

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NHA TRANG**



TRẦN THIÊN LONG

**XÂY DỰNG MÔ HÌNH HỌC MÁY CÓ HUẤN LUYỆN
PHỤC VỤ VIỆC ĐỊNH DANH SỬ DỤNG DỮ LIỆU 2D VÀ 3D**

LUẬN VĂN THẠC SĨ

Ngành đào tạo: Công Nghệ Thông Tin

Mã ngành: 8480201

KHÁNH HÒA - 2024

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC NHA TRANG**



TRẦN THIÊN LONG

**XÂY DỰNG MÔ HÌNH HỌC MÁY CÓ HUẤN LUYỆN
PHỤC VỤ VIỆC ĐỊNH DANH SỬ DỤNG DỮ LIỆU 2D VÀ 3D**

LUẬN VĂN THẠC SĨ

Ngành:	Công nghệ thông tin
Mã số:	8480201
Mã số học viên:	61CH093
Quyết định giao đề tài:	670/QĐ-ĐHNT, ngày 02/07/2021
Quyết định thành lập hội đồng:	
Ngày bảo vệ:	
Người hướng dẫn khoa học:	
TS. Đinh Đồng Lương	
Chủ tịch Hội đồng	
Phòng ĐT Sau đại học:	

KHÁNH HÒA – 2024

LỜI CAM ĐOAN

Tôi xin cam đoan mọi kết quả của đề tài: “Xây dựng mô hình học máy có huấn luyện phục vụ việc định danh sử dụng dữ liệu 2D và 3D” là công trình nghiên cứu của cá nhân tôi và chưa từng được công bố trong bất cứ công trình khoa học của tác giả nào khác cho tới thời điểm này.

Khánh Hòa, ngày tháng năm 2024

Tác giả luận văn

(ký và ghi rõ họ tên)

LỜI CẢM ƠN

Trong suốt quá trình thực hiện và hoàn thành luận văn “*Xây dựng mô hình học máy có huấn luyện phục vụ việc định danh sử dụng dữ liệu 2D và 3D*”, tác giả đã nhận được sự chỉ bảo và hướng dẫn tận tình về chuyên môn lẫn những kiến thức chuyên sâu từ TS. Đinh Đồng Lương. Luận văn này đã giúp tác giả cập nhật thêm được nhiều kiến thức bổ ích nhằm nâng cao về chuyên môn cũng như ứng dụng vào công việc của mình về sau này. Tác giả xin được gửi lời cảm ơn chân thành, sâu sắc nhất đến thầy Đinh Đồng Lương.

Bên cạnh đó, tác giả cũng xin chân thành cảm ơn quý Thầy, Cô khoa Công nghệ Thông tin trường Đại học Nha Trang đã luôn quan tâm, giúp đỡ, tạo mọi điều kiện thuận lợi cho tác giả trong quá trình nghiên cứu và hoàn thành luận văn.

Một lần nữa, xin chân thành cảm ơn!

Nha Trang, ngày tháng năm 2024

Tác giả luận văn

(Ký và ghi rõ họ tên)

MỤC LỤC

LỜI CAM ĐOAN	ii
LỜI CẢM ƠN	iii
DANH MỤC THUẬT NGỮ VIẾT TẮT	vi
DANH MỤC HÌNH	vii
TRÍCH YẾU LUẬN VĂN.....	viii
MỞ ĐẦU.....	1
CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI.....	2
1.1. Giới thiệu chung	2
1.2. Các nghiên cứu liên quan	3
1.3. Các kiến thức liên quan	5
1.3.1. Trích chọn đặc trưng.....	5
1.3.2. Học máy	7
1.3.3. Học sâu	9
1.3.4. Kỹ thuật HOG (Histograms of Oriented Gradients).....	11
1.3.5. Kỹ thuật Hu Moments	14
1.3.6. Kỹ thuật Haar-like Features	17
1.3.6. Rút trích đặc trưng từ đám mây điểm (Point Cloud)	20
1.3.7. Sử dụng các đặc trưng hình học cho dữ liệu 3D.....	23
1.3.8. Mô hình SVM (Support Vector Machine).....	26
1.3.9. Mô hình mạng nơ-ron tích chập (CNN)	29
1.3.10. Đặc trưng của dữ liệu hình ảnh 2D	31
1.3.11. Đặc trưng của dữ liệu 3D.....	32
CHƯƠNG 2: PHÂN TÍCH BÀI TOÁN ĐỊNH DANH ĐỐI TƯỢNG DỰA TRÊN DỮ LIỆU 2D VÀ 3D.....	35
2.1. Tổng quan về bài toán định danh đối tượng	35
2.2. Quy trình định danh đối tượng sử dụng dữ liệu 2D và 3D	35
2.3. Thu thập dữ liệu 2D và 3D từ Realsense Camera:.....	38
2.4. Trích chọn đặc trưng	39
2.5 Mô hình định danh/phân lớp đối tượng:	41
CHƯƠNG 3: THỬ NGHIỆM MÔ HÌNH VÀ ĐÁNH GIÁ KẾT QUẢ	43
3.1. Kết quả thực nghiệm.....	43
3.1.1. Thu thập dữ liệu 2D và 3D và xử lý	43

3.1.2. Thử nghiệm: Kịch bản 1:	46
3.1.3. Thử nghiệm: Kịch bản 2:	47
3.1.4. Thử nghiệm: Kịch bản 3:	48
3.2. Đánh giá kết quả mô hình học máy	48
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	50
TÀI LIỆU THAM KHẢO.....	51

DANH MỤC THUẬT NGỮ VIẾT TẮT

Thuật ngữ	Nghĩa tiếng Anh	Nghĩa Tiếng Việt
2D	2-Dimensional	Hình ảnh 2 chiều
3D	3-Dimensional	Hình ảnh 3 chiều
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
GNNs	Graph Neural Networks	Mạng nơ-ron đồ thị
HOG	Histogram of Oriented Gradient	Biểu đồ của gradient định hướng
LBP	Local Binary Patterns	Mẫu nhị phân cục bộ
MRI	Magnetic Resonance Imaging	Chụp cộng hưởng từ
NLP	Natural Language Processing	Xử lý ngôn ngữ tự nhiên
RNN	Recurrent Neural Networks	Mạng nơ-ron hồi tiếp
SVM	Support Vector Machine	Máy vector hỗ trợ

DANH MỤC HÌNH

Hình 1.1. Mô tả kết quả của thuật toán phát hiện cạnh sử dụng toán tử Canny	6
Hình 1.2. Lập trình truyền thống và học máy	8
Hình 1.3. Tổng quan về mô hình học máy	8
Hình 1.4. Hình ảnh vận động viên được chia thành các lưới ô vuông	12
Hình 1.5. Mapping độ lớn gradients với các bins.....	13
Hình 1.6. Sơ đồ mô tả chức năng nhận dạng hoa trong ảnh.....	16
Hình 1.7. Cửa sổ trượt được đặt ngay ngắn vừa gọn để nhìn được toàn bộ ảnh .	18
Hình 1.8. Quy trình nhận diện khuôn mặt của thuật toán Haar	19
Hình 1.9. Mô hình SVM	26
Hình 2.1. Quy trình định danh đối tượng sử dụng dữ liệu 2D và 3D	36
Hình 2.2. Đặc trưng HOG được trích xuất từ hình ảnh 2D	39
Hình 2.3. Đặc trưng Hu moments được trích xuất từ hình ảnh 2D	40
Hình 3.1. Bộ dữ liệu khuôn mặt được lấy từ Realsence Camara	43
Hình 3.2. Bộ dữ liệu hình con mèo.....	43
Hình 3.3. Hình ảnh dữ liệu 3D.....	44
Hình 3.4. Giao diện của phần mềm	44
Hình 3.5. Các file mô hình đã được chương trình huấn luyện	45
Hình 3.6. Kết quả dự đoán của phần mềm	46
Hình 3.7. Kết quả dự đoán dữ liệu hình ảnh trong kịch bản 1.....	47
Hình 3.8. Kết quả dự đoán dữ liệu hình ảnh trong kịch bản 2.....	48
Hình 3.9. Kết quả dự đoán dữ liệu hình ảnh trong kịch bản 3.....	48

TRÍCH YẾU LUẬN VĂN

Tên đề tài: “*Xây dựng mô hình học máy có huấn luyện phục vụ việc định danh sử dụng dữ liệu 2D và 3D*”

Họ và tên học viên: **Trần Thiên Long**

Mã học viên: 61CH093

Lớp: Công nghệ thông tin

Từ khóa: Học máy, học sâu, mô hình 2D và 3D

Nội dung tóm tắt:

Trong bối cảnh công nghệ phát triển nhanh chóng, việc sử dụng máy học để giải quyết các vấn đề phức tạp đang ngày càng phổ biến và quan trọng. Đề tài “Xây dựng mô hình học máy có huấn luyện phục vụ việc định danh sử dụng dữ liệu 2D và 3D” nhằm mục tiêu tận dụng sức mạnh của máy học để xây dựng một hệ thống có khả năng nhận diện đối tượng dựa trên dữ liệu hình ảnh 2D và 3D. Dữ liệu được thu thập từ camera RealSense – một thiết bị mạnh mẽ có khả năng cung cấp cả thông tin độ sâu và hình ảnh màu sắc, phục vụ cho việc phân tích chi tiết hơn.

Mục tiêu nghiên cứu:

Mục tiêu nghiên cứu của bài toán là xây dựng mô hình học máy phục vụ việc định danh sử dụng dữ liệu huấn luyện 2D và 3D thông qua nền tảng SDK Intel RealSense nhằm phân tích, trích xuất đặc trưng đáp ứng được các nội dung: Nghiên cứu và tạo ứng dụng nhận diện khuôn mặt dựa trên nền tảng SDK RealSense của Intel và xây dựng mô hình học máy để định danh dữ liệu huấn luyện 2D và 3D.

Đề tài hi vọng sẽ đánh giá hiệu quả của việc ứng dụng dữ liệu huấn luyện 2D và 3D từ nền tảng SDK Intel RealSense trong việc nhận diện và định danh khuôn mặt. Nghiên cứu không chỉ tập trung vào khía cạnh kỹ thuật của việc xây dựng mô hình học máy mà còn phân tích khả năng trích xuất đặc trưng và áp dụng trên các dữ liệu thực tế. Ngoài ra, đề tài cũng có thể góp phần quan trọng vào việc cải thiện an ninh, hỗ trợ các ứng dụng trong xác thực danh tính tại các hệ thống bảo mật hoặc nơi công cộng.

Phương pháp nghiên cứu:

Đề tài nghiên cứu tập trung vào việc thu thập và nghiên cứu các tài liệu liên quan đến các kỹ thuật huấn luyện và định danh dữ liệu nhằm xây dựng nền tảng lý thuyết cho mô hình học máy. Việc nghiên cứu sẽ bao gồm việc phân tích các phương pháp và kỹ

thuật tiên tiến trong xây dựng mô hình máy học để áp dụng hiệu quả vào quá trình định danh dữ liệu.

Phần thực nghiệm của nghiên cứu sử dụng SDK RealSense của Intel để xây dựng mô hình học máy phục vụ mục tiêu định danh dữ liệu thông qua camera. Các dữ liệu thu thập được sẽ được xử lý và đưa vào cơ sở dữ liệu nhằm đảm bảo độ chính xác và tính ứng dụng thực tiễn của mô hình trong quá trình nhận diện và định danh.

Kết quả nghiên cứu:

Qua quá trình nghiên cứu và thực nghiệm, tôi đã xây dựng mô hình máy học phát hiện, phân tích và thu thập dữ liệu định danh khuôn mặt 2D và 3D từ RealSense Camera. Kết quả thực nghiệm cho thấy, với các mô hình được xây dựng trong công cụ đạt độ chính xác cao, đặc biệt khi kết hợp các đặc trưng từ dữ liệu 2D và 3D, thể hiện được tiềm năng trong việc tối ưu hóa hệ thống định danh trong thực tế.

MỞ ĐẦU

Lý do chọn đề tài

Sự phát triển mạnh mẽ của công nghệ thông tin và trí tuệ nhân tạo đã mở ra những tiềm năng to lớn trong nhiều lĩnh vực, đặc biệt là việc nhận diện và định danh đối tượng qua hình ảnh. Trong bối cảnh đó, việc sử dụng máy học kết hợp dữ liệu 2D và 3D để định danh đối tượng đang trở thành một xu hướng nghiên cứu được quan tâm. Camera RealSense là một công cụ mạnh mẽ, cung cấp cả dữ liệu hình ảnh màu sắc (2D) và thông tin độ sâu (3D), giúp cải thiện độ chính xác trong việc nhận diện đối tượng.

Nhằm đáp ứng nhu cầu ứng dụng trong các hệ thống thực tế, như nhận diện đối tượng trong an ninh, giám sát, và hỗ trợ công tác quản lý, tôi quyết định chọn đề tài "Xây dựng mô hình học máy phục vụ việc định danh sử dụng dữ liệu 2D và 3D từ RealSense Camera" làm luận văn tốt nghiệp Thạc sĩ của mình. Đề tài này không chỉ giúp khám phá các ứng dụng mới của học máy mà còn hướng tới việc nâng cao chất lượng phân tích và xử lý dữ liệu đa chiều.

Mục đích nghiên cứu

- Tìm hiểu và nghiên cứu phương pháp thu thập, xử lý dữ liệu 2D và 3D từ RealSense Camera.
- Xây dựng và đánh giá mô hình học máy cho nhiệm vụ định danh đối tượng sử dụng dữ liệu 2D và 3D.

Đối tượng nghiên cứu

- Các mô hình học máy ứng dụng cho dữ liệu hình ảnh 2D và 3D.
- Kỹ thuật thu thập và xử lý dữ liệu từ RealSense Camera.

Phạm vi nghiên cứu

- Kiến trúc và kỹ thuật xử lý dữ liệu của RealSense Camera.
- Các thuật toán học máy trong nhận diện đối tượng từ dữ liệu 2D và 3D.

Phương pháp nghiên cứu

- Phương pháp tổng hợp: Tổng hợp và nghiên cứu các tài liệu liên quan đến học máy, xử lý dữ liệu hình ảnh 2D và 3D, và các nghiên cứu trước đó về camera RealSense.
- Phương pháp thu thập và xử lý thông tin: Thu thập dữ liệu từ RealSense Camera và sử dụng các công cụ phân tích học máy để xây dựng và đánh giá mô hình định danh đối tượng.

CHƯƠNG 1: TỔNG QUAN VỀ ĐỀ TÀI

1.1. Giới thiệu chung

Ngày nay, với sự phát triển mạnh mẽ của nền khoa học máy tính và công nghệ thông tin, các hệ thống nhận diện và phân tích dữ liệu ngày càng trở thành một phần thiết yếu trong đời sống hiện đại. Những ứng dụng của việc nhận dạng và xử lý dữ liệu hình ảnh từ camera không chỉ giới hạn trong việc định danh các đối tượng mà còn được triển khai rộng rãi trong các lĩnh vực khác như bảo mật, giám sát an ninh, dịch vụ khách hàng, điều khiển robot tự động, và thậm chí trong lĩnh vực y tế. Đặc biệt, việc sử dụng camera tiên tiến để thu thập dữ liệu hình ảnh nhằm phục vụ các mục đích như phát hiện khuôn mặt, phân tích cử chỉ, xác thực người dùng hay hỗ trợ giám sát từ xa đã và đang đóng góp tích cực vào sự phát triển của xã hội.

Cùng với sự bùng nổ của trí tuệ nhân tạo, việc ứng dụng các mô hình học máy để dự đoán, phân tích và nhận diện ngày càng trở nên phổ biến và quan trọng. Học máy, với khả năng tự động hóa việc học và tối ưu hóa quá trình phân tích dữ liệu từ những nguồn lớn, mang đến những giải pháp ưu việt trong việc dự đoán và định danh các dữ liệu từ hình ảnh. Hệ thống học máy không ngừng cải tiến thông qua việc thu thập và xử lý các tập dữ liệu phong phú, giúp các hệ thống này học và nâng cao độ chính xác trong việc nhận dạng và phân tích thông tin.

Trong bối cảnh hiện tại, sự kết hợp giữa việc thu thập dữ liệu và trí tuệ nhân tạo đang mở ra nhiều cơ hội cho những ứng dụng thực tế trong đời sống. Tuy nhiên, lĩnh vực này vẫn còn mới mẻ và đầy thách thức, đặc biệt khi phải đối mặt với các yêu cầu ngày càng cao về độ chính xác, bảo mật và hiệu năng. Việc thu thập, xử lý và định danh dữ liệu từ các thiết bị ghi nhận hình ảnh và cảm biến tiên tiến như RealSense Camera không chỉ là một thử thách kỹ thuật mà còn là một vấn đề mang tính chiến lược trong nghiên cứu khoa học và ứng dụng thực tiễn.

RealSense Camera, với khả năng cung cấp dữ liệu hình ảnh 2D và 3D, mang lại lợi thế đáng kể trong việc xây dựng các hệ thống nhận diện. Khả năng thu thập dữ liệu đa chiều giúp tăng cường độ chính xác, cho phép các mô hình học máy phân tích dữ liệu từ nhiều góc độ, từ đó cải thiện khả năng định danh và nhận dạng đối tượng. Trong các ứng dụng bảo mật, công nghệ này cho phép xây dựng các hệ thống nhận diện khuôn mặt có độ tin cậy cao, giảm thiểu sai sót và tăng cường bảo vệ người dùng. Trong lĩnh vực

tự động hóa, dữ liệu 2D và 3D giúp hệ thống phân tích chi tiết hơn và phản hồi chính xác hơn trong các môi trường làm việc khác nhau.

Xuất phát từ nhu cầu thực tế, đồng thời nhận thức được tiềm năng ứng dụng lớn từ việc khai thác công nghệ này, đề tài ***“Xây dựng mô hình học máy có huấn luyện phục vụ việc định danh sử dụng dữ liệu 2D và 3D”*** được lựa chọn nhằm nghiên cứu và phát triển các giải pháp hiệu quả cho việc định danh dữ liệu. Thông qua đề tài này, mục tiêu không chỉ dừng lại ở việc xây dựng một mô hình nhận diện mạnh mẽ mà còn là việc đưa công nghệ vào thực tế, phục vụ cho các nhu cầu đời sống hiện đại trong nhiều lĩnh vực khác nhau, từ bảo mật, giám sát an ninh, điều khiển tự động cho đến các ứng dụng thông minh khác trong tương lai.

1.2. Các nghiên cứu liên quan

Sự phát triển vượt bậc của công nghệ cảm biến và các mô hình học máy đã tạo ra một cuộc cách mạng trong lĩnh vực nhận diện và định danh đối tượng. Camera Intel RealSense, với khả năng ghi nhận dữ liệu 2D và 3D, đã trở thành một công cụ đắc lực cho việc phát triển các hệ thống định danh thông minh, mang lại sự chính xác và hiệu quả vượt trội. Các nghiên cứu về lĩnh vực này không ngừng được triển khai nhằm cải thiện khả năng nhận diện, từ đó mở rộng ứng dụng trong nhiều lĩnh vực như bảo mật, giám sát, chăm sóc sức khỏe, và các hệ thống tự động hóa. Dưới đây là một số nghiên cứu tiêu biểu, giúp hình dung rõ hơn về các giải pháp và công nghệ đã được triển khai trong lĩnh vực này.

