

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG**

UNG NHO DÃI

**NGHIÊN CỨU TRÍCH CHỌN ĐẶC TÍNH
TRONG NHẬN DẠNG HÀNH ĐỘNG NGƯỜI
TRONG KHÔNG GIAN 3D**

LUẬN VĂN THẠC SĨ KỸ THUẬT

ĐÀ NẴNG - Năm 2015

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG**

UNG NHO DÃI

**NGHIÊN CỨU TRÍCH CHỌN ĐẶC TÍNH
TRONG NHẬN DẠNG HÀNH ĐỘNG NGƯỜI
TRONG KHÔNG GIAN 3D**

**Chuyên ngành: KHOA HỌC MÁY TÍNH
Mã số: 60.48.01**

LUẬN VĂN THẠC SĨ KỸ THUẬT

Người hướng dẫn khoa học: TS. PHẠM MINH TUẤN

ĐÀ NẴNG - Năm 2015

LỜI CẢM ƠN

Tôi chân thành cảm ơn

LỜI CAM ĐOAN

Tôi xin cam đoan :

- a. Những nội dung trong luận văn này là do tôi thực hiện dưới sự hướng dẫn trực tiếp của thầy giáo TS. Phạm Minh Tuấn.*
- b. Mọi tham khảo dùng trong luận văn đều được trích dẫn rõ ràng và trung thực tên tác giả, tên công trình, thời gian, địa điểm công bố.*
- c. Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo, hay gian trá, tôi xin chịu hoàn toàn trách nhiệm.*

Tác giả

UNG NHỎ DÃI

MỤC LỤC

Lời cam đoan	
Mục lục	i
Danh mục các từ viết tắt	ii
Danh mục các bảng biểu	v
Danh mục các hình vẽ	vi
MỞ ĐẦU	3
NGHIÊN CỨU TỔNG QUAN.....	9
I. Nhận dạng hành động người trong không gian 3D.....	9
I.1. Các phương pháp thu thập dữ liệu chuyển động 3D	9
I.1.1. <i>Phương pháp sử dụng stereo camera</i>	9
I.1.2. <i>Phương pháp sử dụng Mocap</i>	11
I.1.3. <i>Phương pháp sử dụng range sensor</i>	12
I.2. Các phương pháp học máy thường sử dụng	12
I.2.1. <i>Máy vector hỗ trợ</i>	13
I.2.2. <i>Mô hình Markov ẩn</i>	14
II. Hệ thống chụp chuyển động – Mocap	14
II.1. Mocap	15
II.2. Dữ liệu thu được từ Mocap.....	16
II.3. Cấu trúc Acclaim.....	16
II.3.1. <i>Cấu trúc tệp ASF</i>	16
II.3.2. <i>Cấu trúc tệp AMC</i>	18
III. Trích chọn, lựa chọn đặc tính	19
III.1. Phương pháp phân tích thành phần chính – PCA.....	19
III.2. Phương pháp phân tích biệt thức tuyến tính – LDA.....	22
III.3. Các phương pháp lựa chọn đặc tính	22
GIẢI PHÁP ĐỀ XUẤT.....	23
I. Mô hình tổng quan.....	23
II. Trích chọn đặc tính	23
III. Giải thuật huấn luyện.....	23
III.1. Tiêu đề mục con	23
III.1.1. <i>Tiêu đề mục con nhỏ hơn</i>	23
TRIỂN KHAI VÀ ĐÁNH GIÁ KẾT QUẢ.....	25
I. Dữ liệu nhận dạng.....	25
II. Môi trường triển khai.....	25
III. Kết quả thực nghiệm.....	25
IV. Đánh giá.....	25
IV.1. Tiêu đề mục con	25
IV.1.1. <i>Tiêu đề mục con nhỏ hơn</i>	25
KẾT LUẬN.....	27

PHỤ LỤC.....	II
TÀI LIỆU THAM KHẢO	V

DANH MỤC CÁC CHỮ VIẾT TẮT

ACL	Agent Communication Language
AMS	Agent Management System
AP	Agent Platform
API	Application Programming Interface
CF	Communication Failitator

DANH MỤC CÁC BẢNG BIỂU

DANH MỤC CÁC HÌNH VẼ

Hình 1	Error! Bookmark not defined.
--------------	------------------------------

MỞ ĐẦU

1. Lý do chọn đề tài

a. Bối cảnh chung

Từ những năm 80 của thế kỷ trước, nhận dạng hành động (activity recognition) [1] đã thu hút rất nhiều quan tâm, nghiên cứu của các nhà khoa học. Nó được sử dụng rộng rãi trong nhiều ứng dụng và trong các lĩnh vực khác như y học, xã hội học, giao tiếp người máy. Nhận dạng hành động được chia làm hai loại chính: loại thứ nhất sử dụng cảm biến (sensor-based) loại thứ hai sử dụng hình ảnh (vision-based) [2].

Hoạt động nhận dạng sử dụng cảm biến kết hợp sự đa dạng của mạng lưới thiết bị cảm biến với việc khai phá dữ liệu và học máy để mô hình hoạt động của con người. Cấu hình các thiết bị di động hiện nay đủ mạnh để thu thập dữ liệu từ nhiều loại cảm biến khác nhau và xử lý các dữ liệu đó để có thể đưa ra ước lượng về năng lượng cần thiết cho các hoạt động hàng ngày của con người. Các nhà nghiên cứu tin rằng, với sự phát triển mạnh mẽ của các loại thiết bị và các loại cảm biến, việc theo dõi và nhận dạng hoạt động của con người sẽ trở nên dễ dàng hơn [3].

Vấn đề quan trọng và thách thức nhất đối với nhận dạng hành động là nhận biết được hành động của con người thông qua hình ảnh từ hệ thống các camera. Kỹ thuật chủ yếu được dùng để nhận dạng từ hình ảnh là thị giác máy tính (vision computer). Có rất nhiều phương pháp đã được áp dụng trong nhận dạng hành động dựa vào hình ảnh như optical flow, bộ lọc Kalman, mô hình Markov ẩn, ... sử dụng các dữ liệu khác nhau từ camera, sóng âm (stereo) và hồng ngoại [1, 2, 4].



