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 :*

* 1. *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.*
  2. *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.*
  3. *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 NHO DÃI**

**MỤC LỤC**

MỞ ĐẦU 1

NGHIÊN CỨU TỔNG QUAN 9

1. Nhận dạng hành động người trong không gian 3D 9

1.1. Các phương pháp thu thập dữ liệu chuyển động 3D 9

1.1.1. Phương pháp sử dụng stereo camera 9

1.1.2. Phương pháp sử dụng Mocap 11

1.1.3. Phương pháp sử dụng range sensor 12

1.2. Các phương pháp học máy thường sử dụng 13

1.2.1. Máy vectơ hỗ trợ 13

1.2.2. Mô hình Markov ẩn 14

2. Hệ thống chụp chuyển động – Mocap 15

2.1. Mocap 15

2.2. Dữ liệu thu được từ Mocap 17

2.3. Cấu trúc Acclaim 17

2.3.1. Cấu trúc tệp ASF 18

2.3.2. Cấu trúc tệp AMC 20

3. Trích chọn, lựa chọn đặc tính 21

3.1. Phương pháp phân tích thành phần chính – PCA 22

3.2. Phương pháp phân tích biệt thức tuyến tính – LDA 26

3.3. Sequential Forward Selection - SFS 26

GIẢI PHÁP ĐỀ XUẤT 27

1. Mô hình tổng quan 27

2. Trích chọn đặc tính 27

3. Giải thuật huấn luyện 27

3.1. Tiêu đề mục con 27

3.1.1. Tiêu đề mục con nhỏ hơn 27

TRIỂN KHAI VÀ ĐÁNH GIÁ KẾT QUẢ 29

1. Dữ liệu nhận dạng 29

2. Môi trường triển khai 29

3. Kết quả thực nghiệm 29

4. Đánh giá 29

4.1. Tiêu đề mục con 29

4.1.1. Tiêu đề mục con nhỏ hơn 29

KẾT LUẬN 31

PHỤ LỤC ii

TÀI LIỆU THAM KHẢO v

DANH MỤC CÁC CHŨ 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** ii](#_Toc271487141)

# 

# MỞ ĐẦU

1. **Lý do chọn đề tài**

**a. Bối cảnh chung**

Từ những năm 80 của thế kỷ trước, nhận dạng hành động người (human activity recognition) 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).

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 lại cảm biến, việc theo dõi và nhận dạng hoạt động của con người sẽ trở nên dễ dàng hơn.

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.

Gần đây, một số nhà nghiên cứu đã sử dụng camera RGBD (Red, Green, Blue, Depth) như Kinect[[1]](#footnote-2) 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.



**Microsoft Kinect Camera**

**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. 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), đã có một hướng tiếp cận mới trong nhận dạng hành động người, đó là sử dụng dữ liệu chuyển động 3D. Trong 20 năm trở lại, một số phương pháp chính để thu thập dữ liệu 3D hoặc là sử hệ thống chụp chuyển động dựa vào marker[[2]](#footnote-3) như là Mocap[[3]](#footnote-4) (Motion Capture) hoặc là dùng sóng âm (stereo) - chụp hình ảnh 2D từ nhiều hướng khác nhau để dựng thành mô hình 3D.



**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 đã nghiên cứu, thực nghiệm trên 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]. L. Fengjun 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]. K. Dana đã phát triển một mô hình học tăng cường mới dựa trên mô hình Markov ẩn [11].



**Kết quả nhận dùng bảy loại đặc tính do L. Fengjun [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. Các hạn chế đó là: dữ liệu chuyển động 3D là dữ liệu phức tạp, có số lượng thuộc tính lớn dẫn đến chi phí tính toán lớn do đó hiệu năng sẽ không cao và tỉ lệ nhận dạng đúng thấp, đặc biệt với những hoạt động phức tạp.