Kết hợp HOG và CNN trong theo dõi đối tượng đa camera: Nghiên cứu này nhằm tối ưu hóa khả năng theo dõi và phát hiện đối tượng trong môi trường có nhiều camera không chồng lấp vùng quan sát. L. Kalake, Y. Dong, W. Wan and Li Hou (2022) [4], đã sử dụng HOG (Histogram of Oriented Gradients) để trích xuất đặc trưng từ hình ảnh, sau đó kết hợp với mạng nơ-ron tích chập CNN (Convolutional Neural Network) để tăng cường khả năng nhận diện và theo dõi đối tượng. HOG giúp mô hình nhận diện chi tiết các cạnh và cấu trúc của đối tượng, trong khi CNN xử lý dữ liệu phức tạp, cho phép hệ thống hoạt động hiệu quả hơn trong môi trường có độ phức tạp cao như giám sát an ninh hoặc phân tích hành vi đối tượng.

So sánh hiệu quả của các phương pháp Haar-like, HOG và LBP trong nhận diện khuôn mặt: A. Adouani, Wiem Mimoun Ben Henia; Z. Lachiri (2019) [5] thực hiện so sánh chi tiết các thuật toán Haar-like Features, HOG và LBP trong nhiệm vụ nhận diện

khuôn mặt. Kết quả cho thấy việc sử dụng HOG kết hợp với SVM mang lại độ chính xác cao hơn trong nhận diện khuôn mặt, nhờ khả năng mô tả chi tiết các đặc trưng của khuôn mặt như đường biên và cấu trúc tổng thể. Haar-like Features cho thấy hiệu quả trong phát hiện nhanh các đặc trưng cơ bản của khuôn mặt, nhưng không có độ chính xác cao bằng khi xử lý các hình ảnh phức tạp. LBP, trong khi đó, phù hợp với các ứng dụng nhận diện khuôn mặt trong điều kiện thay đổi ánh sáng nhưng có giới hạn trong độ chính xác khi dữ liệu hình ảnh có độ phức tạp cao.

Giám sát thông minh như một dịch vụ mạng biên: từ Haar-Cascade, SVM đến CNN: Mục tiêu chính của nghiên cứu là triển khai các thuật toán nhận diện như Haar-Cascade, HOG kết hợp với SVM và CNN trên các thiết bị mạng biên để phát hiện con người trong thời gian thực. Các thiết bị mạng biên có khả năng xử lý dữ liệu nhanh, nhưng thường bị giới hạn về tài nguyên. Do đó, A. Adouani, Wiem Mimoun Ben Henia; Z. Lachiri (2019) [6] tìm cách tối ưu hóa khả năng nhận diện mà không làm tiêu tốn quá nhiều tài nguyên. Haar-Cascade được dùng để nhận diện nhanh các đặc trưng cơ bản, trong khi CNN xử lý các đặc điểm phức tạp để đảm bảo độ chính xác và khả năng nhận diện trong môi trường thực tế.

So sánh hiệu suất của HOG và Haar Cascade với CNN trong hệ thống an ninh gia đình: MZ Asy'ari, S. Filbert, ZL Sukra (2022) [7] đã so sánh hiệu suất của các thuật toán HOG và Haar Cascade khi kết hợp với CNN trong việc xây dựng hệ thống an ninh gia đình thông minh. RealSense Camera được sử dụng để thu thập dữ liệu hình ảnh 2D và 3D từ môi trường thực tế, giúp tăng cường độ chính xác trong việc phát hiện và nhận diện các đối tượng. HOG cho thấy hiệu quả trong việc mô tả chi tiết các đặc điểm của đối tượng, trong khi Haar Cascade xử lý nhanh các hình ảnh đầu vào. CNN, nhờ khả năng học sâu, giúp cải thiện độ chính xác tổng thể và xử lý dữ liệu từ môi trường phức tạp hơn như ánh sáng kém hoặc góc nhìn khó khăn.

Nhận diện khuôn mặt dựa trên CNN, HOG và Haar Cascade sử dụng Raspberry Pi 4 Model B: Nghiên cứu triển khai các thuật toán nhận diện khuôn mặt như CNN, HOG và Haar Cascade trên nền tảng nhúng Raspberry Pi 4, đồng thời sử dụng RealSense Camera để thu thập dữ liệu hình ảnh 2D và 3D. H Belmajdoub, R Mafamane, YB Karfa, M Ouadou, K Minaoui (2023) đã kiểm tra và so sánh hiệu quả nhận diện trong điều kiện tài nguyên hạn chế. Kết quả nghiên cứu cho thấy rằng, nhờ vào việc tận dụng

dữ liệu đa chiều từ cảm biến RealSense, các mô hình học máy có thể đạt được độ chính xác cao và hiệu suất ổn định ngay cả trên các thiết bị nhỏ gọn.

Trong nghiên cứu này, mô hình sẽ cho phép lựa chọn các phương pháp trích chọn đặc trưng như Harr, HOG... và các phương pháp phân lớp như SVN, KMM. Việc kết hợp linh hoạt giữa các phương pháp trích chọn đặc trưng và các phương pháp phân lớp giúp mô hình có khả năng điều chỉnh và tối ưu theo từng bài toán cụ thể, từ đó nâng cao độ chính xác, hiệu suất và khả năng thích nghi trong môi trường ứng dụng thực tế.

1.3. Các kiến thức liên quan

1.3.1. Trích chọn đặc trưng

Trích chọn đặc trưng là quá trình chuyển đổi tập dữ liệu thô ban đầu thành tập các thuộc tính (features) có thể giúp biểu diễn tập dữ liệu ban đầu tốt hơn, tạo điều kiện để giải quyết các bài toán dễ dàng hơn, giúp tương thích với từng mô hình dự đoán cụ thể, cũng như cải thiện độ chính xác của mô hình dự đoán hiện tại.

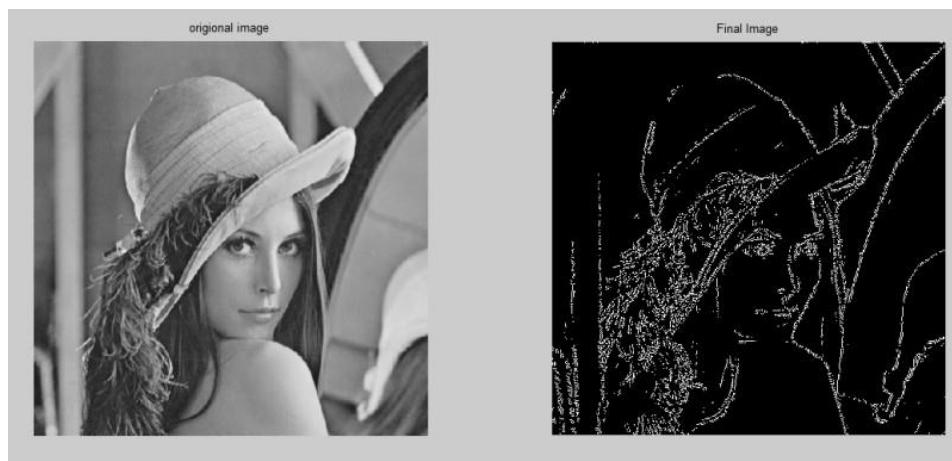
Trích chọn đặc trưng cố gắng biểu diễn tốt nhất tập dữ liệu ban đầu sao cho tương thích với mô hình dự đoán bạn đang sử dụng. Xét bài toán dự đoán sinh viên này có khả năng bỏ học là bao nhiêu phần trăm. Thông thường, ta sẽ sử dụng tất cả các thuộc tính liên quan đến sinh viên đó để áp dụng cho bài toán phân lớp, mà các thuộc tính này thường rất nhiều từ 20-50 cột thuộc tính. Tuy nhiên, khi đưa toàn bộ thuộc tính này vào mô hình phân lớp của mình (ví dụ mô hình cây quyết định), thời gian để máy training rất lâu, đồng thời kết quả dự đoán có độ chính xác thấp. Thay vì làm như vậy, ta sử dụng kỹ thuật Trích chọn đặc trưng để chọn ra một số thuộc tính phù hợp hơn như có vay mượn để đóng học phí không, số điểm đầu vào là bao nhiêu, quá trình tiến bộ trong học tập là bao nhiêu. Ngoài ra, ta có thể thu thập thêm các thuộc tính cần thiết khác để bổ sung vào tập dữ liệu ban đầu như có tham gia nhiều hoạt động ngoại khóa hay không, có được hỗ trợ vào kí túc xá hay không. Thì số thuộc tính đưa vào mô hình phân lớp được giảm đi đáng kể giúp tốc độ để máy training nhanh hơn. Hơn nữa, nhờ biểu diễn tốt tập dữ liệu ban đầu mà độ chính xác mô hình phân lớp được cải thiện đáng kể.

Do tập dữ liệu trên thực tế phức tạp hơn rất nhiều so với các giả định trong nghiên cứu, thêm vào đó các bài toán trong thực tế luôn luôn biến đổi và đòi hỏi những nhà khoa học dữ liệu phải biết thích nghi trong từng trường hợp để đưa ra cách xây dựng mô hình phù hợp nhất.

Vì vậy, việc thêm bớt, chỉnh sửa các thuộc tính cho tập dữ liệu ban đầu sao cho có thể cải thiện tốc độ tính toán cũng như nâng cao độ chính xác của mô hình là một nghệ thuật. Để đạt được trình độ này, đòi hỏi chúng ta phải va chạm nhiều các bài toán thực tế, đồng thời học hỏi từ cộng đồng để nâng cao kiến thức cũng như kinh nghiệm cho bản thân.

Các thuộc tính trong tập dữ liệu ảnh hưởng trực tiếp đến mô hình dự đoán, do đó ta cần xác định tốt cấu trúc của các thuộc tính sao cho diễn đạt hiệu quả nhất bản chất của tập dữ liệu.

Mặc dù bạn không chọn được mô hình dự đoán tốt nhất nhưng ta vẫn có thể đạt được kết quả dự đoán cao. Hầu hết các thuật toán sẽ tự điều chỉnh các thông số phù hợp theo cấu trúc các thuộc tính hiện tại. Tuy nhiên, việc có được tập thuộc tính tinh gọn sẽ góp phần làm đơn giản hoá độ phức tạp tính toán của mô hình nhờ vậy mà tính toán được nhanh hơn và dễ dàng để diễn giải cho người dùng. Ví dụ, khi sử dụng mô hình cây quyết định, nếu ta sử dụng quá nhiều thuộc tính vào quá trình dự đoán, mặc dù cho kết quả rất tốt tuy nhiên, người dùng sẽ rất khó quan sát và diễn giải kết quả dự đoán. Trong quá trình tối ưu hoá tham số, mặc dù không đạt được mục tiêu này nhưng với tập thuộc tính được thiết kế tốt, ta vẫn có thể đạt được kết quả dự đoán cao. Ta không cần phải cực lực tìm kiếm mô hình nào phù hợp nhất cũng như bộ trọng số được tối ưu nhất. Chỉ với tập thuộc tính được thiết kế tốt, ta đã mô tả được tập dữ liệu hiện có cũng như tiếp cận với bài toán thực tế dễ dàng và rõ ràng hơn rất nhiều.



Hình 1.1. Mô tả kết quả của thuật toán phát hiện cạnh sử dụng toán tử Canny [3]

Đối với dữ liệu dạng bảng tính, thuật ngữ observations (các mẫu quan sát) hay instance (các mẫu thể hiện) là các dòng dữ liệu trong bảng tính. Mỗi dòng dữ liệu nào gồm có nhiều variables (biến) hay attributes (thuộc tính) là các cột của bảng tính. [16]

Feature khác với attribute mặc dù cũng mang nghĩa thuộc tính nhưng feature mang nhiều ngữ nghĩa hơn. Ví dụ trong thị giác máy tính, một bức ảnh là một observation gồm nhiều attributes là các điểm pixel. Tuy nhiên, feature ở đây có thể là các đường biên cạnh hay dáng dấp hình học trong bức ảnh. Trong xử lý ngôn ngữ tự nhiên một văn bản hay một tweet sẽ là một observation gồm nhiều attributes là các từ trong đó. Thế nhưng, feature ở đây có thể là một ngữ hay số lượng từ phân bố trong đó. Trong nhận dạng tiếng nói, một người phát ra âm thanh là observation, attributes sẽ là các sóng âm dưới dạng tần số, còn feature sẽ là các nhóm âm thanh được gom lại để phân tích.

Ta có thể thực hiện việc đánh giá độ hữu dụng của các feature để tiền xử lý dữ liệu trước khi đưa vào xây dựng mô hình dự đoán. Các feature này sau khi được ước lượng và đánh giá sẽ có một điểm số để xếp hạng. Những feature có điểm xếp hạng cao sẽ được chọn ra để đưa vào training, còn lại những feature có điểm xếp hạng thấp sẽ bị lược bỏ.

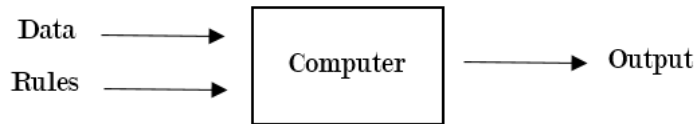
Các feature có tầm quan trọng khác nhau, tùy theo từng bài toán mà có thuộc tính này hữu ích hơn các thuộc tính kia và cần loại bỏ bớt các thuộc tính không liên quan đến bài toán. Có một số feature cần thiết để góp phần vào cải thiện độ chính xác của thuật toán cũng có các feature dư thừa không phù hợp với bài toán hiện tại.

1.3.2. Học máy

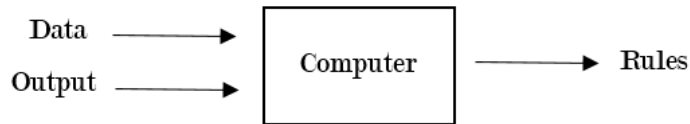
Học máy (Machine Learning) là một lĩnh vực trong trí tuệ nhân tạo, chuyên sâu vào việc phát triển các hệ thống có khả năng học hỏi từ dữ liệu và tự cải thiện hiệu suất thực thi mà không cần lập trình rõ ràng các quy tắc cụ thể. Với sự gia tăng khối lượng dữ liệu và sự phát triển của các phương pháp tính toán mạnh mẽ, học máy đã trở thành một công nghệ mang tính cách mạng, tạo nên những biến đổi quan trọng trong nhiều lĩnh vực như y tế, tài chính, giao thông, và công nghiệp [1].

Học máy có thể hiểu đơn giản là việc cung cấp cho máy tính khả năng "học" từ dữ liệu và tự động điều chỉnh hành vi dựa trên những gì đã học. Điều này khác biệt với lập trình truyền thống, nơi các lập trình viên phải mô tả chính xác cách thức mà chương trình sẽ hoạt động. Thay vào đó, trong học máy, các mô hình và thuật toán được phát triển để học từ dữ liệu đầu vào, tối ưu hóa kết quả dự đoán và đưa ra các quyết định với độ chính xác cao.

Traditional Programming



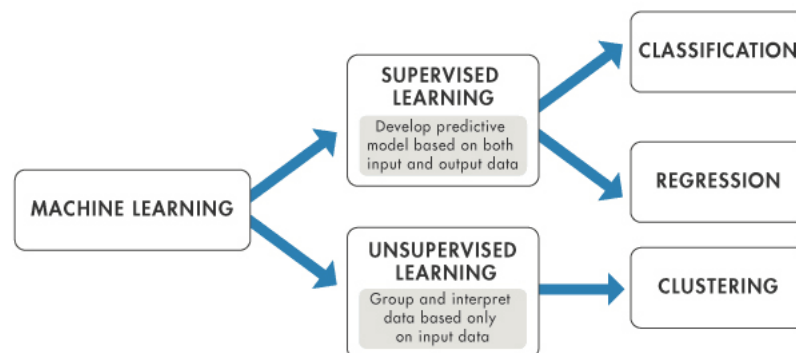
Machine Learning



Hình 1.2. Lập trình truyền thống và học máy [13]

Học máy có thể được chia thành ba loại chính:

- Học có giám sát (Supervised Learning): Là phương pháp học mà dữ liệu đầu vào được cung cấp với các nhãn (labels) rõ ràng. Mục tiêu của các thuật toán trong loại hình này là học cách ánh xạ dữ liệu đầu vào đến các đầu ra chính xác. Ví dụ điển hình bao gồm nhận diện hình ảnh, phân loại văn bản, và dự đoán giá trị bất động sản.
- Học không giám sát (Unsupervised Learning): Ở đây, dữ liệu đầu vào không có nhãn hoặc không có kết quả cụ thể, và hệ thống phải tìm cách phát hiện ra các mẫu hoặc mối quan hệ trong dữ liệu. Một số ứng dụng phổ biến bao gồm phân cụm dữ liệu và giảm chiều dữ liệu.



Hình 1.3. Tổng quan về mô hình học máy [14]

- Học tăng cường (Reinforcement Learning): Đây là loại hình mà mô hình được đào tạo để đưa ra các hành động tối ưu trong một môi trường nhất định nhằm đạt được kết quả tốt nhất. Nó được ứng dụng rộng rãi trong việc huấn luyện các hệ thống điều khiển tự động, robot, và trò chơi điện tử.

Học máy đã tác động mạnh mẽ đến nhiều ngành công nghiệp khác nhau. Trong y tế, học máy giúp chẩn đoán bệnh từ hình ảnh, dự đoán nguy cơ bệnh tật và hỗ trợ cá nhân hóa phác đồ điều trị. Trong tài chính, công nghệ này hỗ trợ phân tích rủi ro, phát hiện gian lận và tối ưu hóa đầu tư. Gần đây, sự kết hợp của học máy và dữ liệu lớn (Big Data) đã cho phép nhiều doanh nghiệp đưa ra các quyết định nhanh chóng, chính xác hơn và tăng cường năng lực cạnh tranh.

1.3.3. Học sâu

Học sâu (deep learning) đã trở thành một trong những lĩnh vực phát triển nhanh nhất và đóng vai trò cốt lõi trong lĩnh vực trí tuệ nhân tạo. Học sâu là một nhánh của học máy tập trung vào việc sử dụng các mạng nơ-ron nhân tạo nhiều lớp nhằm mô phỏng cách thức hoạt động của não người. Nhờ khả năng tự động học và trích xuất đặc trưng từ dữ liệu, học sâu đã tạo ra bước nhảy vọt về khả năng xử lý dữ liệu phức tạp, đặc biệt là trong các lĩnh vực như nhận diện hình ảnh, xử lý ngôn ngữ tự nhiên, lái xe tự động và nhiều ứng dụng thực tiễn khác.

Bản chất của học sâu nằm ở khả năng của các mạng nơ-ron có thể học từ các biểu diễn dữ liệu thông qua các tầng sâu, từ đó cho phép máy tính có khả năng phân tích và đưa ra các quyết định chính xác cao. Các mô hình học sâu hiện đại thường bao gồm hàng triệu, thậm chí hàng tỷ tham số, yêu cầu một lượng lớn dữ liệu và tài nguyên tính toán mạnh mẽ để huấn luyện.

Một trong những cột mốc quan trọng của học sâu là sự ra đời của các kiến trúc mạng nơ-ron như mạng nơ-ron tích chập, mạng nơ-ron hồi tiếp, và gần đây là Transformer. Những kiến trúc này đã giúp tăng cường khả năng nhận diện và phân loại các đối tượng, tối ưu hóa các nhiệm vụ dự đoán chuỗi.

Với sự phát triển không ngừng, học sâu ngày càng chứng minh được vai trò quan trọng của mình trong các lĩnh vực công nghiệp và nghiên cứu khoa học. Chính vì vậy, việc nghiên cứu, tìm hiểu và áp dụng các mô hình học sâu vào các vấn đề thực tiễn trở nên cần thiết hơn bao giờ hết, nhằm đáp ứng nhu cầu phát triển của xã hội và công nghệ trong thời đại số.

