Xác nhận chỉnh sửa luận văn (sau khi bảo vệ xong)

LỜI CAM ĐOAN

Tôi xin cam đoan các kết quả của đề tài "*Xây dựng công cụ tự động nhận dạng cảm xúc dựa trên mặt người*" là công trình nghiên cứu của cá nhân tôi và chưa từng được công bố trong bất cứ công trình khoa học nào khác tính tới thời điểm hiện tại. Các số liệu, kết quả được nêu trong luận văn là trung thực. Các thông tin tham khảo đã được trích dẫn và ghi nguồn đầy đủ. Nếu không đúng như đã nêu trên, tôi xin chịu trách nhiệm về đề tài của mình.

Cần Thơ ngày 31 tháng 7 năm 2018
Người cam đoan

Trầm Vũ Kiệt

LỜI CẢM ƠN

Tôi phép gửi lời cảm ơn chân thành nhất đến Thầy Trương Quốc Bảo - người đã tận tình giúp đỡ và hướng dẫn để tôi có thể hoàn thành đề tài luận văn một cách tốt nhất có thể. Tuy trong thời gian thực hiện luận văn đã gặp phải rất nhiều khó khăn, trở ngại nhưng với sự hỗ trợ của Thầy Trương Quốc Bảo đã tạo cho em niềm tin và kiến thức để thực hiện luận văn.

Tôi xin gửi lời cảm ơn trân quý đến nhà Trường Đại học Cần Thơ, khoa Sau đại học và Khoa Công nghệ thông tin và Truyền thông đã phối hợp tổ chức khóa đào tạo cao học chuyên ngành Hệ thống thông tin để tôi có cơ hội được học tập và nâng cao kiến thức, trình độ chuyên môn, phát triển bản thân.

Bên cạnh đó, tôi cũng xin gửi lời cảm ơn đến gia đình, người thân và bạn bè, những người đã luôn ở bên cạnh và là nguồn động lực để tôi có thể vượt qua những khó khăn về tâm lý, thể trạng để thực hiện tốt luận văn. Tuy nhiên, do sự hạn chế về mặt thời gian, không gian, vị trí địa lý và kinh nghiệm nghiên cứu cá nhân còn non trẻ nên đề tài còn nhiều thiếu sót, rất mong nhận được sự đóng góp và đánh giá của quý Thầy Cô và các bạn học viên.

MỤC LỤC

DANH MỤC CÁC CHỮ VIẾT TẮT VÀ KÝ HIỆU

Từ viết tắt	Cụm từ đầy đủ	Ý nghĩa
AAM	Active Appearance Model	Thuật toán thị giác máy tính so khớp mô hình thống kê của hình dạng đối tượng
ASM	Active Shape Model	Mô hình thống kê về hình dạng đối tượng
PCA	Principal Component Analysis	Phân tích thành phần chính
AUs	Action Units	Các đơn vị vận động
OpenCV	Open Source Computer Vision Library	Thư viện thị giác máy tính mã nguồn mở
SVM	Support Vector Machine	Máy học véc tơ hỗ trợ
ANNs	Artificial Neutral Networks	Mạng nơ ron nhân tạo
HOG	Histogram of Oriented Gradients	Bộ mô tả tính năng dùng để phát hiện đối tượng. Một bộ đặc trưng của các hình dạng đối tượng.
BSD	Berkeley Software Distribution	Bản quyền phân phối phần mềm mã nguồn mở.
DNA	Deoxyribonucleic Acid	Phân tử mang thông tin di truyền
LBP	Local Binary Pattern	Mẫu nhị phân cục bộ. Là một loại của bộ mô tả trực quan dùng trong phân lớp trong thị giác máy tính
CUDA	Compute Unified Device Architecture	Một phiên bản mở rộng của ngôn ngữ lập trình C, được tạo ra bởi nVidia.
GPU	Graphics Processing Unit	Chíp hay bộ xử lý máy tính logic có thể lập trình được, đặc biệt dùng cho các chức năng hiển thị
API	Application Program Interface	Là một tập các chương trình máy tính, giao thức, và công cụ để xây dựng các ứng dụng phần mềm. Xác định các thành phần của phần mềm ảnh hưởng lẫn nhau như thế nào.
Dlib		Một thư viện C++ hiện đại chứa các thuật toán máy học và công cụ để tạo ra các phần mềm phức tạp.
MLP	Multilayer Perception	Mạng nơ-ron đa tầng truyền thẳng
CNN	Convolutional Neutral Network	Mạng nơ-ron tích chập

DANH MỤC BẢNG

DANH MỤC HÌNH

TÓM TẮT

Trong đề tài này, các thuật toán xử lý ảnh và máy học được sử dụng để tự động phát hiện gương mặt người và nhận dạng cảm xúc con người, áp dụng kết hợp đặc trưng cục bộ HOG, đặc trưng Haar-like và máy học véc-tơ hỗ trợ hay mạng nơ-ron nhân tạo. Cơ sở dữ liệu ảnh được dùng để huấn luyện bao gồm JAFFE, Cohn-Kanade và các hình ảnh cá nhân. Hệ thống được xây dựng từ đề tài có khả năng phát hiện và nhận dạng hầu hết các loại cảm xúc cơ bản nhất của con người (vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm) trong thời gian thực. Độ chính xác trung bình của quá trình huấn luyện là *96.4*% và kiểm thử là *90.6*%. Kết quả thể hiện khả năng có thể tích hợp máy học véc-tơ hỗ trợ hay mạng nơ-ron nhân tạo để phát hiện cảm xúc con người trong các ứng dụng thực tế.

Từ khóa: *Nhận dạng cảm xúc gương mặt, Mô hình hình dáng hoạt động, Máy học véc-tơ hỗ trợ, Mạng nơ-ron nhân tạo.*

ABSTRACT

In this thesis, the image processing and machine learning algorithms was used to detect human face and recognize human emotion automatically that implements Histogram of Gradients local features, Haar-like features and Support Véc-tơ Machine or Artificial Neural Network. The databases are included JAFFE, Cohn-Kanade and my own unique images. The system is able to detect human face and recognize almost seven common human universal emotions in real time, includes happiness, sadness, suprises, anger, fear, disgust and neutral (no emotion). An overall training accuracy of *96.4*% and test accuracy of *90.6*% is archieved. The consenquence represents the capable of using Support Vector Machine or Artificial Neutral Network for emotion recognizing in reality projects.

Keywords: *Facial Emotion Recognition, Active Shape Model, Support Vector Machine, Artificial Neural Network.*

LỜI MỞ ĐẦU

CHƯƠNG 1: GIỚI THIỆU

1.1 Lý do chọn đề tài

Khi trò chuyện hoặc quan sát một người nào đó, chúng ta dễ dàng nhận ra cảm xúc hiện tại của họ. Họ vui khi nở một nụ cười trên môi và hai bên gò má nâng lên cao hơn; khi họ cảm thấy khó chịu thì thường nheo hai chân mày lại; hoặc miệng sẽ há thật to, mắt sẽ mở thật to khi nhận được một thông tin nào đó khiến học bất ngờ; và cảm xúc thông dụng nhất là trung tính, có nghĩa là không có cảm xúc tương ứng với các cơ trên gương mặt luôn đặt ở trạng thái bình thường. Vậy làm như thế nào để máy tính cũng có khả năng này như là con người.

Trong giao tiếp, bên cạnh ngôn ngữ cơ thể thì gương mặt là một trong những kênh truyền thông phi ngôn ngữ quan trọng nhất. Ngoài việc biểu hiện cảm xúc, các cử chỉ trên gương mặt còn mang đến nhiều thông tin khác như truyền đạt một tín hiệu, thông điệp giao tiếp (nháy mắt, chớp mắt liên tục) hay là biểu hiện trong các trạng thái đau đớn, khó chịu, các trường hợp bệnh lí, trầm cảm, rối loạn cảm xúc, tự kỷ. Vì vậy, trong nhiều thế kỷ qua, vấn đề nghiên cứu để nhận dạng những cảm xúc thông qua biểu hiện gương mặt đã được thực hiện và không ngừng phát triển bởi các nhà khoa học trên khắp thế giới.

Biểu cảm trên gương mặt là biểu hiện có thể nhìn thấy bằng mắt thường, hiểu được những trạng thái tình cảm, hoạt động nhận thức, tính cách và tâm lí của một người, dự đoán ý định của người đó, đóng góp hơn *55%* hiệu quả trong hoạt động giao tiếp [*theo Mehrabian*]. Những biểu hiện gương mặt cùng với ngôn ngữ cơ thể giúp người nghe có thể hình dung thêm, hiểu rõ hơn về ngữ cảnh đang nói và nắm bắt sâu hơn vấn đề, ý nghĩa và người nói muốn truyền đạt.

Do đó, biểu cảm gương mặt đóng một vai trò vô cùng quan trọng trong tương tác người – máy, đang được nghiên cứu và cải tiến để hỗ trợ cho những ứng dụng thực tế, giúp con người giải quyết các vấn đề dường như là không thể, như đánh giá độ trung thành của một nhân viên đối với công ty, tìm kiếm và truy bắt tội phạm từ hình ảnh vệ tinh, đánh giá phản ứng và sự hài lòng của khách hàng đối với các sản phẩm, dịch vụ của công ty hay các hệ thống hỗ trợ trong y học.

1.2 Mục tiêu đề tài

Mục tiêu của đề tài là tìm hiểu về các loại cảm xúc của con người, các đặc trưng của gương mặt người và sự tương quan giữa cảm xúc và các đặc trưng đó. Nghiên cứu các kỹ thuật, phương pháp, thuật toán để thực hiện các công việc cụ thể trong toàn bộ

quá trình hoạt động của hệ thống, như đặc trưng HOG, mô hình ASM, AAM, các đơn vị vận động trên gương mặt AUs. Thông qua đó, thực hiện huấn luyện được một tập dữ liệu với SVM hay ANN, có khả năng nhận dạng được cảm xúc của con người thông qua gương mặt. Cuối cùng, xây dựng được một hệ thống nhận dạng cảm xúc con người dựa trên các hình ảnh đầu hay video đầu vào.

1.3 Đối tượng, phạm vi nghiên cứu của đề tài

Đối tượng nghiên cứu: phương pháp tiền xử lý hình ảnh, chuyển đổi không gian màu của ảnh, căn chỉnh ảnh; trích rút đặc trưng dựa các kỹ thuật nhận dạng và xác định vị trí của các thành phần trên gương mặt như đặc trưng HOG, ASM/AAM, PCA, AUs; phân lớp và nhận dạng với mô hình máy học véc-tơ hỗ trợ hay là mạng nơ ron nhân tạo, sử dụng thư viện hỗ trợ OpenCV, Dlib.

Phạm vi nghiên cứu: cơ sở lý thuyết và ứng dụng thực tiễn của các phương pháp, kỹ thuật hỗ trợ nhận dạng gương mặt để áp dụng vào hệ thống, dữ liệu đầu vào là các ảnh tĩnh, hoặc video.

1.4 Phương pháp nghiên cứu

Nghiên cứu cơ sở lý thuyết của các kỹ thuật hỗ trợ nhận dạng cảm xúc dựa trên mặt người. Nghiên cứu và so sánh kết quả thực nghiệm với những đề tài cùng chủ đề đã được thực hiện trước đó. Xây dựng một chương trình thực tế để kiểm tra độ chính xác và khả năng mở rộng của đề tài. Ứng dụng kết quả nghiên cứu vào một ngữ cảnh thực tế của đời sống xã hội.

1.5 Kết quả dự kiến

Tìm hiểu được các loại cảm xúc cơ bản của con người trên toàn thế giới; có kiến thức nền tảng về các phương pháp hỗ trợ nhận dạng cảm xúc dựa trên mặt người, xây dựng chương trình kiểm tra độ chính xác với kết quả chấp nhận được và ứng dụng vào một lĩnh vực cụ thể.

1.6 Bố cục luận văn

Bố cục của luận văn được trình bày với năm chương như sau:

Chương 1 giới thiệu về lý do chọn đề tài, các mục tiêu cần đạt được của đề tài, đối tượng nghiên cứu và phạm vi nghiên cứu của đề tài, các phương pháp nghiên cứu của đề tài, cùng với đó là các kết quả dự kiến và bố cục của luận văn.

Chương 2 sẽ trình bày tổng quan về lịch sử giải quyết vấn đề, tính cấp thiết của đề tài; giới thiệu các loại cảm xúc phổ biến của con người và các đặc trưng tương ứng trên

gương mặt của các loại cảm xúc đó; tập trung làm rõ một số phương pháp giúp nhận dạng cảm xúc dựa trên mặt người và cơ sở lý thuyết của các thuật toán, khái niệm là thành phần cốt lõi trong quá trình nhận dạng cảm xúc dựa trên mặt người.

Nội dung nghiên cứu của đề tài, các nghiên cứu liên quan ở trong và ngoài nước, định hướng giải quyết luận văn và quy trình thực hiện luận văn sẽ được trình bày chi tiết từng bước trong *Chương 3*.

Chương 4 trình bày các công việc cài đặt thực nghiệm thuật toán nghiên cứu, thể hiện các kết quả và đánh giá thành quả đạt được.

Phần kết luận, những thách thức trong lĩnh vực nhận dạng cảm xúc và hướng phát triển của đề tài được đưa ra trong *Chương 5*.

Cuối cùng là phần Tài liệu tham khảo.

CHƯƠNG 2: TỔNG QUAN

2.1 Lịch sử giải quyết vấn đề

Thế giới đang bước vào cuộc cách mạng công nghiệp 4.0 với những bước chuyển biến tích cực và mạnh mẽ của tất cả các ngành nghề trong mọi lĩnh vực của đời sống hiện đại. Cùng với đó là sự phát triển vượt bậc của khoa học và công nghệ, sự hỗ trợ thiết yếu của công nghệ thông tin trong việc tin học hóa và đơn giản hóa các quy trình làm việc của mọi lĩnh vực khác nhau. Để thực hiện được điều đó, cần có sự tương tác giữa người và máy để máy tính có thể hiểu và thực hiện những công việc tự động theo mong muốn của con người.

Thật vậy, để máy tính có thể giao tiếp với con người thì chúng ta cần có những phương pháp và kỹ thuật cụ thể, một trong số chúng là khả năng nhận dạng được cảm xúc của con người. Một trong những phương pháp vật lý và có hiệu quả để nhận dạng được cảm xúc của con người là thông qua cả biểu cảm trên gương mặt. Đây là chủ đề chính trong các nghiên cứu của các nhà khoa học ở nhiều năm trước đây, và hiện tại nó vẫn được xem là một đề tài hấp dẫn vì tính ứng dụng và độ phổ biến.

Các ứng dụng thực tế của việc nhận dạng cảm xúc là rất đa dạng và hữu ích. Trong đời sống, các ứng dụng di động nhận dạng được cảm xúc của người dùng để gán các biểu tượng cảm xúc tương ứng như snow, magic, polygram đang được giới trẻ rất ưa chuộng. Trong y học, các bác sĩ có thể theo dõi các thay đổi cảm xúc của bệnh nhân để đưa ra những chẩn đoán bệnh chính xác và điều trị bệnh hiệu quả với các ca thần kinh hay rối loạn cảm xúc. Trong thương mại, các tập đoàn, doanh nghiệp, nhà sản xuất có

thể thu thập, phân tích và thống kê cảm xúc của khách hàng để đưa ra những quyết định kịp thời và đúng đắn nhằm tối ưu hóa lợi nhuận.

2.2 Tính cấp thiết của đề tài

Từ những ứng dụng thiết thực trên, việc nhận dạng cảm xúc của con người dựa vào biểu cảm trên gương mặt là một chủ đề rất hay, tuy không mới nhưng tính khoa học và thực tiễn cao. Các nhà nghiên cứu đi trước đã có những công trình nghiên cứu hay với các phương pháp nghiên cứu khác nhau đã cho kết quả và độ chính xác nhất định, đây cũng làm một động lực để các đề tài sau có cơ sở khoa học để so sánh và đánh giá, cải tiến hơn nữa hiệu quả mang lại.

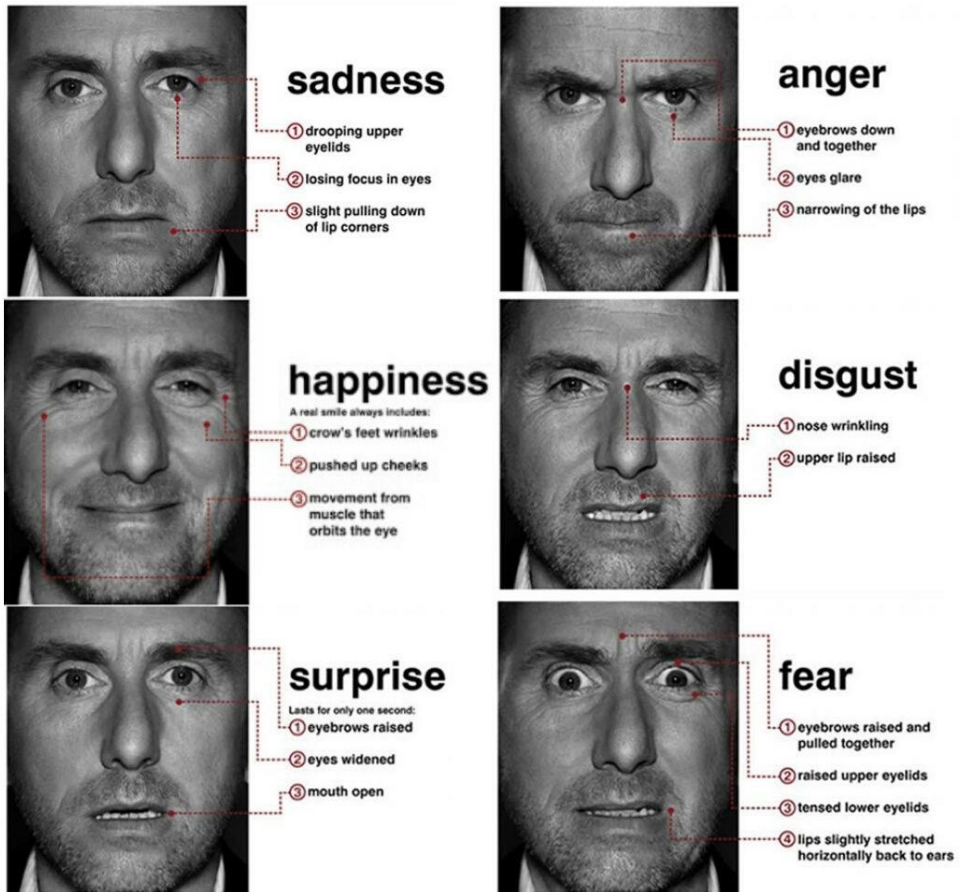
Với sự phát triển và phổ biến của mạng xã hội và công nghệ thông tin như hiện nay, việc tạo ra các ứng dụng để hiểu được người dùng hơn cả người dùng hiểu chính họ là một điều tuyệt vời. Và việc nhận dạng cảm xúc của con người dựa trên gương mặt là một nền tảng cốt lõi cho các ứng dụng này.

2.3 Các loại cảm xúc của con người và các đặc trưng trên gương mặt người

2.3.1 Các trạng thái cảm xúc của con người

Cảm xúc có tác động rất lớn đến cuộc sống của con người. Cảm xúc là một thứ rất phức tạp, có thể thay đổi nhanh chóng, một người có thể có hơn một cảm xúc tại một thời điểm. Vì vậy, cảm xúc là gì và con người có tất cả bao nhiêu cảm xúc là các câu hỏi chưa có một đáp án thuyết phục. Theo Paul Ekman và Wallace V. Freisen, có sáu cảm xúc cơ bản với tất cả các nền văn hóa khác nhau trên thế giới, bao gồm vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm.