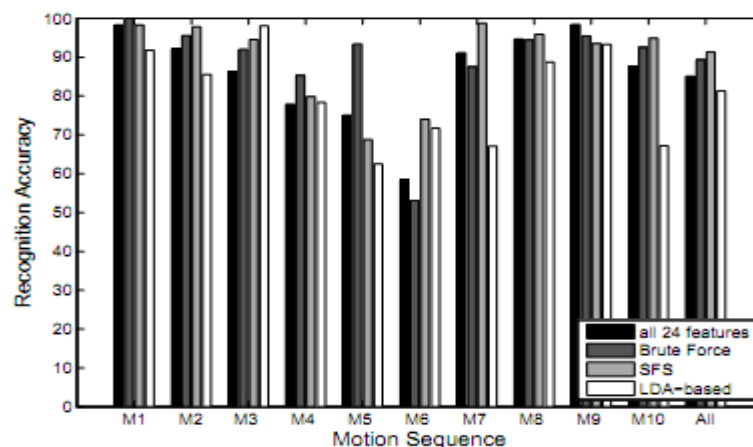
Hình 1. Microsoft Kinect Camera

Gần đây, một số nhà nghiên cứu đã sử dụng camera RGBD (Red, Green, Blue, Depth) như Microsoft Kinect [3] (xem hình 1) để nhận dạng hoạt động của con người. Dữ liệu thu được từ các thiết bị chuyên dụng này là dữ liệu chuyển động 3D của cơ thể người. Những dữ liệu này sẽ là dữ liệu huấn luyện hữu ích cho các mô hình nhận dạng hành động.

b. Các phương pháp trước đây

Những nghiên cứu gần đây trong lĩnh vực nhận dạng hoạt động người chủ yếu tập trung vào nghiên cứu và nhận dạng từ những video được quay bởi các camera thông dụng [2, 4]. Khó khăn lớn nhất đối với dữ liệu từ camera thông dụng là chỉ quay được ở một hướng, dẫn đến sự thiếu hụt dữ liệu, nếu kết hợp nhiều camera thì vẫn không đảm bảo thu được toàn bộ hoạt động, đồng thời giảm hiệu năng của quá trình nhận dạng. Mặc dù đã có rất nhiều nỗ lực trong những thập kỷ qua, lĩnh vực nhận dạng hoạt động người từ dữ liệu video vẫn còn nhiều khó khăn, thách thức.

Từ sau sự ra đời của các thiết bị cảm biến chiều sâu (depth sensor) [3], đã có một hướng tiếp cận mới trong nhận dạng hành động người, đó là sử dụng dữ liệu chuyển động 3D [5, 6, 7]. Trong 20 năm trở lại, một số phương pháp chính để thu thập dữ liệu 3D hoặc là sử dụng hệ thống chụp chuyển động dựa vào marker như là Mocap (Motion Capture) [8] hoặc là dùng sóng âm (stereo) - chụp hình ảnh 2D từ nhiều hướng khác nhau để dựng thành mô hình 3D [9].



Hình 2. So sánh kết quả giữa các phương pháp trích chọn đặc tính

Sau khi đã thu thập được dữ liệu 3D, có rất nhiều phương pháp đã được đề xuất để hoàn thành quá trình nhận dạng. Điểm chung của các phương pháp này là cố gắng làm giảm số lượng thuộc tính của dữ liệu nhận dạng trước khi xây dựng mô hình huấn luyện. D. Gehrig và T. Schultz đã nghiên cứu, thực nghiệm ba phương pháp trích chọn và lựa chọn đặc tính khác nhau (Brute Force, SFS, LDA) và đã giảm đáng kể hiệu năng của quá trình nhận dạng so với dữ liệu ban đầu [10] (xem hình 2). L. Fengjun và N. RamaKant phân tích và đưa ra bảy loại đặc tính khác nhau dựa vào tư thế và sự kết hợp giữa các khớp xương, sau đó dựng mô hình huấn luyện và nhận dạng dùng Markov ẩn [5] (xem hình 3). K. Dana, T. Wataru và N. Yoshihiko đã phát triển một mô hình học tăng cường mới dựa trên mô hình Markov ẩn [11].

action	walk	run	j upward	j forward	stand	sit	bow	lie
Exp.1	94.1%	95.5%	92.2%	91.2%	91.8%	92.4%	89.8%	88.7%
Exp.2	89.0%	91.3%	87.3%	86.6%	87.9%	90.5%	86.0%	84.8%
action	stand2sit	sit2stand	stand2bow	bow2stand	stand2lie	lie2stand	sit2lie	lie2sit
Exp.1	89.7%	89.8%	89.0%	88.3%	92.4%	88.2%	91.2%	91.8%
Exp.2	84.7%	86.6%	84.8%	86.5%	88.7%	84.5%	86.6%	86.1%
action	wave hand	point	lower arm	lift arm	nod	shake head		
Exp.1	95.8%	94.2%	92.7%	92.3%	97.9%	96.7%		
Exp.2	91.3%	92.8%	89.2%	89.4%	95.1%	94.8%		

Hình 3. Kết quả nhận dạng bảy loại đặc tính do L. Fengjun và N.

RamaKant [5] đề xuất

c. Những vấn đề tồn tại

Việc xây dựng mô hình nhận dạng sử dụng dữ liệu chuyển động 3D vẫn còn nhiều điểm chưa tốt về hiệu năng cũng như chi phí.

- Dữ liệu chuyển động 3D là dữ liệu phức tạp, có số lượng thuộc tính lớn
- Chi phí tính toán lớn
- Hiệu năng không cao
- Tỷ lệ nhận dạng đúng thấp, đặc biệt với những hoạt động phức tạp

Do vậy giải pháp đề xuất là nghiên cứu, thực nghiệm và sánh các phương pháp trích chọn và lựa chọn đặc tính khác nhau, từ đó đưa ra phương pháp hiệu quả nhất cho bài toán nhận dạng hoạt động người, góp phần nâng cao hiệu năng cũng như độ chính xác của mô hình.

2. Mục tiêu và nhiệm vụ

Mục tiêu của đề tài là nghiên cứu đề xuất phương pháp trích chọn đặc tính tối ưu nhất cho mô hình nhận dạng hành động người dựa trên dữ liệu chuyển động 3D.

Các hoạt động cụ thể:

- Nghiên cứu hệ thống chụp chuyển động và phân tích dữ liệu chuyển động 3D
- Sử dụng các phương pháp trích chọn đặc tính (feature extraction) và lựa chọn đặc tính (feature selection) làm giảm độ phức tạp và độ lớn của dữ liệu
- Sử dụng mô hình SVM (Support Vector Machine) xây dựng mô hình nhận dạng hành động người dựa trên dữ liệu trích chọn được; trên cơ sở đó thực nghiệm và so sánh kết quả giữa các phương pháp nhằm đề xuất phương án tốt nhất.

