồn dữ liệu lấy từ 3 tập dữ liệu được dùng nhiều trong các nghiên cứu phát hiện ngã như mô tả trong phần 2.4.3, để từ đó có cái nhìn ban đầu về đặc trưng dựa vào phân phối tần suất góc của các mảnh nối hai điểm dữ liệu liên tiếp.

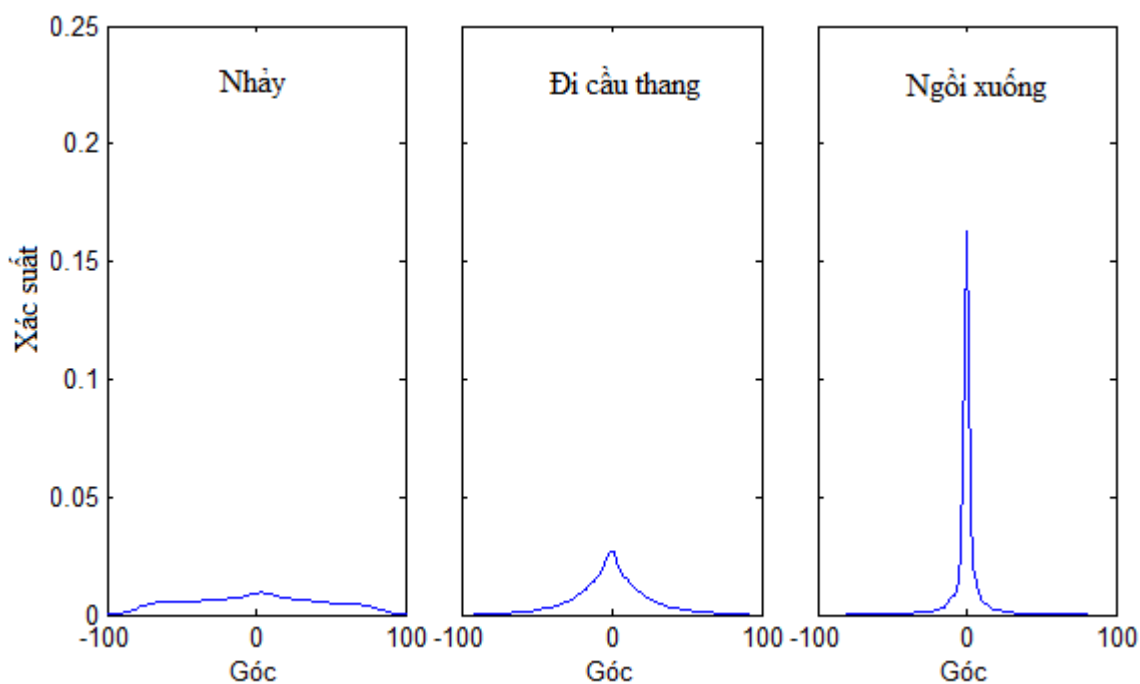
Dữ liệu ngã được phân thành 3 nhóm: ngã trước, ngã sau và ngã sang bên cạnh. Hình 2.2 cho thấy 3 kiểu ngã về cơ bản có phân phối xác suất tương tự nhau. So sánh với 6 hoạt động khác trong hình 2.3 và 2.4, có thể thấy rằng phân phối xác suất theo góc của các hoạt động này khác hẳn nhau, trừ *ngã* và hoạt động *ngồi*. Điều này cho thấy việc sử dụng phân phối tần suất góc của các mảnh tín hiệu nối hai điểm có thể giúp phân biệt được nhiều hoạt động. *Ngồi* và *ngã* có phân phối xác suất khá tương tự nhau do *ngồi* có tính chất tín hiệu gia tốc tương tự ngã. Các nghiên cứu trước như [2,84] cũng chỉ ra rằng đứng-ngồi, ngồi-nằm và một số hoạt động khác có tính chất về tín hiệu gia tốc tương tự *ngã* rất khó để có thể phân biệt được. Để làm được điều này cần tìm ra các đặc trưng mới của ngã khác biệt so với các hoạt động đó. Phần 2.2 tiếp theo sẽ mô tả chi tiết về các ưu điểm của đặc trưng HALF để thấy được bên cạnh khả năng phân biệt ngã với các hoạt động khác, HALF có thể phân biệt tốt *ngã* và các *hoạt động tương tự ngã*.



Hình 2.2. Phân phối xác suất góc với 180 cụm (góc trong khoảng $(-90^0, 90^0)$) cho sự kiện ngã trước, ngã sau và ngã sang bên cạnh



Hình 2.3. Phân phối xác suất góc với 180 cụm (góc trong khoảng $(-90^0, 90^0)$) cho hoạt động đứng, đi bộ và chạy bộ



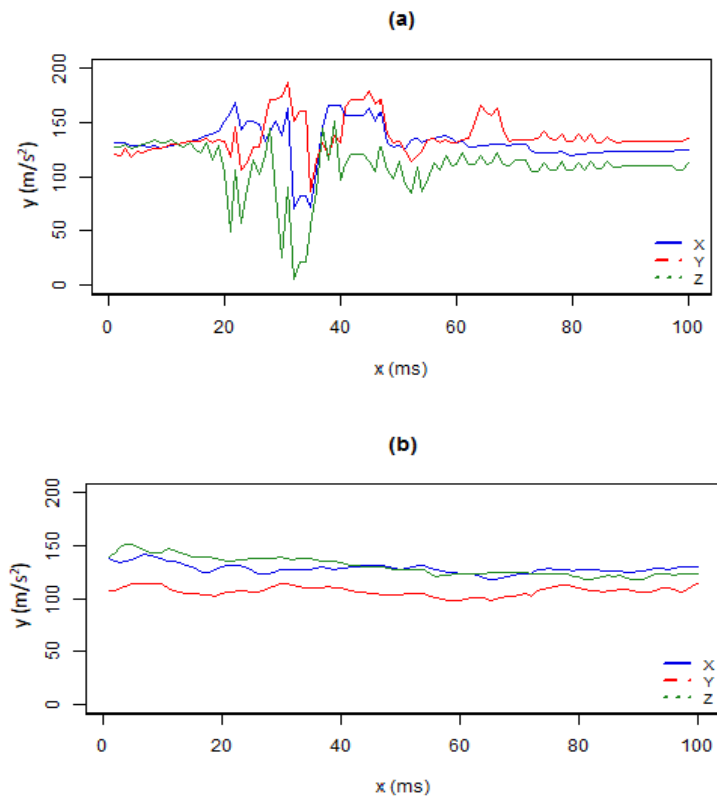
Hình 2.4. Phân phối xác suất góc với 180 cụm (góc trong khoảng $(-90^0, 90^0)$) cho hoạt động nhảy, đi cầu thang và ngồi xuống

2.3. CÁC ĐẶC TRƯNG HALF

Xét dữ liệu đầu hệ thống vào là các luồng dữ liệu gia tốc theo 3 trục x , y , z , thuật toán nhận dạng hoạt động người thông qua các bước sau: phân đoạn dữ liệu, trích xuất đặc trưng và phân loại. Các đặc trưng HALF sẽ được trích xuất trên các khung hay cửa sổ (frame/window) dữ liệu đầu vào có độ dài cố định được tạo ra từ bước phân đoạn dữ liệu trước đó, trên các luồng dữ liệu cảm biến gia tốc ban đầu. Sau đó các đặc trưng này sẽ được sử dụng làm dữ liệu đầu vào cho bước phân loại.

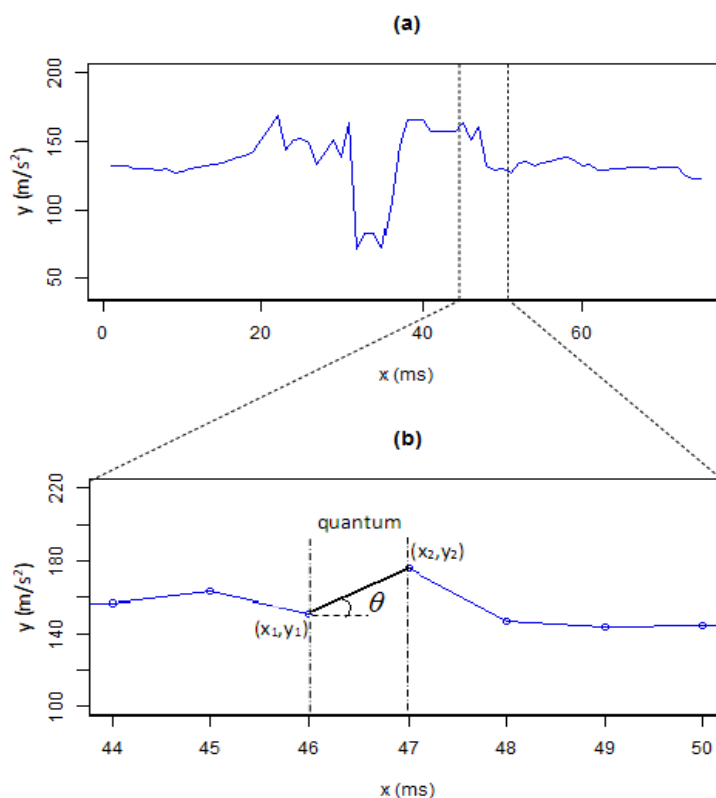
Phần dưới đây sẽ mô tả chi tiết về thuật toán trích xuất đặc trưng và một vài khía cạnh quan trọng khác của phương pháp đề xuất.

2.3.1. Thuật toán trích xuất đặc trưng



Hình 2.5. Ví dụ về mẫu tín hiệu nhiễu (a) và không nhiễu (b)

Xét biểu diễn hình học trong biểu đồ đường (line chart) của các khung dữ liệu cảm biến như hình 2.6. Đầu tiên, mỗi khung tín hiệu được chia thành các mảnh nhỏ hơn (*quantum*). Mỗi *mảnh* có độ dài l là một tham số được định nghĩa trước đó. Trong biểu diễn hình học theo biểu đồ đường, góc của mảnh được tạo bởi véc-tơ nối giữa điểm dữ liệu đầu và cuối của mảnh với trục ngang thời gian x như trên hình 2.6. Sau khi tính toán góc của mỗi mảnh, cần thực hiện tính một phân phối tần suất (histogram) của các góc nhờ việc chia toàn bộ khoảng giá trị góc sang dạng khoảng (interval), được gọi là các *ngăn* (bin), và đếm số các mảnh có góc thuộc từng *ngăn*. Khoảng giá trị của *ngăn* được học thông qua việc sử dụng các phương pháp phân cụm như k -means để nhóm các góc của mảnh trong các khung trong tập huấn luyện thành các cụm. Hoặc có thể xác định nhanh chóng khoảng giá trị của *ngăn* qua việc gán các giá trị cụ thể theo một cách thức nào đó. Phân phối tần suất đã tạo có thể được sử dụng làm đặc trưng cho khung đang xét.

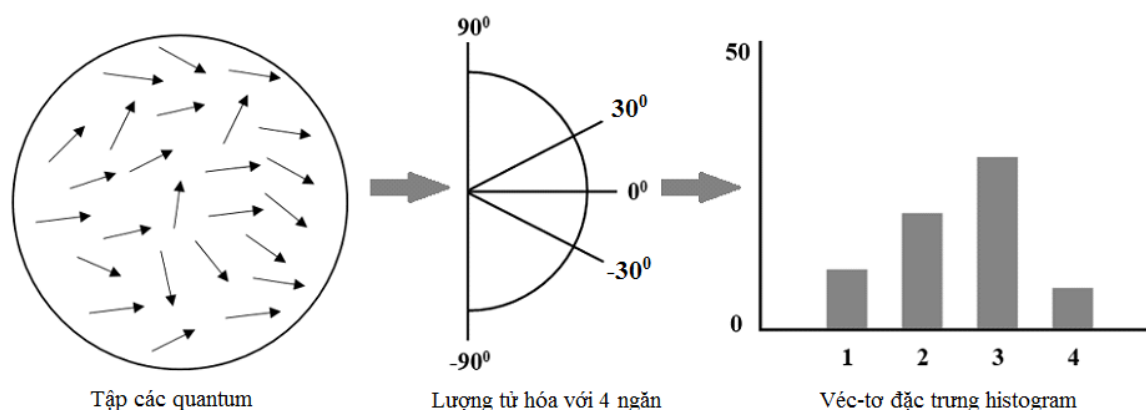


Hình 2.6. Ví dụ về một mảnh (quantum) và hướng của nó

Quá trình trích xuất đặc trưng được mô tả chi tiết như sau.

Bước 1: Tính góc của mảnh. Xét một khung tín hiệu trên trục x trong luồng tín hiệu cảm biến gia tốc theo 3 trục và các mảnh của nó (xem minh họa trong hình 2.6). Chú ý rằng luồng tín hiệu được mô tả trong hệ tọa độ xy với x theo mili giây và y theo m/s^2 . Hình 2.6(a) mô tả một ví dụ về khung cắt ra từ chuỗi tín hiệu cảm biến, từ mẫu tín hiệu ngã 2.6(a) tương ứng. Và hình 2.6(b) mô tả ví dụ về một mảnh cùng hướng của nó. Góc $\theta(q)$ của một mảnh q được xác định là góc giữa véc-tơ nối điểm đầu (x_i, y_i) và điểm cuối (x_{i+1}, y_{i+1}) và trục x (hình 2.6), và có thể được tính theo công thức:

$$\theta(q) = \tan^{-1} \left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right) \quad (2.1)$$



Hình 2.7. Một mẫu rời rạc hóa 4 ngăn và véc-tơ đặc trưng phân phối tần suất

Giá trị của góc nằm trong khoảng $(-90^0, 90^0)$. Thực chất, giá trị này chính là độ khác nhau (hay xấp xỉ đạo hàm) giữa hai điểm dữ liệu (x_i, y_i) và (x_{i+1}, y_{i+1}) và được tính theo đơn vị đo góc.

Khi cài đặt, ở bước phân loại, ta có thể giảm bớt chi phí tính góc theo công thức 2.1 bằng cách dựa vào giá trị trong hàm $\tan^{-1}()$ để ánh xạ trực tiếp vào ngăn tương ứng mà không cần phải tính cụ thể $\tan^{-1}()$. Việc này có thể dễ dàng thực hiện bằng cách sử dụng các ngăn đã được xác định về ranh giới (sau khi thực hiện phân cụm ở phần **Xác định các ngăn bằng phương pháp phân cụm** bên dưới) để tính ngược lại các khoảng giá trị (tính $\tan()$) trung gian, và sau đó các giá trị nằm trong hàm $\tan^{-1}()$ sẽ được ánh xạ vào khoảng giá trị trung gian tương ứng thông qua tìm kiếm nhị phân (do các khoảng đã được sắp xếp theo thứ tự về ranh giới một cách liên tục).

Bước 2: Rời rạc hóa. Để tạo phân phối tần suất, trước tiên cần chia giá trị các góc xác định ở bước 1 vào các ngăn. Các ngăn này được tính trước đó, bằng cách phân chia 180 độ thành M ngăn (với các khoảng giá trị của ngăn có thể khác nhau). Đây là quá trình rời rạc hóa mức độ khác biệt về tín hiệu giữa hai điểm liên nhau. Việc rời rạc hóa, tức là sự lựa chọn số lượng và ranh giới các ngăn có ảnh hưởng tới chất lượng biểu diễn cửa sổ và nhận dạng. Lựa chọn ngăn quá nhỏ sẽ gây nhiễu do các góc đặc trưng cho mỗi loại chuyển động có thể rơi vào hai ngăn khác

nhau. Ngược lại, lựa chọn ngăn quá lớn có thể dẫn tới việc các góc đặc trưng cho các chuyển động khác nhau lại rơi vào cùng một ngăn và do vậy đặc trưng không còn tính đại diện cho chuyển động. Tương tự như vậy, việc lựa chọn ranh giới của các ngăn không phù hợp cũng có thể gây ra hậu quả như trên. Việc rời rạc hóa được thực hiện bằng hai cách, thủ công hoặc tự động, và hiệu quả sẽ được xác định nhờ thực nghiệm. Tuy nhiên, việc rời rạc hóa theo cách thủ công có nhiều hạn chế như không thuận tiện và khó tối ưu các ngăn, cho nên ở đây chỉ tập trung vào cách tính toán tự động.

Như vậy, một ngăn sẽ chứa một nhóm các mảnh có góc tương tự nhau. Hình 2.7 đưa ra một ví dụ về trường hợp lượng tử hóa 4 ngăn: $(-90^0, -30^0)$, $(-30^0, 0^0)$, $(0^0, 30^0)$ và $(30^0, 90^0)$. Việc tính toán tự động có thể thực hiện thông qua các phương pháp phân cụm không giám sát. Phần sau sẽ mô tả chi tiết cách xác định các ngăn bằng phương pháp phân cụm. Khi các ngăn được xác định, việc rời rạc hóa được thực hiện bằng cách gán góc $\theta(q)$ của mỗi mảnh q vào một trong các ngăn.

Bước 3: Tính phân phối tần suất. Số các mảnh thuộc mỗi ngăn được đếm để tạo ra véc-tơ phân phối tần suất độ dài M . Véc-tơ phân phối tần suất này có các phần tử là các số đã đếm được trong mỗi ngăn (xem ví dụ hình bên phải hình 2.7). Sau đó, nối 3 véc-tơ được tính theo cách này theo 3 luồng tín hiệu gia tốc (theo 3 trục x, y, z) để tạo ra véc-tơ tổng có độ dài $3xM$ và sử dụng véc-tơ tổng này làm véc-tơ đặc trưng cho khung đang xét.

Xác định các ngăn bằng phương pháp phân cụm. Đầu tiên, thực hiện tính góc của tất cả các mảnh thuộc các khung tín hiệu trong tập huấn luyện. Sau đó, sử dụng phương pháp phân cụm như k -means để phân các góc thành M cụm (lý do chọn k -means là vì đây là một phương pháp phân cụm tự động đơn giản và hoạt động hiệu quả, thường dùng trong các phương pháp nhận dạng hoạt động sử dụng túi từ (Bag-of-Word) như [20,55,151]). Tuy nhiên, để gán góc của một mảnh vào một ngăn cần thực hiện gán chỉ số cụm bằng phương pháp tìm kiếm hàng xóm gần nhất. Bước này có thể được cải tiến thông qua việc tính khoảng giá trị của M ngăn

một cách trực tiếp như sau. Xét một ngăn thứ i (khác ngăn đầu và ngăn cuối). Giá trị thấp nhất của khoảng giá trị được tính bằng giá trị trung bình của tâm cụm thứ $(i-1)$ và cụm thứ i . Giá trị cao nhất của khoảng giá trị được tính bằng giá trị trung bình của tâm cụm thứ i và cụm thứ $(i+1)$. Khoảng giá trị của ngăn đầu tiên tính từ 0 tới giá trị trung bình của tâm cụm thứ nhất và thứ hai. Khoảng giá trị của ngăn cuối M tính từ giá trị trung bình của tâm 2 cụm cuối cùng tới 90^0 .

Về lựa chọn số cụm tối ưu cho thuật toán k -means, theo như khảo sát của tác giả về các phương pháp nhận dạng hoạt động sử dụng túi từ (Bag-of-Word) như [20,55,151], việc này chỉ có thể được thực hiện dựa trên thực nghiệm kiểm tra chéo trên dữ liệu huấn luyện (như được mô tả cụ thể trong phần thực nghiệm mục 2.4.4).

Thuật toán trích xuất đặc trưng HALF được mô tả trong sơ đồ thuật toán *HALF_Extraction* dưới đây như trong Hình 2.8. Chú ý là cần thực hiện xác định các ngăn trước đó sử dụng phương pháp phân cụm *Bin_Calculation* (Hình 2.9) để làm đầu vào cho thuật toán trích xuất đặc trưng.

Tên chương trình: HALF_Extraction

Ký hiệu: $Q_t = \{q_i\}$ – tập các mảnh thuộc khung tín hiệu s_t

(x_i, y_i) – điểm dữ liệu trong khung tín hiệu s_t , trong đó x_i là giá trị của điểm thời gian trong khung và y_i là giá trị gia tốc tương ứng.

Đầu vào: Khung tín hiệu s_t , độ dài lát tín hiệu bằng l , số ngăn M , khoảng giá trị các ngăn (sv_j, ev_j)

Đầu ra: Véc-tơ đặc trưng $f = (c_i)$, với $i = 1, \dots, 3 \times M$

Chương trình:

1. Tách khung tín hiệu đầu vào thành 3 véc-tơ theo 3 trục x, y, z : s_1, s_2, s_3

2. Lặp lại với $t = 1, 2, 3$

a) Chia khung tín hiệu s_t thành tập Q_t gồm k mảnh độ dài l liên tiếp nhau $q_i = \{(x_i, y_i), (x_{i+1}, y_{i+1})\}$, với $i = 1, \dots, k-1$ và $k = |s_t| \text{ div } l$.

b) Tính góc $\theta(q_i)$ cho mọi mảnh q_i , với $i = 1, \dots, k-1$, trong Q_t theo:

$$\theta(q_i) = \tan^{-1} \left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right)$$

c) Rời rạc hóa các $\theta(q_i)$ vào M ngăn cho trước:

Nếu $\theta(q_i)$ nằm trong khoảng (sv_j, ev_j) , gán q_i vào ngăn j

d) Tính phân phối tần suất:

Đếm số mảnh trong mỗi ngăn để tạo thành véc-tơ phân phối tần suất đặc trưng độ dài M : $f_j = \{c_j\}$. Với $j = 1, \dots, M$.

3. Trả về véc-tơ đặc trưng $f = \{f_j\} = \{c_j\}$, với $t = 1, 2, 3$ và $i = 1, \dots, 3 \times M$.

Hình 2.8. Thuật toán trích xuất đặc trưng HALF cho một khung tín hiệu

Tên chương trình: Bin_Calculation

Ký hiệu: $center_i$ là giá trị trung tâm của cụm thứ i

Đầu vào: $Q = \{q_i\}$ – tập các mảnh thuộc các khung tín hiệu trong tập huấn luyện, số ngăn M

Đầu ra: Các khoảng giá trị của M ngăn: (sv_j, ev_j) với $j = 1, \dots, M$

Chương trình:

1. Sử dụng k -means phân cụm các góc $\theta(q_i)$ thành M cụm với q_i thuộc Q

2. Xác định khoảng của M ngăn

$$sv_1 = 0^0; ev_1 = (center_0 + center_1)/2.$$

$$sv_M = (center_{(M-1)} + center_M)/2; ev_M = 90^0.$$

$$sv_i = (center_{(i-1)} + center_i)/2; ev_i = (center_i + center_{(i+1)})/2, \text{ với } i = 2, \dots, k-1.$$

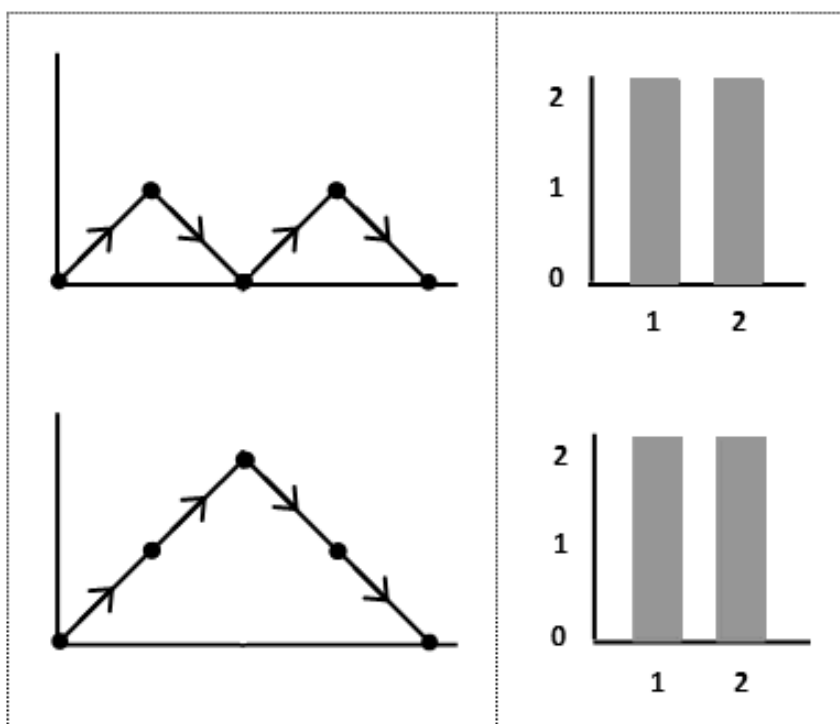
3. Trả về M khoảng giá trị của M ngăn (sv_j, ev_j) với $j = 1, \dots, M$

Hình 2.9. Thuật toán tự động xác định các ngăn sử dụng phương pháp phân cụm

2.3.2. Biểu diễn đặc trưng đa mức

Ý tưởng đằng sau cách sử dụng các đặc trưng này là các khung tương ứng với các hoạt động sẽ có đường cong tín hiệu tương tự nhau và do đó dẫn tới số lượng các mảnh theo nhóm các hướng là tương đương nhau. Vì vậy, những đặc trưng này sẽ cung cấp một mô tả xấp xỉ cho các luồng tín hiệu, tính toán hiệu quả cũng như phù hợp với một số lớn các giải thuật phân lớp. Vấn đề là các phân phối tần suất góc của các mảnh không chứa các thông tin về vị trí tương đối của mảnh trong một cửa sổ. Ví dụ dưới đây minh họa vấn đề này.

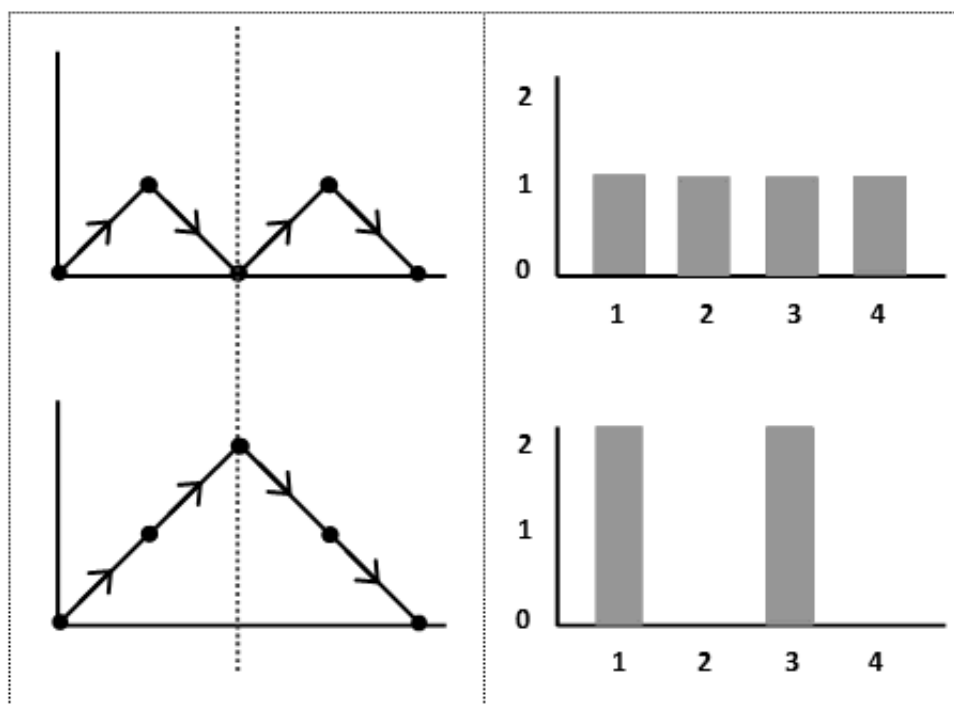
Xét 2 khung dữ liệu của 2 hoạt động khác nhau như trong hình 2.10.