Vào thế kỷ IV trước công nguyên, Aristole đã nhận định có 14 loại cảm xúc cơ bản nhất, bao gồm: hài lòng, tử tế, tranh đua, ganh tị, đáng thương, câm phẫn, sợ hãi, bình tĩnh, tự tin, thù địch, giận dữ, xấu hổ, bằng hữu và vô liêm sĩ. Những năm gần đây, sáu là con số đại diện cho số lượng cảm xúc cơ bản nhất của con người được các nhà tâm lý học công bố, gồm có: vui vẻ, buồn bã, ngạc nhiên, sợ hãi, ghê tởm, giận dữ.



Hình 1: Sáu cảm xúc phổ biến nhất của con người với các biểu hiện cụ thể của từng thành phần trên gương mặt

Ở một nhận định khác, TS. Rachael Jack của đại học Glasgow, Vương quốc Anh cho rằng con người chỉ có bốn loại cảm xúc cơ bản khi mà những biểu hiện của các cơ trên gương mặt trong hai cảm xúc sợ hãi và ngạc nhiên là như nhau, tương tự với giận dữ và ghê tởm. Thật sự là hai loại cảm xúc trong từng cặp ở trên có sự giống nhau trong quá trình vận động các nhóm cơ mặt để hình thành cảm xúc, chúng chỉ khác nhau khi được thể hiện đầy đủ và xong xuôi, vì thực chất chúng là những cảm xúc khác nhau.

Đề tài được thực hiện nhận dạng sáu loại cảm xúc cơ bản nhất của con người (vui vẻ, buồn bã, ngạc nhiên, sợ hãi, ghê tởm, giận dữ).

2.3.2 Đặc trưng của gương mặt người

Gương mặt là một phần của cơ thể người, bộ phận trung tâm để bộ lộ cảm xúc và là phương tiện truyền tải cảm xúc giữa người với người. Gương mặt người có các thành phần đặc trưng giống nhau như chân mày, mắt, mũi, miệng, v.v. nhưng mỗi

người lại có một gương mặt khác nhau và là duy nhất. Do đó, gương mặt là đặc trưng tốt nhất để phân biệt một người với những người khác.

Gương mặt là nơi biểu hiện của những cảm xúc. Một nụ cười thể hiện cho một niềm vui, một cái chau mày đồng nghĩa với việc không tán thành hay khó chịu. Chính vì thế, nhận dạng cảm xúc của người đối diện qua gương mặt có một vai trò quan trọng trong giao tiếp. Con người có thể đọc được cảm xúc của người khác nhờ biểu cảm trên gương mặt của người đó, từ đó dự đoán được khả năng xảy ra của các hành vi tiếp theo.

Cơ mặt đóng một vai trò nổi bật trong việc mô tả cảm xúc con người, cùng với các đặc trưng khác trên gương mặt mang đến sự đa dạng trong sự biểu lộ nhiều cảm xúc khác nhau.

2.3.3 Nhận dạng cảm xúc dựa trên mặt người

Gương mặt là nơi cảm xúc được bộc lộ rõ nhất và dễ dàng nhận thấy nhất. Chính vì thế các biểu hiện trên gương mặt người được sử dụng như là một nền tảng cốt lõi để nhận dạng cảm xúc con người.

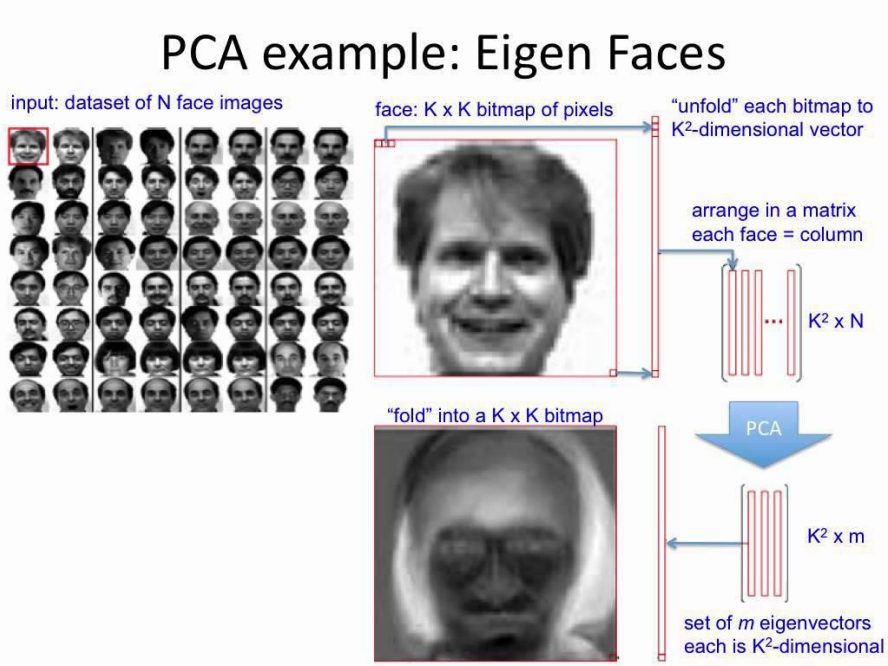
Tương ứng với từng cảm xúc riêng biệt là những biểu hiện khác nhau trên gương mặt. Như vậy, với một loại cảm xúc nhất định sẽ có một tập các biểu hiện nhất định của các nhóm cơ và các đặc trưng khác trên gương mặt.

Đối với một người, việc quan sát gương mặt và nhận dạng cảm xúc của một người khác là một điều rất dễ dàng. Vậy đối với máy tính thì như thế nào. Chúng cần được học, được huấn luyện để có được khả năng này như con người.

2.4 Các phương pháp giúp nhận dạng cảm xúc dựa trên gương mặt

2.4.1 Phương pháp dựa trên đặc trưng của gương mặt

Sử dụng phương pháp phân tích thành phần chính **PCA**. Phương pháp này sẽ lấy ra được các thành phần chính của gương mặt (thể hiện bằng các véc-tơ riêng - eigenvectors) trong tập ảnh huấn luyện và tạo ra không gian mặt. Tiếp tục sử dụng các thuật toán máy học để huấn luyện tạo thành các tập dữ liệu huấn luyện là các lớp tương ứng với các loại cảm xúc cơ bản.



Hình 2: Ví dụ về Faces và EigenFaces của nó

PCA là một công cụ mạnh mẽ cho việc xác định hình dạng của gương mặt, phân tích các thành phần trên gương mặt. **PCA** là một công cụ giảm chiều dữ liệu, có khả năng cắt giảm một tập hợp lớn các biến thành một tập nhỏ hơn mà vẫn giữ được hầu hết các thông tin quan trọng của tập hợp ban đầu.

2.4.2 Phương pháp sử dụng các đơn vị vận động trên gương mặt

Cảm xúc được xác định dựa trên sự chuyển động của các đơn vị trên khuôn mặt, được gọi là các **Action Units**. Ban đầu, nó được tạo ra bởi Carl-Herman Hjortsjö với 23 AUs vào năm 1970, sau đó được phát triển bởi Paul Ekman và Wallace Friesen. Có tất cả 64 action unit tương ứng với 64 biểu hiện khác nhau của các nhóm cơ trên gương mặt [Ekman và Friesen 1978]. Việc xác định cảm xúc chỉ đơn giản là việc xác định có bao nhiêu action unit cùng xuất hiện trên gương mặt tại một thời điểm, và sự kết hợp của chúng sẽ cho ra một cảm xúc duy nhất. Ví dụ: cảm xúc vui là kết quả của sự kết hợp hai action unit 6 (má nâng lên) và 12 (góc ở mép môi đưa lên cao), cảm xúc buồn gồm có các action unit 1 (vầng trán nâng lên), 4 (chân mày hạ xuống) và 15 (góc ở mép môi đưa hạ xuống) biểu hiện đồng thời. [(1)]

Một số ví dự về các đơn vị vận động trên gương mặt:

Hình 3: Một số ví vụ về các đơn vị vận động (action units) trên khuôn mặt người

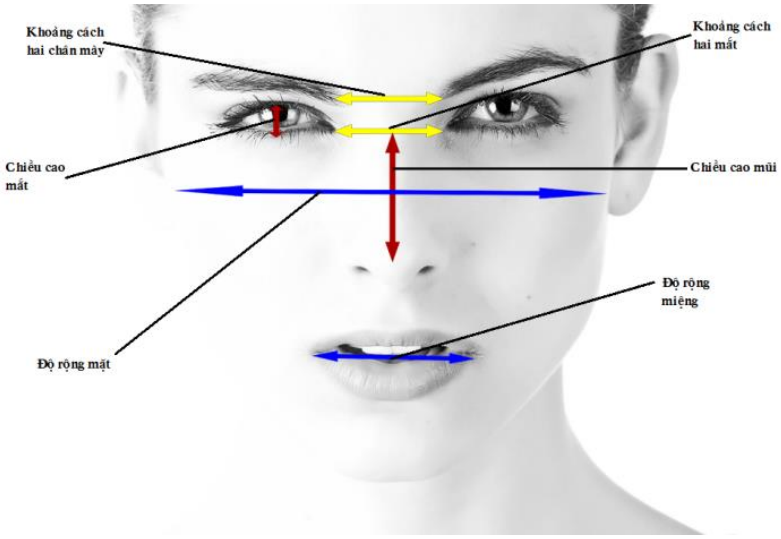
Ưu điểm của **AUs** là một phương pháp rất dễ hiểu và dễ thực hiện, thuật toán đơn giản, dễ tiếp cận với người sử dụng. Ứng dụng được cho các ứng dụng nhận dạng thời gian thực.

Nhược điểm của phương pháp này là mang tính chủ quan cao và vấn đề độ chính xác thực hiện cần được cải thiện.

2.4.3 Sử dụng mô hình AAM kết hợp tương quan điểm.

Đầu tiên, mô hình AAM được sử dụng để phát hiện vùng mặt, sau đó dựa vào sự phân bố hiện tại của các landmark và tỉ lệ giữa chúng để biết được tương ứng với loại cảm xúc nào. Ý tưởng chính là: Nếu có thể xác định được chính xác tọa độ các điểm trên khuôn mặt thì có thể dựa vào tương quan các điểm đó để nhận dạng cảm xúc.

Ví dụ một cách chi tiết được trình bày như hình bên dưới:



Hình 4: Các vị trí tương quan điểm của các thành phần chính của gương mặt người.

Gọi:

$$R1 = \frac{\text{Độ } rộng\ miệng}{\text{Độ } rộng\ gương\ mặt}$$

$$R2 = \frac{Chiều\ cao\ mắt}{Chiều\ cao\ mũi}$$

$$R3 = \frac{Khoảng\ cách\ hai\ chân\ mày}{Khoảng\ cách\ hai\ mắt}$$

Với R1: Khi gương mặt ở trạng thái bình thường, độ rộng miệng luôn nhỏ hơn độ rộng gương mặt, nên R1 luôn nhỏ hơn 1. Khi cười, miệng của chúng ta sẽ rộng hơn bình thường và độ rộng miệng tăng lên, độ rộng gương mặt vẫn giữ nguyên, kéo theo R1 sẽ tăng lên.

Với R2: Khi gương mặt ở trạng thái bình thường, chiều cao mắt luôn nhỏ hơn chiều cao của mũi, nên R2 luôn nhỏ hơn 1. Khi chúng ta nhạc nhiên vì một điều gì đó, mắt chúng ta sẽ mở to và rộng hơn, đồng thời là miệng há to làm cho độ rộng miệng nhỏ lại. Điều đó làm cho R2 tăng lên và R1 giảm đi.

Với R3: Khi gương mặt ở trạng thái bình thường, khoảng chắc giữa hai chân mày và khoảng cách giữa hai mắt là tương đương nhau, nên R3 xấp xỉ 1. Khi chúng ta giận dữ hoặc khó chịu với một điều gì đó, chúng ta thường nheo mày lại và làm cho khoảng cách giữa hai chân mày thu ngắn lại, kéo theo R3 giảm.

Để chi tiết hơn, các trường hợp được miêu tả cụ thể trong bảng bên dưới (đây chỉ là một số trường hợp đại diện cho thuật toán):

Bảng 1: Cảm xúc tương ứng với các tương quan điểm

Cảm xúc	R1	R2	R3
Trung tính	*Bình thường*	*Bình thường*	*Bình thường*
Vui vẻ	**Tăng**	*Bình thường*	Giảm
Ngạc nhiên	*Bình thường*	**Tăng**	*Bình thường*
Giận dữ	**Tăng**	*Bình thường*	Giảm
Buồn bã	Giảm	*Bình thường*	Giảm

Phương pháp này có *ưu điểm* rất lớn khi được sử dụng trong nhận dạng cảm xúc thời gian thực vì có tốc độ xử lý nhanh, thường được dùng trong nhận dạng cảm xúc người dùng qua video hay người dùng thực tế. Ít bị ảnh hưởng của yếu tố nền khuôn mặt lên kết quả nhận dạng. Việc nhận dạng có thể được thực hiện trong điều kiện gương mặt chịu ảnh hưởng của các phép biến hình (như phép tịnh tiến, tỉ lệ, xoay).

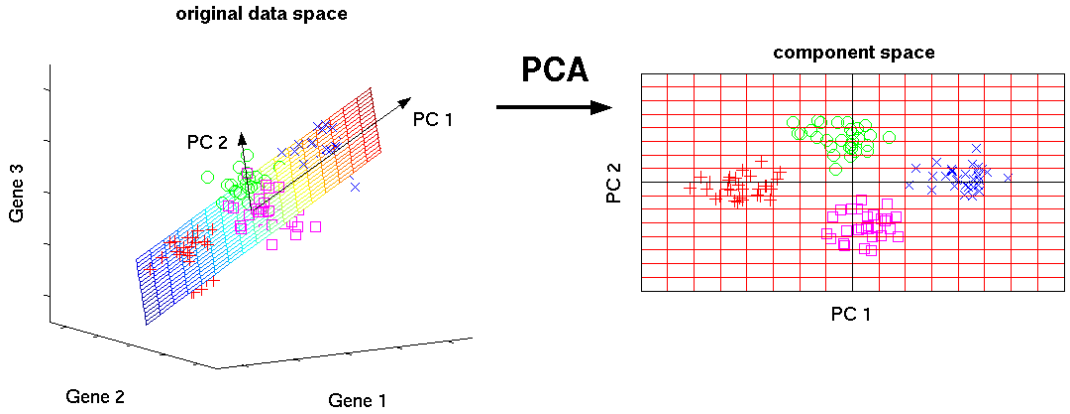
Một *hạn chế* của phương pháp này là việc xác định được ngưỡng tỉ lệ để nhận định đó là cảm xúc nào. Cần cải tiến thuật toán xác định từng điểm của gương mặt để tăng tốc độ nhận dạng và độ chính xác của hệ thống.

2.5 Cơ sở lý thuyết

2.5.1 Principal Component Analysis

PCA (Phân tích thành phần chính) là một trong những phương pháp phân tích dữ liệu nhiều biến (chiều) đơn giản nhất, là một kỹ thuật dùng để làm nổi bật sự thay đổi và đưa ra các mô hình chính yếu trong tập dữ liệu, mô hình hóa tập dữ liệu, giúp cho việc phân tích và khai phá tập dữ liệu trở nên dễ dàng và chính xác.

Phần lớn dữ liệu thực tế đều là dữ liệu nhiều chiều: dữ liệu hình ảnh, video, audio, văn bản, chữ viết tay, cảm biến. Trong trường hợp này, phương pháp phân tích thành phần chính được sử dụng để giảm chiều dữ liệu, trích rút các các thành phần chính cần được giữ lại mà vẫn đảm bảo đầy đủ thông tin của dữ liệu ban đầu. **PCA** là một kỹ thuật có tính ứng dụng cao trong việc nhận dạng gương mặt, phổ biến để phát hiện các mẫu của0020dữ liệu nhiều chiều.



Hình 5: Giảm chiều dữ liệu từ ba chiều về hai chiều trong phương pháp phân tích thành phần chính PCA

Các tính năng chính của **PCA** được trình bày như sau:

- Giảm số chiều của dữ liệu quan sát. Được sử dụng khi các mẫu và thông tin trong dữ liệu khó nhận ra trong không gian đa chiều ban đầu.
- **PCA** xây dựng một không gian mới với trục tọa độ mới có số chiều ít hơn, nhưng có khả năng thể hiện dữ liệu tương đương hoặc tốt hơn không gian cũ, nghĩa là đảm bảo độ biến thiên của dữ liệu trên mỗi chiều mới.
- Các trục tọa độ trong không gian mới là tổ hợp tuyến tính của các trục tọa độ ở không gian cũ, được xây dựng sao cho độ biến thiên của dữ liệu trên mỗi trục là lớn nhất có thể.
- Trong không gian mới, các thông tin của dữ liệu có thể được phát hiện ở một khía cạnh khác mà ở không gian cũ không tìm thấy được, vì các liên kết tìm ẩn của dữ liệu có thể được khai phá.

Một số khái niệm toán học quan trọng được sử dụng trong **PCA**: độ lệch chuẩn (*standard deviation*), phương sai (*variance*), hiệp phương sai (*covariation*), giá trị riêng (*aigenvalue*) và vector riêng (*eigenvector*). Tất cả các khái niệm này được trình bày cụ thể ở phần bên dưới.

2.5.1.1 Độ lệch chuẩn

Độ lệch chuẩn là một đại lượng dùng để đo khoảng cách giữa các phần tử trong tập dữ liệu. Độ lệch chuẩn còn được hiểu là khoảng cách trung bình từ trung bình mẫu ($\overline{X}$) đến các điểm của dữ liệu. Nó cho thấy sự chênh lệch về giá trị của từng phần tử so với giá trị trung bình.

Giả sử ta có tập dữ liệu $X = [x_1, x_2, ..., x_n]$ có n phần tử, phần tử thứ i là x_i. Vậy trung bình mẫu $\overline{X}$ là một giá trị nằm giữa của tập X, được tính theo công thức:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Công thức tính độ lệch chuẩn (s):

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}}$$

2.5.1.2 Phương sai

Phương sai là bình phương của độ lệch chuẩn. Phương sai là một đại lượng khác dùng để biểu diễn dữ liệu: đo khoảng cách giữa các phần tử trong tập dữ liệu. Phương sai của một biết ngẫu nhiên là thước đo sự phân tán thống kê của biến đó, nó hàm ý giá trị của biến đó thường ở cách giá trị kỳ vọng là bao xa.

Công thức tính phương sai (s^2):

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}$$

2.5.1.3 Hiệp phương sai

Hai khái niệm *độ lệch chuẩn* và *phương sai* được sử dụng để biểu diễn dữ liệu một chiều, nhưng dữ liệu trong thực tế thường có nhiều hơn một chiều và có sự liên hệ với nhau mật thiết. Do đó, đại lượng *hiệp phương sai* ra đời để có thể tính toán và biểu diễn được dữ liệu đa chiều.

Hiệp phương sai thực chất chỉ tính toán được sự biến thiên của hai chiều dữ liệu. Nên ta có thể tính từng cặp chiều dữ liệu cho toàn bộ chiều của tập dữ liệu.

Hiệp phương sai là độ đo sự biến thiên cùng nhau của *hai biến* ngẫu nhiên thay vì đo mức độ biến thiên của *một biến* ngẫu nhiên – công việc của *phương sai*.

Công thức tính hiệp phương sai *cov(X, Y)*:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

2.5.1.4 Véc-tơ riêng

Với điều kiện số cột của ma trận thứ nhất (**A**) bằng với số dòng của ma trận thứ hai (**B**), ta có thể nhân hai ma trận với nhau theo thứ tự tương ứng **AxB**. Kết quả của

phép nhân ma trận này có một số trường hợp đặc biệt, véc-tơ đầu ra là một bội số của véc-tơ gốc, và chúng được gọi là các *véc-tơ riêng* (*eigenvector*).

Các tính chất của một *véc-tơ riêng*:

- Chỉ có các ma trận vuông (kích thước *nxn*) thì mới có véc-tơ riêng.
- Nhưng, không phải mọi ma trận vuông đều có véc-tơ riêng.
- Nếu một ma trận vuông (kích thước *nxn*) có véc-tơ riêng, thì sẽ có số lượng véc-tơ riêng là *n*.
- Tất cả các véc-tơ riêng của một ma trận đều trực giao với nhau.
- Véc-tơ riêng có tính biến đổi, khi nhân với một số thì kết quả sau khi nhân với ma trận chuyển đổi vẫn là véc-tơ ban đầu.

