

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC SƯ PHẠM THÀNH PHỐ HỒ CHÍ MINH**

Nguyễn Thị Anh Thư

**NHẬN DẠNG CẢM XÚC CỦA NGƯỜI DỰA
TRÊN ẢNH NHIỆT**

**Chuyên ngành : Khoa học máy tính
Mã số : 8480101**

LUẬN VĂN THẠC SĨ MÁY TÍNH

**NGƯỜI HƯỚNG DẪN KHOA HỌC:
TS. NGUYỄN VIẾT HƯNG**

301 - 9394

THƯ VIỆN
Trường Đại-Học Sư-Phạm
TP. HỒ-CHÍ-MINH

Thành phố Hồ Chí Minh - 2018

LỜI CAM ĐOAN

Tôi tên Nguyễn Thị Anh Thư, học viên cao học khóa 27, chuyên ngành khoa học máy tính, trường đại học Sư phạm Thành phố Hồ Chí Minh. Mã số học viên KHMT-16-012.

Tôi xin cam đoan luận văn “Nhận dạng cảm xúc của người dựa trên ảnh nhiệt” là do tôi tìm hiểu, nghiên cứu và thực hiện dưới sự hướng dẫn của TS. Nguyễn Viết Hưng. Luận văn không có sự sao chép từ các tài liệu, công trình nghiên cứu khác mà không ghi rõ nguồn trong tài liệu tham khảo.

Tôi xin chịu trách nhiệm về kết quả thực hiện và lời cam đoan này.

Tp Hồ Chí Minh, ngày 25 tháng 09 năm 2018

LỜI CẢM ƠN

Đầu tiên, tôi xin kính gửi lời cảm ơn đến quý Thầy đã trực tiếp giảng dạy và truyền đạt kiến thức cho tôi trong suốt khóa học cao học chuyên ngành khoa học máy tính vừa qua tại trường Đại học Sư phạm Thành phố Hồ Chí Minh.

Đặc biệt, tôi xin chân thành gửi lời cảm ơn sâu sắc đến Thầy của tôi, TS. Nguyễn Viết Hưng, người đã định hướng, giúp đỡ, trực tiếp hướng dẫn và tận tình chỉ bảo tôi trong suốt quá trình tìm hiểu, tiếp cận, nghiên cứu, xây dựng và hoàn thiện luận văn này.

Tôi cũng xin được cảm ơn sự hỗ trợ đến từ Phòng Sau Đại học, Khoa Công nghệ thông tin trường Đại học Sư phạm Thành phố Hồ Chí Minh và các anh chị học viên cùng khóa trong thời gian tôi học tập tại trường.

Cuối cùng, tôi muốn gửi đến ba mẹ tôi cũng như gia đình nhỏ của tôi lòng biết ơn và tình cảm yêu thương nhất. Nếu không có sự chia sẻ, đồng cảm và ủng hộ từ gia đình thì tôi đã không thể có đủ niềm tin, động lực để hoàn thành khóa học và luận văn của mình.

Tp Hồ Chí Minh, ngày 25 tháng 09 năm 2018

MỤC LỤC

Trang

Lời cam đoan

Lời cảm ơn

Mục lục

Danh mục các ký hiệu và chữ viết tắt

Danh mục các bảng biểu

Danh mục các hình vẽ, biểu đồ

MỞ ĐẦU 1

Chương 1. CƠ SỞ LÝ THUYẾT VÀ NGHIÊN CỨU LIÊN QUAN 5

 1.1. Tổng quan tình hình nghiên cứu 5

 1.1.1. Các công trình nghiên cứu liên quan 5

 1.1.2. Một số thách thức trong lĩnh vực nghiên cứu nhận dạng cảm xúc
 người 10

 1.2. Thông tin về ảnh nhiệt 10

 1.2.1. Giới thiệu ảnh nhiệt 10

 1.2.2. Ưu điểm 12

 1.2.3. Hạn chế 13

 1.3. Một số phương pháp rút trích đặc trưng sử dụng trong nhận dạng cảm
 xúc 13

 1.3.1. Active Appearance Model (AAM) 13

 1.3.2. Local Binary Pattern (LBP) 15

 1.4. Một số phương pháp sử dụng phân loại cảm xúc 17

 1.4.1. Linear Discriminant Analysis (LDA) 17

 1.4.2. Support Vector Machines (SVM) 18

 1.4.3. Principal Component Analysis (PCA) 20

1.4.4. Eigenspace Method based on Class features (EMC).....	22
Chương 2. NHẬN DẠNG CẢM XÚC CỦA NGƯỜI DỰA TRÊN ẢNH NHIỆT.....	25
2.1. Cơ sở dữ liệu ảnh nhiệt Kotani Thermal Facial Emotion (KTFE)	25
2.2. Phương pháp nhận dạng	26
2.2.1. Tìm các vùng nhiệt quan tâm (ROI)	26
2.2.2. Nhận dạng cảm xúc.....	28
Chương 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ.....	31
3.1. Môi trường thực nghiệm	31
3.2. Dữ liệu đầu vào	31
3.3. Phân tích đánh giá	31
3.3.1. Kết quả thực nghiệm.....	32
3.3.2. So sánh kết quả thực nghiệm với kết quả của phương pháp [35]... ..	36
3.3.3. So sánh kết quả thực nghiệm với kết quả của phương pháp [38]... ..	38
Chương 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	41
CÔNG TRÌNH CÔNG BỐ.....	42
TÀI LIỆU THAM KHẢO	43
PHỤ LỤC	49

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Chữ viết tắt	Nguyên mẫu	Điễn giải
AAM	Active Appearance Model	Mô hình xuất hiện tích cực
AU	Action Units	Đơn vị hành động, xác định độ co của cơ mặt
CNN	Convolutional Neural Networks	Mô hình mạng nơ-ron tích chập
DBM	Deep Boltzmann Machine	Mô hình máy Boltzman học sâu
ECG	Electrocardiogram	Tín hiệu điện tâm đồ
EEG	Electroencephalogram	Tín hiệu điện não đồ
EMC	Eigenspace Method based on Class features	Phương pháp không gian riêng dựa trên các đặc trưng của lớp
FACS	Facial Action Coding System	Hệ thống mã hóa hành động mặt
FTFP	Facial Thermal Feature Points	Cơ sở các điểm đặc trưng nhiệt mặt
HMM	Hidden Markov Models	Mô hình Markov ẩn
HOG	Histogram of Oriented Gradients	Biểu đồ của các gradient định hướng
KTFE	Kotani Thermal Facial Emotion	Cơ sở dữ liệu ảnh nhiệt
LBP	Local Binary Pattern	Mô hình nhị phân cục bộ
LDA	Linear Discriminant Analysis	Phương pháp phân tích phân biệt tuyến tính

LDP	Local Directional Pattern	Mô hình định hướng cục bộ
n-EMC	Norm-Eigenspace method based on class Feature	Phương pháp không gian riêng chuẩn dựa trên đặc trưng lớp
NVIE	Natural Visible and Infrared facial Expression	Cơ sở dữ liệu ảnh nhiệt
PCA	Principal Component Analysis	Phương pháp phân tích thành phần chính
PTSD	Posttraumatic Stress Disorder	Chấn thương tâm lý
ROI	Region of Interest	Khu vực quan tâm
SVM	Support Vector Machines	Phương pháp Máy vector hỗ trợ
t-PCA	thermal-Principal Component Analysis	Phương pháp phân tích thành phần chính nhiệt
t-ROI	Thermal Region of Interest	Các khu vực nhiệt quan tâm

DANH MỤC CÁC BẢNG BIỂU

Bảng 1.1. Một số cơ sở dữ liệu ảnh nhiệt [42].....	12
Bảng 3.1. Ma trận kết quả với PCA. Từ trái sang phải (trên xuống dưới): giận dữ (An), ghê tởm (Di), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne), buồn rầu (Sa) và ngạc nhiên (Su).....	33
Bảng 3.2. Ma trận kết quả với EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), ghê tởm (Di), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne), buồn rầu (Sa) và ngạc nhiên (Su).....	33
Bảng 3.3. Ma trận kết quả với PCA-EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), ghê tởm (Di), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne), buồn rầu (Sa) và ngạc nhiên (Su).....	34
Bảng 3.4. So sánh phân tích với phương pháp [35].....	37
Bảng 3.5. Ma trận kết quả với PCA. Từ trái sang phải (trên xuống dưới): giận dữ (An), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne) và buồn rầu (Sa).....	39
Bảng 3.6. Ma trận kết quả với EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne) và buồn rầu (Sa).....	39
Bảng 3.7. Ma trận kết quả với PCA-EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne) và buồn rầu (Sa).....	40
Bảng 3.8. So sánh phân tích với phương pháp [38].....	40

DANH MỤC CÁC HÌNH VẼ, BIỂU ĐỒ

Hình 1.1. Ví dụ về đặc trưng hình học của mặt gồm lông mày, mắt, mũi, miệng [17]	6
Hình 1.2. Sơ đồ quy trình thực hiện của phương pháp [35]	9
Hình 1.3. Trích xuất ROI trong phương pháp [38]	9
Hình 1.4. Ví dụ về AAM [47]	14
Hình 1.5. Ví dụ về LBP cơ bản. Lấy giá trị điểm trung tâm làm ngưỡng xác định các giá trị điểm xung quanh chuyển về 0,1. Từ giá trị nhị phân nhận được tính toán giá trị thập phân tương ứng [22]	16
Hình 1.6. Các vector đặc trưng LBP được trích xuất từ các vùng khác nhau trên hình ảnh khuôn mặt [22]	16
Hình 1.7. Siêu phẳng tách với khoảng cách lề cực đại [44]	19
Hình 1.8. Ví dụ về PCA [45]	21
Hình 1.9. Ví dụ về EMC [46]	23
Hình 2.1. Mẫu ảnh nhiệt và ảnh nhìn thấy của bảy cảm xúc[42]	25
Hình 2.2. Ví dụ về ROI [51]	26
Hình 2.3. Nhận dạng cảm xúc với ROI và PCA	28
Hình 2.4. Nhận dạng cảm xúc với ROI và EMC	30
Hình 3.1. Ví dụ hình ảnh nhiệt của bảy cảm xúc (bình thường, tức giận, hạnh phúc, buồn, sợ hãi, ghê tởm, ngạc nhiên) [42]	31
Biểu đồ 3.1. Kết quả nhận dạng cảm xúc của PCA	34
Biểu đồ 3.2. Kết quả nhận dạng cảm xúc của EMC	35
Biểu đồ 3.3. Kết quả nhận dạng cảm xúc của PCA-EMC	35
Biểu đồ 3.4. So sánh giữa PCA, EMC và PCA-EMC	36
Biểu đồ 3.5. So sánh giữa phương pháp đề xuất và phương pháp [35] cho PCA	37
Biểu đồ 3.6. So sánh giữa phương pháp đề xuất và phương pháp [35] cho EMC	38

MỞ ĐẦU

1. Lý do chọn đề tài

Cảm xúc đóng một vai trò quan trọng trong giao tiếp phi ngôn ngữ, là một trong những cơ sở để hiểu được hành vi của con người. Có thể nhận dạng cảm xúc thông qua nhiều hình thức khác nhau, và đặc biệt là thông qua biểu hiện trên khuôn mặt [1], [6].

Trong xu hướng phát triển của khoa học kỹ thuật, công nghệ thông tin, nhận dạng cảm xúc trở thành một vấn đề đã và đang thu hút được sự quan tâm, nghiên cứu của các nhà khoa học. Đầu những năm 1970, Paul Ekman và các đồng nghiệp của ông đã thực hiện các nghiên cứu sâu rộng về các biểu hiện trên khuôn mặt con người [2]. Trong suốt những thập kỷ sau đó, phương pháp nhận dạng cảm xúc đã được ứng dụng phổ biến trong cuộc sống, các hệ thống phát hiện trạng thái buồn ngủ dựa vào cảm xúc trên khuôn mặt được phát triển để cảnh báo cho người lái xe khi thấy dấu hiệu buồn ngủ, mệt mỏi [48], các phần mềm điều khiển dựa vào cảm xúc, các ứng dụng trò chơi video thông minh [49], tương tác người và robot [50], v.v.

Tuy nhiên, do sự phức tạp của khuôn mặt người, biểu cảm của cảm xúc và quan điểm về hành vi cho thấy nghiên cứu về lĩnh vực này vẫn còn là một thách thức lớn [3], [4], [7], [8].

Ngoài ra, các hệ thống dựa trên các hình ảnh nhìn thấy phải đối mặt với những thách thức như: chất lượng ảnh không tốt do điều kiện ánh sáng kém, nhiều người không biểu hiện nên không thể nhận biết được cảm xúc thật, v.v. Tất cả những nguyên nhân đó đều có thể làm giảm độ chính xác của kết quả nghiên cứu. Việc sử dụng hình ảnh nhiệt là một trong những giải pháp khắc phục những hạn chế của ảnh nhìn thấy được [9]. Hình ảnh nhiệt không nhạy cảm với điều kiện ánh sáng. Bên cạnh đó, không phải toàn bộ ảnh có tác động thay đổi nhiệt độ khi cảm xúc thay đổi nên việc rút trích đặc trưng vùng quan

trọng là rất cần thiết. Có thể nói, sử dụng ảnh nhiệt là một kỹ thuật đầy hứa hẹn trong việc nhận dạng cảm xúc thực tế hơn.

