

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

ĐẶNG TẤN ĐÔNG

**ỨNG DỤNG HIDDEN MARKOV MODELS
TRONG NHẬN DẠNG HÀNH ĐỘNG
CON NGƯỜI DỰA TRÊN CẤU TRÚC
ACCLAIM**

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: TS. Đặng Hoài Phương

Phản biện 2: PGS.TS. Võ Thanh Tú

Luận văn sẽ được bảo vệ trước Hội đồng chấm Luận văn tốt nghiệp thạc sĩ Khoa học máy tính họp tại Đại học Đà Nẵng vào ngày 18 tháng 7 năm 2015.

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin - Học liệu, Đại học Đà Nẵng*
- Thư viện trường Đại học Bách Khoa, Đại học Đà Nẵng*

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Hiện nay, công nghệ thông tin ngày càng phát triển mạnh mẽ và rộng rãi, tương tác giữa con người và thiết bị trở nên cần thiết và quan trọng. Việc tạo ra các hệ thống có khả năng nhận định thông tin là một trong những hướng đi có rất nhiều thành công. Từ các dạng thông tin rời rạc thu nhận được, hệ thống sẽ phân tách thành các mẫu thông tin riêng biệt, sau đó biểu diễn, sắp xếp, phân loại và dùng kỹ thuật xử lý để nhận ra ý nghĩa của các thông tin đó. Ứng dụng của lý thuyết nhận dạng mẫu đã được áp dụng cho các lĩnh vực: Nhận dạng tín hiệu, nhận dạng tiếng nói, nhận dạng khuôn mặt, nhận dạng hành động, nhận dạng vân tay... Trong đó nhận dạng hành động của con người là một trong những vấn đề được các nhà khoa học nghiên cứu và ứng dụng trong thực tiễn.

Những năm 1980, các hệ thống nhận dạng hành động gặp phải không ít những khó khăn: Cấu hình phần cứng máy tính rất thấp kém, chất lượng các thiết bị thu nhận không cao và chưa được tập trung nghiên cứu nhiều. HMM (Hidden Markov Models) ban đầu được nghiên cứu cho lĩnh vực nhận dạng tiếng nói, nhưng kỹ thuật này đã nhanh chóng được phát triển và ứng dụng trong các bài toán nhận dạng. Trong đó nhận dạng hành động con người chiếm vị trí quan trọng trong lĩnh vực này. Đã có rất nhiều kỹ thuật được áp dụng để nhận dạng hành động con người như: dựa vào hình ảnh như Optical flow, bộ lọc Kalman, mô hình Markov ẩn, mô hình SVM (Support Vector Machine) ... sử dụng các dữ liệu khác nhau từ camera, stereo và hồng ngoại. Nhưng vẫn còn hạn chế ở quá trình nhận dạng và thiếu

hạt dữ liệu dẫn đến hiệu năng chưa cao. Và gần đây khi công nghệ 3D càng ngày càng phát triển, thì việc áp dụng nó vào trong việc nhận dạng đang được các nhà khoa học tập trung nghiên cứu. Trong đó cấu trúc Acclaim được xây dựng bởi đại học Carnegie Mellon (CMU) là cơ sở dữ liệu được xây dựng đầy đủ tất cả các chuyển động của con người trên mô hình 3D gần như là hoàn chỉnh nhất. Chính vì vậy tôi đã chọn đề tài “Ứng dụng Hidden Markov Models trong nhận dạng hành động người dựa trên cấu trúc Acclaim” với mong muốn phần nào áp dụng vào bài toán thực tế.

2. Mục tiêu và nhiệm vụ nghiên cứu

2.1. Mục tiêu

Ứng dụng mô hình Markov ẩn trong học máy để nhận dạng hành động của con người dựa trên cấu trúc Acclaim.

2.2. Nhiệm vụ của đề tài

Tìm hiểu về nhận dạng hành động, một số phương pháp máy học trong nhận dạng. Chú trọng đến việc tìm hiểu, phân tích, thiết kế, triển khai xây dựng hệ thống nhận dạng hành động của con người trên mô hình Markov ẩn dựa trên cấu trúc Acclaim.

3. Đối tượng và phạm vi nghiên cứu

3.1. Đối tượng nghiên cứu

Đối tượng nghiên cứu của đề tài này là cấu trúc Acclaim, mô hình Markov ẩn để nhận dạng hành động của con người.

3.2. Phạm vi nghiên cứu

- Nghiên cứu nhận dạng hành động sử dụng HMM (Hidden Markov Models).

- Phân tích cấu trúc file Acclaim trong ứng dụng nhận dạng hành động của con người.

- Xây dựng chương trình nhận dạng hành động sử dụng HMM

(Hidden Markov Models).

4. Phương pháp nghiên cứu

Phương pháp nghiên cứu, sử dụng hai phương pháp chính là nghiên cứu lý thuyết và nghiên cứu thực nghiệm.

4.1. Phương pháp lý thuyết

- Tìm hiểu ngôn ngữ lập trình C# trong môi trường .NET.
- Tìm hiểu về nhận dạng hành động, các phương pháp máy học, trong đó chú trọng đến nghiên cứu mô hình Markov ẩn.
- Các tài liệu về phương pháp và thuật toán nhận dạng.
- Tìm hiểu, nghiên cứu về HMM (Hidden Markov Models).

4.2. Phương pháp thực nghiệm

- Xây dựng chương trình ứng dụng HMM (Hidden Markov Models) để nhận dạng hành động người bằng mô hình Acclaim.

