

**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: 60.48.01.01**

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

Đà Nẵng – Năm 2015

Công trình được hoàn thành tại

ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: **TS. PHẠM MINH TUẤN**

Phản biện 1:

Phản biện 2:

Luận văn sẽ được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp
thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày
tháng..... năm

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin-Học liệu, Đại học Đà Nẵng
- Trung tâm Học liệu, Đại học Đà Nẵng

MỞ ĐẦU

1. Lý do chọn đề tài

1.1. 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 loại 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 Kinect¹ 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.



Hình 1. Microsoft Kinect Camera

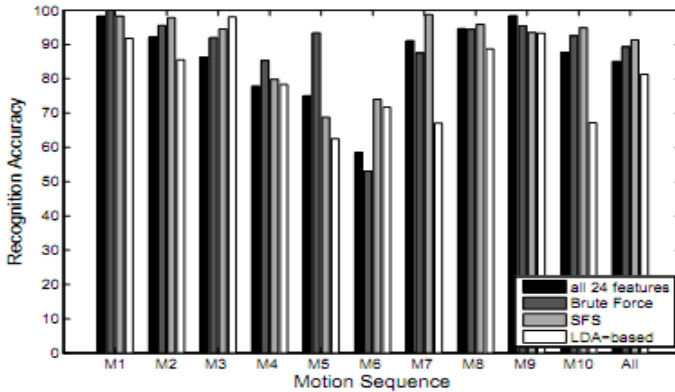
1.2. Các phương pháp trước đây

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 dụng

¹Kinect 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.

thống chụp chuyển động dựa vào marker² như là Mocap³ (Motion Capture) hoặc là dùng stereo camera⁴ - chụp hình ảnh 2D từ nhiều hướng khác nhau để dựng thành mô hình 3D.



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 đã nghiên cứu, thực nghiệm trên ba phương pháp trích chọn và lựa chọn đặc tính khác nhau (Brute Force, SFS, LDA) (xem hình 0.2) và đã giảm đáng kể hiệu năng của quá trình nhận dạng so với dữ liệu ban đầu. 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 (xem hình 0.3). K.

² Marker là một thiết bị đánh dấu, được gắn lên đối tượng cần theo dõi.

³ Mocap (Motion Capture) là hệ thống chuyên biệt dùng để chụp chuyển động.

⁴ Stereo camera là camera có hai hay nhiều ống kính với cảm biến ảnh hoặc khung phim riêng biệt cho mỗi ống kính.

Dana đã phát triển một mô hình học tăng cường mới dựa trên mô hình Markov ẩn.

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 dùng bảy loại đặc tính do L. Fengjun đề xuất

1.3. 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.

2. 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.

3. Đối tượng và phạm vi nghiên cứu

3.1. 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 Acclaim⁵ (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.

3.2. 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 vector hỗ trợ (SVM) trong học máy.

4. Phương pháp nghiên cứu

4.1. 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 vector hỗ trợ (SVM) và các phương pháp trích chọn, lựa chọn đặc tính.

⁵ 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.

4.2. 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 Framework⁶ 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.

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

5.1. Về mặt lý thuyết

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.

5.2. Về mặt thực tiễn

Đề xuất, 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.

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 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.

⁶Accord.NET Framework là thư viện mã nguồn mở tổng hợp các vấn đề trong học máy (<http://accord-framework.net/>)

Chương 2: Giải pháp đề xuất

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

Chương 3: Thực nghiệm và đánh giá kết quả

Chương này trình bày chi tiết quá trình thực nghiệm bao gồm môi trường thực nghiệm, các giai đoạn thực nghiệm và kết quả thực nghiệm qua từng giai đoạn, từ đó đưa ra các nhận xét, đánh giá.

CHƯƠNG 1

NGHIÊN CỨU TỔNG QUAN

1.1. 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. 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.