Hình 2.10. Các đặc trưng phân phối tần suất giống nhau (các hình bên phải) được trích xuất từ 2 khung (các hình bên trái) không phân đoạn

Giả sử cần tạo một phân phối tần suất có 2 ngăn với kích cỡ là 90 độ. Khoảng giá trị của ngăn thứ nhất là từ -90 độ tới 0 và của ngăn thứ 2 là từ 0 tới 90 độ. Sử dụng thuật toán trích xuất đặc trưng ở phần trước sẽ trích xuất đặc trưng của 2 khung thành 2 véc-tơ phân phối tần suất đều có giá trị là (2,2). Hai véc-tơ phân phối tần suất có giá trị giống nhau được trích xuất từ 2 khung dữ liệu khác nhau. Như vậy trong trường hợp này, thuật toán trích xuất đặc trưng đề xuất không hiệu quả trong việc phân biệt 2 hoạt động khác nhau. Vấn đề này xảy ra do các đặc trưng không chứa các thông tin về vị trí tương đối của các mảnh trong từng đoạn.

Để giải quyết vấn đề xuất hiện trong ví dụ này, có thể trích xuất các đặc trưng cục bộ thay vì trích xuất một đặc trưng trên toàn bộ khung. Chẳng hạn, có thể chia khung thành 2 đoạn (sử dụng đường gạch nối trong hình 2.11) và trích xuất đặc trưng trên từng đoạn, rồi sau đó gộp 2 véc-tơ phân phối tần suất cục bộ này thành một véc-tơ tổng để nhận được 2 véc-tơ đặc trưng hoàn toàn khác nhau. Một véc-tơ là (1,1,1,1) và một véc-tơ là (2,0,0,2).



Hình 2.11. Các đặc trưng khác nhau (các hình bên phải) được trích xuất từ 2 khung (các hình bên trái), mỗi khung được chia thành 2 đoạn (bởi đường gạch nối ở giữa)

Như vậy, với cách tính đặc trưng cục bộ, đặc trưng tạo ra có thể nắm giữ các thông tin về vị trí tương đối của mảnh trong một cửa sổ. Đây chính là thuật toán biểu diễn đặc trưng đa mức hiệu quả hơn thuật toán ban đầu (như mô tả trong phần 2.3.1). Với thuật toán này, chúng ta có được các đặc trưng HALF tốt hơn (như kết quả cho thấy trong phần thử nghiệm). Thuật toán được mô tả như sau.

Xét một khung dữ liệu w . Chia khung dữ liệu này thành K đoạn sw_i nhỏ hơn và liên tiếp nhau, có kích thước bằng nhau ($i = 1..K$). Với mỗi đoạn sw_i trong khung, trích xuất đặc trưng HALF (đã mô tả trong phần 2.3.1) để tạo ra véc-tơ đặc trưng cục bộ. Véc-tơ đặc trưng của cả khung được tạo ra bằng cách nối K véc-tơ đặc trưng cục bộ từ K đoạn trong khung. Sơ đồ thuật toán được trình bày trong Hình 2.12 dưới đây. Chú ý rằng cũng cần thực hiện xác định các ngăn trước đó với phương pháp phân cụm Bin_Calculation (Hình 2.9) để làm đầu vào cho thuật toán.

Tên chương trình: Multilevel_Representation

Đầu vào: Khung tín hiệu w , độ dài lát tín hiệu bằng l , số ngăn M , khoảng giá trị các ngăn (sv_j, ev_j)

Đầu ra: Véc-tơ đặc trưng F độ dài $3 \times K \times M$

Chương trình:

1. Chia w thành K đoạn tín hiệu nhỏ hơn, liên tiếp nhau, độ dài bằng nhau sw_i
2. Trích xuất đặc trưng HALF cho K đoạn dữ liệu sw_i , $i = 1..K$

$$f_i = \text{HALF_Extraction}(sw_i, l, M, (sv_j, ev_j)), \text{ với } f_i \text{ có độ dài } 3 \times M$$

3. Nối K véc-tơ f_i thành véc-tơ đặc trưng tổng $F = \{f_i\}$
4. Trả về véc-tơ F , độ dài $3 \times K \times M$

Hình 2.12. Thuật toán biểu diễn đặc trưng đa mức cho một khung tín hiệu

Trong quá trình trích xuất đặc trưng, các tham số như độ dài mảnh l , số các ngăn M và số các đoạn K rất quan trọng. Do tốc độ lấy mẫu cho các hoạt động sử dụng cảm biến gia tốc thường từ 50 Hz đến 100 Hz, nghiên cứu này chỉ sử dụng mảnh với độ dài bằng $l = 1$ bởi vì với độ dài $l = 2$ hay lớn hơn, tốc độ lấy mẫu sẽ có thể nhỏ hơn 50 Hz và như vậy, hệ thống có thể không đáp ứng được cho các ứng dụng chứa các hoạt động thay đổi nhanh như ngã. Khi tốc độ lấy mẫu lớn hơn, tham số l có thể xác định thông qua thực nghiệm. Hai tham số còn lại có thể xác định được thông qua thực nghiệm hay tối ưu hóa sau đó. Cách chọn tham số về số các ngăn M và số các đoạn K phù hợp cho phép phân tách rõ ràng các tín hiệu hoạt động khác biệt nhau, ví dụ như ngã và các hoạt động khác. Chú ý rằng, tham số M và K càng lớn sẽ gây ảnh hưởng tới tốc độ tính toán của hệ thống do số cụm M tăng làm chậm thuật toán phân cụm k -means, còn số đoạn K tăng làm số chiều của đặc trưng cũng tăng.

2.3.3. Độ phức tạp thuật toán

Độ phức tạp tính toán cho phương pháp trích xuất đặc trưng HALF đề xuất (trong mục 2.3.1) được tính bằng số các phép tính phải thực hiện khi thực hiện thuật toán, tức là số các phép tính cần để *xây dựng mô hình* và *gán nhãn cho dữ liệu đầu vào* (dự đoán mẫu đầu vào). Độ phức tạp thuật toán đề xuất khi *xây dựng mô hình* bằng tổng độ phức tạp thuật toán của các bước trong thuật toán, gồm: tự động xác định các ngăn bằng phương pháp phân cụm - *Bin_Calculation* (hình 2.9, ở đây dựa trên *k-means*), các bước trong thuật toán trích xuất đặc trưng *HALF_Extraction* (hình 2.8, với tách khung tín hiệu thành 3 véc-tơ theo 3 trục, chia nhỏ khung tín hiệu, tính góc cho mảnh, rời rạc hóa góc của mỗi mảnh và tính phân phối tần suất).

Cho N là tổng số mảnh của tập dữ liệu huấn luyện đầu vào, độ phức tạp của thuật toán *Bin_Calculation* (hình 2.9) chính là độ phức tạp của thuật toán phân cụm dựa trên *k-means* $O(NMi)$, với i là số lần lặp *k-means* phải thực hiện khi chạy, M là số cụm (cluster), cộng với độ phức tạp của bước xác định khoảng của M ngăn $O(M)$. Để thấy tổng độ phức tạp thuật toán *Bin_Calculation* là $O(NMi)$. Độ phức tạp thuật toán trích xuất đặc trưng *HALF_Extraction* (hình 2.8) gồm độ phức tạp của bước tách khung tín hiệu là không đáng kể, độ phức tạp của bước tính góc cho các mảnh là $O(N)$, độ phức tạp thuật toán của bước rời rạc hóa vào M ngăn cho trước là $O(M)$ và độ phức tạp của bước tính phân phối tần suất là $O(N)$ do chỉ thực hiện đếm số mảnh có trong từng cụm. Như vậy, có thể thấy rằng độ phức tạp của bước xây dựng mô hình học đặc trưng là $O(NMi)$. Tuy nhiên, trong trường hợp xây dựng mô hình học theo kiểu ngoại tuyến (offline) thì chỉ cần quan tâm tới tốc độ của việc *gán nhãn cho dữ liệu đầu vào* ở dưới, như thực nghiệm tiến hành ở phần sau.

Trong trường hợp *gán nhãn dữ liệu đầu vào*, độ phức tạp thuật toán bằng tổng độ phức tạp thuật toán của tất cả các bước trừ tự động phân cụm (*Bin_Calculation*). Để thấy độ phức tạp thuật toán là $O(MN)$ với M là số cụm và N là số mảnh đầu vào đối với cách tìm kiếm hàng xóm gần nhất. Tuy nhiên, với các ngăn đã xác định giới hạn đầu cuối và đã sắp xếp theo thứ tự (xem thuật toán

HALF_Extraction hình 2.8) thì việc tìm kiếm chỉ mất $O(\log M)$ nếu sử dụng tìm kiếm nhị phân. Trong thực tế, số cụm M nhỏ hơn 100 nên độ phức tạp thuật toán sẽ là $O(N)$, tức là tuyến tính với độ dài khung. Với độ phức tạp thuật toán tuyến tính, hơn hẳn phương pháp uốn thời gian động DTW [101] với độ phức tạp $O(N^2)$ đã từng được sử dụng trong các hệ thống thời gian thực như trong công trình nghiên cứu số 2 của tác giả, HALF rất hứa hẹn đối với các hệ thống tính toán theo thời gian thực.

Chú ý là độ phức tạp thuật toán không bị ảnh hưởng nhiều bởi số đoạn K trong một khung do số mảnh đầu vào dữ liệu không thay đổi khi số đoạn thay đổi. Do đó khi sử dụng thuật toán biểu diễn đặc trưng đa mức *Multilevel_Representation* (hình 2.12) ứng dụng *HALF_Extraction* (hình 2.8) để trích xuất đặc trưng đa mức hiệu quả hơn thì độ phức tạp thuật toán cũng không tăng nhiều.

Với độ phức tạp tính toán như trên của thuật toán HALF, hệ thống nhận dạng hoạt động người dựa trên HALF như phát hiện ngã có thể dễ dàng được triển khai trên các hệ thống cá nhân thông minh đeo trên người trong thực tế, ví dụ các điện thoại thông minh như Samsung Galaxy S3 hay đồng hồ thông minh như Fitbit Surge được đề cập tới trong phần thực nghiệm.

2.4. THỬ NGHIỆM

Để đánh giá hiệu quả của các đặc trưng HALF đã đề xuất, phần này trình bày thử nghiệm phương pháp nhận dạng hoạt động người sử dụng các đặc trưng HALF, cụ thể là phương pháp phân biệt ngã và các hoạt động khác (còn gọi là phát hiện ngã) sử dụng bộ cảm biến gia tốc đeo trên người. Các đặc trưng HALF này được trích xuất từ các mảnh thuộc các cửa sổ thời gian trên chuỗi dữ liệu gia tốc liên tục lấy từ bộ cảm biến gia tốc. Các đặc trưng đã trích xuất được sử dụng làm dữ liệu đầu vào cho bộ phân loại SVM (đã được huấn luyện), từ đó phân loại các khung của chuỗi dữ liệu đầu vào của hệ thống thành *ngã* hoặc *không phải ngã*. Phương pháp đề xuất được đánh giá một cách chặt chẽ trên 3 tập dữ liệu đã công bố, chứa dữ liệu ngã và các hoạt động khác của cuộc sống hàng ngày (bao gồm số lượng nhiều đáng

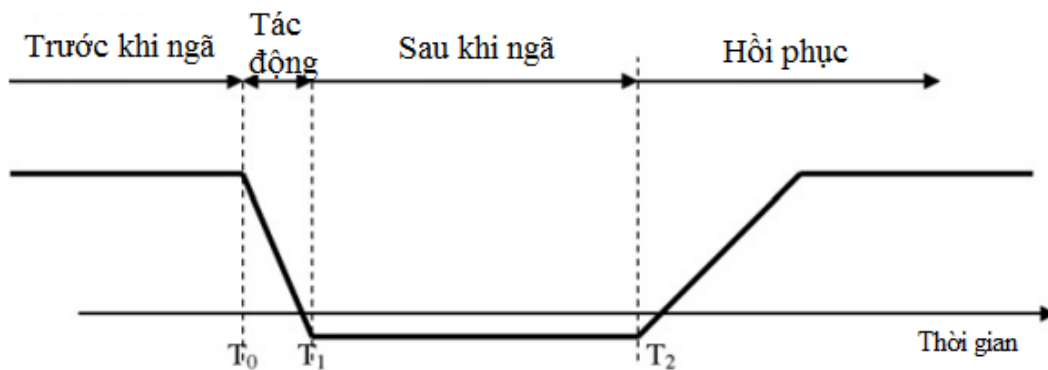
kể đối với hoạt động có ngã). Kết quả cho thấy ngã có thể được phát hiện với độ chính xác cao trên 3 tập dữ liệu thử nghiệm. Các thử nghiệm cũng chứng minh hiệu năng vượt trội của phương pháp đề xuất khi so sánh với hai phương pháp phát hiện ngã khác.

2.4.1. Giới thiệu một số nghiên cứu phát hiện ngã

Các kỹ thuật phát hiện ngã sử dụng trong các hệ thống dùng cảm biến mang trên người có thể chia làm hai loại chính [61]: dựa trên ngưỡng và dựa trên học máy. Có một lượng lớn các hệ thống sử dụng các thuật toán dựa trên ngưỡng do có khả năng phát hiện ngã chính xác, nhanh và độ phức tạp thấp. Trong các phương pháp dựa trên ngưỡng, một sự kiện ngã được phát hiện khi một hay một vài đặc trưng như các đỉnh cao hay thấp (acceleration peaks or valley), hay các đặc trưng về hình dạng khác đạt đến ngưỡng định trước. Ví dụ, Lindeman và các cộng sự [87] sử dụng các cảm biến gia tốc, được gắn vào thiết bị đeo sau tai và phát hiện ngã bằng cách sử dụng hướng của đầu trong không gian cùng vận tốc ngay trước khi tiếp xúc mặt đất và gây nên tác động. Kangas và cộng sự [66] đề xuất bốn ngưỡng bao gồm véc-tơ tổng, véc-tơ tổng động, gia tốc thẳng đứng và độ khác biệt giữa giá trị gia tốc cực đại và cực tiểu để phát hiện ngã. Bourke và cộng sự trình bày hai thuật toán phát hiện ngã riêng biệt dựa trên ngưỡng của các tín hiệu gia tốc [23] và tín hiệu con quay hồi chuyển [22]. Lai và cộng sự [76] sử dụng ngưỡng để phân biệt các trạng thái tĩnh và động thông qua gia tốc của 3 trục. Mặc dù các phương pháp thuộc loại này có độ phức tạp thấp và thực hiện đơn giản, hệ thống xây dựng lên thường gặp vấn đề nghiêm trọng là tỉ lệ cảnh báo giả cao. Hơn nữa, cách tiếp cận này được chứng minh là không hiệu quả trong nhiều tình huống (xem [16]) do phụ thuộc vào đặc trưng di chuyển của từng người dùng, từ đó dẫn tới khả năng khái quát hóa thấp [60].

Dù có độ phức tạp cao hơn phương pháp dựa trên ngưỡng, các hệ thống phát hiện ngã gần đây thường dựa trên các phương pháp dựa trên học máy do có độ tin cậy cao. Thay vì sử dụng các ngưỡng cố định, các phương pháp dựa trên học máy

sử dụng các bộ phân loại như máy véc-tơ tựa (SVM) để phân biệt ngã và các hoạt động khác. Cách tiếp cận này cũng thường cho tỉ lệ phát hiện ngã tốt hơn [61]. Ví dụ, Khan [69] đề xuất sử dụng một tập gồm nhiều đặc trưng thống kê cho các tín hiệu gia tốc tổng, gia tốc phương ngang, phương thẳng đứng và vận tốc góc với bộ phân lớp dựa trên HMM và kết quả đạt được là rất khả quan. Jantaraprim và cộng sự [64] sử dụng SVM cùng các đặc trưng gọi là cực đại-cực tiểu trong thời gian ngắn (short time min-max) cho việc phát hiện ngã và có được độ chính xác cao. Tuy vậy, các đặc trưng đề xuất chỉ mới hạn chế khai thác đặc tính thuộc pha trước khi ngã (pre fall phase) và pha tác động (critical phase) (xem hình 2.13) dẫn tới độ chính xác hệ thống có thể bị ảnh hưởng khi các hoạt động hàng ngày tương tự ngã như đứng-ngồi, ngồi-nằm xuất hiện trong tập dữ liệu. Trong [7], Albert và đồng sự đề xuất một tập đặc trưng thống kê lớn để phát hiện ngã và đạt được kết quả cao. Tương tự, các tác giả trong [106] đưa ra một tập rất nhiều đặc trưng dựa trên các đặc trưng thống kê (tạo thành véc-tơ đặc trưng 234 chiều), sau đó rút gọn tập đặc trưng với PCA để phát hiện ngã với độ chính xác rất cao. Tuy vậy, độ phức tạp của tập đặc trưng lớn dẫn tới tính khả thi của các phương pháp này bị hạn chế do khó có thể tính toán nhanh trên hệ thống có năng lực xử lý thấp. Pham và cộng sự trong [110] sử dụng tập các đặc trưng thống kê đơn giản hơn nhưng cũng rất hiệu quả với bộ phân lớp SVM để phát hiện ngã và nhận dạng một số ADL. Tuy các đặc trưng này đã chứng tỏ khả năng trong việc phân biệt một số các hoạt động thường ngày, nhưng cũng chưa khai thác được hết các đặc tính của sự kiện ngã. Đồng thời, các phương pháp nêu trên dù có được tỉ lệ phát hiện ngã cao nhưng mới chỉ thử nghiệm hạn chế trên tập dữ liệu riêng của tác giả, do đó khó có thể kiểm chứng.



Hình 2.13. Bốn pha của sự kiện ngã ([103]): trước khi ngã, tác động, sau khi ngã, hồi phục

Qua khảo sát các phương pháp phát hiện ngã dựa trên học máy cho thấy, chìa khóa cho sự thành công của các phương pháp này là phải xây dựng được mô tả đặc trưng phù hợp cho dữ liệu cảm biến. Về mặt lý tưởng, để đạt được mức phát hiện ngã có độ chính xác cao, các đặc trưng được thiết kế để phân biệt rõ ngã và các hoạt động có tín hiệu tương tự ngã, đồng thời hoạt động tốt trên nhiều cá nhân khác nhau. Nghiên cứu này đề xuất một phương pháp phát hiện ngã dựa trên tập đặc trưng mới HALF hiệu quả và sử dụng bộ phân lớp SVM. Quá trình đánh giá nghiêm ngặt trên 3 tập dữ liệu đã công bố công khai đã chứng tỏ rằng phương pháp đề xuất sử dụng các đặc trưng HALF có hiệu quả cao hơn những phương pháp khả thi khác trong phần thử nghiệm để phân biệt ngã với các ADL.

2.4.2. Phương pháp phát hiện người ngã

Thuật toán đề xuất thực hiện phát hiện ngã bằng cách phân loại một cửa sổ tín hiệu hoạt động thành “ngã” hoặc “không ngã”. Dữ liệu đầu vào là 3 luồng tín hiệu gia tốc theo ba trục x , y , z . Thuật toán phát hiện ngã gồm 3 bước phổ biến: phân đoạn dữ liệu, trích xuất đặc trưng và phân lớp.

a) Phân đoạn dữ liệu

Dữ liệu từ bộ cảm biến biến thiên liên tục theo thời gian nên để xác định khoảng thời gian nào có nhiều khả năng người đeo bộ cảm biến bị ngã, dữ liệu được phân đoạn thành các khung cửa sổ trượt (sliding window hay frame). Do kích thước

khung cửa sổ trượt thường ảnh hưởng đáng kể đến hiệu suất của hệ thống nhận dạng [17] nên thông thường cần thử nghiệm các khung cửa sổ trượt với các kích thước khác nhau để tìm ra giá trị tối ưu.

Để chọn được độ dài cửa sổ phù hợp cần thực hiện một thực nghiệm dựa trên một tập con trên tập dữ liệu hiện có. Các độ dài cửa sổ được thay đổi với nhiều giá trị khác nhau. Đối với trường hợp phát hiện ngã, thông thường người ta lựa chọn các giá trị từ khoảng 1 giây tới 6 giây [110] để tìm được độ dài phù hợp nhất, do trong khoảng này cửa sổ có thể ghi nhận được toàn bộ sự kiện ngã. Độ dài cửa sổ được chọn phải đáp ứng được yêu cầu về độ trễ cho hệ thống khi xử lý liên tục trong thời gian thực, đồng thời giúp cho hệ thống có được độ chính xác cao. Nếu cửa sổ ngắn quá có thể không cung cấp đủ thông tin để mô tả đầy đủ sự kiện ngã. Ngược lại, các cửa sổ quá dài có thể chứa nhiều hoạt động trong một cửa sổ thời gian duy nhất [118], khiến nhận dạng ngã không chính xác.

b) Trích xuất đặc trưng

Ở đây, các đặc trưng từ mỗi khung được trích xuất và sau đó được dùng làm dữ liệu đầu vào cho bước phân lớp. Thử nghiệm này sử dụng các đặc trưng HALF đã đề xuất ở phần trên. Các véc-tơ đặc trưng thu được sẽ được dùng để huấn luyện thuật toán nhận dạng mẫu. Phần tiếp theo sẽ trình bày về bộ phân lớp SVM dùng cho phát hiện ngã trong thời gian thực.

c) Phân lớp

Trong nghiên cứu này, máy véc-tơ tựa (SVM) được sử dụng làm bộ phân lớp do độ chính xác vượt trội của SVM trong các ứng dụng dự đoán hoạt động người và trong một số lĩnh vực ứng dụng khác. SVM có thể đạt được kết quả tốt cho các dữ liệu chưa biết nhờ sử dụng hai kỹ thuật: (i) gán tập đặc trưng đầu vào với không gian đặc trưng mới (thường có số chiều cao hơn) thông qua các hàm nhân; và (ii) tìm trong không gian đặc trưng mới này một siêu phẳng với lề cực đại, cho phép phân tách các mẫu âm tính khỏi các mẫu dương tính.

Trước khi huấn luyện và dự đoán với SVM, cần chuẩn hóa các véc-tơ để tất cả các giá trị đặc trưng nằm trong khoảng $[0, 1]$. Đồng thời, việc thử nghiệm với một số dạng hàm nhân và chọn các tham số tốt nhất, bằng cách sử dụng thủ tục tìm kiếm lưới (grid-search) hay dùng, cũng được thực hiện.

Do các dữ liệu huấn luyện cho mô hình phát hiện ngã mất cân bằng và dễ gây ra hiện tượng quá vừa dữ liệu (như trong dữ liệu thực nghiệm mô tả trong bảng 2.1), nghiên cứu này sử dụng các tham số để quy định trọng số cho hai lớp trong SVM, tức là gán một giá trị phạt cho lớp *không phải ngã* do có số mẫu lớn hơn nhiều lớp *ngã*, đồng thời sử dụng kỹ thuật lấy mẫu SMOTE [30] để tăng lượng mẫu dữ liệu đầu vào của SVM của lớp có số mẫu ít một cách hợp lý, giúp SVM hoạt động hiệu quả hơn. Phương pháp này gọi là SDC [6], được chứng minh là rất phù hợp cho các trường hợp khi có tập dữ liệu không quá lớn [133].

d) Các cấu hình thử nghiệm khác

Một độ đo phù hợp giúp đánh giá đúng hiệu năng của một hệ thống và ngược lại. Nghiên cứu này dựa trên nhiều độ đo khác nhau, được sử dụng trong nhiều bài toán phân lớp (kể cả phát hiện ngã) [60,133,140], bao gồm độ chính xác tổng thể (overall accuracy), độ nhạy (sensitivity), độ đặc hiệu (specificity).

Độ chính xác tổng thể là một độ đo đơn giản thường được sử dụng, được tính bằng tỉ lệ giữa số khung được phân loại chính xác trên tổng số khung, theo công thức:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.2)$$

Trong đó, TP (True Positive) là tỉ lệ đo số lần hệ thống phát hiện ngã và số lần thực tế ngã. TN (True Negative) là tỉ lệ đo số hệ thống phát hiện là không ngã và số lần thực tế không ngã. FP (False Positive) là tỉ lệ đo số lần hệ thống phát hiện ngã và số lần thực tế không ngã. FN (False Negative) là tỉ lệ đo số lần số lần hệ thống phát hiện không ngã và số lần thực tế ngã.

Tuy nhiên, để đánh giá hiệu năng một hệ thống phân lớp không cân bằng như phát hiện ngã thì độ chính xác không phải là một độ đo thực sự hiệu quả do ảnh hưởng của hai lớp tới kết quả phân loại là không cân bằng [133,140]. Điều này là do số lượng mẫu không cân bằng. Thông thường, trong các nghiên cứu về phát hiện ngã [1,16] thì các độ đo thường được dùng (ngoài độ chính xác tổng thể) để đánh giá hệ thống là độ nhạy (sensitivity), độ đặc hiệu (specificity) do độ nhạy có thể mô tả chỉ duy nhất về lớp có ít mẫu (*ngã*) và độ đặc hiệu có thể mô tả được chỉ duy nhất về lớp chiếm đa số mẫu (*không phải ngã*). Độ nhạy được tính bằng số lần phát hiện ngã đúng trên số lần hệ thống phát hiện ngã, qua đó xác định được khả năng *phát hiện ngã chính xác* của hệ thống. Độ đặc hiệu được tính bằng số lần phát hiện sự kiện thực sự không phải ngã trên tổng số sự kiện không phải ngã, qua đó xác định được khả năng tránh cảnh báo giả của hệ thống (phát hiện sự kiện không phải ngã chính xác). Hai độ đo này được tính bằng các công thức:

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2.4)$$

Do các mẫu dữ liệu ADL và ngã phân bố không cân bằng và số lượng mẫu dữ liệu ngã ít trong các tập dữ liệu dùng để đánh giá nên ở đây sử dụng kiểm tra chéo trên 4 phần dữ liệu (4-fold cross validation) để đánh giá hiệu năng cho tất cả các phương pháp và cấu hình. Các độ đo được tính theo từng phần dữ liệu (fold) trong tất cả các thực nghiệm. Với mỗi phần dữ liệu đó, 10% dữ liệu trong tập huấn luyện (lấy mẫu theo từng lớp – stratified sampling) được sử dụng làm tập kiểm chứng hay tập tối ưu (validation set) để tối ưu hóa các tham số và sử dụng những tham số này cho tập thử nghiệm hay tập kiểm thử (test set).