Học sâu đã chứng minh hiệu quả vượt trội trong nhiều lĩnh vực khác nhau, nhờ khả năng học và trích xuất đặc trưng từ dữ liệu một cách mạnh mẽ. Dưới đây là một số ứng dụng phổ biến của học sâu:

- Nhận diện hình ảnh và thị giác máy tính: Nhận diện khuôn mặt: Học sâu được sử dụng rộng rãi trong các hệ thống nhận diện khuôn mặt như mở khóa bằng khuôn mặt, bảo mật, giám sát và trong các ứng dụng công nghệ cao; Phân loại hình ảnh: Mô hình học sâu có thể phân loại và nhận diện các đối tượng trong ảnh với độ chính xác cao. Điều này đã được ứng dụng trong y tế (phát hiện khối u, ung thư qua ảnh X-quang), nông nghiệp (phân loại cây trồng, vật nuôi), và nhiều lĩnh vực khác; Xe tự lái: Học sâu giúp xe tự lái nhận diện được người đi bộ, biển báo giao thông và các phương tiện khác trên đường; Xử lý ngôn ngữ tự nhiên (NLP); Dịch máy: Các mô hình học sâu được áp dụng để cải thiện hiệu suất và chất lượng dịch ngôn ngữ tự động, như Google Dịch; Trợ lý ảo: Các trợ lý như Siri, Google Assistant và Alexa sử dụng học sâu để hiểu và trả lời các câu hỏi của người dùng.

- Phân tích văn bản: Học sâu được ứng dụng trong việc phân tích cảm xúc, nhận diện chủ đề hoặc trích xuất thông tin từ văn bản; Xử lý giọng nói và nhận diện giọng nói; Chuyển giọng nói thành văn bản (Speech-to-Text): Học sâu giúp các hệ thống chuyển đổi giọng nói sang văn bản với độ chính xác cao, ứng dụng trong việc chuyển biên cuộc họp hay tạo bản ghi âm; Nhận diện giọng nói: Công nghệ này được ứng dụng trong việc xác thực người dùng qua giọng nói và điều khiển thiết bị thông qua lệnh thoại.

Lĩnh vực Y tế: Chẩn đoán hình ảnh: Học sâu hỗ trợ trong việc phát hiện bệnh thông qua hình ảnh y khoa như X-quang, MRI giúp chẩn đoán bệnh nhanh chóng và chính xác hơn; Dự đoán bệnh và điều trị cá nhân hóa: Các mô hình học sâu giúp phân tích dữ liệu bệnh nhân và dự đoán các bệnh tiềm ẩn, từ đó cung cấp liệu pháp điều trị phù hợp.

- Lĩnh vực tài chính: Phát hiện gian lận: Học sâu giúp nhận diện các giao dịch bất thường và phát hiện gian lận trong thời gian thực; Dự báo thị trường: Các mô hình dự đoán giá chứng khoán, xu hướng thị trường và phân tích rủi ro sử dụng học sâu.

- Lĩnh vực giải trí: Tạo nội dung tự động: Học sâu có thể được sử dụng để tạo ra văn bản, nhạc, hình ảnh hoặc video mới, chẳng hạn như các ứng dụng tạo bài hát theo phong cách nhạc cụ thể hoặc hình ảnh nghệ thuật; Đề xuất nội dung: Các dịch vụ như Netflix, YouTube, và Spotify sử dụng học sâu để phân tích sở thích người dùng và đưa ra các gợi ý nội dung phù hợp.

Thương mại điện tử: Tối ưu hóa trải nghiệm khách hàng: Học sâu được áp dụng để cá nhân hóa trải nghiệm mua sắm, hiển thị các sản phẩm được gợi ý dựa trên lịch sử

tìm kiếm và mua sắm của người dùng; Phân tích hành vi người dùng: Các mô hình học sâu giúp các doanh nghiệp hiểu rõ hơn về hành vi người dùng, từ đó điều chỉnh chiến lược kinh doanh hiệu quả hơn.

Sản xuất và Robot tự động hóa: Robot công nghiệp: Học sâu được sử dụng để điều khiển robot tự động hóa trong sản xuất, từ việc lắp ráp đến kiểm tra chất lượng sản phẩm; Bảo trì dự đoán: Phân tích dữ liệu từ cảm biến của các thiết bị để dự đoán và ngăn chặn hỏng hóc.

An ninh mạng: Phát hiện mã độc: Học sâu giúp nhận diện các mã độc hoặc các hành vi đáng ngờ trên mạng, nâng cao khả năng bảo mật hệ thống; Hệ thống phòng thủ tự động: Phát hiện và ngăn chặn các cuộc tấn công mạng trong thời gian thực.

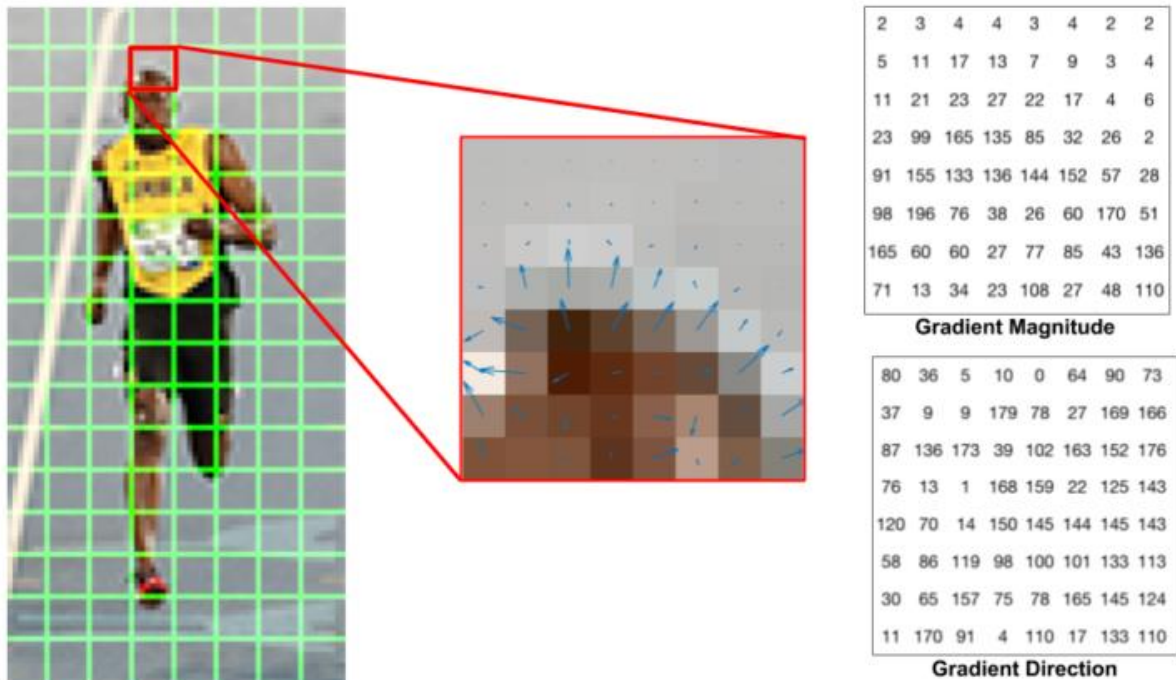
1.3.4. Kỹ thuật HOG (Histograms of Oriented Gradients)

Ý tưởng kỹ thuật HOG (Histograms of Oriented Gradients) được đề xuất dựa trên nguyên tắc cơ bản là sự phân bố hướng của gradient trong hình ảnh có thể mô tả hình dạng của đối tượng. Mục tiêu chính của HOG là trích xuất các đặc trưng của hình ảnh thông qua việc phân tích các gradient tại mỗi pixel, từ đó tạo ra một tập hợp các đặc trưng có khả năng nhận diện các đối tượng cụ thể, như khuôn mặt.

Trong bài toán nhận diện khuôn mặt, HOG giúp phát hiện các đặc trưng như cạnh của mắt, mũi, miệng và đường viền của khuôn mặt bằng cách sử dụng gradient theo các hướng khác nhau. Việc tổng hợp các gradient theo các hướng này giúp tạo ra một mô hình mô tả chi tiết hình dạng khuôn mặt, bất kể các yếu tố như ánh sáng, góc nhìn hay kích thước.

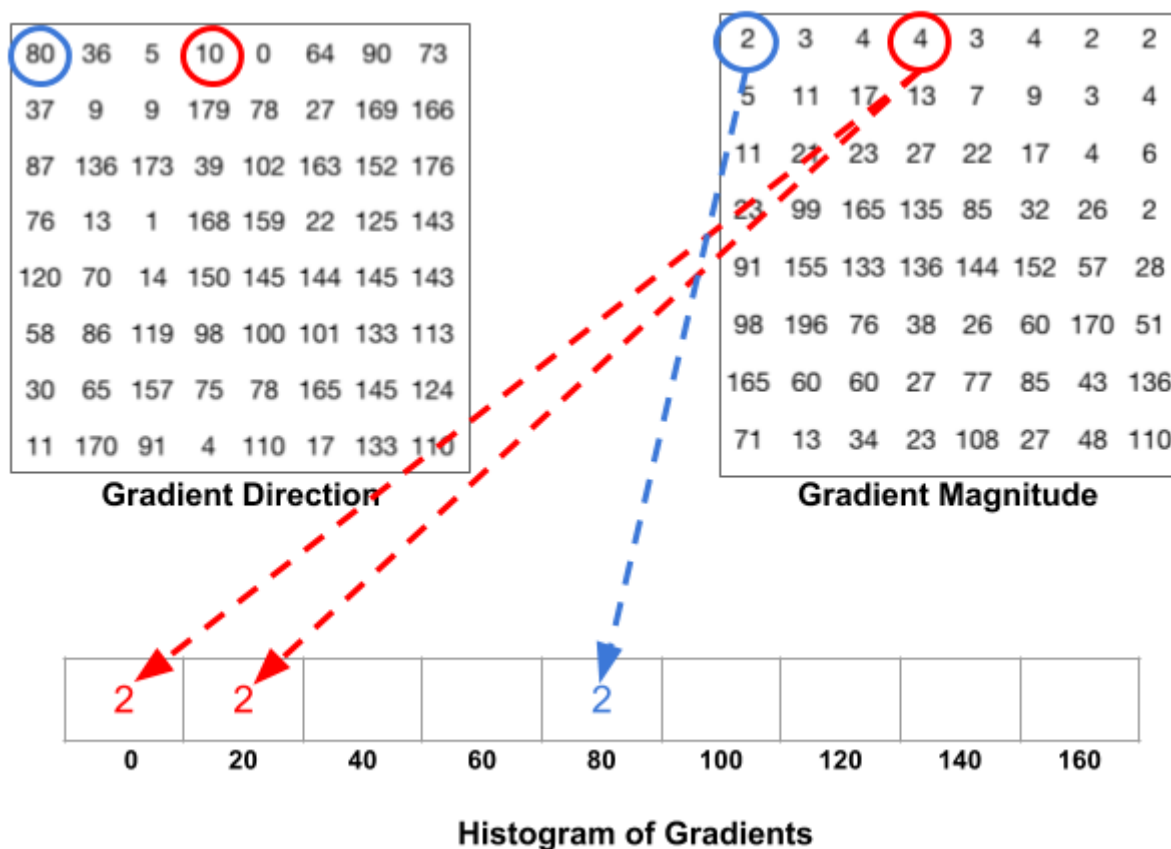
- Thuật toán HOG bao gồm các bước chính sau:

+ Chia hình ảnh thành các ô nhỏ (cells): Hình ảnh được chia thành các ô nhỏ (cells) có kích thước nhất định (thường là 8x8 pixel). Tại mỗi ô, gradient của từng pixel được tính toán bằng cách sử dụng bộ lọc Sobel để xác định độ dốc cường độ ánh sáng.



Hình 1.4. Hình ảnh vận động viên được chia thành các lưới ô vuông, mỗi ô vuông có kích thước 8x8 pixels. Trên mỗi ô chúng ta thực hiện tính đạo hàm thông qua bộ lọc Sobel để thu được 2 ma trận bên phải là gradient magnitude và gradient direction. [12]

+ **Tạo histogram của các gradient:** Trong mỗi ô, các gradient được gom nhóm theo các hướng cụ thể (ví dụ: 0, 45, 90, 135 độ), từ đó tạo ra một biểu đồ histogram của các hướng gradient. Mỗi bin trong histogram sẽ chứa số lượng gradient thuộc hướng tương ứng.



Hình 1.5. Mapping độ lớn gradients với các bins.

+ Chuẩn hóa histogram: Các histogram của các ô sẽ được chuẩn hóa để giảm thiểu sự thay đổi về độ sáng hoặc độ tương phản trong hình ảnh, giúp HOG không bị ảnh hưởng bởi các yếu tố như ánh sáng hoặc bóng tối.

+ Kết hợp các histogram: Các histogram từ các ô nhỏ sẽ được kết hợp thành một vector đặc trưng tổng thể, đại diện cho các đặc điểm hình dạng của toàn bộ khuôn mặt trong hình ảnh.

- Dạng phức tạp

Đối với những bài toán phức tạp như nhận diện khuôn mặt trong các điều kiện ánh sáng hoặc góc nhìn khác nhau, HOG có thể gặp phải một số vấn đề với những phương pháp đơn giản, do đó cần các biến thể của thuật toán để xử lý các tình huống khó hơn. Trong dạng phức tạp hơn của HOG, có thể bổ sung các bước sau:

- Tăng cường chuẩn hóa vùng (block normalization): Thay vì chuẩn hóa riêng lẻ cho từng ô nhỏ, có thể nhóm các ô lại thành một khối lớn hơn và thực hiện chuẩn hóa trên toàn khối (block normalization). Cách này giúp hệ thống trở nên ổn định hơn khi hình ảnh có sự biến dạng hoặc thay đổi góc nhìn.

- Sử dụng đa cấp độ phân giải (multi-scale approach): Ảnh khuôn mặt được chia thành các cấp độ phân giải khác nhau để hệ thống có thể nhận diện đối tượng ngay cả khi đối tượng ở xa hoặc gần, với các kích thước khác nhau. Điều này giúp HOG hoạt động tốt hơn trong các tình huống mà đối tượng có sự thay đổi về kích thước hoặc vị trí trong khung hình.

Kỹ thuật HOG là một công cụ mạnh mẽ và hiệu quả trong việc nhận diện đối tượng, đặc biệt là khuôn mặt, nhờ vào khả năng phát hiện và trích xuất các đặc trưng hình dạng của đối tượng một cách chi tiết. HOG không chỉ hoạt động tốt trong các bài toán nhận diện khuôn mặt mà còn được áp dụng rộng rãi trong nhiều bài toán khác như nhận diện vật thể, phân tích hình dáng, và phát hiện biên.

Tuy nhiên, HOG có một số nhược điểm, như độ nhạy với sự thay đổi về góc nhìn và ánh sáng, đặc biệt khi đối tượng có sự biến dạng lớn. Do đó, trong nhiều trường hợp, HOG cần được kết hợp với các phương pháp học sâu hoặc các kỹ thuật khác để nâng cao hiệu quả nhận diện trong môi trường thực tế phức tạp.

1.3.5. Kỹ thuật Hu Moments

Ý tưởng kỹ thuật **Hu Moments** được dựa trên các moment hình học không đổi, cho phép nhận diện các đối tượng dựa trên hình dạng của chúng, bất kể sự thay đổi về góc quay, dịch chuyển hay tỷ lệ. Các moment này đặc biệt hữu ích trong các ứng dụng nhận diện đối tượng, chẳng hạn như nhận diện khuôn mặt, vì chúng mô tả được hình dáng tổng quát và không bị ảnh hưởng bởi sự biến đổi của hình ảnh.

Đối với bài toán nhận diện khuôn mặt, Hu Moments giúp phát hiện các đặc trưng hình học quan trọng của khuôn mặt như đường viền của khuôn mặt, hình dáng mắt, mũi, miệng. Những đặc điểm này không thay đổi nhiều khi khuôn mặt quay hoặc bị dịch chuyển, điều này làm cho Hu Moments trở nên hiệu quả trong các tình huống nhận diện với nhiều biến đổi.

Moment cơ bản nhất của một hình ảnh có thể được biểu diễn bằng công thức:

$$M = \sum_x \sum_y I(x, y)$$

Tổng này chỉ đơn giản là tính tổng các giá trị điểm ảnh trong hình. Đối với hình ảnh nhị phân, điều này tương đương với việc đếm số lượng điểm ảnh màu trắng hoặc đo diện tích vùng trắng trong hình ảnh. Tuy nhiên, moment này chỉ có thể cung cấp một

thông tin rất đơn giản về hình dạng và không đủ mạnh để phân biệt giữa các hình dạng khác nhau.

Moment phức tạp hơn có thể được định nghĩa như sau:

$$M_{ij} = \sum_x \sum_y x^i y^j I(x, y)$$

Moment này chính là tổng cường độ của các điểm ảnh trong ảnh. Đối với hình ảnh nhị phân, nó có thể được hiểu là:

- Số lượng điểm ảnh trắng ($I(x, y) = 1$).
- Diện tích vùng trắng trong ảnh.

Ví dụ, xét ba hình ảnh nhị phân: hình chữ S ban đầu (S0), chữ S xoay (S5), và chữ K (K0). Moment của hình S và S đã xoay sẽ rất gần nhau, trong khi moment của chữ K sẽ khác biệt. Điều này cho thấy moment có thể nắm bắt được một số đặc điểm của hình dạng.

Tuy nhiên, moment cơ bản này chỉ là một điều kiện cần chứ chưa đủ để xác định hai hình giống nhau. Chúng ta có thể dễ dàng tạo ra hai hình ảnh có moment như nhau nhưng hình dạng hoàn toàn khác.

Ví dụ, để xác định tâm khối (centroid) của một khối hình dạng trong hình ảnh nhị phân, chúng ta có thể sử dụng moment như sau:

$$\bar{x} = \frac{M_{10}}{M_{00}}, \quad \bar{y} = \frac{M_{01}}{M_{00}}$$

Đây là công thức để xác định tọa độ của tâm khối (centroid), giúp xác định vị trí trung tâm của hình dạng trong hình ảnh.

Moment trung tâm là dạng moment đặc biệt được điều chỉnh sao cho bất biến đối với dịch chuyển. Nó được tính bằng công thức:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j I(x, y)$$

Moment trung tâm giúp chúng ta nhận ra rằng dù hình dạng có di chuyển trong hình ảnh, giá trị moment sẽ không thay đổi. Để thêm vào khả năng bất biến đối với tỉ lệ, chúng ta có moment trung tâm chuẩn hóa

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(i+j)/2+1}}$$

Hu Moments là tập hợp 7 giá trị tính toán từ moment trung tâm và có tính chất bất biến đối với các phép dịch chuyển, tỉ lệ và quay. Sáu moment đầu tiên là bất biến với dịch chuyển, tỉ lệ và quay; moment thứ bảy có dấu thay đổi khi hình ảnh bị phản chiếu.