Xuất phát từ các cơ sở đó, chúng tôi chọn đề tài “**Nhận dạng cảm xúc của người dựa trên ảnh nhiệt**”. Luận văn làm rõ cơ sở lý thuyết và các nghiên cứu liên quan cũng như thực nghiệm, đánh giá kết quả đạt được trong thực tế và khả năng phát triển của đề tài. Quan trọng hơn đề tài mong muốn đóng góp một phần nghiên cứu về sự cải tiến quá trình rút trích đặc trưng tốt hơn trong bài toán nhận dạng cảm xúc.

2. Mục tiêu và nhiệm vụ nghiên cứu

Mục tiêu: Mục tiêu của đề tài là xây dựng một mô hình rút trích đặc trưng của ảnh nhiệt hỗ trợ cho việc nhận dạng cảm xúc của người dựa trên ảnh nhiệt của khuôn mặt người.

Nhiệm vụ: Để đạt mục tiêu trên, đề tài thực hiện những nhiệm vụ sau đây:

Thứ nhất, tìm hiểu tổng quan tình hình nghiên cứu về nhận dạng, phân tích và dự đoán cảm xúc của người bằng ảnh thông thường và ảnh nhiệt.

Thứ hai, nghiên cứu các cơ sở lý thuyết của đề tài.

Thứ ba, xây dựng và giải quyết bài toán nhận dạng cảm xúc của người dựa trên ảnh nhiệt.

Thứ tư, tiến hành thực nghiệm chương trình và đánh giá kết quả đạt được.

Thứ năm, đúc kết kinh nghiệm qua quá trình nghiên cứu và đề xuất hướng phát triển trong tương lai của đề tài.

3. Đóng góp của luận văn

Đóng góp chính của luận văn là xây dựng mô hình rút trích đặc trưng mới, rút trích đặc trưng trong vùng quan tâm của ảnh nhiệt trên khuôn mặt người.

4. Đối tượng và phạm vi nghiên cứu

Đối tượng: Trong đề tài này, dữ liệu được xử lý là ảnh nhiệt của mặt người.

Phạm vi nghiên cứu: Nhận dạng bảy cảm xúc cơ bản: buồn rầu, hạnh phúc, sợ hãi, giận dữ, ngạc nhiên, ghê tởm và bình thường.

5. Phương pháp nghiên cứu

Phương pháp nghiên cứu lý thuyết

- Tìm hiểu các công trình nghiên cứu liên quan.
- Tìm hiểu các phương pháp rút trích đặc trưng và phân loại.
- Tìm hiểu về bài toán ước lượng cảm xúc của người.
- Tìm hiểu về hình ảnh nhiệt và rút trích đặc trưng vùng quan tâm của ảnh nhiệt.

Phương pháp nghiên cứu thực nghiệm

- Tiến hành phân tích và cài đặt.
- So sánh và đánh giá kết quả đạt được.

6. Ý nghĩa khoa học và thực tiễn

Về mặt lý thuyết:

- Ứng dụng thành công công nghệ thị giác máy tính vào trong thực tế. Nhận dạng cảm xúc của người dựa trên ảnh nhiệt cho kết quả chính xác cao trong những điều kiện thu nhận thay đổi về tỉ lệ, ánh sáng, đặc biệt là với dữ liệu khuôn mặt của người.

- Tạo tiền đề cho các nghiên cứu dùng thị giác máy tính và tâm lý học tiếp theo trong tương lai.

Về mặt thực tiễn:

- Cung cấp các thuật toán hỗ trợ cho điều khiển robot, các thiết bị phát hiện và nhận dạng cảm xúc khuôn mặt người, các hệ thống tư vấn, đánh giá độ hài lòng của khách hàng thông qua nét mặt, các ứng dụng trong tương tác

người - máy, khoa học máy tính, khoa học hành vi, tâm lý học và khoa học y khoa, v.v.

7. Cấu trúc luận văn

Nội dung luận văn được xây dựng gồm các phần sau:

Chương mở đầu

Chương này giới thiệu tổng quan về đề tài gồm các nội dung như: lý do chọn đề tài, mục tiêu và nhiệm vụ nghiên cứu, đối tượng và phạm vi nghiên cứu, phương pháp nghiên cứu, ý nghĩa khoa học và thực tiễn cũng như cấu trúc chung của đề tài.

Chương 1. Cơ sở lý thuyết và một số nghiên cứu liên quan

Chương này giới thiệu tổng quan về tình hình nghiên cứu và các lý thuyết cơ bản về nhận dạng cảm xúc, tìm hiểu về các đặc trưng của ảnh nhiệt, cơ sở dữ liệu ảnh nhiệt, phương pháp rút trích đặc trưng, phân loại, v.v. Những kiến thức cơ bản này là tiền đề để áp dụng vào bài toán nhận dạng cảm xúc của người dựa trên ảnh nhiệt.

Chương 2. Nhận dạng cảm xúc của người dựa trên ảnh nhiệt

Chương này xây dựng và giải quyết bài toán nhận dạng cảm xúc sử dụng các lý thuyết cơ bản đã nêu ở chương một.

Chương 3. Thực nghiệm và đánh giá

Chương này phân tích về ưu điểm, nhược điểm, so sánh và đánh giá kết quả đạt được khi thực hiện chương trình.

Chương 4. Kết luận và hướng phát triển

Chương này tổng kết lại những gì đã đạt được và chưa đạt được sau khi nghiên cứu và tiến hành thực nghiệm. Từ đó nêu lên những hướng nghiên cứu và phát triển tiếp theo trong tương lai.

Chương 1. CƠ SỞ LÝ THUYẾT VÀ CÁC NGHIÊN CỨU LIÊN QUAN

1.1. Tổng quan tình hình nghiên cứu

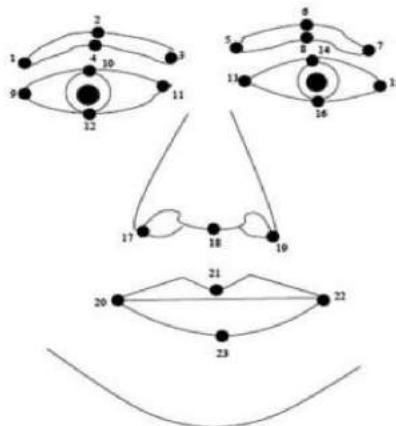
1.1.1. Các công trình nghiên cứu liên quan

Cuộc cách mạng công nghiệp lần thứ tư (hay được gọi với cái tên cuộc cách mạng công nghiệp 4.0) được biết đến trong thời gian gần đây đã tác động mạnh mẽ trên nhiều lĩnh vực, với sự xuất hiện của robot trí tuệ nhân tạo mang lại nhiều ứng dụng trong xã hội [10], [11], [12], [13]. Để sự tương tác giữa người và máy có hiệu quả, chính xác và thiết thực hơn, hệ thống nhận dạng cảm xúc trở thành một vấn đề đã và đang thu hút được sự quan tâm, nghiên cứu của các nhà khoa học. Một số tín hiệu sinh học đã được nghiên cứu để dự đoán cảm xúc của người như: ECG [14], thể tích máu [15], EEG [16], v.v. Trong nghiên cứu của Mehrabian [6] chỉ ra rằng trong các tín hiệu truyền đạt thông tin thì lời nói chỉ đóng góp 7%, ngữ điệu đóng góp 38%, trong khi biểu hiện trên khuôn mặt đóng góp 55% hiệu quả. Điều này cho thấy biểu hiện trên khuôn mặt tạo thành phương thức chính trong giao tiếp và cảm xúc của con người. Khảo sát [7], các tác giả đã cho chúng ta một cách nhìn tổng quát về lịch sử, xu hướng và một số ứng dụng trong nhận dạng biểu hiện khuôn mặt.

Hầu hết các phương pháp tiếp cận bài toán nhận dạng cảm xúc đều chuyển đổi dữ liệu đến các không gian khác có mô tả tốt hơn và dễ dàng phân loại cảm xúc hơn. Trích xuất đặc trưng là cách tìm các đặc trưng phù hợp để thể hiện tốt nhất cảm xúc của con người. Dựa trên các đặc trưng này để thực hiện phân loại cảm xúc. Trích xuất đặc trưng là chìa khóa để trích xuất các thông tin hữu ích của khuôn mặt bằng cách giảm số chiều của không gian dữ liệu sao cho lượng thông tin sau khi trích xuất vẫn đảm bảo các đặc trưng của dữ liệu ban đầu, trích xuất đặc trưng tốt làm giảm chi phí và tăng độ chính xác

cho quá trình nhận dạng. Có thể phân thành hai phương pháp chính: Phương pháp dựa trên việc xuất hiện đặc trưng và phương pháp dựa trên đặc trưng hình học.

Phương pháp trích xuất đặc trưng dựa trên hình học, trích xuất thông tin đặc trưng bằng hình dạng, khoảng cách và vị trí của các thành phần trên khuôn mặt. Công việc đầu tiên trong lĩnh vực này là hệ thống mã hóa hành động mặt (FACS), được thiết kế bởi Ekman [17]. FACS mã hóa một biểu hiện khuôn mặt trong 44 đơn vị hành động trên khuôn mặt (AU). Ngoài ra các phương pháp dựa trên đặc trưng hình học còn sử dụng mô hình xuất hiện tích cực (AAM) hoặc các biến thể của nó [47]. Vị trí của các vùng trên khuôn mặt này sau đó được sử dụng theo nhiều cách khác nhau để trích xuất hình dạng các đặc điểm khuôn mặt và chuyển động các đặc điểm khuôn mặt khi cảm xúc thay đổi. Choi và cộng sự [18] sử dụng AAM để nhận biết biểu hiện khuôn mặt. Các thuật toán cải tiến AAM khác nhau được so sánh và đánh giá trong [19]. Đối với [20], [21] các tác giả cũng đã đưa ra một cách tiếp cận dựa trên đặc trưng hình học để lập mô hình, theo dõi và nhận dạng các biểu hiện khuôn mặt.



Hình 1.1. Ví dụ về đặc trưng hình học của mặt gồm lông mày, mắt, mũi, miệng [17]

Các đặc trưng dựa trên hình học mô tả hình dạng của khuôn mặt và các thành phần của khuôn mặt trong khi các đặc điểm dựa trên xuất hiện đặc trưng được tạo ra từ biểu hiện trên khuôn mặt. Các đặc trưng xuất hiện đã được sử dụng thành công để nhận dạng cảm xúc như: toán tử mô hình nhị phân cục bộ (LBP) [22], biểu đồ của gradient định hướng (HOG) [23], [24], cục bộ mô hình định hướng (LDP) [25], thông tin kết cấu dựa trên bộ lọc Gabor [26], phân tích thành phần chính (PCA) [27] và phân tích phân biệt tuyến tính (LDA) [28].

Nhìn chung các đặc trưng hình học có hiệu quả để mô tả trực quan biểu hiện khuôn mặt nhưng chúng nhạy cảm với nhiễu và có những khó khăn khi mô tả một số thay đổi tinh tế trên khuôn mặt như nếp nhăn hoặc thay đổi cấu trúc da. Sử dụng dựa trên việc xuất hiện đặc trưng ổn định hơn khi có nhiễu, cho phép phát hiện biểu hiện khuôn mặt hoàn chỉnh hơn, đặc biệt quan trọng đối với việc phát hiện các biểu hiện rất nhỏ.

Với những thành quả đã đạt được, cho đến nay có nhiều phương pháp khác nhau để trích xuất đặc trưng. Từ các phương pháp kinh điển rất phổ biến cho đến các kỹ thuật học tập sâu gần đây phát triển mạnh mẽ đã được sử dụng học các đặc trưng cùng với mô hình nhận dạng của nó như: CNN, DBM, v.v. Một số nghiên cứu và ứng dụng đã đạt độ chính xác hơn 80% trong phân tích biểu hiện trên khuôn mặt [4] để nhận dạng cảm xúc.

Tuy nhiên, với những hệ thống này thường nhận dạng trên ảnh nhìn thấy được. Điều đó dẫn đến kết quả không chính xác trong những trường hợp như: thiếu ánh sáng, ảnh bị mờ hoặc người không có biểu cảm trên khuôn mặt thì cũng khó để nhận dạng được cảm xúc thật. Đây là thách thức lớn đòi hỏi cần có những hướng nghiên cứu và tiếp cận mới để khắc phục những hạn chế của ảnh nhìn thấy được nhằm tăng tỉ lệ chính xác hơn trong nhận dạng cảm xúc.

Những tiến bộ gần đây trong ảnh hồng ngoại nhiệt đã giúp chúng ta có được tín hiệu sinh lý, nhiệt độ cơ thể, thông qua các phương tiện không xâm

nhập và không tiếp xúc [29]. Khi một cảm xúc xảy ra, nhiệt độ trên khuôn mặt xuất hiện do lưu lượng máu mà cơ thể phát ra qua các mạch máu ở dưới da [30], sự thay đổi này có thể đủ điều kiện và định lượng để có thể nghiên cứu cảm xúc thông qua ảnh nhiệt.