Các thử nghiệm trong phần này sử dụng LibSVM với hàm nhân RBF (do thực nghiệm cho thấy RBF có kết quả tốt nhất). Tham số C và $gamma$ của SVM được chọn thông qua tìm kiếm lưới (grid search) trên tập tối ưu. Tham số quy định

trong số $w1$ và $w2$ cho 2 lớp tương ứng *ngã* và *không phải ngã* được chọn khi huấn luyện mô hình là $w1 = 1$, $w2 = (\text{số mẫu huấn luyện không phải ngã}) / (\text{số mẫu ngã})$. Tất cả các thử nghiệm được chạy trên một máy tính chạy hệ điều hành Windows với bộ xử lý 2,8 GHz dual-core và RAM 8GB.

2.4.3. Tập dữ liệu thử nghiệm

Phương pháp đề xuất được đánh giá thử nghiệm trên 3 tập dữ liệu ngã công khai từ 3 tổ chức nghiên cứu khác nhau, bao gồm: tập dữ liệu DLR [41] của Viện truyền thông và đạo hàng của trung tâm hàng không vũ trụ Đức, tập dữ liệu MobiFall2 [139] của phòng thí nghiệm tin sinh và sức khỏe điện tử của Viện giáo dục công nghệ Crete và tập dữ liệu tFall [96] của trường đại học Zaragoza. Tất cả các tập dữ liệu trên chứa nhiều dữ liệu cảm biến về các sự kiện ngã và hoạt động thường ngày khác.

- **DLR [154]**: tập dữ liệu này chứa dữ liệu đã gán nhãn về ngã và 6 hoạt động thường ngày (ADL). Tổng số có 16 nam và nữ với độ tuổi từ 23 tới 50 thực hiện các kiểu ngã và ADL như “ngồi xuống”, “đứng lên”, “đi bộ”, “chạy” và “nằm xuống”. Mỗi người tham gia đeo thiết bị đo gắn trong (Inertial Measurement Unit) Xsens MTx cùng một bộ theo dõi gắn trên dây thắt lưng. Tần số lấy mẫu của các cảm biến gia tốc được đặt ở mức 100 Hz.
- **MobiFall2 [21]**: tập dữ liệu này có 4 kiểu ngã và 9 kiểu ADL, được thực hiện bởi 11 tình nguyện viên gồm 6 nam và 5 nữ, độ tuổi từ 22 tới 36. Có nhiều mẫu ngã như “ngã nằm ra đằng trước”, “ngã quỳ rồi nằm ra đằng trước”, “ngã nằm sang bên”, “ngã ngồi ra đằng sau”, cùng với 9 loại hoạt động hàng ngày như “đứng lên”, “đi bộ”, “chạy bộ”, “nhảy”, “đi lên cầu thang”, “đi xuống cầu thang”, “ngồi ghế”, “bước vào xe” và “bước ra xe”. Từng người tham gia thử nghiệm đặt điện thoại di động Samsung Galaxy S3 trong túi quần/áo. Tần số lấy mẫu của cảm biến gia tốc là 87 Hz.

- **tFall** [155]: tập dữ liệu này gồm 8 kiểu ngã và một số hoạt động sống thực tế. Dữ liệu về các hoạt động này được thu thập từ 7 nam và 3 nữ, tuổi từ 20 tới 42 trong điều kiện sống thực, trong chính môi trường sống của chính họ và kéo dài trong suốt quãng thời gian ít nhất một tuần. Tập dữ liệu ngã thử nghiệm bao gồm “ngã ra phía trước”, “ngã thẳng ra phía trước”, “ngã ra sau”, “ngã sang bên trái”, “ngã sang bên phải”, “ngồi trên không khí”, “bất tỉnh”, và “ngã ra phía trước vào vật cản”. Người tham gia sử dụng điện thoại Samsung Galaxy Mini để thu thập dữ liệu ADL nhưng chỉ lấy các dữ liệu với đỉnh gia tốc trên ngưỡng 1,5g. Khi thu thập dữ liệu ngã, người tham gia để điện thoại trong 2 túi quần/áo. Tần số lấy mẫu cho cảm biến gia tốc được thiết lập ở mức 50 Hz. Sau tiền xử lý, tập dữ liệu có tổng cộng 7816 mẫu ADL và 503 mẫu ngã, mỗi mẫu có độ dài 6 giây. Đây là tập dữ liệu lớn nhất sử dụng trong nghiên cứu này.

Để dữ liệu có cùng một kiểu định dạng khi đánh giá thuật toán, các tập dữ liệu được thực hiện tiền xử lý sử dụng cửa sổ trượt 6 giây, độ chồng lấn 50%, mỗi khung được gán nhãn thành ADL hay ngã, với các đỉnh tín hiệu gia tốc vượt ngưỡng 1,5g (tương tự tập dữ liệu tFall). Đỉnh tín hiệu gia tốc của một khung dữ liệu là giá trị cực đại của độ lớn tín hiệu gia tốc của khung dữ liệu. Sau quá trình tiền xử lý, thu được 2771 khung ADL và 36 khung dữ liệu ngã từ tập dữ liệu DLR, 1832 khung ADL và 288 khung dữ liệu ngã từ MobiFall2, và 7816 khung ADL và 503 khung dữ liệu ngã từ tFall. Thông tin về các tập dữ liệu được tóm tắt trong bảng 2.1.

Bảng 2.1. Thông tin tóm tắt các tập dữ liệu

	DLR	MobiFall2	tFall
Người tham gia	16 nam và nữ (độ tuổi 23-50)	6 nam và 5 nữ (độ tuổi 22-36)	7 nam và 3 nữ (độ tuổi 20-42)
Số kiểu ngã	Không xác định	4	8
Số kiểu ADL	6	9	Không xác định
Số khung dữ liệu ngã	36	288	503
Số khung dữ liệu ADL	2771	1832	7816
Tần số lấy mẫu	100 Hz	87 Hz	50 Hz
Khoảng giá trị	7g	2g	2g
Vị trí đặt cảm biến	Thắt lưng	Túi quần/áo	Túi quần/áo

2.4.4. Kết quả thử nghiệm và đánh giá

a) Ảnh hưởng của các khía cạnh khác nhau trong thuật toán phát hiện ngã tới độ chính xác hệ thống

Phần này thực hiện các thử nghiệm để kiểm tra sự ảnh hưởng các tham số cũng như cấu hình thuật toán tới độ chính xác trong việc phát hiện ngã. Các đặc trưng sẽ được kiểm tra so sánh bao gồm: (a) các đặc trưng được trích xuất trên toàn bộ khung cửa sổ thời gian, sử dụng các ngăn có độ lớn giống nhau nhờ chọn thủ công; (b) các đặc trưng được trích xuất trên toàn bộ khung cửa sổ thời gian với cách chọn ngăn tự động sử dụng phương pháp phân cụm k -means; (c) các đặc trưng được trích xuất trên các đoạn của khung cửa sổ thời gian với cách chọn ngăn tự động sử dụng phương pháp phân cụm k -means, sau đó gộp lại thành đặc trưng tổng. Để minh họa khả năng ứng dụng rộng rãi của các thuật toán đã đề xuất, các tham số được tối ưu hóa chỉ trên một tập dữ liệu là tFall, sử dụng tập tối ưu dành riêng (hold-out validation set), sau đó sử dụng các tham số tối ưu này trên cả 3 tập dữ liệu, cho tất cả các thử nghiệm. Các tham số tối ưu thu được gồm: số ngăn $M = 60$

(chọn từ tập {10, 30, 60, 90, 120, 180}) và số đoạn trong một khung $K = 8$ (chọn từ tập {4, 8, 16, 32, 64}). Mỗi phương pháp đều được áp dụng kiểm tra chéo và sử dụng SVM làm bộ phân lớp. Bảng 2.2 dưới đây tóm lược kết quả đạt được trên tập dữ liệu tFall.

Bảng 2.2. So sánh các kết quả phát hiện ngã thử nghiệm trên tập dữ liệu tFall

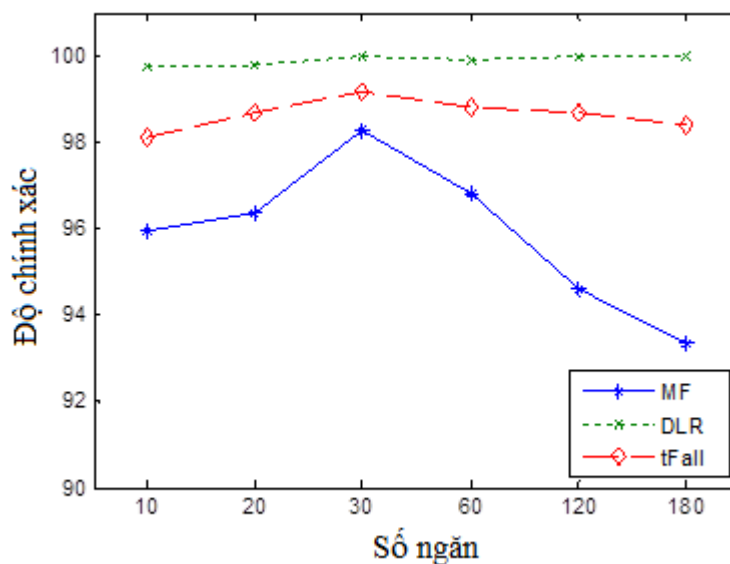
Phương pháp	Độ chính xác	Độ nhạy	Độ đặc hiệu
(a) Trích xuất trên cả khung + chọn ngăn thủ công	95,1% ± 0,35%	97,36% ± 0,27%	71,74% ± 1,25%
(b) Trích xuất trên cả khung + chọn ngăn tự động	97,47% ± 0,24%	98,75% ± 0,17%	86,45% ± 1,15%
(c) Trích xuất từng đoạn trong khung + chọn ngăn tự động	98,71% ± 0,32%	99,40% ± 0,27%	93,75% ± 1,89%

Chú ý là tất cả các phương pháp trong bảng 2.2 đều đạt độ chính xác cao (trên 95%). Để nhận thấy việc sử dụng tính toán ngăn tự động với một thuật toán phân cụm giúp tăng độ chính xác đáng kể ((b) tăng khoảng 2% so với (a)). Kết quả thậm chí còn tốt hơn với (c), cao hơn 3,5% so với (a). Có thể giải thích các kết quả này là với việc sử dụng các đặc trưng cục bộ, các thông tin về vị trí tương đối của các mảnh trong khung cửa sổ thời gian được giữ lại. Kết quả là phương pháp này giúp giữ được các đặc tính của ngã, cụ thể là các đặc trưng của các pha khác nhau trong một sự kiện ngã (pha trước khi ngã, pha tác động, pha sau khi ngã và pha hồi phục). Các đặc trưng này góp phần làm phân biệt rõ các sự kiện ngã và các hoạt động hàng ngày khác, kể cả các hoạt động có tín hiệu tương tự ngã. Dù kết quả trên bảng trên cho thấy độ nhạy của (c) không tăng nhiều lắm so với các phương pháp trích xuất đặc trưng còn lại nhưng độ đặc hiệu lại có được mức tăng đáng kể. Điều này chỉ ra rằng (c) có thể tránh được nhiều cảnh báo giả hơn so với (a) và (b).

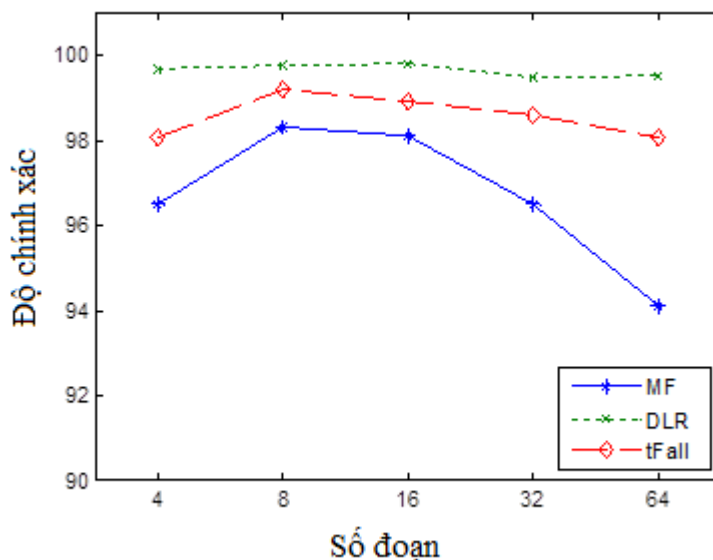
b) Ảnh hưởng của các tham số tới hiệu năng

Phần này chỉ xem xét phương pháp trích xuất đặc trưng cho kết quả tốt nhất là (c): tính ngăn tự động bằng kỹ thuật phân cụm và trên các phân đoạn trong một khung cửa sổ thời gian (HALF). Hai tham số cần xem xét trong phương pháp là số ngăn M và số phân đoạn K . Để đánh giá tác động của số ngăn đến hiệu năng của phương pháp, giá trị của số ngăn được thay đổi trong khi vẫn giữ nguyên số đoạn K không đổi như trong thử nghiệm trước đó. Sau đó thực hiện vẽ đồ chính xác trên cả 3 tập dữ liệu cùng số ngăn M và số đoạn K để xem xét sự thay đổi của hiệu năng trong hình 2.14 và hình 2.15. Như mô tả trong hình 2.14, độ chính xác tăng khi số các ngăn M tăng từ 10 đến 30 và sau đó giảm, đặc biệt giảm mạnh trong trường hợp tập dữ liệu tFall.

Hình 2.15 cho thấy độ chính xác của hệ thống phát hiện ngã tăng khi số đoạn tăng từ 4 đến 6, giảm chậm dần khi số đoạn tới 16 và rồi sau đó giảm mạnh.



Hình 2.14. Ảnh hưởng của số ngăn tới độ chính xác hệ thống



Hình 2.15. Ảnh hưởng của số đoạn trong khung cửa sổ thời gian tới độ chính xác hệ thống

c) So sánh các phương pháp phát hiện ngã

Tiếp theo đây, phương pháp đã đề xuất được so sánh với ba phương pháp khác được trình bày bởi Phạm và cộng sự [110], Jantaraprim và cộng sự [64] và Özdemir [106], sử dụng bộ phân lớp SVM trên cả 3 tập dữ liệu đã đề cập ở trên. Theo khảo sát của tác giả, hai phương pháp phát hiện ngã đầu tiên này là hai trong số những phương pháp dựa trên học máy có độ chính xác rất cao và khả thi cho các hệ thống nhỏ gọn mang trên người và yêu cầu hoạt động theo thời gian thực. Phương pháp phát hiện ngã thứ ba tuy cần nhiều tính toán hơn nhưng có độ chính xác cao hơn, đồng thời có thời gian tính toán là chấp nhận được trong một số ứng dụng ngã. Các thuật toán được cài đặt lại trong Matlab và sử dụng các giá trị tham số tốt nhất như đã mô tả trong bài báo của các tác giả này. Các kết quả được mô tả trong bảng 2.3 và được trình bày với cấu trúc độ chính xác trung bình \pm độ lệch chuẩn.

Bảng 2.3. Các kết quả phát hiện ngã trên 3 tập dữ liệu đánh giá

Phương pháp	DLR	MobiFall2	tFall
(1)	98,11% ± 0,20%	96,51% ± 0,41%	95,62% ± 0,53%
(2)	98,83% ± 0,50%	96,22% ± 0,34%	96,18% ± 0,46%
(3)	99,93% ± 0,08%	99,20% ± 0,45%	98,70% ± 0,32%
(4)	99,27% ± 0,19%	98,76% ± 0,23%	97,92% ± 0,41%

(1): Pham và cộng sự [110]; (2): Jantaraprim và cộng sự [64]; (3): Phương pháp đã đề xuất;
 (4): Özdemir và cộng sự [106]

Có thể thấy rằng độ chính xác thu được của tất cả các phương pháp trên tập dữ liệu DLR là rất cao (trên 98%) và kết quả nhận dạng trên tập dữ liệu tFall là kém nhất. Kết quả này là hợp lý do DLR có ít kiểu hoạt động hơn các tập dữ liệu khác dẫn tới phân biệt ngã dễ hơn. Trong khi đó tFall lại chứa nhiều hoạt động thời gian thực cũng như số kiểu ngã nhiều làm việc phân biệt ngã khó. Kết quả cũng chỉ ra rằng phương pháp đã đề xuất có kết quả cao nhất trên cả 3 tập dữ liệu. Mức tăng của độ chính xác với phương pháp đề xuất so với (1) và (2) là đáng kể trên cả 3 tập dữ liệu, còn so với (4) là xấp xỉ. Còn kết quả của hai phương pháp (1) và (2) là tương đương nhau và đồng thời cũng kém hơn (4). Điều này là do cả hai phương pháp (1) (sử dụng tập đặc trưng thống kê đơn giản) và (2) (sử dụng đặc trưng short-time min-max) đều không khai thác tốt đặc trưng của sự kiện ngã. Phương pháp (4) cũng sử dụng các đặc trưng thống kê như (1) nhưng có độ chính xác cao hơn do sử dụng nhiều đặc trưng có thể hiệu quả với phân biệt ngã hơn với skewness, kurtosis, tự tương quan và chuyển đổi Fourier. Trong khi đó, phương pháp do tác giả đề xuất (3), sử dụng các đặc trưng HALF có khả năng tách biệt rõ ràng các hoạt động ngã và các hoạt động hàng ngày khác nhờ độ khác biệt về thông tin góc của các mảnh trong tín hiệu ngã và các hoạt động khác (nguyên nhân là vì sự thay đổi đặc trưng tín hiệu giữa các pha của ngã so với các hoạt động khác) (xem hình 2.1). Cùng với các phân phối dựa trên góc, việc sử dụng các đặc trưng cục bộ cũng góp phần khai thác tốt các thông tin về vị trí tương đối của các mảnh trong khung cửa sổ thời gian

được giữ lại, dẫn tới nắm giữ được các đặc tính của ngã trong tất cả các pha. Các thông tin này giúp phân biệt rõ hơn các sự kiện ngã và ADL, kể cả các hoạt động có tín hiệu tương tự ngã. Ví dụ như các hoạt động đứng-ngồi, ngồi-nằm có dạng tín hiệu tương tự với ngã tại các pha trước khi ngã, tác động và thậm chí một phần của pha sau khi ngã nhưng sẽ khác biệt về phần tín hiệu trong pha hồi phục.

Về mặt tốc độ, hệ thống đề xuất thực hiện phát hiện ngã nhanh, đáp ứng được cho các hệ thống yêu cầu thời gian thực do có độ phức tạp tính toán thấp (xem phần 2.3.3). Tốc độ phát hiện ngã của hệ thống đề xuất được so sánh với ba phương pháp trong [64], [110] và [106], đo trên 1000 mẫu dữ liệu trong tập tFall.

Bảng 2.4. Tốc độ phát hiện ngã trên tập dữ liệu tFall với 1000 mẫu

Phương pháp	Tập dữ liệu tFall (1000 mẫu)
(1)	2,86 giây
(2)	1,15 giây
(3)	3,01 giây
(4)	15,02 giây

(1): Phạm và cộng sự [110]; (2): Jantaraprim và cộng sự [64]; (3): Phương pháp đã đề xuất; (4): Özdemir và cộng sự [106]

Bảng 2.4 cho thấy phương pháp (2) có tốc độ phát hiện ngã nhanh nhất trong cả 4 phương pháp. Phương pháp đề xuất có tốc độ kém (1) với tập các đặc trưng thống kê đơn giản một chút nhưng có tốc độ nhanh hơn gấp nhiều lần so với (4) với độ chính xác tương đương. (2) nhanh nhất do phương pháp trích xuất đặc trưng short time min-max có cách tính toán đơn giản hơn các phương pháp khác. Phương pháp (4) với các đặc trưng thống kê phức tạp gồm trung bình, phương sai, skewness, kurtosis, tự tương quan và chuyển đổi Fourier, cùng với bước thực hiện PCA khiến cho thời gian xử lý lâu hơn rất nhiều.

Theo số liệu kết quả về tốc độ như trên bảng 2.4 và phân tích về độ phức tạp thuật toán, có thể thấy được rằng phương pháp trích xuất đặc trưng HALF với 3 bước tính góc, rời rạc hóa và tính phân phối tần suất có khả năng tính toán nhanh do độ phức tạp thấp, phù hợp với các thiết bị trợ giúp cá nhân thông minh, nhỏ gọn, có yêu cầu xử lý thời gian thực. Ví dụ như thiết bị đồng hồ thông minh Fitbit Surge sử dụng chip Cortex M3 48MHz có năng lực tính toán kém 250 lần so với cấu hình thử nghiệm cũng có thể phát hiện ngã được sau 0,75 giây khi có dữ liệu đầu vào.

2.5. KẾT LUẬN CHƯƠNG

Chương này đã trình bày một phương pháp nhận dạng hoạt động sử dụng các đặc trưng mới HALF. Các đặc trưng này đơn giản và có thể tính toán nhanh chóng, đồng thời có khả năng phân biệt tốt một số hoạt động riêng lẻ trong nhận dạng hoạt động người, chẳng hạn như phân biệt ngã và các hoạt động khác. Các đặc trưng HALF được tính toán bằng cách gộp nhiều véc-tơ đặc trưng phân phối tần suất cục bộ, được tạo nên thông qua việc rời rạc hóa các góc của các mảnh trong mỗi đoạn cục bộ của khung cửa sổ thời gian. Các ngăn của phân phối tần suất được xác định một cách tự động, dựa trên các kỹ thuật phân cụm không giám sát. Cách tiếp cận này cho phép ứng dụng cho nhiều loại dữ liệu gia tốc khác nhau mà không bị ảnh hưởng bởi các tính chất như tham số lấy mẫu, khoảng giá trị gia tốc hay vị trí đặt cảm biến. Đặc trưng đã đề xuất được thử nghiệm bằng cách sử dụng máy véc-tơ tựa với 3 tập dữ liệu đã công bố. Các kết quả thử nghiệm trên 3 tập dữ liệu đã công bố cho thấy độ chính xác phương pháp phát hiện ngã đã đề xuất trong luận án cao hơn đáng kể so với 2 phương pháp khác [64,110] và tương đương với phương pháp trong [106] nhưng với tốc độ nhanh hơn rất nhiều. Mặc dù chưa được kiểm nghiệm bằng dữ liệu thực từ người lớn tuổi, các kết quả có được cho thấy phương pháp đã đề xuất có khả năng áp dụng rộng rãi trên các môi trường thực tế cần ứng dụng phát hiện ngã. Các kết quả nghiên cứu đã trình bày ở trên cho thấy phương pháp phát hiện ngã dựa trên đặc trưng HALF đã đề xuất có thể *đảm bảo phát hiện ngã chính*

xác cho các hệ thống mang theo người, bị hạn chế về tài nguyên và yêu cầu xử lý theo thời gian thực.

Nội dung trình bày trong chương này được tổng hợp dựa trên kết quả các công trình nghiên cứu số 1, số 2, số 4 và số 6 của tác giả. Công trình nghiên cứu số 1, số 2 là những nghiên cứu cơ sở về ngã dựa trên các đặc trưng thống kê và phương pháp so khớp uốn thời gian động (DTW). Dựa trên kết quả của các nghiên cứu này, tác giả đã cải tiến phương pháp phát hiện ngã sử dụng đặc trưng phân phối tần suất góc của các đoạn nối hai điểm dữ liệu kề (HALF), được mô tả trong công trình nghiên cứu số 4 và số 6. Sự khác biệt đáng chú ý giữa công trình nghiên cứu số 4 và công trình nghiên cứu số 6 là sử dụng cách tiếp cận phân cụm thủ công thay vì phân cụm tự động. Đồng thời, phương pháp đề xuất trong công trình số 6 đã được đánh giá và so sánh một cách tổng quát hơn với việc sử dụng 3 tập dữ liệu đã công bố và được sử dụng rộng rãi trong các nghiên cứu về phát hiện ngã.

CHƯƠNG 3

HỌC CÁC ĐẶC TRƯNG ĐA MỨC MPF

Tăng cường sự chính xác trong nhận dạng hoạt động là yếu tố quan trọng để đảm bảo thành công của các ứng dụng. Để nhận dạng chính xác hoạt động, điều quan trọng là phải xây dựng được tập đặc trưng phù hợp cho dữ liệu cảm biến. Tuy nhiên, rất khó để *đảm bảo hiệu năng hệ thống trong trường hợp cần làm việc với ứng dụng nhận dạng có số hoạt động nhiều*. Chương này đề xuất phương pháp sử dụng một loại đặc trưng chuyển động (motion features) mới, đó là motion primitive forests (MPF) để giải quyết vấn đề đã nêu. Đây là các đặc trưng được kết hợp ngẫu nhiên từ các cây quyết định dựa trên các đặc trưng cục bộ gốc ban đầu, bằng cách phân cụm chúng thành dạng *chuyển động gốc* (motion primitive) (hay còn gọi là từ - word). Các túi (bag) đặc trưng này, được tạo nên thông qua tích lũy các phân phối tần suất của các chuyển động gốc kết quả trên mỗi *khung* (frame) dữ liệu, sau đó được sử dụng để xây dựng mô hình hoạt động. Các thực nghiệm được tiến hành để kiểm tra tính hiệu quả của phương pháp đề xuất dựa trên dữ liệu gia tốc của 3 bộ dữ liệu chuẩn. Trên tất cả 3 bộ dữ liệu, phương pháp đề xuất sử dụng MPF có độ chính xác cao hơn đáng kể so với các phương pháp tốt nhất hiện nay, đồng thời phương pháp này cũng có tốc độ nhanh hơn nhiều trong cả 2 pha huấn luyện và dự đoán, so với phương pháp học đặc trưng k -means. Ngoài ra, phương pháp đề xuất cho kết quả ổn định hơn so với các phương pháp khác khi sử dụng đặc trưng cục bộ, điều này cho thấy khả năng của rừng ngẫu nhiên (random forests) trong việc lựa chọn các đặc trưng cục bộ phù hợp.

3.1. NHẬN DẠNG NHIỀU HOẠT ĐỘNG SỬ DỤNG ĐẶC TRƯNG ĐA MỨC

Như trong bất kỳ bài toán nhận dạng mẫu nào, chìa khóa thành công cho bài toán nhận dạng hoạt động là phải thiết kế được đặc trưng đại diện phù hợp cho dữ liệu cảm biến. Để nhận dạng được chính xác, lý tưởng nhất là các đặc tính phải

được thiết kế để có thể phân biệt được một cách rõ ràng giữa các lớp hoạt động khác nhau và phân biệt tốt cả với nhiều cá nhân khác nhau. Ngoài ra, trong các ứng dụng nhận dạng hoạt động người sử dụng cảm biến, các đặc trưng tốt cần phải được tính toán nhanh chóng để có thể cung cấp dự đoán trong thời gian thực trên các thiết bị có tài nguyên tính toán hạn chế. Tuy nhiên trong thực tế, để xây dựng được các hệ thống nhận dạng hoạt động hiệu quả (chính xác và có tốc độ nhanh) là rất khó khăn, đặc biệt trong các hệ thống cần nhận dạng nhiều hoạt động cùng lúc. Một trong những nguyên nhân quan trọng là do cách thức thực hiện hoạt động đa dạng của con người, dẫn tới sự khó khăn trong việc thiết kế các đặc trưng đại diện phù hợp với dữ liệu cảm biến.