1. **Mục tiêu và nhiệm vụ**

Trước những vấn đề tồn tại phân tích ở trên, luận văn này sẽ nghiên cứu, xây dựng mô hình nhận dạng hành động người từ dữ liệu chuyển động 3D; trong đó trọng tâm là các phương pháp trích chọn và lựa chọn đặc tính nhằm làm giảm số chiều và độ lớn của dữ liệu, góp phần nâng cao độ chính xác và hiệu năng của mô hình.

1. **Đối tượng và phạm vi nghiên cứu**

**a. Đối tượng nghiên cứu**

Đối tượng nghiên cứu trong luận văn này là mô hình nhận dạng hành động người, cụ thể hơn là dữ liệu chuyển động 3D trong định dạng Acclaim[[4]](#footnote-5) (asf/amc) do trường đại học CMU (Carnegie Mellon University) thu thập bằng Mocap của họ; và các phương pháp trích chọn, lựa chọn đặc tính phù hợp.

**b. Phạm vi nghiên cứu**

Bộ dữ liệu chuyển động 3D do CMU cung cấp có rất nhiều hoạt động khác nhau, luận văn này chỉ nghiên cứu trên một số loại hoạt động đơn giản như đi bộ (walk), chạy (run, jog), nhảy (jump) và khiêu vũ (dance). Ngoài ra, luận văn chỉ tập trung nghiên cứu một số phương pháp trích chọn đặc tính thông dụng như PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis); và dùng máy vectơ hỗ trợ (SVM) trong học máy.

1. **Phương pháp nghiên cứu**

**a. Nghiên cứu lý thuyết**

Về phần lý thuyết, luận văn tập trung nghiên cứu tổng quan về nhận dạng hoạt động người trong không gian 3D, trong đó tập trung vào kỹ thuật thu thập dữ liệu chuyển động 3D bằng phương pháp sử dụng hệ thống chụp chuyển động; máy vectơ hỗ trợ (SVM) và các phương pháp trích chọn, lựa chọn đặc tính.

**b. Nghiên cứu thực nghiệm**

Quá trình nghiên cứu thực nghiệm sử dụng ngôn ngữ lập trình c# trên nền tảng của thư viện học máy mã nguồn mở Accord.NET Framework[[5]](#footnote-6) xây dựng mô hình nhận dạng với các phương pháp trích chọn đặc tính khác nhau so sánh kết quả thu được và đưa ra kết luận.

1. **Ý nghĩa khoa học và thực tiễn của đề tài**

**a. Về mặt lý thuyêt**

Luận văn này củng cố các lý thuyết về nhận dạng hành động người trong không gian 3D, dữ liệu chuyển động 3D, các phương pháp trích chọn, lựa chọn đặc tính.

**b. Về mặt thực tiễn**

Xây dựng mô hình nhận dạng hành động người trong không gian 3D sử dụng các phương pháp trích chọn đặc tính và học máy.

1. **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 đề tài của luận văn. Nội dung chủ yếu xoay quanh các chủ đề chính. Hoạt động nhận dạng hành động người; mô hình chụp chuyển động; các phương pháp học máy; các phương pháp trích chọn và lựa chọn đặc tính.

**Chương 2: Giải pháp đề xuất**

Chương này tập trung vào trình bày và giải thích chi tết mô hình nhận dạng đề xuất và các thành phần trong mô hình.

**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

## 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.

### 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 đông 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 camera[[6]](#footnote-7). Cuối cùng là phương pháp sử dụng range sensor[[7]](#footnote-8). Mục này sẽ trình bày sơ lược từng phương pháp.

#### 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.



**Hình 4. Một stereo camera của Kodak**

Stereo cameara 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 triên mạnh mẽ của range sensor, hầu hết các thiết bị stereo camera đã không còn được sản xuất cho tới ngày nay.

#### 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, củng 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 Database[[8]](#footnote-9), MPI HDM05 Motion Capture Database2[[9]](#footnote-10), CMU Kitchen DataSet[[10]](#footnote-11), LACE Indoor Activity Benchmark Dataset[[11]](#footnote-12), và TUM Kitchen Dataset[[12]](#footnote-13).