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

Nghiên cứu xây dựng một chương trình nhận dạng hành động của người dựa trên mô hình Acclaim có độ chính xác cao.

5.1. Về mặt khoa học

Thực hiện nghiên cứu tổng quan về nhận dạng hành động người bao gồm các hướng tiếp cận nhận dạng hành động, các mô hình và kỹ thuật nhận dạng, tiếp đến trình bày các bước tiền xử lý nhận dạng, phương pháp phân tích tính đặc trưng hành động. Đối với bài toán nhận dạng, nghiên cứu chi tiết, triển khai và ứng dụng mô hình Markov ẩn trong nhận dạng hành động dựa trên mô hình Acclaim .

5.2. Về thực tiễn

- Giúp cho việc nhận dạng hành động của con người chính xác và hiệu quả.
- Triển khai được cho nhiều lĩnh vực khác nhau.

6. Bố cục của luận văn

Chương 1: Nghiên cứu tổng quan

Chương 2: Ứng dụng mô hình Markov ẩn trong nhận dạng hành động của con người dựa trên cấu trúc Acclaim.

Chương 3: Thực nghiệm và đánh giá kết quả.

CHƯƠNG 1

NGHIÊN CỨU TỔNG QUAN

1.1. TỔNG QUAN VỀ HỌC MÁY

1.1.1. Khái quát về học máy

Học máy (còn gọi là máy học) là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc phát triển các kỹ thuật cho phép các máy tính có thể “học”.

1.1.2. Các vấn đề trong học máy

Các vấn đề thường gặp trong học máy là:

- Các ví dụ huấn luyện có tồn tại thuật toán để học hàm đích hay không? Thuật toán nào là tốt nhất?.

- Bao nhiêu dữ liệu huấn luyện thì đủ? Ước lượng các mức tin cậy thế nào? Các tri thức đã biết có tác dụng gì cho quá trình học.

- Chiến lược thí nghiệm đào tạo như thế nào là tốt nhất, cách chọn chiến lược này như thế nào?

- Cách nào tốt nhất để đưa bài toán học về bài toán xấp xỉ hàm? Quá trình này có thể tự động hóa không?

1.1.3. Tổng quan về các phương pháp học máy

a. Học có giám sát (Supervised Learning)

Học có giám sát là phương pháp học với tập dữ liệu huấn luyện ban đầu hoàn toàn được gán nhãn từ trước, sử dụng cho lớp bài toán phân lớp, phân loại (classification).

b. Học không giám sát (Unsupervised Learning)

Là quá trình mô hình hóa một tập dữ liệu, không có sẵn các ví dụ đã được gán nhãn. Học không giám sát là học với tập dữ liệu huấn luyện ban đầu hoàn toàn chưa được gán nhãn. Học không giám sát là phương pháp học sử dụng cho lớp bài toán gom cụm, phân cụm (clustering).

c. Học bán giám sát(Semi-Supervised Learning)

Nội dung chính của học bán giám sát là hệ thống sử dụng một tập dữ liệu huấn luyện (training set) gồm 2 phần: các dữ liệu huấn luyện có nhãn, thường với số lượng rất ít, và các dữ liệu huấn luyện không có nhãn, thường với số lượng rất nhiều. Thực tế cho thấy khi sử dụng kết hợp dữ liệu không có nhãn với một lượng nhất định dữ liệu có nhãn có thể tăng độ chính xác đáng kể.

d. Học tăng cường (Reinforcement Learning)

Các thuật toán học tăng cường cố gắng tìm một chiến lược ánh xạ các trạng thái của thế giới tới các hành động mà agent nên chọn trong các trạng thái đó. Trong học tăng cường không có các cặp dữ liệu vào kết quả đúng, các hành động gần tối ưu cũng không được đánh giá đúng sai một cách tường minh. Hơn nữa, ở đây hoạt động trực tuyến được quan tâm, trong đó có một việc tìm kiếm một sự cân bằng giữa khai phá và khai thác.

1.2. TỔNG QUAN VỀ NHẬN DẠNG

1.2.1. Giới thiệu

Nhận dạng mẫu là một ngành thuộc lĩnh vực học máy. Nó là một tập các phương pháp học có giám sát nhằm phân loại dữ liệu (là các mẫu) dựa trên tiêu chí: kiến thức tiên nghiệm hoặc thông tin thống kê được trích rút từ các mẫu có sẵn. Các mẫu cần phân loại thường được biểu diễn thành các nhóm của các dữ liệu đo đạc hay quan sát được, mỗi nhóm là một điểm ở trong một không gian đa chiều phù hợp. Đó là không gian của các đặc tính để dựa vào đó ta có thể phân loại.

1.2.2. Các phương pháp nhận dạng

a. Đối sánh mẫu

Các kỹ thuật đối sánh có thể nghiên cứu theo ba hướng sau:

- Đối sánh trực tiếp.
- Các mẫu biến dạng và đối sánh mềm.
- Đối sánh giảm nhẹ.

b. K-láng giềng gần nhất (K-Nearest Neighbors)

Phương pháp này sử dụng các vectơ đặc trưng được thu thập trong quá trình huấn luyện để tìm k láng giềng gần trong không gian n-chiều.

c. Máy véc tơ hỗ trợ (Support Vector Machine-SVM)

Có thể mô tả SVM như sau: Cho trước 2 tập dữ liệu học, mỗi tập thuộc 1 lớp cho trước, bộ phân lớp SVM sẽ xây dựng mô hình phân lớp dựa trên 2 tập dữ liệu này. Khi có mẫu mới được đưa vào, bộ phân lớp sẽ đưa ra dự đoán xem mẫu này thuộc lớp nào trong 2 lớp đã định.

d. Mạng nơ ron (Neuron Network-NN)

Mạng nơ ron là một cấu trúc với nhiều bộ xử lý “nơ ron” được kết nối song song và chằng chịt với nhau. Có 2 nhóm chính trong kiến trúc mạng nơ ron: mạng truyền thẳng và mạng lan truyền ngược.

e. Mô hình Markov ẩn (Hidden Markov Model-HMM)

HMM là quá trình Markov với các tham số không biết trước, nhiệm vụ xác định các tham số ẩn. Các tham số của mô hình được rút ra sau đó có thể sử dụng để thực hiện các phân tích kế tiếp.