Một số công trình nghiên cứu gần đây cho thấy rằng các đặc trưng đa mức có khả năng nhận dạng hoạt động tốt [55,151]. Các đặc trưng này được tính toán thông qua việc lượng tử hóa các đặc trưng cục bộ (được trích xuất từ các phần tín hiệu nhỏ của mỗi khung dữ liệu). Các nghiên cứu hiện tại sử dụng các thuật toán học không giám sát như phân cụm k -means, mô hình hỗn hợp Gaussian (GMM), hoặc mô hình chủ đề (topic modeling) [56] để tạo ra các đặc tính mức cao như vậy bằng cách nhóm các mô tả cục bộ thành một dạng được gọi là *chuyển động gốc* (motion primitive hay motion word). Việc sử dụng các *chuyển động gốc* kết hợp với phương pháp túi từ đã mang lại sự cải tiến đáng kể về độ chính xác trong nhận dạng hoạt động người do khả năng bắt giữ các thông tin mô tả cấu trúc mức cao (nhờ tận dụng tốt cả hai loại thông tin cấu trúc toàn cục và cục bộ), kể cả khi việc sử dụng túi từ đánh mất đặc trưng chuỗi thời gian của tín hiệu.

Mặc dù có thể có được kết quả với độ chính xác nhận dạng cao, nhưng các phương pháp trên lại có chi phí tính toán lớn trong cả hai pha huấn luyện và nhận dạng do chi phí để gán mô tả cục bộ (local descriptor) vào chuyển động gốc.

Chương này đề xuất một phương pháp thay thế để xây dựng tập từ vựng chuyển động (motion vocabulary) và thực hiện gán các biểu diễn đặc trưng bằng cách sử dụng một kết hợp lượng nhỏ các cây quyết định hay được gọi là motion

primitive forests (MPF). MPF hoạt động giống như các cây phân nhóm thông qua việc nhóm các mô tả cục bộ tương tự trong các nút lá. Quá trình này được định hướng bởi các nhãn của hoạt động và nhanh hơn nhiều so với k -means. Lợi thế của việc sử dụng cụm cây quyết định trong việc học từ vựng (vocabulary/codebook) đã được minh chứng trong các nghiên cứu xử lý ảnh [100], nhưng đây là lần đầu tiên được áp dụng cho nhận dạng hoạt động dựa trên cảm biến (theo khảo sát về các nghiên cứu liên quan). Đề xuất thứ hai là xây dựng các đặc trưng cục bộ mới đơn giản, phù hợp khi sử dụng với MPF. Các đặc trưng này là các giá trị bao gồm những giá trị thô chưa xử lý của các điểm dữ liệu hoặc các giá trị tổng, hiệu của 2 điểm dữ liệu. Những giá trị này có thể tính toán nhanh, đồng thời có thể giúp cho nhận dạng hoạt động chính xác như các đặc trưng cục bộ phức tạp khác.

Việc thử nghiệm được tiến hành với ba bộ dữ liệu được sử dụng rộng rãi trong các nghiên cứu nhận dạng hoạt động và kết quả cho thấy phương pháp đã đề xuất ở đây không những thực hiện tốt hơn k -means trong việc học đặc trưng mà còn có độ chính xác cao hơn các phương pháp tốt nhất hiện nay, đề xuất trong [148] và [114].

3.2. CÁC NGHIÊN CỨU CÓ LIÊN QUAN

Hầu hết các phương pháp nhận dạng hoạt động hiện nay đều dựa trên các đặc trưng thông kê hoặc đặc trưng theo tần số được thiết kế thủ công, trích xuất từ các cửa sổ thời gian (hay khung) có độ dài xác định [17]. Các đặc trưng có thể là cục bộ [151], toàn cục [44,51,56,114,148], hoặc kết hợp cả hai loại véc-tơ đặc trưng như [152]. Ví dụ như Zhang và cộng sự [150] đề xuất một lược đồ chọn đặc trưng theo thứ bậc để tăng cường hiệu năng nhận dạng. Một số các nghiên cứu khác sử dụng phương pháp học đặc trưng để có thể tự động học các đặc trưng trích xuất từ dữ liệu cảm biến sử dụng kỹ thuật phân tích thành phần chính (PCA) [51] hoặc Restricted Boltzman Machine (RBM) [114]. Một phương pháp học đặc trưng hiệu quả khác là Empirical Cumulative Distribution Function (ECDF) dựa trên PCA cũng được đề xuất trong [114] và CNN trong [148]. Các nghiên cứu này đã công bố những kết

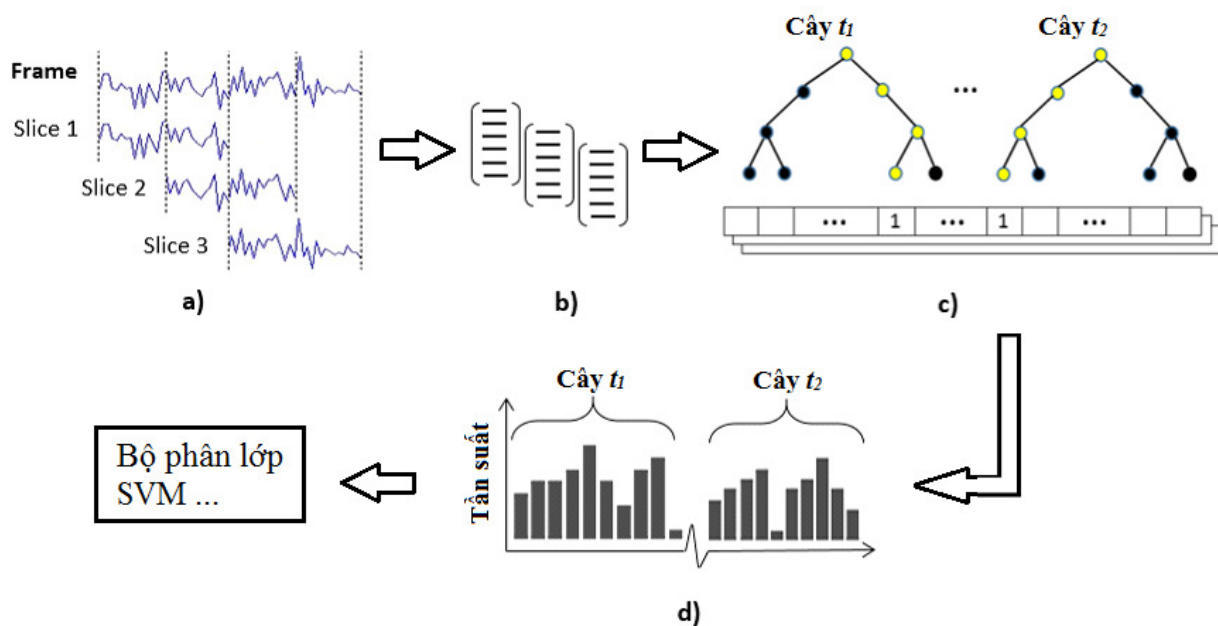
quả đáng kể trong việc nâng cao hiệu năng của hệ thống nhận dạng. Một phương pháp học đặc trưng hiệu quả khác dựa trên phương pháp túi từ cũng được sử dụng trong [56,151]. Phương pháp túi từ ban đầu được sử dụng trong phân nhóm văn bản và phân loại ảnh. Trong nghiên cứu của mình [56], Huynh sử dụng túi từ để phân loại một tập phức tạp gồm 34 hoạt động với nhiều biến đổi, kết hợp thêm với mô hình chủ đề (dựa vào Latent Dirichlet Allocation). Phương pháp túi từ đã chứng tỏ được sự hiệu quả khi có thể phân loại được 10 hoạt động hàng ngày với độ chính xác 73%.

3.3. PHƯƠNG PHÁP HỌC ĐẶC TRƯNG MPF

Phần này trình bày phương pháp mới cho nhận dạng hoạt động người - MPF, bao gồm các nội dung: tổng quan phương pháp MPF và các thành phần liên quan khác như rừng ngẫu nhiên, MPF, đặc trưng cục bộ cùng các bộ phân lớp.

3.3.1. Sơ đồ hoạt động

Sơ đồ của phương pháp đề xuất được minh họa trong hình 3.1. Đầu vào là chuỗi dữ liệu cảm biến liên tục 3 chiều từ thiết bị cảm biến gia tốc được gắn trên cơ thể người dùng. Bằng cách sử dụng một cửa sổ trượt, chuỗi dữ liệu cảm biến liên tục được phân đoạn thành các khung có kích thước bằng nhau, trong đó chiều dài của chúng có thời lượng dài hơn khoảng thời gian cho một hoạt động bất kỳ. Mỗi khung sau đó sẽ được phân chia thành các lát tín hiệu (hay các slice) bằng nhau (có thể chồng lấn) để mỗi lát tín hiệu nhỏ hơn nhiều so với khung (frame) của nó (hình 3.1(a)). Từ mỗi lát tín hiệu, trích xuất các đặc trưng để tạo thành một véc-tơ đặc trưng cục bộ. Các đặc trưng này có thể là các đặc trưng thống kê đơn giản như mean (trung bình), variance (phương sai), hay các đặc trưng phức tạp như skewness, kurtosis, v.v. (hình 3.1(b)). Trong giai đoạn huấn luyện, nếu một khung chứa một hoạt động thì nhãn của hoạt động đó sẽ được gán cho toàn bộ khung cũng như các lát tín hiệu của nó. Trong pha dự đoán, bài toán trở thành dự đoán nhãn lớp hoạt động cho mỗi khung.



Hình 3.1. (a) 3 lát tín hiệu (slice) được phân đoạn từ một khung hoạt động (frame) với độ chồng lấn là 50%, (b) các đặc trưng được trích xuất từ mỗi lát tín hiệu, (c) MPF với vai trò là từ vựng chuyển động, (d) phân phối tần suất của chuyển động gốc (motion primitive histogram)

Các véc-tơ đặc trưng cục bộ từ tất cả các lớp hoạt động (và lớp nền) được huấn luyện, sau đó được gộp lại với nhau và lượng tử hóa để tạo thành từ vựng chuyển động gốc (motion primitive vocabulary) hay là bộ mã (codebook). Đây là quá trình gán một chỉ số cho mỗi véc-tơ đặc trưng cục bộ để các véc-tơ đặc trưng tương tự nhau sẽ có xác suất cao có cùng chung chỉ số. Chất lượng của việc xây dựng chuyển động gốc là rất quan trọng đối với độ chính xác của việc nhận dạng. Các cách tiếp cận trước đây cho nhận dạng hoạt động thường sử dụng các phương pháp học không giám sát như phân cụm k -means để thực hiện việc này, bằng cách gán chỉ số nhờ tìm kiếm hàng xóm gần nhất (xem [57,151]). Phương pháp này có chi phí tính toán rất cao.

Phương pháp được đề xuất ở đây sử dụng rừng ngẫu nhiên (nhờ kết hợp ngẫu nhiên các cây quyết định) sử dụng đầu vào là các véc-tơ đặc trưng cục bộ và sau đó phân cụm chúng sao cho các véc-tơ tương tự nhau thuộc về cùng một nút lá (với xác suất cao). Các rừng ngẫu nhiên này, được gọi là motion primitive forests

(MPF), có tốc độ phân cụm nhanh hơn k -means và phương pháp gán hàng xóm gần nhất. Đồng thời có khả năng sinh ra các chuyển động gốc có độ phân biệt cao hơn kể cả khi có số lượng lớn các lát tin hiệu không nằm trong các khung có chứa hoạt động (background slice). Kết quả này có được chủ yếu là do các nhãn hoạt động được dùng trong việc định hướng xây dựng các cây quyết định.

Trong bước tiếp theo, các chuyển động gốc được học bởi MPF được kết hợp với mô hình túi từ. Các phân phối tần suất của các chỉ số nút lá cho mỗi khung được tính toán và sau đó được dùng làm đặc trưng đầu vào cho các thuật toán phân lớp như k -NN hoặc SVM.

Phần còn lại của chương này sẽ trình bày chi tiết MPF và các thành phần khác trong phương pháp đã đề xuất.

3.3.2. Rừng ngẫu nhiên

Do được ứng dụng để xây dựng MPF nên cần một mô tả tóm tắt về phương pháp rừng ngẫu nhiên [25] ở đây. Rừng ngẫu nhiên dựa trên kỹ thuật học kết hợp, là một phương pháp tạo ra nhiều bộ phân lớp và kết hợp các kết quả của chúng. Rừng ngẫu nhiên sẽ tạo ra nhiều cây quyết định mà cụ thể là thuật toán phân lớp cây phân loại và hồi quy (CART) [26], sử dụng kỹ thuật bagging. Mỗi cây được huấn luyện sử dụng một mẫu dữ liệu mới (bootstrap), lấy từ tập dữ liệu gốc ban đầu và tìm kiếm trong một nhóm nhỏ ngẫu nhiên các thuộc tính trong tập thuộc tính đầu vào để phân chia cho mức tiếp theo của cây. CART là cây quyết định nhị phân, được tạo nên bằng cách liên tiếp phân chia dữ liệu tại mỗi nút thành các nút nhỏ hơn, bắt đầu từ nút gốc chứa toàn bộ dữ liệu đầu vào. Nhờ chia nhỏ không gian tìm kiếm theo cách này, việc phân loại được thực hiện rất nhanh chóng kể cả đối với trường hợp không gian thuộc tính lớn. Tham số đầu vào của hàm phân chia bao gồm các thuộc tính được chọn trong mỗi lần phân chia. Rừng ngẫu nhiên không hạn chế số lượng cây tạo ra cũng như sự mở rộng của cây. Tuy nhiên, cần chọn số cây sẽ được sinh để đảm bảo mỗi một thuộc tính sẽ được kiểm tra một vài lần. Để xây dựng tập huấn

luyện và phương pháp kiểm tra, kỹ thuật “out-of-bag” được sử dụng trong rừng ngẫu nhiên.

Giả thiết một rừng ngẫu nhiên gồm một tập T cây quyết định như mô tả trong hình 3.1(c). Mẫu dữ liệu đi vào một cây quyết định theo các nhánh đệ quy trái hoặc phải tùy thuộc vào chức năng phân chia đã được học trên các thuộc tính cho đến khi tới được nút lá. Việc gán nhãn cuối cùng được quyết định bằng cách lấy giá trị trung bình trên tất cả T cây để có thể thu được phân lớp chính xác và tốt nhất. Ý tưởng này chính là theo nguyên tắc bỏ phiếu. Việc sinh ra nhiều cây quyết định từ một mẫu dữ liệu nhằm đa dạng hóa các “phiếu bầu”. Việc áp dụng kỹ thuật bagging để sinh ra các mẫu dữ liệu hay tập thuộc tính, hoặc việc rẽ nhánh ngẫu nhiên giúp làm tăng số lượng loại hình cũng như phiếu bầu, từ đó có được cái nhìn đa chiều, chi tiết hơn và do vậy sẽ cho kết luận có độ chính xác cao hơn, gần với thực tế hơn. Thuật toán rừng ngẫu nhiên nhanh và có thể làm việc tốt với các dữ liệu có mất mát hay không cân bằng, nhưng có thể bị quá vừa dữ liệu (over-fitting) nếu có nhiều nhiều.

Giống CART, rừng ngẫu nhiên sử dụng độ đo Gini cho độ pha tạp thông tin (impurity) để tìm ra tập phân chia tốt nhất, là tập có độ pha tạp thông tin tối thiểu tại mọi nút [25]. Độ pha tạp thông tin Gini là một cách đo phân bố nhãn của các lớp trong một nút, lấy giá trị trong khoảng $[0, 1]$, trong đó giá trị 0 có được khi mọi phân tử trong một nút đều thuộc về cùng một lớp.

Hình 3.2 minh họa thuật toán rừng ngẫu nhiên.

Tên chương trình: Random_Forest

Đầu vào: T là số cây được tạo, N là số mẫu trong tập dữ liệu

Đầu ra: $Y_{\text{rf}}(x)$ – nhãn dự báo

Chương trình:

1. Lặp lại với $t = 1, \dots, T$
 - a. Tạo mẫu dữ liệu mỗi N' với kích thước N từ tập dữ liệu gốc – có thay thế.
 - b. Tạo cây phân lớp từ mỗi N'
 - c. Lặp lại với $i = 1, \dots, \text{số nút}$
 - i. Lấy ngẫu nhiên m_{try} thuộc tính từ M thuộc tính ban đầu
 - ii. Chọn tập phân chia tốt nhất (best split) trong tập m_{try} thuộc tính
 - d. $y_t(x)$ = lớp dự báo của cây thứ t
2. Trả về $Y_{\text{rf}}(x)$ = đầu phiếu đa số $\{y_t(x)\}^T$ (lớp dự báo cuối cùng nhờ đầu phiếu đa số)

Hình 3.2. Thuật toán rừng ngẫu nhiên cho phân lớp

Rừng ngẫu nhiên đã được chứng minh là có hiệu quả trong cả nhiệm vụ phân lớp và phân cụm [100,123]. Nghiên cứu này sử dụng rừng ngẫu nhiên để thực hiện phân cụm và ánh xạ véc-tơ đặc trưng cục bộ thành các chuyển động gốc.

Xây dựng rừng ngẫu nhiên

Cho một tập dữ liệu N , mỗi phần tử được đại diện bởi D đặc trưng $A = \{a_1, \dots, a_D\}$ và có cùng nhãn của mỗi hoạt động trong một khung. Mỗi cây sẽ được huấn luyện trên một tập con N' được lấy ra ngẫu nhiên từ N . Cây được xây dựng đệ quy từ trên xuống dưới. Dữ liệu huấn luyện N_p của nút p được phân chia thành 2 tập con tách rời nhau N_l trái và N_r phải theo một ngưỡng c của hàm chia f của véc-tơ đặc trưng a :

$$N_l = \{n \in N_p \mid f(a_n) < c\} \quad (3.1)$$

$$N_r = N_p \setminus N_l \quad (3.2)$$

Tại mỗi nút chia (*split node*) sinh ra một nhóm gồm m thuộc tính đề xuất cho hàm f một cách ngẫu nhiên và tìm kiếm trên tất cả các giá trị c có thể có, và chọn ra một giá trị để từ đó có được chỉ số Gini cao nhất:

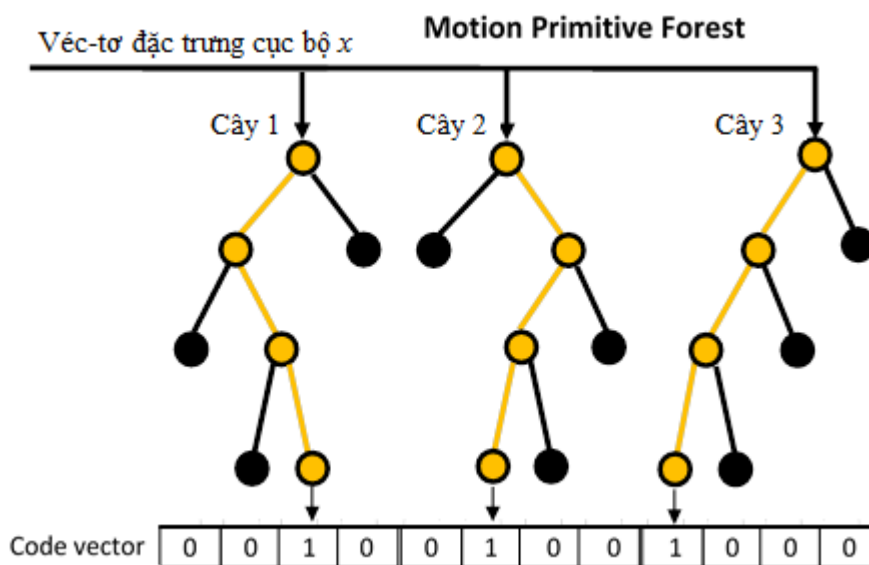
$$\Delta I_G(N_p) = I_G(N_p) - \frac{|N_l|}{|N_p|} I_G(N_l) - \frac{|N_r|}{|N_p|} I_G(N_r) \quad (3.3)$$

Trong đó $I_G(N)$ là chỉ số Gini cho dữ liệu huấn luyện N . Trong suốt quá trình học, dữ liệu huấn luyện được chia đệ quy tại mỗi nút p cho đến khi tất cả các phần tử của N_p là của cùng một lớp hoặc ít hơn một số lượng nhất định.

3.3.3. MPF

MPF (motion primitive forests) là rừng ngẫu nhiên được sử dụng để phân cụm và ánh xạ các véc-tơ đặc trưng cục bộ thành chuyển động gốc (hình 3.1(c)). Trong rừng ngẫu nhiên chuẩn [25], việc biểu quyết (voting) trên cụm các cây được sử dụng để làm giảm sự thay đổi nhãn lớp. Ở đây không sử dụng biểu quyết nhưng sẽ xử lý mỗi nút lá từ mỗi cây như là một chuyển động gốc riêng biệt. Nói cách khác, các nút lá xác định một phân vùng, và mỗi nút lá tương ứng với một cụm các véc-tơ đặc trưng cục bộ tương tự nhau.

Đối với mỗi véc-tơ đặc trưng cục bộ đầu vào, MPF trả về một tập các chỉ số của các nút lá, mỗi chỉ số nằm trong một cây. Các chỉ số nút lá này được dùng để tạo thành véc-tơ mã (code vector) như minh họa trong hình 3.3, trong đó “1” là véc-tơ đầu vào đã được ánh xạ tương ứng với chuyển động gốc.



Hình 3.3. Véc-tơ mã (code vector) tạo ra từ một rừng ngẫu nhiên với 3 cây cho một véc-tơ đặc trưng cục bộ. Đường dẫn đến các nút lá kết quả được hiển thị bằng màu vàng.

Sau đó, theo cách tiếp cận từ [48], thực hiện tổng hợp các véc-tơ mã của tất cả các lát tin hiệu thuộc về một khung để tạo thành véc-tơ phân phối tần suất các chuyển động gốc cho mỗi khung này (Hình 3.1(d)). Véc-tơ phân phối tần suất sau khi đã chuẩn hóa được sử dụng làm đặc trưng của khung trong bộ phân lớp cuối cùng. Do vậy ở đây bộ phân lớp cuối cùng sẽ có nhiệm vụ biểu quyết (voting) chứ không phải rừng ngẫu nhiên như các trường hợp khác.

Trong phương pháp đã đề xuất, kích thước từ vựng bằng với tổng số các nút lá của tất cả các cây trong rừng. Để kiểm soát kích thước từ vựng, cần thực hiện kiểm soát kích thước của các cây bằng cách cắt tỉa chúng cho đến khi đạt số lượng các nút lá mong muốn. Dưới đây sẽ trình bày minh họa thuật toán MPF (xem hình 3.4).

Tên chương trình: MPF_Extraction

MPF là rừng ngẫu nhiên đã được chỉnh sửa, sau khi huấn luyện sẽ trả về danh sách các nút lá, bằng 1 nếu nút lá là đầu ra và bằng 0 trong trường hợp ngược lại.

Đầu vào: $trained_mpf$ là MPF đã huấn luyện, x_i là mảng véc-tơ đặc trưng cục bộ đầu vào của MPF tính cho các lát tín hiệu trong cùng một khung hoạt động, với $i = 1, \dots, n_s$; n_s là số lát tín hiệu của 1 khung.

Đầu ra: véc-tơ đặc trưng f_{MPF} cho khung đầu vào

Chương trình:

1. Lặp lại với $i = 1, \dots, n_s$

Tính véc-tơ mã thứ i , độ dài bằng số nút lá của MPF:

$$code_vector_i = trained_mpf(x_i)$$

2. Véc-tơ phân phối tần suất $f_{Histogram} = \text{sum}(code_vector_i)$ với $i = 1, \dots, n_s$

3. Chuẩn hóa $f_{Histogram}$ trong khoảng $[0, 1]$

Hình 3.4. Thuật toán trích xuất đặc trưng MPF cho mảng véc-tơ đặc trưng cục bộ đầu vào thuộc một khung hoạt động

3.3.4. Đặc trưng cục bộ

Các loại đặc trưng cục bộ khác nhau (như mô tả trong phần 1.4.1) đã được đề xuất cho nhận dạng hoạt động, ví dụ như đặc trưng thống kê, đặc trưng vật lý, đặc trưng cho các thông số môi trường hay dấu hiệu sống. Chú ý rằng, do lát tín hiệu có độ dài ngắn và chỉ là một phần rất nhỏ trong một khung hoạt động, không có khả năng bắt giữ được đầy đủ thông tin của một hoạt động, cho nên các đặc trưng cần phải được lựa chọn một cách hợp lý. Ví dụ, các đặc trưng thống kê phức tạp như skewness hay kurtosis là không phù hợp do không thể tin cậy khi tính toán trong đoạn có độ dài quá nhỏ [151]. Trong số các đặc trưng đã khảo sát, các đặc trưng

thống kê đơn giản và đặc trưng vật lý là phù hợp nhất để diễn tả cho các lát tín hiệu nhỏ [151] và do đó sẽ được sử dụng trong các thử nghiệm trong chương này.

Tất cả các đặc trưng ở trên yêu cầu một số tính toán trên các điểm dữ liệu của một lát tín hiệu. Để có thể đơn giản hơn việc tính toán, các đặc trưng cục bộ mới được giới thiệu, trong đó yêu cầu rất ít hoặc không cần tính toán bởi vì chúng là giá trị thô chưa xử lý của các điểm dữ liệu hoặc là các tổng hay hiệu của giá trị của 2 điểm dữ liệu. Cụ thể, phương pháp này sử dụng ba loại đặc trưng đơn giản sau đây: (i) giá trị $p_{i,j}$ của điểm dữ liệu với chỉ số i trong trục j (đặc trưng đơn lẻ - unary feature), (ii) tổng $p_{i_1,j_1} + p_{i_2,j_2}$ (đặc trưng tổng - sum feature), và (iii) hiệu $p_{i_1,j_1} - p_{i_2,j_2}$ (đặc trưng hiệu - difference feature) của một cặp các điểm dữ liệu từ các trục j_1 và j_2 , (j_1 và j_2 có thể giống hoặc khác nhau), và $i_1 - i_2 = 1$ (tổng và hiệu của 2 điểm kề nhau). Giả sử 1 khung có độ dài l_s , thì véc-tơ đặc trưng cho 1 trục sẽ có độ dài là $(3 \times l_s - 2)$. Như trình bày trong phần thử nghiệm, các đặc trưng đơn giản này, khi được kết hợp với MPF, sẽ cung cấp độ chính xác có thể so sánh được với các đặc trưng phức tạp hơn, trong khi chi phí tính toán lại thấp hơn nhiều.