Hu Moments được định nghĩa bằng công thức sau:

$$h_0 = \eta_{20} + \eta_{02}$$

$$h_1 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$h_2 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

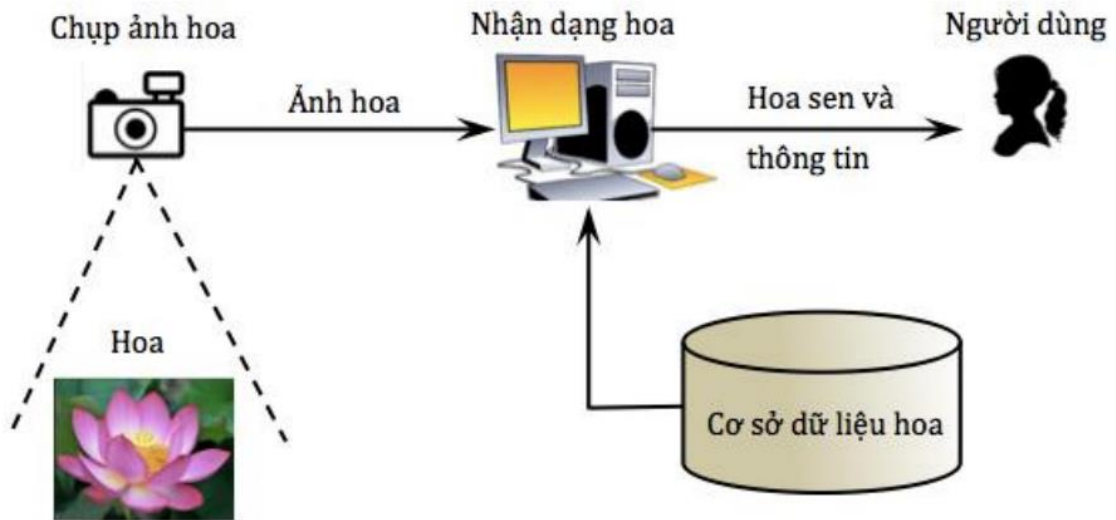
$$h_3 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$h_4 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$h_5 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$h_6 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

Các moment Hu cung cấp thông tin về hình dạng theo cách bất biến với dịch chuyển, tỉ lệ và quay, giúp ta so sánh hình dạng một cách hiệu quả.



Hình 1.6. Sơ đồ mô tả chức năng nhận dạng hoa trong ảnh [2]

Trong các tình huống nhận diện khuôn mặt phức tạp, chẳng hạn như khi khuôn mặt bị biến dạng hoặc có sự thay đổi lớn về góc quay, kỹ thuật Hu Moments có thể mở rộng bằng cách kết hợp thêm các bước xử lý sau:

+ Sử dụng thêm các moment bậc cao: Trong các trường hợp phức tạp, có thể cần tính thêm các moment bậc cao hơn để có được thông tin chi tiết hơn về hình dạng khuôn mặt, giúp hệ thống nhận diện tốt hơn các đặc điểm nhỏ hoặc khó phát hiện.

+ Kết hợp với các kỹ thuật khác: Để tăng độ chính xác khi nhận diện trong các tình huống phức tạp, Hu Moments có thể được kết hợp với các phương pháp khác như HOG hoặc các mô hình học sâu như CNN. Điều này giúp Hu Moments bù đắp nhược điểm khi đối tượng có nhiều biến đổi không gian lớn.

+ Sử dụng phân tích đa tỷ lệ: Hình ảnh khuôn mặt có thể được chia thành nhiều cấp độ phân giải khác nhau để tính toán Hu Moments ở các cấp độ chi tiết khác nhau, từ đó có thể phát hiện các đặc trưng nhỏ mà phương pháp Hu Moments tiêu chuẩn có thể bỏ qua.

Kỹ thuật Hu Moments là một phương pháp mạnh mẽ trong việc nhận diện khuôn mặt nhờ khả năng trích xuất các đặc trưng hình học ổn định và không bị ảnh hưởng bởi các biến đổi hình học. Nhờ vào các moment không đổi, Hu Moments có thể nhận diện được khuôn mặt ngay cả khi đối tượng bị xoay hoặc co giãn.

Tuy nhiên, nhược điểm của Hu Moments là chúng không phản ứng tốt với các biến đổi phức tạp như thay đổi góc nhìn lớn hoặc biến dạng tự nhiên của khuôn mặt trong các điều kiện thực tế khác nhau. Để vượt qua những hạn chế này, việc kết hợp Hu Moments với các kỹ thuật trích xuất đặc trưng khác như HOG hoặc các mô hình học sâu hiện đại có thể giúp tăng cường khả năng nhận diện.

1.3.6. Kỹ thuật Haar-like Features

Ý tưởng kỹ thuật Haar-like Features là một phương pháp hiệu quả để trích xuất các đặc trưng hình ảnh bằng cách phát hiện các vùng sáng và tối trong hình ảnh. Ý tưởng cơ bản của kỹ thuật này là sử dụng các bộ lọc đơn giản dựa trên các đặc trưng hình học như các cạnh, đường viền, hoặc vùng sáng-tối tương phản để phát hiện các đặc điểm cụ thể trên khuôn mặt, chẳng hạn như mắt, mũi, miệng.

Phương pháp này rất hiệu quả trong nhận diện khuôn mặt vì khuôn mặt con người có các đặc trưng cố định về hình dạng và sự phân bố sáng-tối. Ví dụ, vùng mắt thường tối hơn so với trán và gò má, hoặc khu vực mũi có sự thay đổi về độ sáng so với các vùng lân cận. Bằng cách sử dụng các bộ lọc Haar-like, hệ thống có thể phát hiện nhanh chóng các vùng này và sử dụng chúng để xác định vị trí của khuôn mặt.

Các dạng thuật toán Haar-like Features hoạt động dựa trên việc áp dụng các bộ lọc đơn giản lên hình ảnh để phát hiện các đặc trưng hình học. Các bước cơ bản của thuật toán bao gồm:

- Chia hình ảnh thành các vùng: Hình ảnh khuôn mặt được chia thành các vùng nhỏ hơn, tại mỗi vùng này, các đặc trưng Haar được áp dụng để trích xuất thông tin.

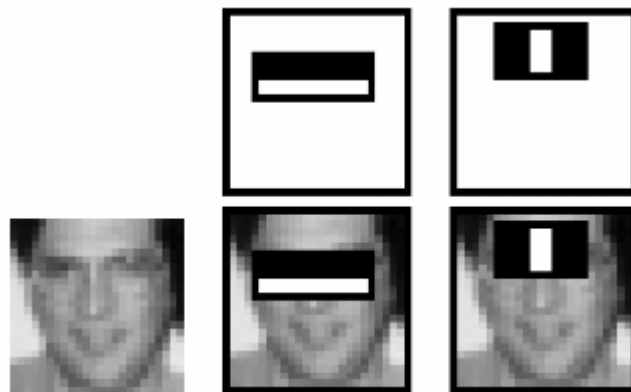
- Bộ lọc Haar: Bộ lọc Haar so sánh sự khác biệt về cường độ sáng giữa các vùng lân cận trong hình ảnh. Một số dạng bộ lọc Haar thông dụng bao gồm:

+ Bộ lọc cạnh (Edge features): Phát hiện các cạnh dọc hoặc ngang bằng cách so sánh sự khác biệt về độ sáng giữa hai vùng kề nhau.

+ Bộ lọc đường (Line features): Phát hiện các đường ngang hoặc dọc bằng cách so sánh sự khác biệt giữa ba vùng liên tiếp.

+ Bộ lọc bốn vùng (Four-rectangle features): So sánh sự khác biệt giữa bốn vùng trong một hình vuông, thường được sử dụng để phát hiện các vùng có sự phân bố sáng-tối phức tạp hơn.

- Tính toán giá trị Haar-like Features: Sau khi áp dụng các bộ lọc, giá trị của mỗi đặc trưng Haar được tính bằng hiệu số giữa cường độ sáng của các vùng tương phản. Ví dụ, sự khác biệt giữa vùng tối của mắt và vùng sáng của gò má sẽ cho ra một giá trị đặc trưng Haar.



Hình 1.7. Cửa sổ trượt được đặt ngay ngắn vừa gọn để nhìn được toàn bộ ảnh [11]

- Cascade Classifier: Các đặc trưng Haar không được áp dụng một lần duy nhất mà theo dạng xếp lớp. Những đặc trưng cơ bản trước tiên sẽ giúp loại bỏ các vùng không có khuôn mặt nhanh chóng, sau đó các đặc trưng phức tạp hơn sẽ được áp dụng để phát hiện chính xác khuôn mặt.

Dạng phức tạp

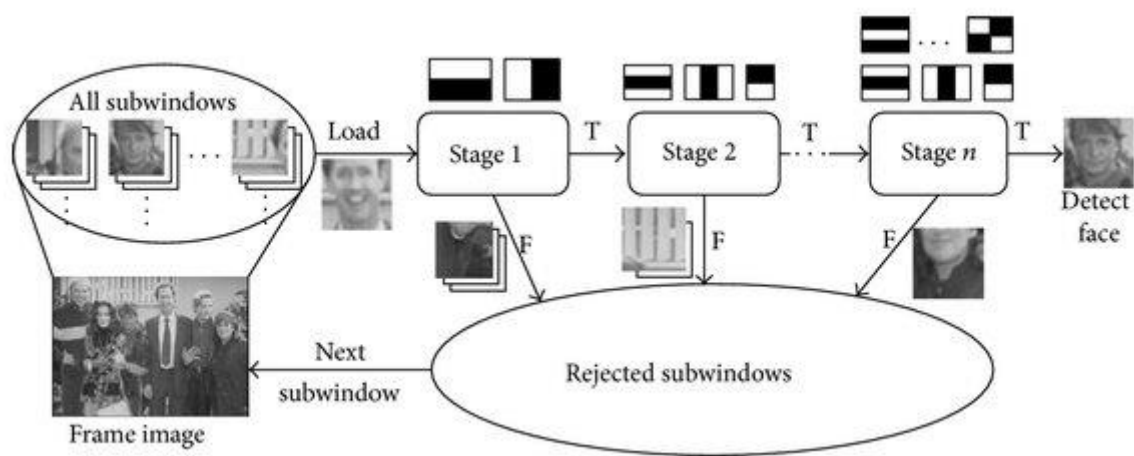
Khi xử lý các trường hợp nhận diện khuôn mặt trong các điều kiện phức tạp như góc quay hoặc thay đổi ánh sáng, kỹ thuật Haar-like Features cần phải được mở rộng bằng cách thêm các yếu tố như:

- Phân tích đa tỷ lệ: Hình ảnh được phân tích ở nhiều kích thước khác nhau để đảm bảo rằng các khuôn mặt ở nhiều khoảng cách khác nhau trong khung hình đều có thể được phát hiện. Điều này giúp Haar-like Features phát hiện khuôn mặt ở các tỷ lệ lớn nhỏ khác nhau.

- Cascade Classifier nâng cao: Khi đối mặt với những hình ảnh phức tạp, hệ thống cần sử dụng nhiều lớp phân loại hơn. Các đặc trưng phức tạp hơn (chẳng hạn như bộ lọc bốn vùng hoặc các tổ hợp đặc trưng) được áp dụng để phân biệt rõ hơn giữa khuôn mặt và các vật thể khác trong hình ảnh.

Các tác giả cũng đề xuất một phương pháp được gọi là ảnh tích phân có thể đạt được tính toán tương tự bằng cách thực hiện các phép toán chỉ trên bốn điểm ảnh. Có thể sử dụng hàng nghìn hình ảnh được gán nhãn để chuyển đổi chúng thành bản đồ tính năng tầng Haar và đào tạo nó bằng cách sử dụng mô hình học máy để tạo bộ phân loại nhận diện khuôn mặt.

Để hiểu hơn về cách Haar Cascade hoạt động thì mình sẽ đi sâu hơn về từng bước nhận diện khuôn mặt của Haar Cascade



Hình 1.8. Quy trình nhận diện khuôn mặt của thuật toán Haar

- Bước 1: Hình ảnh (đã được gửi đến bộ phân loại) được chia thành các phần nhỏ (hoặc các cửa sổ con như trong hình minh họa).
- Bước 2: Chúng tôi đặt N không có bộ dò theo cách xếp tầng trong đó mỗi bộ phát hiện sự kết hợp của các loại đặc trưng khác nhau từ các hình ảnh (ví dụ: đường thẳng, cạnh, hình tròn, hình vuông) được truyền qua. Giả sử khi việc trích xuất đối tượng địa lý được thực hiện, mỗi phần phụ được gán một giá trị tin cậy.

- Bước 3: Hình ảnh (hoặc hình ảnh phụ) có độ tin cậy cao nhất được phát hiện dưới dạng khuôn mặt và được gửi đến bộ tích lũy trong khi phần còn lại bị từ chối. Do đó, Cascade tìm nạp khung hình / hình ảnh tiếp theo nếu còn lại và bắt đầu lại quá trình.

Kết luận kỹ thuật Haar-like Features là một phương pháp mạnh mẽ và hiệu quả trong việc phát hiện khuôn mặt, đặc biệt là nhờ vào tính đơn giản và khả năng xử lý nhanh chóng. Bằng cách sử dụng các đặc trưng hình học đơn giản, kỹ thuật này có thể phát hiện các vùng đặc trưng trên khuôn mặt, chẳng hạn như mắt, mũi và miệng, ngay cả khi đối mặt với các biến đổi về ánh sáng và góc nhìn.

Tuy nhiên, hạn chế của Haar-like Features là khó phát hiện chính xác trong những điều kiện rất phức tạp, chẳng hạn như khi khuôn mặt bị che khuất một phần hoặc có sự biến dạng mạnh. Để vượt qua những hạn chế này, kỹ thuật Haar-like Features thường được kết hợp với các phương pháp khác như HOG hoặc mạng nơ-ron tích chập (CNN) để tăng cường độ chính xác trong nhận diện khuôn mặt.

1.3.6. Rút trích đặc trưng từ đám mây điểm (Point Cloud)

Đám mây điểm (Point Cloud) là tập hợp các điểm dữ liệu được sắp xếp trong không gian ba chiều (3D) và được thu thập thông qua các cảm biến như Lidar, máy quét laser 3D, hoặc camera độ sâu như Intel RealSense. Đám mây điểm là cấu trúc dữ liệu phổ biến trong lĩnh vực thị giác máy tính, bản đồ 3D, robot và công nghệ thực tế ảo/ thực tế tăng cường (AR/VR). Trong hệ tọa độ 3D, mỗi điểm chứa thông tin không gian về tọa độ (x, y, z) và có thể bao gồm các thông tin bổ sung như màu sắc, cường độ ánh sáng phản chiếu, hoặc thuộc tính bề mặt. Mặc dù rất hữu ích, đám mây điểm có thể chứa hàng triệu điểm, dẫn đến độ phức tạp cao trong xử lý và phân tích. Do đó, việc rút trích đặc trưng từ dữ liệu này là bước cần thiết để làm giảm độ phức tạp, trích xuất thông tin quan trọng và tối ưu hóa hiệu suất của các thuật toán học máy.

Đặc trưng là các thông tin biểu diễn cho dữ liệu, thường là các thông số hoặc mô hình giúp phân loại, nhận dạng và tái tạo lại đối tượng từ đám mây điểm. Việc rút trích đặc trưng đóng vai trò quan trọng trong xử lý và khai thác đám mây điểm bởi nó giúp đơn giản hóa dữ liệu mà vẫn bảo toàn các thông tin cần thiết. Mục tiêu chính là làm cho dữ liệu trở nên dễ hiểu và có thể xử lý được trong bối cảnh các bài toán khác nhau như nhận diện đối tượng 3D, tái cấu trúc bề mặt, và phân vùng không gian.

Để minh họa quá trình rút trích đặc trưng cục bộ từ đám mây điểm (Point Cloud), hình ảnh dưới đây thể hiện một đám mây điểm 3D của một đối tượng, trong đó các điểm được tô màu dựa trên giá trị của một đặc trưng cục bộ, chẳng hạn như độ cong (curvature) hoặc vector pháp tuyến (normal vector).

Trong hình, mỗi điểm trong đám mây được gán một màu sắc tương ứng với giá trị của đặc trưng cục bộ tại vị trí đó. Việc tô màu này giúp trực quan hóa các thuộc tính hình học cục bộ của bề mặt đối tượng, hỗ trợ trong các ứng dụng như nhận dạng hình dạng, phân đoạn bề mặt và tái tạo mô hình 3D.

Các phương pháp rút trích đặc trưng có thể được phân thành hai nhóm chính: đặc trưng cục bộ và đặc trưng toàn cục.

Đặc trưng cục bộ tập trung vào thông tin xung quanh một điểm cụ thể và vùng lân cận gần nó. Chúng được sử dụng để mô tả các thuộc tính hình học của đối tượng như độ cong, độ nhẵn và hướng bề mặt tại vị trí đó. Một số phương pháp phổ biến bao gồm:

- FPFH (Fast Point Feature Histogram): Đây là một trong những bộ mô tả hình dạng phổ biến nhất để rút trích các đặc trưng hình học cục bộ. FPFH phân tích các mối quan hệ giữa điểm trung tâm và các điểm lân cận, giúp mô tả hướng của
- SHOT (Signature of Histograms of Orientations): Bộ mô tả này kết hợp giữa đặc trưng hình dạng và hướng, cho phép biểu diễn thông tin cục bộ quanh mỗi điểm.
- Normal Vector: Vector pháp tuyến tại mỗi điểm cung cấp thông tin về hướng bề mặt cục bộ và giúp xác định các vùng góc cạnh, cong hoặc nhẵn.
- Curvature (Độ cong): Độ cong của một vùng nhỏ quanh điểm giúp mô tả đặc tính hình học của bề mặt lân cận và phân biệt các vùng cong so với các bề mặt phẳng.

Đặc trưng toàn cục biểu diễn thông tin của toàn bộ đám mây điểm, mô tả cấu trúc tổng thể và hình dạng không gian của đối tượng. Các phương pháp phổ biến bao gồm:

- PointNet và các biến thể: Điển hình như PointNet++, đây là các mạng neural học sâu có khả năng xử lý trực tiếp dữ liệu đám mây điểm mà không cần chuyển đổi sang các dạng lưới hay voxel. PointNet trích xuất đặc trưng toàn cục thông qua một chuỗi các tầng học dựa trên hàm pooling, cho phép tổng hợp thông tin từ tất cả các điểm.

- Voxelization và Octree: Các phương pháp này chia đám mây điểm thành các lưới nhỏ hoặc phân đoạn thành các cấu trúc phân cấp để dễ dàng rút trích các đặc trưng tổng quát.
- Global Descriptors: Một số bộ mô tả khác như Spin Images và Global Histograms cung cấp các cách biểu diễn toàn cục của đám mây điểm thông qua phân tích các mối quan hệ không gian giữa các điểm hoặc cụm điểm.

Trước khi thực hiện rút trích đặc trưng, các kỹ thuật tiền xử lý thường được áp dụng để làm sạch và chuẩn hóa dữ liệu đám mây điểm nhằm đảm bảo tính chính xác và hiệu quả. Các kỹ thuật này bao gồm:

- Lọc nhiễu (Noise Filtering): Loại bỏ các điểm ngoại lai và dữ liệu không mong muốn, giúp tăng độ tin cậy của quá trình rút trích đặc trưng.
- Giảm điểm (Downsampling): Giảm số lượng điểm trong đám mây điểm bằng cách chọn các điểm đại diện, làm giảm tải tính toán mà không làm mất đi thông tin quan trọng.
- Căn chỉnh (Alignment) và Đăng ký (Registration): Đám mây điểm cần được căn chỉnh sao cho chúng khớp với hệ tọa độ chung, đặc biệt quan trọng trong việc kết hợp dữ liệu từ nhiều cảm biến hoặc trong quá trình tái tạo các đối tượng.