2.5.1.5 Giá trị riêng

Giá trị riêng là một khái niệm song song với véc-tơ riêng. Một giá trị riêng nhân với hai véc-tơ riêng bằng nhau được gọi là giá trị riêng ứng với một véc-tơ riêng. Các véc-tơ riêng cũng là tiêu chuẩn để chọn ra các giá trị riêng thỏa mãn với yêu cầu của một bài toán nào đó.

Giá trị riêng là nghiệm của phương trình đặc trưng: $\det(\mathrm{A} - \lambda \mathrm{I}) = 0$ với A là ma trận vuông đầu vào, λ là biến giá trị riêng cần tìm. Một giá trị riêng có thể có nhiều véc-tơ riêng nhưng mỗi véc-tơ riêng chỉ ứng với một giá trị riêng duy nhất.

Phương pháp tìm *giá trị riêng* và *véc-tơ riêng*:

- **Bước 1**: Giải phương trình đặc trưng tìm *giá trị riêng*: $\det(\mathrm{A} - \lambda \mathrm{I}) = 0$
- **Bước 2**: Giải hệ phương trình thuần nhất tìm *véc-tơ riêng* u_i ứng với *giá trị riêng* λ_i: $\det(\mathrm{A} - \lambda \mathrm{I})u = 0$

2.5.1.6 Các bước thực hiện cơ bản của PCA

Bước 1: Lấy dữ liệu đầu vào. Tính trung bình mẫu, hay còn gọi là vector kỳ vọng ($\hat{x}$) của toàn bộ dữ liệu.

Công thức: $\hat{x} = \frac{1}{N}\sum_{n=1}^{N} x_n$

Bước 2: Trừ mỗi điểm dữ liệu đi một lượng vector kỳ vọng của toàn bộ dữ liệu. Xét chiều dữ liệu ở chiều x đều có một giá trị trung bình mẫu. Thực hiện trừ lần lượt các giá trị chiều x cho trung bình mẫu.

Công thức: $\hat{x}_n = x_n - \hat{x}$

Bước 3: Tính ma trận hiệp phương sai.

Công thức: $S = \frac{1}{N}\hat{X}\hat{X}^T$

Bước 4: Tính các vector riêng và giá trị riêng của ma trận hiệp phương sai. Sắp xếp kết quả theo thứ tự giảm dần của giá trị riêng.

Bước 5: Chọn ra các thành phần chính. Chọn ra **K** vector riêng ứng với **K** giá trị riêng lớn nhất để xây dựng ma trận U_K có các cột tạo thành một hệ trực giao. **K** vector này còn được gọi là các thành phần chính, tạo thành không gian con gần với phân bố của dữ liệu ban đầu đã chuẩn hóa. Tùy vào số lượng thành phần chính yêu cầu, lấy lần lượt các thành phần (các vector riêng) tương ứng có các giá trị riêng cao nhất.

Bước 6: Chiếu dữ liệu ban đầu đã chuẩn hóa xuống không gian con tìm được. Dữ liệu mới chính là tọa độ của các điểm trong không gian mới.

$$Z = U_K^T \hat{X}$$

2.5.1.7 Một số hạn chế của PCA

PCA chỉ làm việc với dữ liệu số, vì vậy cần phải có một bước tiền xử lý nếu dữ liệu đầu vào không phải là số.

Độ chính xác của thuật toán còn phụ thuộc nhiều vào điệu kiện ngoại cảnh như ánh sáng, phông nền.

Nhạy cảm với các điểm outlier/extreme. Khi góc mặt nghiêng hay quá xa với webcam hay công cụ ghi hình, thuật toán có thể sẽ cho ra kết quả nhận dạng sai.

Do PCA hoàn toàn dựa trên các biến đổi tuyến tính, nên không phù hợp với các mô hình phi tuyến.

Để đạt được độ chính xác cao hơn, cần sử dụng nhiều hình ảnh để huấn luyện, kéo theo tốc độ xử lý sẽ chậm hơn.

Phương pháp PCA không phù hợp với các ứng dụng đòi hỏi xứ lý nhanh và thời gian thực.

2.5.2 Mô hình ASM/AAM

Một trong những bước quan trọng trong nhận dạng cảm xúc gương mặt là định vị chính xác được các điểm điều khiển thể hiện trạng thái khuôn mặt. Và hai mô hình ASM/ AAM có chức năng để thực hiện công việc này. Các biến dạng bị ràng buộc bởi PDM (Mô hình phân phối điểm) để chỉ thay đổi trong các cách nhìn thấy trong tập

huấn luyện và gán nhãn. Hình dạng của đối tượng được trình bày bằng một tập các điểm. Mục đích của thuật toán là để so khớp mô hình với một hình ảnh hoàn toàn mới.

Kỹ thuật này được sử dụng rộng rãi để phân tích hình ảnh gương mặt hỗ trợ trong nhận dạng, y học, điều khiển robot.

2.5.2.1 Mô hình ASM

ASM là một mô hình thống kê của hình dạng đối tượng, thực hiện vòng lặp biến dạng để so khớp với một hình ảnh mới của đối tượng. **ASM** được phát triển bởi Tim Cootes, Chris Taylor vào năm 1995. Các hình dáng của đối tượng được ràng buộc bởi mô hình phân phối điểm (**PDM**), mô hình hình dạng thống kê chỉ thay đổi theo các cách đã được nhìn thấy trong tập các dữ liệu đã được gán nhãn từ trước. Hình dáng của đối tượng được biểu diễn bởi một tập các điểm được điều khiển bởi mô hình hình dáng. Mục đích chính của mô hình **ASM** là để so khớp mô hình với hình ảnh mới.

Mô hình **ASM** đại diện cho một mô hình biến dạng tham số, là một mô hình thống kê các thay đổi tổng thể của hình dáng đối tượng được xây dựng từ một tập huấn luyện. Mô hình này được gọi là mô hình phân phối điểm **PDM**, sau khi được xây dựng, nó được sử dụng để so khớp một mô hình hay một mẫu mới và chưa được nhận dạng của đối tượng được gán nhãn trước đó trong tập huấn luyện. Mô hình phân phối điểm được xây dựng phổ biến nhất là bởi thuật toán phân tích thành phần chính **PCA**.

Các bước xây dựng mô hình phân phối điểm được mô tả tổng quát như sau:

Bản thân hình dáng của đối tượng (như hình ảnh) được biểu diễn như một đa giác n điểm trong tọa độ ảnh:

$$X = (x_1, y_1, \dots, x_{n-1}, y_{n-1}, x_n, y_n)^T$$

Bước 1: Tính phương sai của hình dạng X. Để tính toán chính xác, cần chuyển đổi X về một khung dạng bình thường với các tham số được đặt ra bao gồm: phép dịch chuyển t_x, t_y, phép biến đổi tỉ lệ s và phép quay θ. Ta được công thức như sau:

$$x = T_{t_x,t_y,s,\theta}(X)$$

Bước 2: Tính giá trị trung bình của x:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Bước 3: Tính độ lệch chuẩn của mỗi hình dạng so với giá trị trung bình:

$$dx_i = x_i - \bar{x}$$

Bước 4: Tính ma trận hiệp phương sai theo công thức:

$$cov = \frac{1}{n}\sum_{i=1}^{n} dx_i dx_i^T$$

Bước 5: Trục số chính của tập hợp các điểm hai chiều bây giờ được xem như là các véc-tơ riêng của ma trân hiệp phương sai p_i. Giả sử λ_i được ký hiệu cho giá trị riêng thứ i của p_i, ta có được:

$$\sum p_i = \lambda_i p_i$$

Bước 6: Sắp xếp các véc-tơ riêng theo thứ tự giảm dần của các giá trị riêng tương ứng, ta được ma trận P:

$$P = [p_1 \dots p_{2n}]$$

Bước 7: Một thể hiện của hình dạng đối tượng có thể được phát sinh bằng cách biến đổi giá trị trung bình với sự kết hợp tuyến tính của các véc-tơ riêng:

$$x = \bar{x} + Pb$$

Bước 8: Bây giờ thể hiện của mô hình được định nghĩa bởi véc-tơ v của nó:

$$v = \{t_x, t_y, s, \theta, b\}$$

Với b là tham số thể hiện số chiều của không gian sau khi tập huấn luyện được biến đổi trực giao.

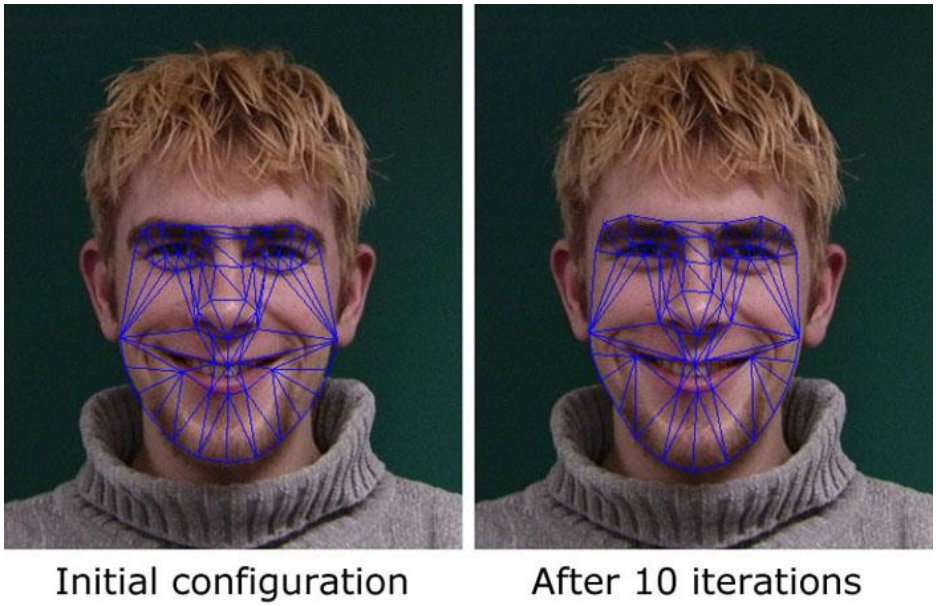
2.5.2.2 Mô hình AAM

$$\det(\mathrm{A} - \lambda \mathrm{I}) = 0$$

AAM là một mô hình cải tiến từ ASM. Một trong những nhược điểm của ASM là chỉ sử dụng các ràng buộc về hình dạng đối tượng với nhau với một vài thông tin về cấu trúc hình ảnh, và không tận dụng lợi thế của tất cả thông tin có sẵn, các kết cấu trên các đối tượng mục tiêu. Và vấn đề này có thể giải quyết với AAM.

AAM đầu tiên được ra mắt bởi Edwards, Cootes và Taylor trong đề tài phân tích khuôn mặt tại Hội nghị quốc tế lần thứ 3 về nhận dạng gương mặt và cử chỉ vào năm 1998. Hướng tiếp cận này đã được ứng dụng rộng rãi trong theo dõi và so khớp trong y học.

AAM là một thuật toán phổ biến trong lĩnh vực thị giác máy tính, mục tiêu tối ưu một mô hình thống kê hình ảnh thể hiện của đối tượng vào một ảnh đầu vào mới. Kết quả của quá trình tối ưu là một bộ điểm điều khiển thể hiện cấu trúc của một đối tượng đã được học có các tọa độ tương ứng với thể hiện trong ảnh đầu vào của đối tượng; bên cạnh đó là một bộ các tham số mô hình thống kê đã được ước lượng, được sử dụng để tái cấu trúc hình dạng, kết cấu hình ảnh của đối tượng tương ứng một cách tương đối.



Hình 6: Ví dụ về mô hình AAM áp lên gương mặt người

Mô hình thống kê của đối tượng cần đảm bảo có thể mô tả được những biến thể về hình dạng và kết cấu hình ảnh, mối tương quan giữa chúng. Vấn đề chính yếu trong phương pháp này là việc xây dựng mô hình thống kê cho đối tượng ảnh và việc thiết ké thuật toán tối ưu cho tìm kiếm.

Mô hình hình dạng của đối tượng được biểu diễn bởi một tập hợp có thứ tự các điểm điều khiển.

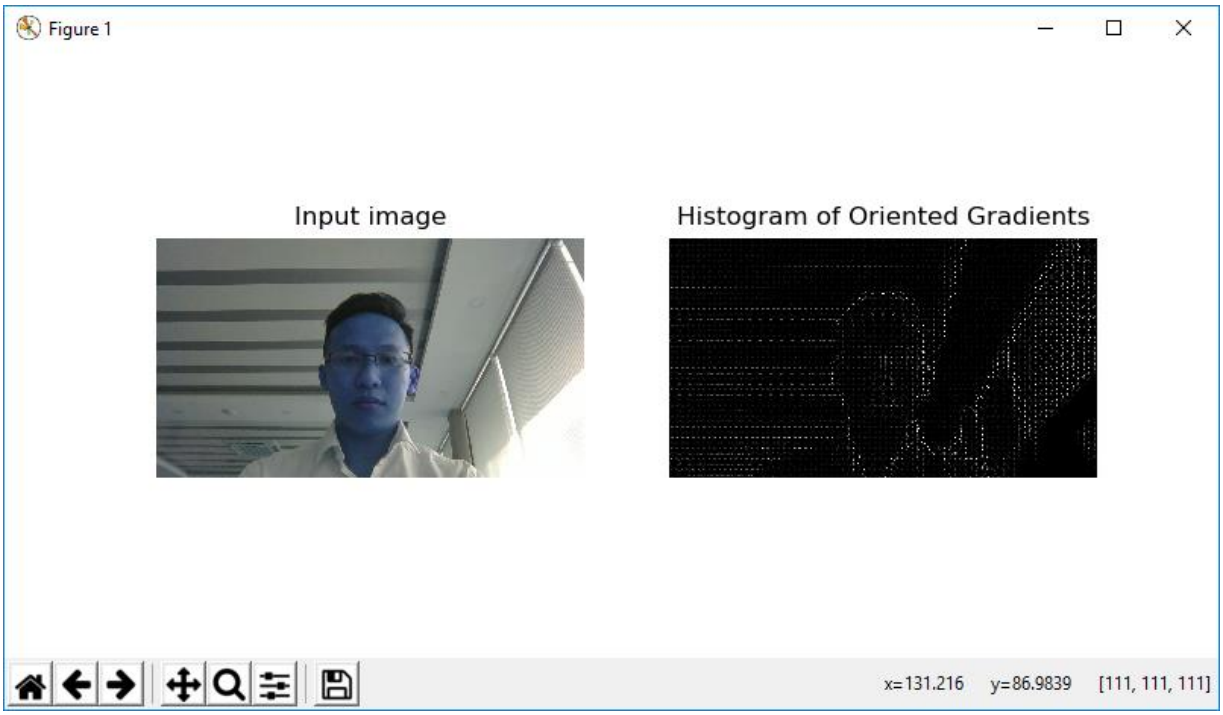
2.5.3 Đặc trưng HOG

HOG *(Histogram of Oriented Gradients)* là một bộ mô tả tính năng được sử dụng để phát hiện đối tượng trong thị giác máy tính và xử lý ảnh. **HOG** được tính toán trên một lưới dày đặc các ô và chuẩn hóa sự tương phản giữa các khối để nâng cao độ chính xác. **HOG** được dùng chủ yếu để phát hiện và mô tả hình dạng của một đối tượng trong ảnh

Ban đầu, đặc trưng HOG được thiết kế để phát hiện đối tượng người trong dữ liệu hình ảnh, sau đó được cải thiện và phát triển rộng rãi hơn trong lĩnh vực phát hiện

đối tượng nói chung. Ý tưởng chính của thuật toán này là dựa trên việc đếm số lần xuất hiện của các hướng gradient (*gradient orientation*) trong các vùng cục bộ của ảnh. (2 p. 11)

Các thông số về hình dáng và bề ngoài của các đối tượng cục bộ trong ảnh được mô tả bằng cách sử dụng thông tin phân bố của các gradients cường độ (intensity gradients) và các hướng của cạnh (edge directions). Thuật toán **HOG** được tiến hành bằng cách chia nhỏ một bức ảnh thành các vùng con nhỏ hơn, gọi là các ô (cell), và thực hiện tính từng biểu đồ về hướng (histogram of gradients) cho từng điểm trong ô. Khi tổng hợp các biểu đồ này lại sẽ được biểu đồ biểu diễn hình ảnh ban đầu. (2 p. 13)



Hình 7: Ví dụ về đặc trưng HOG với hình ảnh gương mặt

Để tăng cường hiệu quả nhận dạng, thay vì làm việc trên từng ô thì chúng ta làm việc trên từng khối (block) chứa các ô. Các biểu đồ cục bộ được chuẩn hóa về độ tương phản bằng cách tính một ngưỡng cường độ của khối. Giá trị ngưỡng này được sử dụng để chuẩn hóa tất cả các ô trong khối. Việc này cho kết quả là các vector đặc trưng có tính bất biến cao hơn với các ảnh hưởng của điều kiện ánh sáng.

Bài toán tính toán **HOG** thông thường gồm năm bước chính. Mục đích là tìm một vector HOG cho ảnh đầu vào.

Bước 1: Chuẩn hóa hình ảnh trước khi xử lý.

Bước 2: Tính gradient theo x và y.

Nhận chập ảnh gốc với hai nhân một chiều Dx = [-1 0 1] và Dy = $[1\ 0\ -1]^T$. Việc này tương ứng với việc lấy đạo hàm của anh theo hai chiều Ox và Oy.

Với ảnh I đầu vào, ta tính được hai ảnh đạo hàm riêng theo hai hướng với công thức: Ix = I*Dx và Iy = I*Dy

Tính cường độ ảnh: G = sqrt($I_x^2 + I_y^2$) và hướng của ảnh: $\theta = \arctan(Iy, Ix)$

Dựa vào G và θ sẽ tính được một biểu đồ cường độ gradient, với các cột dựa trên θ và trọng số dựa trên G

Bước 3: Thống kê thành phần véc-tor cùng trọng số trong mỗi ô và vẽ một histogram cho mỗi ô.

Bước 4: Chuẩn hóa các khối.

Chia hình ảnh theo các khối, mỗi khối chưa các ô. Các khối này thường có kích thước là 2x2 hoặc 3x3 để dễ tính toán. Các khối này sẽ chồng lên nhau. Tiếp theo, tiến hành thu thập và ghép các biểu đồ của từng ô trong khối.

Gọi v là véc-tơ cần chuẩn hóa chứa tất cả các histogram của một khối, $||v_k||$ là giá trị chuẩn của nó theo các chuẩn k = 1, k = 2 và hằng số nhỏ e. Khi đó, các giá trị chuẩn hóa có thể được tính bằng một trong ba công thức:

$$\text{L2-norm: } f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}}$$

$$\text{L1-norm: } f = \frac{v}{\|v\|_1 + e}$$

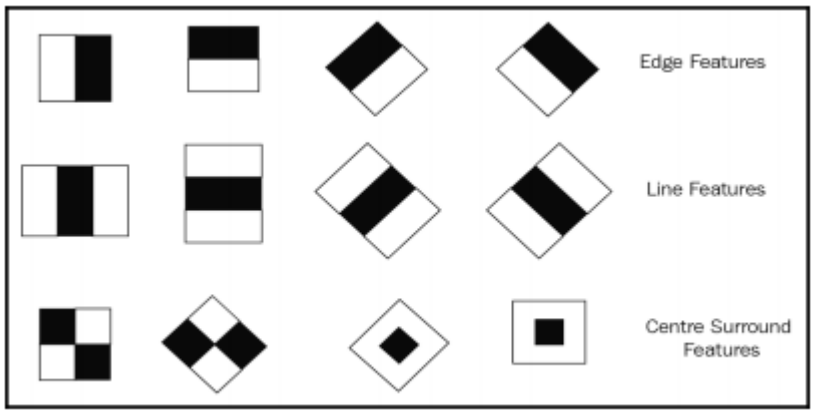
$$\text{L1-sqrt: } f = \sqrt{\frac{v}{\|v\|_1 + e}}$$

Bước 5: Thu thập tất cả các biểu đồ cường độ gradient định hướng của các khối để tạo ra vector tính năng cuối cùng.