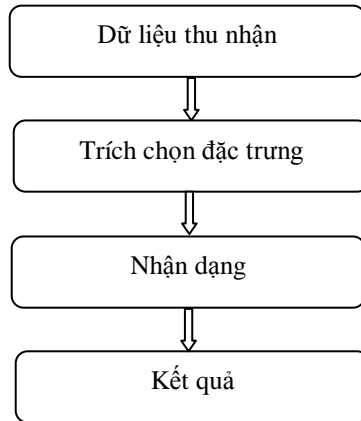
1.2.3. Tiêu chí đánh giá của một hệ thống nhận dạng

a. Độ chính xác

b. Tính thích nghi

c. Thời gian hoạt động

1.2.4. Mô hình tổng quát hệ thống nhận dạng hành động



1.2.5. Khó khăn trong việc nhận dạng hành động

Khó khăn cơ bản là hành động luôn biến thiên theo thời gian và có sự khác biệt giữa hành động người tùy độ tuổi, tốc độ chuyển động, môi trường... Xác định hành động nào là chính, hành động nào không có ích với nhận dạng hành động.

1.3. ỨNG DỤNG CỦA NHẬN DẠNG HÀNH ĐỘNG

1.3.1. Thực tế ảo

Các ứng dụng thực tế ảo gia 8ang đã kiểm nghiệm một trong những mức độ hấp dẫn lớn nhất trong lĩnh vực máy tính. Tương tác thực tế ảo sử dụng cử chỉ cho phép thao tác các đối tượng thực sử dụng các bàn tay, cho các tương tác 3D, hoặc 2D mô phỏng 3D.

1.3.2. Điều khiển robot

Ứng dụng điều khiển robot thường nằm trong lĩnh vực thăm dò không gian và các dự án nghiên cứu quân sự.

1.3.3. Trò chơi

Bàn tay của một người chơi hoặc vị trí cơ thể để điều khiển chuyển động và định hướng đối tượng trò chơi tương tác, chẳng hạn

như đua xe hơi, chơi tennis, đánh cầu lông, lướt ván....

1.3.4. Ngôn ngữ ký hiệu

Ngôn ngữ cử chỉ, hành động là trường hợp quan trọng của cử chỉ giao tiếp. Đầu tiên là một hệ thống trên máy để bàn, hành động của người thực hiện được lưu giữ thông qua một camera phía trước.

1.4. MỘT SỐ CÔNG TRÌNH ĐÃ NGHIÊN CỨU LIÊN QUAN

1.5. GIỚI THIỆU CÔNG NGHỆ MOCAP (MOTION CAPTURE)

MoCap là kỹ thuật thu lại chuyển động thật của đối tượng được sử dụng trong các lĩnh vực điện ảnh, quân đội, giải trí,... Đặc biệt công nghệ này được sử dụng nhiều nhất trong lĩnh vực điện ảnh.

Có thể phân loại các phương pháp MoCap như sau:

- Optical Passive (hệ thống quang học bị động).
- Optical Active (hệ thống quang học chủ động).
- Electromagnetic (hệ thống cảm biến từ).
- Acoustic (hệ thống cảm biến âm thanh).
- Electromechanical (hệ thống cảm biến cơ học).
- Optical Fabric (hệ thống cảm biến sợi quang)

1.6. CẤU TRÚC ACCLAIM

1.6.1. Giới thiệu

Acclaim là công ty games có trụ sở tại Beverly Hills, California. Acclaim Game có rất nhiều nghiên cứu trong lĩnh vực theo dõi chuyển động. Họ đã phát triển và sử dụng cấu trúc riêng của mình trong việc lưu trữ dữ liệu chuyển động 3D.

Cấu trúc Acclaim bao gồm hai tệp: Tệp thứ nhất lưu trữ cấu trúc khung tổng quan của đối tượng, tệp thứ hai lưu trữ 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 khung bộ xương

được nối với nhau bởi 30 khớp trên cơ thể, tệp thứ hai là vị trí của các khớp xương ứng với mỗi mốc thời gian. Tệp cấu trúc là gọi là ASF (Acclaim Skeleton File), còn tệp chuyển động được gọi là AMC (Acclaim Motion Capture).

1.6.2. Cấu trúc tập tin ASF

Trong các tập tin ASF (Acclaim Skeleton File), một cơ sở tư thế được xác định cho các bộ xương đó là trạng thái khởi đầu cho các dữ liệu chuyển động.

Kết cấu của một file ASF bao gồm: “version”, “unit”, “documentation”, “root”, “bonedata”, “id”, “name”, “direction”, “length”, “axis”, “dof”, “limits”.

1.6.3. Cấu trúc tập tin AMC

AMC chứa các dữ liệu chuyển động cho một bộ xương được xác định bởi một tập tin ASF. Các dữ liệu chuyển động được đưa ra một mẫu tại một thời điểm. Mỗi mẫu bao gồm một số dòng, một đoạn trên mỗi dòng, có chứa các dữ liệu.