3. Đối tượng và phạm vi nghiên cứu

a. Đối tượng nghiên cứu

- Cấu trúc dữ liệu chuyển động 3D (Acclaim)
- Các phương pháp trích chọn đặc tính (feature extraction) và lựa chọn đặc tính (feature selection)
- Mô hình SVM

b. Phạm vi nghiên cứu

- Dữ liệu đầu vào của đề tài này là dữ liệu trên cấu trúc Acclaim (asf/amc) được xây dựng và công bố bởi Đại học Carnegie Mellon (CMU) [8]
- Mô hình nhận dạng sử dụng trong đề tài này là mô hình SVM
- Phương pháp trích chọn đặc tính sử dụng trong đề tài này là: PCA, LDA

4. Phương pháp nghiên cứu

a. Nghiên cứu lý thuyết

- Cấu trúc Acclaim (asf/amc)
- Các phương pháp trích chọn đặc tính: PCA, LDA
- Mô hình SVM

b. Nghiên cứu thực nghiệm

- Thực nghiệm trên mô hình nhận dạng với nhiều phương pháp khác nhau và thu thập, phân tích kết quả
- So sánh, đánh giá kết quả

5. Ý nghĩa khoa học và thực tiễn của đề tài

a. Về mặt lý thuyết

- Lý thuyết về nhận dạng
- Lý thuyết về dữ liệu chuyển động 3D
- Các lý thuyết trích chọn đặc tính và lựa chọn đặc tính

b. Về mặt thực tiễn

- Góp phần hoàn thiện và nâng cao hiệu quả, độ chính xác của mô hình nhận dạng hành động người

- Xây dựng được ứng dụng nhận dạng hành động người với độ chính xác cao

6. Bố cục của luận văn

Ngoài phần mở đầu và kết luận, luận văn gồm có ba chương với các nội dung chính như sau:

Chương 1: Nghiên cứu tổng quan

Chương này trình bày tổng quan các vấn đề liên quan đến luận văn. Nội dung chủ yếu xoay quanh ba chủ đề chính. Thứ nhất là hoạt động nhận dạng hành động người. Thứ hai là mô hình chụp chuyển động. Cuối cùng là các phương pháp trích chọn và lựa chọn đặc tính.

Chương 2: Giải pháp đề xuất

Chương này tập trung vào trình bày quá trình phân tích và chuẩn hóa dữ liệu. Dựa trên các phương pháp trích chọn đặc tính và dữ liệu đã xử lý đề xuất, xây dựng mô hình nhận dạng hành động người sử dụng máy vectơ hỗ trợ.

Chương 3: Thực nghiệm và đánh giá kết quả

Dựa trên mô hình đề xuất ở chương hai, xây dựng chương trình thực nghiệm, sau đó phân tích và đánh giá kết quả.

CHƯƠNG 1

NGHIÊN CỨU TỔNG QUAN

I. Nhận dạng hành động người trong không gian 3D

Từ những năm 1980, nhận dạng hành động người luôn là lĩnh vực quan trọng trong các nghiên cứu về thị giác máy tính. Có rất nhiều hướng tiếp cận khác nhau đã được đề xuất và phần lớn trong số đó sử dụng hình ảnh từ các camera thông dụng. Các phương pháp thu thập dữ liệu chuyển động 3D đã ra đời và phát triển mạnh mẽ trong các thập niên gần đây. Phần này khái quát các kỹ thuật thu thập dữ liệu chuyển động 3D cũng như các phương pháp học máy phổ biến trong nhận dạng hành động người.

I.1. Các phương pháp thu thập dữ liệu chuyển động 3D

Từ sau sự ra đời của các thiết bị cảm biến chiều sâu (depth sensor), hoạt động nghiên cứu trên dữ liệu 3D đã có những chuyển biến tích cực. Nhìn chung có ba phương pháp chính trong việc thu thập dữ liệu chuyển động 3D. Đầu tiên là phương pháp sử dụng hệ thống chụp chuyển động dựa vào các marker như là Mocap (Motion Capture). Kế đến là phương pháp sử dụng stereo camera¹. Cuối cùng là phương pháp sử dụng range sensor². Mục này sẽ trình bày sơ lược từng phương pháp.

I.1.1. Phương pháp sử dụng stereo camera

Thu thập dữ liệu 3D từ stereo camera là một trong những nghiên cứu kinh điển trong lĩnh vực thị giác máy tính. Trước đây các thiết bị range sensor rất đắt đỏ và cồng kềnh, do đó stereo camera đã thu hút được sự quan tâm của các nhà nghiên cứu trong việc xây dựng các hệ thống thị giác. Một stereo camera được trang bị hai hay nhiều ống kính với cảm biến ảnh hoặc khung phim riêng biệt cho mỗi ống kính. Điều đó cho phép camera có thể mô phỏng thị giác của con người, đây là tiền đề cho khả năng xây dựng dữ liệu hình ảnh 3D.

¹ Stereo camera là camera có hai hay nhiều ống kính với cảm biến ảnh hoặc khung phim riêng biệt cho mỗi ống kính.

² Range sensor hay range camera là một loại cảm biến dùng để tạo ra một loại hình ảnh, trong đó giá trị của các điểm ảnh là khoảng cách của nó tới một điểm nhất định nào đó. Ảnh được tạo bởi range sensor gọi là range image.



Hình 4. Một stereo camera của Kodak

Stereo camera có vai trò đặc biệt quan trọng trong các lĩnh vực như người máy (robotics) và được ứng dụng rộng rãi trong giải trí, truyền thông và các hệ thống tự hành. Điển hình như kỹ thuật stereo trắc quang của TS. Maria Petrou [17]. Ông phát triển một hệ thống gồm một camera cố định và ba đèn chiếu sáng đối tượng từ các góc độ khác nhau. Tất cả dữ liệu được kết hợp thành một dạng dữ liệu 3D bằng cách phân tích các vùng tối sáng khác nhau. Nó đã được ứng dụng để tìm lỗi của các sản phẩm công nghiệp và mô phỏng mô hình 3D của khuôn mặt con người.



Hình 5. Sputnik stereo camera

Vì sự phức tạp của hình học, việc thu thập dữ liệu 3D từ stereo camera vẫn còn là một nhiệm vụ đầy thử thách. Với sự ra đời và phát triển mạnh mẽ của range sensor, hầu hết các thiết bị stereo camera đã không còn được sản xuất cho tới ngày nay.

1.1.2. Phương pháp sử dụng Mocap

Kỹ thuật thu thập dữ liệu 3D tiếp theo là sử dụng hệ thống chụp chuyển động – Mocap. Nó là một phương pháp quan trọng trong việc theo dõi và phân tích cấu trúc hình thể của con người. Mocap được sử dụng rộng rãi trong hoạt hình và trò chơi điện tử. Ngoài ra, người ta còn sử dụng Mocap để phân tích và hoàn thiện các động tác trong thể thao, khiêu vũ, cũng như giám sát tiến độ phục hồi trong vật lý trị liệu.



Hình 6. Một hệ thống chụp chuyển động

Có rất nhiều cách khác nhau để xây dựng một Mocap. Phổ biến nhất là sử dụng các marker cảm quang cố định trên chủ thể (thường là tại các khớp), đồng thời bố trí nhiều camera xung quanh để ghi nhận tọa độ và sự chuyển động của các marker khi chủ thể chuyển động. Ngoài ra, một số hệ thống sử dụng thẻ RFID (Radio Frequency Identification) hay các loại thẻ từ khác để thay thế cho marker. Chi tiết về hệ thống chụp chuyển động của CMU (Carnegie Mellon University) sẽ được trình bày trong phần II.