#### Phương pháp sử dụng range sensor

Range sensor hay range camera là một loại cảm biến dùng để tạo ra range image. Range image làm một dạng hình ảnh trong đó giá trị của các điểm ảnh là khoản 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 khoản 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.

### 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.

#### Máy vectơ hỗ trợ

Máy vectơ hỗ trợ (Support Vector Machine – SVM) làm một giải thuật học máy dựa trên lý thuyết học thống kê. Bài toán cơ bản của SVM là bài toán phân loại hai lớp: Cho trước n điểm trong không gian d chiều (mỗi điểm thuộc vào một lớp kí hiệu là +1 hoặc -1, mục đích của giải thuật SVM là tìm một siểu phẳng (hyperplane) phân hoạch tối ưu cho phép chia các điểm này thành hai phần sao cho các điểm cùng một lớp nằm về một phía với siêu phẳng này.

Xét tập dữ liệu mẫu có thể tách rời tuyến tính với và . Siêu phẳng tối ưu phân tập dữ liệu này thành hai lớp là siêu phẳng có thể tách rời dữ liệu thành hai lớp riêng biệt với lề (margin) lớn nhất. Tức là, cần tìm siêu phẳng và hai siêu phẳng H1, H2 hỗ trợ song song với H và có cùng khoảng cách đến H. Với điều kiện không có phần tử nào của tập mẫu nằm giữa H1 và H2, khi đó:

với y = 1

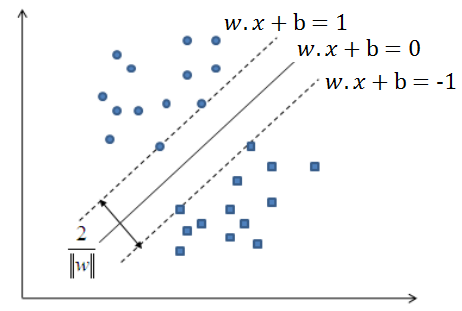
với y = -1

Kết hợp hai điều kiện trên ta có .

Khoảng cách của siêu phẳng H1 và H2 đến H là . Ta cần tìm siêu phẳng H với lề lớn nhất, tức là giải bài toán tối ưu tìm với ràng buộc . Bài toán này có thể chuyển sang bài toán tương đương dễ giải hơn là với ràng buộc . Lời giải cho bài toán tối ưu này là cực tiểu hóa hàm Lagrange:

Trong đó là các hệ số Lagrange, . Sau đó người ta chuyển thành bài toán đối ngẫu là cực đại hóa hàm :

Từ đó giải để tìm được các giá trị tối ưu cho w, b và . Về sau, việc phân loại một mẫu mới chỉ là việc kiểm tra hàm dấu sign(w.x + b). Hình xxx là một minh họa siêu phẳng với lề cực đại trong không gian hai chiều.



Siêu phẳng với lề cực đại trong không gian hai chiều

#### 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. XXX biểu diễn các chuyển tiếp trạng thái trong HMM. Trong đó x1, x2, x3 là các trạng thái trong mô hình; là xác suất chuyển tiếp từ trạng thái *i* sang trạng thái *j*; b1, b2, b3 là các xác suất đầu ra, y1, y2, y3 là dữ liệu quan sát.



**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.

## 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. 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.

### Mocap

Để theo dõi chuyển động của các đối tượng, các nhà nghiên cứu tại đại học Carnigie Mellon đã xây dựng hệ thống gồm 12 camera hồng ngoại MX-40[[13]](#footnote-14) 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 nhu 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. Vị trí các marker được mô tả như hình xxx. Các camera sẽ định vị marker bằng sóng hồng ngoại. Tín hiệu thu được từ hệ thống camera được xử lý và cho ra kết quả cuối cùng là dữ liệu dạng mô hình hóa 3D của cơ thể người. Có rất nhiều cấu trúc định dạng khác nhau được sử dụng như asf/amc, vsk/v, c3d, bvh, txt. Có một số phần mềm hỗ trợ việc chuyển đổi qua lại giữa các định dạng này[[14]](#footnote-15).