CHƯƠNG 2

ỨNG DỤNG MÔ HÌNH MARKOV ẨN TRONG NHẬN DẠNG HÀNH ĐỘNG CỦA CON NGƯỜI TRÊN CẤU TRÚC ACCLAIM

2.1. MÔ HÌNH MARKOV ẨN

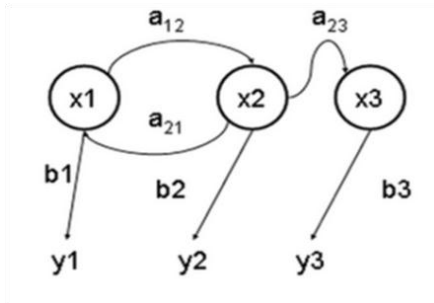
2.1.1. Mô hình Markov (Markov Model-MM)

Mô hình Markov còn gọi là mô hình Markov điển hình, hay mô hình Markov “hiện”. Trong mô hình này, trạng thái được quan sát trực tiếp bởi người quan sát, do đó các xác suất chuyển tiếp trạng thái là các tham số duy nhất.

Mô hình Markov được ứng dụng trong rất nhiều lĩnh vực như: nhận dạng gene, phân loại protein, dự báo giá, nhận dạng giọng nói, xử lý tín hiệu... Tuy nhiên, Markov cũng có những hạn chế như: các giá trị chuyển trạng thái là các giá trị áp đặt sẵn, trong khi đối tượng và dữ liệu quan sát luôn biến đổi theo thời gian. Do đó, để khắc phục nhược điểm này, chúng ta nghiên cứu mô hình Markov ẩn.

2.1.2. Mô hình Markov ẩn (Hidden Markov Model-HMM)

Trong một mô hình Markov điển hình, trạng thái được quan sát trực tiếp bởi người quan sát, và vì vậy các xác suất chuyển tiếp trạng thái là các tham số duy nhất.



Hình 2.2 Mô hình Markov ẩn

x_i : Các trạng thái trong mô hình Markov.

a_{ij} : Các xác suất chuyển tiếp.

b_{ij} : Các xác suất đầu ra.

y_i : Các dữ liệu quan sát.

Trong hầu hết các nghiên cứu và ứng dụng của mô hình Markov ẩn, đều đi đến việc giải quyết ba bài toán cơ bản sau:

Bài toán 1: Cho chuỗi tín hiệu quan sát $O = \{O_1, O_2, \dots, O_T\}$ mô hình Markov ẩn $\lambda = (A, B, \pi)$. Tính $P(O|\lambda)$ là xác suất xảy ra chuỗi tín hiệu quan sát O từ mô hình λ ?

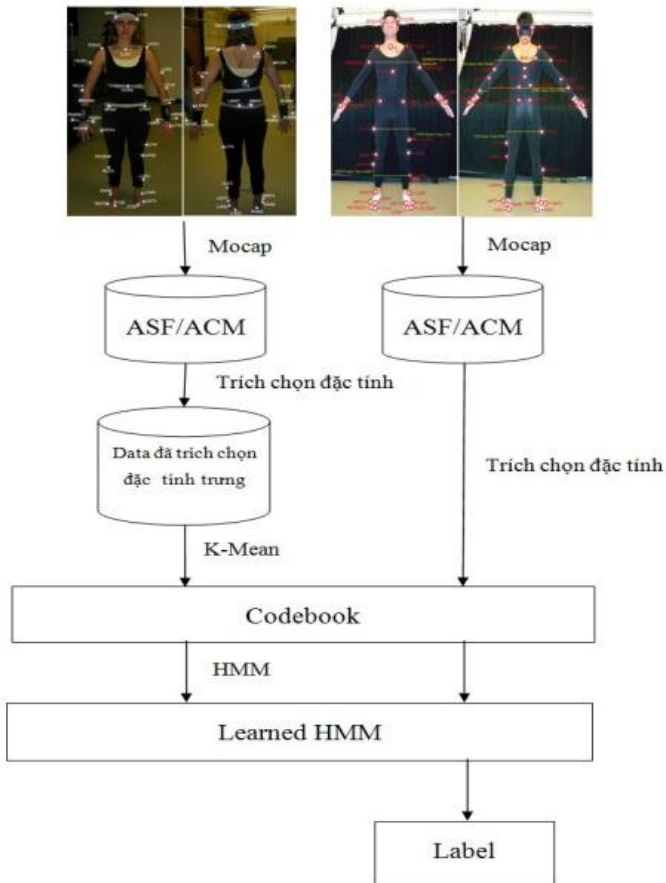
Bài toán 2: Cho chuỗi tín hiệu quan sát $O = \{O_1, O_2, \dots, O_T\}$ và mô hình Markov ẩn, tìm chuỗi trạng thái $Q = \{Q_1, Q_2, \dots, Q_T\}$ sao cho xác suất tương ứng với chuỗi tín hiệu quan sát được O lớn nhất? (với bài toán này, chúng ta cung cấp cho mô hình các tham số, tìm dãy các trạng thái ẩn có khả năng lớn nhất mà có thể sinh ra dãy đầu ta đã cung cấp, sử dụng thuật toán Viterbi).

Bài toán 3: Xây dựng mô hình Markov ẩn $\lambda = (A, B, \pi)$ sao cho $P(O|\lambda)$ là lớn nhất? (với bài toán này, chúng ta cung cấp dãy đầu ra, tìm tập hợp có khả năng nhất của chuyển tiếp trạng thái và các xác suất đầu ra, thực hiện thay đổi λ , sử dụng thuật toán Baum-Welch).