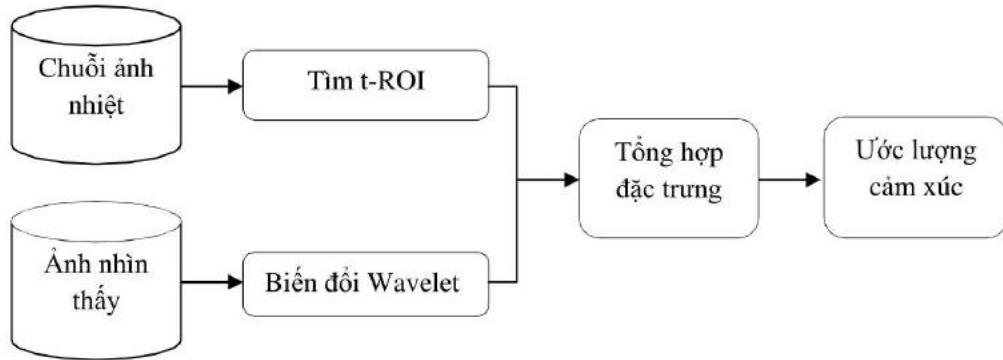
M.M.Khan và các cộng sự của ông [31] đề xuất sử dụng các điểm đặc trưng nhiệt mặt (FTFP) tỷ lệ chính xác dao động từ 66.3% đến 83.8% với năm loại cảm xúc.

Di Giacinto cùng cộng sự [32] khám phá cảm xúc sợ hãi ở những bệnh nhân bị rối loạn tâm lý sau chấn thương (PTSD). Nó đã được tìm thấy rằng nhiệt độ trên mặt ở bệnh nhân PTSD là thấp hơn so với những người được kiểm soát.

Trong [33] các tác giả phân loại các cảm xúc: ngạc nhiên, hạnh phúc, bình thường từ hai yếu tố đầu vào. Dữ liệu đầu vào đầu tiên bao gồm các lựa chọn của một tập hợp các vùng thích hợp khi thực hiện trích xuất đặc trưng, đầu vào thứ hai là ma trận Co-occurrence mức xám được sử dụng để tính toán các bộ mô tả vùng của ảnh nhiệt.

Liu và Wang [34] đã phân tích dữ liệu chuỗi nhiệt độ mặt, tính toán thống kê và các tính toán biểu đồ khác biệt về nhiệt độ. Hơn nữa, mô hình Hidden Markov (HMM) được sử dụng để phân biệt hạnh phúc, ghê tởm và sợ hãi với tỷ lệ công nhận tương ứng là: 68.11%, 57.14% và 52.30%. Họ đã sử dụng các mẫu từ cơ sở dữ liệu NVIE (biểu hiện trên khuôn mặt tự nhiên và ảnh nhiệt) để đánh giá kết quả.

Nguyen.H, Chen.F, Kotani.K và Le.B [35] đề xuất phương pháp kết hợp ảnh nhìn thấy và chuỗi ảnh nhiệt để ước lượng bảy cảm xúc cơ bản. Kết quả được đánh giá với hai phương pháp phân loại t-PCA và n-EMC trên tập cơ sở dữ liệu KTFE.

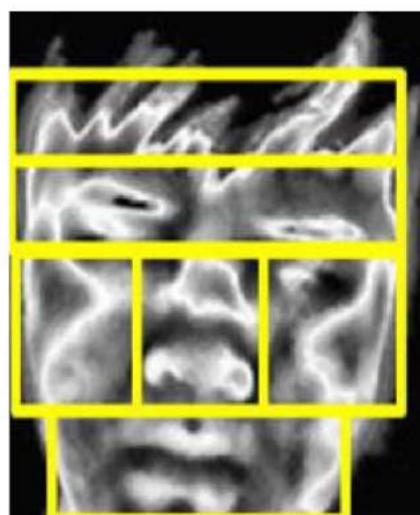


Hình 1.2. Sơ đồ quy trình thực hiện của phương pháp [35]

B.R.Nhan và T.Chau [36] nhận dạng cảm xúc dựa trên phân tích tần số thời gian từ chuỗi thời gian trung bình của năm vùng quan tâm trên mặt nhiệt.

Shan He và các cộng sự [37] đề xuất phương pháp nhận dạng cảm xúc với Deep Boltzmann Machine (DBM) trên cơ sở dữ liệu NVIE đạt tỷ lệ công nhận là 51.3% .

Trong bài báo [38], một phương pháp nhận dạng cảm xúc dựa trên những thời điểm bất biến của Hu và phân loại bằng cách sử dụng SVM nhiều lớp. Thuật toán được kiểm tra trên cơ sở dữ liệu KTFE công nhận năm cảm xúc cơ bản gồm: giận dữ, sợ hãi, hạnh phúc, buồn rầu và bình thường với độ chính xác 87,50%.



Hình 1.3. Trích xuất ROI trong phương pháp [38]

Tất cả các nghiên cứu cho thấy hình ảnh nhiệt không nhạy cảm với điều kiện ánh sáng, không chịu tác động bởi điều kiện môi trường xung quanh, có thể được sử dụng để phát hiện sự thay đổi nhiệt độ trên khuôn mặt. Bên cạnh đó, không phải toàn bộ ảnh có tác động thay đổi nhiệt độ khi cảm xúc thay đổi nên việc rút trích đặc trưng quan trọng là rất cần thiết. Như vậy sử dụng ảnh nhiệt là một kỹ thuật đầy hứa hẹn trong việc nhận dạng cảm xúc con người chính xác hơn và thực tế hơn.

1.1.2. Một số thách thức trong lĩnh vực nghiên cứu nhận dạng cảm xúc người

Có thể nói xác định cảm xúc của người là một bài toán khó, vì con người ngoài bảy cảm xúc cơ bản còn rất nhiều cảm xúc đa dạng khác đan xen lẫn nhau. Làm thế nào để nhận dạng cảm xúc chính xác của người trong thời gian thực vẫn còn là một thách thức rất lớn.

Sử dụng ảnh nhiệt để nhận dạng cảm xúc của người có thể khắc phục được những hạn chế của ảnh nhìn thấy được. Tuy nhiên hiện nay các cơ sở dữ liệu ảnh nhiệt chưa nhiều và chưa phổ biến. Điều này cho thấy cần có nhiều nghiên cứu sâu rộng hơn để nâng cao khả năng tích hợp các trạng thái sinh lý bên trong với sự nhận thức tự động các biểu hiện trên khuôn mặt bên ngoài để có được tập cơ sở dữ liệu đa dạng, phù hợp, hiệu quả trong nhận dạng cảm xúc thực tế hơn.

Ngoài ra, việc kết hợp các phương pháp khác nhau để nhận dạng cảm xúc giúp cải thiện tỷ lệ chính xác hơn cũng là một bài toán cần được nghiên cứu và phát triển mạnh mẽ trong tương lai.

1.2. Thông tin về ảnh nhiệt

1.2.1. Giới thiệu ảnh nhiệt

Tất cả các vật thể, cả tự nhiên lẫn nhân tạo, đều phát ra năng lượng nhiệt hồng ngoại. Bằng cách phát hiện sự khác biệt nhiệt độ rất tinh vi, công nghệ hình ảnh nhiệt cho thấy những gì mà mắt thường không nhìn thấy được.

Máy ảnh nhiệt thường phát hiện bức xạ trong dải hồng ngoại dài của phổ điện từ (khoảng 9.000–14.000 nm hoặc 9–14 μm) và tạo ra hình ảnh của bức xạ đó, được gọi là thermograms. Lượng bức xạ phát ra bởi một vật thể tăng theo nhiệt độ; do đó, phép đo nhiệt cho phép người ta thấy các biến đổi về nhiệt độ. Khi camera chụp ảnh nhiệt, các vật thể ám áp nổi bật so với các nền lạnh hơn, con người và các động vật máu nóng khác trở nên dễ dàng nhìn thấy ngay cả trong bóng tối hoàn toàn và điều kiện môi trường đầy thách thức.

Hình ảnh nhiệt không gây hại cho cơ thể con người. Do đó nó đã được đề xuất như là một nguồn thông tin thay thế để phát hiện, công nhận khuôn mặt và ước tính của cảm xúc.

So với số lượng cơ sở dữ liệu có thể nhìn thấy hiện tại phổ biến rất nhiều thì chỉ có rất ít cơ sở dữ liệu mặt nhiệt được phát triển để hỗ trợ nghiên cứu về biểu hiện và cảm xúc trên khuôn mặt người. IRIS [39] và NIST / Equinox [40], bao gồm các cặp ảnh được gắn nhãn với ba cảm xúc chính đặt ra dưới nhiều ánh sáng và tư thế đầu khác nhau. Cơ sở dữ liệu Natural Visible and Infrared facial Expression (NVIE) chứa 215 đối tượng, mỗi đối tượng hiển thị bảy cảm xúc, là một cơ sở dữ liệu rất tốt thích ứng với dữ liệu nhiệt và tự phát [41]. Các cảm xúc tự phát được kích hoạt thông qua phương tiện nghe nhìn, nhưng không phải tất cả chúng đều có mặt cho mỗi chủ thể. Trong các đối tượng bộ dữ liệu Kotani Thermal Facial Emotion (KTFE) hiển thị các chuyển động được đặt ra và tự phát, cũng được kích hoạt thông qua phương tiện nghe nhìn [42].

Bảng 1.1. Một số cơ sở dữ liệu ảnh nhiệt [42]

Cơ sở dữ liệu	Số lượng đối tượng	Dải sóng camera nhiệt	Loại cảm xúc
NIST/Equinox	600	8-12μm, 3-5μm	Mỉm cười, cau mày và ngạc nhiên.
IRIS	30	7-14μm	Ngạc nhiên, cười và giận dữ.
NVIE	215	8-14μm	Hạnh phúc, giận dữ, bình thường, ghê tởm, sợ hãi, buồn rầu và ngạc nhiên.
KTFE	30	8-14μm	Hạnh phúc, giận dữ, bình thường, ghê tởm, sợ hãi, buồn rầu và ngạc nhiên.

1.2.2. Ưu điểm

- Thiết bị hồng ngoại hiện đại cho phép đo nhiệt độ không xâm lấn. Điều này có thể đặc biệt hữu ích cho các ứng dụng chẩn đoán y khoa và tâm lý theo điều kiện khi bệnh nhân không thể hoặc không muốn hợp tác.
- Hình ảnh nhiệt là bất biến đối với điều kiện ánh sáng và chiều sáng.
- Các máy ảnh hồng ngoại mới nhất có độ nhạy cao với bất kỳ biến đổi nhiệt nào trên da người. Những máy ảnh này có khả năng cảm biến lên đến ± 0,05°C biến thể nhiệt.
- Hình ảnh nhiệt an toàn, không gây hại cho người, có thể cung cấp cả thông tin hình ảnh và sinh lý cho hệ thống nhận dạng cảm xúc con người.

1.2.3. Hạn chế

Mặc dù có nhiều ưu điểm đáng kể khi sử dụng hình ảnh nhiệt, nhưng nó có một số hạn chế:

- Dữ liệu nhiệt có thể thay đổi cùng với nhiệt độ cơ thể gây ra bởi nhiệt độ môi trường biến đổi.
- Sự hiện diện của kính đeo mắt có thể dẫn đến mất thông tin hữu ích xung quanh mắt. Kính mờ đục với ảnh nhiệt, hoàn toàn che khuất các phần nằm phía sau chúng. Do đó, độ nhạy của hình ảnh nhiệt bị giảm bởi các vật trên khuôn mặt.
- Có một số vùng mặt không tiếp nhận những thay đổi về cảm xúc.

1.3. Một số phương pháp rút trích đặc trưng sử dụng trong nhận dạng cảm xúc

1.3.1. Active Appearance Model (AAM)

Ý tưởng chính của thuật toán AAM [47] trích chọn đặc trưng biểu cảm khuôn mặt là mô hình hóa đối tượng bởi một mô hình hình dạng và một mô hình kết cấu bì mặt ảnh bao quanh tập điểm mô hình hình dạng tương ứng. Thuật toán này học mối tương quan giữa độ điều chỉnh vị trí hình dạng thông qua các tham số mô hình với độ sai khác của ảnh mới và ảnh khuôn mặt tạo ra được từ mô hình.

Thuật toán AAM cơ bản có hai quy trình: mô hình hóa khuôn mặt và tối ưu hóa thuật toán để tìm mô hình thể hiện hình dạng tốt nhất cho một ảnh khuôn mặt đầu vào mới dựa trên mô hình đã xây dựng được.

- Mô hình hóa khuôn mặt:

Đầu tiên, hình dạng được lập mô hình dựa trên các mốc được gắn nhãn. Trong ảnh hai chiều, ta có thể mô tả n điểm điều khiển $\{(x_i, y_i)\}$ thành vector $2n$ thành phần $h = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)^T$



Hình 1.4. Ví dụ về AAM [47]

Để chuẩn hóa các đối tượng hình dạng này vào chung một không gian tọa độ, các tác giả sử dụng phương pháp Procrustes Analysis và chiếu lên không gian con được tạo bởi PCA.

$$h = \bar{h} + p_s b_s \quad (1.1)$$

- + \bar{h} là hình dạng trung bình.
- + p_s là tập vector không gian con xây dựng từ PCA.
- + b_s là tập các tham số hình dạng trong mô hình.

Tiếp theo, mô hình kết cấu được tạo ra:

$$g = \frac{(g_i - \beta \cdot 1)}{\alpha} \quad (1.2)$$

- + α và β là giá trị trung bình và phương sai của kết cấu g .
- + $1 = [1, 1, \dots, 1]^T$ là các vector có chiều dài giống nhau của g_i

Tương tự, ta thực hiện mô hình hóa các dữ liệu kết cấu hình ảnh đã được chuẩn hóa với PCA

$$g = \bar{g} + p_g b_g \quad (1.3)$$

- + \bar{g} biểu thị kết cấu trung bình.
- + p_g là cơ sở không gian con được xây dựng.
- + b_g là hình chiếu tương ứng của mẫu kết cấu được biểu diễn.