2.5.4 **Đặc trưng Haar-like**

Đặc trưng Haar-like là tập hợp những đặc trưng đường dọc, đường ngang, đường chéo, … của một đối tượng hình ảnh, ở đây là gương mặt. Các đặc trưng này được lưu trữ trong một file định dạng XML. File này sẽ được sử dụng để kiểm tra gương mặt có xuất hiện trong hình ảnh hay là không.

Hiện nay có các file XML được sử dụng nhiều để phát hiện gưởng mặt như haarcascade_frontalface_default.xml, haarcascade_frontalface_alt.xml, và để phát hiện vùng mắt như haarcascade_eye.xml, haarcascade_lefteye_2splits.xml, haarcascade_righteye_2splits.xml. Các file này được cung cấp sẵn trong thư mục opencv/sources/data của thư viện mã nguồn mở OpenCV.



Hình 8: Danh sách các đặc trưng hình học của đặc trưng Haar-like

Hình trên là danh sách các đặc trưng cơ bản của đặc trưng Haar-like theo thứ tự từ trên xuống tương ứng là: Đặc trưng cạnh (Edge Features), Đặc trưng đường (Line Features) và Đặc trưng xung quanh tâm (Centre Surround Features).

2.5.5 Open CV

OpenCV (Open Source Computer Vision Library) là một thư viện nguồn mở của thị giác máy tính, xử lý ảnh và máy học, gồm các hàm chức năng tính toán được tích hợp sẵn, được phân phối dưới giấy phép BSD. **OpenCV** được thiết kế để tính toán hiệu quả hơn, thuận tiện hơn và tập trung chủ yếu vào các ứng dụng thời gian thực.

Thư viện **OpenCV** được sử dụng rộng rãi trong các ứng dụng như: kiểm tra và giám sát tự động, robot và xe hơi tự hành, phân tích hình ảnh y tế, tìm kiếm và phục hồi hình ảnh hay video, thực tế ảo, và nhiều ứng dụng khác.

OpenCV được bắt đầu phát triển tại Intel vào năm 1999 bởi Gary Bradsky và phát hành phiên bản đầu tiên năm 2000. **OpenCV** hỗ trợ đa dạng ngôn ngữ lập trình như C++, Python, Java, v.v. và đa nền tảng bao gồm Windows, Linux, OS X, Android và iOS. Cùng với đó là vấn đề hỗ trợ giao diện người dùng dựa trên CUDA và OpenCL được phát triển tích cực giúp cải thiện tốc độ xử lí của GPU. (3).

Với từng ngôn ngữ lập trình sẽ có các API khác nhau được sử dụng, OpenCV-Python là một Python API của OpenCV, dùng cho ngôn ngữ lập trình Python. Nó kết hợp được các tính năng tốt nhất của OpenCV C++ API và ngôn ngữ lập trình Python.

2.5.6 Máy học SVM

2.5.6.1 Giới thiệu SVM

SVM (Support Vector Machines) là một máy học vector hỗ trợ, được Vapnik nghiên cứu từ những năm 1965, đến những năm 1990 thì giải thuật chính thức được phát triển mạnh, trở thành công cụ hữu hiệu và phổ biến của lĩnh vực máy học, nhận dạng và khai phá dữ liệu. **SVM** đã được áp dụng thành công trong rất nhiều lĩnh vực như nhận dạng gương mặt người, phân loại văn bản, phân loại bệnh, … Bằng việc kết hợp với phương pháp hàm nhân, **SVM** cung cấp các mô hình hiệu quả và chính xác cho các vấn đề phân lớp, hồi quy tuyến tính và phi tuyến trong thực tế. Giải thuật **SVM** nhận đầu vào là một hàm nhân (kernel function) sẽ tạo ra một mô hình mới mà không cần đến bất kỳ sự thay đổi nào từ mã chương trình. Giải thuật học dẫn đến việc giải bài toán quy hoạch toàn phương, luôn có kết quả tối ưu toàn cục. **SVM** là một trong những giải thuật quan trọng của khai mỏ dữ liệu. (4)

Trong kỹ thuật **SVM**, không gian dữ liệu nhận vào ban đầu sẽ được ánh xạ và không gian đặc trưng, và trong không gian đặc trưng này thì mặt siêu phẳng phân chia dữ liệu tối ưu sẽ được xác định.

2.5.6.2. Giới thiệu về phân lớp dữ liệu

Phân lớp dữ liệu là một kỹ thuật quan trọng trong khai phá dữ liệu và được sử dụng rộng rãi nhất bên cạnh kỹ thuật hồi quy.

Mục đích: Để dự đoán và gán nhãn phân lớp cho các bộ dữ liệu mới hoặc mẫu mới. Với đầu vào là một tập các mẫu dữ liệu huấn luyện có riêng từng nhãn phân lớp và đầu ra là một bộ phân lớp dựa trên tập huấn luyện hoặc nhãn phân lớp. Phân lowpsd dữ liệu dựa trên tập huấn luyện và các gái trị trong một thuộc tính phân lớp và dùng nó để xác định lớp cho dữ liệu mới. (5 p. 8)

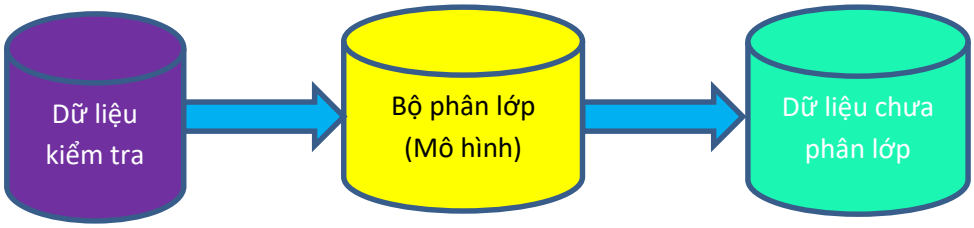
Kỹ thuật phần lớp dữ liệu gồm hai bước cơ bản:

Bước 1: Xây dựng mô hình từ tập huấn luyện

Bước 2: Kiểm tra tính đúng đắn của mô hình và sử dụng mô hình để phân lớp dữ liệu mới. (5)

Phân lớp cho những đối tượng mới hoặc chưa được phân lớp.

Đánh giá độ chính xác của mô hình: Lớp đã biết của một mẫu dữ liệu đem kiểm tra được so sánh với kết quả thu được từ mô hình. Tỉ lệ chính xác được tính bằng phần trăm các mẫu dữ liệu được phân lớp đúng bởi mô hình trong số các lần kiểm tra. (5 p. 9)



2.5.6.3. Vì sao sử dụng SVM trong phân lớp dữ liệu?

SVM rất hiệu quả trong việc giải quyết các bài toán với dữ liệu có số chiều quan sát lớn, như ảnh của dữ liệu gen, tế báo, ADN.

SVM giải quyết vấn đề overfitting rất tốt (dữ liệu có nhiễu và tách rời nhóm, hoặc số lượng dữ liệu huấn luyện quá ít).

Tốc độ phân lớp nhanh, hiệu suất tổng hợp tốt và hiệu năng tính toán cao. (5 p. 9)

2.5.6.4 Ứng dụng SVM vào đề tài

SVM được sử dụng để phân lớp cảm xúc, phân ra sáu loại cảm xúc riêng biệt với từng tiêu chí khác nhau, để tạo thành một tập huấn luyện. Sau đó, dự vào tập huấn luyện này để nhận dạng hình ảnh mới bất kỳ có cảm xúc như thế nào.

Máy học véc-tơ hỗ trợ được sử dụng với nhiều tập dữ liệu khác nhau để đánh giá sự chính xác một cách tương đối.

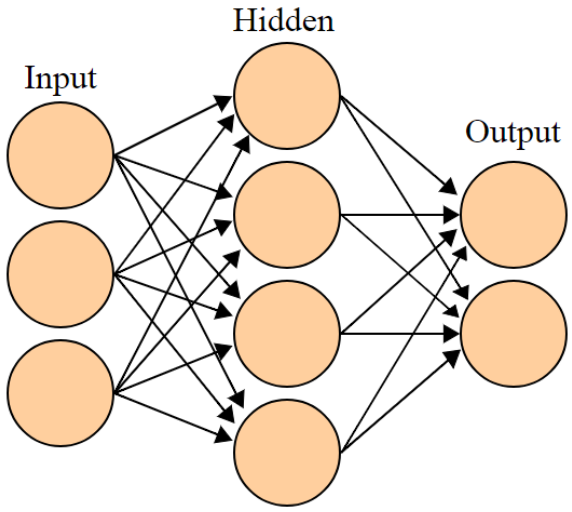
2.5.7 Mạng nơ-ron nhân tạo

2.5.7.1 Giới thiệu mạng nơ-ron nhân tạo

Mạng nơ-ron nhân tạo (**ANNs** – *Artificial Neural Networks*) là một họ phương pháp tính toán tổng quát mô hình hóa hoạt động của hệ thần kinh con người. Là một mạng phức tạp kết nối các đơn vị tính toán lại với nhau, trong đó mỗi đơn vị tính toán

là một nơ-ron nhân tạo, có thể có nhiều đầu vào, như chỉ có một đầu ra duy nhất cuối cùng.

Mạng nơ-ron nhân tạo là một giải thuật học có giám sát. Mạng nơ-ron nhân tạo là một mô hình tính toán được xây dựng mô phỏng theo mạng nơ-ron sinh học, bao gồm một nhóm các nơ-ron nhân tạo (các nút) nối với nhau, và thông tin được xử lý bằng cách truyền theo các kết nối (connection) và tính giá trị mới tại các nút (mỗi nút có vùng nhớ riêng của mình). Các nút này chỉ xử lý thông tin trên bộ dữ liệu của riêng nó và các thông tin đầu vào được truyền tới từ các kết nối. (4 p. 19)



Hình 9: Mô hình mạng nơ-ron nhân tạo 3 tầng: tầng đầu vào, tầng ẩn và tầng đầu ra

Trong đa số trường hợp, mạng nơ-ron nhân tọa là một hệ thống thích ứng có khả năng tự thay đổi cấu trúc dựa trên các thông tin bên ngoài hay bên trong mạng trong quá trình học. Mạng nơ-ron nhân tạo là một trong những kỹ thuật xử lý dữ liệu hiện đại, cho phép lấy được lượng thông tin tối đa từ dữ liệu như nhận dạng, phân loại, dự báo, xây dựng mô hình, nghiên cứu về suy nghĩ của con người và cách để tạo ra trí thông minh nhân tạo. (4 p. 20)

Do giải thuật đơn giản, mạng nơ-ron nhân tạo được cài đặt khá dễ dàng trên hệ thống nhúng. Tuy nhiên, quá trình huấn luyện mạng nơ-ron tốn nhiều thời gian, do phải huấn luyện nhiều lần vì kết quả thu được chỉ là tối ưu cục bộ. Một khó khăn khác nữa là mạng nơ-ron chỉ làm việc với dữ liệu số, vì thế cần phải có một bước tiền xử lý nếu dữ liệu không phải là số. Ngoài ra, kết quả của mạng nơ-ron không dễ hiểu chút nào, rất khó để giải thích kết quả của mạng nơ-ron với dữ liệu đầu vào được cho. Dù đã có nhiều nghiên cứu về kiến trúc mạng nơ-ron, nhưng vấn đề thiết kế mạng nơ-ron để phù hợp với từng ứng dụng là một chủ đề được đang quan tâm và nghiên cứu. (4 p. 47)

Đa số các mạng nơ-ron đều có quy tắc học riêng của mình mà thông qua đó thì trọng số của các liên kết được điều chỉnh dựa trên dữ liệu. Hay mạng nơ-ron học trên các dữ liệu sẽ có khả năng tổng quát hóa tri thức và có khả năng đưa ra nhận thức của mình cho những trường hợp xảy ra trong tương lai.

Một hạn chế của ANN là các nơ-ron của nó thường ở trạng thái nghỉ trong suốt quá trình đào tạo, nó chỉ làm việc khi được kích thích.

2.5.7.2 Lịch sử ra đời và phát triển của mạng nơ-ron nhân tạo

Trong những năm cuối thế kỉ XIX và đầu thế kỉ XX, sự manh nha bắt đầu từ việc nghiên cứu của các nhà khoa học Hermann von Hemholtz, Ernst Mach, Ivan Pavlov trong các ngành Vật lý học, Tâm lý học và Thần kinh học. Các công trình nghiên cứu này chủ yếu đi sâu vào các lý thuyết tổng quát về học, nhìn và lập luận mà chưa đưa ra các mô hình toán học và thuật toán cụ thể để mô tả hoạt động của các nơ-ron.

Vào những năm 1940, Warren McCulloch và Walter Pitts đã chỉ ra rằng về nguyên tắc, mạng của các nơ-ron nhân tạo có thể tính toán bất kì một hàm số học hay logic nào trong công trình nghiên cứu của mình. Donald Hebb đã phát biểu rằng thuyết lập luận cổ điển của Pavlov là hiện thực do các thuộc tính của từng nơ-ron riêng biệt.

Cuối những năm 1950, Frank Rosenblatt đã phát minh ra mạng nhận thức và luật học tương ứng để áp dụng vào ứng dụng đầu tiên của các nơ-ron nhân tạo. Mạng này có khả năng nhận dạng các mẫu, thành quả này đã mở ra một niềm hy vọng mới cho công trình nghiên cứu mạng nơ-ron. Tuy nhiên có những hạn chế ban đầu là chỉ giải quyết được một số hữu hạn các bài toán. Cũng vào khoảng thời gian này, Bernard Widrow và Marcian Hoff đã cho ra đời một thuật toán học mới và sử dụng nó để huấn luyện cho các mạng nơ-ron tuyến tính thích nghi. Luật học này có tên Widrow – Hoff và vẫn được ứng dụng ở thời điểm hiện tại. Tuy nhiên, luật học này cũng mắc phải một vấn đề là các mạng nhận thức chỉ có khả năng giải quyết được các bài toán tuyến tính.

Những kết quả nghiên cứu của Minsky Papert về sự trở ngại trên làm cho công cuộc nghiên cứu về mạng nơ-ron bị chững lại một thập kỉ vào những năm 1970, nguyên nhân là không có được phần cứng và máy tính đủ mạng để làm các thao tác thực nghiệm. Tuy nhiên vẫn có những phát kiến quan trọng ra đời về phát triển một loại mạng mới có thể độc lập hoạt động như một bộ nhớ của Teuvo Kohonen và James Anderson, khảo sát các mạng tự tổ chức của Stephen Grossberg.

Bước vào những năm 1980, cũng với sự ra đời của PC là sự phát triển mạnh mẽ của các công trình nghiên cứu mạng nơ-ron, các khái niệm mới được ra đời như mạng hồi qui và mạng lan truyền ngược.

*2.5.7.**3. Ứng dụng của mạng nơ-ron***

Trong quá trình nghiên cứu, phát triển và triển khai, mạng nơ-ron nhân tạo được ứng dụng nhiều để thực hiện các công việc nhận dạng, phân lớp, điều khiển và dự báo trong đa lĩnh vực:

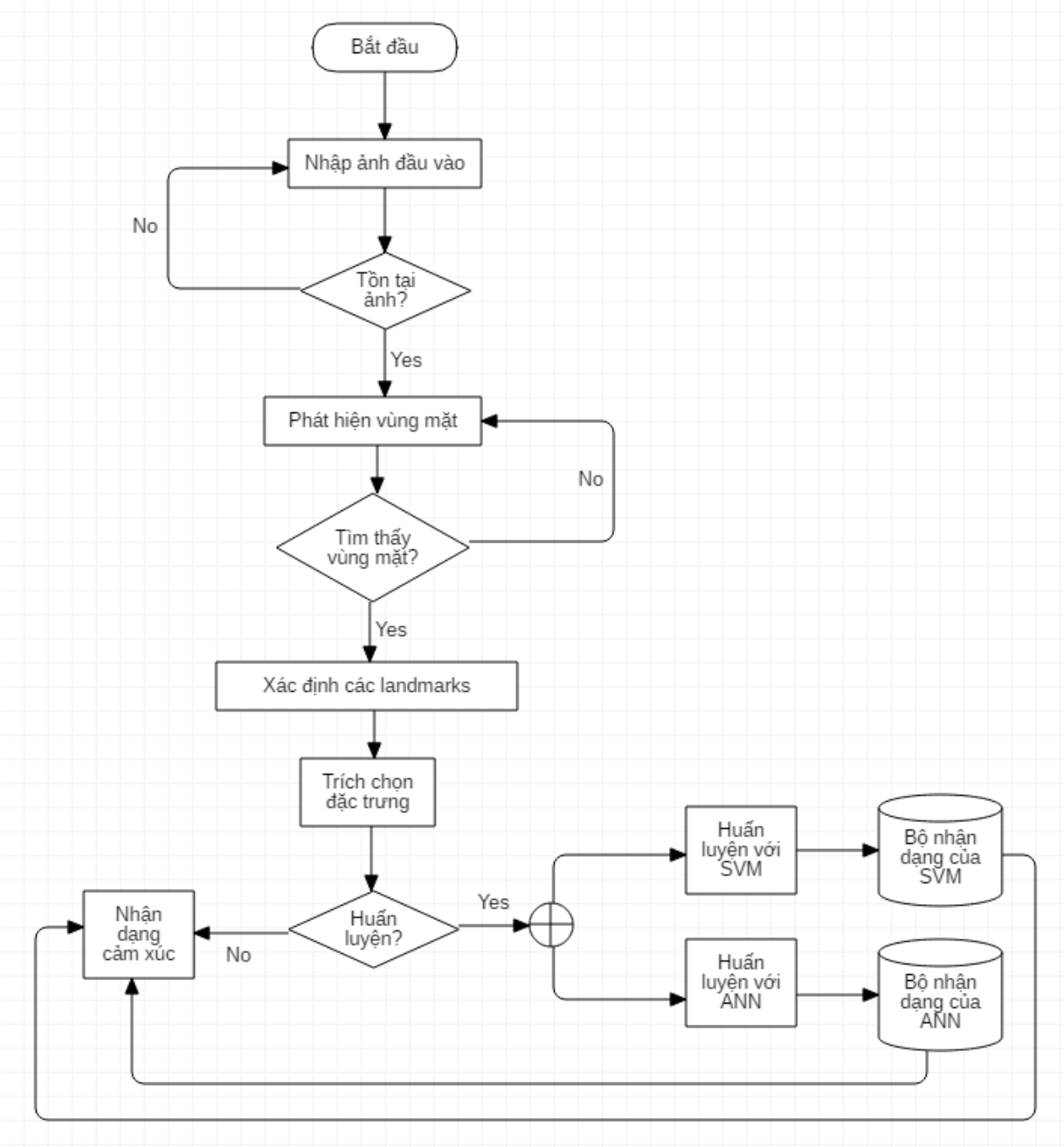
- Trong lĩnh vực tài chính, có các ứng dụng như định giá bất động sản, cho vay, kiểm tra tài sản cầm cố, đánh giá mức độ hợp tác, phân tích đường tín dụng, chương trình thương mại qua giấy tờ, phân tích tài chính liên doanh, dự báo tỉ giá tiền tệ.
- Về ngân hàng, ứng dụng tính tiền của thẻ tín dụng, bộ độc séc và các tài liệu.
- Trong lĩnh vực giải trí, được ứng dụng vào các bộ phim hoạt hình, các hiệu ứng điện ảnh, các ứng dụng di động tương tác cảm xúc người dùng, nhận dạng gương mặt.
- Mạng nơ-ron nhân tạo có thể đánh giá việc áp dụng chính sách, tối ưu hóa sản phẩm trong ngành bảo hiểm.
- Điện tử học: dự báo mã tuần tự, sơ đồ chíp IC, điều khiển tiến trình, phân tích nguyên nhân hư hỏng chíp, nhận dạng tiếng nói.
- Trong lĩnh vực quốc phòng – an ninh: định vị, phát hiện vũ khí, dò tìm mục tiêu, phát hiện đối tượng, nhận dạng nét mặt, điều tra tội phạm, các bộ cảm biến thế hệ mới, xử lý ảnh radar.
- Ở các vấn đề tự động hóa: tiến trình ô tô tự động, các bộ phân tích hoạt động của xe hơi.
- Về hàng không: Phi công tự động, đường bay giả lập, các hệ thống điều khiển lái máy bay, bộ phát hiện lỗi.