2.2. ỨNG DỤNG MÔ HÌNH MARKOV ẨN TRONG NHẬN DẠNG HÀNH ĐỘNG CỦA CON NGƯỜI DỰA TRÊN CẤU TRÚC ACCLAIM

2.2.1. Giới thiệu bài toán

2.2.2. Hệ thống nhận dạng hành động đề xuất



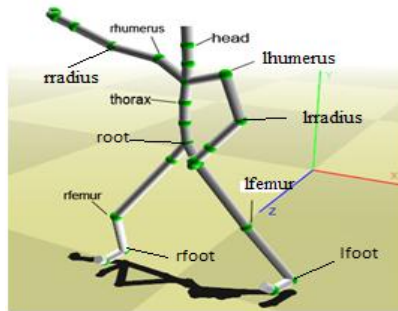
Hình 2.4: Hệ thống nhận dạng hành động đề xuất

2.2.3. Dữ liệu đầu vào

Dữ liệu đưa vào là một tập tin hiển thị các hành động được lưu dưới dạng .c3d. Các tập tin .c3d được tạo nên bằng công nghệ Mocap. Cấu trúc C3D rất phức tạp nên đề xuất sử dụng cấu trúc Acclaim. Cấu trúc Acclaim bao gồm 2 loại tập tin ở định dạng .asf/.amc. Tập tin .asf chứa thông tin chi tiết của khung xương. Tập tin .amc chứa các thông tin chuyển động của các đoạn xương. Dữ liệu dưới dạng văn bản của 2 tập amc/asf sẽ dễ đọc và dễ sử dụng hơn rất nhiều.

2.2.4. Trích chọn đặc tính

Việc trích chọn các đặc tính của dữ liệu rất quan trọng cho việc huấn luyện theo mô hình, với kích thước dữ liệu đã được chuẩn hóa theo một khuôn dạng, khi thực hiện việc trích chọn các đặc tính để sử dụng cho hệ thống nhận dạng gồm rất nhiều cách trích chọn khác nhau. Trong mô hình HMM ở luận văn này tôi thực hiện công việc trích chọn 11 khớp trong 30 khớp [hình phụ lục 1] của mô hình Acclaim để thực hiện công việc nhận dạng của mình. Và đồng thời đưa ra lựa chọn thích hợp với số khớp trích chọn để được kết quả nhận dạng tốt hơn. Trong luận văn sử dụng 11 khớp: “root”, “thorax”, “head”, “rhumerus”, “lhumerus”, “rfemur”, “lfemur”, “lfoot”, “rfoot”, “rradius”, “lradius”.



Hình 2.5. Tên danh sách của 11 khớp trích chọn

2.2.5. Lượng tử hóa vector dựa trên K-mean

Để chuyển các dữ liệu liên tục thành dữ liệu rời rạc cho HMM rời rạc, luận văn này đề xuất sử dụng phương pháp lượng tử hóa vector. Quá trình lượng tử hóa vector (VQ) cho phép các mô hình rời rạc biểu diễn những quan sát liên tục bằng cách ánh xạ mỗi quan sát liên tục thành một mẫu tượng nằm trong bảng ký hiệu rời rạc. Cụ thể, các vector liên tục m -chiều được phân bố thành một số tế bào (codebook) và ánh xạ một quan sát liên tục thành một nhãn của tế bào mà nó nằm trong đó. Tiến trình VQ cho phép mô hình rời rạc, với dữ liệu huấn luyện không lớn, biểu diễn những quan sát liên tục.

Nhiều độ đo khoảng cách khác và các hàm tiêu chuẩn khác đã được đề nghị. Tuy nhiên luận văn này chỉ dùng phương pháp phân nhóm thông thường nhất – thuật toán k -means.

Trong luận văn việc lượng tử hóa vector là một quá trình lập bản đồ của một chuỗi vector rời rạc gồm m chiều liên tục [14] $\mathbf{O} = \{\mathbf{v}_1, \dots, \mathbf{v}_T\}$, $\mathbf{v}_t \in \mathbf{R}^m$ một chuỗi kích thước của chỉ số codebook $\hat{\mathbf{O}} = \{\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_T\}$, $\hat{\mathbf{v}}_t \in \mathbf{N}$ tại đó codebook $\mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_K\}$, $\mathbf{c}_k \in \mathbf{R}^m$ và K trọng tâm của \mathbf{c}_i . Việc chọn một chuỗi liên tục tới codebook là tìm khoảng cách tối thiểu nếu codebook \mathbf{C} được sinh ra.

$$\hat{\mathbf{v}}_i = \operatorname{argmin}_k d(\mathbf{v}_i, \mathbf{c}_k), \forall i \in [1, \dots, T] \quad (2.6)$$

Tại đó $d(\mathbf{v}_i, \mathbf{c}_k) = \|\mathbf{v}_i - \mathbf{c}_k\|^2$ là khoảng cách Euclide bình phương. Có rất nhiều cách để tạo ra các codebook. Phần này mô tả một phương pháp cơ bản để tạo ra \mathbf{C} dùng k -means clustering.