Hiện nay, có một số bộ dữ liệu chuyển động 3D được thu thập bởi hệ thống chụp chuyển động như CMU Motion Capture Database³, MPI HDM05 Motion Capture Database⁴, CMU Kitchen DataSet⁵, LACE Indoor Activity Benchmark Dataset⁶, và TUM Kitchen Dataset⁷.

³ <http://mocap.cs.cmu.edu/>.

⁴ <http://www.mpi-inf.mpg.de/resources/HDM05/>.

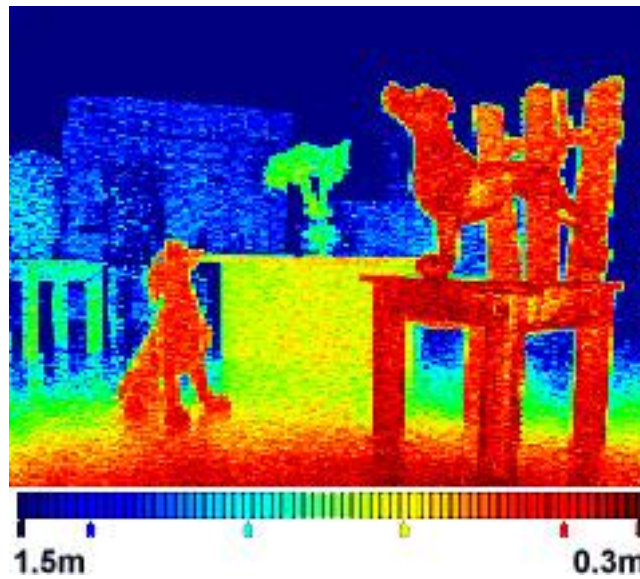
⁵ <http://kitchen.cs.cmu.edu/>.

⁶ <http://www.cs.rochester.edu/%18spark/muri/>.

⁷ <http://ias.in.tum.de/software/kitchen-activity-data/>.

1.1.3. Phương pháp sử dụng range sensor

Range sensor hay range camera là một loại cảm biến dùng để tạo ra range image. Range image là một dạng hình ảnh trong đó giá trị của các điểm ảnh là khoảng cách của nó tới một điểm cố định nào đó (cảm biến). Hình 7 là một ví dụ minh họa range image, những vùng càng gần với cảm biến được thể hiện bằng màu đỏ, vùng xa hơn là màu vàng và xa nhất là màu xanh. Dựa vào các khoảng cách này, chúng ta có thể dễ dàng dựng được cấu trúc ba chiều của vật thể.



Hình 7. Một ví dụ minh họa cho range image

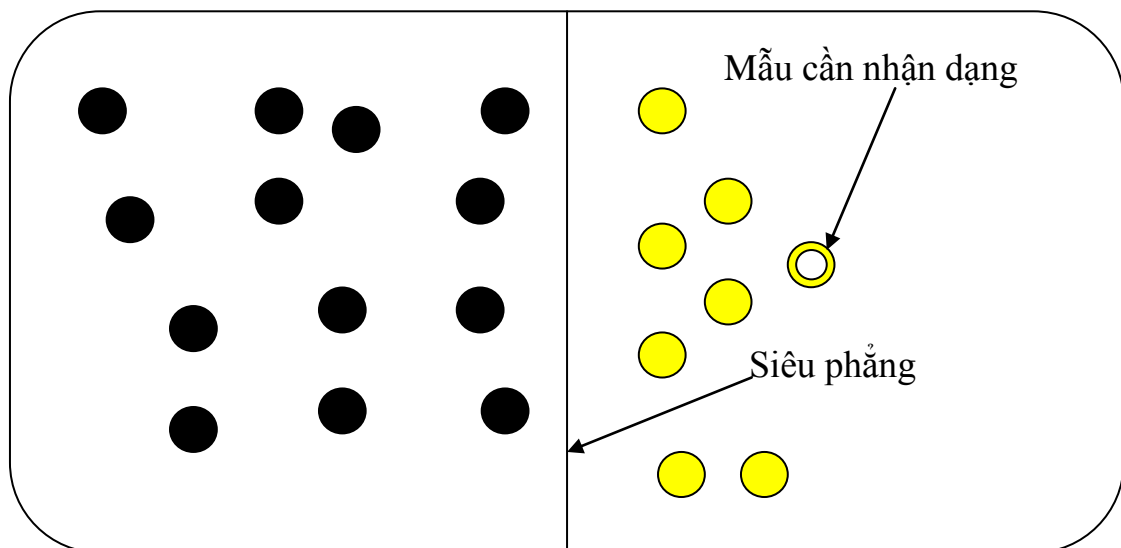
Ngày nay một số công ty sử dụng range sensor kết hợp với các kỹ thuật khác đã tạo ra những thiết bị có thể dựng mô hình khung xương của chủ thể trong thời gian thực. Điển hình là Microsoft Kinect.

I.2. Các phương pháp học máy thường sử dụng

Với tập hợp dữ liệu chuyển động 3D đã được gắn nhãn, phần lớn các nhà nghiên cứu áp dụng phương pháp học có giám sát để xây dựng mô hình nhận dạng. Các phương pháp thường dùng là máy vectơ hỗ trợ (SVM), mô hình Markov ẩn (HMM), kết hợp giữa SVM và HMM. Phần này sẽ giới thiệu hai phương pháp phổ biến nhất: SVM và HMM.

1.2.1. Máy vector hỗ trợ

Máy vector hỗ trợ (Support Vector Machine) là một phương pháp phân lớp dựa trên lý thuyết học thống kê và khoa học máy tính, gồm một tập hợp các phương pháp học máy liên quan đến nhau để phân lớp dữ liệu (classification) và phân tích hồi quy (regression analysis). SVM dạng chuẩn nhận dữ liệu vào và phân loại chúng vào hai lớp khác nhau. Do đó SVM là một thuật toán phân loại nhị phân. Với một bộ dữ liệu huấn luyện thuộc hai thể loại cho trước, thuật toán SVM xây dựng một mô hình SVM để phân loại các ví dụ khác vào hai thể loại đó. Một mô hình SVM là một cách biểu diễn các điểm trong không gian và lựa chọn ranh giới giữa hai thể loại sao cho khoảng cách từ các điểm tới ranh giới là xa nhất có thể. Các điểm mới cũng được biểu diễn trong cùng một không gian và được thuật toán dự đoán thuộc một trong hai thể loại tùy vào vị trí của nó nằm ở phía nào của ranh giới.