1.1.1. 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 camera. Cuối cùng là phương pháp sử dụng range sensor⁷. Mục này sẽ trình bày chi tiết từng phương pháp.

a. Phương pháp sử dụng stereo camera

Một stereo camera được trang bị hai hay nhiều ống kính với cảm biến ảnh hoặc phim riêng biệt cho mỗi ống kính. Điều đó

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 khoảng cách của nó tới một điểm nhất định nào đó. Ảnh được tạo bởi range sensor gọi là range image.

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

b. 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 điện ảnh, hoạt hình và trò chơi điện tử

c. 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).

1.1.2. Các phương pháp học máy thường sử dụng

Học máy (machine learning) có rất nhiều phương pháp, thuật toán khác nhau. Có nhiều cách để phân loại các thuật toán học máy, cách phân loại phổ biến nhất chia học máy thành ba loại chính: học có giám sát, học không giám sát và học bán giám sát.

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 vector hỗ trợ (SVM), mô hình Markov ẩn (HMM), kết hợp giữa SVM và HMM. Phần tiếp theo sẽ giới thiệu hai phương pháp phổ biến nhất: SVM và HMM.

a. Máy vector hỗ trợ

Máy vector hỗ trợ (Support Vector Machine – SVM) là 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 siêu 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.

b. 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

1.2. 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.

1.2.1. 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-40 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 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 này⁸.

1.2.2. 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*.

1.2.3. Cấu trúc Acclaim

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.

a. Cấu trúc tệp ASF

b. Cấu trúc tệp AMC

⁸ <http://mocap.cs.cmu.edu/resources.php>

1.3. 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. Mục đích chung của trích chọn hay lựa chọn đặc tính 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 - FE) và lựa chọn đặc tính (feature selection - FS). Sự khác nhau cơ bản giữa hai phương pháp này là trích chọn đặc tính sử dụng phương pháp biến đổi tuyến tính hoặc phi tuyến tính để biến đổi dữ liệu sang không gian mới, trong khi đó lựa chọn đặc tính chọn một tập con của tập dữ liệu cho trước. Một bên biến đổi dữ liệu, một bên không biến đổi. 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 phần chính (PCA), biệt thức tuyến tính (LDA). Nội dung chính của từng phương pháp được trình bày ở phần tiếp theo.

1.3.1. 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.

1.3.2. Phương pháp phân tích biệt thức tuyến tính – LDA

Phân tích biệt thức tuyến tính (Linear Discriminant Analysis – LDA) là phương pháp được sử dụng trong thống kê và học máy để giải quyết bài toán phân lớp hoặc trích chọn đặc tính. Trong phân loại hai lớp dữ liệu, LDA tìm kiếm trục đường thẳng sao cho khi tất cả dữ liệu của 2 lớp ánh xạ lên trục này có độ phân ly 2 lớp là cao nhất. Độ phân ly các lớp dữ liệu được định nghĩa bởi tỷ lệ phương sai giữa các lớp và phương sai giữa các dữ liệu trong từng lớp.

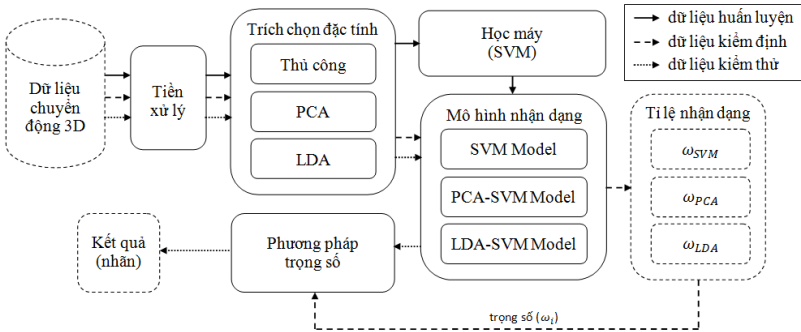
1.3.3. Sử dụng hàm nhân

Các phương pháp trích chọn đặc tính được trình bày ở trên chỉ có thể áp dụng hiệu quả với dữ liệu tuyến tính đơn giản, không phù hợp với các bài toán có dữ liệu phi tuyến, phức tạp. Do đó người ta dùng một phép biến đổi để chuyển dữ liệu phi tuyến tính thành tuyến tính. Phép biến đổi đó gọi là hàm nhân (kernel). Phương pháp PCA kết hợp hàm nhân gọi là KPCA (Kernel Principal Components Analysis), phương pháp LDA kết hợp hàm nhân gọi là KDA (Kernel Discriminant Analysis). Các phương pháp trích chọn đặc tính sử dụng trong luận văn này đều được kết hợp với hàm nhân.