Bảng 3.1. Các đặc trưng cục bộ

Tập đặc trưng	Các đặc trưng
1	Trung bình, độ lệch chuẩn, trung bình bình phương, trung bình đạo hàm, tỷ lệ vượt qua trung bình.
2	Trung bình và phương sai của cường độ chuyển động, diện tích độ lớn tín hiệu chuẩn hóa, trị riêng của các hướng chính, vận tốc trung bình theo hướng di chuyển và trung bình năng lượng gia tốc.
3	Đặc trưng đơn lẻ (unary), tổng (sum) và hiệu (difference)

Khi phân cụm k -means được dùng để lượng tử hóa véc-tơ đặc trưng cục bộ, điều quan trọng là phải chọn được các đặc trưng cục bộ phù hợp. Các đặc trưng không phù hợp sẽ cho kết quả về khoảng cách giữa các véc-tơ không chính xác, dẫn tới việc phân cụm kém. Ở đây, một dự đoán được đưa ra là MPF có khả năng chọn ra các đặc trưng cục bộ phù hợp trong khi học hàm phân chia. Để xác minh điều này, thực hiện nhóm các đặc trưng thành 3 tập như trong bảng 3.1, và xây dựng MPF bằng cách sử dụng mỗi tập trong chúng. Tập đặc trưng đầu tiên bao gồm 5 đặc trưng thông kê đơn giản và phổ biến sử dụng cho dữ liệu chuỗi thời gian như trong nhiều nghiên cứu về đặc trưng đa mức, trong khi tập đặc trưng thứ hai phức tạp hơn (gồm 6 đặc trưng) do sử dụng các đặc trưng vật lý được dẫn xuất từ các tham số vật lý của các chuyển động của con người [150]. Chi tiết về các đặc trưng này được mô tả trong phần 1.4.1 của chương 1. Tập đặc trưng thứ ba bao gồm các đặc trưng đơn giản đã được đề xuất và số lượng các đặc trưng phụ thuộc vào độ dài của lát tín hiệu.

3.3.5. Bộ phân lớp

Nhiều bộ phân lớp đã được sử dụng trong nhận dạng hoạt động bao gồm cây quyết định, phương pháp Bayesian, k -NN, mạng nơron, mô hình Markov và kết hợp các bộ phân lớp [78]. Thực nghiệm này sử dụng hai bộ phân lớp: phương pháp phân lớp phi tham số k -NN do sự phổ biến của nó trong hoạt động nhận dạng; và máy véc-tơ tựa SVM do đã được chứng minh là có độ chính xác trong dự báo rất tốt trong nhiều miền ứng dụng, bao gồm cả nhận dạng hoạt động [146].

Đối với bộ phân lớp k -NN, nghiên cứu này sẽ chọn $k = 1$ để dễ so sánh và đánh giá các tham số. Đối với bộ phân lớp SVM, trước khi huấn luyện và dự đoán, cần chuẩn hóa các véc-tơ để tất cả các giá trị đặc trưng nằm trong khoảng $[0, 1]$. Đồng thời cũng tiến hành thử nghiệm với một vài dạng hàm nhân và chọn các thông số tốt nhất bằng cách sử dụng thủ tục tìm kiếm lưới (grid-search) hay được dùng.

3.3.6. Độ phức tạp thuật toán

Độ phức tạp tính toán cho phương pháp đề xuất là số các phép tính cần để xây dựng mô hình và gán nhãn cho đặc trưng cục bộ đầu vào (dự đoán mẫu đầu vào). Ở đây sẽ đánh giá độ phức tạp thuật toán của thuật toán MPF đề xuất (mục 3.3.3) và so sánh với thuật toán k -means. Đây là sở cứ để chứng minh MPF có tốc độ nhanh hơn các phương pháp dựa trên k -means.

Độ phức tạp thuật toán trong trường hợp xấu nhất khi xây dựng một cây quyết định CART là $O(pN^2 \log N)$ [26], với N là tổng số lát tin hiệu của tập dữ liệu đầu vào (số mẫu của dữ liệu đầu vào cho cây), p là số thuộc tính đầu vào và D là số thuộc tính được chọn ngẫu nhiên tại mỗi nút. Khi đó, độ phức tạp khi xây dựng rừng ngẫu nhiên với T cây là $O(TD\tilde{N}^2 \log \tilde{N})$ trước khi cắt tỉa các cây, với $\tilde{N} = 0,632N$, do trong thực tế trung bình số mẫu mỗi chọn ra để xây dựng cây là 63,2%. Đối với k -means, độ phức tạp thuật toán là $O(pNki)$, với i là số lần lặp k -means phải thực hiện khi chạy, k là số cụm (cluster) [146]. Trong thực tế, phương pháp đề xuất sử dụng số cây M nhỏ hơn 10, và lấy D theo mặc định khi xây dựng rừng ngẫu nhiên, với $D = \sqrt{p}$ [25] nên có độ phức tạp chỉ là $O(\sqrt{p} \tilde{N}^2 \log \tilde{N})$.

Độ phức tạp thuật toán trong trường hợp xấu nhất cho việc gán nhãn một đặc trưng cục bộ của một rừng với T cây là $O(T \log N)$ [25]. Trong trường hợp ở đây số cây nhỏ hơn 10 (xem thực nghiệm) nên có thể bỏ qua, dẫn tới độ phức tạp yêu cầu là $O(\log N)$, trong khi k -means có độ phức tạp thuật toán là $O(kp)$ [146]. Như vậy, có thể thấy rằng thuật toán được đề xuất chạy nhanh hơn rất nhiều so với trường hợp dùng k -means trong việc gán nhãn cho các đặc trưng cục bộ đầu vào.

3.4. TẬP DỮ LIỆU THỬ NGHIỆM

Phần này trình bày về ba tập dữ liệu được sử dụng trong thử nghiệm, việc thiết lập thử nghiệm và trình bày chi tiết các thử nghiệm để đưa ra những phân tích và đánh giá về phương pháp đề xuất trong các phần sau.

3.4.1. Activity Prediction (AP)

Tập dữ liệu AP [74] là một tập dữ liệu đóng (close-dataset) (không có hoạt động nền), trong đó bao gồm dữ liệu gia tốc cho 6 hoạt động hàng ngày gồm “chạy bộ”, “đi bộ”, “đi lên cầu thang”, “đi xuống cầu thang”, “ngồi xuống”, “đứng lên”, được thực hiện bởi 36 cá nhân. Các dữ liệu được thu thập từ điện thoại di động được đặt trong túi của người sử dụng theo các thiết lập thử nghiệm và cảm biến được lấy mẫu ở mức 20 Hz. Trong thử nghiệm ở dưới đây, tổng cộng tập dữ liệu AP có 29000 khung. Đây là tập dữ liệu lớn nhất sử dụng trong thử nghiệm.

3.4.2. Opportunity (OP)

Tập dữ liệu OP [119] là một bộ dữ liệu mở (open-dataset) (bao gồm cả các hoạt động chưa được biết (unknown), chứa các dữ liệu từ nhiều cảm biến đeo trên người hoặc nhúng vào các đối tượng được thao tác bởi người dùng. Người dùng thực hiện các hoạt động khác nhau trong một môi trường nhà bếp chẳng hạn như “mở tủ lạnh”, “đóng tủ lạnh”, “lau bàn”, “uống nước”, “chuẩn bị món ăn”, v.v. và tổng cộng có 11 hoạt động đã được chú thích. Nghiên cứu sử dụng một tập con của tập dữ liệu, trong đó bao gồm dữ liệu từ một bộ cảm biến đeo trên tay phải của một đối tượng. Các tập con tương tự cũng được sử dụng bởi [114] và [148]. Dữ liệu được lấy mẫu với tần số 64 Hz, tạo ra khoảng 4200 khung hoạt động, bao gồm 10 hoạt động có nhãn riêng và một hoạt động không biết trước.

3.4.3. Skoda (SK)

Tập dữ liệu Skoda Mini Checkpoint (SK) [147] là một tập dữ liệu mở được thu thập từ nhiều cảm biến gia tốc gắn trên người một công nhân trong dây chuyền lắp ráp các linh kiện trong nhà máy sản xuất xe hơi. Tập dữ liệu SK chứa 46 hoạt động lắp ráp như “mở mui xe”, “đóng cửa bên trái”, “kiểm tra tay lái”, v.v. Thử nghiệm này được giới hạn bằng cách lấy dữ liệu thuộc về một cảm biến được đeo trên cánh tay phải và tập dữ liệu có được gồm 10 hoạt động cộng với các hoạt động

chưa được biết (unknown). Với tốc độ lấy mẫu 48 Hz, tập dữ liệu bao gồm 7500 khung. Tập con tương tự cũng được sử dụng bởi [114] và [148].

3.5. THỬ NGHIỆM VÀ KẾT QUẢ

3.5.1. Phương pháp thử nghiệm

Thử nghiệm này dùng độ đo là độ chính xác tổng thể để đánh giá phương pháp đề xuất và so sánh với các phương pháp khác. Độ chính xác tổng thể được tính bằng tỷ lệ số khung được phân loại chính xác trên tổng số khung. Đây là tham số về hiệu năng được sử dụng rộng rãi nhất trong nhận dạng hoạt động vì có thể dễ dàng so sánh giữa các phương pháp khác nhau, với nhiều loại hoạt động [28,114,151].

Để so sánh với các phương pháp trong [114] và [148], các phương pháp và thiết lập cũng được đánh giá bằng cách kiểm tra chéo trên 10 phần dữ liệu (10-fold cross validation) cho các tập dữ liệu AP và OP, kiểm tra chéo trên 4 phần dữ liệu (4-fold cross validation) trên tập dữ liệu SK. Với tất cả các thực nghiệm, độ chính xác trung bình trên các phần (fold) sẽ được tính. Với mỗi phần, 10% của tập dữ liệu huấn luyện (qua lấy mẫu phân tầng) được lấy ra để làm tập kiểm tra cho việc hiệu chỉnh các tham số và sau đó các tham số này sẽ được sử dụng cho tập kiểm tra.

Thử nghiệm này sử dụng LibSVM [29] với hàm nhân RBF (thực nghiệm cho thấy hàm nhân RBF có kết quả tốt nhất) và cài đặt 1-NN. Thông số C và γ của SVM được chọn bằng phương pháp grid-search trên tập kiểm tra. MPF được cài đặt bằng cách sửa đổi cài đặt rừng ngẫu nhiên được cung cấp bởi Breiman, ở đó giữ số lượng các đặc trưng được chọn tại từng phần được chia như thiết lập mặc định ban đầu (tức là căn bậc hai của số các đặc trưng). Tất cả các thử nghiệm được chạy trên một máy tính chạy hệ điều hành Windows với bộ xử lý 2,8 GHz dual-core và RAM 8GB.

3.5.2. Kết quả thử nghiệm và đánh giá

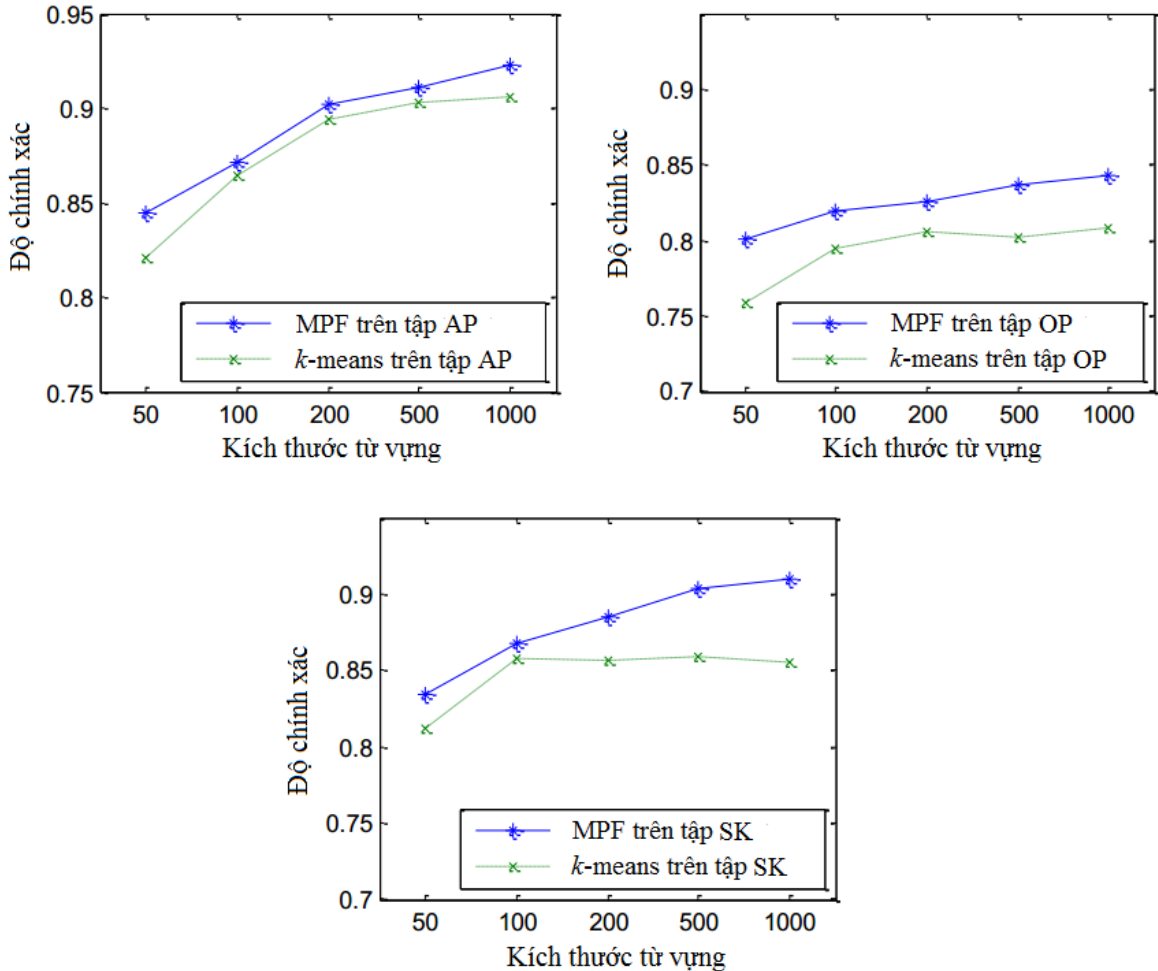
a) So sánh với phương pháp học đặc trưng sử dụng k -means

Thử nghiệm đầu tiên được thiết kế để so sánh MPF với phương pháp học đặc trưng dựa trên k -means, trong đó phương pháp phân cụm k -means được sử dụng để xây dựng từ vựng và tìm kiếm hàng xóm gần nhất để gán chuyển động gốc [55,71,151]. Để so sánh, một khung làm việc được dùng chung cho cả hai phương pháp, và chỉ thay phương pháp phân cụm k -means và gán hàng xóm gần nhất bởi MPF. Kết quả so sánh cho thấy mức cải thiện về độ chính xác tăng lên khá nhiều khi dùng phương pháp MPF.

Để việc so sánh được công bằng, trong quá trình thử nghiệm, hầu hết các thiết lập tương tự như khi dùng trong [151] và [55] sẽ được giữ lại. Cụ thể là sẽ sử dụng tập đặc trưng 1 trong bảng 3.1, giống như tập đặc trưng thống kê được sử dụng bởi Zhang và Sawchuk [151] và dùng SVM làm bộ phân lớp. Đồng thời cũng sử dụng lát tin hiệu với cùng kích thước và không chồng lấn như trong mô tả về nghiên cứu của các tác giả này. Các thông số khác được tối ưu hóa riêng cho từng phương pháp bằng cách sử dụng các tập kiểm tra.

Hình 3.5 cho thấy độ chính xác của MPF và k -means cho các kích cỡ khác nhau của từ vựng (kích cỡ từ vựng chính là bằng với số các nút lá trong trường hợp MPF và số các cụm trong trường hợp k -means) trên mỗi tập dữ liệu. Như đã trình bày, kích thước từ vựng tốt nhất thay đổi trên tập dữ liệu nhưng MPF có độ chính xác tốt hơn so với k -means trong mọi trường hợp. Kết quả này có được chủ yếu là do các nhãn hoạt động được dùng trong việc định hướng xây dựng các cây quyết định trong MPF. Nhờ vậy, MPF có khả năng sinh ra các chuyển động gốc có độ phân biệt cao hơn kể cả khi có số lượng lớn các lát tin hiệu không nằm trong các khung có chứa hoạt động (background slice). Chính vì vậy mà có sự khác biệt đáng kể về độ chính xác giữa hai phương pháp cho tập dữ liệu OP và tập dữ liệu SK. Đối với trường hợp không chứa hoạt động nền (trường hợp tập dữ liệu AP) thì độ chính xác của hai phương pháp MPF và k -means là không nhiều. Như vậy, MPF sẽ thích hợp sử dụng trong thực tế hơn k -means do hoạt động hàng ngày rất đa dạng (không phải tập dữ liệu đóng). Có một chú ý là MPF có xu hướng thích hợp với kích thước

từ vựng lớn hơn k -means. Ví dụ, trên tập dữ liệu SK, độ chính xác của k -means giảm khi kích thước vượt quá 100, trong khi độ chính xác của MPF tiếp tục tăng.



Hình 3.5. So sánh các thuật toán xây dựng từ vựng trên ba tập dữ liệu

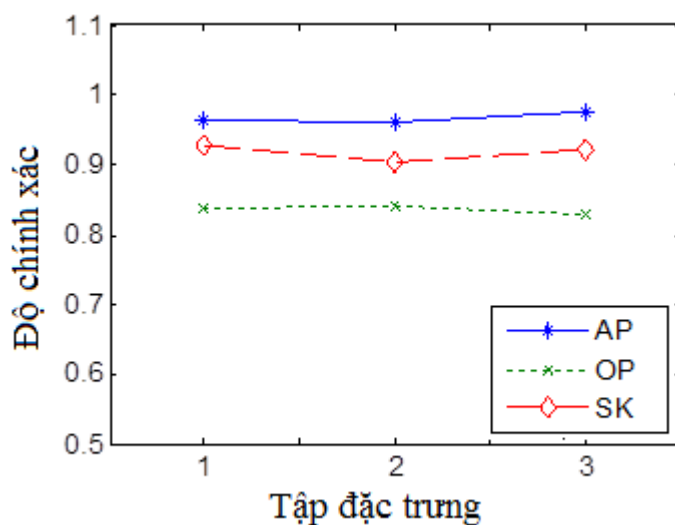
Ở đây không so sánh trực tiếp thời gian chạy của hai phương pháp vì MPF được cài đặt bằng C++ và k -means được cài đặt bằng ngôn ngữ Matlab, do đó kết quả chi tiết không được đưa ra. Tuy nhiên, có thể nhận thấy rằng trong khi MPF có thời gian chạy tăng không đáng kể khi kích thước từ vựng tăng lên, thì thời gian chạy của k -means lại tăng tuyến tính. Với kích thước là 200, thời gian chạy của k -means đã lớn hơn nhiều so với MPF (580 giây so với 54 giây, trên tập dữ liệu SK).

Kết quả này có được là do việc tìm kiếm và gán biểu diễn đặc trưng sử dụng phương pháp k hàng xóm gần nhất trong k -means có chi phí tốn kém, trong khi

trong MPF, do không gian tìm kiếm được chia nhỏ khi xây dựng các cây quyết định nhị phân trong rừng MPF, việc tìm kiếm và gán biểu diễn đặc trưng được thực hiện rất nhanh chóng kể cả đối với trường hợp số chiều đầu vào của dữ liệu lớn. Kết quả này cũng phản ánh đúng mức độ phức tạp thuật toán của cả hai thuật toán đã được trình bày trong mục 3.3.6.

b) Ảnh hưởng của các đặc trưng cục bộ

Các thử nghiệm tiếp theo được thiết kế để đánh giá ảnh hưởng của các loại đặc trưng cục bộ khác nhau. Để chứng minh tính tổng quát của MPF, thực hiện điều chỉnh các tham số chỉ cho tập dữ liệu SK bằng cách sử dụng tập kiểm tra riêng (hold-out set), và sử dụng các tham số này cho cả ba tập dữ liệu. Các thông số điều chỉnh là: kích thước lát tin hiệu là 8 với 87,5% chồng lấn, số cây là 5, số nút lá trên cây là 500. Phương pháp kiểm tra chéo được thực hiện lần lượt cho mỗi tập đặc trưng trong bảng 3.1 với bộ phân lớp SVM.



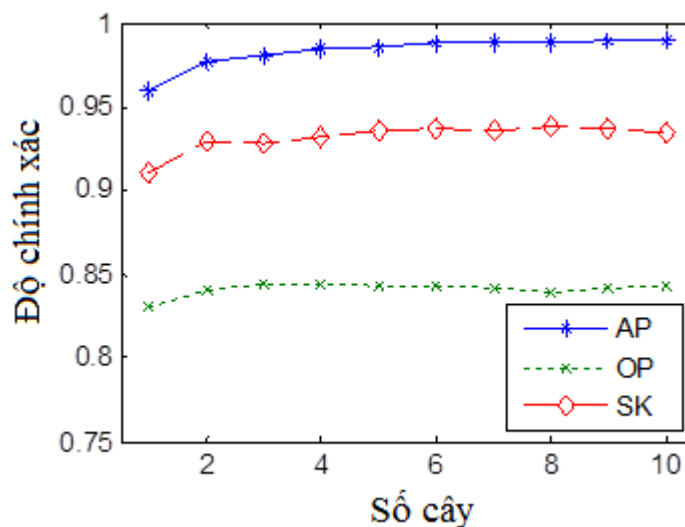
Hình 3.6. Ảnh hưởng của các đặc trưng cục bộ tới độ chính xác

Hình 3.6 thể hiện độ chính xác đạt được của từng tập đặc trưng trên 3 tập dữ liệu. Kết quả cho thấy rằng các kiểu đặc trưng đang xem xét đều có độ chính xác tương đối giống nhau với độ dao động rất nhỏ, trên toàn bộ các tập dữ liệu. Kết quả này khá quan trọng vì nó cho thấy hai ý nghĩa. Thứ nhất, MPF ít nhạy cảm với việc

lựa chọn các đặc trưng cục bộ, có thể vì các cây có thể tự chọn ra được các đặc trưng hợp lý tại các nút. Thứ hai, các đặc trưng đơn giản từ tập đặc trưng số 3 cũng có độ chính xác tương đương như các tập đặc trưng thống kê và vật lý phức tạp hơn, trong khi tốc độ tính toán lại nhanh hơn.

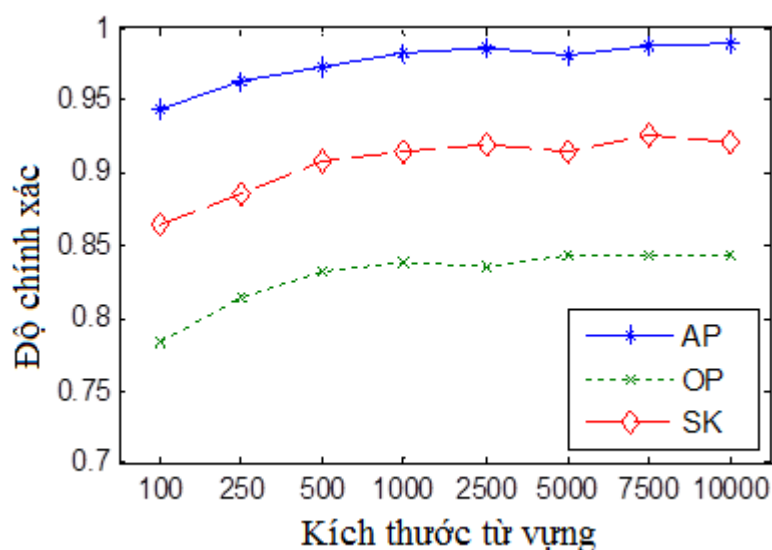
c) Ảnh hưởng của số lượng cây và kích thước của từ vựng

Thử nghiệm tiếp theo sẽ xem xét sự ảnh hưởng của 2 tham số quan trọng đến độ chính xác, đó là số lượng cây trong rừng và kích thước của từ vựng. Ở đây sử dụng tập đặc trưng thứ 3 và giữ các tham số khác giống như các thử nghiệm trước và chỉ thay đổi số lượng cây và kích thước từ vựng.



Hình 3.7. Ảnh hưởng của số lượng cây tới độ chính xác

Hình 3.7 cho biết độ chính xác khi phân lớp với số lượng cây thay đổi. Như thấy trên hình, độ chính xác tăng khi số lượng cây tăng từ 1 đến 5 trên các tập dữ liệu AP và SK, và từ 1 đến 4 trên tập dữ liệu OP. Sau đó, độ chính xác có xu hướng ổn định lại và việc sử dụng nhiều cây hơn cũng không mang lại sự cải thiện đáng kể về độ chính xác. Do vậy, có thể kết luận rằng một tập hợp nhỏ tầm 5 cây là đủ để giảm phương sai (variance) và đạt được độ chính xác cao khi nhận dạng.



Hình 3.8. Ảnh hưởng của kích thước từ vựng tới độ chính xác

Để xem xét ảnh hưởng của kích thước từ vựng đối với hiệu năng phân lớp, số lượng các nút lá sử dụng khi huấn luyện MPF được thay đổi để theo dõi biến đổi của độ chính xác. Như thấy trong hình 3.8, độ chính xác trên 3 tập dữ liệu tăng khi kích thước từ vựng tăng tới 2500 và sau đó đi vào ổn định. Lưu ý là, kích thước này lớn hơn nhiều so với các báo cáo trước đây khi sử dụng k -means hoặc phân cụm GMM [55,151]. Kết quả trên cũng chỉ ra rằng MPF không phải là rất nhạy cảm với số lượng cây và kích thước từ vựng, miễn là chúng đủ lớn.

d) So sánh với các phương pháp khác

Thử nghiệm cuối cùng sẽ so sánh độ chính xác nhận dạng của MPF khi sử dụng các đặc trưng cục bộ đơn giản với 2 phương pháp được cho là tốt nhất hiện nay. Phương pháp thứ nhất là học đặc trưng sử dụng PCA và ECDF (viết gọn lại là PCA+ECDF) kết hợp với bộ phân lớp 1-NN của [114]. Phương pháp thứ hai sử dụng Convolutional Neural Networks (CNN) với partial weight sharing cho cả việc học đặc trưng và phân lớp [148]. Các phương pháp trên được cài đặt lại trong Matlab và các giá trị tham số tốt nhất được giữ lại như trong báo cáo của các tác giả này. Đồng thời MPF cũng được sử dụng để chạy với cả 1-NN và SVM.

Bảng 3.2. So sánh độ chính xác trong phân lớp

	AP	OP	SK
MPF + 1-NN	97,93%	84,17%	92,58%
MPF + SVM	98,48%	85,68%	95,08%
PCA + ECDF	95,75%	79,39%	90,48%
CNN partial weight sharing	96,21%	77,34%	86,73%

Bảng 3.2 tóm tắt độ chính xác của các phương pháp. Tất cả các phương pháp đều đạt được độ chính xác cao trên tập dữ liệu AP (hơn 95%) trong khi tập dữ liệu OP cho thấy độ chính xác kém nhất. Cả hai biến thể của MPF (với 1-NN và SVM), luôn đạt được độ chính xác cao hơn PCA+ECDF và CNN trong cả ba tập dữ liệu. SVM cho kết quả tốt hơn so với 1-NN, điều này là hợp lý vì số chiều của dữ liệu đầu vào (số chiều các véc-tơ phân phối tần suất) là khá cao (SVM là bộ phân lớp hoạt động tốt khi làm việc với dữ liệu có số chiều lớn). Những cải tiến về độ chính xác của MPF so với PCA+ECDF và CNN là đặc biệt đáng chú ý trên các tập dữ liệu OP và SK (gần 10% sự khác biệt giữa MPF và CNN). Kết quả này là do các nhãn hoạt động được dùng trong việc định hướng xây dựng các cây quyết định trong MPF. Nhờ vậy, MPF có khả năng sinh ra các chuyển động gốc có độ phân biệt cao hơn kể cả khi có số lượng lớn các lát tín hiệu không nằm trong các khung có chứa hoạt động (background slice), kéo theo độ chính xác cao trong nhận dạng hoạt động. Kết quả cũng cho thấy không có sự chênh lệch rõ ràng giữa CNN và PCA+ECDF.