**Vị trí của một số marker**

### Dữ liệu thu được từ Mocap

Dữ liệu thu được từ Mocap là dữ liệu dưới dạng mô hình hóa 3D của cơ thể người. Có rất nhiều định dạng khác nhau được dùng để lưu trữ loại dữ liệu này như asf/amc, bvh, c3d, vsk/v, txt. Trong quá trình thu nhận tín hiệu, các marker được định vị ở những vị trí định sẵn cho trước. Tín hiệu thu được từ các camera sẽ được phần mềm của Vincon (Vicon Bodybuilder) tổng hợp và mô hình hóa thành dữ liệu 3D.

Dưới góc độ người dùng có hai loại dữ liệu chính:

Thứ nhất là dạng dữ liệu nhị phân chứa vị trí của các marker trong không gian 3D với định dạng tệp là c3d. Định dạng này chỉ mô tả vị trí 3D của các marker trong quá trình chuyển động của đối tượng, nhưng lại không cho biết mối liên hệ giữa các marker với nhau.

Dạng dữ liệu thứ hai chi tiết hơn, thường là một cặp gồm hai loại tệp văn bản (text) ở định dạng *.asf/.amc* hoặc *.vsk/.v*. Tệp thứ nhất chứa các thông tin về chi tiết về mô hình 3D của khung xương và các khớp nối; trong đó bao gồm các kết nối, vị trí, độ dài và hướng của các đoạn xương và độ tự do của các khớp. Tệp còn lại chứa thông tin chuyển động của mỗi khớp. Nếu hành động của cùng một đối tượng được ghi nhận bằng nhiều clip thì sẽ có nhiều tệp *.amc* hoặc *.v* tương ứng với một tệp *.asf* hoặc *.vsk*. Nhìn chung dữ liệu dưới dạng văn bản (text) dễ đọc và dễ sử dụng hơn so với dữ liệu nhị phân. Luận văn này chọn dữ liệu dưới dạng asf/amc làm dữ liệu đầu vào để nghiên cứu. Phần tiếp theo sẽ trình bày chi tiết cấu trúc loại dữ liệu này.

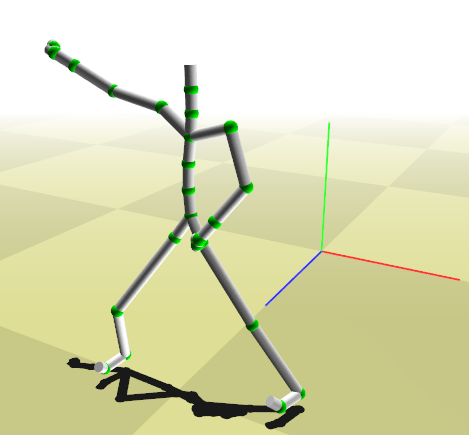
### Cấu trúc Acclaim

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 thư nhất lưu trữ cấu trúc tổng quan của đối tượng, tiệp thứ hai chứa dư 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).

#### Cấu trúc tệp ASF

Tệp cấu trúc ASF mô tả sự gắn kết giữa các xương trong cơ thể và độ tự do (degrees of freedom) của các khớp. ASF chính là trạng thái ban đầu của dữ liệu chuyển động. Hình XXX mô tả mô hình 3D của bộ xương người được dựng lại bằng tệp *.asf* (nguồn http://run.usc.edu/).



Mô hình 3D của xương người được dựng lại từ tệp .asf

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 củng có thể chứa giá trị mặc định cho các đại lượng.
* Mục **:documentation** lưu trữ các thông 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 tù 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.