CHƯƠNG 2

GIẢI PHÁP ĐỀ XUẤT

Hình 2.1 mô tả tổng quan hệ thống đề xuất, bao gồm năm khối chức năng chính: tiền xử lý, trích chọn đặc tính, học máy, mô hình nhận dạng và phương pháp trọng số. Có ba phương pháp chính trong trích chọn đặc tính: lựa chọn thủ công, PCA và LDA. Mô hình nhận dạng chính là mô hình học máy xây dựng được từ dữ liệu huấn luyện. Mỗi phương pháp trích chọn đặc tính xây dựng được một mô hình nhận dạng.



Hình 2.1. Mô hình hệ thống đề xuất

Dữ liệu đầu vào của hệ thống được chọn lựa trong cơ sở dữ liệu của CMU Mocap và được phân chia ngẫu nhiên thành ba nhóm: nhóm thứ nhất được dùng làm dữ liệu huấn luyện (training data) cho giải thuật học máy, nhóm thứ hai được dùng để kiểm định độ chính xác của mỗi mô hình học máy ứng với một phương pháp trích chọn đặc tính khác nhau (gọi là nhóm dữ liệu kiểm định - validating data), nhóm dữ liệu cuối cùng là dữ liệu kiểm thử (testing data) dùng để kiểm tra và đánh giá kết quả của hệ thống.

Trong hệ thống đề xuất, có tất cả ba luồng dữ liệu chính. Thứ nhất là luồng dữ liệu huấn luyện, được thể hiện bằng hình mũi tên nét liền. Kết quả cuối cùng của luồng dữ liệu huấn luyện là mô hình nhận dạng xây dựng được sau khi đã áp dụng giải thuật học máy. Thứ hai là luồng dữ liệu kiểm định, được thể hiện bằng hình mũi tên nét đứt. Kết quả cuối cùng của luồng dữ liệu kiểm định là tỉ lệ nhận dạng tương ứng với mỗi phương pháp trích chọn đặc tính. Các tỉ lệ này sẽ là đầu vào cho phương pháp trọng số. Thứ ba là luồng dữ liệu kiểm thử (hay bất kỳ dữ liệu mới nào), được mô tả bằng hình mũi tên chấm liền. Dữ liệu của luồng này sau khi qua các khối chức năng tiền xử lý, trích chọn đặc tính, mô hình nhận dạng, phương pháp trọng số sẽ cho ra kết quả là nhãn của hành động cần nhận dạng. Chức năng học máy chỉ được sử dụng bởi luồng dữ liệu huấn luyện, các luồng dữ liệu còn lại chỉ sử dụng mô hình nhận dạng xây dựng bởi học máy.

Các mục sau sẽ trình bày chi tiết từng khối chức năng trong hệ thống đồng thời giải thích vai trò của chúng với mỗi luồng dữ liệu khác nhau.

2.1. TIỀN XỬ LÝ

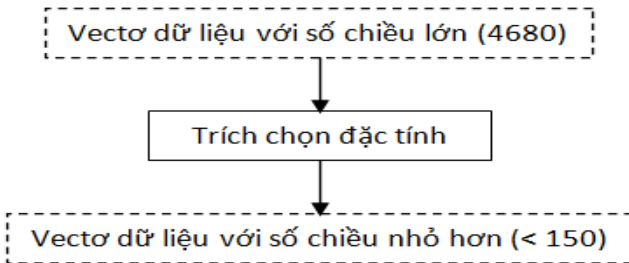
Tín hiệu video được quay bởi một loại camera chuyên dụng MX-40, có tốc độ 120Hz, nghĩa là dữ liệu thu được sẽ bao gồm 120 khung hình (frame) trong một giây. Tuy nhiên thời gian thực hiện của mỗi loại hành động trong mỗi thí nghiệm là khác nhau, do đó dữ liệu thu được có độ dài ngắn khác nhau ứng với số lượng nhiều hay ít các khung hình.