Ngày nay, các ứng dụng của mạng nơ-ron nhân tạo đã ngày càng phát triển và rộng khắp trong hầu hết các lĩnh vực của đời sống xã hội, góp phần tích cực và mạnh mẽ vào cuộc cách mạng công nghiệp 4.0 và trí tuệ nhân tạo.

2.5.7.4 Mạng nơ-ron tích chập

CHƯƠNG 3: NỘI DUNG NGHIÊN CỨU

3.1 Sơ đồ tổng quan các thành phần chính của hệ thống nhận dạng biểu cảm gương mặt
Sơ đồ dòng xử lý hệ thống được trình bày như sau:



3.2 Các nghiên cứu liên quan

Việc ứng dụng xử lý ảnh để nhận diện cảm xúc đã có một số luận văn và đề tài nghiên cứu trước đây như:

Bài báo khoa học phân tích các đặc điểm gương mặt của hệ thống phát hiện cảm xúc, do hai tác giả N Dharmesh và Mausmi Kulshreshtha đến từ khoa Khoa học máy tính Viện Công nghệ Veermata Jijabai, Mumbai, Ấn Độ. Nghiên cứu này trình bày hệ thống để phát hiện trạng thái cảm xúc con người sử dụng kỹ thuật AUs. Hệ thống tự động phát hiện gương mặt từ ảnh chụp và nhận dạng bảy loại cảm xúc cơ bản của con người như: *vui, buồn, ngạc nhiên, sợ hãi, ghê tởm, giận dữ và bình thường*. Bài báo tập

trung vào việc phát hiện cảm xúc tự động của các điểm tính năng và nhận dạng cảm xúc từ ảnh gương mặt kỹ thuật số hai chiều. Hệ thống này sử dụng cơ sở dữ liệu JAFFE để kiểm thử (6).

Tiếp theo là bài nghiên cứu của Monika Dubey và Giáo sư Lokesh Singh, thuộc khoa Kỹ thuật và Khoa học Máy tính, Viện công nghệ thông tin Technocrats, Bhopal, Ấn Độ. Mục đích của bài báo là giới thiệu sự cần thiết và các ứng dụng của nhận diện cảm xúc gương mặt. Giữa hai hình thức giao thiếp bằng lời nói và không bằng lời nói thì giao tiếp không bằng lời nói giữ một vai trò cực kỳ quan trọng. Nó thể hiện trạng thái của người dùng và lấp đầy mạch cảm xúc của tình huống giao tiếp. Nội dung bài nghiên cứu bao gồm giới thiệu hệ thống nhận dạng cảm xúc mặt người, ứng dụng, so sánh các kỹ thuật nhận dạng cảm xúc phổ biến (7).

Một bài báo của Việt Nam, đến từ trường đại học Khoa học và đại học Công nghệ thông tin quốc gia mang tên "An Efficient Real-Time Emotion Detection Using Camera and Facial Landmarks", của các tác giả Binh T. Nguyen, Minh H. Trinh, Tan V. Phan và Hien D. Nguyen. Bài báo trình bày một tiếp cận tiềm năng về phát hiện cảm xúc con người thời gian thực. Với mỗi cảm xúc được phát hiện từ camera, hình ảnh sẽ được trích xuất các landmarks của gương mặt tương ứng, kiểm tra nhiều đặc điểm và mô hình khác nhau để dự đoán cảm xúc của con người (8). Hạn chế của bài nghiên cứu này là chỉ áp dụng với ba loại cảm xúc chính là tích cực, bình thường và tiêu cực.

Một thành tựu nghiên cứu khác đến từ trường đại học Standford, tác giả là James Pao, với đề tài mang tên "Emotion Detection Through Facial Feature Recognition". Tác giả nhận định con người thể hiện các cảm xúc của họ thường thông qua biểu cảm gương mặt. Một thuật toán sẽ giúp phát hiện, nhân dạng, đánh giá các loại cảm xúc này và cho phép tự động nhận dạng cảm xúc của con người trong hình ảnh và video. Đề tài trích xuất các tính năng và nhận dạng cảm xúc gương mặt sử dụng bộ phát hiện Viola-Jones và Harris để lấy gương mặt và cảm xúc trong ảnh. Sử dụng PCA, HOG, SVM để huấn luyện và phân lớp thành bảy lớp cảm xúc cơ bản. Cách tiếp cận này cho phép phân lớp nhanh từ các phép chiếu kiểm tra hình ảnh được tính toán bằng vector riêng. Bước đầu cho kết quả tốt 5/7 cảm xúc cơ bản và dễ phân biệt, độ chính xác khi kiểm thử là 81% (9).

Nhìn chung, các đề tài nghiên cứu trên đều sử dụng ngưỡng để phân lớp và nhận dạng cảm xúc dựa vào đặc trưng của toàn bộ gương mặt hoặc ngưỡng của từng thành phần gương mặt như mắt, mũi, miệng, … và với cách tiếp cận này thì thường cho kết quả có độ chính xác chưa cao do một số nguyên nhân khác nhau. Có những gương mặt người có hình dáng và kích thước khác nhau, cảm xúc khó nắm bắt với những ảnh

gương mặt nghiêng, không cố định. Độ chính xác của hệ thống còn phụ thuộc vào quá trình phát hiện đầy đủ và chính xác vùng mặt. Do đó, cần có một phương pháp tổng quát sử dụng các kỹ thuật máy học để nhận dạng cảm xúc con người.

3.3 Định hướng giải quyết của luận văn

Trong thực tế, để dự đoán được cảm xúc hiện tại của một người, đầu tiên chúng ta cần nhìn vào gượng mặt của người đó. Cụ thể hơn, chúng ta quan sát các thành phần chính trên gương mặt như miệng, hai mắt, má, chân mày, …. Trong quá trình này, bộ não đang thu thập dữ liệu để nhận dạng cảm xúc. Và như một cách tự nhiên hay đã được học ở một nơi nào đó, chúng ta biết được với các biểu hiện như thế nào của các thành phần trên gương mặt sẽ thể hiện cảm xúc gì tương ứng.

Quá trình nhận dạng cảm xúc gương mặt trên máy tính cũng được thực hiện tương tự như những gì diễn ra trong bộ não con người, gồm ba bước chính:

- **Bước 1**: Thu nhận dữ liệu hình ảnh. Tìm gương mặt trong ảnh. Xác định các landmarks trên gương mặt. Trích chọn các đặc trưng của các thành phần của gương mặt.
- **Bước 2**: So khớp các đặc trưng vừa trích rút với bộ huấn luyện nhận dạng cảm xúc đã được huấn luyện từ trước.
- **Bước 3**: Nhận dạng cảm xúc gương mặt: Kiểm tra sự chính xác của bộ nhận dạng vừa huấn luyện bằng việc tiến hành nhận dạng cảm xúc của những dữ liệu hình ảnh mới. Hiển thị kết quả nhận dạng.

Để thực hiện công việc trên, vấn đề đầu tiên như đã đề cập ở *bước 2* là phải xây dựng được một tập huấn luyện từ trước, có khả năng nhận dạng được các loại cảm xúc cơ bản của con người dựa vào gương mặt, với các bước cơ bản sau:

- **Bước 1**: Thu thập dữ liệu. Dữ liệu là hình ảnh của các gương mặt người với những biểu cảm cần nhận dạng. Số lượng hình ảnh phải đủ lớn để có thể đảm bảo độ chính xác dữ bộ huấn luyện.
- **Bước 2**: Huấn luyện bộ nhận dạng. Thực hiện công việc dạy cho máy tính biết với những biểu hiện như thế nào của các thành phần chính trên gương mặt thì sẽ tương ứng với những cảm xúc gì.
- **Bước 3**: Nhận dạng cảm xúc gương mặt với bộ dữ liệu kiểm thử để kiểm tra sự chính xác của bộ nhận dạng vừa huấn luyện bằng việc tiến hành nhận dạng cảm xúc của những dữ liệu mới. Phải đảm bảo độ chính xác hệ thống nằm trong vùng chấp nhận được.

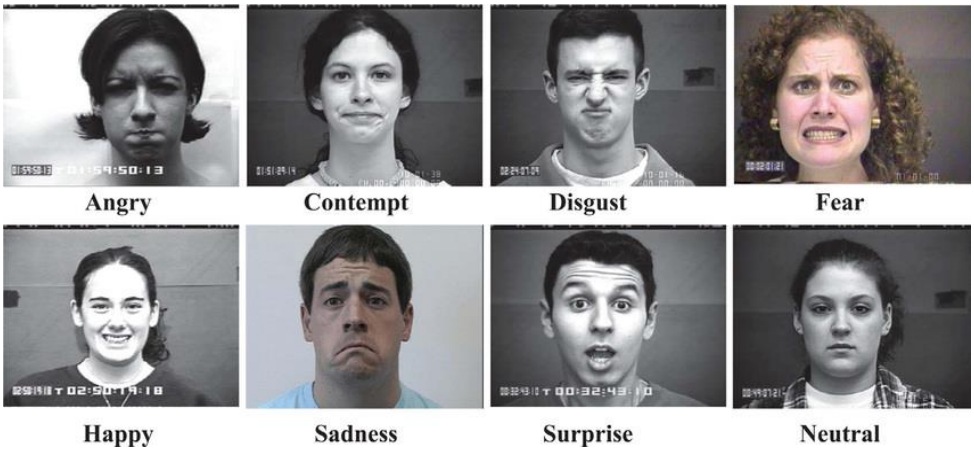
3.4 Quy trình thực hiện luận văn

3.4.1 Chuẩn bị dữ liệu huấn luyện

Dữ liệu được dùng để huấn luyện là tập dữ liệu Kohn-Kanade, JAFFE (Japanese Femal Facial Expression), và một số hình ảnh tự thu thập.

Cơ sở dữ liệu thứ nhất: **Cohn-Kanade** (CK và CK+): Là cơ sở dữ liệu biểu cảm gương mặt dựa trên các đơn vị vận động gương mặt, được dùng cho nghiên cứu trong tự động phân tích và tổng hợp hình ảnh gương mặt, và cho các lĩnh vực học về cảm xúc. Cohn-Kanade hiện tại có hai phiên bản:

- *Phiên bản 1:* **CK**. Là phiên bản đầu tiên, bao gồm 486 chuỗi hình ảnh cảm xúc gương mặt của 97 đối tượng. Mỗi chuỗi hình ảnh cảm xúc bắt đầu với một cảm xúc trung tính và tăng dần để đạt đến mức độ cảm xúc mạnh mẽ nhất. Từng hình ảnh cảm xúc trong mỗi chuỗi được mã hóa và gán nhãn. (10)
- *Phiên bản 2:* **CK+**. Chứa 593 chuỗi hình ảnh (327 chuỗi có các nhãn cảm xúc riêng biệt), mỗi hình ảnh thể hiện một trong tám loại cảm xúc: vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm, trung tính và khinh thường. Các hình ảnh hầu hết là ảnh mức xám với độ phân giải quy định là 640x490 pixels.



Hình 10: Ví dụ về các hình ảnh gương mặt thể hiện cảm xúc trong cơ sở dữ liệu Cohn-Kanade

Cơ sở dữ liệu thứ hai: **JAFFE**: Kho dữ liệu chứa 213 hình ảnh của 7 biểu cảm gương mặt (vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm và trung tính) được thể hiện bởi 10 người mẫu nữ đến từ Nhật Bản. Mỗi hình ảnh được đánh giá trên 6 tính từ cảm xúc bởi 60 người Nhật. (11)

Hình 11: Các hình ảnh gương mặt biểu hiện bảy cảm xúc của một người mẫu đại diện trong tập dữ liệu JAFFE

Cơ sở dữ liệu thứ ba: **Fer2013**. Là một tập dữ liệu ảnh nguồn mở, được bắt đầu tạo ra trong một dự án của Pierre-Luc Carrier và Aaron Courville, sau đó được chia sẻ rộng rãi trên Kaggle. Tập dữ liệu này chứa 35.887 ảnh mức xám có kích thước 48x48 của các gương mặt thể hiện bảy cảm xúc cơ bản. Theo quy ước, các cảm xúc được gán nhãn trong tập dữ liệu như sau:

Bảng 2: Bảy cảm xúc cơ bản được gán nhãn tương ứng trong tập dữ liệu Fer2013

Cảm xúc		**Nhãn**	**Số lượng ảnh**
Cảm xúc giận dữ	*Angry*	0	*4593*
Cảm xúc ghê tởm	*Disgust*	1	*547*
Cảm xúc sợ hãi	*Fear*	2	*5121*
Cảm xúc vui vẻ	*Happy*	3	*8989*
Cảm xúc buồn bã	*Sad*	4	*6077*
Cảm xúc ngạc nhiên	*Surprise*	5	*4002*
Không cảm xúc	*Neutral*	6	*6198*

Cơ sở dữ liệu thứ tư: Những hình ảnh tự sưu tầm của cá nhân.

Hình 12: Những hình ảnh trong tập dữ liệu cá nhân đã được chuẩn hóa

Vấn đề được đặt ra rất đơn giản là càng nhiều hình ảnh được sử dụng để huấn luyện thì hiệu quả tập huấn luyện càng cao. Tuy nhiên cần quan tâm đến tốc độ thực thi của chương trình và yêu cầu của hệ thống.

Thực hiện tổ chức thư mực như sau: có một thư mục cha tên là *"training-data"*. Thư mục này chứa sáu thư mục con tương ứng là sáu cảm xúc của con người: *vui vẻ, buồn bã, ngạc nhiên, giận dữ, sợ hãi, kinh tởm*. Mỗi thư mục con sẽ chứa hình ảnh gương mặt của con người với biểu lộ cảm xúc tương ứng với tên thư mục con. Các hình ảnh được đánh số từ 1 cho đến số lượng hình ảnh có trong thư mục. Được mô tả cụ thể như hình bên dưới:

```
training-data
|----------------happiness
|                |------ 1.jpg
|                |------ 2.jpg
|                |------ …
|                |------ n.jpg
|----------------sadness
|                |------ 1.jpg
|                |------ 2.jpg
|                |------ …
|                |------ n.jpg
|----------------surprise
|                |------ 1.jpg
|                |------ 2.jpg
|                |------ …
|----------------anger
|                |------ 1.jpg
|                |------ 2.jpg
|                |------ …
|                |------ n.jpg
|----------------fear
|                |------ 1.jpg
|                |------ 2.jpg
|                |------ …
|                |------ n.jpg
|----------------disgust
|                |------ 1.jpg
|                |------ 2.jpg
|                |------ …
|                |------ n.jpg
```

|------ n.jpg	|----------neutral |------ 1.jpg |------ 2.jpg |------ … |------ n.jpg |----------

Tiếp theo, chúng ta cần có một thư mục *"test-data"* chứa các hình ảnh sẽ được sử dụng để kiểm tra độ chính xác của chương trình nhận dạng cảm xúc sau khi đã huấn luyện thành công. Những hình ảnh trong bộ dữ liệu kiểm thử có thể là các hình ảnh được sử dụng để huấn luyện bộ nhận dạng ban đầu, cũng có thể là các hình ảnh mới với các cảm xúc được thể hiện bởi các nhân vật mới.

Bên cạnh đó, cần chuẩn bị các mô-đun, thư viện cần thiết phục vụ cho quá trình lập trình, huấn luyện và nhận dạng. Trong quá trình thực hiện, cần phải cài đặt và sử dụng nhiều module và thư viện khác nhau tương ứng với từng chức năng khác năng để đảm bảo thực hiện được công việc chung của bài toán. Dưới đây sẽ giới thiệu một số đại diện cơ bản và thông dụng nhất được sử dụng:

- **numpy**: là một mô-đun tuyệt vời của Python, hỗ trợ cho việc tính toán trên các ma trận một cách dễ dàng. Nó có các hàm mạnh mẽ được tích hợp để xử lý các mảng nhiều chiều, chuyển đổi các danh sách trong Python thành các mảng numpy phục vụ cho việc tính toán và nhận dạng. Sử dụng dòng lệnh dưới đây: "*import numpy as np*" để tương tác với thư viện numpy.
- **pandas**: là một thư viện phân tích dữ liệu của Python, là một thư viện mã nguồn mở và được phát hành dưới giấy phép BSD, nó cung cấp các cấu trúc dữ liệu và công cụ phân tích dữ liệu dễ sử dụng, hiệu quả cao cho ngôn ngữ lập trình Python.
- **cv2**: mô-đun OpenCV sử dụng cho ngôn ngữ lập trình Python, được dùng để phát hiện và nhận dạng gương mặt. Để sử dụng mô-đun cv2 trong Python, ta thực hiện chèn cv2 vào mã chương trình như một thư viện bằng dòng lệnh: "*import cv2*".
- **os**: là một mô-đun của Python, cung cấp một cách linh hoạt để sử dụng các chức năng phục thuộc vào hệ điều hành. Ví dụ như mô-đun *os.path* được sử dụng để thao tác với các đường dẫn hệ thống, như là đường dẫn trỏ tới file hình ảnh đầu vào. Trong trường hợp này, mô-đun này được sử dụng để đọc thư mục huấn luyện và tên file. Để sử dụng, chèn dòng lệnh "*import os*" vào mã chương trình.

3.4.2 Phát hiện vùng mặt với OpenCV

Phát hiện gương mặt là bước đầu tiên của bài toán nhận dạng gương mặt cũng như nhận dạng cảm xúc. Hiện nay có rất nhiều phương pháp dùng để phát hiện gương mặt người trong bức ảnh, dựa vào tính chất thì các phương pháp được chia ra thành hai hướng tiếp cận chính như sau:

Hướng tiếp cận dựa trên các đặc trưng cơ bản:

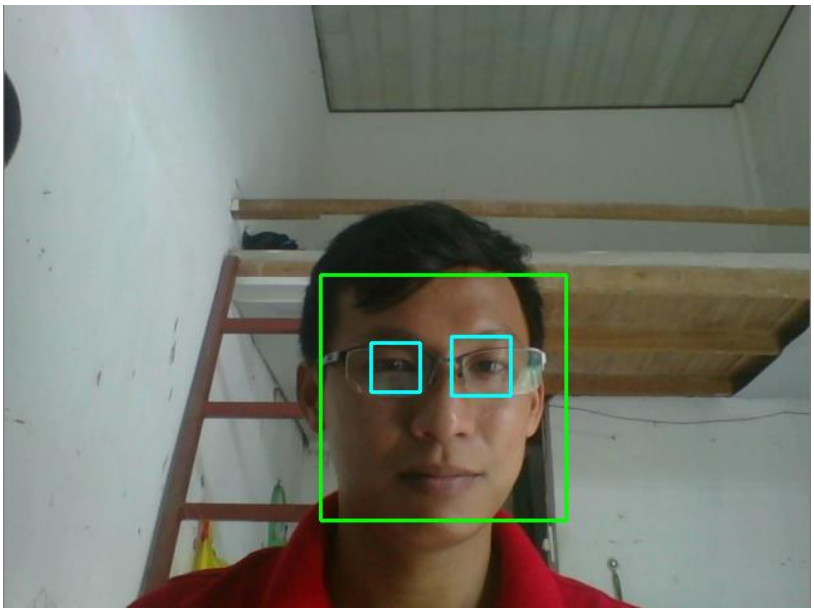
- Đây là phương pháp chủ yếu dựa trên những hiểu biết của con người về gương mặt. Đó là dựa trên những bộ phận chính cấu tạo nên khuôn mặt như mắt, mũi, miệng, chân mày, gò má, cằm và hình dạng cũng như kết cấu của gương mặt.
- Trong cách tiếp cận này có hai hướng tiếp cận nhỏ hơn:
  - Từ dưới lên (*Bottom-up*): Xác định từng đặc trưng riêng biệt, nhóm chúng lại với nhau để tạo nên đặc trưng chung. Ưu điểm của cách tiếp cận này là không bị ảnh hưởng bởi hướng hay di chuyển của gương mặt. Nhược điểm ở chỗ các đặc trưng dễ bị ảnh hưởng bởi các yếu tố ngoại cảnh như ánh sáng, nhiễu.
  - Từ trên xuống (*Top-down*): Đầu tiên tạo ra một mẫu chuẩn của khuôn mặt (2D hoặc 3D), sau đó áp mẫu này vào ảnh chứa gương mặt bằng việc tìm kiếm trên toàn bộ ảnh. Với cách tiếp cận này, có các mô hình điển hình như ASM/ AAM.