**Một đoạn tệp ASF**

#### Cấu trúc tệp AMC

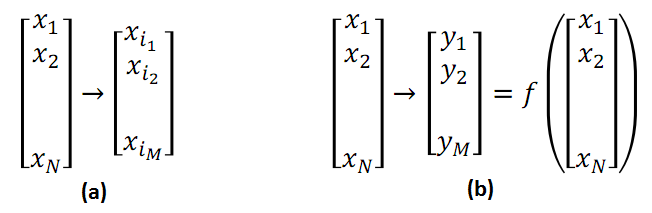
Tệp chuyển động AMC chứa các đại lượng có thể thay đổi trong hệ thống các khớp xương được định nghĩa ở tệp ASF. Là dữ liệu chuyển động theo thời gian nên trên tệp AMC được tạo thành bởi nhiều frame, mỗi frame thể hiện dữ liệu cho một thời điểm. Dữ liệu bên trong mỗi frame là giá trị theo thứ tự của đại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



**Một đoạn tệp AMC**

## 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, lực chọn đặc tính. Mục đích chung của trích chọn hay lựa chọn đặc tích là làm giảm độ lớn của dữ liệu, hay nói cách khác là làm giảm số chiều của dữ liệu. Quá trình làm giảm số chiều của dữ liệu được chia làm hai loại chính: trích chọn đặc tính (feature extraction - FA) và lựa chọn đặc tính (feature selection - FS). Sự khác nhau cơ bản giữa hai phương pháp này là trích chọn đặc tích biến đổi dữ liệu từ không gian cũ sang không gian mới ít chiều hớn, trong khi đó lựa chọn đặc tính chọn một tập con của tập dữ liệu cho trước. Hình xxx mô tả sự khác nhau giữa hai phương pháp này. Kết quả của lựa chọn đặc tính (a) là một tập con của dữ liệu ban đầu, kết quả của trích chọn đặc tính (b) là một tập dữ liệu trong không gian mới ít chiều hơn .



So sánh giữa lựa chon đặc tính (a) và trích chọn đặc tính (b)

Đối với bài toán nhận dạng, mục đích của hai phương pháp này là giống nhau nên từ đây, trong luận văn này xin dùng “trích chọn đặc tính” làm tên gọi chung cho cả hai phương pháp.

Có rất nhiều phương pháp trích chọn đặc tính khác nhau được đề xuất sử dụng trong nhận dạng hành động. Luận văn này chọn một số phương pháp tiêu biểu để nghiên cứu như: phân tích thành phân chính (PCA), biệt thức tuyến tính (LDA), sequential forward selection (SFS). Nội dung chính của từng phương pháp như sau:

### 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.

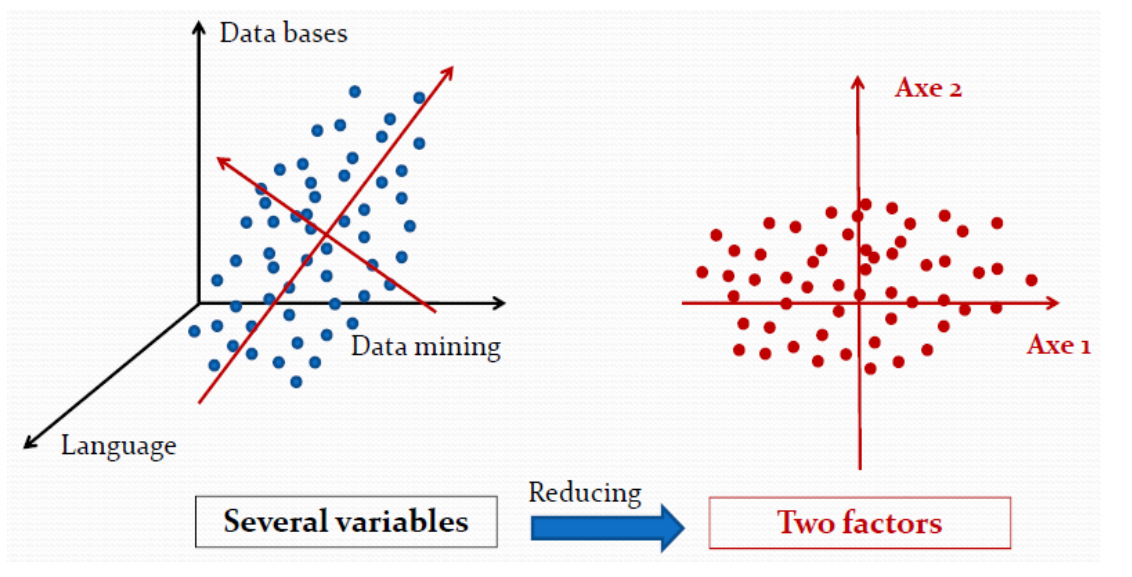
Trong không gian mới, các liên kết tiềm ẩn của dữ liệu có thể được khám phá, mà nếu đặt trong không gian cũ thì khó phát hiện hơn, hoặc những liên kết như thế không thể hiện rõ.