3.6. KẾT LUẬN CHƯƠNG

Chương này đã trình bày motion primitive forests, là loại đặc trưng chuyển động mới, hiệu quả trong nhận dạng hoạt động người. Các rừng (forests) này lượng tử hóa các véc-tơ đặc trưng cục bộ của các đoạn dữ liệu nhỏ theo cách có giám sát để tạo ra từ vựng chuyển động gốc. Bằng cách sử dụng dữ liệu có nhãn, phương

pháp này có thể tạo ra chuyển động gốc có độ phân biệt cao. Ngoài ra, phương pháp đề xuất có thể tự động lựa chọn các đặc trưng cục bộ phù hợp, do đó ít nhạy cảm hơn với những cách thức thủ công dựa trên thuật giải heuristic hay dùng để lựa chọn các đặc trưng cục bộ. Và đồng thời, phương pháp này có thể áp dụng một cách trực tiếp vào MPF trên các đặc trưng rất đơn giản như các giá trị nguồn của các điểm dữ liệu mà vẫn cải tiến được độ chính xác nhận dạng. Kết hợp với phương pháp túi từ, MPF có sự cải thiện hơn về độ chính xác nhận dạng so với các phương pháp nhận dạng tốt nhất hiện nay khi được đánh giá trên ba tập dữ liệu chuẩn. Kết quả cho thấy phương pháp đề xuất là một phương pháp nhận dạng hoạt động người có *độ chính xác cao kể cả trong trường hợp cần nhận dạng nhiều hoạt động và có tốc độ tính toán tương đối nhanh.*

Trong tương lai, có thể xem xét việc thử nghiệm học không giám sát đối với rừng ngẫu nhiên (unsupervised learning of forests). Ví dụ như lựa chọn các hàm chia để có thể chia dữ liệu đồng đều nhất. Ngoài ra, cũng có thể sử dụng thêm các chỉ số của các nút khác nút lá để có được sự phân cấp về mức độ diễn tả chi tiết cho dữ liệu đặc trưng đầu vào.

Nội dung trình bày trong chương này được thể hiện hoàn toàn trong công trình nghiên cứu số 7 của tác giả.

CHƯƠNG 4

XÂY DỰNG MỘT SỐ ỨNG DỤNG NHẬN DẠNG HOẠT ĐỘNG

Chương này trình bày hai ứng dụng thử nghiệm cho nhận dạng hoạt động người trong thời gian thực dựa trên các đặc trưng đã đề xuất trong chương 2. Ứng dụng thứ nhất là hệ thống phát hiện người ngã trong thời gian thực sử dụng cảm biến gia tốc tích hợp trong thiết bị Wii Remote. Hệ thống có giá rẻ, dễ triển khai và có tính khả thi cao. Ứng dụng thứ hai là hệ thống xác thực người dùng dựa trên chữ ký 3D, sử dụng cảm biến gia tốc có sẵn trong các thiết bị di động cầm tay thông minh. Phương pháp xác thực mới kết hợp ưu điểm của chữ ký và cảm biến gia tốc trên thiết bị di động, giúp cho người dùng có thể đăng nhập hệ thống hay ứng dụng một cách thuận tiện hơn so với các cách nhập liệu cũ như sử dụng màn hình hay bàn phím.

4.1. PHÁT HIỆN NGƯỜI NGÃ TRONG THỜI GIAN THỰC

4.1.1. Các giải pháp phát hiện người ngã

Trong những năm gần đây, nhiều giải pháp được đưa ra để giải quyết vấn đề phát hiện người ngã kịp thời [22,39,42,50,128]. Có hai giải pháp thương mại đáng được quan tâm. Giải pháp thứ nhất là SmartCane [145], gồm một cây gậy đỡ bình thường với bộ cảm biến không dây. Giao tiếp dữ liệu cảm biến được chuyển tiếp đến một thiết bị cá nhân để xử lý. Tuy nhiên, SmartCane chỉ phát hiện được người ngã khi đem theo gậy bên mình. Giải pháp thứ hai cũng là một hệ thống phát hiện ngã thương mại khá nổi tiếng, gọi là Brickhouse [153]. Hệ thống có một cảm biến kèm theo người và một hệ thống trợ giúp từ xa. Tuy vậy chi phí cho các hệ thống này khá cao: phần thiết bị của Brickhouse là 199,95\$ và phần dịch vụ theo dõi là 419,40\$/năm [153].

Các kỹ thuật phát hiện ngã tự động có thể chia làm hai loại chính. Loại thứ nhất gồm các giải pháp sử dụng trang bị trong môi trường. Chẳng hạn như máy camera thường hoặc hồng ngoại để theo dõi hình ảnh các hoạt động người [59,76,87], cảm biến âm thanh và hình ảnh [76], hay cảm biến rung chân của nền nhà [39,42]. Loại thứ hai là dựa trên cảm biến gia tốc mang trên người [23,39]. Các giải pháp loại thứ nhất có tỉ lệ phát hiện cao, tuy nhiên cần có máy quay hoặc phải thiết lập môi trường hoạt động trước đó và được sử dụng để phát hiện được người ngã trong một không gian nhỏ hẹp như trong nhà hay trong bếp. Giải pháp này thường có chi phí cao và thường bị ảnh hưởng bởi ánh sáng hoặc nhiễu âm thanh từ môi trường. Trong khi đó, giải pháp dùng cảm biến mang theo người hầu hết không cần cài đặt hoặc cấu hình môi trường trước, rất dễ dàng trong việc triển khai và sử dụng. Thêm vào đó, phạm vi hoạt động của người được theo dõi không bị giới hạn do hệ thống được gắn theo người và có thể hoạt động ở bất kỳ nơi đâu. Giải pháp này mang lại sự thuận tiện cho người dùng, đáp ứng nhiều tình huống để phát hiện ngã, do đó có thể áp dụng tốt cho các hoạt động một mình, trong thời gian dài kể cả trường hợp khi có nguy cơ xảy ra ngã cao. Giải pháp thuộc loại thứ hai được sử dụng và triển khai nhiều hơn bởi vì hiện nay có rất nhiều thiết bị, đặc biệt là thiết bị di động có trang bị cảm biến gia tốc.

Phần này trình bày một hệ thống tự động phát hiện ngã sử dụng bộ cảm biến gia tốc, dựa trên các đặc trưng đề xuất trong chương 2. Cảm biến gia tốc này được tích hợp sẵn trong Wii Remote [157], một thiết bị thường được sử dụng trong các trò chơi tương tác và có sẵn trên thị trường với giá thành tương đối rẻ (khoảng 25\$). Hơn nữa, thời lượng pin sử dụng rất khả quan với khoảng 30 tiếng hoạt động, hơn hẳn so với nhiều điện thoại di động thông minh có kích hoạt cảm biến gia tốc hoạt động liên tục. Hệ thống cho kết quả rất khả quan với tỉ lệ phát hiện cao và chi phí thấp hơn so với các hệ thống phát hiện ngã khác.

Với mục tiêu là thực hiện một thử nghiệm đơn giản, giá rẻ, giúp phát hiện ngã trong thời gian thực, hệ thống này sử dụng một thiết bị mới cùng cảm biến mới và có vị trí đặt khác với dữ liệu huấn luyện trong chương 2. Do đó việc lấy dữ liệu

huấn luyện mới cho mô hình phát hiện ngã thời gian thực là cần thiết để đảm bảo sự chính xác của hệ thống. Các phần tiếp theo sẽ mô tả về phần cứng và môi trường thu thập dữ liệu, thuật toán sử dụng, dữ liệu thử nghiệm cho hệ thống nhận dạng ngã thời gian thực và đánh giá kết quả. Ngoài ra, ở đây cũng mô tả ngắn gọn về phần mềm phát hiện ngã sử dụng cho người dùng.

4.1.2. Phần cứng và môi trường thu thập dữ liệu

Trái ngược với hầu hết các cảm biến gia tốc thường được sử dụng trong các nghiên cứu phòng thí nghiệm hoặc trên thị trường nhưng tương đối đắt tiền, hoặc triển khai phức tạp (tức là đòi hỏi một trạm cơ sở để giao tiếp với máy tính), Wii Remote tương đối rẻ, có sẵn trên thị trường, triển khai dễ dàng và đơn giản do có thể giao tiếp với máy tính thông qua một Bluetooth dongle. Cả hai thiết bị Wii Remote và Bluetooth dongle đều rẻ tiền và có sẵn trong các cửa hàng kinh doanh trò chơi.

Wii Remote [157] là hệ thống cảm biến không dây cho người tiêu dùng và bộ điều khiển trò chơi, hỗ trợ hai chức năng: (i) phát hiện nhập liệu từ người dùng thông qua một cảm biến gia tốc nhúng bên trong; và (ii) trao đổi dữ liệu thông qua Bluetooth. Một thiết bị Wii Remote bao gồm một bảng mạch in (được đóng gói trong một vỏ trắng như trong hình 4.1), trong đó chứa một cảm biến gia tốc AXDL 330 và một chip BCM2042 Broadcom tích hợp toàn bộ cấu hình, ứng dụng, và ngăn xếp giao thức Bluetooth. Dựa trên công nghệ MEMS, cảm biến gia tốc AXDL 330 có kích cỡ rất nhỏ, công suất thấp, 3 trục gia tốc với đầu ra điện áp được điều hòa về mặt tín hiệu. Các cảm biến gia tốc AXDL 330 có thể cảm nhận được gia tốc trên ba trục với khoảng giá trị toàn thang đo tối thiểu là $\pm 3g$. Trong khi gia tốc tĩnh của trọng trường có thể được sử dụng để cảm nhận độ nghiêng trong các ứng dụng, gia tốc động có thể đo được thông qua định lượng chuyển động, sóc và rung.

Bo mạch Broadcom BCM2042 là một hệ thống SoC được tích hợp một bộ vi xử lý 8051 trên mạch, bộ nhớ truy cập ngẫu nhiên/bộ nhớ chỉ đọc, cấu hình thiết bị tiếp con người (HID), ứng dụng, và ngăn xếp giao thức Bluetooth. Hơn nữa, bo

mạch này cũng tích hợp nhiều thiết bị ngoại vi và một cổng mở rộng. Bộ phận nhập liệu của Wii Remote bao gồm các nút, một bộ cảm biến hồng ngoại và một cảm biến gia tốc. Các bộ cảm biến hồng ngoại được nhúng vào trong một máy ảnh trong đó phát hiện ánh sáng hồng ngoại đến từ một thanh cảm biến bên ngoài. Gia tốc được đo theo các trục x , y , z (liên quan đến gia tốc) và ba hướng chuyển động (x , y , z) có thể được tính được thông qua góc nghiêng. Dữ liệu đầu vào qua các thiết bị nhập liệu của Wii Remote, trong đó có cảm biến gia tốc được truyền đến một máy chủ Bluetooth thông qua các giao thức Bluetooth HID chuẩn. Các giá trị gia tốc được truyền với tần số lấy mẫu 100 Hz.



Hình 4.1. Wii Remote (bên trái nhất), bo mạch Broadcom BCM2042 và vị trí đeo Wii Remote

Trong hệ thống này, các đối tượng đeo Wii Remote trên hông do cảm biến đeo bên hông tốt cho việc phát hiện các hoạt động ngã, chạy, đi bộ và leo cầu thang.

4.1.3. Thuật toán phát hiện ngã

Hệ thống sử dụng đặc trưng đề xuất trong chương 2 và phương pháp phát hiện ngã đã đề xuất trong phần 2.4.1. Các dữ liệu lấy được từ cảm biến gia tốc trên Wii Remote được truyền qua Bluetooth tới phần mềm xử lý trung tâm trên máy tính. Tại đây, luồng tín hiệu đầu vào sau khi tiền xử lý sẽ được phân đoạn thành các khung dữ liệu, thực hiện trích xuất đặc trưng và đưa ra dự đoán. Hệ thống được huấn luyện sử dụng tập dữ liệu ngã mô tả ở phần 4.1.4, với các tham số là số ngăn M và số đoạn trong một khung K được tối ưu hóa trên tập dữ liệu đầu vào, sử dụng

tập tối ưu dành riêng (hold-out validation set). Riêng đối với bước tiền xử lý dữ liệu, hệ thống cần áp dụng các phương pháp cụ thể để xử lý tín hiệu bị nhiễu cho hiệu quả. Đồng thời, lựa chọn độ dài cửa sổ trượt cho bước phân đoạn cũng được tiến hành để tìm ra độ dài phù hợp, như được mô tả trong phần dưới đây.

Tiền xử lý dữ liệu. Trong các hệ thống nhận dạng hoạt động sử dụng cảm biến mang trên người, dữ liệu lấy được từ các cảm biến thường bị nhiễu và không rõ ràng. Lý tưởng nhất, ở tần số lấy mẫu 100 Hz, mỗi giây cảm biến cho ra 100 mẫu của gia tốc với 3 giá trị trên 3 trục x , y , z . Trong thực tế, có nhiều yếu tố có thể gây ra sự mất mát các mẫu giá trị như sự ảnh hưởng do các vật dụng kim loại đặt giữa các cảm biến và máy thu, hay cảm biến vô ý bị nhúng vào nước, v.v. hoặc cũng có thể là do các tác động bên ngoài vào chuyển động làm cho chuyển động của con người trở nên không bình thường (giả sử là chuyển động bị tay chạm vào). Ngoài ra, các cảm biến có thể tự sinh ra nhiễu tùy vào chất lượng của nó (ví dụ như sinh ra các giá trị quá lớn hay nhỏ). Trong trường hợp như vậy, người ta thường sử dụng một bộ lọc để loại bỏ nhiễu, sau đó sinh ra giá trị phù hợp bù lại cho mẫu bị mất. Ở đây, các bộ lọc dữ liệu bao gồm bộ lọc thông thấp để loại bỏ các mẫu có giá trị thấp bất thường và bộ lọc thông cao để lọc ra các mẫu có giá trị cao bất thường. Sau đó, các mẫu được nhóm vào các khung hay cửa sổ thời gian. Nếu một khung chứa ít hơn 75% số mẫu so với thông thường, nó sẽ bị loại bỏ bởi vì không đủ thông tin để phân lớp các hoạt động. Ngược lại, khung được lấy mẫu lại bằng cách sử dụng phương pháp nội suy cubic spline [24] để bù vào mẫu bị mất.

Phân đoạn dữ liệu. Nhiều nghiên cứu trước đó đã cho thấy độ dài của cửa sổ trượt có ảnh hưởng quan trọng tới hiệu năng của thuật toán nhận dạng hoạt động [17,110]. Để chọn được độ dài cửa sổ phù hợp cần thực hiện một thực nghiệm dựa trên một tập con trên tập dữ liệu hiện có. Các độ dài cửa sổ được thay đổi với nhiều giá trị khác nhau, gồm 1 giây, 1,2 giây, 1,5 giây, 1,8 giây, 2 giây, 2,5 giây, 3 giây và 6 giây, và độ dài phù hợp nhất được chọn là 1,8 giây. Lý do để chọn độ dài cửa sổ như vậy là vì độ dài 1,8 giây cho phép tránh trễ khi xử lý liên tục trong thời gian thực, đồng thời có được độ chính xác cao.

4.1.4. Tập dữ liệu thử nghiệm

Tập dữ liệu huấn luyện cho ứng dụng là dữ liệu được thu thập từ 12 đối tượng là sinh viên Học viện Công nghệ Bưu chính Viễn thông. Tất cả các đối tượng tham gia đeo Wii ở cùng một vị trí (vùng thắt lưng) và cùng hướng. Dữ liệu cảm biến được truyền đến và ghi lại (logging) trên một máy tính. Các đối tượng được yêu cầu thực hiện 12 hoạt động hàng ngày bao gồm đi bộ, nhảy, đi lên cầu thang, đi xuống cầu thang, chạy, vươn người, lau nhà, đánh máy, ngồi xuống lúc đang đứng, đứng lên lúc đang ngồi, đánh răng, hút bụi và 12 tư thế ngã khác nhau (cố ý).

Các hoạt động không yêu cầu phải thực hiện theo thứ tự, đồng thời không hạn chế thời gian cho từng hoạt động thực hiện đối với mỗi chủ thể. Mỗi hoạt động được yêu cầu phải được thực hiện một cách tự nhiên nhất có thể. Ngoài việc ghi lại dữ liệu cảm biến, video của các đối tượng thực hiện các hoạt động cũng được ghi lại.

Để đồng bộ hóa các đoạn video thu thập với dữ liệu gia tốc từ thiết bị Wii Remote, vào thời điểm bắt đầu của mỗi phiên hoạt động, đối tượng được yêu cầu lắc cơ thể và tay 3 lần để tách biệt các đoạn tín hiệu. Các đối tượng thử nghiệm được hiển thị trên video. Ngoài ra, mốc thời gian của các mẫu dữ liệu cũng ghi lại vào tập tin nhật ký dữ liệu gia tốc.

Các đối tượng đã được đưa cho một danh sách gồm 12 nhãn các hoạt động để chú thích các đoạn video thu được bằng cách sử dụng công cụ chú thích video ELAN Multimedia Annotator. Các chuyển động không phải là một trong 12 hoạt động hay ngã được gán nhãn là “unknown” (không rõ).

Như vậy, số hoạt động ngã trong tập dữ liệu là 144 và các hoạt động còn lại thuộc về 12 hoạt động hàng ngày khác (không phải ngã). Vì nhiều hoạt động ADL khác tạo ra tín hiệu tương tự với ngã, ví dụ như nhảy, đang ngồi rồi đứng, và rất nhiều nhiều, bộ dữ liệu này là một thách thức đối với một hệ thống phát hiện ngã cần độ chính xác cao.

4.1.5. Kết quả thử nghiệm

Thử nghiệm hệ thống phát hiện ngã trên tập dữ liệu đã mô tả trong phần 4.1.4. cho kết quả khá cao, với độ chính xác là $96,25\% \pm 0,38\%$, độ nhạy là $97,67\% \pm 0,23\%$ và độ đặc hiệu là $84,34\% \pm 2,32\%$. Kết quả thử nghiệm như vậy tuy không phải là tốt nhất so với kết quả đánh giá phương pháp phát hiện ngã trong chương 2 nhưng cũng không ngoài dự đoán, do tập dữ liệu đánh giá tương đối nhiều nhiễu như đã chỉ ra trong phần 4.1.4.

4.1.6. Phần mềm

Phần mềm xử lý trên máy tính được xây dựng để thu nhận tín hiệu cảm biến gia tốc trên Wii Remote qua Wii dongle (giao diện USB), truyền qua Bluetooth. Trình điều khiển thiết bị được cấu hình qua phần mềm Bluesolein. Chương trình phần mềm được lập trình bằng ngôn ngữ Java và sử dụng thư viện giao tiếp với Wii Remote có sẵn là wiiusej.jar từ dự án <https://code.google.com/archive/p/wiiusej/>. Giao diện chính của chương trình như trong hình dưới đây.



(a)



(b)

Hình 4.2. (a) Giao diện chính của phần mềm phát hiện ngã tự động và (b) thông báo khi phát hiện sự kiện ngã

4.2. XÁC THỰC NGƯỜI DÙNG VỚI CHỮ KÝ 3D – SIGVER3D

Phần này trình bày một hệ thống xác thực người dùng theo thời gian thực thuận tiện cho các thiết bị di động có sẵn thiết bị cảm biến gia tốc, gọi là SigVer3D. Hệ thống này hoạt động bằng cách phân tích các luồng tín hiệu gia tốc trả về từ cảm biến trên thiết bị di động khi người dùng sử dụng thiết bị để ký chữ ký của họ lên không gian. Chữ ký này được gọi là chữ ký 3D. Bài toán xác thực ở đây được chuyển thành bài toán phân lớp nhị phân và thực hiện huấn luyện các bộ phân lớp

SVM để xác định các lần đăng nhập thành công. Với việc sử dụng phương pháp trích xuất đặc trưng hiệu quả đã đề xuất trong chương 2, kết quả thử nghiệm cho thấy hệ thống tính toán nhanh trên các thiết bị có khả năng tính toán hạn chế và hiệu quả khi đánh giá trên bộ dữ liệu cử chỉ được thu thập nghiêm ngặt từ một nhóm các đối tượng. Cùng với sự phổ biến của các bộ gia tốc trong các thiết bị di động, hệ thống này cung cấp một sự bổ sung tiềm năng vào các hệ thống xác thực trên các thiết bị di động hiện có.

4.2.1. Vấn đề xác thực cho các hệ thống mang trên người

Ngày nay, thiết bị di động ngày càng phát triển và hỗ trợ con người trong mọi mặt cuộc sống. Hàng loạt các thiết bị di động mang được trên người như điện thoại thông minh, Google Glass, Nike+ FuelBand, hay Pepple Smartwatch. Các thiết bị này được mang theo người mọi lúc mọi nơi và thông thường mang theo rất nhiều dữ liệu cá nhân quan trọng. Ngoài ra, chúng quá nhỏ khiến cho việc nhập liệu rất khó khăn. Điều này gây ra hạn chế tới việc xác thực. Ví dụ như các thiết bị như cảm biến di động cực nhỏ thậm chí còn không có màn hình cảm ứng để nhập liệu. Vấn đề này thúc đẩy các nhà nghiên cứu tìm ra phương thức mới thuận tiện, linh hoạt mà vẫn đảm bảo an toàn cho các thiết bị có những hạn chế như vậy.

Một trong các phương thức được đề xuất là sử dụng các hệ thống xác thực dựa trên sinh trắc học, thông thường sẽ dựa trên các đặc tính sinh lý học hay hành vi của người dùng. Ví dụ như mặt, móng mắt [46], vân tay [93], bàn tay [120], cách thức gõ bàn phím [35], chữ ký tay [113], giọng nói [97], cách bước đi [43], v.v. Nhiều hệ thống xác thực hoạt động tốt, ví dụ như các hệ thống nhận dạng mặt hay vân tay, nhưng vẫn có thể bị vượt qua [93]. Một cách khác là sử dụng chữ ký. Đây là một phương thức xác thực dựa trên đặc tính hành vi của người dùng, rất phổ biến và truyền thống cho việc xác thực người dùng trong ngân hàng, các giao dịch cá nhân. Có hai loại xác thực dựa trên chữ ký, bao gồm xác thực tĩnh dựa trên ảnh và xác thực động dựa trên các đặc tính động như hướng, nét, lực nhấn, tốc độ ký. Xác thực động thường được sử dụng với các thiết bị có màn hình cảm ứng. Tuy nhiên

chữ ký tĩnh cũng dễ giả mạo do người kiểm tra bằng mắt thường rất khó để ý sự khác biệt giữa các chữ ký chính chủ và chữ ký của kẻ giả mạo chuyên nghiệp. Chữ ký động lại bị phụ thuộc nhiều vào các thiết bị nhập như màn hình cảm ứng.

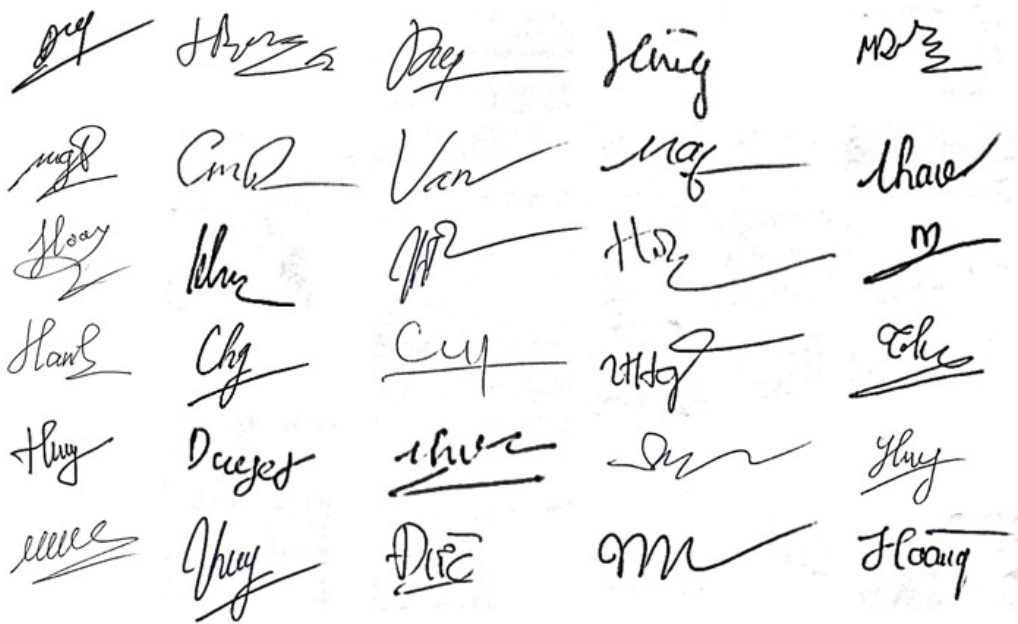
Tác giả đề xuất một phương thức xác thực mới cho người dùng thiết bị di động với chữ ký 3D, được gọi là SigVer3D. Phương thức này có cả ưu điểm về tính an toàn của chữ ký và thuận tiện cho người dùng khi muốn “đăng nhập” vào thiết bị di động khi không có các thiết bị nhập liệu truyền thống như màn hình cảm ứng, hay bàn phím. Phương pháp đề xuất sử dụng các dữ liệu cảm biến gia tốc như sau. Người dùng cầm thiết bị di động và vẽ chữ ký của họ trong không gian để đăng nhập thông qua việc di chuyển thiết bị. Khi người dùng thực hiện ký, các dữ liệu gia tốc được sinh ra và được hệ thống sử dụng để kiểm tra người dùng. Phương pháp này cho phép khắc phục hạn chế của những thiết bị di động không tiện trong việc đăng nhập (như màn hình nhỏ, nhập liệu khó khăn, hoặc thậm chí không có màn hình, bàn phím). Một lý do để chọn chữ ký 3D làm mật khẩu là nó dễ nhớ và dễ thực hiện, đồng thời khó giả mạo bởi người khác do sử dụng các đặc trưng sinh trắc học và vật lý bổ sung để làm tăng tính an toàn cho việc xác thực chữ ký.

4.2.2. Phần cứng và môi trường thu thập dữ liệu

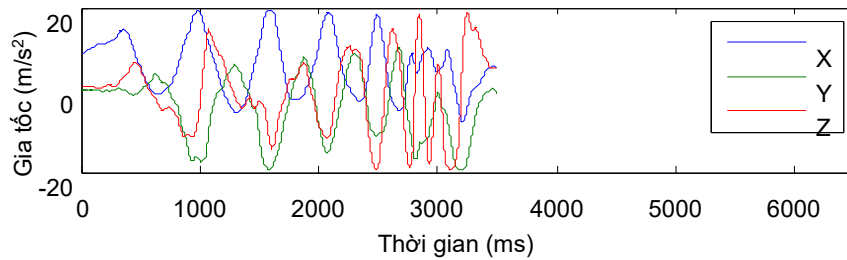
Hệ thống này được cài đặt trên một điện thoại di động thông minh Samsung S3 (xem hình 4.3) có trang bị một cảm biến gia tốc gắn trong dùng để cảm nhận gia tốc theo ba trục. Cảm biến gia tốc tích hợp bên trong Samsung S3 là một phần của mô-đun LSM330DLC, có thể đo các giá trị gia tốc với khoảng đo đầy đủ là $\pm 2g$ [156] và tốc độ lấy mẫu trong khoảng 20-120 Hz. Mỗi mẫu tín hiệu là một bộ ba của 3 giá trị theo 3 trục x, y, z . Trong hệ thống này, tốc độ lấy mẫu của bộ cảm biến gia tốc được thiết lập tại mức 100 Hz.