Cuối cùng, kết hợp giữa hình dạng và kết cấu được phân tích bởi PCA và không gian con xuất hiện:

$$h = \bar{h} + Q_s \cdot c \quad (1.4)$$

$$g = \bar{g} + Q_g \cdot c \quad (1.5)$$

+ c là một vector của các tham số xuất hiện kiểm soát cả hình dạng và kết cấu.

+ Q_s và Q_g là các ma trận mô tả các phương thức biến đổi bắt nguồn từ tập huấn luyện.

- Tối ưu hóa thuật toán:

Tính ma trận R .

Khởi tạo giá trị h .

Khởi tạo tham số $K=\{1, 0.5, 0.25, 0.125, 0.0625\}$.

Thực hiện lặp:

+ Tính vector độ lệch r .

+ Tính $E = r^T r$.

+ Tính $\partial p = Rr(p)$.

+ Với mỗi $k \in K$, cập nhật $p' = p + k^* \partial p$.

+ Tính r' .

+ Tính $E' = r'^T r'$.

+ Nếu $E' < E$, cập nhật $p = p'$, cập nhật độ thay đổi giá trị lỗi hiện tại và sang bước lặp tiếp, ngược lại thử với các k khác.

Thuật toán AAM bắt biến với các phép biến đổi tương tự đó là phép quay, phép dịch chuyển và phép biến đổi tỷ lệ. Nó thể hiện được các biến thể về hình dạng của khuôn mặt. Khi khuôn mặt bị thay đổi bởi các phép biến đổi tuyến tính tương tự đó thì các đặc trưng trích chọn được cũng không thay đổi.

1.3.2. Local Binary Pattern (LBP)

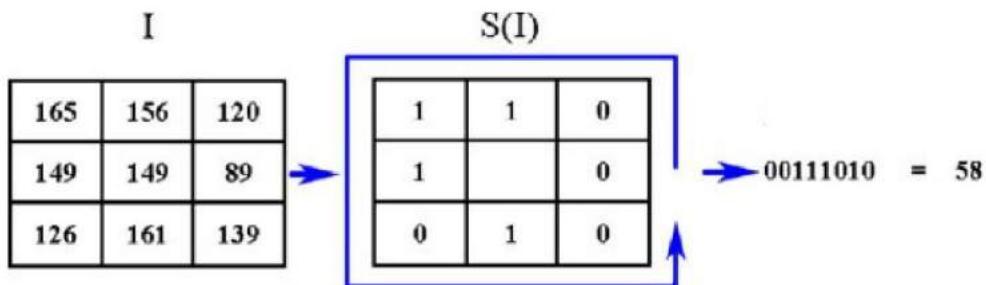
LBP đã được chứng minh là thành công như một kết cấu mô tả trong nhiều ứng dụng thị giác máy tính. Các thuộc tính quan trọng nhất của LBP là khả năng chống lại sự thay đổi ánh sáng và sự tính toán đơn giản. Các toán tử

gán các pixel f_p ($P = 0,..,7$) của một hình ảnh bằng cách làm nổi bật vùng lân cận 3×3 của mỗi pixel với giá trị của pixel trung tâm f_c và xem xét kết quả dưới dạng số nhị phân $S(f_p - f_c)$.

$$S(f_p - f_c) = \begin{cases} 1 & \text{khi } f_p \geq f_c \\ 0 & \text{khi } f_p < f_c \end{cases} \quad (1.6)$$

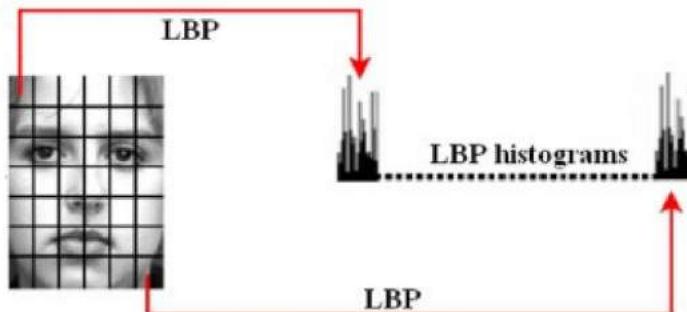
Sau đó, bằng cách gán một hệ số nhị thức 2^P cho mỗi $S(f_p - f_c)$, LBP được tính như sau:

$$LBP = \sum_{p=0}^7 S(f_p - f_c) 2^p \quad (1.7)$$



Hình 1.5. Ví dụ về LBP cơ bản. Lấy giá trị điểm trung tâm làm ngưỡng xác định các giá trị điểm xung quanh chuyển về 0,1. Từ giá trị nhị phân nhận được tính toán giá trị thập phân tương ứng [22]

Toán tử LBP phát hiện nhiều nguyên tố kết cấu khác nhau (tại chỗ, đầu dòng, cạnh, góc), thường được tích lũy thành một biểu đồ trên một vùng để thu thập thông tin kết cấu cục bộ.



Hình 1.6. Các vector đặc trưng LBP được trích xuất từ các vùng khác nhau trên hình ảnh khuôn mặt [22]

1.4. Một số phương pháp sử dụng phân loại cảm xúc

1.4.1. Linear Discriminant Analysis (LDA)

Thuật toán LDA [43] dựa trên phân tích phân loại phi tuyến của Fisher là phương pháp tính toán chuyển đổi tối đa hóa sự phân tán giữa các lớp trong khi giảm thiểu phân tán trong lớp. LDA cho phép đánh giá khách quan về tầm quan trọng của thông tin hình ảnh trong các đặc điểm khác nhau của khuôn mặt để xác định khuôn mặt người. LDA cũng cho chúng ta một tập hợp nhỏ các đặc trưng mà thực hiện phù hợp nhất thông tin cho các mục đích phân loại.

Mục tiêu của LDA là thực hiện giảm chiều trong khi vẫn giữ càng nhiều các thông tin phân biệt giữa các lớp càng tốt. Thực hiện bằng cách xem xét các phân tán trong các lớp và giữa các lớp. Nó có khả năng thay đổi hình ảnh từ những biến đổi do các nguồn khác nhau.

Cho tập dữ liệu gồm các lớp ảnh i thuộc bộ dữ liệu C . Ta có y là số mẫu có trong tập C . Cho M_i là số lượng mẫu trong mỗi lớp i ; $i=1, 2, \dots, C$. Cho μ_i là vector trung bình trong các lớp i ; $i=1, 2, \dots, C$.

$$\mu = \frac{1}{C} \sum_{y=1}^C \mu_y \quad (1.8)$$

trong đó, μ : trung bình của toàn tập dữ liệu C .

Gọi S_w là ma trận phân tán bên trong lớp (các phần tử trong lớp) và S_B là ma trận phân tán giữa các lớp thuộc C .

Ta có:

$$S_w = \sum_{i=1}^C \sum_{j=1}^{M_i} (y_j - \mu_i)(y_j - \mu_i)^T \quad (1.9)$$

$$S_B = \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T \quad (1.10)$$

Phương pháp LDA sẽ tìm giá trị W để cực đại hóa hàm mục tiêu $H_{(W)}$:

$$H_{(W)} = \frac{W S_B W^T}{W S_w W^T} \quad (1.11)$$

1.4.2. Support Vector Machines (SVM)

Ý tưởng chính của thuật toán này là ánh xạ (tuyến tính hoặc phi tuyến) dữ liệu vào không gian các vector đặc trưng, ở đó một siêu phẳng tối ưu được tìm ra để tách dữ liệu thuộc hai lớp khác nhau. Mục đích của phương pháp SVM là tìm được khoảng cách biên lớn nhất. Dựa trên một kết hợp có các trọng số của một tập con nhỏ các vector huấn luyện, các vector này được gọi là support vector. Ước lượng siêu phẳng thì tương đương giải một bài toán tuyến tính bậc hai.

SVM có hai đặc trưng cơ bản:

- Kết hợp với các dữ liệu có ý nghĩa về mặt vật lý, do vậy dễ dàng giải thích được một cách tường minh.
- Cần một tập các mẫu huấn luyện rất nhỏ.

Theo [44], hàm mục tiêu của một máy phân lớp SVM nhị phân có thể được phát biểu như sau:

$$g(x) = w \cdot \Phi(x) + b \quad (1.12)$$

Trong đó:

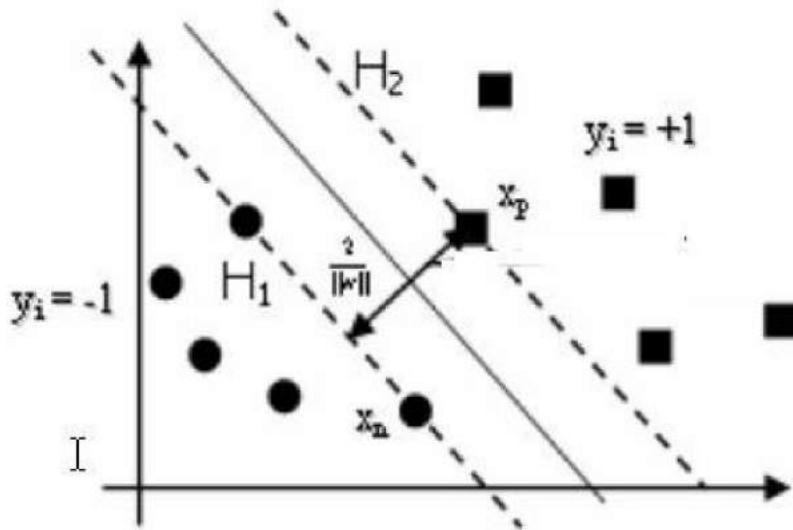
$$x \in \mathbb{R}^D$$

- w là vector chuẩn của siêu phẳng phân cách trong không gian đặc trưng được sinh ra từ ánh xạ $\Phi(x): \mathbb{R}^D \rightarrow \mathbb{R}^F (F > D)$, $\Phi(x)$ có thể tuyến tính hoặc phi tuyến.

- b là độ lệch so với gốc tọa độ.

SVM được thiết kế cho bài toán phân lớp nhị phân, do đó dấu của $g(x)$ cho biết vector x thuộc lớp +1 hay lớp -1.

Cho tập mẫu $\{(x_1, y_1), \dots, (x_N, y_N)\}$ trong đó $x_i \in \mathbb{R}^D$ và $y_i \in \{\pm 1\}$, mục tiêu của phương pháp phân lớp SVM là tìm một siêu phẳng phân cách sao cho khoảng cách biên giữa hai lớp là lớn nhất.



Hình 1.7. Siêu phẳng tách với khoảng cách lề cực đại [44]

Khoảng cách lề giữa hai lớp là: $\frac{2}{\|w\|^2}$

Việc huấn luyện SVM chính là giải bài toán tối ưu có ràng buộc:

$$\min_{w, \xi, b} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \right\} \quad (1.13)$$

sao cho $y_i(w \cdot \Phi(x_i) + b) \geq 1 - \xi_i ; \xi_i \geq 0 \forall i$, C là tham số nói lỏng điều kiện phân lớp.

Bằng cách thêm các nhân tử Lagrange α , bài toán trở thành:

$$\min_{w, \xi, b} \max_{\alpha, \beta} \left\{ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i (w \cdot \Phi(x_i) + b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i \xi_i \right\} \quad (1.14)$$

Giải bài toán đối ngẫu của [1.14]:

- Cực đại hóa theo α_i :

$$\tilde{L}(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \Phi(x_i) \Phi(x_j) \quad (1.15)$$

thỏa mãn: $0 \leq \alpha_i \leq C, \forall i$ và $\sum_{i=1}^n \alpha_i y_i = 0$

- Giải phương trình ta được:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \Phi(x_i) \quad (1.16)$$

Mỗi mẫu huấn luyện x_i tương ứng với một hệ số Lagrange α_i . Sau khi huấn luyện, các mẫu có $\alpha_i > 0$ được gọi là vector hỗ trợ.

Thay (1.16) vào (1.12), ta có :

$$g(x) = \sum_{i=1}^N \alpha_i y_i \Phi(x_i) \cdot \Phi(x) + b \quad (1.17)$$

Giả sử $\Phi(x_i) \cdot \Phi(x_j) = K(x_i, x_j)$. Nghĩa là, tích vô hướng trong không gian đặc trưng tương đương với một hàm nhân K của không gian đầu vào. Vì vậy, ta không cần phải tính trực tiếp các giá trị $\Phi(x_i), \Phi(x_j)$ mà chỉ cần tính tích vô hướng $\langle \Phi(x_i) \cdot \Phi(x_j) \rangle$ gián tiếp thông qua hàm nhân $K(x_i, x_j)$. Nếu sử dụng hàm nhân là hàm tuyến tính thì SVM được gọi là SVM tuyến tính, ngược lại thì gọi là SVM phi tuyến.

Như vậy, hàm mục tiêu cho bài toán phân lớp SVM nhị phân có dạng:

$$g(x) = \sum_{i=1}^N \alpha_i y_i K(x_i, x) + b \quad (1.18)$$

Chỉ có các vector hỗ trợ tham gia vào việc xây dựng siêu phẳng phân cách, do đó một mẫu x sẽ được phân lớp theo hàm quyết định:

$$f(x) = \operatorname{sgn} \left(\sum_{i=1}^{N_{SV}} \alpha_i y_i K(x_i, x) + b \right) \quad (1.19)$$

trong đó, N_{SV} là số vector hỗ trợ.