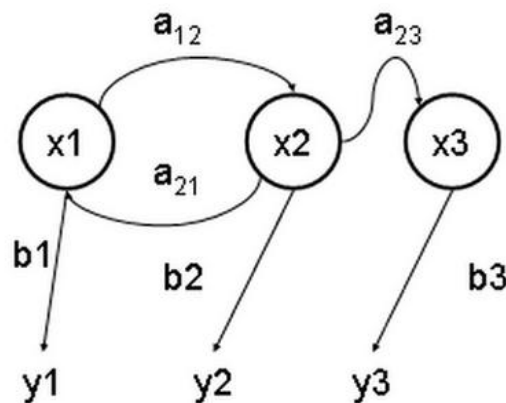


Hình 8. Siêu phẳng phân hoạch tập mẫu

Ý tưởng chính của SVM là tìm một siêu phẳng phân cách tối ưu để tách hai lớp, phân hoạch tập mẫu dựa vào các phân lớp của chúng sao cho khoảng cách giữa hai lớp đó đạt cực đại. Khoảng cách này được xác định bởi các vector hỗ trợ (SV – support vector). Các SV này được lọc ra từ tập mẫu huấn luyện bằng cách giải một bài toán tối ưu lồi, điều đó chính là tìm ra miền phân bố của từng lớp, để từ đó xác định được phân lớp của một mẫu cần nhận dạng.

1.2.2. Mô hình Markov ẩn

Mô hình Markov ẩn (Hidden Markov Model) là một mô hình xác suất hữu hạn trạng thái theo kiểu phát sinh tiến trình bằng cách định nghĩa xác suất liên kết trên các chuỗi quan sát. Mỗi chuỗi quan sát được sinh ra bởi một chuỗi các phép chuyển trạng thái, bắt đầu từ trạng thái khởi đầu cho đến khi thu được trạng thái kết thúc. Tại mỗi trạng thái mỗi phần tử của chuỗi quan sát được phát sinh ngẫu nhiên trước khi chuyển sang trạng thái tiếp theo. Hình 9 biểu diễn các chuyển tiếp trạng thái trong HMM. (x là các trạng thái trong mô hình, a là các xác suất chuyển tiếp, b là các xác suất đầu ra, y là dữ liệu quan sát).



Hình 9. Các chuyển tiếp trạng thái trong HMM

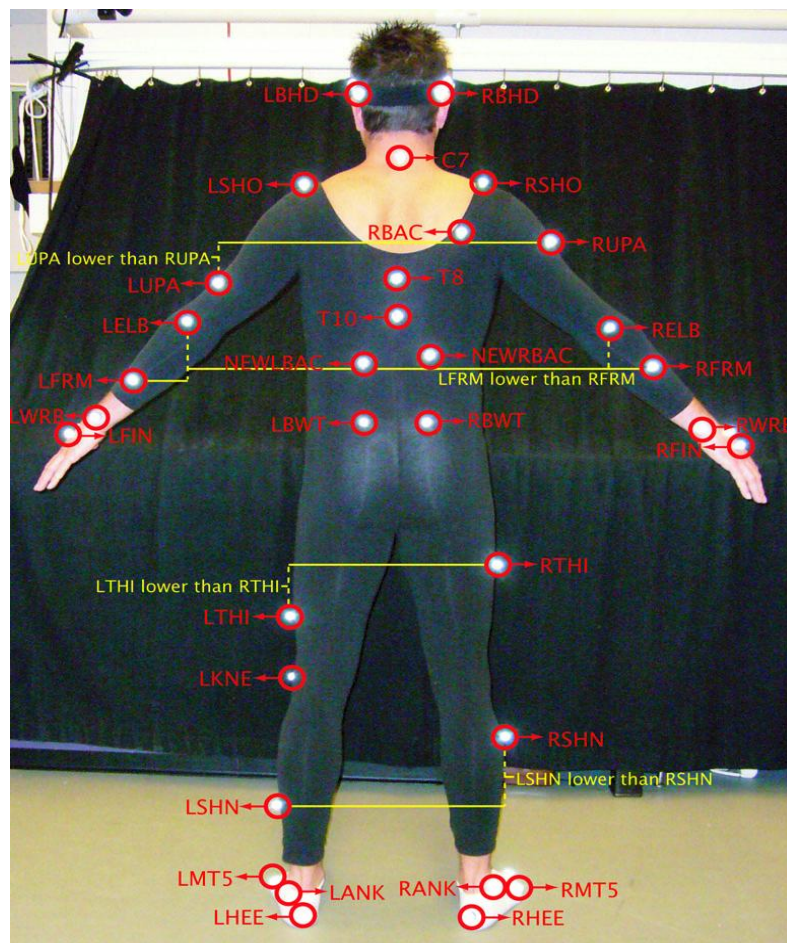
Các trạng thái của HMM được xem là ẩn bên trong mô hình vì tại mỗi thời điểm chỉ nhìn thấy các kí hiệu quan sát còn các trạng thái cũng như sự chuyển đổi trạng thái được vận hành ẩn bên trong mô hình.

II. Hệ thống chụp chuyển động – Mocap

Dữ liệu đầu vào trong quá trình thực nghiệm của luận văn là dữ liệu chuyển động 3D (ở định dạng Acclaim - asf/amc) được thu thập từ hệ thống chụp chuyển động (Mocap) của trường đại học CMU. Do đó phần này sẽ trình bày chi tiết về hệ thống chụp chuyển động của CMU cũng như dữ liệu thu được từ hệ thống này. Cuối cùng là chi tiết về cấu trúc dữ liệu Acclaim.

II.1. Mocap

Để theo dõi chuyển động của các đối tượng, các nhà nghiên cứu tại đại học Carnegie Mellon đã xây dựng hệ thống gồm 12 camera hồng ngoại MX-40⁸ lắp đặt xung quanh một không gian hình chữ nhật có kích thước 3m x 8m. Các chuyển động diễn ra trong vùng này sẽ được ghi lại. Đối với các chuyển động đòi hỏi sự tỉ mỉ và chi tiết như chuyển động của tay, các camera có thể di chuyển vào gần hơn. Người ta sử dụng một bộ quần áo đặc biệt có gắn 41 marker bên trên, chủ thể phải mặc bộ quần áo này và di chuyển trong vùng ghi nhận. Các camera sẽ nhận ra các marker bằng sóng hồng ngoại. Dữ liệu chuyển động thu được từ hệ thống camera được xử lý và cho ra kết quả cuối cùng là dữ liệu 3D. Có rất nhiều cấu trúc định dạng khác nhau để lưu trữ dữ liệu 3D như asf/amc, vsk/v, c3d, bvh, txt.



Hình 10. Vị trí các marker

⁸ MX-40 là sản phẩm của Vicon – một công ty chuyên cung cấp các thiết bị dùng trong hệ thống chụp chuyển động. MX-40 có thể quay với tốc độ 120Hz tức là ghi được 120 khung hình trong một giây.

II.2. Dữ liệu thu được từ Mocap

Dữ liệu thu được từ Mocap là dữ liệu 3D dưới dạng...