Hướng tiếp cận dựa trên diện mạo:

- Ý tưởng chính là phân một bước ảnh vào hai lớp *là mặt* hoặc *không là mặt*, nếu bức ảnh có chứa gương mặt người thì được phân vào lớp *là mặt* và ngược lại.
- Để làm được điều đó, phương pháp này phải học từ một tập ảnh huấn luyện mẫu để xác định như thế nào là gương mặt người. Phương pháp gồm các bước chính:
  - Sử dụng một số phương pháp biểu diễn khuôn mặt mặt LBP, Gabor để tạo ra bộ phân lớp *là mặt* và *không là mặt*.
  - Dùng một cửa sổ có kích thước cố định quét trên toàn bộ bức ảnh đầu vào ở các vị trí và tỉ lệ khác nhau, hoặc trên toàn bộ bức ảnh.
  - Tìm ra và xử lý các trường hợp trùng lặp.

Trong hai hướng tiếp cận trên, hướng tiếp cận dựa trên diện mạo có tính ưu việt hơn vì không phụ thuộc nhiều vào hướng của đầu.

Vùng mặt được phát hiện như hình bên dưới:



Hình 13: Vùng mặt được phát hiện trong hình vuông có viền màu xanh.

3.4.3 Xác định các landmark gương mặt với thư viện Dlib

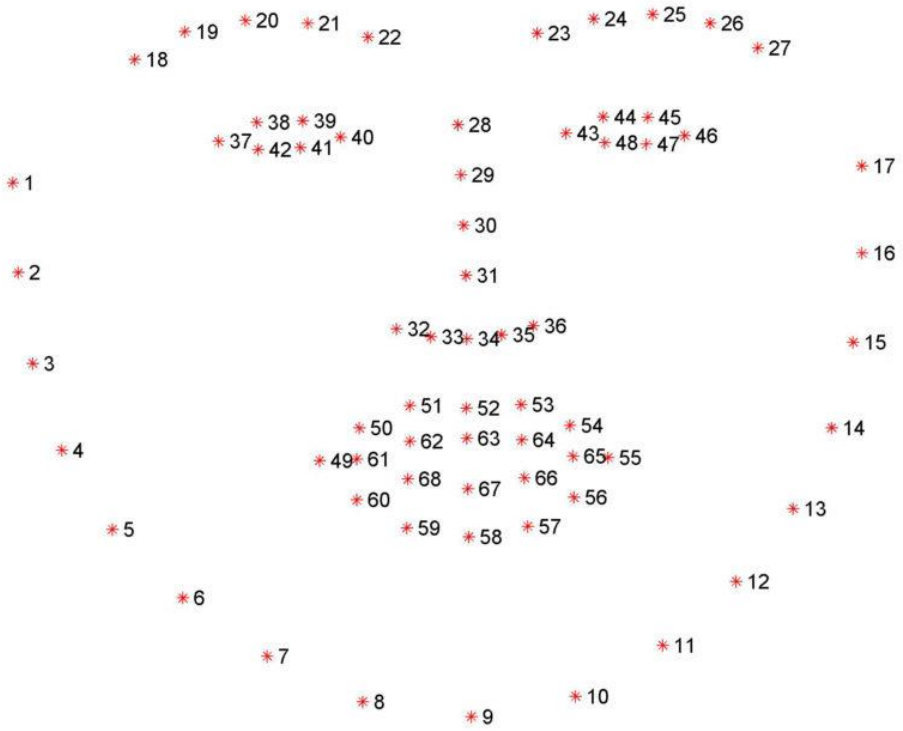
Gương mặt người có nhiều hình dáng khác nhau như mặt tròn, mặt vuông, mặt trái xoan, mặt chữ điền, nhưng hầu hết một gương mặt người đều có các thành phần cơ bản như hai mắt, hai chân mày, mũi, miệng, cằm. Các landmark dùng để xác định vị trí và thể hiện các vùng nổi bật của gương mặt cũng là quá trình xác định các thành phần trên trong một bức ảnh.

Các landmark gương mặt đã được ứng dụng thành công trong căn chỉnh gương mặt, định vị vị trí đầu, phát hiện nháy mắt. Xác định các landmark gương mặt là một bài toán con của bài toán dự đoán hình dạng khuôn mặt, có nghĩa là cần xác định được những thành phần quan trọng nào trong bức ảnh để tạo nên hình dạng gương mặt người.

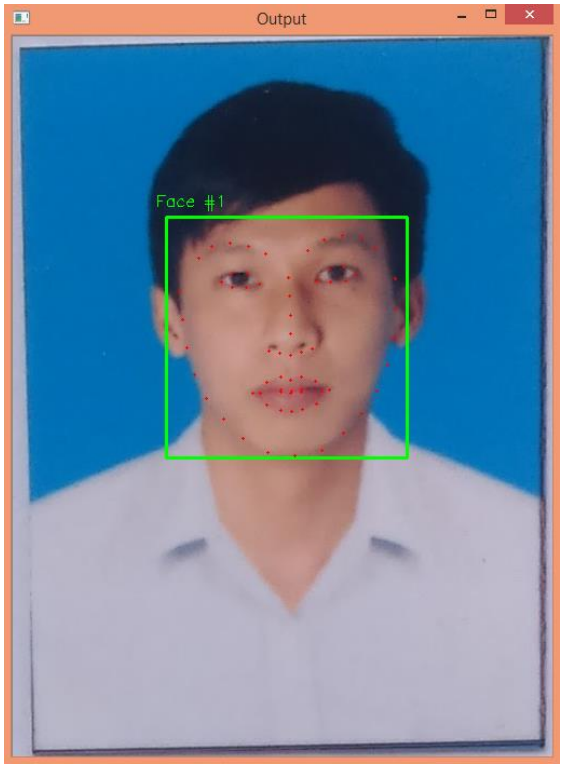
Việc xác định các landmark gồm có hai bước:

- Xác định được vị trí gương mặt trong ảnh.
- Xác định được các thành phần tạo nên cấu trúc gương mặt.

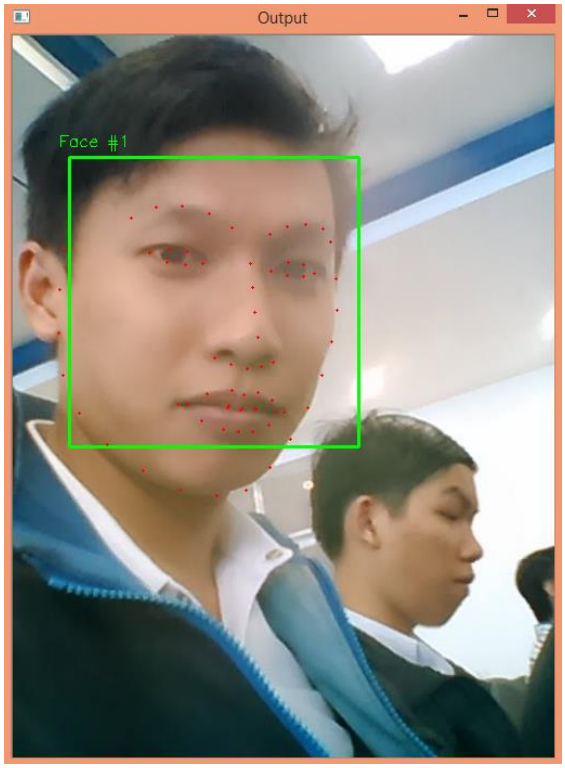
Bộ xác định các landmark gương mặt của thư viện Dlib sẽ xác định 68 điểm chính theo tọa độ *(x, y)* cấu thành gương mặt người:



Hình 14: Toàn bộ 68 landmarks gương mặt của thư viện Dlib

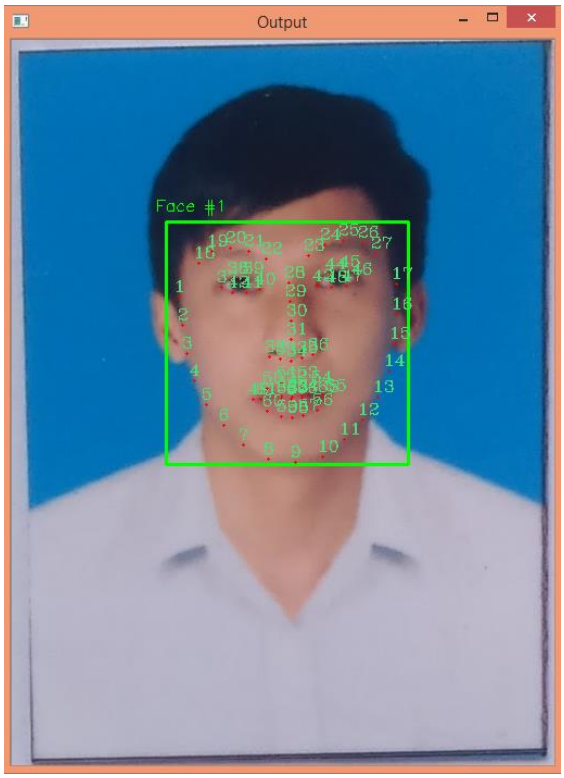


Hình 15: Các landmarks được phát hiện của gương mặt trực diện

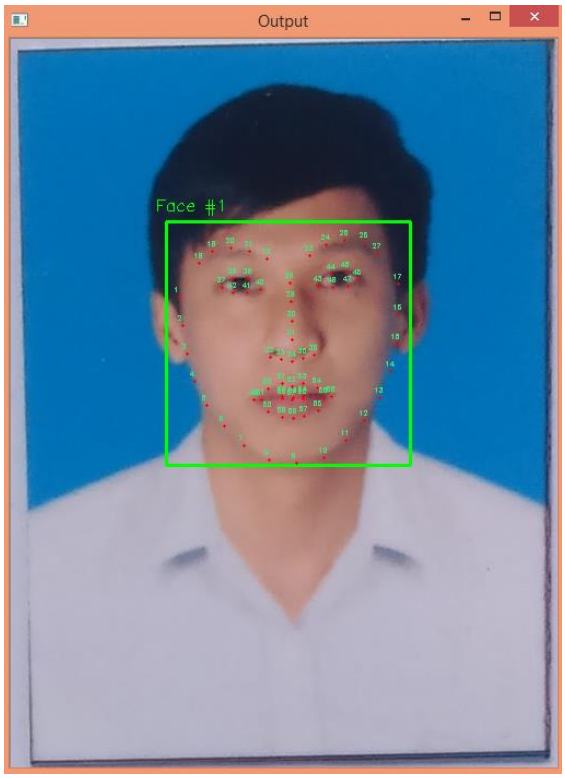


Hình 16: Các landmarks được phát hiện của gương mặt nghiêng

Dlib là thư viện phần mềm mã nguồn mở được viết bằng ngôn ngữ lập trình C++, chạy được trên nhiều nền tảng và được tạo ra bởi Davis King. Dlib được sử dụng nhiều trong lĩnh vực máy học và thị giác máy tính. Dlib



Hình 17: Mô hình Dlib áp lên gương mặt với các landmarks được phát hiện và đánh số lớn



Hình 18: Mô hình Dlib áp lên gương mặt với các landmarks được phát hiện và đánh số nhỏ

3.4.4 Rút trích đặc trưng thành phần gương mặt

Bài toán rút trích đặc trưng trên ảnh gương mặt người là bài toán cơ bản và quan trọng trong xử lý ảnh và thị giác máy tính. Đầu ra của công việc này là đầu vào cho bài toán nhận dạng gương mặt, nhận dạng cảm xúc.

Sử dụng đặc trưng HOG để rút trích đặc trưng của từng thành phần của gương mặt. Thư viện Dlib được dùng với đặc trưng HOG và bộ phân lớp LibSVM để trích rút ra các đặc trưng của gương mặt. Dlib chỉ hoạt động chính xác với các gương mặt trực diện cũng là một hạn chế nhất định của phương pháp này.

3.4.5 Huấn luyện với SVM

Trường hợp 1: Sử dụng tập dữ liệu Cohn-Kanade

Bước 1: Tổ chức dữ liệu. Đầu tiên, cấu trúc thư mục của tập dữ liệu Cohn-Kanade được giới thiệu. Tập dữ liệu Cohn-Kanade được tổ chức thành các thư mục (được đặt

tên như *S010*, *S011*, …), mỗi thư mục là tập hợp các ảnh của một người mẫu. Trong mỗi thư mục này có chứa các thư mục con (như *001*, *002*, …). Mỗi thư mục con sẽ chứa các hình ảnh thể hiện cảm xúc ở trạng thái bình thường (không cảm xúc) và tăng dần lên theo từng ảnh tới một cảm xúc nhất định được thể hiện rõ rang nhất. Thông tin chi tiết hơn được thể hiện như hình bên dưới:

Cohn-Kanade C+
|-----**S010**
| |------*001*
| |------ S010_001_01594215.png
| |------ S010_001_01594216.png
| |------……..
| |------ S010_001_01594226.png
| |------*002*
| |------ …
| |------*007*
|-----**S011**
| |------*001*
| |------*002*
| |------ …
| |------*007*
|-----…….
| |------ *001*
| |------ *002*
| |------ …
| |------ *007*
|-----**S100**
| |------*001*
| |------ S100_001_00533804.png
| |------ S100_001_00533805.png
| |------……..
| |------ S100_001_00533817.png
| |------*002*
| |------ …
| |------*007*
|-----**S101**
| |------*001*
| |------*002*
| |------ …
| |------*007*
|-----**S102**
| |------*001*
| |------*002*
| |------ …
| |------*007*
|----------

Tiếp theo, thực hiện sắp xếp các ảnh vào các thư mục với cảm xúc tương ứng là tên thư mục. Sau đó, đối với từng thư mục cảm xúc sẽ lấy ngẫu nhiên *80%* số lượng ảnh để huấn luyện và *20%* số lượng ảnh để kiểm thử tập huấn luyện.

Bước 2: Tiến hành huấn luyện để xây dựng bộ nhận dạng cảm xúc. Sử dụng thư viện OpenCV để phát hiện vùng chứa gương mặt. Dùng thư viện Dlib để xác định các landmarks – các điểm đặc trưng trên gương mặt được phát hiện. Dựa vào các landmarks này, tiến hành xây dựng tập huấn luyện, phân thành các lớp với các đặc trưng khác nhau tương ứng với từng cảm xúc khác nhau.

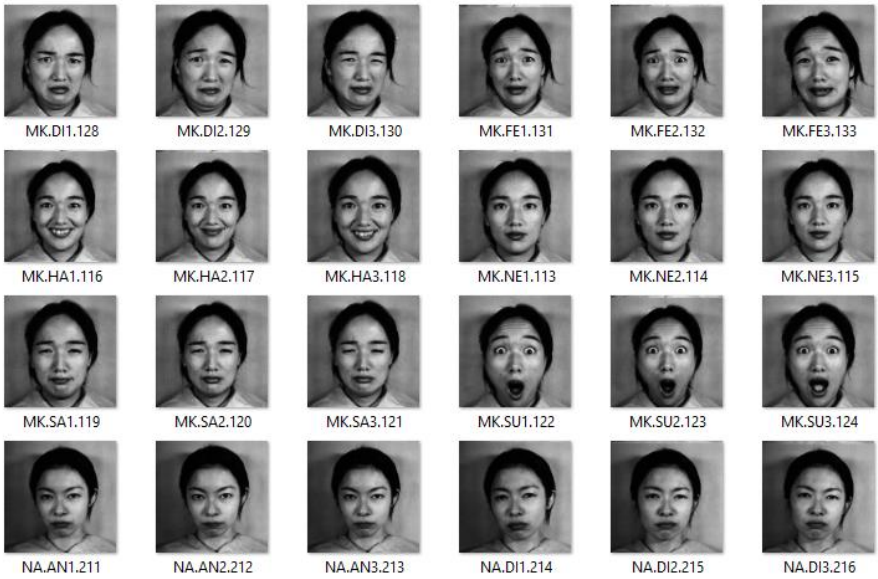
Chọn giá trị các tham số bằng file grid.py của thư viện LibSVM. Các kết quả chi tiết được trình bày trong phần kiểm thử và kết quả.

Bước 3: Kiểm thử tập huấn luyện. Kiểm tra xem với số lượng hình ảnh đầu vào là 20% tập dữ liệu ngẫu nhiên ban đầu thì mô hình nhận dạng vừa mới huấn luyện sẽ cho kết quả chính xác là bao nhiêu phần trăm. Bên cạnh đó, có thể sử dụng webcam để ghi

nhận hình ảnh người dùng và nhận dạng cảm xúc trực tiếp, thời gian thực, hiển thị cảm xúc ngay lập tức cho người dùng.

Trường hợp 2: Sử dụng tập dữ liệu JAFFE. Quy trình các bước được thực hiện tương tự như với tập dữ liệu Cohn-Kanade.

Bước 1: Tổ chức lại dữ liệu. Tập dữ liệu JAFFE được biểu diễn như hình bên dưới:



Hình 19: Cách đặt tên các file ảnh trong tập dữ liệu JAFFE

Cần phải sắp xếp các hình ảnh này vào đúng thư mục với tên thư mục biểu diễn cảm xúc mà nó thuộc về. Quan sát cách đặt tên sẽ nhận thấy được, hai ký tự thứ tư và thứ năm tính từ trái sang phải là hai ký tự đại diện cho cảm xúc của hình ảnh tương ứng. Ví dụ: *HA* là viết tắt của *happy*, có nghĩa là cảm xúc vui vẻ, *SA* là đại diện cho cảm xúc buồn và là viết tắt của từ *sad*. Chi tiết hơn sẽ được trình bày trong bảng bên dưới:

Bảng 3: Bảng mô tả ký hiệu của các nhãn và cảm xúc tương ứng trong tập dữ liệu JAFFE

Ký hiệu	Ý nghĩa	Cảm xúc
HA	Happy	Vui vẻ
SA	Sad	Buồn bã
SU	Surprise	Ngạc nhiên
AN	Anger	Giận dữ
FE	Fear	Sợ hãi
DI	Disgust	Kinh tởm
NE	Neutral	Không cảm xúc

Bước 2: Huấn luyện mô hình nhận dạng. Sử dụng đầu vào là các hình ảnh đã được sắp xếp ở bước 1. Các tham số dùng cho hàm huấn luyện gồm có: C=0.01 (kiểu float, là độ lỗi mô hình), kernel=poly (xác định loại hàm nhân được sử dụng trong thuật toán, bao gồm các giá trị: 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed', mặc định là 'rbf'), decision function shape=ovo (one vs one, giá trị mặc định là ovr – one vs rest, là hàm quyết định của hình dạng phân chia tập dữ liệu, hay cách thức để phân tách bộ dữ liệu), probability=True (mặc định là False, ước tính xác suất).

Bước 3: Kiểm tra độ chính xác của bộ nhận dạng vừa tạo ra ở bước 2. Đánh giá và đưa ra giải pháp cải thiện mô hình huấn luyện bằng cách thay đổi trọng số, giá trị của các tham số trong hàm huấn luyện, kiểm tra lại độ chính xác của các hình ảnh trong thư mục cảm xúc tương ứng, thay đổi cách thức huấn luyện. Kết quả huấn luyện và kiểm thử tập huấn luyện được trình bày trong bảng 4.