Ngoài ra, mô hình bộ xương 3D của con người có một số lượng lớn các đoạn xương, kết hợp với độ tự do của mỗi khớp sẽ làm tăng số chiều của thuộc tính.

Trước sự không đồng bộ của dữ liệu và độ lớn của số lượng thuộc tính, hệ thống cần có chức năng tiền xử lý để phân nào giải quyết hai vấn đề trên.

2.2. TRÍCH CHỌN ĐẶC TÍNH

Nội dung chính của luận văn này là nghiên cứu giai đoạn trích chọn đặc tính. Luận văn này sử dụng một số phương pháp trích chọn đặc tính phổ biến và kết hợp các phương pháp với nhau sử dụng trọng số. Hình 2.5 mô tả quá trình trích chọn đặc tính chung cho các phương pháp. Mục đích chung là tìm vector biểu diễn dữ liệu với số chiều nhỏ hơn dữ liệu ban đầu nhưng vẫn đảm bảo thể hiện đầy đủ các đặc trưng của dữ liệu.



Hình 2.5. Mô tả quá trình trích chọn đặc tính

a. Lựa chọn thủ công

Mô hình 3D cơ thể người được cấu thành từ tất cả 29 xương khác nhau. Tuy nhiên rất nhiều trong số đó không phải là đặc trưng của một số loại hành động. Do đó luận văn này sẽ tiến hành thực

nghiệm với một tập con các xương trong tổng số 29 xương ở trên. Quá trình thực nghiệm sẽ tiến hành với một số các xương cơ bản trước, sau đó sẽ từng bước thêm vào các xương khác nhau để tìm ra tập con thích hợp nhất.

b. PCA

PCA là phương pháp thích hợp để ứng dụng trong vector dữ liệu có số chiều lớn như dữ liệu của bài toán nhận dạng hành động. Vì dữ liệu đã được xử lý ở giai đoạn trước nên việc áp dụng PCA vào trong bài toán nhận dạng hành động hoàn toàn giống như áp dụng trên các bài toán khác. Vấn đề cần giải quyết ở đây là tìm số lượng vector riêng (hay nói cách khác là số chiều của dữ liệu sau trích chọn) thích hợp sao cho độ chính xác của mô hình là lớn nhất. Vấn đề này có thể đơn giản được giải quyết bằng thực nghiệm. Chương trình sẽ thử lần lượt các giá trị từ nhỏ đến lớn cho đến khi tìm được giá trị thích hợp.

c. LDA

Khác với PCA, LDA không chỉ là một phương pháp trích chọn đặc tính thông thường mà còn là một thuật toán phân loại, nghĩa là nó có khả năng phân loại dữ liệu. Do đó LDA cần một bộ dữ liệu huấn luyện (có gán nhãn) cho thuật toán trích chọn đặc tính. Nếu như tham số cần tìm của PCA là số vector riêng thì với LDA đó là số lượng các đặc trưng khác nhau của dữ liệu. Tham số này cũng được xác định thông qua thực nghiệm.