II.3. Cấu trúc Acclaim

Acclaim là một công ty game (2006 – 2010) có rất nhiều nghiên cứu trong lĩnh vực theo dõi chuyển động. Họ phát triển và sử dụng cấu trúc riêng trong việc lưu trữ dữ liệu chuyển động 3D. Đó là một bộ gồm hai tệp văn bản. Sau này, họ công bố và cho phép sử dụng rộng rãi cấu trúc đó trên toàn thế giới. Có rất nhiều công ty, tổ chức sử dụng cấu trúc Acclaim làm dữ liệu đầu ra cho các hệ thống theo dõi chuyển động của họ.

Acclaim là một bộ gồm hai tệp văn bản có cấu trúc. Tệp thứ nhất lưu trữ cấu trúc tổng quan của đối tượng, tệp thứ hai chứa dữ liệu chuyển động tương ứng của đối tượng theo thời gian. Nếu là chuyển động của con người thì tệp thứ nhất thể hiện một bộ xương được nối với nhau bởi các khớp, tệp thứ hai là vị trí của các khớp xương ứng với mỗi mốc thời gian. Tệp cấu trúc là ASF (Acclaim Skeleton File), tệp chuyển động là AMC (Acclaim Motion Capture).

II.3.1. Cấu trúc tệp ASF

Tệp cấu trúc ASF mô tả sự gắn kết giữa các xương trong cơ thể và bậc tự do (degrees of freedom) của các khớp. ASF chính là trạng thái ban đầu của dữ liệu chuyển động. Thành phần cụ thể trong ASF được mô tả như sau:

- Các chú thích được bắt đầu bởi dấu thăng (#).
- Các từ khóa bắt đầu bởi dấu hai chấm (:). Từ khóa có thể được dùng cho các giá trị toàn cục hoặc bắt đầu cho một mục dữ liệu.
- Từ khóa **:version** cho biết phiên bản hiện tại của tệp tin
- Từ khóa **:name** được dùng để đặt tên cho dữ liệu, tên này có thể khác với tên tệp.
- Mục **:units** định nghĩa đơn vị cho một số loại dữ liệu trong tệp. Nó cũng có thể chứa giá trị mặc định cho các đại lượng.
- Mục **:documentation** lưu trữ các thông tin thêm về dữ liệu

- Mục **:root** định nghĩa một khớp xương đặc biệt của ASF, đây chính là nút gốc trong hệ thống cây với nút là các khớp và nhánh là các đoạn xương giữa các khớp. Từ khóa **axis** trong mục **:root** cho biết thứ tự quay của các trục tọa độ X, Y, Z của nút gốc. Từ khóa **order** cho biết các kênh chuyển động của nút gốc cũng như thứ tự xuất hiện của các đại lượng này trong tệp AMC. Hai từ khóa còn lại **position** và **orientation** lưu giữ tọa độ và phương hướng ban đầu của nút gốc, thường thì giá trị của các đại lượng này bằng không (0).
- Mục **:bonedata** là nơi chứa thông tin chi tiết của mỗi đoạn xương trong cây hệ thống. Các thông tin của mỗi đoạn xương được đặt trong một cặp từ khóa **begin** và **end**. Với mỗi cặp, chúng ta có: **id**, **name**, **direction** là hướng của xương, **length** là độ dài, **axis** là góc quay của trục tọa độ tương đối của mỗi đoạn xương, **dof** là độ tự do của đoạn xương, **limits** là giới hạn quay của các khớp.
- Mục cuối cùng **:hierarchy** định nghĩa sự liên kết giữa các khớp xương để tạo nên một bộ xương hoàn chỉnh.

```

1  # AST/ASF file generated using VICON BodyLanguage
2  # -----
3  :version 1.10
4  :name VICON
5  :units
6    mass 1.0
7    length 0.45
8    angle deg
9  :documentation
10   .ast/.asf automatically generated from VICON data using
11   VICON BodyBuilder and BodyLanguage model FoxedUp or BRILLIANT.MOD
12  :root
13    order TX TY TZ RX RY RZ
14    axis XYZ
15    position 0 0 0
16    orientation 0 0 0
17  :bonedata
18    begin
19      id 1
20      name lhipjoint
21      direction 0.693937 -0.600361 0.397515
22      length 2.37164
23      axis 0 0 0 XYZ
24    end
25    begin
26      id 2
27      name lfemur
28      direction 0.34202 -0.939693 0
29      length 6.18497
30      axis 0 0 20 XYZ
31      dof rx ry rz
32      limits (-160.0 20.0)

```

mal text file length : 7264 lines : 339

Hình 11. Một đoạn tệp ASF

II.3.2. Cấu trúc tệp AMC

Tệp chuyển động AMC chứa các đại lượng có thể thay đổi trong hệ thống các khớp xương được định nghĩa ở tệp ASF. Là dữ liệu chuyển động theo thời gian nên trên tệp AMC được tạo thành bởi nhiều frame, mỗi frame thể hiện dữ liệu cho một thời điểm. Dữ liệu bên trong mỗi frame là giá trị theo thứ tự của đại lượng dof trên tệp ASF.

```

1  #!OML:ASF F:\VICON\USERDATA\INSTALL\rory3\rory3.ASF
2  :FULLY-SPECIFIED
3  :DEGREES
4  1
5  root -0.307087 17.6356 -28.2214 2.36076 1.44212 -4.54601
6  lowerback 15.4094 -0.182495 1.65268
7  upperback 1.54579 0.0318172 -0.110122
8  thorax -6.9977 -0.0335751 -1.06068
9  lowerneck -3.24163 -0.676991 -1.34632
10 upperneck -9.28199 -0.818331 1.08102
11 head -2.3551 -0.388697 0.578143
12 rclavicle 1.74931e-014 -4.77083e-015
13 rhumerus -42.2757 19.3184 -90.6312
14 rradius 79.2191
15 rwrist 2.46902
16 rhand -35.8906 32.487
17 rfingers 7.12502
18 rthumb -9.00425 2.69918
19 lclavicle 1.74931e-014 -4.77083e-015
20 lhumerus -46.581 -10.5126 91.072
21 lradius 108.082
22 lwrist 30.7395
23 lhand -39.5085 13.512
24 lfingers 7.12502
25 lthumb -12.4939 43.1185
26 rfemur 4.30283 -1.72433 25.7796
27 rtibia 82.7602
28 rfoot 27.83 -8.73877
29 rtoes 20.2614
30 lfemur -27.49 -2.09007 -20.1015
31 ltibia 38.398
32 lfoot -7.19848 -5.78026
33 ltoes 5.97973
34 2
35 root -0.303728 17.5624 -27.7253 2.02549 1.77071 -4.33872

```

Hình 12. Một đoạn tệp AMC

III. Trích chọn, lựa chọn đặc tính

Một khâu quan trọng trong quá trình xây dựng mô hình nhận dạng hành động người là trích chọn, lựa chọn đặc tính.có rất nhiều phương pháp, ...