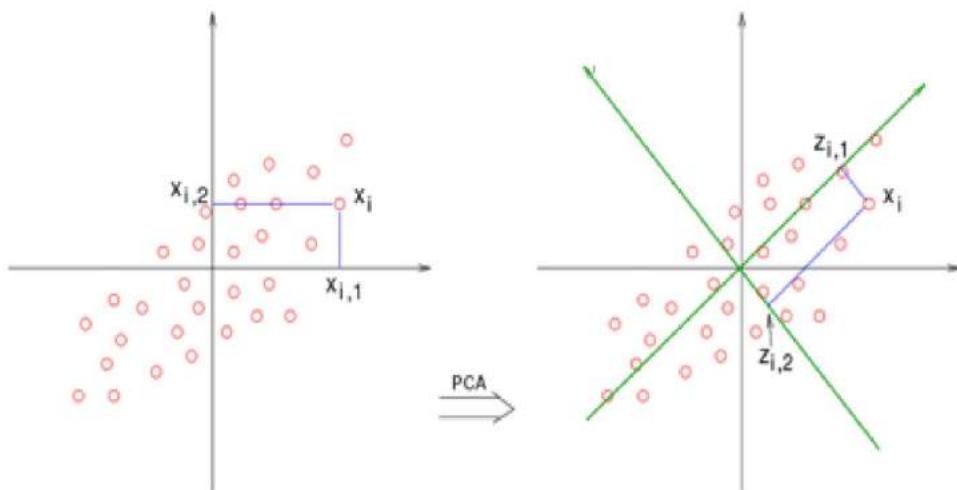
1.4.3. Principal Component Analysis (PCA)

Với PCA [45], mục tiêu là xây dựng một không gian khuôn mặt, bao gồm các vector cơ bản được gọi là các thành phần chính, mô tả tốt hơn hình ảnh khuôn mặt gốc. PCA có lợi thế hơn các thuật toán khác về tốc độ và tính đơn giản của nó. Khi các mẫu trong dữ liệu rất khó nhận ra trong không gian nhiều chiều thì PCA là một công cụ mạnh để phân tích chúng.

PCA tìm được các đặc tính tiêu biểu của đối tượng cần nhận dạng mà không cần phải xác định các thành phần và mối quan hệ giữa các thành phần

đó. Thuật toán có thể thực hiện tốt với các ảnh có độ phân giải cao, do PCA sẽ thu gọn ảnh thành một ảnh có kích thước nhỏ hơn. PCA có thể kết hợp với các phương pháp khác để mang lại hiệu quả nhận dạng cao hơn như: mạng Nơron, SVM, v.v.

PCA phân loại theo chiều phân bố lớn nhất của tập vector. Tuy nhiên, chiều phân bố lớn nhất không phải lúc nào cũng mang lại hiệu quả tốt nhất cho bài toán nhận dạng. Đây là nhược điểm cơ bản của PCA. PCA rất nhạy với vấn đề nhiễu.



Hình 1.8. Ví dụ về PCA [45]

Thuật toán PCA được thực hiện như sau:

Giả sử ta có một tập dữ liệu đào tạo gồm M hình ảnh biểu hiện trên khuôn mặt :

$$F = \{\Gamma_1, \Gamma_2, \dots, \Gamma_M\} \quad (1.20)$$

Tính trung bình dữ liệu đào tạo :

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \quad (1.21)$$

Tính sai lệch của dữ liệu đầu vào so với trung bình:

$$\Phi_i = \Gamma_i - \Psi \quad (1.22)$$

Tính ma trận hiệp phương sai:

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = AA^T; A = \{\Phi_1, \Phi_2, \dots, \Phi_M\} \quad (1.23)$$

Vì ma trận C có kích thước lớn (NxN) nên để tìm vector riêng u_i của C ta tìm vector riêng và trị riêng của ma trận C' :

$$C' = A^T A \quad (1.24)$$

$$C'_{m,n} = \Phi_m^T \Phi_n \quad (1.25)$$

Gọi v_i là vector riêng của C' thì vector riêng của C là :

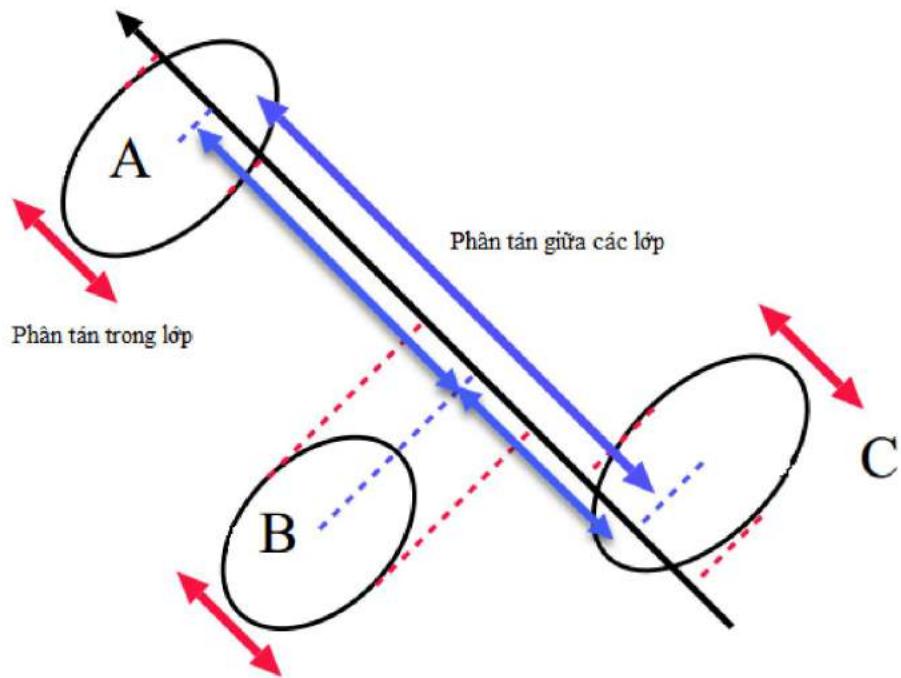
$$u_i = \sum_{k=1}^M v_{ik} \Phi_k; i = 1..M \quad (1.26)$$

Sau khi tìm được các vector riêng và trị riêng, các ảnh trong tập cơ sở dữ liệu sẽ được chiếu lên không gian các vector riêng này để tạo ra vector đặc trưng. Kích thước của vector đặc trưng nhỏ hơn nhiều nhưng vẫn mang nhiều nhất thông tin so với hình ảnh gốc.

1.4.4. Eigenspace Method based on Class features (EMC)

Trong [46], các tác giả đề xuất phương pháp không gian riêng dựa trên các đặc trưng của lớp (EMC) để phân tích các biểu thức khuôn mặt. Chọn một phép biến đổi các vector đặc trưng để tối đa hóa sự khác biệt các phương sai giữa các lớp và các phương sai trong lớp.

EMC cho tỷ lệ phân loại chính xác cao trong trường hợp sử dụng số lượng nhỏ các vector riêng. Kết quả này là do sự mất cân bằng phương sai trong lớp của mỗi chiều.



Hình 1.9. Ví dụ về EMC [46]

Gọi F là tập hợp các lớp biểu thức, khuôn mặt M_f được đưa ra cho mỗi lớp $f \in F$ và x_{fm} là một vector N -chiều, $m = \overline{1, M_f}$, z_{fm} là giá trị của x_{fm} trên một dòng.

Xấp xỉ x_{fm} được tính như sau:

$$\tilde{x}_{fm} = \sum_{k=1}^K z_{kfm} \Phi_k + \bar{x} \quad (1.27)$$

Trong đó, $\Phi_k = (\Phi_{1k}, \Phi_{2k}, \dots, \Phi_{Nk})'$, $k = \overline{1, K}$ và tất cả các Φ_k đều trực giao nhau.

Tính trung bình mặt theo công thức:

$$\bar{x} = \frac{1}{M} \sum_{f \in F} \sum_{m=1}^{M_f} x_{fm}, \text{ với } M = \sum_{f \in F} M_f \quad (1.28)$$

Tính Φ_k để trung bình lỗi nhỏ nhất

$$\varepsilon^2(\Phi_k) = \frac{1}{M} \sum_{f \in F} \sum_{m=1}^{M_f} \left\| \bar{x}_f - \tilde{x}_{fm} \right\|^2 \quad (1.29)$$

Tính z_{kfm} theo công thức:

$$z_{kfm} = \left(x_{fm} - \bar{x}_f \right)^t \Phi_k \quad (1.30)$$

Gọi S_{Bk} và S_{Wk} là phương sai giữa các lớp và phương sai bên trong lớp của Z_{kfm} cho tất cả $f \in F$ tương ứng. Mỗi trị riêng λ_k cho thấy sự khác biệt giữa S_{Bk} và S_{Wk} .

$$\begin{aligned} \lambda_k &= S_{Bk} - S_{Wk} \\ &= \frac{1}{M} \sum_{f \in F} M_f \left(\bar{x}_f - \bar{x} \right) \left(\bar{x}_f - \bar{x} \right)^T - \frac{1}{M} \sum_{f \in F} \sum_{f' \in F} M_f \left(\bar{x}_{fm} - \bar{x}_f \right) \left(\bar{x}_{fm} - \bar{x}_f \right)^T \end{aligned} \quad (1.31)$$

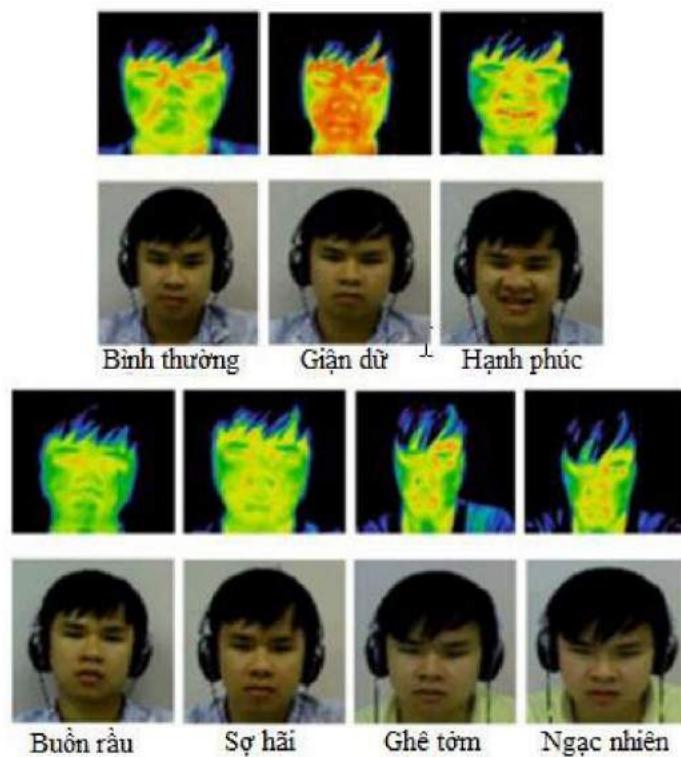
Nếu sự khác biệt được coi là mức độ tách biệt của mỗi $f \in F$, giá trị của λ_k càng lớn thì càng có nhiều Φ_k tương ứng cho các vector đặc trưng ban đầu.

Chương 2. NHẬN DẠNG CẢM XÚC CỦA NGƯỜI DỰA TRÊN ẢNH NHIỆT

2.1. Cơ sở dữ liệu ảnh nhiệt Kotani Thermal Facial Emotion (KTFE)

Hiện nay, có nhiều cơ sở dữ liệu ảnh nhìn thấy rất được phổ biến rộng rãi. Trong khi đó, chỉ có rất ít cơ sở dữ liệu ảnh nhiệt được khai thác và sử dụng. Ngoài ra, có một số cơ sở dữ liệu ảnh nhiệt gặp vấn đề khi thiết kế trong thu thập dữ liệu, đó là hiện tượng trễ thời gian. Do đó, chúng tôi đề xuất sử dụng cơ sở dữ liệu cảm xúc và khuôn mặt nhiệt Kotani Thermal Facial Emotion (KTFE) [42] để nghiên cứu phân tích biểu hiện khuôn mặt trở nên thực tế hơn.

KTFE chứa bảy cảm xúc tự phát của 30 đối tượng, bao gồm 130 GB các video cảm xúc trên khuôn mặt có thể nhìn thấy và nhiệt, cơ sở dữ liệu hình ảnh biểu hiện khuôn mặt và cơ sở dữ liệu hình ảnh khuôn mặt nhiệt.



Hình 2.1. Mẫu ảnh nhiệt và ảnh nhìn thấy của bảy cảm xúc [42]

KTFE có một số ưu điểm:

- Thứ nhất, đây là một trong những video có thể nhìn thấy và nhiệt tự nhiên đầu tiên. Những cơ sở dữ liệu này sẽ cho phép các nhà nghiên cứu biểu hiện trên khuôn mặt và cảm xúc để có nhiều cách tiếp cận thực tế hơn.
- Thứ hai, cơ sở dữ liệu này đã khắc phục lỗi trễ thời gian mà cơ sở dữ liệu cũ gặp khi thực hiện các thiết lập thử nghiệm.
- Thứ ba, đã có một số nghiên cứu trong dữ liệu KTFE và thu được một số kết quả để hỗ trợ các nhà nghiên cứu sử dụng cơ sở dữ liệu này.