Hình 4.3. Samsung S3 (trái) và cách người dùng cầm điện thoại để ký trong không gian (phải)

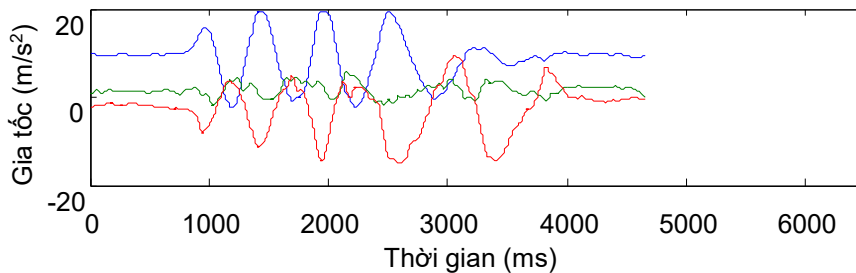


Hình 4.4. Các chữ ký trong tập dữ liệu



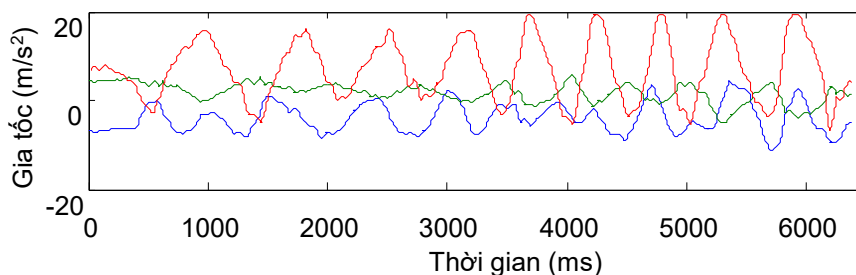
Chy

(a) Một mẫu chữ ký gốc đơn giản: hình và dữ liệu gia tốc tương ứng



ngp

(b) Một mẫu chữ ký gốc phức tạp: hình và dữ liệu gia tốc tương ứng



Hoay

(c) Một mẫu chữ ký gốc phức tạp khác: hình và dữ liệu gia tốc tương ứng

Hình 4.5. Các chữ ký gốc (hình phải) và biểu diễn gia tốc tương ứng (hình trái)

4.2.3. Thuật toán xác thực người dùng

Bài toán xác thực người dùng ở đây thực chất là bài toán phân lớp nhị phân: phát hiện chữ ký giả mạo và chữ ký chính chủ. Luồng dữ liệu lấy được từ cảm biến gia tốc trên điện thoại được tiền xử lý để loại bỏ nhiễu, thực hiện trích xuất đặc trưng và đưa ra dự đoán. Chi tiết các bước được mô tả như sau.

a) Tiền xử lý dữ liệu

Trong nghiên cứu này, tốc độ lấy mẫu của bộ cảm biến gia tốc được thiết lập tại mức 100 Hz. Tuy nhiên, dữ liệu cảm biến thường bị nhiễu và không thể sử dụng trực tiếp được. Vì vậy, một số kỹ thuật tiền xử lý tín hiệu cảm biến (như mô tả trong phần 4.1.3 trên đây) cũng được sử dụng để loại bỏ nhiễu và nội suy các mẫu giá trị bị thiếu hoặc lỗi. Ngoài ra, đoạn dữ liệu đầu tiên và cuối cùng của các mẫu tín hiệu chữ ký cũng được cắt bỏ do các đoạn tín hiệu này thường chứa nhiễu.

b) Phân đoạn dữ liệu và trích xuất đặc trưng

Các đặc trưng được sử dụng là các đặc trưng đã đề xuất trong chương 2 với các tham số là số ngăn M và số đoạn trong một khung K được tối ưu hóa trên tập dữ liệu đầu vào, sử dụng tập tối ưu dành riêng (hold-out validation set). Các đặc trưng trích xuất sẽ được sử dụng làm dữ liệu đầu vào cho bộ phân lớp.

c) Phân lớp

Do sự hiệu quả của máy véc-tơ tựa SVM trong các nghiên cứu nhận dạng hoạt động người, nghiên cứu này cũng sử dụng SVM với hàm nhân RBF để làm bộ phân lớp. Các tham số của bộ phân lớp được điều chỉnh tự động bằng hàm tìm kiếm lưới.

d) Các cấu hình thử nghiệm khác

Đánh giá một hệ thống kiểm tra đòi hỏi sự phân tích của hai loại lỗi: tỷ lệ từ chối sai (FRR - false rejection rate) cho biết tỷ lệ của chữ ký gốc bị từ chối nhầm bởi hệ thống, và tỷ lệ chấp nhận sai (FAR - false accepted rate) cho biết tỷ lệ chữ ký giả mạo được chấp nhận bởi hệ thống. FAR và FRR được định nghĩa như sau:

$$FRR = \frac{FN}{FN + TP} \quad (4.1)$$

$$FAR = \frac{FP}{FP + TN} \quad (4.2)$$

Trong đó, FP là dương tính giả (chữ ký giả mạo được chấp nhận), TP là dương tính thực (chữ ký gốc được phát hiện), FN là âm tính giả (chữ ký gốc bị từ chối không đúng), và TN là âm tính thực (chữ ký giả mạo bị phát hiện).

Hai độ đo FAR và FRR được sử dụng để đánh giá hiệu năng của phương pháp cho rất nhiều hệ thống xác thực. Ngoài ra, một biện pháp đo hiệu quả khác là sử dụng đường cân bằng lỗi (hay gọi là ROC - receiver operating characteristics curve) để nhận thấy sự thay đổi của một lỗi đối với các lỗi khác như thế nào. Đường ROC thể hiện một đặc trưng trực quan của sự cân bằng giữa FAR và FRR. Một đường ROC thường được tóm tắt bởi khu vực nằm dưới ROC (gọi là AUC - Area Under ROC). Một số đo khác là tỉ lệ lỗi tương đương (EER - equal error rate) cũng thường được sử dụng để đo lường hiệu quả hoạt động của các hệ thống kiểm tra, là tỷ lệ mà tại đó FAR bằng FRR. Ngược lại với đường ROC, giá trị EER càng thấp thì hệ thống kiểm tra càng chính xác và đáng tin cậy hơn [89].

4.2.4. Thu thập dữ liệu thử nghiệm

Có 30 đối tượng tham gia trong thử nghiệm này thực hiện việc ký chữ ký 3D của họ. Một ứng dụng ghi nhật ký dữ liệu được phát triển dùng cho nghiên cứu này, chạy trên điện thoại thông minh Samsung Galaxy S3 và thực hiện ghi lại dữ liệu gia tốc theo thời gian thực. Người dùng bắt đầu thực hiện ký chữ ký 3D của họ bằng cách chạm và giữ ngón tay lên trên màn hình điện thoại. Các giá trị gia tốc lấy từ cảm biến gia tốc được ghi lại vào một tập tin nhật ký kể từ khi ngón tay bắt đầu chạm vào màn hình cho đến khi ngón tay rời khỏi màn hình. Cách người dùng cầm điện thoại và ký vào không gian được mô tả như hình 4.3 (bên phải). Quá trình này được thực hiện giống nhau đối với tất cả các đối tượng, kể cả những người đóng vai làm kẻ giả mạo, cố gắng bắt chước chữ ký của người dùng chính chủ. Hình 4.4 cho thấy chữ ký trên giấy của các đối tượng tham gia thử nghiệm này, và hình 4.5 mô tả một số mẫu chữ ký và các tín hiệu gia tốc đầu ra tương ứng.

Mỗi đối tượng được yêu cầu ký chữ ký của chính họ trên một tờ giấy và thực hiện các thao tác ký chữ ký 3D của mình 10 lần theo cách đã mô tả ở trên. Sau đó,

đối với từng đối tượng chính chủ, 10 đối tượng khác được lựa chọn ngẫu nhiên từ các đối tượng còn lại để đóng vai trò là kẻ giả mạo. Mỗi kẻ giả mạo được cho trước chữ ký trên giấy của đối tượng đã thực hiện chữ ký 3D của họ và thử ký bắt chước 5 lần trong không gian. Khi đó, với mỗi đối tượng, chúng ta có 10 chữ ký 3D gốc và 50 chữ ký 3D giả mạo. Do vậy, dữ liệu thu được bao gồm 1800 mẫu chữ ký 3D trong đó 300 là gốc và 1500 mẫu là chữ ký 3D giả mạo.

4.2.5. Kết quả thử nghiệm

Kết quả thử nghiệm cho thấy hệ thống xác thực có độ chính xác tương đối cao và khá tin cậy. Giá trị AUC trung bình đạt được với giá trị là 98,3%, và giá trị EER thấp, tại đó $FRR = FAR = 1,4\%$. Điều này có nghĩa là người dùng gốc thực hiện đăng nhập 1000 lần thì có 14 lần không thành công, trong khi đó người giả mạo thực hiện ký 1000 lần thì thành công 14 lần. Kết quả này đã cho thấy rằng chữ ký động với việc sử dụng các đặc trưng sinh trắc học và vật lý bổ sung đã làm tăng tính an toàn cho việc xác thực. Hai người có đặc tính sinh lý khác nhau, kể cả bắt chước động tác ký giống nhau cũng khó có thể thành công.

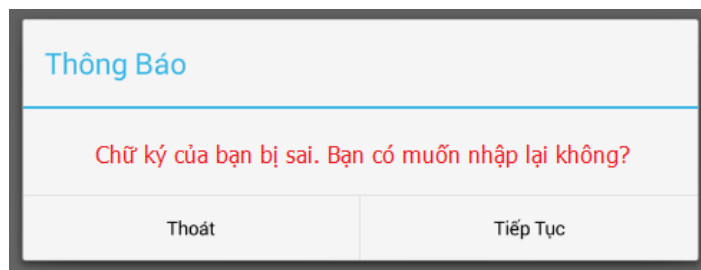
4.2.6. Phần mềm

Phần mềm xác thực chữ ký 3D được xây dựng trên điện thoại di động có hệ điều hành Android phiên bản 4.2, có trang bị cảm biến gia tốc. Cảm biến gia tốc sẽ được tự động kích hoạt và tần số lấy mẫu của cảm biến gia tốc được thiết lập tại mức 100 Hz trong chương trình. Chương trình phần mềm bao gồm một giao diện mô phỏng một giao diện đăng nhập cho người dùng. Người dùng có thể xác thực thông qua cách thức nhập tài khoản người dùng bằng ký tự hoặc sử dụng chữ ký 3D. Trong trường hợp dùng chữ ký 3D để đăng nhập, người dùng lấy mẫu chữ ký 3D 10 lần, hệ thống sẽ tự động phát hiện xem chữ ký đúng hay sai. Trong chương trình, có một chức năng “Thiết lập chữ ký 3D” để lấy mẫu chữ ký 3D của người dùng nhằm huấn luyện cho hệ thống. Chức năng này cần được thực hiện trước khi người dùng có thể thực hiện đăng nhập bằng chữ ký 3D của mình với điện thoại di động. Cách thức lấy chữ ký được trình bày như trong mô tả lấy dữ liệu ở mục 4.2.4

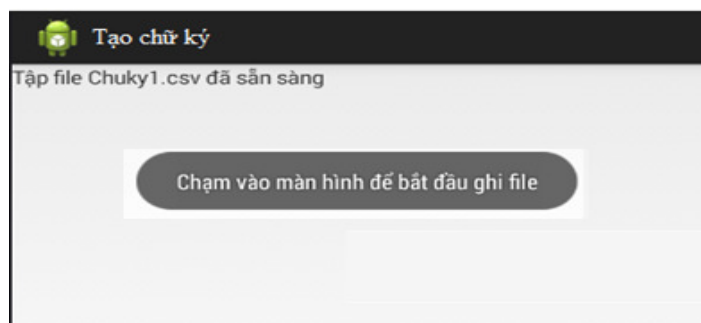
và cần phải lấy 10 mẫu. Hình 4.6 dưới đây trình bày một số hình ảnh minh họa phần mềm.



(a)



(b)



(c)



(d)

Hình 4.6. Các hình ảnh minh họa cho phần mềm “Xác thực bằng chữ ký 3D”

4.3. KẾT LUẬN CHƯƠNG

Chương này đã trình bày hai ứng dụng nhận dạng hoạt động người trong thời gian thực hữu ích dựa trên các đặc trưng đã đề xuất trong chương 2. Thứ nhất là hệ thống phát hiện người ngã trong thời gian thực dễ triển khai, và có tính ứng dụng cao do sử dụng các phần cứng phổ biến có sẵn với giá thành rẻ. Thứ hai là hệ thống xác thực người dùng dựa trên chữ ký 3D. Đây là một phương pháp xác thực mới kết hợp ưu điểm của chữ ký và cảm biến gia tốc trên thiết bị di động, giúp cho người dùng có thể đăng nhập hệ thống hay ứng dụng một cách thuận tiện, không bị nhiều hạn chế do những giới hạn của cách thức nhập liệu cũ phụ thuộc vào màn hình hay bàn phím. Phương pháp xác thực này cũng là một phương án bổ sung tốt cho các phương pháp xác thực hiện tại trên các thiết bị di động hoặc thậm chí cho các ứng dụng giao dịch thực tế bằng chữ ký viết tay.

Hai ứng dụng đã trình bày ở trên cũng chứng minh rằng phương pháp nhận dạng hoạt động sử dụng đặc trưng HALF đề xuất trong chương 2 không những có thể *đảm bảo nhận dạng hoạt động chính xác mà còn có khả năng ứng dụng tốt trong thực tế do có khả năng xây dựng các hệ thống yêu cầu thời gian thực với tài nguyên hạn chế.*

Nội dung trình bày trong chương này được tổng hợp dựa trên kết quả các công trình nghiên cứu số 4, số 5 và số 6 của tác giả. Công trình nghiên cứu số 4 mô tả ngắn gọn ứng dụng phát hiện ngã sử dụng đặc trưng HALF nhưng với phương pháp phân cụm thủ công, tuy nhiên ứng dụng trong chương này sử dụng đặc trưng với phương pháp phân cụm tự động mô tả trong công trình nghiên cứu số 6. Công trình nghiên cứu số 5 mô tả cụ thể ứng dụng chữ ký 3D sử dụng đặc trưng HALF cũng với phương pháp phân cụm thủ công, được thay thế bằng phương pháp phân cụm tự động trong công trình nghiên cứu số 6 có ưu điểm hơn.

KẾT LUẬN

Nhận dạng hoạt động người là một chủ đề nghiên cứu quan trọng trong nhiều lĩnh vực như tính toán nhận biết ngữ cảnh, tính toán khắp nơi, tương tác người-máy, tính toán di động. Luận án này đã nghiên cứu một hướng tiếp cận về chủ đề này, đó là nhận dạng hoạt động sử dụng cảm biến mang trên người.

Dù có chung nhiều vấn đề về mặt phương pháp luận với các lĩnh vực nghiên cứu khác như thị giác máy, xử lý ngôn ngữ tự nhiên, hay nhận dạng tiếng nói, nhận dạng hoạt động dựa trên cảm biến mang theo người vẫn có nhiều khó khăn riêng, đòi hỏi các phương pháp tính toán chuyên biệt kể cả khi tận dụng các phương pháp hiện đang sử dụng trong các lĩnh vực khác. Nguyên nhân là vì bên cạnh những vấn đề chung của bài toán nhận dạng mẫu, xuất hiện một số vấn đề mang tính đặc thù đối với nhận dạng hoạt động sử dụng cảm biến mang trên người. Dựa vào các nghiên cứu cơ bản về phương pháp nhận dạng hoạt động sử dụng cảm biến mang trên người, luận án tập trung giải quyết hai vấn đề quan trọng tồn tại trong lĩnh vực này, đó là *đảm bảo nhận dạng hoạt động chính xác cho các hệ thống đòi hỏi thời gian thực bị hạn chế về tài nguyên và xây dựng phương pháp nhận dạng hiệu quả với nhiều loại hoạt động người*.

Đối với vấn đề *đảm bảo nhận dạng hoạt động chính xác cho các hệ thống đòi hỏi thời gian thực bị hạn chế về tài nguyên*, luận án đề xuất phương pháp trích xuất đặc trưng mới có *tốc độ nhanh* cho các ứng dụng cần nhận dạng một số *hoạt động riêng lẻ*, sử dụng cảm biến mang theo người với yêu cầu xử lý *thời gian thực*. Phương pháp trích xuất đặc trưng mới, được gọi là HALF, có thể phân biệt tốt các hoạt động có đặc tính dữ liệu tương tự nhau, phù hợp cho nhiều loại dữ liệu trong các hệ thống nhận dạng hoạt động sử dụng cảm biến đeo trên người, thích hợp cho các hệ thống di động thông minh, nhỏ gọn với năng lực tính toán thấp, đặc biệt là các hệ thống đòi hỏi thời gian thực. Việc thử nghiệm và đánh giá kết quả của phương pháp được thực hiện trên các tập dữ liệu đã công bố rộng rãi. Kết quả cho

thấy, phương pháp đề xuất cho kết quả tốt hơn so với một số phương pháp mới được đề xuất gần đây.

Đối với vấn đề *xây dựng phương pháp nhận dạng hiệu quả cho nhiều loại hoạt động người*, luận án đề xuất phương pháp trích xuất đặc trưng tự động dựa trên học đặc trưng cho các bài toán nhận dạng hoạt động người. Các đặc trưng đề xuất là các đặc trưng đa mức MPF có khả năng cải thiện về độ chính xác trong nhận dạng hoạt động người so với các đặc trưng đa mức kiểu cũ, đồng thời khắc phục được các hạn chế về mặt tốc độ khi xử lý. Việc thử nghiệm và đánh giá kết quả được tiến hành trên 3 tập dữ liệu đã công bố rộng rãi. Kết quả thử nghiệm cho thấy phương pháp đề xuất có kết quả tốt hơn đáng kể so với một số phương pháp nhận dạng hoạt động người tiên tiến nhất hiện nay.

Cuối cùng, luận án cũng mô tả hai ứng dụng nhận dạng hoạt động người khả thi được xây dựng dựa trên các phương pháp đề xuất, bao gồm: *ứng dụng phát hiện ngã trong thời gian thực* và *ứng dụng xác thực người dùng sử dụng chữ ký 3D*. Hệ thống phát hiện ngã sử dụng các bộ cảm biến đeo trên người có chi phí thấp và dễ triển khai, có thể phát hiện ngã một cách chính xác trong thời gian thực. Hệ thống xác thực người dùng cho phép các ứng dụng cài đặt trên thiết bị di động thông minh có thể xác thực người dùng thông qua hành động “ký tên” trong không gian trong thời gian thực. Các hệ thống này đều có hiệu năng tốt khi thử nghiệm với tập dữ liệu của người dùng.

DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ

TẠP CHÍ KHOA HỌC

- [1] Pham Cuong, **Diep Nguyen Ngoc**, Phuong Tu Minh (2013), A wearable sensor based approach to real-time fall detection and fine-grained activity recognition, *Journal of Mobile Multimedia*, Rinton Press, Incorporated. **9**(1-2), p. 15–26.
- [2] Đinh Khắc Quân, Phạm Văn Cường, **Nguyễn Ngọc Diệp**, Từ Minh Phương (2013), Tự động phát hiện người ngã trong thời gian thực sử dụng bộ cảm biến gia tốc, *Tạp chí Khoa học và công nghệ*, Viện Hàn lâm KHCN Việt Nam. **51**(1A), 108-116.
- [3] **Nguyễn Ngọc Diệp**, Hà Quang Tấn (2015), Nhận dạng cử chỉ người trong thời gian thực bằng thiết bị di động, *Tạp chí Khoa học và công nghệ*, Viện Hàn lâm KHCN Việt Nam. **53**(2C), 1-13.

HỘI NGHỊ KHOA HỌC

- [4] **Diep Nguyen Ngoc**, Pham Cuong, Phuong Tu Minh (2014), A classifier based approach to real-time fall detection using low-cost wearable sensors, *Proceedings of the Fifth Symposium on Information and Communication Technology*, ACM. p. 14–20.
- [5] **Diep Nguyen Ngoc**, Pham Cuong, Phuong Tu Minh (2015), SigVer3D: Accelerometer Based Verification of 3-D Signatures on Mobile Devices, *Knowledge and Systems Engineering: Proceedings of the 6th International Conference KSE 2014*, Springer. p. 353–365.
- [6] **Diep Nguyen Ngoc**, Pham Cuong, Phuong Tu Minh (2016), An Orientation Histogram based Approach for Fall Detection using Wearable Sensors, *Proceedings of the 14th Pacific Rim International Conference on Artificial Intelligence*, Springer. p. 354-366.
- [7] **Diep Nguyen Ngoc**, Pham Cuong, Phuong Tu Minh (2016), Motion Primitive Forests for Human Activity Recognition using Wearable Sensors, *Proceedings of the 14th Pacific Rim International Conference on Artificial Intelligence*, Springer. p. 340-353.

TÀI LIỆU THAM KHẢO

- [1] Abbate, S., Avvenuti, M., Bonatesta, F., Cola, G., Corsini, P., Vecchio, A. (2012), A smartphone-based fall detection system, *Pervasive and Mobile Computing*, Elsevier B.V. **8**(6), p. 883–99.
- [2] Abbate, S., Avvenuti, M., Cola, G., Corsini, P., Light, J., Vecchio, A. Recognition of false alarms in fall detection systems (2011), *2011 IEEE Consumer Communications and Networking Conference (CCNC)*, p. 23–28.
- [3] Abdel-Hamid, O., Mohamed, A., Jiang, H., Penn, G. (2012), Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition, *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, p. 4277–80.
- [4] Abowd, D., Dey, A.K., Orr, R., Brotherton, J. (1998), Context-awareness in wearable and ubiquitous computing, *Virtual Reality*, Springer. **3**(3), p. 200–11.
- [5] Ahmed, I., Ye, Y., Bhattacharya, S., Asokan, N., Jacucci, G., Nurmi, P. et al. (2015), Checksum gestures: continuous gestures as an out-of-band channel for secure pairing, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 391–401.
- [6] Akbani, R., Kwek, S., Japkowicz, N. (2004), Applying support vector machines to imbalanced datasets, *European Conference on Machine Learning*, p. 39–50.
- [7] Albert, M. V., Kording, K., Herrmann, M., Jayaraman, A. (2012), Fall classification by machine learning using mobile phones, *PLoS ONE*, **7**(5), p. 3–8. <http://dx.doi.org/10.1371/journal.pone.0036556>
- [8] Alberts, J.L., Hirsch, J.R., Koop, M.M., Schindler, D.D., Kana, D.E., Linder, S.M. et al. (2015), Using accelerometer and gyroscopic measures to quantify

- postural stability, *Journal of Athletic Training*, National Athletic Trainers Association. **50**(6), p. 578–88.
- [9] Altun, K., Barshan, B. (2010), Human activity recognition using inertial/magnetic sensor units, *Human Behavior Understanding*, Springer. p. 38–51.
- [10] Amft, O., Kusserow, M., Tröster, G. (2007), Probabilistic parsing of dietary activity events, *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*, p. 242–7.
- [11] Aminian, K., Dadashi, F., Mariani, B., Lenoble-Hoskovec, C., Santos-Eggimann, B., Büla, C.J. (2014), Gait analysis using shoe-worn inertial sensors: how is foot clearance related to walking speed?, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 481–5.
- [12] Anliker, U., Ward, J.A., Lukowicz, P., Troster, G., Dolveck, F., Baer, M. et al. (2004), AMON: a wearable multiparameter medical monitoring and alert system, *Information Technology in Biomedicine, IEEE Transactions on*, **8**(4), p. 415–27. <http://dx.doi.org/10.1109/TITB.2004.837888>
- [13] Ashbrook, D., Starner, T. (2010), MAGIC: A Motion Gesture Design Tool, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA. p. 2159–68.
- [14] Ashbrook, D., Starner, T. (2003), Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users, *Personal Ubiquitous Comput*, Springer-Verlag, London, UK, UK. **7**(5), p. 275–86.
- [15] Bachlin, M., Roggen, D., Troster, G., Plotnik, M., Inbar, N., Meidan, I. et al. (2009), Potentials of Enhanced Context Awareness in Wearable Assistants for Parkinson’s Disease Patients with the Freezing of Gait Syndrome, *Proceedings of the 2009 International Symposium on Wearable Computers*,

IEEE Computer Society, Washington, DC, USA. p. 123–30.

- [16] Bagala, F., Becker, C., Cappello, A., Chiari, L., Aminian, K., Hausdorff, J.M. et al. (2012), Evaluation of accelerometer-based fall detection algorithms on real-world falls, *PLoS ONE*, **7**(5), p. 1–9.
- [17] Bao, L., Intille, S.S. (2004), Activity Recognition from User-Annotated Acceleration Data, *Pervasive Computing*, p. 1–17.
- [18] Bengio, Y. (2009), Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, Now Publishers Inc. **2**(1), p. 1–127.
- [19] Bhattacharya, S., Nurmi, P., Hammerla, N., Plötz, T. (2014), Using unlabeled data in a sparse-coding framework for human activity recognition, *Pervasive and Mobile Computing*, Elsevier. **15**, p. 242–62.
- [20] Blanke, U., Schiele, B. (2009), Daily routine recognition through activity spotting, *Location and Context Awareness*, Springer. p. 192–206.
- [21] BMI Lab. BMI Lab. MobiFall2 dataset: <http://www.bmi.teicrete.gr/index.php/research/mobifall> [Internet], (accessed on 12/22/2015).
- [22] Bourke, A.K., Lyons, G.M. (2008), A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor, **30**, p. 84–90.
- [23] Bourke, A.K., O’Brien, J. V, Lyons, G.M. (2007), Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm, *Gait & Posture*, **26**(2), p. 194–9.
- [24] Brady, S., Dunne, L.E., Tynan, R., Diamond, D., Smyth, B., O’Hare, G.M.P. (2005), Garment-based monitoring of respiration rate using a foam pressure sensor, *Wearable Computers, 2005 Proceedings Ninth IEEE International Symposium on*, p. 214–5.
- [25] Breiman, L. (2001), Random forests, *Machine Learning*, Springer. **45**(1), p.

5–32.