2.3. HỌC MÁY

Về cơ bản, SVM được dùng cho các bài toán tuyến tính, có nghĩa là tồn tại một siêu phẳng tối ưu và lẽ cực đại phân chia hoàn

toàn hai các lớp dữ liệu. Tuy nhiên trong rất nhiều trường hợp miền trong của lễ trong tập dữ liệu huấn luyện có thể có một số lượng nhỏ các điểm, dẫn đến việc không thể phân chia tập dữ liệu bằng một siêu phẳng tuyến tính. Để giải quyết các bài toán phi tuyến tính người ta dùng một kỹ thuật để biến đổi siêu phẳng từ phi tuyến tính trở thành tuyến tính, phép biến đổi này gọi là hàm nhân. Một số hàm nhân thường dùng là hàm nhân tuyến tính (linear kernel) và hàm nhân đa thức (polynomial kernel). Luận văn này sử dụng hàm nhân tuyến tính cho thuật toán SVM.

2.4. MÔ HÌNH NHẬN DẠNG

Mô hình nhận dạng ở đây chính là mô hình xây dựng được từ dữ liệu huấn luyện sau khi áp dụng giải thuật học máy. Với mỗi phương pháp trích chọn đặc tính sẽ có một mô hình nhận dạng tương ứng. Mô hình nhận dạng sẽ được sử dụng bởi luồng dữ liệu kiểm định và luồng dữ liệu kiểm thử.

2.5. PHƯƠNG PHÁP TRỌNG SỐ

Trọng số là độ chính xác của các mô hình nhận dạng sau khi kiểm định bằng dữ liệu kiểm định. Ví dụ, với một hành động cần nhận dạng, nếu mỗi phương pháp cho ra một kết quả khác nhau thì kết quả cuối cùng được chọn là kết quả của phương pháp có trọng số cao nhất. Trong trường hợp khác, nếu có hai hay nhiều phương pháp cho ra cùng một hành động thì xác suất để chọn hành động đó làm kết quả cuối cùng bằng tổng các trọng số. Phương pháp trọng số nhận đầu vào là các tỉ lệ nhận dạng thu được khi sử dụng dữ liệu kiểm định và kết quả của mỗi mô hình nhận dạng. Đầu ra của phương pháp trọng số là nhãn của hành động cần nhận dạng.

Cụ thể hàm xác suất và cách gán nhãn cho hành động cần nhận dạng được mô tả như sau: Giả sử ta có n phương pháp trích chọn đặc tính và m loại hành động khác nhau (cụ thể trong luận văn này $n = 3$, $m = 4$). Các phương pháp được đánh số thứ tự từ 1 đến n ; các loại hành động được gán nhãn từ 1 đến m . Gọi f_i là nhãn của hành động nhận dạng từ mô hình sử dụng phương pháp trích chọn đặc tính i , ω_i là trọng số của phương pháp trích chọn đặc tính i , $i = \{1..n\}$, $f_i \in \{1..m\}$. Gọi p_j là xác suất để gán nhãn j cho hành động cần nhận dạng, $j = \{1..m\}$. Khi đó xác suất p_j được xác định như sau:

$$p_j = \sum_{i|f_i=j} \omega_i$$

Nhãn \hat{f} cần tìm được xác định như sau:

$$\hat{f} = \underset{j}{\operatorname{argMax}}(p_j)$$

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.

3.1. MÔI TRƯỜNG THỰC NGHIỆM

3.1.1. Dữ liệu sử dụng

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.