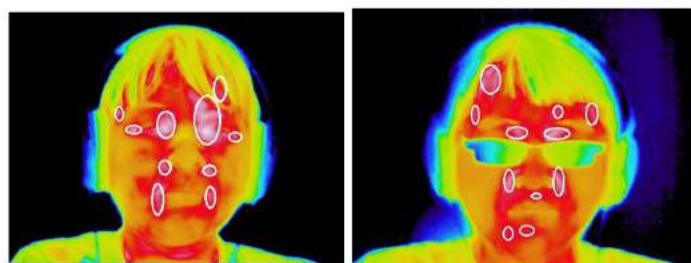
Các kết quả về dữ liệu nhiệt cho chúng ta một tương lai đầy hứa hẹn về nghiên cứu trên khuôn mặt tốt hơn.

2.2. Phương pháp nhận dạng

Trong phần này, chúng tôi đề xuất một phương pháp trích xuất các vùng quan tâm (ROI) trên dữ liệu ảnh nhiệt. Sau đó, chúng tôi tiến hành công nhận cảm xúc với các phương pháp kinh điển trong lĩnh vực thị giác máy tính bao gồm: PCA, EMC và kết hợp PCA-EMC trên cơ sở dữ liệu KTFE.

2.2.1. Tìm các vùng nhiệt quan tâm (ROI)

Việc xác định ROI là rất hữu ích và cần thiết. Nó xác định các khu vực nơi nhiệt độ tăng hoặc giảm đáng kể khi cảm xúc thay đổi và tập trung ở trán, xung quanh mắt, gò má và hàm trên. Đây là một yếu tố quan trọng để làm cho việc xác định các cảm xúc chính xác hơn. Quan trọng hơn, nó giúp chúng ta hiểu rõ về bản chất của mối quan hệ giữa nhiệt độ và cảm xúc.



Hình 2.2. Ví dụ về ROI [51]

Từ ma trận nhiệt ban đầu, sau khi tiền xử lý, ROI được xác định như sau:

Gọi h, g, f là các hàm ánh xạ được xác định:

$$\begin{aligned} h: F \rightarrow Tc & ; \quad g: F \rightarrow Gr \quad ; \quad f: Gr \rightarrow Ic \\ (i, j) \mapsto h(i, j) & \quad (i, j) \mapsto g(i, j) \quad (i, j) \mapsto f(i, j) \end{aligned} \quad (2.1)$$

Trong đó F là không gian mặt; Tc là không gian nhiệt độ; Gr là không gian mặt trong thang đo xám; Ic là cường độ của pixel trong thang đo xám.

$$\Delta T_F = T_{Max}^F - T_{Min}^F \quad (2.2)$$

$$\delta T_F = \Delta T_F / T_{mgr}^{Ic} \quad (2.3)$$

$$T_{mgr}^{Ic} = \max(f(i, j)) \quad (2.4)$$

$$L_k^F = \left\{ (i, j) \in F \mid T_{Min}^F + \delta T_F * g(i, j) \leq h(i, j) \leq T_{Max}^F - \delta T_F * g(i, j) \right\} \quad (2.5)$$

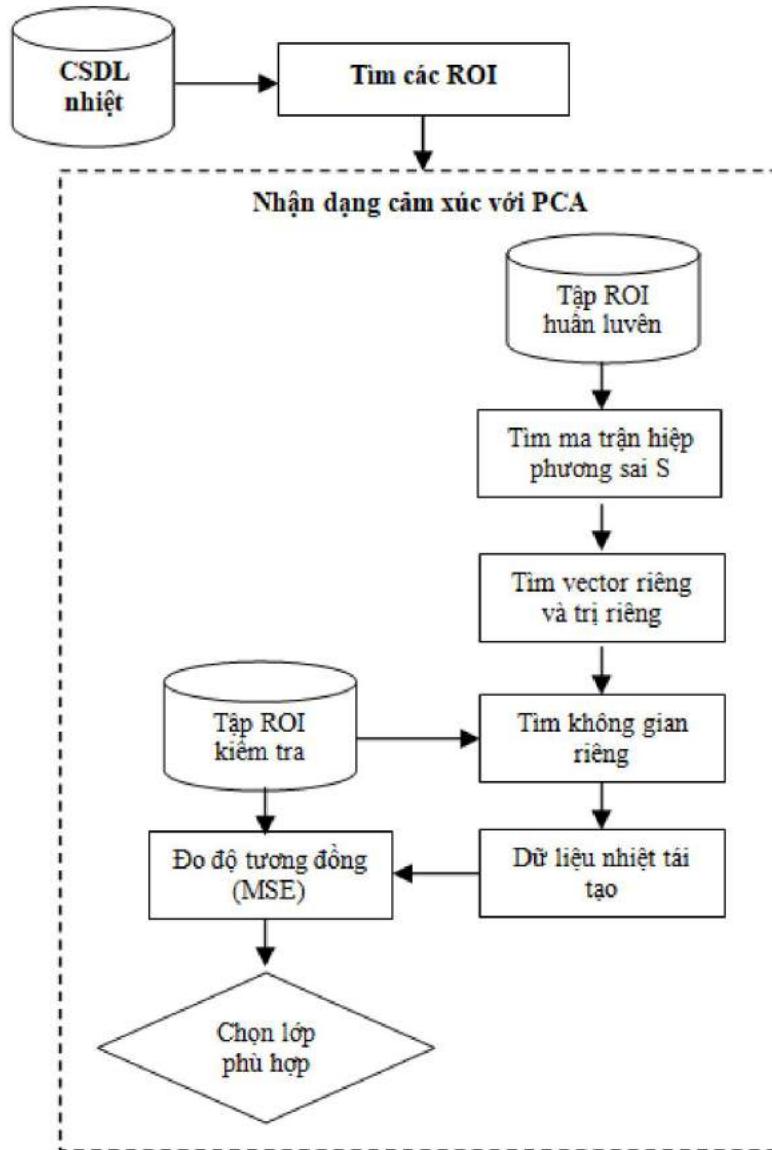
$$k \in \overline{1, K}$$

Với T_{Max}^F, T_{Min}^F là nhiệt độ lớn nhất và nhỏ nhất của mỗi mặt trong khung k ; T_{mgr}^{Ic} là cường độ cao nhất trong thang đo xám, K là số khung.

Sau khi tìm thấy ROI, chúng tôi sử dụng PCA, EMC và PCA-EMC để nhận dạng bảy cảm xúc cơ bản.

2.2.2. Nhận dạng cảm xúc

⊕ Sử dụng PCA



Hình 2.3. Nhận dạng cảm xúc với ROI và PCA

Để nhận dạng cảm xúc bằng cách sử dụng PCA [45], chúng tôi chia tập huấn thành bảy lớp và tính toán không gian riêng của mỗi lớp như sau:

Bước 1: Mỗi khung dữ liệu nhiệt được xác định từ ma trận hai chiều tới vector một chiều. Với khung dữ liệu nhiệt M_f như dữ liệu huấn luyện, chúng tôi chuyển đổi các dữ liệu này thành các vector cột tương ứng.

Bước 2: Tìm ma trận hiệp phương sai đại diện cho sự phân tán trên trung bình của dữ liệu đào tạo.

$$S = \frac{1}{M_f} \sum_{k=1}^{M_f} (x_k - \bar{x})(x_k - \bar{x})^T \quad (2.6)$$

$$\bar{x} = \frac{1}{M_f} \sum_{k=1}^{M_f} x_k \quad (2.7)$$

Với x_k là một vector N chiều, M_f là tổng số khung của mỗi lớp và S là ma trận hiệp phương sai.

Bước 3: Tính các trị riêng và các vector riêng tương ứng.

Bước 4: Chọn các vector riêng và giá trị riêng lớn nhất.

Đối với mỗi dữ liệu thử nghiệm, chúng tôi chiếu nó vào không gian riêng của mỗi lớp và lấy dữ liệu nhiệt tái tạo từ mỗi không gian riêng. Sử dụng sai số trung bình, đo độ tương đồng giữa dữ liệu nhiệt đầu vào và dữ liệu nhiệt tái tạo, chúng ta có thể chọn một lớp thích hợp cho dữ liệu nhiệt đầu vào là mức tối thiểu của phương sai.

Sử dụng EMC

Với EMC [46], chúng tôi chia đào tạo thành bảy lớp và tạo ra một không gian riêng. Đối với mỗi dữ liệu thử nghiệm nhiệt, chúng tôi chiếu nó vào không gian riêng của mỗi lớp và một cảm xúc được chọn nếu nó có góc cosin lớn nhất giữa vector thu được sau khi chiếu và không gian riêng của mỗi lớp.

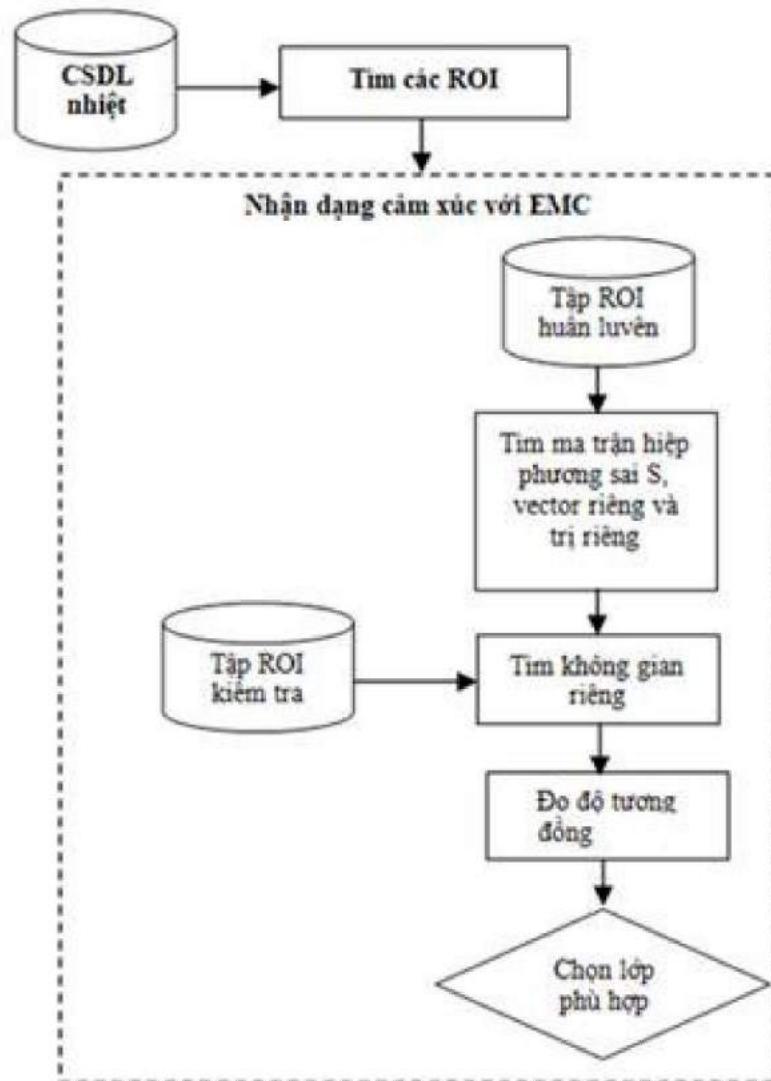
$$S = S_B - S_W \quad (2.8)$$

$$S_W = \frac{1}{M} \sum_{f \in F} \sum_{m=1}^{M_f} M_f (\bar{x}_{fm} - \bar{x}_f)(\bar{x}_{fm} - \bar{x}_f)^T \quad (2.9)$$

$$S_B = \frac{1}{M} \sum_{f \in F} M_f (\bar{x}_f - \bar{x})(\bar{x}_f - \bar{x})^T \quad (2.10)$$

$$\bar{x}_f = \frac{1}{M} \sum_{m=1}^{M_f} x_{fm}; \quad \bar{x} = \frac{1}{M} \sum_{f \in F} \sum_{m=1}^{M_f} x_{fm} \quad (2.11)$$

Trong đó F là tập hợp các lớp biểu thức, khuôn mặt M_f được đưa ra cho mỗi lớp $f \in F$ và x_{fm} là một vector N-chiều, $m = 1, \dots, M_f$



Hình 2.4. Nhận dạng cảm xúc với ROI và EMC

💡 Sử dụng kết hợp PCA-EMC

Với PCA-EMC, chúng tôi sử dụng PCA để giảm kích thước và sau đó áp dụng EMC để thử nghiệm.

Chương 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ

3.1. Môi trường thực nghiệm

Về thông tin máy tính chạy thực nghiệm:

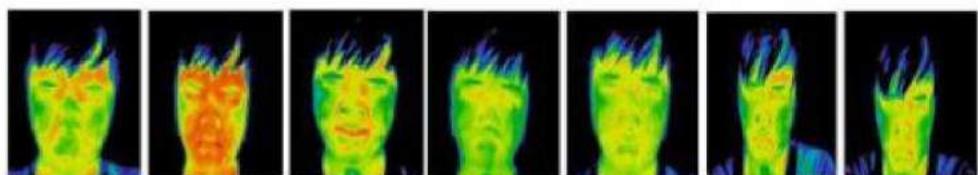
- Hệ điều hành: Window 10 – 64 bit.
- Bộ vi xử lý: Intel(R) Core(TM) i5-7500 CPU @ 3.40GHz 3.40GHz.
- Bộ nhớ RAM: 16.0 GB.