**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 xxx



**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 . 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 gián 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:

,

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

**Normed PCA** mang tất các các thuộc tính về cùng một gốc tọa độ, đồng thời chuẩn hóa về cùng một quãng có độ lệch chuẩn (standard deviation) là 1:

,

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

Thông thường, Normed PCA hay được dùng, sau bước tiền xử lí, ma trận 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 :

Do là tích của ma trận với một chuyển vị của nó nên là ma trận positive semidefinite kích thước . Hơn nữa có p trị riêng .

Tiếp theo, PCA tìm trị riêng và vector riêng tương ứng của V, sắp xếp theo thứ tự giảm dần của trị riêng. Giả sử p trị riêng của **V** là , và vector riêng tương ứng là , …, . Khi đó các trục của không gian mới chính là các vector riêng ở trên, đương nhiên các vector riêng hoàn toàn độc lập tuyến tính (nghĩa là trực 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 **H**, mà thông thường chỉ từ *k* vector riêng đầu tiên.

Như vậy gọi  **|**

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

### Phương pháp phân tích biệt thức tuyến tính – LDA

Phân tích biệt thức tuyến tính (Linear Discriminant Analysis – LDA)

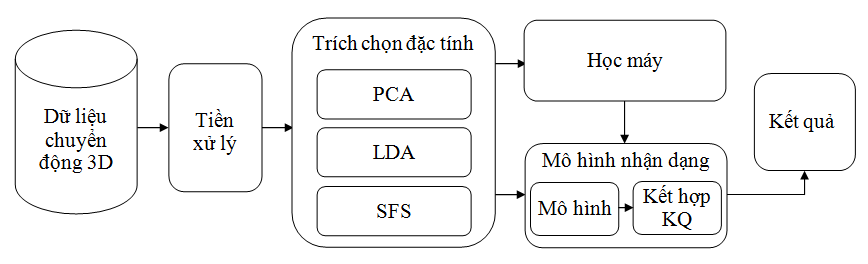
### Sequential Forward Selection - SFS

CHƯƠNG 2

# GIẢI PHÁP ĐỀ XUẤT

Giải pháp đề xuất được trình bày trong luận văn này sử dụng kết hợp hai loại mô hình: mô hình chức năng và mô hình luồng dữ liệu. Mô hình chức năng mô tả các khối chức năng chính của hệ thống và mối liên hệ giữa chúng, trong khi đó mô hình dữ liệu mô tả đường đi của dữ liệu, các trạng thái dữ liệu trước và sau các khối chức năng. Dữ liệu đầu vào của hệ thống được chọn lựa trong cở sở dữ liệu của CMU Mocap và được phân chia ngẫu nhiên thành ba loại: loại thứ nhất được dùng làm dữ liệu huấn luyện cho giải thuật học máy, loại thứ hai được dùng để kiểm định độ chính các của mỗi mô hình học máy ứng với một phương pháp trích chọn đặc tính khác nhau (gọi là dữ liệu kiểm định), loại dữ liệu cuối cùng là dữ liệu kiểm thử dùng để kiểm tra và đánh giá kết quả của hệ thống. Mỗi loại dữ liệu tương tác với hệ thống theo một luồng khác nhau, do đó sẽ có ba sơ đồ dữ luồng liệu tương ứng.