Cho một tập huấn luyện $\mathbf{S}_{train} = \{\mathbf{O}_1, \dots, \mathbf{O}_N\}$, tại đó $N = |\mathbf{S}_{train}|$ là số lượng mẫu $\mathbf{O}_i = \{\mathbf{v}_{i,1}, \dots, \mathbf{v}_{i,T}\}$, $\mathbf{v}_{i,t} \in \mathbf{R}^m$. Một codebook \mathbf{C} sẽ được tính bằng cách tối giản những công việc sau:

$$\min_{u, \mathbf{c}} \sum_{k=1}^K \sum_{i=1}^N \sum_{t=1}^T u_{k,i,t} d(\mathbf{v}_{i,t}, \mathbf{c}_k)$$

$$\text{s.t. } \sum_{k=1}^K u_{k,i,t} = 1, u_{k,i,t} \in \{0,1\}, \quad (2.7)$$

Ở đó $d(\mathbf{v}_{i,t}, \mathbf{c}_k) = \mathbf{v}_{i,t} - \mathbf{c}_k^2$ là khoảng cách Euclide bình phương giữa vector $\mathbf{v}_{i,t}$ và các k^{th} trọng tâm codebook \mathbf{c}_k .

2.2.6. Nhận dạng

Luận văn này tập trung vào khả năng giải bài toán thứ 3 của HMMs rời rạc. Nó có nghĩa là phương pháp này sử dụng DHMMs cho phân loại thứ tự. Các DHMMs đã được xác định bởi các thiết lập sau đây của tham số:

$$\lambda = \{A, B, \pi\}$$

Trong đó: A là phân bố xác suất chuyển trạng thái được đưa ra trong các hình thức của một ma trận $A = \{a_{ij}\}$. B là biểu tượng của sự quan sát (index codebook) phân bố xác suất được đưa ra trong các hình thức của một ma trận. $B = \{b_j(k)\}$ và π là sự phân bố trạng thái ban đầu.

Để sử dụng DHMMs, các quan sát liên tục $O = \{v_1, \dots, v_T\}$, $v_t \in \mathbb{R}^m$ là vector được lượng tử hóa nhưng bộ những chuỗi quan sát riêng biệt. $\hat{O} = \{\hat{v}_1, \dots, \hat{v}_T\}$, $\hat{v}_t \in \mathbb{N}$ sử dụng codebook. Trong trường hợp của vấn đề phân loại, ta cần tạo ra một tập các mô hình và chuyên về mỗi mô hình để nhận ra mỗi người trong số các lớp học riêng biệt. Các thông số λ_i của các mô hình i^{th} có thể được trained với các thuật toán Baum-Welch. Sau khi tất cả các mô hình đã được trained, xác suất của chuỗi không rõ lớp có thể được tính cho mỗi mô hình. Như mỗi mô hình chuyên môn trong một lớp nhất định, một trong những kết quả đầu ra có xác suất cao nhất có thể được sử dụng để xác định các lớp học có khả năng đặt cho chuỗi mới.

CHƯƠNG 3

THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

3.1. MÔI TRƯỜNG THỰC NGHIỆM

Để cài đặt chương trình, thực nghiệm và đánh giá kết quả. Chương trình được cài đặt và chạy thử nghiệm trên hệ thống máy tính:

- Phần cứng : CPU core i3 tốc độ 2,53GHz; RAM 4GB.

- Phần mềm : Hệ điều hành Windows 7 (32bit); database từ trường đại học Carnegie Mellon; Ngôn ngữ lập trình C# nằm trong bộ Microsoft Visual Studio 2010, sử dụng một số thư viện mã nguồn mở của Accord.NET.

3.2. DỮ LIỆU THỰC NGHIỆM

Để đánh giá kết quả của việc nhận dạng hành động của con người dựa trên cấu trúc Acclaim. Trong luận văn này chúng ta tiến hành chạy thử nghiệm cơ sở dữ liệu training và testing ở dạng cấu trúc Acclaim của đại học Carnegie Mellon. Luận văn này sử dụng dữ liệu 3D Acclaim [35] với 5 loại hành động sau:

- + Chạy (run)
- + Đi bộ (walk)
- + Jump (nhảy)
- + Múa (dance)
- + Đấm bốc (boxing)

3.3. KẾT QUẢ THỰC NGHIỆM

Luận văn chia dữ liệu ngẫu nhiên thành 2 loại : loại thứ nhất được dùng làm dữ liệu huấn luyện cho giải thuật học máy, loại thứ hai được dùng để kiểm thử như bảng 3.1.

Bảng 3.6. Thống kê dữ liệu huấn luyện và kiểm thử

Hành động	Run	Walk	Jump	Dance	Boxing
Dữ liệu huấn luyện	35	120	62	30	19
Dữ liệu kiểm thử	13	30	23	15	18
Tất cả	48	150	85	45	37
Tổng	365				

Tất cả các frame của các file được bóc tách thuộc tính bằng cách sử dụng hướng quay của khớp đối với 11 bone “root”, “thorax”, “head”, “rhumerus”, “lhumerus”, “rfemur”, “lfemur”, “lfoot”, “rfoot”, “rradius”, “lradius” và đưa về dạng vector $3 * 11 = 33$ phần tử. Sau đó huấn luyện các hành động từ dữ liệu huấn luyện sử dụng HMM rời rạc và có lượng tử hóa vector bằng K-mean với các tham số : state (số trạng thái ẩn), và số symbol của HMM và số nhóm (k) của K-mean sẽ được thay đổi để được mô hình nhận dạng tối ưu nhất.

Luận văn này tiến hành thực nghiệm với các nhận dạng 2 lớp dữ liệu, 4 lớp dữ liệu và 5 lớp dữ liệu.

- Đối với việc nhận dạng 2 lớp dữ liệu, luận văn tiến hành huấn luyện và kiểm thử với 2 loại cặp hành động:

+ Lớp 1 là “chạy” và lớp 2 là “đi bộ”.