Về ngôn ngữ lập trình:

- Sử dụng ngôn ngữ lập trình Matlab, phiên bản Matlab R2016a.

3.2. Dữ liệu đầu vào

Đối với cơ sở dữ liệu, chúng tôi sử dụng cơ sở dữ liệu Kotani Thermal Facial Emotions (KTFE) [42]. Cơ sở dữ liệu này có 30 đối tượng gồm người Việt, người Nhật, người Thái từ 11 đến 32 tuổi với bảy cảm xúc.



Hình 3.1. Ví dụ hình ảnh nhiệt của bảy cảm xúc (bình thường, giận dữ, hạnh phúc, buồn rầu, sợ hãi, ghê tởm, ngạc nhiên) [42]

Trong các thí nghiệm, chúng tôi chỉ sử dụng dữ liệu nhiệt để nhận dạng cảm xúc của người. Từ 130 GB dữ liệu nhiệt, chúng tôi trích xuất 27,2 GB dữ liệu nhiệt cho bảy cảm xúc. Chúng tôi sử dụng 70% dữ liệu cho huấn luyện và thử nghiệm là 30%.

3.3. Phân tích đánh giá

Để đánh giá khách quan hiệu suất của phương pháp nghiên cứu được đề xuất, trong phần này chúng tôi tiến hành phân tích kết quả thực nghiệm đạt được. Tiếp theo, chúng tôi so sánh phương pháp đề xuất với phương pháp trong [35] và [38]. Với [35], so sánh kết quả trên cùng cơ sở dữ liệu và cùng

phương pháp. Với [38], so sánh kết quả trên cùng cơ sở dữ liệu nhưng phương pháp khác nhau.

3.3.1. Kết quả thực nghiệm

Bảng 3.1 trình bày kết quả nhận dạng cảm xúc với PCA cho dữ liệu nhiệt sử dụng ROI. Kết quả cho thấy tỷ lệ chính xác rất cao ở mức 90.42%, tỷ lệ lỗi chỉ là 9.59%. Đặc biệt, giận dữ đạt 92.5%, hạnh phúc đạt 97.44% và bình thường đạt 98.64%.

Bảng 3.2 trình bày kết quả nhận dạng cảm xúc với EMC cho dữ liệu nhiệt sử dụng ROI. Kết quả cho thấy giận dữ đạt 93.75%, PCA cao hơn là 1.25%, ghê tởm 91.11%, trong khi PCA chỉ đạt 83.25%, tăng 7.86%, buồn rầu đạt 90.32%, tăng 4.22% và ngạc nhiên đạt 94.12%, tăng 5.9% so với PCA. Vì vậy, tỷ lệ chính xác là 89.99%, tỷ lệ lỗi chỉ là 10.01%.

Bảng 3.3 trình bày kết quả nhận dạng cảm xúc với PCA-EMC cho dữ liệu nhiệt sử dụng ROI. Kết quả cho thấy tỷ lệ chính xác là 85.94%, tỷ lệ lỗi chỉ là 14.06%. So với PCA, mặc dù mức trung bình thấp hơn, tỷ lệ sơ hãi tăng 0,28% và buồn rầu tăng 2.17%. So với EMC, hạnh phúc tăng 0.97% và bình thường là 0.39%.

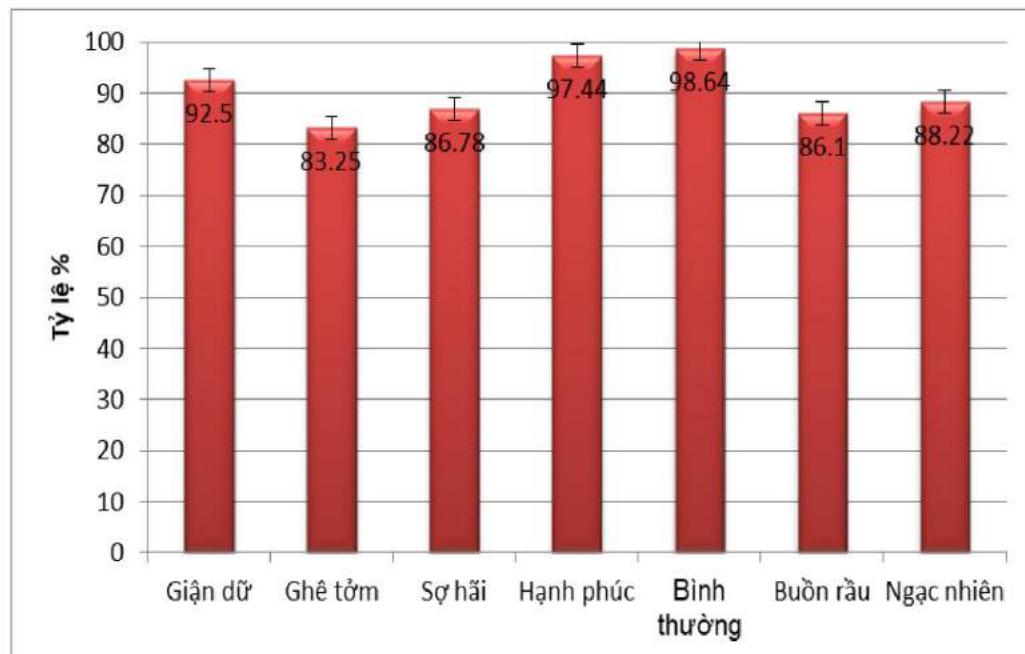
Như vậy, từ kết quả thu được trong bảng: 3.1, 3.2 và 3.3, không có ngạc nhiên trong bất kỳ trường hợp giận dữ nào, không có ghê tởm được công nhận là hạnh phúc và buồn rầu, không có buồn rầu nào được công nhận là ngạc nhiên. Dựa trên kết quả, chúng tôi tin rằng dữ liệu nhiệt là thông tin bổ sung quan trọng để hỗ trợ nhận dạng cảm xúc người chính xác hơn.

Bảng 3.1. Ma trận kết quả với PCA. Từ trái sang phải (trên xuống dưới): giận dữ (An), ghê tởm (Di), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne), buồn rầu (Sa) và ngạc nhiên (Su).

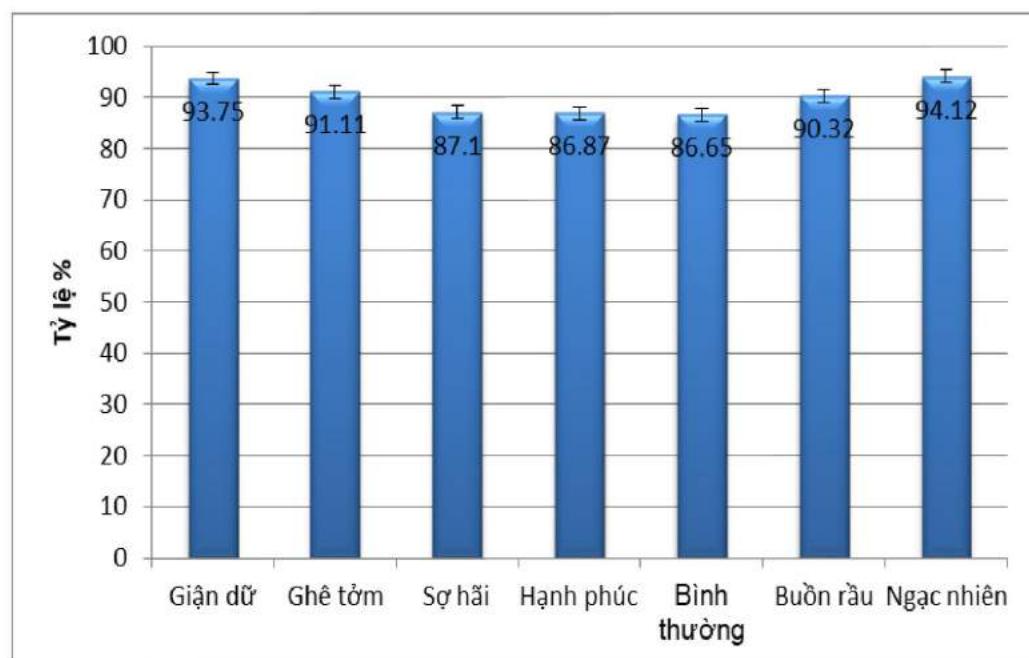
Bảng 3.2. Ma trận kết quả với EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), ghê tởm (Di), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne), buồn rầu (Sa) và ngạc nhiên (Su).

Bảng 3.3. Ma trận kết quả với PCA-EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), ghê tởm (Di), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne), buồn rầu (Sa) và ngạc nhiên (Su).

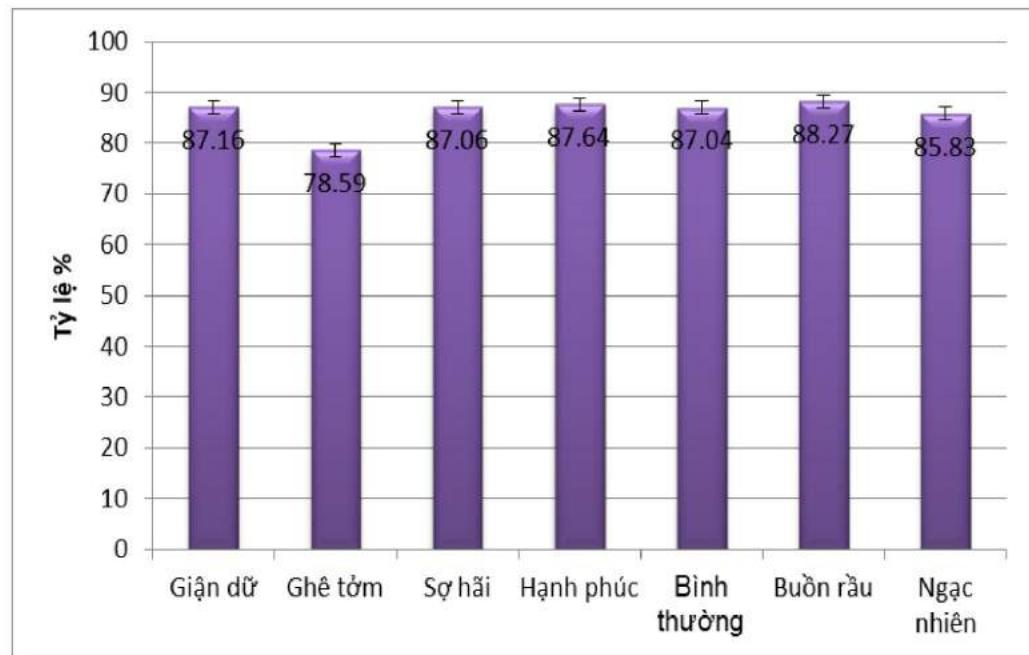
	An	Di	Fe	Ha	Ne	Sa	Su
An	87.16%	2.14%	1.81%	1.43%	2.27%	0%	1.76%
Di	5.24%	78.59%	0%	0%	4.23%	0%	3.67%
Fe	2.86%	5.18%	87.06%	5.44%	1.42%	5.15%	3.43%
Ha	4.74%	1.22%	3.86%	87.64%	1.8%	2.86%	1.88%
Ne	0%	4.17%	0.71%	0.37%	87.04%	2.75%	3.43%
Sa	0%	0.67%	5.49%	2.97%	2.85%	88.27%	0%
Su	0%	8.03%	1.07%	2.15%	0.38%	0.97%	85.83%
Avg.							85.94%



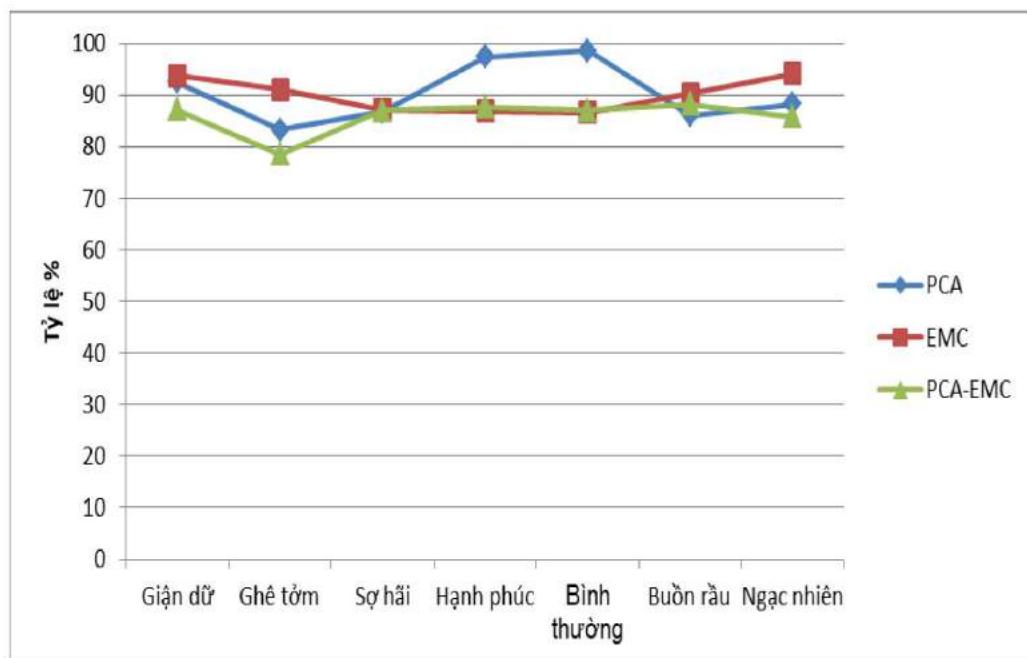
Biểu đồ 3.1. Kết quả nhận dạng cảm xúc của PCA