Bảng 4: Kết quả huấn luyện bộ nhận dạng sử dụng mô hình SVM và hai tập dữ liệu Cohn-Kanade, JAFFE

Tập dữ liệu		**Số mẫu phân lớp đúng**	**Số mẫu phân lớp sai**
Cohn-Kanade	Tập huấn luyện (4234 mẫu) Tập kiểm tra (1057 mẫu)	3,748/4,234 (chiếm 88.54%)	486/4,234 (chiếm 11.46%)
JAFFE	Tập huấn luyện (169 mẫu) Tập kiếm tra (42 mẫu)	97/169 (chiếm 57.14%)	72/169 (chiếm 42.86%)

So với tập dữ liệu JAFFE thì tập dữ liệu Cohn-Kanade được sử dụng với mô hình huấn luyện SVM cho kết quả huấn luyện và kiểm thử tốt hơn. Sự sai lệch này là do số lượng hình ảnh trong tập dữ liệu JAFFE ít hơn, chỉ bằng 1/24 so với tập dữ liệu Cohn-Kanade. Độ chính xác có thể thay đổi khi tăng số lượng ảnh dùng để huấn luyện và thực nghiệm để thay đổi tham số và giá trị thuộc tính phù hợp.

Trường hợp 3: Sử dụng SVM với đặc trưng HOG

Các bước thực hiện của trường hợp huấn luyện SVM với đặc trung HOG được trình bảy như sau:

Bước 1: Tạo vector đặc trưng của từng cảm xúc. Với mỗi hình ảnh trong từng thư mục cảm xúc sẽ được trích đặc trưng HOG. Sau đó xây dựng véc-tơ đặc trưng cho loại cảm xúc đó.

Bước 2: Tạo vector đặc trưng tổng của tất cả cảm xúc, được dùng để huấn luyện SVM.

Bước 3: Tạo nhãn cho từng cảm xúc để dễ phân lớp. Độ dài mảng chứa nhãn tương ứng với độ dài véc-tơ đặc trưng ở bước 2.

Bước 4: Tạo nhãn tổng. Tổng hợp ma trận gán gãn ở bước 3.

Bước 5: Huấn luyện SVM. Tiến hành sử dụng ma trận đặc trưng ở bước 3 và ma trận nhãn tương ứng ở bước 4 để huấn luyện bộ nhận dạng cảm xúc.

Bước 6: Nhận dạng. Dựa vào mô hình vừa huấn luyện để nhận dạng. Kết quả thu được với độ chính xác chưa cao. Tiến hành điều chỉnh tham số để tìm ra tham số phù hợp cho ra độ chính xác tốt nhất. Dưới đây là một số phép thử được thực hiện.

STT	C	gamma	kernel	Precision	Recall
1	10,000	0.0001	rbf	0.30	0.28
2	1	0.0001	poly	0.07	0.27
3	0.0001	0.001	poly	0.07	0.27
4	1	0.2	rbf	0.45	0.28
5	1	0.3	poly	0.24	0.26
6	1	1	rbf	0.21	0.27

Tham số tốt nhất được tìm thấy cho tập huấn luyện là:

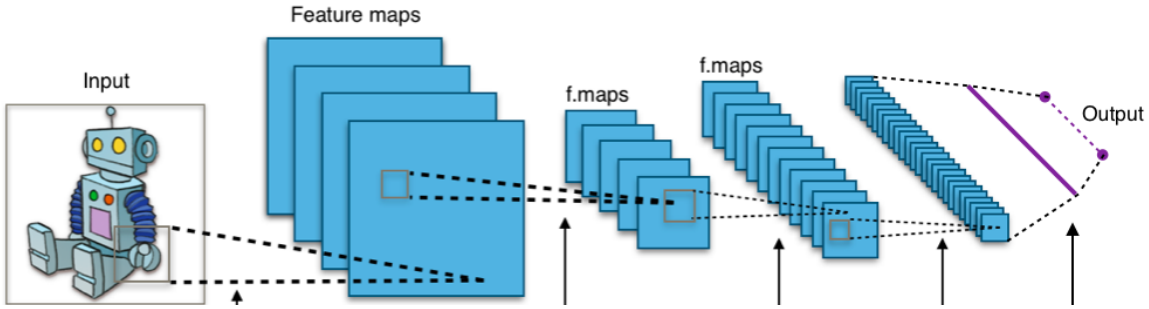
{'C': 1, 'gamma': 0.2, 'kernel': 'rbf'}

Các giá trị nhận được: precision là 0.45 và recal là 0.28. Độ chính xác là 45%.

3.4.6 Huấn luyện với ANN

Sử dụng mạng nơ-ron tích chập *(Convolutional Neural Networks)* để thực hiện giải thuật học sâu, dạy cho hệ thống có khả năng nhận dạng được cảm xúc của con người. CNN là một mạng nơ-ron Multilayer Perception đa tầng với cấu trúc đặc biệt, là một trường hợp đặc trưng và phổ biến của mạng nơ-ron nhân tạo.

MLP là mạng nơ-ron đa tầng truyền thẳng, bao gồm một tầng đầu vào, một tầng đầu ra và không giới hạn số tầng ẩn (các tầng không phải là tầng đầu vào và tầng đầu ra). Số lượng tầng ẩn được điều chỉnh để độ chính xác mô hình đạt được ở mức độ tốt nhất có thể, thông thường chỉ dao động từ một đến năm tầng. CNN được gọi là mạng nơ-ron tích chập, là một mô hình học sâu tiên tiến, cấu trúc đặc biệt của mạng MLP được thể hiện ở quá trình *Convolution* và *Pooling*. Hình 20 minh họa cấu trúc của một



Hình 20: Minh họa mô hình mạng nơ-ron Convolutional Neutral Networks

mạng CNN.

Các hình vuông thể hiện cho các mảng dữ liệu ở các tầng, hai quá trình tích chập và tổng hợp được thực hiện xen kẽ nhau qua các tầng này. Tầng đầu vào không chứa các nơ-ron mà chứa các giá trị ban đầu được cung cấp bởi người dùng. Giá trị này được xử lý trong mạng và tổng hợp thành một đặc trưng đại diện cho một lớp. Các lớp này tương ứng với các loại cảm xúc cần phải phân biệt của yêu cầu hệ thống.

Quá trình huấn luyện bộ nhận dạng được thực hiện lần lượt với hai tập dữ liệu Cohn-Kanade và JAFFE với những quy trình chi tiết trong từng trường hợp.

Trường hợp 1: Huấn luyện bộ nhận dạng với CNN và tập dữ liệu Cohn-Kanade

Bước 1: Tổ chức lại dữ liệu. Như đã trình bày ở trên, tập dữ liệu Cohn-Kanade được tổ chức thành các thư mục tương ứng với một người mẫu (được đặt tên như *S010, S011*, …), trong mỗi thư mục có các thư mục con (như *001, 002*, …). chứa các ảnh biểu hiện cảm xúc. Tiến hành sắp xếp lại các ảnh này vào các thư mục tương ứng với cảm xúc được thể hiện trong ảnh, tạm gọi là thu mục cảm xúc. Cấu trúc thư mục được tổ chức như bảng 5.

Bảng 5: Cách tổ chức thư mục của tập dữ liệu Cohn-Kanade

training-data |----------anger | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg |----------contempt | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg |----------disgust | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg |----------fear | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg	|----------happy | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg |----------neutral | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg |----------sadness | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg |----------surprise | |------ 1.jpg | |------ 2.jpg | |------ … | |------ n.jpg |----------

Sau đó, với từng thư mục cảm xúc sẽ lấy ngẫu nhiên 80% số lượng ảnh để huấn luyện và 20% số lượng ảnh để kiểm thử tập huấn luyện.

Bước 2: Tiến hành huấn luyện để xây dựng bộ nhận dạng. Sử dụng thư viện OpenCV để phát hiện vùng chứa gương mặt. Dùng thư viện Dlib để xác định các landmarks – các điểm đặc trưng trên gương mặt phát hiện được. Dựa vào các landmarks này, tiến hành xây dựng tập huấn luyện, phân thành các lớp với các đặc trưng khác nhau tương ứng với từng cảm xúc khác nhau.

Bước 3: Kiểm thử tập huấn luyện. Kiểm tra xem với số lượng hình ảnh đầu vào của tập dữ liệu ngẫu nhiên ban đầu thì bộ nhận dạng cảm xúc vừa mới huấn luyện sẽ cho kết quả chính xác bao nhiêu. Sau đó đưa ra các giải pháp điều chỉnh giá trị các tham số để tối ưu hóa mô hình nhận dạng. Có thể sử dụng webcam để ghi nhận hình ảnh người dùng và nhận dạng cảm xúc thời gian thực, trực tiếp hiển thị cảm xúc cho người dùng.

Trường hợp 2: Huấn luyện bộ nhận dạng với CNN và tập dữ liệu JAFFE. Quy trình các bước được thực hiện tương tự như với tập dữ liệu Cohn-Kanade.

Bước 1: Tổ chức lại dữ liệu tương tự như với tập dữ liệu Cohn-Kanade, sắp xếp các hình ảnh này vào đúng thư mục với tên thư mục biểu diễn cảm xúc mà nó thuộc về. Tuy nhiên, chỉ có thể nhận dạng sáu loại cảm xúc cơ bản vì JAFFE không có cảm xúc khinh miệt.

Bước 2: Huấn luyện mô hình nhận dạng. Sử dụng đầu vào là các hình ảnh đã được sắp xếp ở bước 1. Các tham số dùng cho hàm huấn luyện gồm có: C=0.01, kernel=poly, decision function shape=ovo, probability=True.

Bước 3: Kiểm tra độ chính xác của bộ nhận dạng vừa tạo ra ở bước 2. Đánh giá và đưa ra giải pháp cần thiết để cải thiện mô hình huấn luyện. Kết quả huấn luyện được trình bày trong bảng 6.

Bảng 6: Kết quả huấn luyện bộ nhận dạng sử dụng mô hình CNN và hai tập dữ liệu Cohn-Kanade, JAFFE

Tập dữ liệu		**Số mẫu phân lớp đúng**	**Số mẫu phân lớp sai**
Cohn-Kanade	Tập huấn luyện (4234 mẫu) Tập kiểm tra (1058 mẫu)	4,209/4,234 (chiếm 99.41%)	25/4,234 (chiếm 0.59%)
JAFFE	Tập huấn luyện (171 mẫu)	150/171	21/171

	Tập kiểm tra (42 mẫu)	(chiếm 87.72 %)	(chiếm 12.28%)

So với tập dữ liệu JAFFE thì tập dữ liệu Cohn-Kanade cho kết quả huấn luyện và kiểm thử tốt hơn khi được dùng với mô hình CNN. Sự sai lệch này là do số lượng hình ảnh trong tập dữ liệu JAFFE tương đối ít, nếu số lượng hình ảnh của JAFFE nhiều hơn thì độ chính xác sẽ được cải thiện.

Đối với tập dữ liệu Cohn-Kanade, khi được sử dụng với mô hình huấn luyện CNN cho kết quả huấn luyện tốt hơn phương pháp SVM. Tương tự như vậy, CNN dùng tập dữ liệu JAFFE cho ra kết quả cao hơn so với SVM.

3.4.7 Nhận dạng cảm xúc

Sử dụng ảnh tĩnh và video để kiểm tra độ chính xác của mô hình huấn luyện. Đối với ảnh tĩnh, kiểm tra từng hình ảnh có đối tượng gương mặt biểu hiện một loại cảm xúc nào đó, và mong đợi hệ thống sẽ cho ra kết quả cảm xúc đúng với cảm xúc mà người sử dụng mong muốn. Song song đó, có thể dùng video có chứa các đối tượng người và biểu cảm của họ để thực hiện công việc nhận dạng cảm xúc; hoặc là sử dụng webcam của máy tính để phát hiện cảm xúc thời gian thực.

CHƯƠNG 4: THỰC NGHIỆM VÀ ĐÁNH GIÁ

4.1 Yêu cầu phần cứng và phần mềm

4.1.1 Yêu cầu phần cứng
Phần cứng phải đảm bảo các yêu cầu sau:

- Laptop, desktop có cấu hình tương đối mạnh.
- Bộ nhớ RAM: > 2GB

4.1.2 Yêu cầu phần mềm
Về phần mềm, gồm các yêu cầu sau:

- Hệ điều hành: Windows 7 trở lên.
- Ngôn ngữ lập trình: Python
- Sử dụng công cụ phát triển phần mềm IDE: PyCharm
- Thư viện mã nguồn mở: OpenCV, Dlib, Numpy, Matplotlib và các thư viện hỗ trợ khác.

4.2 Giao diện chương trình

Chương trình được trình bày giao diện như sau:

- Công cụ sử dụng để tạo giao diện người dùng: QtPy Designer.

- Đảm bảo dễ hiểu và dễ dùng đối với người sử dụng phần mềm.
- Giao diện chính:

4.3 Kiểm thử và kết quả

4.3.1 Một số hình ảnh nhận dạng

Với ảnh tĩnh:

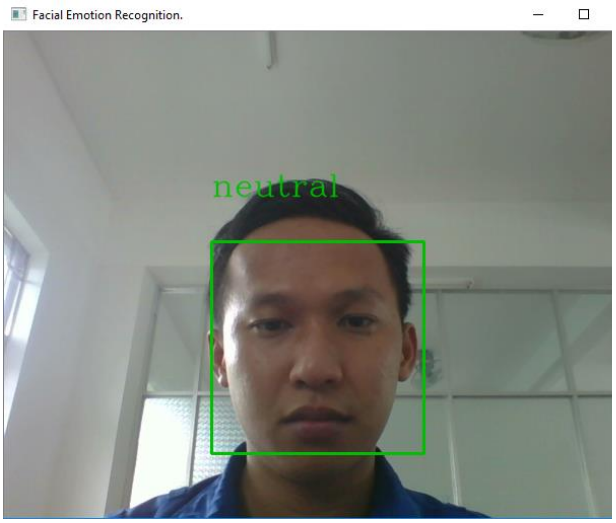
Bảng 7: Nhận dạng cảm xúc của hình ảnh tĩnh với mạng nơ-ron tích chập

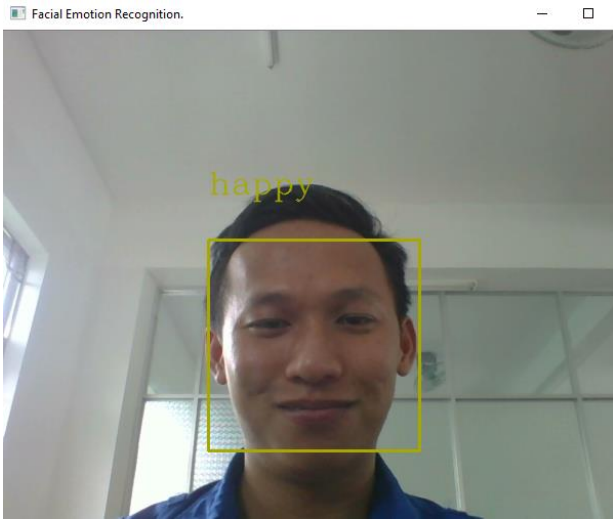
result = [[7.5705007e-02 7.7710385e-05 8.1867181e-02 4.7940958e-01 5.8311332e-02
7.8736812e-02 2.2589242e-01]]

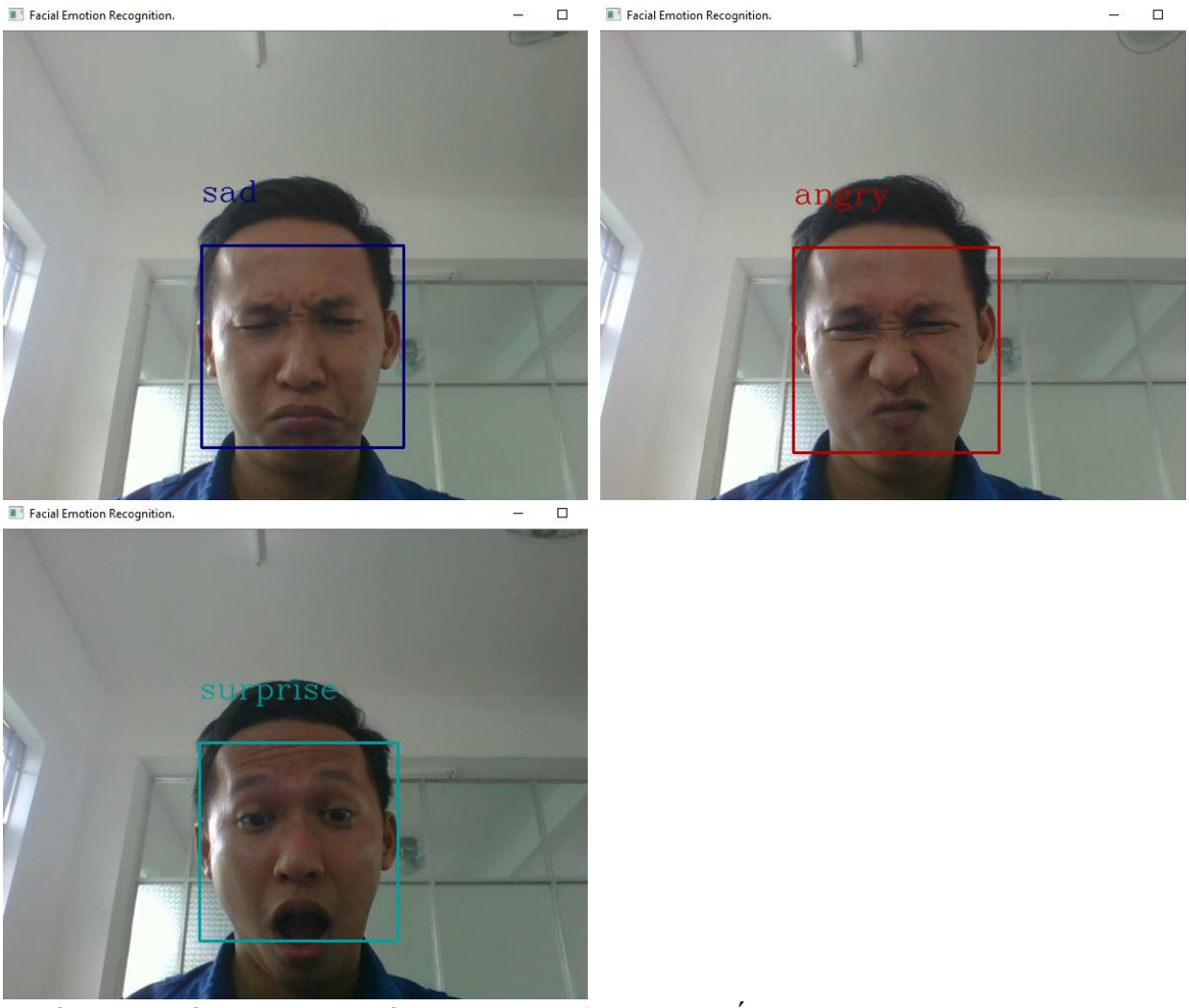
biggest = 0.47940958

emotion = happy

Với ảnh từ video webcam máy tính:







Hình 21: Hình ảnh cảm xúc nhận dạng được trực tiếp từ webcam sử dụng bộ nhận dạng mạng nơ-ron CNN

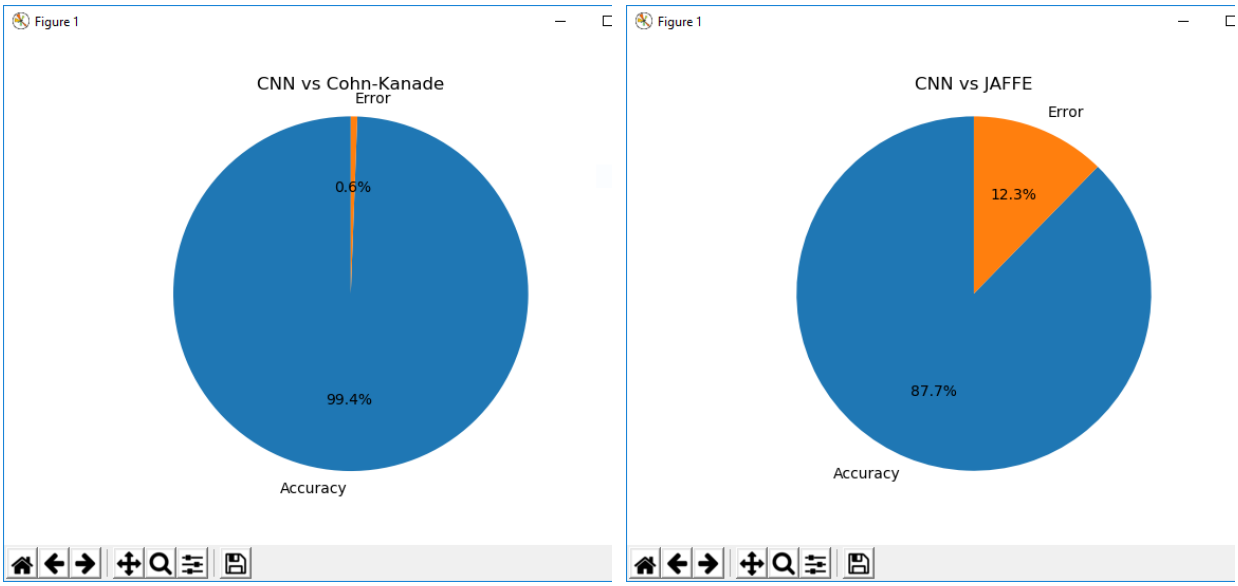
4.3.2 Một số thử nghiệm mô hình và các tham số

Thực hiện chạy thử nghiệm mô hình huấn luyện SVM với các tham số khác nhau sẽ cho ra các kết quả tương ứng. Tìm ra các tham số tối ưu nhất cho thuật toán sao cho độ chính xác là cao nhất.