Hình xxx là mô hình chức năng, bao gồm bốn khối chức năng chính: tiền xử lý, trích chọn đặc tính, học máy và mô hình nhận dạng. Hình xxx, xxx, xxx lần lược là mô hình luồng dữ liệu tương ứng với tương tác của dữ liệu huấn luyện, dữ liệu kiểm định và dữ liệu kiểm thử với các khối chức năng của hệ thống. Các mục sau sẽ trình bày chi tiết từng thành phần của mỗi loại mô hình.



Mô hình chức năng

## Trích chọn đặc tính

## Giải thuật huấn luyện

CHƯƠNG 3

# THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Quá trình thực nghiệm được chia làm hai giai đoạn. Giai đoạn thứ nhất xây dựng mô hình nhận dạng với dữ liệu đầu vào ứng với các phương pháp trích chọn đặc tính khác nhau. Dữ liệu huấn luyện được sử dụng trong quá trình xây dựng mô hình. Bên cạnh đó sử dụng dữ liệu kiểm định để kiểm tra mức độ chính xác của mô hình (tính theo phần trăm). Ở giai đoạn này, một số giá trị tham số đầu vào được thay đổi bằng thực nghiệm để tìm ra giá trị phù hợp cho kết quả (độ chính xác) cao nhất ứng với mỗi phương pháp. Kết quả thu được ở giai đoạn thực nghiệm này là độ chính xác của mô hình tương ứng với các phương pháp trích chọn đặc tính, kết quả này chính là trọng số để xây dựng mô hình hoàn chỉnh trong giai đoạn thực nghiệm thứ hai.

Với các trọng số thu được trong giai đoạn thứ nhất, giai đoạn thực nghiệm thứ hai kết hợp tất cả các phương pháp trích chọn đặc tính theo tỉ lệ tương ứng để tạo nên mô hình nhận dạng hoàn chỉnh. Lúc này dữ liệu kiểm thử sẽ được sử dụng để đánh giá kết quả. Các phần sau sẽ mô tả chi tiết hai giai đoạn thực nghiệm và kết quả thu được.