III.1. Phương pháp phân tích thành phần chính – PCA

Phương pháp phân tích thành phần chính (Principal Components Analysis - PCA) là một thuật toán thống kê sử dụng phép biến đổi trực giao để biến đổi một tập hợp dữ liệu từ một không gian nhiều chiều sang một không gian mới ít chiều hơn nhằm tối ưu hóa việc thể hiện sự biến thiên của dữ liệu (maximize the variability).

Các đặc tính tốt của PCA:

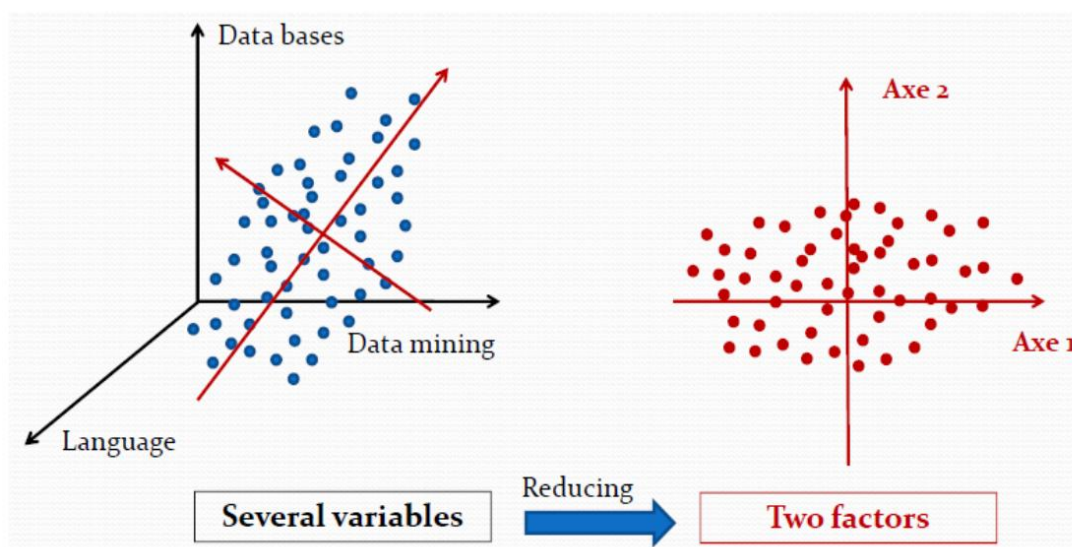
- Giúp giảm số chiều của dữ liệu
- Thay vì giữ lại các trục tọa độ của không gian cũ, PCA xây dựng một không gian mới ít chiều hơn, nhưng lại có khả năng biểu diễn dữ liệu tốt tương đương không gian cũ, nghĩa là đảm bảo độ biến thiên (variability) của dữ liệu trên mỗi chiều mới.
- Các trục tọa độ trong không gian mới là tổ hợp tuyến tính của không gian cũ, do đó về mặt ngữ nghĩa, PCA xây dựng các thuộc tính mới dựa trên các thuộc tính hiện có. Điểm hay là những thuộc tính này vẫn biểu diễn tốt dữ liệu ban đầu.
- Trong không gian mới, các liên kết tiềm ẩn của dữ liệu có thể được khám phá, mà nếu đặt trong không gian cũ thì khó phát hiện hơn, hoặc những liên kết như thế không thể hiện rõ.



Hình 13. Minh họa PCA. Phép chiếu lên các trục tọa độ khác nhau có thể cho cách nhìn rất khác nhau về cùng một dữ liệu

Một ví dụ kinh điển để minh họa PCA là hình ảnh về con lạc đà. Cùng là một con lạc đà nhưng nếu nhìn từ bên hông thì ta có được đầy đủ thông tin nhất, trong khi nhìn từ phía trước thì thật khó để nói nó là con lạc đà.

Một ví dụ thuyết phục hơn được minh họa như hình sau.



Hình 14. Minh họa PCA

Giả sử tập dữ liệu ban đầu (tập điểm màu xanh) được quan sát trong không gian ba chiều (trục màu đen) như hình bên trái. Rõ ràng ba trục này không biểu diễn được tốt nhất mức độ biến thiên của dữ liệu. PCA do đó sẽ tìm hệ trục tọa độ mới (là hệ trục màu đỏ trong hình bên trái). Sau khi tìm được không gian mới, dữ liệu sẽ được chuyển qua không gian này để được biểu diễn như trong hình bên phải. Rõ ràng hình bên trái

chỉ cần hai trục tọa độ nhưng biểu diễn tốt hơn độ biến thiên của dữ liệu so với hệ trục ba chiều ban đầu.

Một điểm rất hay nữa của PCA là các trục tọa độ trong không gian mới luôn đảm bảo trực giao đôi một với nhau, mặt dù trong không gian ban đầu, các trục có thể không trực giao.

Thuật toán PCA:

Về cơ bản, thuật toán PCA gồm có ba bước: tiền xử lí, xây dựng không gian mới, chuyển dữ liệu từ không gian ban đầu sang không gian mới. Cho ma trận $\mathbf{X} = \{x_{ij}\} \in \mathbf{R}^{n \times p}$. Các bước của PCA lần lượt như sau:

Bước 1: Tiền xử lí

Dữ liệu ban đầu có thể có giá trị thay đổi bất thường. Ví dụ một thuộc tính (feature) có giá trị thay đổi trong khoảng (0, 1) nhưng trên thuộc tính khác lại biến thiên trong đoạn (-100, 100). Rõ ràng cần phải có một bước tiền xử lí để chuẩn hóa các giá trị trên các cột của ma trận X. Có hai cách tiền xử lí thường được dùng cho PCA là Centered PCA và Normed PCA.

Centered PCA mang tất cả các thuộc tính (các cột của X) về cùng một gốc tọa độ:

$$\begin{aligned}\widehat{\mathbf{X}} &= \{\widehat{x}_{ij}\}, \\ \widehat{x}_{ij} &= \frac{x_{ij} - g_j}{\sqrt{n}} \quad (1a)\end{aligned}$$

Trong đó n là số dòng của X, g_j là mean của cột thứ j của X, được tính như sau:

$$g_j = \frac{\sum_{i=1}^n x_{ij}}{n}.$$

Normed PCA mang tất cả các feature về cùng một gốc tọa độ, đồng thời chuẩn hóa về cùng một quãng có độ lệch chuẩn (standard deviation) bằng 1:

$$\begin{aligned}\widehat{\mathbf{X}} &= \{\widehat{x}_{ij}\}, \\ \widehat{x}_{ij} &= \frac{x_{ij} - g_j}{\sqrt{n\sigma_j}} \quad (1b)\end{aligned}$$

Trong đó σ_j là độ lệch chuẩn của cột thứ j trong X.

Thông thường, Normed PCA hay được dùng, sau bước tiền xử lí, ma trận $\widehat{\mathbf{X}}$ sẽ là đầu vào cho bước tiếp theo.