3.1.2. Môi trường triển khai

3.2. CÁC GIAI ĐOẠN THỰC NGHIỆM

3.2.1. Giai đoạn thứ nhất

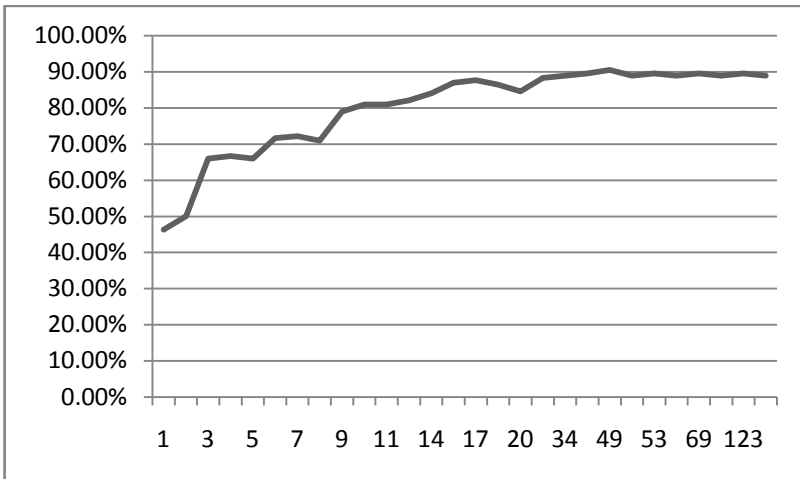
Mục đích của giai đoạn này là tìm các tham số thích hợp cho mỗi phương pháp trích chọn đặc tính để mô hình nhận dạng xây dựng được có độ chính xác cao nhất. Độ chính xác của mỗi mô hình được kiểm nghiệm bằng tập dữ liệu kiểm định. Sau đây là các kết quả tương ứng cho từng phương pháp.

a. Lựa chọn thủ công

Bảng 3.2. Kết quả thực nghiệm với phương pháp lựa chọn thủ công

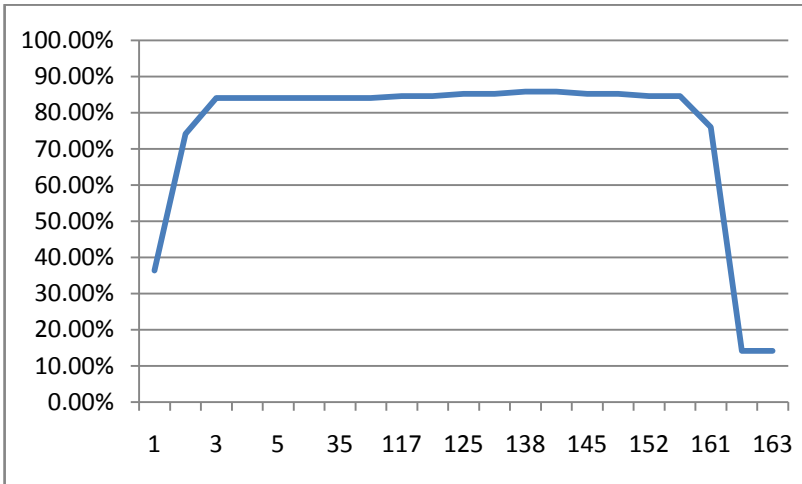
Số xương	Run	Walk	Jump	Dance	Trung bình
3	0.0%	93.3%	14.3%	18.2%	49.4%
4	0.0%	93.3%	14.3%	13.6%	48.8%
7	0.0%	94.7%	30.9%	13.6%	53.7%
13	0.0%	96.0%	28.5%	18.2%	54.3%
23	78.3%	98.7%	81.0%	31.9%	82.1%
11	78.3%	98.7%	81.0%	36.4%	82.7%
Tất cả (29)	78.3%	98.7%	81.0%	41.0%	83.3%