Việc áp dụng các phương pháp học sâu trong rút trích đặc trưng từ đám mây điểm đã mở ra những khả năng mới trong phân tích dữ liệu 3D. Một số kỹ thuật tiên tiến được sử dụng nhiều bao gồm:

- Graph Neural Networks (GNNs): GNNs mô hình hóa đám mây điểm dưới dạng đồ thị, cho phép rút trích đặc trưng từ các mối liên kết không gian giữa các điểm.
- Convolutional Neural Networks (CNNs) trên đám mây điểm: Các kỹ thuật như PointCNN và Dynamic Graph CNN cải tiến việc học đặc trưng bằng cách áp dụng các phép biến đổi convolutional đặc biệt, phù hợp với dữ liệu không cấu trúc của đám mây điểm.
- Autoencoders và Generative Models: Các mạng tự mã hóa và mô hình sinh tự động giúp học các đặc trưng tổng quát của đám mây điểm thông qua việc tái tạo và mô phỏng lại dữ liệu.

Ứng dụng của đặc trưng từ đám mây điểm

- Nhận dạng và phân loại đối tượng: Đặc trưng rút trích từ đám mây điểm có thể được sử dụng trong các hệ thống nhận dạng vật thể 3D, phân loại các cấu trúc, hoặc nhận diện hình dạng trong công nghiệp và thương mại.
- Robot và tự động hóa: Các đặc trưng giúp robot nhận diện và xác định môi trường xung quanh, tạo điều kiện cho các hoạt động như điều hướng, lập bản đồ, và tương tác với các đối tượng.
- Kiến trúc, xây dựng và khảo sát: Rút trích đặc trưng từ đám mây điểm giúp xây dựng mô hình 3D chính xác cho các công trình kiến trúc và hạ tầng, hỗ trợ trong việc kiểm tra và lập kế hoạch.
- Y tế và sinh học: Phân tích các cấu trúc y tế 3D từ hình ảnh chụp bằng các thiết bị như MRI.

Việc rút trích đặc trưng từ đám mây điểm là một trong những công đoạn quan trọng trong khai thác và xử lý dữ liệu 3D, nhằm tối ưu hóa thông tin, giúp máy tính và hệ thống học máy hiểu rõ hơn về dữ liệu không gian. Với sự phát triển của công nghệ cảm biến và học sâu, khả năng trích xuất đặc trưng đã được nâng cao đáng kể, mở ra những ứng dụng và hướng đi mới trong tương lai. Nghiên cứu và cải tiến các kỹ thuật này sẽ tiếp tục mang lại những bước đột phá trong lĩnh vực khoa học và công nghiệp.

1.3.7. Sử dụng các đặc trưng hình học cho dữ liệu 3D

Trong xử lý dữ liệu 3D, các đặc trưng hình học đóng vai trò quan trọng trong việc biểu diễn và mô tả các đặc điểm của đối tượng hoặc bề mặt trong không gian ba chiều (3D). Đặc trưng hình học cung cấp thông tin chi tiết về hình dạng, cấu trúc bề mặt và sự biến đổi hình học của đối tượng, giúp các hệ thống máy tính có thể nhận biết, phân loại, và tái tạo dữ liệu một cách chính xác hơn. Hai trong số các đặc trưng hình học phổ biến nhất được sử dụng là Surface normal (vector pháp tuyến bề mặt) và Curvature (độ cong). Chúng không chỉ giúp mô tả chi tiết bề mặt mà còn hỗ trợ trong việc phát hiện các thay đổi hình dạng và phân tích cấu trúc tổng thể của đối tượng.

Vector pháp tuyến bề mặt tại một điểm trong đám mây điểm (Point Cloud) là một vector vuông góc với bề mặt tại điểm đó. Nó thể hiện hướng mà bề mặt “nhìn” ra bên ngoài và cung cấp thông tin quan trọng về đặc tính hình học của bề mặt. Để tính toán vector pháp tuyến, thường áp dụng phương pháp tìm mặt phẳng tiếp xúc gần đúng bằng cách sử dụng các điểm lân cận và tìm một vector vuông góc với mặt phẳng này.

Cách tính toán Surface normal: Tính toán vector pháp tuyến có thể thực hiện thông qua việc lựa chọn một vùng lân cận các điểm quanh điểm đang xét (ví dụ: bằng cách sử dụng phương pháp tìm kiếm k-láng giềng gần nhất hoặc xác định bán kính lân cận). Sau đó, sử dụng kỹ thuật hồi quy tuyến tính hoặc giải bài toán giá trị riêng để xác định mặt phẳng gần đúng với các điểm này, và vector pháp tuyến là vector chính vuông góc với mặt phẳng đó.

Ứng dụng của Surface normal: Nhận diện hình dạng và bề mặt: Surface normal cung cấp thông tin chi tiết về cấu trúc bề mặt, giúp nhận diện các đặc điểm của vật thể như vùng lõm, lồi, hoặc bề mặt góc cạnh.

- Phân đoạn đối tượng: Trong các ứng dụng như thị giác máy tính, việc phân đoạn đối tượng dựa vào hướng vector pháp tuyến giúp tách biệt các phần khác nhau của vật thể dựa trên hướng thay đổi của bề mặt.
- Xử lý đồ họa và ánh sáng: Trong các ứng dụng đồ họa, vector pháp tuyến đóng vai trò quan trọng trong việc xác định cách ánh sáng tương tác với bề mặt, từ đó tạo ra các hiệu ứng ánh sáng và bóng mờ.

Curvature là một đặc trưng hình học mô tả sự thay đổi của bề mặt tại một điểm cụ thể trong không gian 3D. Nó cung cấp thông tin về mức độ "bề cong" hoặc thay đổi hướng của bề mặt tại điểm đó. Curvature giúp mô tả các đặc tính cục bộ của bề mặt, chẳng hạn như góc cạnh, gờ, hoặc các vùng cong. Việc tính toán và sử dụng độ cong có thể giúp phát hiện các điểm đặc biệt hoặc các vùng có tính chất hình học đáng chú ý.

Phân loại độ cong: Có hai loại độ cong chính được sử dụng trong phân tích dữ liệu 3D:

- Độ cong Gaussian: Là tích của hai giá trị độ cong chính tại một điểm, độ cong Gaussian có thể giúp phân biệt giữa các vùng lõm và lồi, hoặc các điểm yên ngựa (vùng mà bề mặt cong theo hai chiều ngược nhau).
- Độ cong trung bình: Là trung bình cộng của hai giá trị độ cong chính, giúp xác định mức độ uốn cong tổng thể của bề mặt tại một điểm.

Cách tính toán Curvature: Việc tính toán độ cong thường liên quan đến phân tích ma trận Hessian hoặc ma trận chính phương của bề mặt tại điểm cần xét. Thông qua việc phân tích các giá trị riêng và vector riêng của ma trận, ta có thể xác định được mức độ cong cục bộ tại mỗi điểm.

Ứng dụng của Curvature:

- Nhận diện các đặc điểm hình học quan trọng: Curvature có thể được sử dụng để phát hiện các điểm góc, gờ hoặc biên trên bề mặt, là những đặc điểm quan trọng trong phân đoạn và nhận diện đối tượng.
- Tái tạo bề mặt và mô hình hóa: Trong các ứng dụng tái tạo bề mặt, độ cong giúp xác định các vùng cần được làm mịn hoặc chỉnh sửa, từ đó cải thiện độ chính xác của mô hình 3D.
- Phân tích hình học và kiểm tra chất lượng: Curvature được sử dụng để kiểm tra tính toàn vẹn của bề mặt, xác định các vùng có sự thay đổi hình học lớn.

Việc kết hợp các đặc trưng Surface normal và Curvature giúp mô tả một cách toàn diện về hình dạng và bề mặt của đối tượng trong không gian 3D. Trong nhiều ứng dụng thực tế, các đặc trưng này không chỉ cung cấp thông tin về hình học mà còn cho phép máy tính hiểu được cấu trúc tổng thể của dữ liệu 3D. Việc rút trích và phân tích các đặc trưng này giúp cải thiện hiệu quả của các hệ thống nhận dạng, phân loại và tái tạo hình ảnh, từ đó tăng cường khả năng ứng dụng của dữ liệu 3D trong các lĩnh vực khác nhau.

Mặc dù Surface normal và Curvature là hai đặc trưng phổ biến và quan trọng, việc tính toán chúng trong dữ liệu 3D phức tạp vẫn đối mặt với nhiều thách thức:

- Độ nhiễu trong dữ liệu: Dữ liệu 3D thu thập từ các cảm biến như Lidar hoặc máy quét laser có thể chứa nhiều điểm nhiễu, làm giảm độ chính xác của đặc trưng hình học.
- Sự không đồng nhất của đám mây điểm: Các điểm trong đám mây điểm có thể không phân bố đều, gây khó khăn trong việc xác định vùng lân cận và tính toán đặc trưng.
- Khối lượng dữ liệu lớn: Xử lý dữ liệu 3D với số lượng lớn điểm đòi hỏi hiệu suất tính toán cao và tối ưu hóa thuật toán.

Ứng dụng thực tiễn của đặc trưng hình học trong dữ liệu 3D

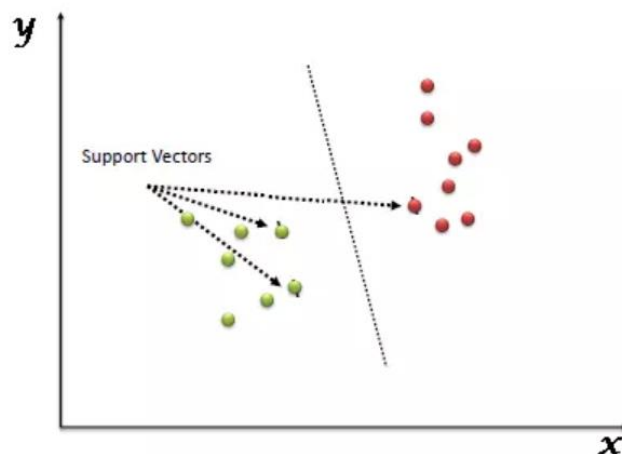
- Robot và tự động hóa: Surface normal và Curvature giúp robot xác định và tương tác với môi trường xung quanh, từ nhận diện vật cản đến thao tác và điều hướng trong không gian 3D.
- Thị giác máy tính: Các đặc trưng hình học hỗ trợ trong việc phân loại, nhận dạng và phân đoạn các đối tượng trong các ứng dụng thị giác.

- Đồ họa và thực tế ảo: Surface normal và Curvature cung cấp nền tảng cho việc tạo ra các mô hình 3D chân thực, tối ưu hóa hiệu ứng ánh sáng và tạo hình ảnh sống động.

Đặc trưng hình học như Surface normal và Curvature là những công cụ quan trọng trong việc phân tích và mô tả dữ liệu 3D. Chúng giúp làm rõ hình dạng, cấu trúc và biến đổi hình học của các đối tượng, từ đó tăng cường hiệu quả trong nhận diện, phân loại và tái tạo dữ liệu. Sự phát triển không ngừng của công nghệ và các thuật toán tính toán sẽ tiếp tục nâng cao khả năng khai thác đặc trưng hình học, góp phần tạo nên những bước đột phá trong lĩnh vực xử lý dữ liệu không gian ba chiều.

1.3.8. Mô hình SVM (Support Vector Machine)

Support Vector Machine (SVM) là một phương pháp học máy có giám sát mạnh mẽ, được sử dụng rộng rãi trong các bài toán phân loại và hồi quy. SVM được thiết kế để tìm kiếm một siêu phẳng tối ưu phân tách các lớp dữ liệu, tối đa hóa khoảng cách giữa các lớp. Phương pháp này đặc biệt phù hợp cho các bài toán có dữ liệu đa chiều và phức tạp, chẳng hạn như dữ liệu hình ảnh và đám mây điểm 3D từ các cảm biến như RealSense Camera. SVM đã chứng minh được khả năng ứng dụng hiệu quả trong các lĩnh vực nhận dạng, phân loại hình ảnh và phân tích dữ liệu không gian ba chiều.



Hình 1.9. Mô hình SVM [9]

Trong không gian hai chiều, một siêu phẳng là một đường thẳng phân chia dữ liệu thành hai tập hợp khác nhau. Trong không gian ba chiều, siêu phẳng trở thành một mặt phẳng, còn trong không gian nhiều chiều hơn (n-dimension), siêu phẳng là một mặt phân chia dữ liệu thành hai lớp riêng biệt. Đối với bài toán phân loại, mục tiêu của SVM là tìm ra một siêu phẳng tối ưu, giúp phân tách các lớp dữ liệu sao cho khoảng cách (hay

còn gọi là margin) giữa siêu phẳng và các điểm dữ liệu gần nhất (các véc-tơ hỗ trợ) là lớn nhất.

Khoảng cách tối đa này chính là một yếu tố quan trọng giúp giảm thiểu sai số tổng quát (generalization error) và tăng cường khả năng phân loại chính xác của mô hình. Điều này làm cho SVM trở thành một công cụ hữu hiệu trong các bài toán phân loại dữ liệu có số chiều lớn và yêu cầu độ chính xác cao.

Các điểm dữ liệu gần nhất với siêu phẳng được gọi là véc-tơ hỗ trợ. Chúng đóng vai trò quyết định trong việc xác định vị trí của siêu phẳng phân tách. Các véc-tơ hỗ trợ này ảnh hưởng trực tiếp đến việc định vị và định hình siêu phẳng. Cách tiếp cận của SVM nhằm tìm ra siêu phẳng tối ưu là tối ưu hóa khoảng cách margin giữa các véc-tơ hỗ trợ và siêu phẳng, giúp tối đa hóa khả năng phân loại chính xác các đối tượng.

Đặc biệt, khi làm việc với dữ liệu không tuyến tính hoặc có sự chồng chéo giữa các lớp dữ liệu, SVM có thể sử dụng các hàm hạt nhân (kernel function) để ánh xạ dữ liệu từ không gian gốc sang một không gian mới có số chiều cao hơn, giúp xác định một siêu phẳng phân tách dữ liệu trong không gian này.

SVM hỗ trợ nhiều loại hàm hạt nhân khác nhau để phù hợp với các dạng dữ liệu khác nhau. Các hàm hạt nhân phổ biến bao gồm:

- Hàm tuyến tính (Linear Kernel): Thích hợp cho các bài toán mà các lớp có thể phân tách bằng một siêu phẳng tuyến tính.
- Hàm đa thức (Polynomial Kernel): Biến đổi dữ liệu sang không gian đa chiều, thích hợp với các bài toán có ranh giới phân chia phi tuyến tính.
- Hàm Gaussian hoặc RBF (Radial Basis Function): Là một trong những hàm hạt nhân phổ biến nhất, thường được sử dụng cho các bài toán phức tạp khi dữ liệu không thể phân tách tốt trong không gian gốc.
- Hàm sigmoid: Tương tự như một mạng neural, thường được sử dụng trong các ứng dụng đặc biệt.

Việc lựa chọn hàm hạt nhân là một yếu tố quan trọng, ảnh hưởng trực tiếp đến hiệu suất của mô hình. Quá trình lựa chọn hàm hạt nhân và tối ưu các tham số đi kèm là một phần không thể thiếu trong việc xây dựng mô hình SVM hiệu quả.

SVM là một công cụ quan trọng trong các hệ thống định danh và phân loại đối tượng. Với dữ liệu hình ảnh 2D và đám mây điểm 3D từ cảm biến RealSense Camera,

SVM có thể cung cấp khả năng phân loại và định danh chính xác thông qua việc học tập từ dữ liệu huấn luyện.

- Khả năng làm việc với dữ liệu có số chiều cao: Dữ liệu 2D và 3D, đặc biệt khi xử lý hình ảnh và đám mây điểm, thường có số chiều lớn và nhiều đặc trưng phức tạp. SVM được biết đến với khả năng xử lý tốt dữ liệu dạng này mà không yêu cầu giảm số chiều. Điều này giúp duy trì thông tin cần thiết cho quá trình phân loại, cải thiện độ chính xác và khả năng phân biệt.
- Khả năng phân loại mạnh mẽ: Khi sử dụng các hàm hạt nhân phi tuyến tính, SVM có thể phân loại dữ liệu trong không gian mới với ranh giới phức tạp, giúp giải quyết các bài toán có sự chồng chéo dữ liệu. Nhờ vậy, nó trở thành công cụ hữu hiệu cho việc nhận dạng và phân loại các đối tượng phức tạp trong không gian ba chiều.

Quá trình xây dựng một mô hình SVM cho định danh đối tượng sử dụng dữ liệu từ cảm biến RealSense Camera bao gồm các bước sau:

- Thu thập và tiền xử lý dữ liệu
 - + Thu thập dữ liệu: Thu thập dữ liệu hình ảnh 2D và 3D từ cảm biến RealSense Camera.
 - + Tiền xử lý: Xử lý dữ liệu bao gồm các bước như loại bỏ nhiễu, chuẩn hóa dữ liệu, và trích xuất các đặc trưng cần thiết.
- Lựa chọn hàm hạt nhân và tối ưu hóa tham số
 - + Lựa chọn hàm hạt nhân phù hợp: Dựa trên đặc điểm của dữ liệu và bài toán, có thể sử dụng các loại hàm hạt nhân khác nhau.
 - + Tối ưu hóa tham số: Sử dụng các kỹ thuật như cross-validation để tối ưu hóa các tham số của mô hình nhằm đạt được hiệu suất cao nhất.
- Huấn luyện mô hình
 - + Chia dữ liệu: Tách dữ liệu thành tập huấn luyện và tập kiểm tra để đảm bảo tính tổng quát của mô hình.
 - + Huấn luyện: Sử dụng tập dữ liệu huấn luyện để tìm ra siêu phẳng tối ưu phân tách các lớp đối tượng.
- Đánh giá mô hình
 - + Sử dụng tập kiểm tra: Đánh giá mô hình trên tập dữ liệu chưa từng được huấn luyện để kiểm tra khả năng tổng quát hóa.

- + Các chỉ số đánh giá: Đánh giá độ chính xác (accuracy), độ nhạy (sensitivity), độ đặc hiệu (specificity), và chỉ số F1-score để đo lường mức độ hiệu quả của mô hình.
- Một số thách thức khi triển khai mô hình SVM với dữ liệu 2D và 3D bao gồm:
 - + Kích thước dữ liệu lớn: Việc xử lý các tập dữ liệu lớn và phức tạp có thể làm tăng thời gian tính toán và yêu cầu tài nguyên cao.
 - + Lựa chọn thông số phù hợp: Lựa chọn hàm hạt nhân và các tham số tối ưu là một vấn đề quan trọng, ảnh hưởng lớn đến hiệu suất của mô hình.

Giải pháp: Có thể sử dụng các phương pháp tối ưu hóa như cross-validation, hoặc kết hợp với các thuật toán giảm chiều dữ liệu để cải thiện khả năng xử lý.

Mô hình SVM là một công cụ mạnh mẽ cho bài toán phân loại và định danh đối tượng, đặc biệt khi làm việc với dữ liệu hình ảnh 2D và đám mây điểm 3D từ cảm biến RealSense Camera. Với khả năng phân loại chính xác, SVM đóng vai trò quan trọng trong các ứng dụng nhận diện và phân loại đối tượng, góp phần nâng cao hiệu quả và độ chính xác của các hệ thống học máy.

1.3.9. Mô hình mạng nơ-ron tích chập (CNN)

Mạng nơ-ron tích chập (Convolutional Neural Networks - CNN) là một kiến trúc mạng nơ-ron nhân tạo đặc biệt, được thiết kế để xử lý và phân tích dữ liệu có cấu trúc lưới, chẳng hạn như hình ảnh. CNN đã tạo nên cuộc cách mạng trong lĩnh vực thị giác máy tính (computer vision) và các ứng dụng liên quan đến nhận diện, phân loại và phân tích hình ảnh, video, cũng như dữ liệu không gian 3D. Bằng cách tự động học các đặc trưng từ dữ liệu đầu vào, CNN đã thay đổi cách tiếp cận truyền thống đối với bài toán trích xuất đặc trưng, vốn thường đòi hỏi sự can thiệp của con người trong việc lựa chọn các đặc điểm quan trọng [15].