Bước 2: Xây dựng không gian mới

Tính ma trận hiệp phương sai (covariance) của các thuộc tính trong $\hat{\mathbf{X}}$:

$$\mathbf{V} = \hat{\mathbf{X}}^T \hat{\mathbf{X}} \quad (2)$$

Do là tích của ma trận $\hat{\mathbf{X}}$ với một chuyển vị của nó nên $\mathbf{V} \in \mathbf{R}^{p \times p}$ là ma trận positive semidefinite kích thước $p \times p$. Hơn nữa \mathbf{V} có p trị riêng $\lambda_i \geq 0, i = 1..p$.

Tiếp theo, PCA tìm trị riêng và vector riêng tương ứng của \mathbf{V} , sắp xếp theo thứ tự giảm dần của trị riêng. Giả sử p trị riêng của \mathbf{V} là $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p$ (3), và vector riêng tương ứng là $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ (4). Khi đó các trục của không gian mới chính là các vector riêng \mathbf{u}_i ở trên, đương nhiên các vector riêng hoàn toàn độc lập tuyến tính (nghĩa là trục giao đôi một).

Bước 3: Chuyển dữ liệu từ không gian ban đầu sang không gian mới

Thông thường không gian mới không được xây dựng bằng tất cả các p vector riêng trong (4), mà thông thường chỉ từ k vector riêng đầu tiên.

Như vậy gọi $\mathbf{U} = [\mathbf{u}_1 | \mathbf{u}_2 | \dots | \mathbf{u}_k] \in \mathbf{R}^{p \times k}$.

Khi đó tọa độ các điểm trong hệ tọa độ mới là

$$\mathbf{F} = \hat{\mathbf{X}}\mathbf{U} \quad (5)$$

III.2. Phương pháp phân tích biệt thức tuyến tính – LDA

III.3. Các phương pháp lựa chọn đặc tính

CHƯƠNG 2

GIẢI PHÁP ĐỀ XUẤT

xxx

I. Mô hình tổng quan

II. Trích chọn đặc tính

III. Giải thuật huấn luyện

- Xxx

III.1. Tiêu đề mục con

xxx

III.1.1. Tiêu đề mục con nhỏ hơn

- xxx

- xxx

Chú ý sau mỗi chương cần thêm dấu ngắt chương (Section) :

Insert-Break... Section Break Pages = Next Page

CHƯƠNG 3

TRIỂN KHAI VÀ ĐÁNH GIÁ KẾT QUẢ

xxx

I. Dữ liệu nhận dạng**II. Môi trường triển khai****III. Kết quả thực nghiệm****IV. Đánh giá**

- Xxx

IV.1. Tiêu đề mục con

xxx

IV.1.1. Tiêu đề mục con nhỏ hơn

- xxx

- xxx

KẾT LUẬN

XXX

XXX

PHỤ LỤC

Tất cả các phụ lục nằm ở đây và được đánh số thứ tự cùng tên phụ lục :

xxx

PHỤ LỤC A

xxx

TÀI LIỆU THAM KHẢO

- [1] J.K Aggarwal, Lu Xia (2014), “Human Activity Recognition from 3D Data-A Review”, *Pattern Recognition Letters, Elsevier B.V, USA*.
- [2] Aggarwal J.K, Ryoo M.S (2011), “Human Activity Analysis: A Review”, *ACM Comput. Surv*, page 16.
- [3] Kohei Arai, Rosa Andrie Asmara (2013), “3D Skeleton model derived from Kinect Depth Sensor Camera and its application to walking style quality evaluations”, *IJARAL – International Journal of Advanced in Artificial Intelligence*.
- [4] Turaga P, Chellappa R, Subrahmanian V.S, Udrea O (2008), “Machine Recognition of Human Activities: A survey”, *Circuits Syst. Video Technol. IEEE Trans 18*, pages 1473-1488.
- [5] Fengjun Lv, Ramakant Nevatia (2006), “Recognition and Segmentation of 3D Human Action Using HMM and Multi-class AdaBoost”, *Lecture Notes in Computer Science Volume 3954, 2006, pp 359-372*.
- [6] Rizwan Chaudhry, Ferda Ofli, Gregorij Kurillo, Ruzena Bajcsy, René Vidal (2013), “Bio-inspired Dynamic 3D Discriminative Skeletal Features for Human Action Recognition”, *CVPR-2013*.
- [7] Raviteja Vemulapalli, Felipe Arrate, Rama Chellappa (2014), “Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group”, *CVPR-2014*.
- [8] <http://mocap.cs.cmu.edu/>
- [9] Vasileios Argyriou, Maria Petrou, Svetlana Barsky (2010), “Photometric Stereo with an Arbitrary Number of Illuminants”, *Computer Vision and Image Understanding 114*, pages 887-900.
- [10] D. Gehrig, T. Schultz (2008), “Selecting Relevant Features for Human Motion Recognition”, *ICPR 2008, IEEE*.
- [11] Dana Kulic, Wataru Takano, Yoshihiko Nakamura (2008), “Incremental Learning, Clustering and Hierarchy Formation of Whole Body Motion Patterns using Adaptive Hidden Markov Chains”, *The International Journal of Robotics Research vol. 27 no. 7 761-784*.

- [12] Gita Sukthankar, Katia Sycara (2005), “A Cost Minimization Approach to Human Behavior Recognition”.
- [13] Ahmad Jalal, Shaharyar Kamal, Daijin Kim (2014), “A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments”, *Sensors-2014*.
- [14] Lasitha Piyathilaka, Sarah Kodagoda (2013), “Human Activity Recognition for Domestic Robots”.
- [15] Mi Zhang, Alexander A. Sawchuk (2012), “Motion Primitive-Based Human Activity Recognition Using a Bag-of-Features Approach”.
- [16] Md. Zia Uddin, Nguyen Duc Thang, Jeong Tai Kim, Tae-Seong Kim (2011), “Human Activity Recognition Using Body Joint-Angle Features and Hidden Markov Model”, *ETRI Journal, Volume 33*.
- [17] V. Argyriou, M. Petrou, S. Barsky (2010), “Photometric stereo with an arbitrary number of illuminants”, *CVIU 114* 887–900.

TÓM TẮT LUẬN VĂN THẠC SĨ

Phần này viết tên đề tài (chữ nhỏ 13pt) và :
tóm tắt nội dung LUẬN VĂN, viết ngắn gọn và rõ ràng
(15 đến 20 dòng) cho biết :

Đề tài đã được đặt ra như thế nào (bối cảnh), mục đích (giải thích rõ hơn tên đề tài) và nhiệm vụ phải thực hiện (các mục tiêu cụ thể và kết quả cần có)

HV đã giải quyết vấn đề gì (đã nghiên cứu lý thuyết, thực tiễn như thế nào, đã đề xuất được những giải pháp (biện pháp) hay sáng kiến gì ?).

HV đã giải quyết đến đâu (nêu một số kết quả tiêu biểu).