b. Phương pháp PCA



Hình 3.1. Sự biến thiên của tỉ lệ nhận dạng trong PCA

c. Phương pháp LDA



Hình 3.2. Sự biến thiên của tỉ lệ nhận dạng trong LDA

3.2.2. Giai đoạn thứ hai

Bảng 3.25. Kết quả giai đoạn thực nghiệm thứ hai

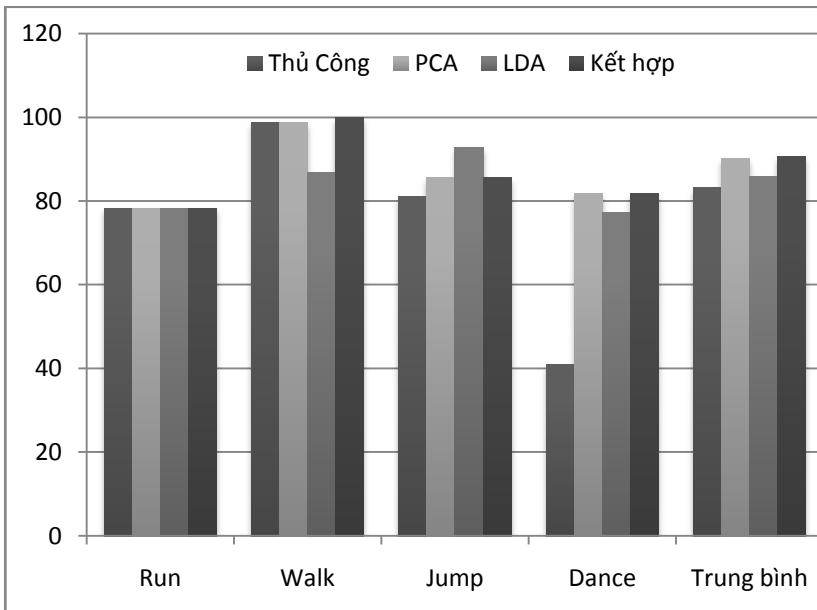
Hoạt động	Run	Walk	Jump	Dance
Run	78.3%	17.4%	4.3%	0.0%
Walk	0.0%	100%	0.0%	0.0%
Jump	0.0%	14.3%	85.7%	0.0%
Dance	0.0%	9.1%	9.1%	81.8%

3.3. ĐÁNH GIÁ

Với bài toán nhận dạng hành động trong không gian 3D, với phương pháp trích chọn đặc tính thủ công có thể sử dụng nhóm có 11 xương để xây dựng mô hình huấn luyện và nhận dạng.

Từ kết quả thực nghiệm với hai phương pháp trích chọn đặc tính PCA, LDA ta thấy: với cùng cỡ sở dữ liệu thì độ chính xác cao nhất của PCA là 90.1%, độ chính xác cao nhất của LDA là 85.8%. Như vậy trong trường hợp này, PCA cho kết quả tốt hơn LDA. Kết quả thực nghiệm cũng cho thấy PCA hiệu quả hơn trong việc tối ưu hóa việc thể hiện sự biên thiên của dữ liệu. Còn LDA hiệu quả hơn trong việc phân ly dữ liệu dựa vào sự đo lường các đặc trưng.

So sánh kết quả giữa áp dụng trích chọn đặc tính và không áp dụng trích chọn đặc tính có thể thấy sự khác biệt lớn. Vì vậy trích chọn đặc tính luôn là một bước quan trọng trong việc xây dựng hệ thống nhận dạng hành động người. Biểu đồ ở hình 3.3 so sánh kết quả giữa các phương pháp.



Hình 3.3. Thống kê kết quả giữa các phương pháp

KẾT LUẬN

Nhận dạng hành động người được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau của cuộc sống. Kết hợp hai phương pháp nghiên cứu lý thuyết và thực nghiệm, luận văn này đã trình bày tổng quan về nhận dạng hành động người trong không gian 3D. Bao gồm các phương pháp thu thập dữ liệu chuyển động 3D, các giải thuật học máy thường sử dụng và đặc biệt là các phương pháp trích chọn đặc tính. Bên cạnh đó, luận văn đã nghiên cứu và đề xuất mô hình nhận dạng kết hợp có trọng số các phương pháp trích chọn đặc tính khác nhau với độ chính xác cao. Kết quả thực nghiệm cho thấy mô hình đề xuất cho kết quả nhận dạng tốt hơn so với mô hình truyền thống.

Ngoài ra quá trình thực nghiệm đã tìm ra được một nhóm các xương có ảnh hưởng lớn đến việc thể hiện các đặc trưng của hành động người trong không gian 3D.

Tuy nhiên, việc kết hợp sử dụng song song các phương pháp đã làm giảm hiệu năng của mô hình nhận dạng. Do vậy hướng nghiên cứu tiếp theo đối với đề tài này là nâng cao hiệu năng của hệ thống khi sử dụng kết hợp nhiều phương pháp khác nhau.