Một mạng CNN điển hình bao gồm ba loại lớp chính: lớp tích chập, lớp gộp và lớp hoàn toàn kết nối. Mỗi lớp đóng một vai trò cụ thể trong quá trình học của mạng và cùng nhau xây dựng một cấu trúc mạnh mẽ để học và trích xuất đặc trưng từ dữ liệu đầu vào.

Lớp tích chập là thành phần cốt lõi của CNN, chịu trách nhiệm thực hiện phép tính tích chập giữa các bộ và dữ liệu đầu vào. Mỗi bộ lọc là một ma trận có kích thước nhỏ hơn dữ liệu đầu vào và di chuyển qua từng điểm ảnh hoặc voxel của dữ liệu để tạo ra một bản đồ đặc trưng. Quá trình này cho phép mạng tự động học và trích xuất các đặc

trung quan trọng như cạnh, góc, hoa văn và các đặc trưng phức tạp hơn trong các tầng sâu hơn. Một ưu điểm lớn của lớp tích chập là tính chia sẻ tham số (parameter sharing), giúp giảm thiểu số lượng tham số cần học và tăng hiệu quả tính toán.

Sau khi dữ liệu được xử lý qua lớp tích chập, một lớp kích hoạt phi tuyến, chẳng hạn như ReLU, được áp dụng để giới thiệu tính phi tuyến vào mô hình. ReLU là một hàm đơn giản nhưng hiệu quả, với đầu ra là giá trị 0 nếu đầu vào nhỏ hơn 0 và giữ nguyên giá trị đầu vào nếu lớn hơn 0. Các hàm kích hoạt phi tuyến giúp mạng học các đặc trưng phức tạp hơn và cải thiện khả năng phân loại.

Lớp gộp thường được sử dụng để giảm kích thước của bản đồ đặc trưng, giúp giảm thiểu số lượng tham số và tính toán trong mạng, đồng thời kiểm soát vấn đề quá khớp. Lớp gộp có hai dạng phổ biến là Max Pooling (giữ lại giá trị lớn nhất trong vùng lọc) và Average Pooling (tính giá trị trung bình của vùng lọc). Max Pooling thường được sử dụng nhiều hơn trong thực tế, do khả năng giữ lại các đặc trưng mạnh mẽ và đáng chú ý hơn trong dữ liệu.

Lớp này được sử dụng để tổng hợp các đặc trưng đã học và thực hiện các dự đoán hoặc phân loại. Trong lớp hoàn toàn kết nối, mỗi nút được kết nối với tất cả các nút trong lớp trước đó, giúp mạng học mối quan hệ giữa các đặc trưng và tạo ra các quyết định cuối cùng.

CNN không chỉ nổi bật trong việc xử lý dữ liệu 2D, chẳng hạn như hình ảnh tĩnh hoặc chuỗi hình ảnh video, mà còn có khả năng xử lý dữ liệu không gian 3D. Điều này đặc biệt quan trọng khi làm việc với dữ liệu từ các thiết bị như RealSense Camera, có khả năng thu thập cả dữ liệu 2D (hình ảnh RGB) và dữ liệu độ sâu 3D. CNN có thể được mở rộng sang các mạng 3D CNN để xử lý dữ liệu ba chiều, cho phép học các đặc trưng không gian như hình dạng, cấu trúc vật thể trong không gian và quan hệ giữa các điểm không gian.

Việc kết hợp CNN với các kỹ thuật trích xuất đặc trưng truyền thống như Histogram of Oriented Gradients (HOG) và Haar-like Features mang lại những lợi ích đáng kể. Trong khi CNN có khả năng tự động học các đặc trưng từ dữ liệu, HOG tập trung vào việc phát hiện các cạnh và hướng, còn Haar-like Features giúp phát hiện các vùng sáng tối nhanh chóng. Sự kết hợp giữa các phương pháp này tạo nên một tập hợp đặc trưng mạnh mẽ, tối ưu cho việc phân loại, nhận dạng và định danh từ dữ liệu 2D và 3D.

Lợi ích của CNN trong xử lý dữ liệu hình ảnh:

- Khả năng học tự động và giảm thiểu sự phụ thuộc vào tiền xử lý: Khác với các phương pháp truyền thống yêu cầu lựa chọn và trích xuất đặc trưng thủ công, CNN có khả năng học tự động các đặc trưng từ dữ liệu thô thông qua quá trình huấn luyện.
- Hiệu quả và tối ưu hóa tính toán: Với việc chia sẻ tham số trong các lớp tích chập và sử dụng các kỹ thuật giảm thiểu số lượng tham số như gộp, CNN đạt được hiệu quả tính toán cao và giảm thiểu nguy cơ quá khớp.
- Tính đa dạng và linh hoạt: CNN có thể dễ dàng điều chỉnh và mở rộng để làm việc với nhiều dạng dữ liệu khác nhau, từ hình ảnh, video đến dữ liệu không gian 3D và chuỗi thời gian.
- Khả năng trích xuất đặc trưng sâu và mạnh mẽ: Các lớp sâu hơn trong CNN có thể học được các đặc trưng phức tạp, chẳng hạn như hình dạng, đối tượng, mối quan hệ không gian, giúp nâng cao khả năng nhận dạng và phân loại dữ liệu.

Mô hình CNN đã được ứng dụng rộng rãi trong nhiều lĩnh vực, từ nhận diện khuôn mặt, phân đoạn hình ảnh, nhận diện vật thể, phân tích video, đến nhận diện sinh trắc học. Trong ngữ cảnh định danh sử dụng dữ liệu 2D và 3D từ RealSense Camera, CNN có khả năng học các đặc trưng từ dữ liệu RGB và độ sâu, giúp tạo ra hệ thống nhận diện và định danh mạnh mẽ với độ chính xác cao.

Kết luận, việc áp dụng CNN trong bài toán định danh dựa trên dữ liệu 2D và 3D từ RealSense Camera mang lại nhiều triển vọng trong các lĩnh vực như an ninh, sinh trắc học, và nhận diện hình ảnh nâng cao. CNN không chỉ cải thiện khả năng trích xuất đặc trưng mà còn giúp tạo ra các hệ thống nhận diện thông minh và tự động.

1.3.10. Đặc trưng của dữ liệu hình ảnh 2D

Dữ liệu 2D từ camera RealSense bao gồm thông tin về màu sắc, chi tiết bề mặt và kết cấu của đối tượng trong một mặt phẳng, được cung cấp bởi cảm biến RGB. Loại dữ liệu này chủ yếu được sử dụng trong các ứng dụng xử lý ảnh truyền thống như nhận diện khuôn mặt, phân loại vật thể, phân đoạn hình ảnh và phân tích môi trường. Cụ thể, dữ liệu 2D có một số đặc điểm đáng chú ý:

- Cung cấp chi tiết bề mặt và màu sắc chính xác: Dữ liệu 2D giúp ghi lại chính xác các đặc điểm bề mặt của đối tượng, bao gồm cả màu sắc và hoa văn. Điều này rất hữu ích trong các ứng dụng yêu cầu phân tích chi tiết về màu sắc hoặc hoa văn

như kiểm tra sản phẩm, nhận diện đối tượng qua bề mặt (như nhận diện gương mặt).

- Phổ biến trong các ứng dụng xử lý ảnh truyền thống: Dữ liệu 2D đã và đang là tiêu chuẩn trong lĩnh vực xử lý ảnh và thị giác máy tính. Với các phương pháp học máy hiện đại như deep learning, dữ liệu 2D được sử dụng để đào tạo các mô hình phân loại, nhận diện và phân đoạn ảnh hiệu quả. Ngoài ra, dữ liệu này cũng dễ dàng thu thập và xử lý, không yêu cầu phần cứng đặc biệt.

Tuy nhiên, dữ liệu 2D từ camera RGB cũng có một số hạn chế quan trọng:

- Thiếu thông tin về chiều sâu: Vì dữ liệu 2D chỉ ghi lại thông tin về mặt phẳng mà không có thông tin về khoảng cách, hệ thống nhận diện sẽ gặp khó khăn khi đối tượng thay đổi khoảng cách hoặc góc nhìn với camera. Điều này đặc biệt ảnh hưởng đến độ chính xác trong việc nhận dạng hoặc xác định vị trí chính xác của đối tượng trong không gian ba chiều.
- Dễ bị ảnh hưởng bởi điều kiện ánh sáng và bóng đổ: Các yếu tố ngoại cảnh như ánh sáng môi trường hoặc bóng đổ có thể ảnh hưởng lớn đến chất lượng và độ chính xác của dữ liệu 2D. Ví dụ, trong trường hợp ánh sáng yếu hoặc quá sáng, dữ liệu 2D có thể thiếu chi tiết hoặc bị nhiễu, dẫn đến việc nhận diện sai lệch. Các hệ thống sử dụng dữ liệu 2D thường phải được điều chỉnh hoặc trang bị các thuật toán xử lý ánh sáng để giảm thiểu các ảnh hưởng này.
- Khả năng hạn chế trong việc nhận diện đối tượng động hoặc khi góc nhìn thay đổi: Khi đối tượng di chuyển hoặc xoay, dữ liệu 2D gặp khó khăn trong việc theo dõi và duy trì độ chính xác. Điều này là do không có dữ liệu về chiều sâu để giúp hệ thống điều chỉnh mô hình nhận diện theo các góc độ khác nhau của đối tượng.

Mặc dù có những hạn chế, dữ liệu 2D vẫn là một thành phần không thể thiếu và quan trọng trong nhiều ứng dụng, đặc biệt là trong các tình huống mà dữ liệu bề mặt và màu sắc đóng vai trò quan trọng hơn là vị trí không gian.

1.3.11. Đặc trưng của dữ liệu 3D

Dữ liệu 3D từ camera RealSense, được cung cấp thông qua cảm biến độ sâu, chứa thông tin ba chiều về không gian và hình dạng của đối tượng. Loại dữ liệu này giúp giải quyết nhiều hạn chế của dữ liệu 2D và mở ra khả năng ứng dụng rộng rãi trong các lĩnh vực như robot tự hành, thực tế ảo tăng cường (AR), và xây dựng mô hình 3D. Dữ liệu 3D có một số đặc điểm quan trọng như sau:

- Cung cấp thông tin về chiều sâu và không gian ba chiều: Với cảm biến độ sâu, dữ liệu 3D có thể xác định chính xác khoảng cách từ camera đến từng điểm trên bề mặt của đối tượng. Điều này giúp xây dựng một mô hình không gian ba chiều, cho phép hệ thống nhận diện đối tượng dựa trên cả kích thước và vị trí của đối tượng. Điều này đặc biệt quan trọng khi hệ thống cần phân biệt giữa các đối tượng có hình dạng tương tự nhưng khác nhau về kích thước hoặc cấu trúc.
- Độ chính xác cao khi góc nhìn và vị trí thay đổi: Dữ liệu 3D có khả năng giữ nguyên độ chính xác ngay cả khi đối tượng quay hoặc di chuyển trong không gian. Khả năng này rất hữu ích trong các ứng dụng nhận diện đối tượng trong môi trường thay đổi, chẳng hạn như robot tự hành hoặc các ứng dụng cần theo dõi chuyển động liên tục của đối tượng. Ví dụ, trong hệ thống kiểm tra chất lượng sản phẩm, dữ liệu 3D cho phép kiểm tra từ nhiều góc độ khác nhau để đảm bảo tính toàn vẹn của sản phẩm.
- Ít bị ảnh hưởng bởi ánh sáng môi trường và bóng đổ: So với dữ liệu 2D, dữ liệu 3D ít phụ thuộc vào điều kiện ánh sáng hơn, nhờ đó mang lại tính ổn định cao trong nhiều môi trường khác nhau, từ trong nhà đến ngoài trời. Điều này làm cho dữ liệu 3D trở thành một lựa chọn lý tưởng trong các ứng dụng yêu cầu hoạt động ổn định và chính xác bất kể điều kiện ánh sáng.
- Khả năng xây dựng bản đồ và nhận dạng không gian chi tiết: Dữ liệu 3D từ cảm biến độ sâu của RealSense cho phép hệ thống tạo ra bản đồ không gian chi tiết của môi trường xung quanh. Đây là một lợi thế lớn trong các ứng dụng như lập bản đồ tự động, xây dựng mô hình kiến trúc 3D, và theo dõi chuyển động của con người. Ví dụ, trong robot tự hành, dữ liệu 3D giúp tạo ra bản đồ môi trường, hỗ trợ trong việc định vị và tránh chướng ngại vật.

Việc kết hợp dữ liệu 2D và 3D từ camera RealSense mang lại nhiều lợi ích cho các hệ thống nhận diện thông minh. Trong các ứng dụng yêu cầu độ chi tiết cao về bề mặt và màu sắc, dữ liệu 2D là lựa chọn hàng đầu. Trong khi đó, với các yêu cầu về thông tin không gian, độ sâu và khả năng chống nhiễu từ môi trường ánh sáng, dữ liệu 3D đóng vai trò quan trọng. Một số ứng dụng điển hình của sự kết hợp này bao gồm:

- Nhận diện đối tượng và phân đoạn hình ảnh 3D: Bằng cách kết hợp dữ liệu RGB và độ sâu, hệ thống có thể nhận diện và phân đoạn đối tượng dựa trên hình dáng,

màu sắc và vị trí trong không gian ba chiều. Điều này giúp tăng cường độ chính xác của các ứng dụng phân loại đối tượng trong môi trường phức tạp.

- Robot tự hành và định vị trong không gian: Dữ liệu 3D giúp robot tự hành xác định vị trí và di chuyển trong không gian một cách an toàn. Đồng thời, dữ liệu 2D hỗ trợ trong việc nhận diện các chi tiết bề mặt của vật thể và con người xung quanh, giúp robot đưa ra phản ứng chính xác và nhanh chóng.
- Thực tế ảo tăng cường (AR) và thực tế ảo (VR): Dữ liệu 3D từ camera RealSense có thể được sử dụng để tạo ra không gian ảo với độ chi tiết cao, hỗ trợ các ứng dụng AR/VR trong việc xây dựng môi trường ảo hoặc tương tác với các đối tượng ảo trong không gian thực.

CHƯƠNG 2: PHÂN TÍCH BÀI TOÁN ĐỊNH DANH ĐỐI TƯỢNG DỰA TRÊN DỮ LIỆU 2D VÀ 3D

2.1. Tổng quan về bài toán định danh đối tượng

Bài toán định danh đối tượng (Object Identification) là một trong những bài toán quan trọng trong lĩnh vực thị giác máy tính (Computer Vision), với mục tiêu xác định chính xác danh tính của đối tượng trong môi trường thực tế. Điều này bao gồm việc sử dụng dữ liệu đầu vào từ các cảm biến hình ảnh và độ sâu như camera RealSense để xây dựng mô hình máy học có khả năng tự động nhận diện và phân loại các đối tượng.

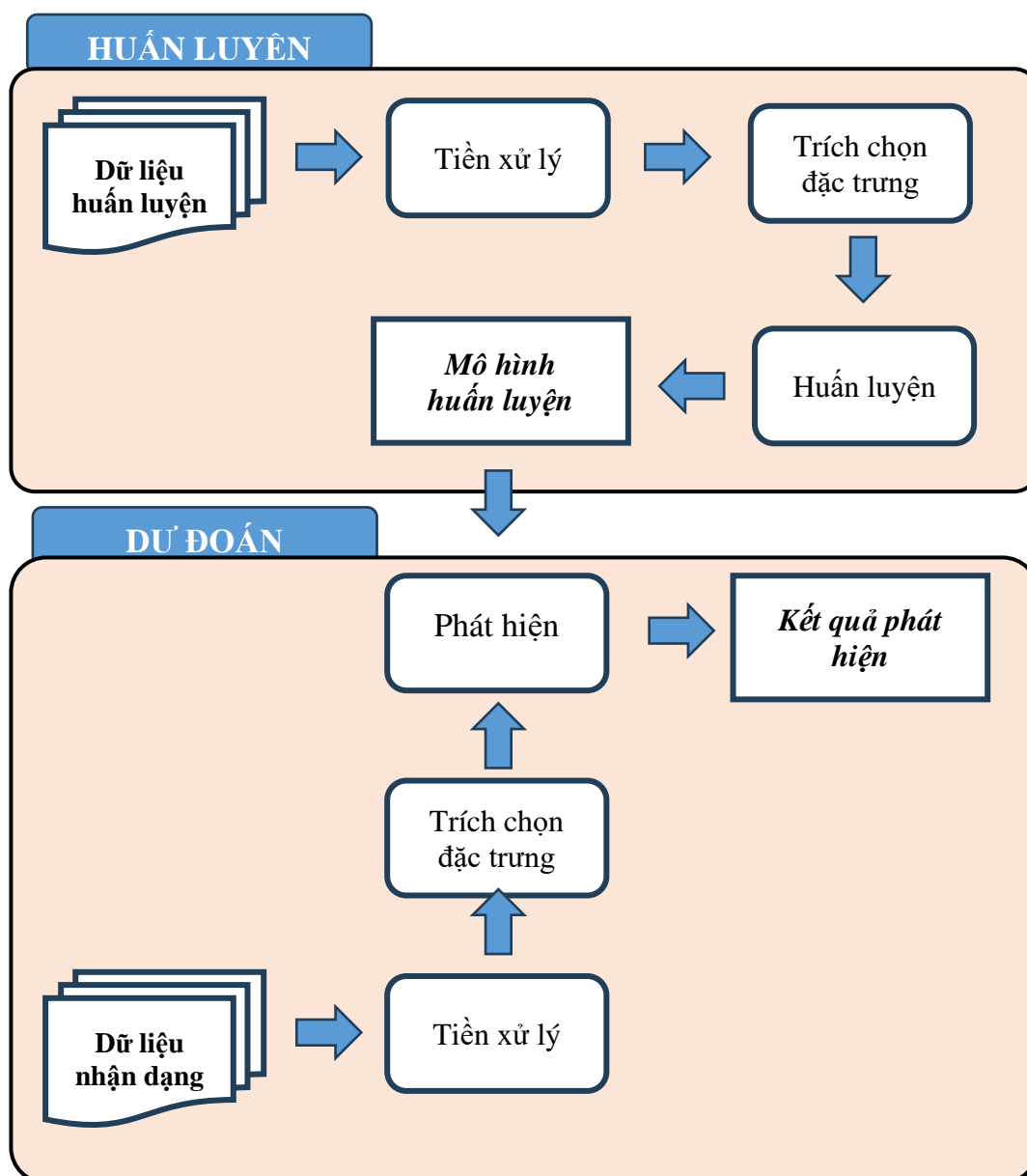
Phương pháp định danh đối tượng bằng dữ liệu 2D và 3D đã và đang được sử dụng rộng rãi trong nhiều lĩnh vực, từ các ứng dụng trong ngành công nghiệp như giám sát sản xuất, kiểm tra chất lượng sản phẩm, cho đến các ứng dụng trong đời sống hằng ngày như hệ thống an ninh, robot tự hành và các hệ thống nhà thông minh. Điểm đặc biệt của bài toán định danh đối tượng là khả năng phân loại chính xác ngay cả khi đối tượng có sự biến dạng, góc nhìn khác nhau, hoặc trong các điều kiện ánh sáng khác nhau.