- [26] Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A. (1984), Classification and regression trees, CRC press.
- [27] Brillinger, D.R. (2001), Time series: data analysis and theory, Siam.
- [28] Bulling, A., Blanke, U., Schiele, B. (2014), A tutorial on human activity recognition using body-worn inertial sensors, *ACM Computing Surveys (CSUR)*, **46**(3), p. 1–33.
- [29] Chang, C.-C., Lin, C.-J. (2011), LIBSVM: a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM. **2**(3), p. 27.
- [30] Chawla, N. V, Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002), SMOTE: synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, **16**, p. 321–57.
- [31] Chen, J., Kam, A.H., Zhang, J., Liu, N., Shue, L. (2005), Bathroom activity monitoring based on sound, *Pervasive Computing*, Springer. p. 47–61.
- [32] Chen, L., Hoey, J., Nugent, C.D., Cook, D.J., Yu, Z. (2012), Sensor-based activity recognition, *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, **42**(6), p. 790–808.
- [33] Chen, L., Khalil, I. (2011), Activity Recognition: Approaches, Practices and Trends, *Activity Recognition in Pervasive Intelligent Environments*, **4**, p. 1–31. http://dx.doi.org/10.2991/978-94-91216-05-3_1
- [34] Chen, Y.-P., Yang, J.-Y., Liou, S.-N., Lee, G.-Y., Wang, J.-S. (2008), Online classifier construction algorithm for human activity detection using a tri-axial accelerometer, *Applied Mathematics and Computation*, Elsevier. **205**(2), p. 849–60.
- [35] Clarke, N.L., Furnell, S.M. (2007), Authenticating mobile phone users using

- keystroke analysis, *International Journal of Information Security*, Springer. **6**(1), p. 1–14.
- [36] Cox, D.R., Oakes, D. (1984), Analysis of survival data, CRC Press.
- [37] Crouter, S.E., Flynn, J.I., Bassett Jr, D.R. (2015), Estimating physical activity in youth using a wrist accelerometer., *Medicine and Science in Sports and Exercise*, **47**(5), p. 944–51.
- [38] Diep, N.N., Pham, C., Phuong, T.M. (2014), A classifier based approach to real-time fall detection using low-cost wearable sensors, *Proceedings of the Fifth Symposium on Information and Communication Technology*, p. 14–20.
- [39] Doukas, C., Maglogiannis, I., Tragas, P., Liapis, D., Yovanof, G. (2007), Patient fall detection using support vector machines, *Artificial Intelligence and Innovations 2007: From Theory to Applications*, Springer. p. 147–56.
- [40] Figo, D., Diniz, P.C., Ferreira, D.R., Cardoso, J.M.P. (2010), Preprocessing techniques for context recognition from accelerometer data, *Personal and Ubiquitous Computing*, Springer. **14**(7), p. 645–62.
- [41] Frank, K., Vera Nadales, M.J., Robertson, P., Pfeifer, T. (2010), Bayesian recognition of motion related activities with inertial sensors, *Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing - Ubicomp '10*, (January 2016), p. 445.
- [42] Fu, Z., Culurciello, E., Lichtsteiner, P., Delbruck, T. (2008), Fall detection using an address-event temporal contrast vision sensor, *Circuits and Systems, 2008 ISCAS 2008 IEEE International Symposium on*, p. 424–7.
- [43] Gafurov, D., Helkala, K., Søndrol, T. (2006), Biometric gait authentication using accelerometer sensor, *Journal of Computers*, **1**(7), p. 51–9.
- [44] Ghasemzadeh, H., Loseu, V., Jafari, R. (2010), Collaborative signal processing for action recognition in body sensor networks, *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in*

Sensor Networks - IPSN '10, p. 244.

- [45] Guyon, I., Elisseeff, A. (2003), An introduction to variable and feature selection, *The Journal of Machine Learning Research*, JMLR. org. **3**, p. 1157–82.
- [46] Hadid, A., Heikkilä, J.Y., Silvén, O., Pietikäinen, M. (2007), Face and eye detection for person authentication in mobile phones, *Distributed Smart Cameras, 2007 ICDCS'07 First ACM/IEEE International Conference on*, p. 101–8.
- [47] Hao, T., Xing, G., Zhou, G. (2015), RunBuddy: a smartphone system for running rhythm monitoring, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 133–44.
- [48] Harris, Z.S. (1954), Distributional structure, *Word*, Taylor & Francis. **10**(2-3), p. 146–62.
- [49] Hartmann, B., Abdulla, L., Mittal, M., Klemmer, S.R. (2007), Authoring Sensor-based Interactions by Demonstration with Direct Manipulation and Pattern Recognition, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, NY, USA. p. 145–54.
- [50] Hazelhoff, L., Han, J., others. (2008), Video-based fall detection in the home using principal component analysis, *Advanced Concepts for Intelligent Vision Systems*, p. 298–309.
- [51] He, Z., Jin, L. (2009), Activity recognition from acceleration data based on discrete cosine transform and svm, *Systems, Man and Cybernetics, 2009 SMC 2009 IEEE International Conference on*, p. 5041–4.
- [52] He, Z.-Y., Jin, L.-W. (2008), Activity recognition from acceleration data using AR model representation and SVM, *Machine Learning and Cybernetics, 2008 International Conference on*, p. 2245–50.
- [53] Hevesi, P., Wille, S., Pirkl, G., Wehn, N., Lukowicz, P. (2014), Monitoring

- household activities and user location with a cheap, unobtrusive thermal sensor array, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 141–5.
- [54] Hoyle, R., Templeman, R., Armes, S., Anthony, D., Crandall, D., Kapadia, A. (2014), Privacy behaviors of lifeloggers using wearable cameras, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 571–82.
- [55] Huynh, T., Blanke, U., Schiele, B. (2007), Scalable recognition of daily activities with wearable sensors, *Location-and Context-Awareness*, Springer. p. 50–67.
- [56] Huynh, T., Fritz, M., Schiele, B. (2008), Discovery of activity patterns using topic models, *Proceedings of the 10th International Conference on Ubiquitous Computing*, p. 10–9.
- [57] Huynh, T., Huynh, T., Fritz, M., Fritz, M., Schiele, B., Schiele, B. (2008), Discovery of activity patterns using topic models, *Proceedings of the 10th International Conference on Ubiquitous Computing - UbiComp '08*, p. 10. <http://dx.doi.org/10.1145/1409635.1409638>
- [58] Huynh, T., Schiele, B. (2005), Analyzing Features for Activity Recognition, *Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence: Innovative Context-Aware Services: Usages and Technologies*, ACM, New York, NY, USA. p. 159–63.
- [59] Hwang, J.Y., Kang, J.M., Jang, Y.W., Kim, H.C. (2004), Development of novel algorithm and real-time monitoring ambulatory system using Bluetooth module for fall detection in the elderly, *Engineering in Medicine and Biology Society, 2004 IEMBS'04 26th Annual International Conference of the IEEE*, p. 2204–7.
- [60] Igual, R., Medrano, C., Plaza, I. (2015), A comparison of public datasets for

- acceleration-based fall detection, *Medical Engineering and Physics*, **37**(9), p. 870–8.
- [61] Igual, R., Medrano, C., Plaza, I. (2013), Challenges , issues and trends in fall detection systems, *BioMedical Engineering OnLine*, BioMedical Engineering OnLine. **12**(1), p. 1.
- [62] Jain, A., Kanhangad, V. (2015), Exploring orientation and accelerometer sensor data for personal authentication in smartphones using touchscreen gestures, *Pattern Recognition Letters*, Elsevier. **68**, p. 351–60.
- [63] Jamil, S., Basalamah, A., Lbath, A., Youssef, M. (2015), Hybrid participatory sensing for analyzing group dynamics in the largest annual religious gathering, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 547–58.
- [64] Jantaraprim, P., Phukpattaranont, P., Limsakul, C., Wongkittisuksa, B. (2012), Fall Detection for the elderly using a Support Vector Machine, *International Journal of Soft Computing and Engineering*, **2**(1), p. 484–90.
- [65] Jeran, S., Steinbrecher, A., Pischon, T. (2016), Prediction of activity related energy expenditure using accelerometer derived physical activity under free-living conditions—a systematic review, *International Journal of Obesity*, Nature Publishing Group.
- [66] Kangas, M., Konttila, A., Lindgren, P., Winblad, I., Jämsä, T. (2008), Comparison of low-complexity fall detection algorithms for body attached accelerometers, *Gait & Posture*, Elsevier. **28**(2), p. 285–91.
- [67] Katz, S., Downs, T.D., Cash, H.R., Grotz, R.C. (1970), Progress in development of the index of ADL, *The Gerontologist*, Oxford University Press. **10**(1 Part 1), p. 20–30.
- [68] Kern, N., Schiele, B., Junker, H., Lukowicz, P., Tröster, G. (2003), Wearable sensing to annotate meeting recordings, *Personal and Ubiquitous Computing*,

Springer. 7(5), p. 263–74.

- [69] Khan, S., Karg, M., Kulic, D., Hoey, J. (2014), X-Factor HMMs for detecting falls in the absence of fall-specific training data, *International Work-Conference on Ambient Assisted Living*, p. 1–9.
- [70] Kohavi, R., John, G.H. (1997), Wrappers for feature subset selection, *Artificial Intelligence*, Elsevier. 97(1), p. 273–324.
- [71] Krause, a., Siewiorek, D.P., Smailagic, A., Farrington, J. (2003), Unsupervised, dynamic identification of physiological and activity context in wearable computing, *Seventh IEEE International Symposium on Wearable Computers, 2003 Proceedings*,.
- [72] Krumm, J., Horvitz, E. (2006), Predestination: Inferring destinations from partial trajectories, *UbiComp 2006: Ubiquitous Computing*, Springer. p. 243–60.
- [73] Kunze, K., Barry, M., Heinz, E., Lukowicz, P., Majoe, D., Gutknecht, J. (2006), Towards Recognizing Tai Chi {Â}?’ An Initial Experiment Using Wearable Sensors, *Applied Wearable Computing (IFAWC), 2006 3rd International Forum on*, p. 1–6.
- [74] Kwapisz, J.R., Weiss, G.M., Moore, S.A. (2011), Activity recognition using cell phone accelerometers, *ACM SigKDD Explorations Newsletter*, ACM. 12(2), p. 74–82.
- [75] Ladha, C., Hammerla, N., Olivier, P., Plötz, T. (2013), ClimbAX: Skill assessment for climbing enthusiasts, *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 235–44.
- [76] Lai, C.-F., Chang, S.-Y., Chao, H.-C., Huang, Y.-M. (2011), Detection of cognitive injured body region using multiple triaxial accelerometers for elderly falling, *Sensors Journal, IEEE, IEEE*. 11(3), p. 763–70.

- [77] Lane, N.D., Georgiev, P. (2015), Can deep learning revolutionize mobile sensing?, *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, p. 117–22.
- [78] Lara, O.D., Labrador, M. a. (2013), A Survey on Human Activity Recognition using Wearable Sensors, *IEEE Communications Surveys & Tutorials*, **15**(3), p. 1192–209.
- [79] LeCun, Y., Bengio, Y., Hinton, G. (2015), Deep learning, *Nature*, Nature Publishing Group. **521**(7553), p. 436–44.
- [80] LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., Huang, F. (2006), A tutorial on energy-based learning, *Predicting Structured Data*, **1**, p. 0.
- [81] Lee, S.-W., Mase, K. (2002), Activity and location recognition using wearable sensors, *IEEE Pervasive Computing*, **1**(3), p. 24–32.
- [82] Lester, J., Choudhury, T., Borriello, G. (2006), A practical approach to recognizing physical activities, *Pervasive Computing*, Springer. p. 1–16.
- [83] Lester, J., Choudhury, T., Kern, N., Borriello, G., Hannaford, B. (2005), A Hybrid Discriminative/Generative Approach for Modeling Human Activities., *IJCAI*, p. 766–72.
- [84] Li, Q., Stankovic, J.A., Hanson, M.A., Barth, A.T., Lach, J. (2009), Accurate , Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information, p. 140–5. <http://dx.doi.org/10.1109/BSN.2009.46>
- [85] Liao, L., Patterson, D.J., Fox, D., Kautz, H. (2007), Learning and inferring transportation routines, *Artificial Intelligence*, Elsevier. **171**(5), p. 311–31.
- [86] Lin, C.-W., Ling, Z.-H. (2007), Automatic fall incident detection in compressed video for intelligent homecare, *Computer Communications and Networks, 2007 ICCCN 2007 Proceedings of 16th International Conference on*, p. 1172–7.

- [87] Lindemann, U., Hock, A., Stuber, M., Keck, W., Becker, C. (2005), Evaluation of a fall detector based on accelerometers: A pilot study, *Medical and Biological Engineering and Computing*, Springer. **43**(5), p. 548–51.
- [88] Liszka, K.J., Mackin, M.A., Lichter, M.J., York, D.W., Pillai, D., Rosenbaum, D.S. (2004), Keeping a beat on the heart, *Pervasive Computing, IEEE*, **3**(4), p. 42–9. <http://dx.doi.org/10.1109/MPRV.2004.10>
- [89] Liu, S., Silverman, M. (2001), A practical guide to biometric security technology, *IT Professional, IEEE*. **3**(1), p. 27–32.
- [90] Logan, B., Healey, J., Philipose, M., Tapia, E.M., Intille, S. (2007), A long-term evaluation of sensing modalities for activity recognition, Springer.
- [91] Lukowicz, P., Hanser, F., Szubski, C., Schobersberger, W. (2006), Detecting and interpreting muscle activity with wearable force sensors, *Lecture Notes in Computer Science*, Springer. **3968**, p. 101–16.
- [92] Lukowicz, P., Ward, J.A., Junker, H., Stäger, M., Tröster, G., Atrash, A. et al. (2004), Recognizing workshop activity using body worn microphones and accelerometers, *Pervasive Computing*, Springer. p. 18–32.
- [93] Matsumoto, T., Matsumoto, H., Yamada, K., Hoshino, S. (2002), Impact of artificial gummy fingers on fingerprint systems, *Electronic Imaging 2002*, p. 275–89.
- [94] Maurer, U., Smailagic, A., Siewiorek, D.P., Deisher, M. (2006), Activity recognition and monitoring using multiple sensors on different body positions, *Wearable and Implantable Body Sensor Networks, 2006 BSN 2006 International Workshop on*, p. 4 – pp.
- [95] Maurtua, I., Kirisci, P.T., Stiefmeier, T., Sbodio, M.L., Witt, H. (2007), A Wearable Computing Prototype for supporting training activities in Automotive Production, *Applied Wearable Computing (IFAWC), 2007 4th International Forum on*, p. 1–12.

- [96] Medrano, C., Igual, R., Plaza, I., Castro, M. (2014), Detecting falls as novelties in acceleration patterns acquired with smartphones, *PLoS ONE*, **9**(4).
- [97] Ming, J., Hazen, T.J., Glass, J.R., Reynolds, D.A. (2007), Robust speaker recognition in noisy conditions, *Audio, Speech, and Language Processing, IEEE Transactions on*, IEEE. **15**(5), p. 1711–23.
- [98] Minnen, D., Starner, T., Essa, I., Isbell, C. (2006), Discovering characteristic actions from on-body sensor data, *Wearable Computers, 2006 10th IEEE International Symposium on*, p. 11–8.
- [99] Mokaya, F., Lucas, R., Noh, H.Y., Zhang, P. (2015), Myovibe: Vibration based wearable muscle activation detection in high mobility exercises, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 27–38.
- [100] Moosmann, F., Triggs, B., Jurie, F. (2007), Fast discriminative visual codebooks using randomized clustering forests, *Twentieth Annual Conference on Neural Information Processing Systems (NIPS'06)*, p. 985–92.
- [101] Myers, C.S., Rabiner, L.R. (1981), A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected-Word Recognition, *Bell System Technical Journal*, Wiley Online Library. **60**(7), p. 1389–409.
- [102] Nils Y., H., Shane, H., Thomas, P. (2016), Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables, *25th International Joint Conference on Artificial Intelligence*,.
- [103] Noury, N., Rumeau, P., Bourke, A.K., ÓLaighin, G., Lundy, J.E. (2008), A proposal for the classification and evaluation of fall detectors, *Irbm*, Elsevier. **29**(6), p. 340–9.
- [104] de Oliveira, R., Cherubini, M., Oliver, N. (2010), MoviPill: Improving Medication Compliance for Elders Using a Mobile Persuasive Social Game,

Proceedings of the 12th ACM International Conference on Ubiquitous Computing, ACM, New York, NY, USA. p. 251–60.

- [105] Oliver, N., Flores-Mangas, F. (2007), HealthGear: automatic sleep apnea detection and monitoring with a mobile phone, *Journal of Communications*, **2**(2), p. 1–9.
- [106] Özdemir, A., Barshan, B. (2014), Detecting Falls with Wearable Sensors Using Machine Learning Techniques, *Sensors*, **14**(6), p. 10691–708. <http://dx.doi.org/10.3390/s140610691>
- [107] Pärkkä, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J., Korhonen, I. (2006), Activity classification using realistic data from wearable sensors, *Information Technology in Biomedicine, IEEE Transactions on*, IEEE. **10**(1), p. 119–28.
- [108] Patterson, D.J., Liao, L., Fox, D., Kautz, H. (2003), Inferring high-level behavior from low-level sensors, *UbiComp 2003: Ubiquitous Computing*, p. 73–89.
- [109] Peng, H., Long, F., Ding, C. (2005), Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE. **27**(8), p. 1226–38.
- [110] Pham, C., Diep, N.N., Phuong, T.M. (2013), A wearable sensor based approach to real-time fall detection and fine-grained activity recognition, *Journal of Mobile Multimedia*, Rinton Press, Incorporated. **9**(1-2), p. 15–26.
- [111] Pierleoni, P., Belli, A., Palma, L., Pellegrini, M., Pernini, L., Valenti, S. (2015), A High Reliability Wearable Device for Elderly Fall Detection, *Sensors Journal, IEEE*, IEEE. **15**(8), p. 4544–53.
- [112] Pirkl, G., Stockinger, K., Kunze, K., Lukowicz, P. (2008), Adapting magnetic resonant coupling based relative positioning technology for wearable activity

- recogniton, *Wearable Computers, 2008 ISWC 2008 12th IEEE International Symposium on*, p. 47–54.
- [113] Plamondon, R., Lorette, G. (1989), Automatic signature verification and writer identification—the state of the art, *Pattern Recognition*, Elsevier. **22**(2), p. 107–31.
- [114] Plötz, T., Hammerla, N.Y., Olivier, P. (2011), Feature learning for activity recognition in ubiquitous computing, *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, p. 1729.
- [115] Plötz, T., Hammerla, N.Y., Rozga, A., Reavis, A., Call, N., Abowd, G.D. (2012), Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities, *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ACM, New York, NY, USA. p. 391–400. <http://dx.doi.org/10.1145/2370216.2370276>
- [116] Rahman, T., Adams, A.T., Ravichandran, R.V., Zhang, M., Patel, S.N., Kientz, J.A. et al. (2015), Dopplesleep: A contactless unobtrusive sleep sensing system using short-range doppler radar, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 39–50.
- [117] Ravi, N., Dandekar, N., Mysore, P., Littman, M.L. (2005), Activity recognition from accelerometer data, *AAAI*, p. 1541–6.
- [118] Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M. (2010), Using mobile phones to determine transportation modes, *ACM Transactions on Sensor Networks (TOSN)*, ACM. **6**(2), p. 13.
- [119] Roggen, D., Forster, K., Calatroni, A., Holleccek, T., Fang, Y., Troster, G. et al. (2009), OPPORTUNITY: Towards opportunistic activity and context recognition systems, *World of Wireless, Mobile and Multimedia Networks Workshops, 2009 WoWMoM 2009 IEEE International Symposium on a*, p. 1–

6.

- [120] Ross, A., Jain, A.K., Pankati, S. (1999), A prototype hand geometry-based verification system, *Proceedings of 2nd Conference on Audio and Video Based Biometric Person Authentication*, p. 166–71.
- [121] Scholl, P.M., Wille, M., Van Laerhoven, K. (2015), Wearables in the wet lab: a laboratory system for capturing and guiding experiments, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 589–99.
- [122] Shieh, W.-Y., Huang, J.-C. (2009), Speedup the multi-camera video-surveillance system for elder falling detection, *Embedded Software and Systems, 2009 ICESS'09 International Conference on*, p. 350–5.
- [123] Shotton, J., Johnson, M., Cipolla, R. (2008), Semantic texton forests for image categorization and segmentation, *Computer Vision and Pattern Recognition, 2008 CVPR 2008 IEEE Conference on*, p. 1–8.
- [124] Somol, P., Novovičová, J., Pudil, P. (2006), Flexible-Hybrid Sequential Floating Search in Statistical Feature Selection, *Proceedings of the 2006 Joint IAPR International Conference on Structural, Syntactic, and Statistical Pattern Recognition*, Springer-Verlag, Berlin, Heidelberg. p. 632–9.
- [125] Starner, T., Weaver, J., Pentland, A. (1998), A wearable computer based american sign language recognizer, *Assistive Technology and Artificial Intelligence*, Springer. p. 84–96.
- [126] Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., Tröster, G. (2008), Wearable activity tracking in car manufacturing, *IEEE Pervasive Computing*, IEEE. (2), p. 42–50.
- [127] Stikic, M., Huynh, T., Laerhoven, K. Van, Schiele, B. (2008), ADL recognition based on the combination of RFID and accelerometer sensing, *Pervasive Computing Technologies for Healthcare, 2008 PervasiveHealth*

2008 Second International Conference on, p. 258–63.

- [128] Stone, E.E., Skubic, M. (2015), Fall detection in homes of older adults using the Microsoft Kinect, *Biomedical and Health Informatics, IEEE Journal of, IEEE*. **19**(1), p. 290–301.
- [129] Sundholm, M., Cheng, J., Zhou, B., Sethi, A., Lukowicz, P. (2014), Smartmat: Recognizing and counting gym exercises with low-cost resistive pressure sensing matrix, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 373–82.
- [130] Sung, J., Ponce, C., Selman, B., Saxena, A. (2012), Unstructured human activity detection from RGBD images, *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, p. 842–9.
- [131] Sung, M., DeVaul, R., Jimenez, S., Gips, J., Pentland, A.S. (2004), Shiver motion and core body temperature classification for wearable soldier health monitoring systems, *Wearable Computers, 2004 ISWC 2004 Eighth International Symposium on*, p. 192–3.
- [132] Sung, M., Marci, C., Pentland, A. (2005), Wearable feedback systems for rehabilitation, *Journal of NeuroEngineering and Rehabilitation*, BioMed Central. **2**(1).
- [133] Tang, Y., Zhang, Y.-Q., Chawla, N. V, Krasser, S. (2009), SVMs modeling for highly imbalanced classification, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, IEEE. **39**(1), p. 281–8.
- [134] Tapia, E.M., Intille, S.S., Haskell, W., Larson, K., Wright, J., King, A. et al. (2007), Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor, *Wearable Computers, 2007 11th IEEE International Symposium on*, p. 37–40.
- [135] Tapia, E.M., Intille, S.S., Larson, K. (2004), Activity recognition in the home using simple and ubiquitous sensors, Ferscha A, and Mattern F, editors.

Springer.

- [136] Taylor, B., Dey, A., Siewiorek, D., Smailagic, A. (2015), Using physiological sensors to detect levels of user frustration induced by system delays, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 517–28.
- [137] Tesselndorf, B., Gravenhorst, F., Arnrich, B., Tröster, G. (2011), An imu-based sensor network to continuously monitor rowing technique on the water, *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2011 Seventh International Conference on*, p. 253–8.
- [138] Thomaz, E., Essa, I., Abowd, G.D. (2015), A practical approach for recognizing eating moments with wrist-mounted inertial sensing, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 1029–40.
- [139] Vavoulas, G., Pediaditis, M., Spanakis, E.G., Tsiknakis, M. (2013), The MobiFall dataset: An initial evaluation of fall detection algorithms using smartphones, *13th IEEE International Conference on BioInformatics and BioEngineering*, p. 1–4.
- [140] Veropoulos, K., Campbell, C., Cristianini, N., others. (1999), Controlling the sensitivity of support vector machines, *Proceedings of the International Joint Conference on AI*, p. 55–60.
- [141] Villalba, E., Ottaviano, M., Arredondo, M.T., Martinez, A., Guillen, S. (2006), Wearable monitoring system for heart failure assessment in a mobile environment, *Computers in Cardiology, 2006*, p. 237–40.
- [142] Vollmer, C., Gross, H.-M., Eggert, J.P. (2013), Learning features for activity recognition with shift-invariant sparse coding, *Artificial Neural Networks and Machine Learning--ICANN 2013*, Springer. p. 367–74.
- [143] Wan, D. (1999), Magic medicine cabinet: A situated portal for consumer

- healthcare, *Handheld and Ubiquitous Computing*, p. 352–5.
- [144] Wang, E.J., Lee, T.-J., Mariakakis, A., Goel, M., Gupta, S., Patel, S.N. (2015), Magnifisense: Inferring device interaction using wrist-worn passive magneto-inductive sensors, *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, p. 15–26.
- [145] Wu, W., Au, L., Jordan, B., Stathopoulos, T., Batalin, M., Kaiser, W. et al. (2008), The smartcane system: an assistive device for geriatrics, *Proceedings of the ICST 3rd International Conference on Body Area Networks*, p. 2.
- [146] Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H. et al. (2008), Top 10 algorithms in data mining, *Knowledge and Information Systems*, Springer. **14**(1), p. 1–37.
- [147] Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L. et al. (2008), Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection, *Wireless Sensor Networks*, Springer. p. 17–33.
- [148] Zeng, M., Nguyen, L.T., Yu, B., Mengshoel, O.J., Zhu, J., Wu, P. et al. (2014), Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors, *6th International Conference on Mobile Computing, Applications and Services*,.
- [149] Zhang, M., Sawchuk, A. a. (2013), Human daily activity recognition with sparse representation using wearable sensors, *IEEE Journal of Biomedical and Health Informatics*, **17**(3), p. 553–60. <http://dx.doi.org/10.1109/JBHI.2013.2253613>
- [150] Zhang, M., Sawchuk, A.A. (2011), A feature selection-based framework for human activity recognition using wearable multimodal sensors, *Proceedings of the 6th International Conference on Body Area Networks*, p. 92–8.
- [151] Zhang, M., Sawchuk, A.A. (2012), Motion primitive-based human activity

recognition using a bag-of-features approach, *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, p. 631–40.

- [152] Zheng, Y., Wong, W., Guan, X., Trost, S. (2013), Physical Activity Recognition from Accelerometer Data Using a Multi-Scale Ensemble Method, *Proceedings of the Twenty-Fifth Innovative Applications of Artificial Intelligence Conference Physical*, p. 1575–81.
- [153] Brickhouse Alert: <http://www.brickhousealert.com/personal-emergency-medical-alarm.html>, (accessed on 5/25/2013).
- [154] DLR. DLR dataset: http://www.dlr.de/kn/en/Portaldata/27/Resources/dokumente/04_abteilungen_fs/kooperative_systeme/high_precision_reference_data/Activity_DataSet.zip [Internet], (accessed on 12/22/2015).
- [155] EduQTech Group. tFall dataset: <http://eduqtech.unizar.es/fall-adl-data/> [Internet], (accessed on 12/22/2015).
- [156] LSM330DLC: <http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/DM00037200.pdf> [Internet], (accessed on 5/25/2014).
- [157] Wii Remote: http://en.wikipedia.org/wiki/Wii_Remote [Internet], (accessed on 5/25/2014).