+ Lớp 1 là “chạy” và lớp 2 là “đấm bốc”.

- Đối với thực nghiệm nhận dạng 4 lớp dữ liệu : lớp 1 là “chạy”, lớp 2 là “đi bộ”, lớp 3 là “nhảy” và lớp 4 là “múa”.

- Đối với thực nghiệm nhận dạng 5 lớp dữ liệu : lớp 1 là “chạy”, lớp 2 là “đi bộ”, lớp 3 là “nhảy”, lớp 4 là “múa”, và lớp 5 là “đấm bốc”.

3.3.1. Kết quả nhận dạng phân loại nhị phân

- Phân loại giữa “chạy” và “đi bộ” .

Bảng 3.7. Kết quả nhận dạng nhị phân với các giá trị khác nhau của run, walk

Symbol	Frame/s	State	Tỉ lệ nhận dạng
100	120	4	97,01%
100	120	5	95,52%
100	60	4	95,65%
100	60	5	94,91%

Bảng 3.7 thể hiện tỉ lệ nhận dạng của các hành động tương ứng với các giá trị của các tham số như symbol, state, frame/s khác nhau. Mỗi hàng của bảng này là kết quả nhận dạng với một giá trị của tham số symbol, state, frame/s. Với nhận dạng hành động “chạy” và “đi bộ” thì cho tỷ lệ nhận dạng khá cao khi ta thay đổi các thuộc tính. Giữa hai hành động “chạy” và “đi bộ” ta thấy nó có nhiều điểm tương đồng giống nhau về hành động nhưng trong suốt thời gian hành động walk thì lúc nào cũng phải có ít nhất là một chân chạm đất, còn đối với “chạy” thì giữa các bước chạy sẽ có một khoảng thời gian cả hai chân đều không chạm đất lơ lửng trên không. Đồng thời walk thường đánh gót chân đất trước tiên và sau đó gập đầu gối xuống, còn “chạy” thì tiếp đất bằng một nửa phía trước của bàn chân và hay đổ người về phía trước. Trong một khoảng thời gian thì “chạy” sẽ đi được khoảng cách xa hơn “đi bộ” nên việc nhận dạng dễ dàng và cho kết quả cao hơn. Việc thay đổi state thì tỉ lệ nhận dạng cũng không thay đổi nhiều.

Tiếp theo ta thử nghiệm ở 2 loại dữ liệu giữa “chạy” (run) với “đấm bốc” (boxing).

*Bảng 3.9. Kết quả nhận dạng nhị phân với các giá trị khác nhau của
run, boxing*

Symbol	Frame/s	State	Tỉ lệ nhận dạng
100	120	4	86,36%
100	120	5	79,54%
100	60	4	84,78%
100	60	5	78,26%

Ở việc nhận dạng giữa “chạy” và “đấm bốc” này thì ta thấy tỉ lệ nhận dạng đã giảm xuống so với giữa “chạy” và “đi bộ”. Vì dữ liệu chỉ trích chọn 11 xương ở hình 2.5 nên việc nhận dạng hành động “đấm bốc” còn nhiều hạn chế ở một số tư thế hành động và vì trong hành động “đấm bốc” thì nó vẫn bao gồm có hành động “chạy”, “đấm bốc” sử dụng kết hợp với di chuyển chân, đầu và thân mình. Đồng thời, hành động “đấm bốc” phức tạp và nhiều cử chỉ hơn “chạy”, “đấm bốc” đa số là cử chỉ ở tay, tốc độ hành động nhanh. Nên tỉ lệ nhận dạng của “đấm bốc” bị giảm.

3.3.2. Kết quả nhận dạng với 4 lớp dữ liệu

Nhận dạng với 4 lớp dữ liệu: lớp 1 là “chạy” (run), lớp 2 là “đi bộ” (walk), lớp 3 là “nhảy” (jump) và lớp 4 là “múa” (dance).

*Bảng 3.11. Kết quả nhận với các giá trị khác nhau của 4 hành động
run, walk, jump, dance*

Symbol	Frame	State	Run	Walk	Jump	Dance	Tỉ lệ nhận dạng
100	120	4	90,90%	94,64%	91,30%	92%	93,04%
100	120	5	90,90%	92,85%	91,30%	100%	93,91%
50	120	4	81,81%	89,28%	60,86%	88%	82,60%

50	120	5	90,90%	89,28%	82,60%	96%	89,56%
100	60	4	84,61%	92,86%	100%	100%	94,87%
100	60	5	84,61%	92,85%	96,65%	96%	93,16%
100	60	4	84,61%	91,07%	69,56%	96%	97,17%
50	60	5	76,92%	81,14%	82,61%	100%	85,47%

Bảng 3.12 là kết quả nhận dạng của 4 hành động “chạy”, “đi bộ”, “nhảy”, và “múa” lần lượt với các thông số symbol, state và frame/s khác nhau. Nên ta thấy hành động “múa” và “chạy” có sự tương quan nhau trong việc nhận dạng. Trong 4 hành động thì ta nhận thấy cả 4 hành động đều tương đồng nhau trong hành động, “múa” là sự kết hợp ở nhiều hành động khác nhau như “chạy”, “đi bộ”. Ở đây việc nhận dạng 4 hành động này thì tỉ lệ nhận dạng cho kết quả cao.

3.3.3. Kết quả nhận dạng với 5 lớp dữ liệu

Nhận dạng với 5 lớp dữ liệu: lớp 1 là “chạy” (chạy), lớp 2 là walk (đi bộ), lớp 3 là jump (nhảy), lớp 4 là dance (múa) và lớp 5 là “đấm bốc” (đấm bốc).