Camera RealSense của Intel là một công cụ mạnh mẽ trong việc hỗ trợ thu thập cả dữ liệu 2D (hình ảnh màu) và dữ liệu 3D (độ sâu). Dữ liệu 2D giúp mô tả chi tiết về màu sắc, hình dạng bề mặt của đối tượng, trong khi dữ liệu 3D cung cấp thêm thông tin về chiều sâu và không gian, từ đó giúp hệ thống định danh có thể nhận diện đối tượng ngay cả trong các điều kiện phức tạp như khi đối tượng bị che khuất hoặc có sự biến dạng về hình dạng. Bằng cách kết hợp dữ liệu 2D và 3D, mô hình máy học có thể đạt được độ chính xác cao hơn trong quá trình định danh.

Bài toán định danh đối tượng thông qua dữ liệu 2D và 3D được xây dựng bằng cách sử dụng các mô hình học máy tiên tiến. Các mô hình này được huấn luyện dựa trên tập dữ liệu lớn, với đầu vào là các hình ảnh 2D và các điểm đám mây (point cloud) 3D của đối tượng. Sau đó, mô hình có thể dự đoán và xác định đối tượng mới chưa từng thấy bằng cách so sánh các đặc điểm rút trích từ dữ liệu đầu vào với các mẫu đã học từ trước.

2.2. Quy trình định danh đối tượng sử dụng dữ liệu 2D và 3D

Quy trình định danh đối tượng sử dụng dữ liệu 2D và 3D từ camera RealSense thường tuân theo các bước sau:



Hình 2.1. Quy trình định danh đối tượng sử dụng dữ liệu 2D và 3D

Các giai đoạn trong việc huấn luyện dữ liệu bao gồm:

Đưa dữ liệu huấn luyện vào phần mềm: Trong bước đầu tiên của giai đoạn huấn luyện, dữ liệu từ camera RealSense sẽ được đưa vào phần mềm chuyên dụng hoặc các công cụ xử lý dữ liệu, chẳng hạn như các thư viện Python như OpenCV hoặc Pyrealsense2 để thu thập dữ liệu từ camera. Bộ dữ liệu này bao gồm hình ảnh 2D từ camera RGB và dữ liệu 3D từ cảm biến chiều sâu, giúp cung cấp thông tin đa chiều về đối tượng cần nhận dạng. Việc chuẩn bị dữ liệu bao gồm cả thao tác thiết lập môi trường thu thập, đảm bảo ánh sáng phù hợp và điều chỉnh khoảng cách từ camera đến đối tượng để tối ưu hóa chất lượng dữ liệu đầu vào.

Tiền xử lý dữ liệu: Dữ liệu thô từ camera RealSense thường chứa nhiều, thông tin dư thừa hoặc sai lệch do điều kiện môi trường và phần cứng. Do đó, bước tiếp theo là tiền xử lý dữ liệu để đảm bảo dữ liệu đầu vào sạch, chuẩn hóa và nhất quán.

Trích chọn đặc trưng: Việc trích chọn đặc trưng là một bước quan trọng nhằm chuyển đổi dữ liệu hình ảnh 2D và dữ liệu 3D thành các tập hợp thông tin có thể sử dụng để huấn luyện mô hình học máy.

Huấn luyện mô hình học máy: Dữ liệu đặc trưng sau khi được trích xuất sẽ được sử dụng để huấn luyện mô hình học máy. Quá trình huấn luyện có thể bao gồm việc áp dụng các thuật toán phân loại và nhận dạng như Support Vector Machine (SVM), mạng nơ-ron nhân tạo hoặc các mô hình học sâu như mạng nơ-ron tích chập.

Xuất mô hình huấn luyện: Sau khi hoàn thành huấn luyện, mô hình đã được tối ưu sẽ được lưu trữ dưới dạng file mô hình để sử dụng trong giai đoạn dự đoán. Việc lưu trữ mô hình bao gồm các thông tin về cấu trúc mô hình, các trọng số đã học được, và các siêu tham số cần thiết cho việc dự đoán sau này.

Các giai đoạn trong việc nhận diện dữ liệu bao gồm:

Thêm dữ liệu nhận dạng: Dữ liệu nhận dạng là các dữ liệu mới cần được phân loại hoặc định danh. Dữ liệu này được thu thập từ camera RealSense và bao gồm hình ảnh 2D và dữ liệu chiều sâu 3D tương tự như dữ liệu huấn luyện.

Tiền xử lý dữ liệu: Dữ liệu nhận dạng cũng cần được tiền xử lý để đảm bảo chất lượng tương tự như trong giai đoạn huấn luyện. Các bước xử lý bao gồm lọc nhiễu, chuẩn hóa dữ liệu và xử lý dữ liệu thiếu để đảm bảo tính nhất quán của dữ liệu.

Trích chọn đặc trưng: Dữ liệu nhận dạng được đưa qua quá trình trích chọn đặc trưng với các phương pháp giống như trong giai đoạn huấn luyện. Việc đảm bảo tính đồng nhất trong quá trình trích chọn là yếu tố quyết định đến hiệu quả của mô hình trong giai đoạn dự đoán.

Nạp mô hình huấn luyện để phát hiện và định danh dữ liệu: Mô hình đã được huấn luyện và lưu trữ sẽ được nạp vào để tiến hành phân loại hoặc định danh dữ liệu nhận dạng. Mô hình sẽ so sánh các đặc trưng từ dữ liệu nhận dạng với dữ liệu đã được học trước đó để đưa ra quyết định về đối tượng.

Xem kết quả phát hiện và định danh: Kết quả của quá trình dự đoán được hiển thị, bao gồm thông tin về đối tượng nhận dạng được, độ tin cậy của kết quả và các thông

tin liên quan. Kết quả này có thể được sử dụng để đánh giá lại hiệu suất của mô hình và cải tiến nếu cần thiết.

2.3. Thu thập dữ liệu 2D và 3D từ Realsense Camera:

Để thu thập dữ liệu cho bài toán định danh đối tượng, việc lựa chọn thiết bị và thiết lập môi trường thu thập dữ liệu cần được chú trọng để đạt hiệu quả tối ưu. Tôi đã sử dụng RealSense Camera Depth D415 để cung cấp dữ liệu màu 2D (hình ảnh RGB) và dữ liệu độ sâu 3D với độ chính xác cao nhờ cảm biến trường nhìn hẹp, phù hợp cho việc thu thập dữ liệu từ các đối tượng tĩnh với chi tiết rõ ràng. Môi trường thu thập dữ liệu nên được thiết lập sao cho ánh sáng ở mức phù hợp, tránh các nguồn sáng quá mạnh hoặc bóng đổ gây nhiễu ảnh. Điều này giúp dữ liệu 2D rõ nét hơn và tăng độ chính xác của dữ liệu 3D. Khoảng cách lý tưởng giữa đối tượng và camera D415 thường từ 0,2 đến 1,5 mét, nhằm đảm bảo đối tượng nằm trong khung hình và dữ liệu thu được có độ phân giải cao. Để đảm bảo tính ổn định, camera nên được cố định trên giá đỡ hoặc chân đế chắc chắn, giảm thiểu rung lắc trong suốt quá trình thu thập dữ liệu.

Để đảm bảo chất lượng dữ liệu thu thập từ camera RealSense D415 cho bài toán định danh đối tượng, cần chú trọng vào các yếu tố như ánh sáng, khoảng cách và không gian thu thập. Đối với ánh sáng, camera hoạt động hiệu quả trong môi trường có ánh sáng tự nhiên hoặc ánh sáng nhân tạo ở mức độ vừa phải, không quá mạnh hay phản chiếu quá mức. Các nguồn sáng mạnh hoặc bóng đổ có thể làm giảm chất lượng dữ liệu hình ảnh 2D và gây nhiễu cho dữ liệu 3D, do đó cần được kiểm soát cẩn thận. Mục tiêu là duy trì một mức ánh sáng ổn định, giúp dữ liệu thu thập được rõ nét, đồng thời tăng độ chính xác trong quá trình định danh đối tượng.

Về khoảng cách, camera hoạt động ở khoảng cách từ 0,2 mét đến 1,5 mét. Điều này giúp đảm bảo dữ liệu thu được có độ phân giải cao và thể hiện được chi tiết của đối tượng trong khung hình. Trong các bài toán định danh, cần chọn khoảng cách sao cho đối tượng luôn nằm gọn trong tầm quan sát của camera, giúp thu thập đầy đủ thông tin mà không bị mất chi tiết quan trọng. Việc duy trì khoảng cách phù hợp sẽ tối ưu hóa khả năng nhận diện và tăng cường hiệu quả của mô hình.

Ngoài ra, không gian thu thập dữ liệu cần đủ rộng để giảm thiểu các yếu tố gây nhiễu như tường gương hoặc bề mặt phản chiếu, vì chúng có thể gây biến dạng cho dữ liệu 3D. Camera D415 nên được đặt cố định trên giá đỡ hoặc chân đế để đảm bảo tính ổn định, tránh rung lắc trong quá trình thu thập dữ liệu. Sự ổn định này giúp duy trì chất

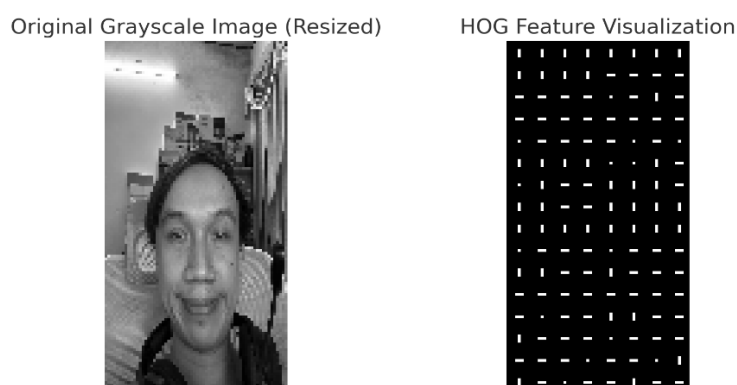
lượng dữ liệu 2D và 3D thu thập được, tạo điều kiện thuận lợi cho quá trình định danh đối tượng chính xác và hiệu quả. Tổng thể, việc tối ưu hóa ánh sáng, khoảng cách và không gian thu thập sẽ giúp khai thác tối đa khả năng của camera trong các ứng dụng nhận diện và định danh.

2.4. Trích chọn đặc trưng

Trong bài toán định danh đối tượng dựa trên dữ liệu từ RealSense Camera, việc trích chọn đặc trưng là một bước quan trọng giúp mô hình có thể nhận diện và phân biệt đối tượng một cách hiệu quả. Trong luận văn, tôi dùng phương pháp trích chọn đặc trưng HOG, Hu Moments và Haar-like để ứng dụng vào bài toán sử dụng hình ảnh 2D.

Trích xuất đặc trưng HOG

HOG là phương pháp trích xuất đặc trưng dựa trên sự biến đổi hướng của các gradient cường độ trong hình ảnh. HOG phân tích và ghi nhận sự thay đổi về hướng và độ mạnh của gradient trong một tập hợp các ô (cells) và khối (blocks) trong hình ảnh, tạo ra một vector đặc trưng đại diện cho cấu trúc của đối tượng trong hình ảnh. Tại bước trích xuất đặc trưng, hình ảnh 2D (RGB) của đối tượng được chuyển đổi sang ảnh xám. Sau đó, ảnh được chia thành các ô nhỏ, mỗi ô sẽ tính toán các hướng gradient. Một histogram của các hướng gradient được tạo ra cho từng ô, và các histogram này được chuẩn hóa để giảm nhạy cảm với ánh sáng và các yếu tố môi trường. Vector đặc trưng HOG cuối cùng sẽ đại diện cho cấu trúc và hình dạng của khuôn mặt hoặc các đối tượng trong hình ảnh.



Hình 2.2. Đặc trưng HOG được trích xuất từ hình ảnh 2D

Trên hình ảnh hiển thị, phía bên trái là ảnh gốc được chuyển sang ảnh xám và thu nhỏ lại để phù hợp cho trích xuất đặc trưng HOG. Phía bên phải là hình ảnh minh họa đặc trưng HOG được trích xuất từ dữ liệu 2D, với các đoạn thẳng hiển thị hướng và độ lớn của gradient trong các ô nhỏ trên ảnh. Đây là cách HOG mô tả cấu trúc và cạnh của

khuôn mặt, giúp mô hình nhận diện dựa vào các đặc điểm nổi bật và hướng gradient trong ảnh

Trích xuất đặc trưng Hu moments

Hu Moments là một tập hợp bảy đặc trưng bất biến dựa trên các moment hình học, được thiết kế để mô tả hình dạng của đối tượng trong không gian 2D. Các đặc trưng này có khả năng bất biến theo phép biến đổi tịnh tiến, xoay, và tỷ lệ, do đó phù hợp cho việc định danh các đối tượng có hình dạng khác nhau. Trong hình ảnh của đối tượng được tiền xử lý để chuyển sang ảnh nhị phân hoặc ảnh xám. Các moment hình học được tính toán để tạo ra các Hu Moments đặc trưng cho hình dạng tổng thể của đối tượng. Hu Moments cung cấp một tập hợp nhỏ các giá trị đặc trưng giúp mô hình có thể phân biệt các đối tượng dựa trên hình dạng và cấu trúc của chúng, ngay cả khi có các biến đổi như xoay hoặc thay đổi kích thước.



Hình 2.3. Đặc trưng Hu moments được trích xuất từ hình ảnh 2D

Hình ảnh trên là ảnh nhị phân được sử dụng để tính toán đặc trưng Hu Moments từ dữ liệu 2D. Ảnh này là kết quả của việc chuyển đổi từ ảnh xám sang ảnh nhị phân, qua đó giúp mô tả hình dạng tổng thể của đối tượng bằng cách sử dụng các moment hình học. Các Hu Moments trích xuất từ ảnh này sẽ cung cấp thông tin về hình dạng mà bất biến với các phép biến đổi như dịch chuyển, xoay và tỷ lệ.

Trích xuất đặc trưng Haar-like Features

Haar-like features là các đặc trưng dựa trên sự khác biệt giữa các vùng hình chữ nhật trong hình ảnh. Phương pháp này nổi tiếng nhờ việc áp dụng trong nhận diện khuôn mặt, như trong bộ phát hiện khuôn mặt của Viola-Jones. Các đặc trưng Haar so sánh độ sáng giữa các vùng cạnh nhau để phát hiện các đặc điểm cục bộ như cạnh, góc, hoặc vùng sáng/tối. Hình ảnh của đối tượng được phân tích thành nhiều vùng và các mặt nạ Haar-like được áp dụng để tính toán sự khác biệt giữa các vùng này. Các đặc trưng như

cạnh mũi, hốc mắt hoặc miệng có thể được xác định qua Haar-like features. Để tăng tốc quá trình tính toán, kỹ thuật “cascading” thường được áp dụng để loại bỏ nhanh các vùng không phải là đối tượng quan tâm.

2.5 Mô hình định dạng/phân lớp đối tượng:

Mô hình Support Vector Machine

Support Vector Machine là một thuật toán học máy mạnh mẽ được sử dụng phổ biến trong các bài toán phân loại và hồi quy. Nguyên lý hoạt động của SVM dựa trên việc tìm kiếm một siêu phẳng phân tách tốt nhất trong không gian đặc trưng để tách biệt các lớp dữ liệu. Mục tiêu của SVM là tối ưu hóa khoảng cách, hay còn gọi là "margin", giữa các điểm dữ liệu gần nhất của mỗi lớp và siêu phẳng. Điều này giúp tăng khả năng tổng quát hóa và giảm thiểu sai số trong quá trình phân loại. Trong các trường hợp dữ liệu không tuyến tính, SVM sử dụng các hàm kernel để ánh xạ dữ liệu từ không gian gốc sang không gian đặc trưng cao hơn, nơi các lớp dữ liệu có thể được phân tách dễ dàng hơn. Một số kernel phổ biến bao gồm Linear Kernel cho các bài toán phân loại tuyến tính và RBF Kernel, thường được áp dụng cho dữ liệu phi tuyến tính phức tạp.

Trong bài toán định danh, quá trình sử dụng SVM thường bắt đầu với việc chuẩn bị và trích xuất đặc trưng từ dữ liệu. Dữ liệu có thể bao gồm hình ảnh 2D hoặc thông tin 3D thu được từ các thiết bị như camera RealSense. Các đặc trưng phổ biến được sử dụng có thể là HOG, Hu Moments. Sau khi trích xuất, dữ liệu cần được chuẩn hóa nhằm cải thiện hiệu quả phân loại của SVM. Dữ liệu này sẽ được chia thành tập huấn luyện và tập kiểm tra để đảm bảo mô hình có thể học và đánh giá khả năng phân loại. Trong quá trình huấn luyện, các siêu tham số quan trọng như loại kernel, giá trị điều chỉnh C hay gamma cần được tinh chỉnh để tối ưu hóa hiệu suất mô hình.

SVM có thể được ứng dụng trong các bài toán nhận diện khuôn mặt dựa trên dữ liệu hình ảnh 2D hoặc phân loại đối tượng từ dữ liệu 3D. Đối với các bài toán phức tạp, SVM cũng có thể được mở rộng cho phân loại đa lớp. Ưu điểm nổi bật của SVM nằm ở khả năng xử lý tốt dữ liệu có số lượng đặc trưng lớn và khả năng phân biệt mạnh mẽ nhờ việc tối đa hóa margin giữa các lớp. Tuy nhiên, đối với các bài toán có dữ liệu lớn và nhiều lớp, việc tinh chỉnh các tham số và lựa chọn kernel phù hợp là yếu tố cần thiết để đảm bảo hiệu quả tối ưu. Nhìn chung, SVM cung cấp một giải pháp mạnh mẽ cho việc định danh và phân loại đối tượng, đặc biệt khi dữ liệu đã được xử lý và trích xuất đặc trưng đầy đủ.

Mô hình Convolutional Neural Network

Convolutional Neural Network (CNN) là một loại mạng nơ-ron sâu được thiết kế đặc biệt để xử lý dữ liệu có cấu trúc dạng lưới, chẳng hạn như hình ảnh. CNN đã chứng minh hiệu quả vượt trội trong các bài toán xử lý hình ảnh, phân loại đối tượng, nhận diện khuôn mặt, và nhiều ứng dụng khác trong lĩnh vực thị giác máy tính. Điểm mạnh của CNN nằm ở khả năng tự động trích xuất các đặc trưng từ dữ liệu đầu vào, giúp giảm thiểu sự phụ thuộc vào các kỹ thuật trích xuất đặc trưng thủ công.

CNN hoạt động thông qua một chuỗi các lớp tích chập (convolutional layers) và lớp gộp (pooling layers), cuối cùng là các lớp kết nối đầy đủ (fully connected layers) để thực hiện quá trình phân loại hoặc dự đoán. Các lớp tích chập đóng vai trò như các bộ lọc (filters), quét qua hình ảnh và học các đặc trưng cục bộ như cạnh, góc, và hình dạng phức tạp hơn qua các tầng sâu hơn. Các lớp gộp được sử dụng để giảm kích thước không gian của đặc trưng, từ đó giảm số lượng tham số và tăng tính bất biến đối với phép biến đổi vị trí hoặc thay đổi kích thước của đối tượng trong hình ảnh.

CHƯƠNG 3: THỬ NGHIỆM MÔ HÌNH VÀ ĐÁNH GIÁ KẾT QUẢ

3.1. Kết quả thực nghiệm

3.1.1. Thu thập dữ liệu 2D và 3D và xử lý

Bộ dữ liệu 2D sẽ được lấy trực tiếp từ carema hoặc sử dụng các hình ảnh đã có sẵn để huấn luyện.