Bảng 8: Mô hình huấn luyện SVM với tham số decision_function_shape=ovo

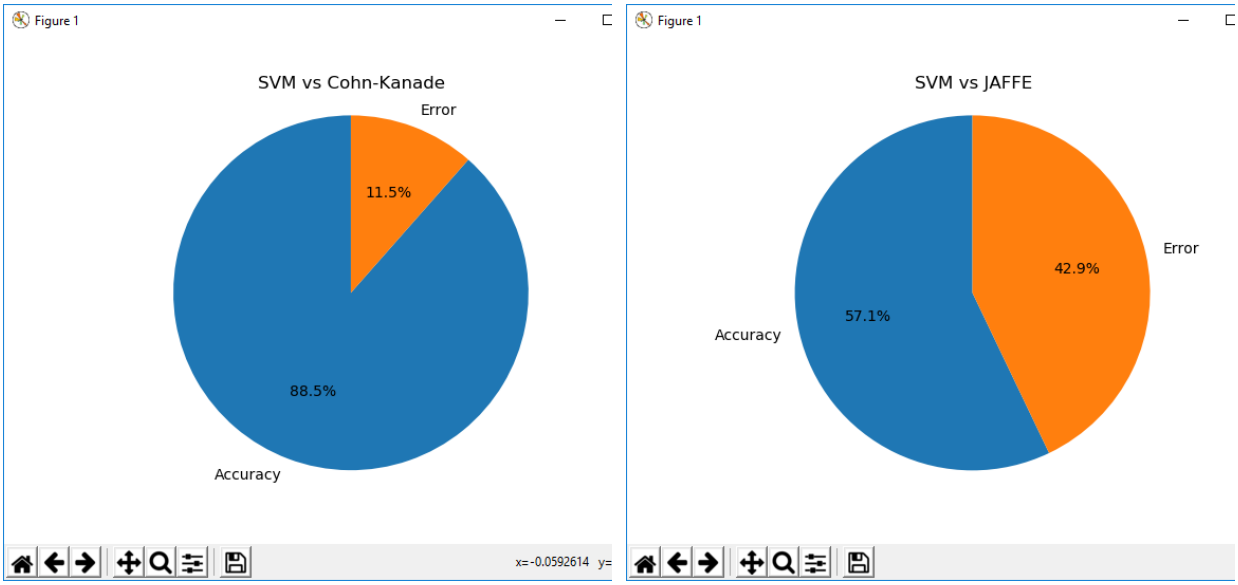
STT	C	kernel	gamma	Precision	Recall	Best Accuracy
1	0.01	rbf	0.2	0.34	0.12	33.99%
2	0.01	rbf	auto	0.34	0.12	33.99%
3	0.01	poly	auto	0.89	0.89	88.54%
4	100,000	rbf	0.01	0.76	0.5	49.72%
5	1	rbf	0.2	0.63	0.35	35.22%

So sánh độ chính xác khi sử dụng một mô hình huấn luyện CNN với hai tập dữ liệu Cohn-Kanada và JAFFE:



Hình 22: So sánh độ chính xác khi sử dụng CNN với Cohn-Kanade và JAFFE

So sánh độ chính xác khi sử dụng một mô hình huấn luyện SVM với hai tập dữ liệu Cohn-Kanade và JAFFE:



Hình 23: So sánh độ chính xác khi sử dụng SVM với Cohn-Kanade và JAFFE

Thực nghiệm cùng tập dữ liệu Cohn-Kanade với phương pháp k-NN. Qua 20 lần chạy, kết quả như sau:

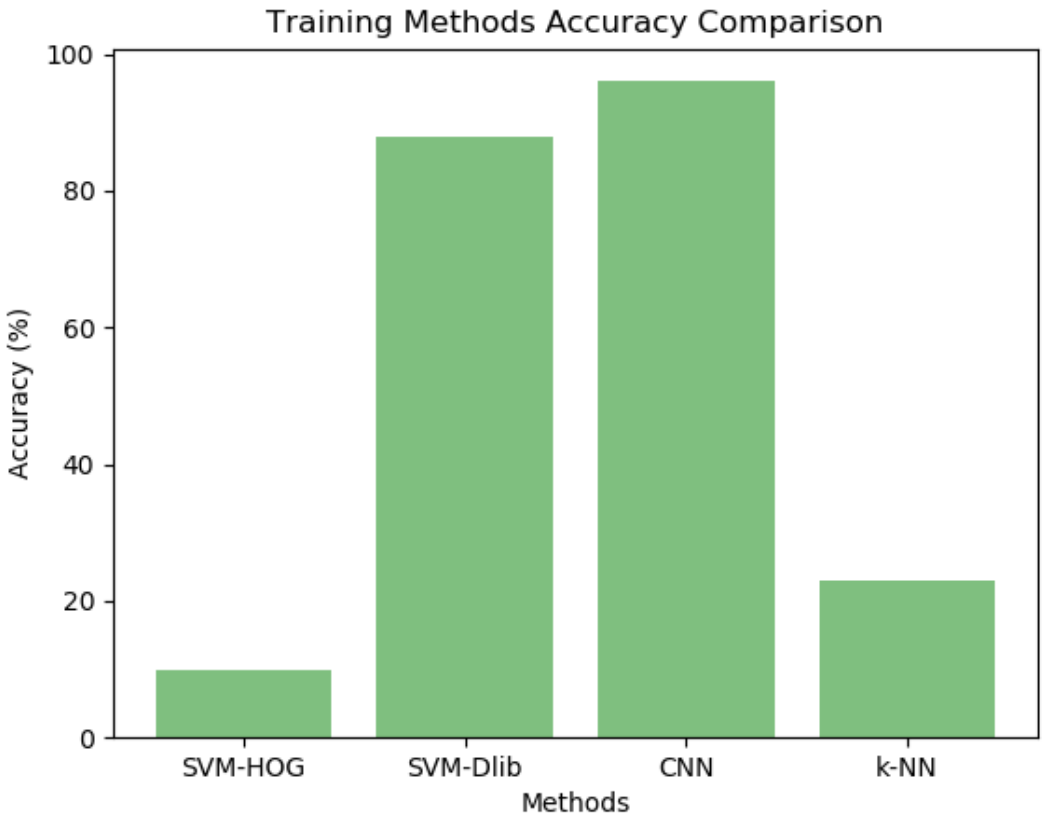
k: 1 scores: 0.2265084075173096	k: 11 scores: 0.22354104846686448
k: 2 scores: 0.21167161226508407	k: 12 scores: 0.2195845697329377
k: 3 scores: 0.21167161226508407	k: 13 scores: 0.228486646884273
k: 4 scores: 0.20870425321463898	k: 14 scores: 0.228486646884273
k: 5 scores: 0.21068249258160238	k: 15 scores: 0.2294757665677547
k: 6 scores: 0.20969337289812068	k: 16 scores: 0.228486646884273
k: 7 scores: 0.228486646884273	k: 17 scores: 0.2304648862512364
k: 8 scores: 0.22354104846686448	k: 18 scores: 0.2304648862512364
k: 9 scores: 0.2265084075173096	k: 19 scores: 0.23837784371909002
k: 10 scores: 0.2274975272007913	

Độ chính xác cao nhất xấp xỉ 23.83% với k = 19

So sánh hiệu quả của các phương pháp nhận dạng cảm xúc khuôn mặt theo độ chính xác (%), thời gian huấn luyện (phút), thời gian nhận dạng (phút):

STT	**Phương pháp**	**Độ chính xác (%)**	**Thời gian huấn luyện (phút)**	**Thời gian nhận dạng (phút)**
1	SVM-HOG	30.01	16.5	2.4
2	SVM-Dlib	88.53	17.8	6.5
3	CNN	97.64	516	10.2
4	k-NN	23.83	2.6	1.2

Biểu đồ so sánh độ chính xác của các tập huấn luyện:



Thời gian huấn luyện của từng phương pháp:

4.4 Đánh giá kết quả đạt được

Đã tìm hiểu được các loại cảm xúc cơ bản của con người, các biểu hiện của gương mặt, sự vận động của các nhóm cơ mặt để biểu hiện một loại cảm xúc nhất định. Tìm hiểu được các phương pháp để trích rút ra các đặc trưng của các thành phần chính của gương mặt như Haar-like, HOG, sự tương quan của các đơn vị vận động trên gương mặt. Biết được cách huấn luyện mô hình SVM, nhận dạng cảm xúc từ mô hình được xây dựng. Tìm hiểu được về mạng nơ-ron nhận tạo, mạng MLP, mạng CNN và các thao tác trên các mạng này để xây dựng bộ nhận dạng và nhận dạng cảm xúc con người thông qua biểu cảm gương mặt. Xây dựng được một hệ thống cơ bản để nhận dạng được cảm xúc của con người thông qua hình ảnh và video thu nhận được.

CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 Kết luận

Qua quá trình thực hiện luận văn, nghiên cứu về các phương pháp, công cụ, kỹ thuật nhận dạng cảm xúc dựa trên gương mặt, học viên đã tìm hiểu và biết được một số thuật toán và cách thức để áp dụng vào bài toán nhận dạng cảm xúc. Một số kết quả chính của luận văn:

- Tìm hiểu được các loại cảm xúc cơ bản của con người trên trái đất, không phân biệt chủng tộc, độ tuổi hay giới tính.
- Trình bày phương pháp phân tích thành phần chính và trích chọn đặc trưng dựa vào PCA, làm đầu vào cho huấn luyện tập dữ liệu với ANN và SVM
- Ứng dụng các công cụ và công nghệ này để thực hiện nhận dạng cảm xúc thông qua biểu cảm gương mặt.
- Thu thập kết quả, phân tích, đánh giá, thống kê, nhận xét về kết quả đặt được khi áp dụng những công cụ khác nhau cho cùng một tập dữ liệu, cùng một bài toán.

Tuy nhiên, đề tài còn nhiều điểm cần phải khắc phục như: Cần phải xây dựng được tập huấn luyện với đủ nhiều số lượng ảnh để cho kết quả chính xác hơn và chấp nhận được. Bên cạnh đó, cần nghiên cứu các phương pháp trích chọn đặc trưng các thành phần của gương mặt người như giải thuật AAM cải tiến, với mục đích chọn ra được chính xác từng thành phần của gương mặt hơn, gia tăng độ chính xác của bài toán. Hơn hết, đề tài cần phải xây dựng được một chương trình hoàn thiện có giao diện tương tác để thân thiện với người dùng hơn.

5.2 Thách thức trong nhận dạng cảm xúc dựa trên mặt người

Nhận dạng cảm xúc con người dựa vào các biểu cảm trên gương mặt là một bài toán khó và mang tính tương đối, vì cảm xúc của con người là vô cùng đa dạng. Ngoài sáu cảm xúc cơ bản, con người còn có rất nhiều cảm xúc khác với sự sai khác không rõ rệt và rất khó phân biệt.

Bên cạnh đó, phương pháp nhận dạng cảm xúc dựa trên gương mặt mang tính tương đối là do chúng ta chỉ dựa vào các nhóm cơ mặt và vị trí các thành phần trên gương mặt để xác định cảm xúc, và cảm xúc này có thể chưa thật sự đúng đắn.

Một khó khăn khác về môi trường thực hiện, hệ thống sẽ cho các kết quả không tốt với độ chính xác thấp nếu ảnh đầu vào có chất lượng kém, điều kiện ánh sáng không tốt, ảnh mờ hay vùng gương mặt khó nhận dạng, kích thước quá nhỏ. (12 p. 11)

5.3 Hướng phát triển

Cải tiến, mở rộng đề tài để có thể nhận dạng thêm được nhiều loại cảm xúc phức tạp hơn của con người.

Nghiên cứu và áp dụng hệ thống vào một ứng dụng thực tế và hữu ích như máy nghe nhạc theo cảm xúc, điều khiển robot theo cảm xúc.

Cải tiến độ chính xác của đề tài với nhiều dữ liệu huấn luyện hơn, sử dụng các thuật toán khác nhau cũng như so sánh và cải tiến kết quả thực hiện.

TÀI LIỆU THAM KHẢO

1. **Wikipedia.** Facial Action Coding System. *Wikipedia.* [Online] April 05, 2018. https://en.wikipedia.org/wiki/Facial_Action_Coding_System.

2. **Nguyễn , Tuân Hữu and Nguyễn, Thủy Văn.** *Xây dựng hệ thống nhận dạng mặt tự động sử dụng LPQ (Local Phase Quantization).* Hải Phòng : Nguyễn Hữu Tuân, 2016. 1.

3. **Team, OpenCV Dev.** OpenCV Documentation. *Introduction to OpenCV - Python Tutorials.* [Online] Novemver 10, 2014. [Cited: April 27, 2018.] https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_setup/py_intro/py_intro.html#intro.

4. **Đỗ, Nghị Thanh and Phạm, Khang Nguyên.** *Giáo trình Nguyên Lý Máy Học.* Cần Thơ : Đại học Cần Thơ, 2012.

5. *Tìm hiểu về Support Vector Machine cho bài toán phân lớp quan điểm.* **Phạm, Sơn Văn.** Hải Phòng : s.n., 2012.

6. *Emotion Detection: A Feature Analysis.* **N Dharmesh and Mausmi, Kulshreshtha.** Issue 7, Mumbai, India : IJASCSE, 2015, Vol. 4. ISSN: 2278 7917.

7. *Automatic Emotion Recognition Using Facial Expression: A Review.* **Monika , Dubey and Prof. Lokesh Singh.** Issue 2, Bhopal, India : IRJET, Feb, 2016, Vol. 3. e-ISSN: 2395 0056 P-ISSN: 2395 0072.

8. *An Efficient Real-Time Emotion Detection Using Camera and Facial Landmarks.* **Binh T. Nguyen, et al.** Da Nang, Viet Nam : IEEE, 2017. Electronic ISBN: 978-1-5090-5401-5.

9. *Emotion Detection Through Facial Feature Extraction.* **James Pao.** California, The USA : s.n.

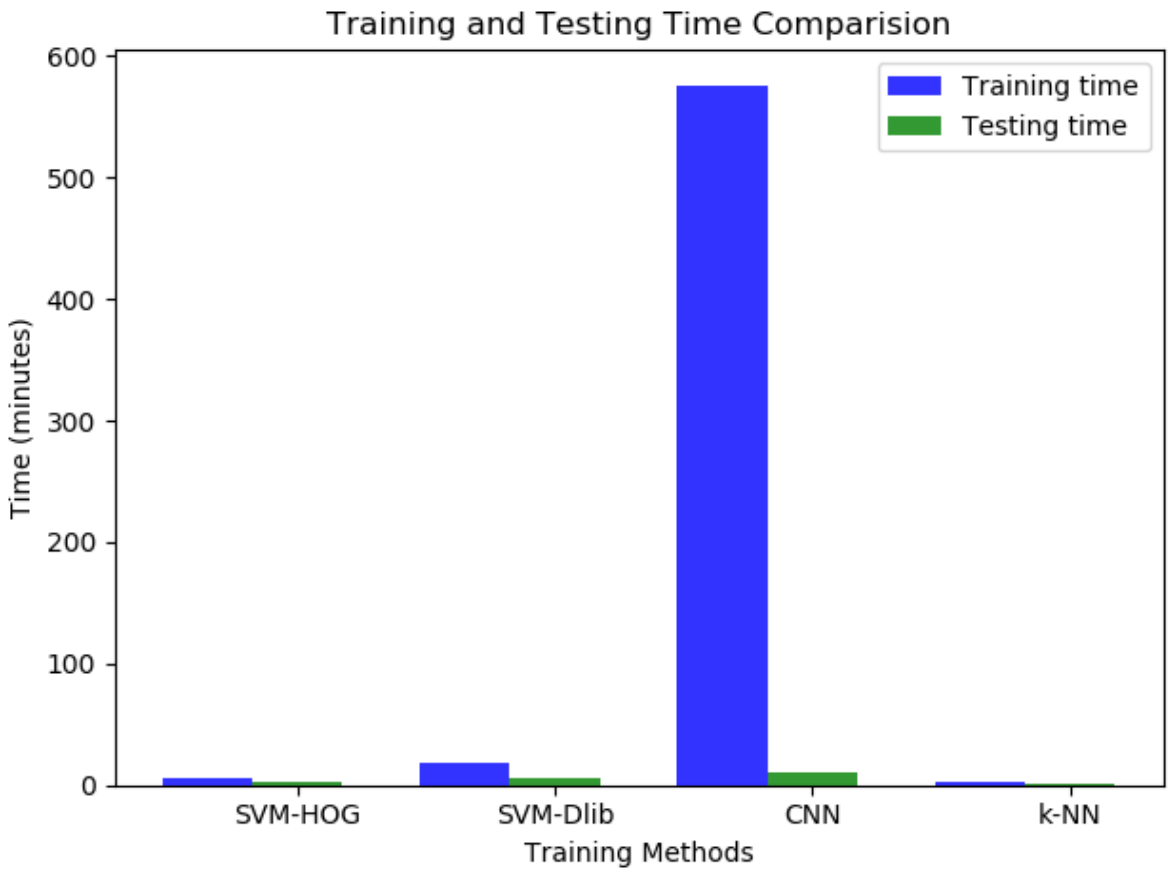
10. **Jeffrey, Cohn.** Affect Analysis Group. *Cohn-Kanade AU-Coded Expression Database.* [Online] Facial expression database. [Cited: 5 23, 2018.]

11. **Michael, Lyons, Miyuki , Kamachi and Jiro, Gyoba.** The Japnese Female Facial Expression (JAFFE) Database. [Online] Psychology Department, Kyushu University. [Cited: 5 23, 2018.] http://www.kasrl.org/jaffe.html.

12. *Nghiên cứu nhận dạng biểu cảm mặt người trong tương tác người máy.* **Nguyễn, Vân Thị Thanh.** Hải phòng : s.n., 2016.

PHỤ LỤC

So sánh thời gian huấn luyện và thời gian kiểm thử của các phương pháp:



Hình 24: So sánh thời gian huấn luyện và thời gian kiểm thử của các phương pháp SVM-HOG, SVM-Dlib, CNN, k-NN