Bảng 3.13. Kết quả nhận với các giá trị khác nhau của 5 hành động run, walk, jump, dance, boxing

Symbol	Frame	State	Run	Walk	Jump	Dance	Boxing	Tỉ lệ nhận dạng
100	120	4	100%	92,86%	95,65%	96%	42,11%	87,31%
100	120	5	81,82%	89,29%	82,60%	100%	52,63%	84,33%
50	120	4	72,73%	82,14%	69,57 %	92%	52,63%	76,87%
50	120	5	81,82%	92,86%	73,91 %	92%	36,84%	80,6%

100	60	4	84.62%	89,29%	82,61 %	100%	57,89%	85,29%
100	60	5	84.62%	92,86%	95,65 %	100%	52,63%	88,24%
50	60	4	100%	92,86%	81,82%	72%	36,84%	78,4%
50	60	5	66.67%	80,36%	86.36%	64.00%	26,32%	69,6%

Bảng 3.13 là kết quả nhận dạng chi tiết của 5 hành động “chạy”, “đi bộ”, “nhảy”, “múa” và “đấm bốc” lần lượt với các thông số symbol, state và frame/s khác nhau. Ở việc nhận dạng này ta thấy được tỷ lệ nhận dạng của “đấm bốc” rất là thấp 26,32%. Lý do cho việc nhận dạng có tỷ lệ thấp này có thể là các động tác của “đấm bốc” có nhiều điểm tương đồng với “chạy” và “nhảy”, hành động “đấm bốc” phức tạp và nhiều cử chỉ hơn các hành động còn lại, “đấm bốc” đa số là cử chỉ ở tay, tốc độ hành động nhanh, độ nghiêng của khung nhiều, trong một chuỗi hành động, “đấm bốc” thì nó vẫn có những hành động như “chạy”, “nhảy”, “đi bộ”... Lý do nữa là do dữ liệu nhận dạng của “đấm bốc” ít nhưng số frame của hành động trên mỗi file .amc lại nhiều. Vì vậy để “đấm bốc” có tỷ lệ nhận dạng cao hơn thì ta cần trích chọn các vị trí các xương ở tay và cổ tay.

3.3. ĐÁNH GIÁ KẾT QUẢ NHẬN DẠNG

Bài toán nhận dạng có nhiều yếu tố phụ thuộc, nên độ chính xác cũng phần nào ảnh hưởng đến kết quả của bài toán, nếu việc trích chọn các đặc tính tốt để xây dựng bộ dữ liệu huấn, hệ thống sẽ đạt kết quả tốt hơn. Để tìm được tham số tốt nhất cho hệ thống, thì số lần thử nghiệm phải tăng lên rất nhiều.

Từ các kết quả thực nghiệm cho thấy, khi dữ liệu tương đối đồng bộ, kết quả nhận dạng với độ chính xác cao. Nhưng với những hành động có độ tương đồng khá cao mà dữ liệu để huấn luyện rất ít nên tỉ lệ nhận dạng sẽ khá thấp, tiêu biểu như hành động “đấm bốc” ở bảng 3.13. Vì vậy cần phải tăng cường huấn luyện với bộ dữ liệu lớn,

khi đó hệ thống sẽ nhận dạng tốt hơn, nhận dạng ở những hành động phức tạp hơn. Với một số kỹ thuật chuẩn bị dữ liệu đầu vào và phân chia tập huấn luyện hợp lý thì tốc độ huấn luyện sẽ được tăng lên đáng kể khi đó hệ thống “học” tốt với dữ liệu huấn luyện nhưng với những đầu vào khác hệ thống cho kết quả không chính xác.

KẾT LUẬN

Trong luận văn này trình bày tổng quát những kiến thức và nền tảng trong lĩnh vực nhận dạng. Hiện nay, lĩnh vực nhận dạng đang được các nước trên thế giới phát triển và áp dụng đã đạt được những thành tựu đáng kể.

Luận văn đã nghiên cứu các đặc trưng của nhận dạng, thuật toán nhận dạng... và được lựa chọn rất kỹ lưỡng. Trong đó thuật toán nhận dạng bằng mô hình Markov ẩn được lựa chọn là phương pháp nhận dạng hành động trong luận văn này vì tính chính xác và tốc độ nhận dạng cao. Từ mô hình Markov ẩn được ứng dụng nhiều trong nhận dạng ký hiệu ngôn ngữ, tiếng nói, cử chỉ tay (là những nhận dạng về hình ảnh động, âm thanh)...tôi đã nghiên cứu mô hình Markov ẩn, cơ sở toán học của nó và đã biết cách vận dụng trong nhận dạng hành động dựa trên cấu trúc của Acclaim.

Trong quá trình thực hiện đề tài, do có nhiều hạn chế nên kết quả đạt được chưa cao. Luận văn chỉ tập trung vào nghiên cứu các các vấn đề cơ bản của, một số phương pháp nhận dạng trong học máy, đặc biệt là mô hình HMM. Trong thời gian tới sẽ cố gắng phát triển thêm những nội dung:

Xây dựng hệ thống nhận dạng hành động hoàn chỉnh hơn theo mô hình HMM, thực hiện thêm nhiều chức năng như: nhận dạng hành động dữ liệu từ file ảnh có sẵn (offline), nhận dạng nhiều hành động xảy ra đồng thời hiển thị các hành động sau khi đã nhận dạng hoàn tất.