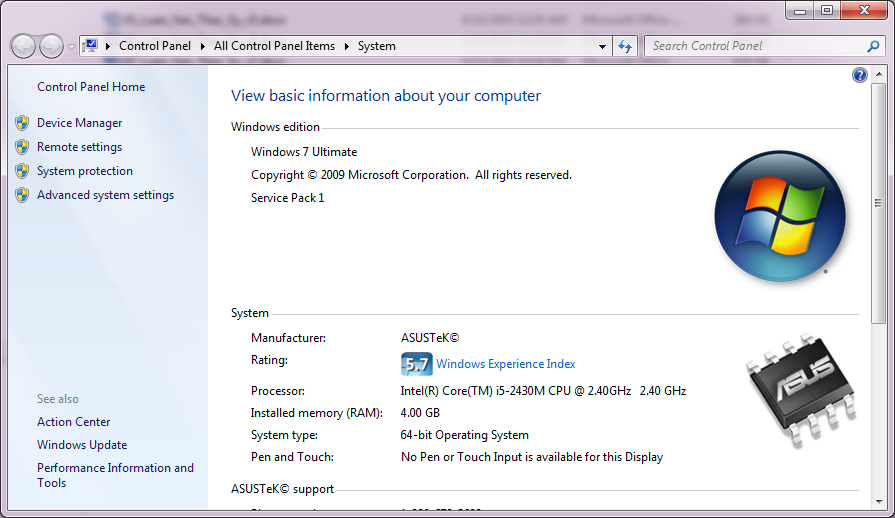
## Thông tin chung

### Dữ liệu sử dụng

Như đã trình bày ở các phần trước, dữ liệu sử dụng trong các giai đoạn thực nghiệm là dữ liệu chọn lọc từ cơ sở dữ liệu của CMU Mocap. Luận văn này chỉ sử dụng dữ liệu thuộc bốn hành động sau: chạy (run), đi (walk), nhảy (jump) và khiêu vũ (dance). Dữ liệu sau khi chọn lọc được chia ngẫu nhiên thành ba nhóm khác nhau: dữ liệu huấn luyện, dữ liệu kiểm định và dữ liệu kiểm thử. Bảng xxx thống kê số lượng dữ liệu của từng hành động trong mỗi nhóm.

### Môi trường triển khai

Quá trình thực nghiệm được thực hiện trên một máy tính xách tay sử dụng hệ điều hành Microsoft Windows 7 (64bit) có cấu hình tương đương một máy tính phổ thông hiện nay. Hình xxx mô tả các thông số kỹ thuật của máy.



Cấu hình máy tính sử dụng trong quá trình thục nghiệm

Ngôn ngữ sử dụng trong chương trình thực nghiệm là ngôn ngữ lập trình c#. Bên cạnh đó, chương trình thực nghiệm có sử dụng mộ số thư viện mã nguồn mở như Accord.NET[[15]](#footnote-16), AForge.NET[[16]](#footnote-17).

## Các giai đoạn thực nghiệm

### Giai đoạn thứ nhất

### Giai đoạn thứ hai

## Kết quả thực nghiệm

## Đánh giá

* Xxx

### Tiêu đề mục con

xxx

#### Tiêu đề mục con nhỏ hơn

* xxx
* xxx

# KẾT LUẬN

xxx

xxx

# PHỤ LỤC

Tấ 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 andits 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), “Recongnition and Segmentation of 3D Human Action Using HMM and Multi-class AdaBoost”, *Lecture Notes in Computer Science Vlolumn 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).

1. Kinect là một thiết bị đầu vào, là cảm biến chuyển động do hãng Microsoft sản xuất dành cho Xbox 360 và máy tính sử dụng hệ điều hành Windows. [↑](#footnote-ref-2)
2. Marker là một thiết bị đánh dấu, được gắn lên đối tượng cần theo dõi. [↑](#footnote-ref-3)
3. Mocap (Motion Capture) là hệ thống chuyên biệt dùng để chụp chuyển động. [↑](#footnote-ref-4)
4. Acclaim là tên một công ty game, đồng thời là tên một loại định dạng dữ liệu chuyển động 3D, một bộ gồm hai tệp tin asf/amc do chính công ty đó đề xuất. [↑](#footnote-ref-5)
5. Accord.NET Framework là thư viện mã nguồn mở tổng hợp các vấn đề trong học máy (http://accord-framework.net/) [↑](#footnote-ref-6)
6. Stereo camera là camera có hai hay nhiều ống kính với cảm biến ảnh hoặc khung phim riêng biệt cho mỗi ống kính. [↑](#footnote-ref-7)
7. 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 khoản cách của nó tới một điểm nhất định nào đó. Ảnh được tạo bởi range senor gọi là range image. [↑](#footnote-ref-8)
8. http://mocap.cs.cmu.edu/. [↑](#footnote-ref-9)
9. http://www.mpi-inf.mpg.de/resources/HDM05/. [↑](#footnote-ref-10)
10. http://kitchen.cs.cmu.edu/. [↑](#footnote-ref-11)
11. http://www.cs.rochester.edu/%18spark/muri/. [↑](#footnote-ref-12)
12. http://ias.in.tum.de/software/kitchen-activity-data/. [↑](#footnote-ref-13)
13. 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. [↑](#footnote-ref-14)
14. http://mocap.cs.cmu.edu/resources.php [↑](#footnote-ref-15)
15. http://accord-framework.net/ [↑](#footnote-ref-16)
16. http://www.aforgenet.com/ [↑](#footnote-ref-17)