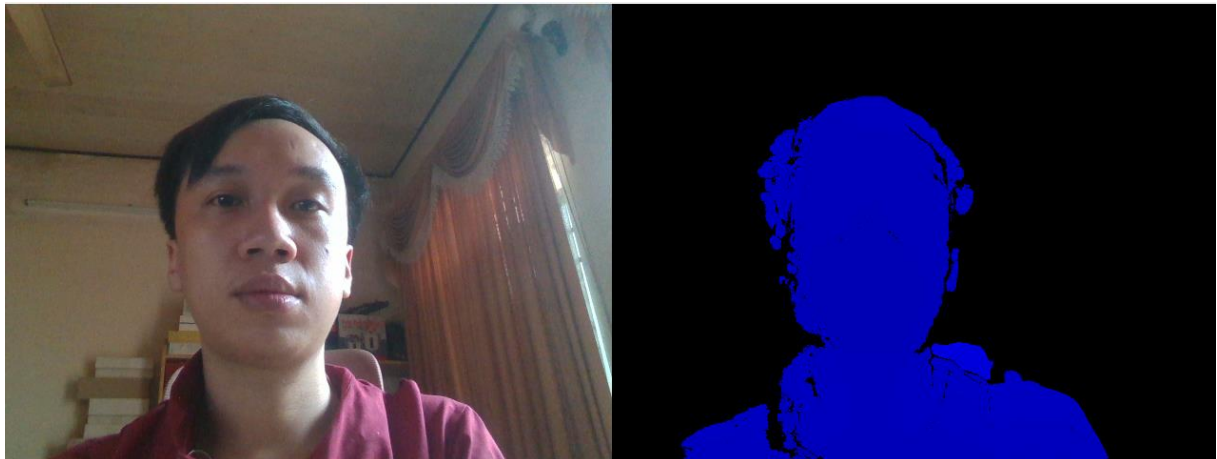


Hình 3.1. Bộ dữ liệu khuôn mặt được lấy từ Realsence Camara



Hình 3.2. Bộ dữ liệu hình con mèo

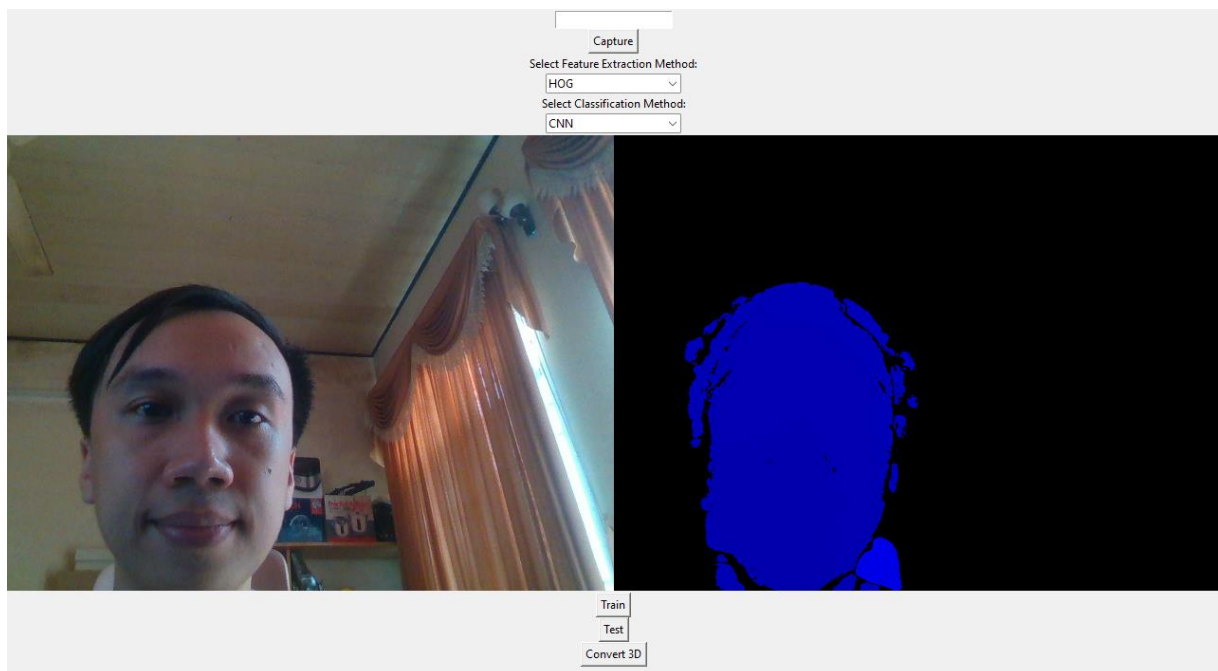
Bộ dữ liệu 3D sẽ được lấy trực tiếp từ camera hoặc sẽ được chuyển đổi trực tiếp không giám sát từ phần mềm.



Hình 3.3. Hình ảnh dữ liệu 3D

Sau khi đã có được dữ liệu hình ảnh 2D và 3D, chương trình sẽ bắt đầu quá trình huấn luyện. Trong phần mềm đã có 3 thuật toán trích xuất đặc trưng HOG, Hu-moments và Haar-like features, tiếp theo là mô hình phân lớp bao gồm: CNN, SVM.

Trong quá trình huấn luyện mô hình này, tôi đã sử dụng nhiều hình ảnh có kích thước khác nhau và batch_size là 16. Chương trình sử dụng thư viện Keras và Tensorflow để thực hiện quá trình huấn luyện và đánh giá mô hình.



Hình 3.4. Giao diện của phần mềm

Chương trình được xây dựng với các chức năng như sau:

- Chức năng chụp ảnh: Chức năng này sẽ chụp dữ liệu thực tế từ camera Realsense của đối tượng. Dữ liệu sau khi đưa vào sẽ chia ra làm 2 hình ảnh là hình ảnh 2D và hình ảnh 3D để đưa vào hệ thống.

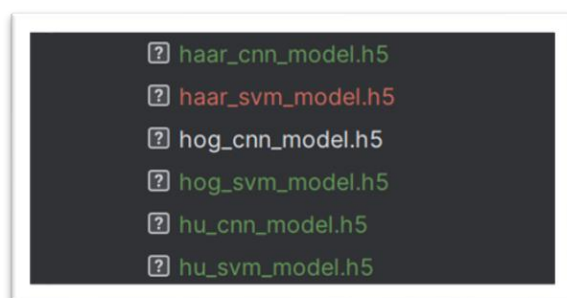
- Chức năng huấn luyện mô hình: Chương trình sẽ cho chọn các thuật toán trích xuất đặc trưng và các thuật toán phân lớp để chạy chức năng huấn luyện mô hình.

- Chức năng dự đoán: Chương trình cho phép upload hình ảnh cần dự đoán để chương trình dự đoán với mô hình huấn luyện đã có sẵn.

- Chức năng chuyển đổi hình ảnh 3D: Chức năng này sẽ chuyển đổi dữ liệu hình ảnh 2D sang 3D với những hình ảnh không lấy trực tiếp được từ camera Realsense.

Sau khi dữ liệu được thu thập được chụp hình sử dụng camera, ta tiến hành quá trình tiền xử lý. Dữ liệu được tiền xử lý bằng cách sẽ chỉnh sửa thành một kích thước duy nhất để bắt đầu quá trình thực hiện mô hình huấn luyện.

Sau khi dữ liệu đã được xử lý, người dùng sẽ chọn thuật toán trích xuất đặc trưng (HOG, Hu-moments, Haar-like features) và thuật toán phân lớp (CNN, SVM). Chương trình sẽ bắt đầu sử dụng trích xuất đặc trưng và thuật toán phân lớp đã được chọn và xuất ra file mô hình được lưu vào trong chương trình.



Hình 3.5. Các file mô hình đã được chương trình huấn luyện

Sau khi đã có được các file mô hình học máy theo mỗi trích xuất đặc trưng và thuật toán phân lớp khác nhau, tôi sẽ bắt đầu dự đoán đối tượng với dữ liệu hình ảnh đầu vào được đưa vào hệ thống.

Sau khi đưa 1 hình ảnh để bắt đầu quá trình dự đoán, phần mềm sẽ xuất ra kết quả dự đoán theo trích xuất đặc trưng và thuật toán phân lớp đã được chọn. Kết quả được hiển thị ở hình bên dưới



Hình 3.6. Kết quả dự đoán của phần mềm

3.1.2. Thử nghiệm: Kịch bản 1:

Kịch bản này, chương trình sẽ tạo mô hình huấn luyện với trích xuất đặc trưng HOG và mô hình CNN.

Chương trình khởi đầu bằng việc thu thập dữ liệu từ Intel RealSense Camera, một thiết bị hiện đại có khả năng cung cấp cả hình ảnh 2D và dữ liệu chiều sâu 3D. Cụ thể, camera sẽ ghi lại:

- Hình ảnh 2D: Các ảnh màu hoặc ảnh grayscale của các đối tượng cần định danh.
- Dữ liệu chiều sâu (3D): Tập hợp các giá trị biểu diễn khoảng cách từ camera đến bề mặt vật thể.

Mỗi ảnh và dữ liệu chiều sâu tương ứng sẽ được gán nhãn theo danh tính hoặc lớp mục tiêu của từng đối tượng. Việc gán nhãn này giúp xác định mục tiêu học có giám sát, tức là mô hình học máy sẽ dự đoán nhãn cho các dữ liệu chưa biết trong tương lai. Sau khi thu thập xong, dữ liệu sẽ được tổ chức trong các thư mục riêng biệt để đảm bảo tính hệ thống. Chẳng hạn, ảnh 2D sẽ lưu trong thư mục “TrainData/2D/” và dữ liệu chiều sâu sẽ lưu trong thư mục “TrainData/3D/”.

Sau khi trích xuất đặc trưng từ cả hai loại dữ liệu, chương trình sẽ tiến hành huấn luyện mô hình bằng cách kết hợp các đặc trưng từ HOG và CNN.

Kết hợp đặc trưng: Vector đặc trưng từ các ảnh 2D và 3D sẽ được ghép lại thành một vector tổng hợp. Vector này được chuẩn hóa để đảm bảo tính đồng nhất giữa các đặc trưng.

Tham số huấn luyện mô hình: Trong chương trình có sử dụng thuật toán tối ưu giảm overfitting, epochs: 10 (10 vòng lặp qua toàn bộ dữ liệu) và batch size: 32. Sau khi hoàn tất, mô hình huấn luyện sẽ được lưu dưới dạng file “hog_cnn_model.h5” trong thư mục “TrainData/”.

Để thử nghiệm mô hình học máy có hiệu quả, chương trình có thể import dữ liệu hình ảnh và cho ra kết quả thử nghiệm.



Hình 3.7. Kết quả dự đoán dữ liệu hình ảnh trong kịch bản 1

3.1.3. Thử nghiệm: Kịch bản 2:

Kịch bản này, chương trình sẽ tạo mô hình huấn luyện với trích xuất đặc trưng Hu-moments và mô hình CNN.

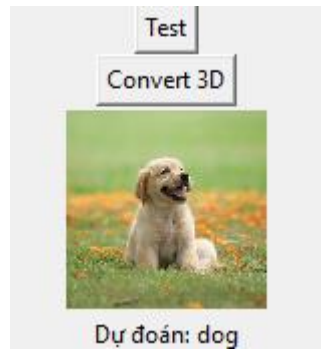
Chương trình khởi đầu bằng việc thu thập dữ liệu từ Intel RealSense Camera, một thiết bị hiện đại có khả năng cung cấp cả hình ảnh 2D và dữ liệu chiều sâu 3D tương tự như kịch bản 1.

Mỗi ảnh và dữ liệu chiều sâu tương ứng sẽ được gán nhãn theo danh tính hoặc lớp mục tiêu của từng đối tượng. Sau đó, ảnh 2D sẽ lưu trong thư mục “TrainData/2D/” và dữ liệu chiều sâu sẽ lưu trong thư mục “TrainData/3D/”.

Sau khi trích xuất đặc trưng từ cả hai loại dữ liệu, chương trình sẽ tiến hành huấn luyện mô hình bằng cách kết hợp các đặc trưng từ Hu-moments và CNN.

Tham số huấn luyện mô hình: Trong chương trình có sử dụng thuật toán tối ưu giảm overfitting, learning rate: 0.0050, momentum: 0.99 để tăng tốc độ hội tụ, dropout: 0.5, epochs: 10 (10 vòng lặp qua toàn bộ dữ liệu) và batch size: 32. Sau khi hoàn tất, mô hình huấn luyện sẽ được lưu dưới dạng file “hu_cnn_model.h5” trong thư mục “TrainData/”.

Để thử nghiệm mô hình học máy có hiệu quả, chương trình có thể import dữ liệu hình ảnh và cho ra kết quả thử nghiệm.



Hình 3.8. Kết quả dự đoán dữ liệu hình ảnh trong kịch bản 2

3.1.4. Thử nghiệm: Kịch bản 3:

Kịch bản này, chương trình sẽ tạo mô hình huấn luyện với trích xuất đặc trưng haar-like features và mô hình CNN.

Chương trình khởi đầu bằng việc thu thập dữ liệu từ Intel RealSense Camera, một thiết bị hiện đại có khả năng cung cấp cả hình ảnh 2D và dữ liệu chiều sâu 3D tương tự như kịch bản 1 và kịch bản 2.

Dữ liệu ảnh 2D sẽ lưu trong thư mục “TrainData/2D/” và dữ liệu chiều sâu sẽ lưu trong thư mục “TrainData/3D/”.

Sau khi trích xuất đặc trưng từ cả hai loại dữ liệu, chương trình sẽ tiến hành huấn luyện mô hình bằng cách kết hợp các đặc trưng từ haar-like features và CNN.

Sau khi hoàn tất, mô hình huấn luyện sẽ được lưu dưới dạng file “haar_cnn_model.h5” trong thư mục “TrainData/”.

Để thử nghiệm mô hình học máy có hiệu quả, chương trình có thể import dữ liệu hình ảnh và cho ra kết quả thử nghiệm.



Hình 3.9. Kết quả dự đoán dữ liệu hình ảnh trong kịch bản 3

3.2. Đánh giá kết quả mô hình học máy

Quá trình dự đoán kết quả từ chương trình tương đối chính xác.

Để cải thiện quá trình định danh đối tượng chính xác hơn nữa, ta cần phải có nhiều dữ liệu đối tượng hơn nữa, hình ảnh cần trực quan trong tập dữ liệu huấn luyện.

Kết quả thử nghiệm các mô hình đã được huấn luyện tỏ ra khá hiệu quả khi ứng dụng tốt để phát hiện đối tượng với nhiều kích thước khác nhau. Mặc dù vậy, do quá trình thu thập dữ liệu trong thời gian tương đối ngắn nên không tránh khỏi việc thiếu dữ liệu dẫn đến mất cân bằng dữ liệu, dữ liệu phân bố giữa các lớp vẫn còn có sự chênh lệch tương đối lớn. Đây cũng là điều cần quan tâm và khắc phục để đạt kết quả thực nghiệm tốt hơn.

Hạn chế:

- Bộ dữ liệu thu thập còn nhỏ và chưa đa dạng.
- Độ chính xác của mô hình định danh cần được cải thiện.
- Thời gian xử lý còn chậm do hạn chế về tài nguyên tính toán.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Qua một thời gian tìm hiểu, nghiên cứu cơ sở lý thuyết cũng như triển khai thực nghiệm, luận văn đã đạt được một số kết quả cụ thể như sau:

Kết quả đạt được

Về mặt lý thuyết:

- Tìm hiểu các mô hình học máy phục vụ việc định danh sử dụng dữ liệu 2D và 3D.
- Nghiên cứu phương pháp thu thập và xử lý dữ liệu từ camera Realsense.

Về mặt ứng dụng:

- Xây dựng được bộ dữ liệu hình ảnh 2D phục vụ cho quá trình định danh.
- Cài đặt và minh họa quá trình định danh bằng mô hình học máy sử dụng dữ liệu từ camera Realsense.

Hướng phát triển của đề tài:

- Mở rộng bộ dữ liệu với nhiều mẫu hơn và đa dạng hơn, sử dụng hình ảnh 3D.
- Áp dụng các phương pháp học máy tiên tiến để nâng cao hiệu suất định danh.
- Tối ưu hóa mô hình để giảm thời gian xử lý và tăng tính ứng dụng thực tế.

Hướng phát triển:

- Tiền xử lý dữ liệu sẽ được chuẩn hóa và giảm nhiễu từ chương trình.
- Nâng cấp và bổ sung thêm các thuật toán trích xuất đặc trưng hoặc các mô hình các đã có sẵn như ImageNet (2D), ShapeNet (3D) để có được nhiều mô hình huấn luyện trong chương trình.

TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt

- [1]. Tiệp, V. H. (2018). Machine learning cơ bản. *Nhà xuất bản Khoa học và Kỹ thuật*.
- [2]. Hoàng Lê Thục Uyên, Nguyễn Văn Đức, Lê Thị Mỹ Hạnh (2018). So sánh đặc trưng moment Hu và biểu đồ gradient có hướng trong nhận dạng tự động hoa cảnh Việt Nam. *Kỷ yếu hội thảo khoa học quốc gia*.

Tài liệu tiếng Anh

- [3]. Anju K S (2015). *Study and Analysis of Edge Detection and Implementation of Fuzzy Set Theory Based Edge Detection Technique in Digital Images*.
- [4]. L. Kalake, Y. Dong, W. Wan and Li Hou (2022). *Enhancing detection quality rate with a combined HOG and CNN for real-time multiple object tracking across non-overlapping multiple cameras*.
- [5]. A. Adouani, Wiem Mimoun Ben Henia; Z. Lachiri (2019). *Comparison of Haar-like, HOG and LBP approaches for face detection in video sequences*.
- [6]. Seyed Yahya Nikouei, Yu Chen, Sejun Song, Ronghua Xu, Baek-Young Choi, Timothy R. Faughnan (2018). *Comparison of Haar-like, HOG and LBP approaches for face detection in video sequences*.
- [7]. MZ Asy'ari, S. Filbert, ZL Sukra (2022). *Histogram of Oriented Gradients (HOG) and Haar Cascade with Convolutional Neural Network (CNN) Performance Comparison in the Application of Edge Home Security System*.
- [8]. H Belmajdoub, R Mafamane, YB Karfa, M Ouadou, K Minaoui (2023). *10 Face Recognition based on CNN, Hog and Haar Cascade Methods using Raspberry Pi v4 Model B*.

Các trang web

- [9]. <https://viblo.asia/p/gioi-thieu-ve-support-vector-machine-svm-6J3ZgPVEImB>, truy cập ngày 01/10/2024.
- [10]. <https://www.gintegro.com/products/geo5-point-cloud>, truy cập ngày 04/10/2024.
- [11]. <https://viblo.asia/p/haar-cascade-la-gi-luan-ve-mot-ky-thuat-chuyen-dung-de-nhan-biet-cac-khuon-mat-trong-anh-E375zamdIGW>, truy cập ngày 05/10/2024.
- [12]. <https://phamdinhhkhanh.github.io/2019/11/22/HOG.html>, truy cập ngày 20/10/2024.

- [13]. <https://nulib-oer.github.io/empirical-methods-polisci/machine-learning.html>, truy cập ngày 15/6/2024.
- [14]. <https://www.datacamp.com/blog/what-is-machine-learning>, truy cập ngày 16/10/2024.
- [15]. <https://pmanigrig.medium.com/introduction-to-convolutional-neural-network-b405c3092e50>, truy cập ngày 17/6/2024.
- [16]. <https://ongxuanhong.wordpress.com/2015/10/29/feature-engineering-la-gi>, truy cập ngày 18/10/2024.