Biểu đồ 3.2. Kết quả nhận dạng cảm xúc của EMC



Biểu đồ 3.3. Kết quả nhận dạng cảm xúc của PCA-EMC



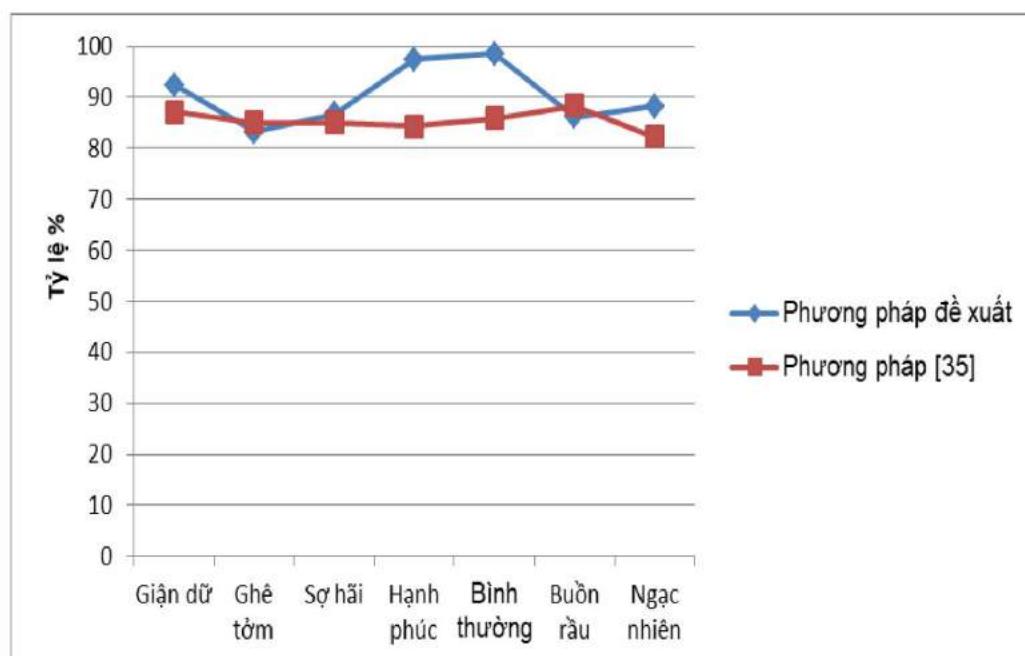
Biểu đồ 3.4. So sánh giữa PCA, EMC và PCA-EMC

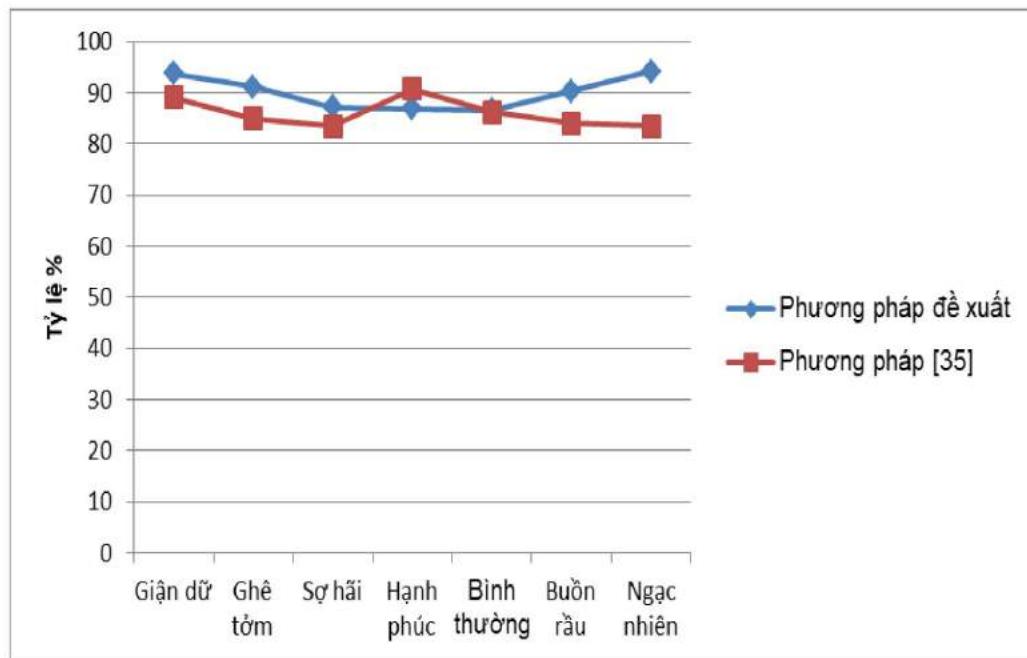
3.3.2. So sánh kết quả thực nghiệm với kết quả của phương pháp [35]

Phương pháp được đề xuất được so sánh với các kết quả thu được bằng cách thực hiện các phương pháp PCA và EMC trên cùng một cơ sở dữ liệu KTFE [35]. Với phương pháp [35], các tác giả đề xuất trích xuất các ROI khá đơn giản. Các ROI tìm thấy không cung cấp thông tin hữu ích thực sự khi cảm xúc thay đổi và có vấn đề với những trở ngại cho nhiệt độ nóng. Bảng 3.4 cho thấy hiệu suất đạt được bằng phương pháp đề xuất tốt hơn so với các phương pháp khác. Bởi vì chúng tôi đã cải thiện ROI, chứa nhiều thông tin hữu ích hơn, cho thấy mối quan hệ tốt hơn giữa nhiệt độ và cảm xúc. Với PCA, tỷ lệ chính xác tăng 4.99% (từ 85.43% lên 90.42%). Trong khi EMC đạt 89.99% (tăng 3.99%). Ngoài ra chúng tôi cũng đã thử nghiệm với PCA-EMC, kết quả đạt 85.94%.

Bảng 3.4. So sánh phân tích với phương pháp [35]

	Tỉ lệ nhận dạng trung bình	
	Phương pháp [35]	Phương pháp đề xuất
PCA	85.43%	90.42%
EMC	86.00%	89.99%
PCA-EMC	-	85.94%

**Biểu đồ 3.5. So sánh giữa phương pháp đề xuất và phương pháp [35] cho PCA**



Biểu đồ 3.6. So sánh giữa phương pháp đề xuất và phương pháp [35] cho EMC

Như vậy, trong cùng điều kiện sử dụng phương pháp phân loại PCA, EMC và PCA-EMC trên cơ sở dữ liệu KTFE thì phương pháp trích xuất ROI của chúng tôi có tỷ lệ nhận dạng trung bình cao hơn so với phương pháp trong [35].

3.3.3. So sánh kết quả thực nghiệm với kết quả của phương pháp [38]

Để phân tích, đánh giá phương pháp đề xuất với phương pháp tác giả sử dụng trong [38], chúng tôi trích xuất 23,5 GB dữ liệu nhiệt trên cơ sở dữ liệu KTFE cho năm cảm xúc tương ứng là: giận dữ, sợ hãi, hạnh phúc, bình thường và buồn rầu. Kết quả được trình bày trong các bảng: 3.5, 3.6 và 3.7

Bảng 3.5. Ma trận kết quả với PCA. Từ trái sang phải (trên xuống dưới): giận dữ (An), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne) và buồn rầu (Sa).

	An	Fe	Ha	Ne	Sa
An	90.8%	2.33%	0%	2.96%	0%
Fe	3.53%	88.75%	2.36%	1.43%	5.59%
Ha	3.91%	3.14%	91.24%	0%	1.36%
Ne	1.76%	0.91%	1.96%	94.41%	0%
Sa	0%	4.87%	4.45%	1.2%	93.04%
Avg.					91.65%

Bảng 3.6. Ma trận kết quả với EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne) và buồn rầu (Sa).

	An	Fe	Ha	Ne	Sa
An	86.36%	0%	0%	8.61%	6.45%
Fe	4.55%	87.1%	7.57%	0%	0%
Ha	4.55%	7.53%	85.29%	0%	0%
Ne	4.55%	0%	0%	86.63%	3.23%
Sa	0%	5.38%	7.14%	4.76%	90.32%
Avg.					87.14%

Bảng 3.7. Ma trận kết quả với PCA-EMC. Từ trái sang phải (trên xuống dưới): giận dữ (An), sợ hãi (Fe), hạnh phúc (Ha), bình thường (Ne) và buồn rầu (Sa).

	An	Fe	Ha	Ne	Sa
An	89.17%	4.74%	1.11%	0.86%	0%
Fe	3.58%	86.04%	6.22%	4.1%	5.15%
Ha	0%	1.07%	86.15%	2.5%	3.5%
Ne	2.35%	3.3%	1.02%	87.3%	2.75%
Sa	4.9%	4.85%	5.49%	5.23%	88.59%
Avg.					87.45%

Bảng 3.8. So sánh phân tích với phương pháp [38]

Phương pháp	Tỷ lệ nhận dạng trung bình
Histogram statistics + Multi Class SVM [38]	81.95%
Moment Invariant + Multi Class SVM [38]	86.74%
Moment invariants + histogram statistics + Multi Class SVM) [38]	87.50%
PCA	91.65%
EMC	87.14%
PCA-EMC	87.45%

Bảng 3.8 cho thấy tỷ lệ nhận dạng trung bình đạt được bằng phương pháp đề xuất tốt hơn so với các phương pháp trong [38] để nhận dạng cảm xúc trên cùng cơ sở dữ liệu KTFE.

Chương 4. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Trong luận văn này, chúng tôi giới thiệu một phương pháp xác định cảm xúc của con người dựa trên hình ảnh nhiệt. Ở đây, mô tả một cách khắc phục ảnh hưởng của ánh sáng môi trường và giảm tác động của kính đeo mắt bằng cách cải thiện việc khai thác khu vực quan tâm cho dữ liệu nhiệt.

Các thuật toán được kiểm tra trên cơ sở dữ liệu KTFE. Bảy cảm xúc cơ bản: giận dữ, ghê tởm, sợ hãi, hạnh phúc, buồn rầu, ngạc nhiên và bình thường được nhận dạng với độ chính xác cao là: 90.42%, 89.99% và 85.94% tương ứng cho: PCA, EMC và PCA- EMC.

Kết quả cho thấy rằng nhiệt độ và cảm xúc có mối liên quan chặt chẽ. Sử dụng thông tin nhiệt độ sẽ tăng tính chính xác cho nhận dạng cảm xúc con người.

Bên cạnh những kết quả đã đạt được, còn có những vấn đề mà thời điểm hiện tại luận văn chưa thực hiện:

- Độ chính xác của từng cảm xúc và từng phương pháp đã được cải thiện với sự điều chỉnh các khu vực quan tâm. Tuy nhiên, trong một số trường hợp bị mất thông tin làm giảm tỷ lệ chính xác. Phương pháp cải thiện ROI vẫn còn đơn giản.

- Nghiên cứu về một số các giải thuật trích chọn đặc trưng và phân lớp dữ liệu khác trên ảnh nhiệt, xử lý ảnh nhiệt với công nghệ học tập sâu,v.v.

Trong tương lai, chúng tôi tiếp tục nghiên cứu thêm về mối quan hệ giữa biểu hiện của người và cảm xúc cũng như mối quan hệ giữa nhiệt độ và cảm xúc. Chúng tôi sẽ tiếp tục cải thiện ROI trên dữ liệu nhiệt để đóng góp kết quả tốt hơn.Thêm vào đó có thể mở rộng kết hợp ảnh nhiệt với ảnh nhìn thấy để tăng thông tin trong nhận dạng cảm xúc hơn.Từ những kiến thức đó, chúng tôi sẽ xây dựng nhiều ứng dụng thực tế hơn có thể hỗ trợ tương tác giữa con người và máy tính tốt hơn.

CÔNG TRÌNH CÔNG BỐ

Thu Nguyen, Khang Tran, Hung Nguyen, “Towards Thermal Region of Interest for Human Emotion Estimation”, *10th International Conference on Knowledge and Systems Engineering, KSE 2018.*

TÀI LIỆU THAM KHẢO

- [1] P.Ekman, editor, “Emotion In the Human Face”. *Cambridge University Press*, New York, NY, 2nd edition, 1982.
- [2] P.Ekman. “Strong evidence for universals in facial expressions: A reply to Russell’s mistaken critique”, *Psychological Bulletin*, pp.268–287, 1994.
- [3] C.Busso, Z.Deng, S.Yildirim, M.Bulut, C.M. Lee, A.Kazemzadeh, S.Lee, U.Neumann, S.Narayanan, “Analysis of emotion recognition using facial expressions, speech and multimodal information”, *In the Proceedings of 6th International Conference on Multimodal Interface, ICMI04, PA*, pp. 205-211, 2004.
- [4] Z.Zeng, M.Pantic, G.I.Roisman, T.S. Huang: “A Survey of Affect recognition methods: Audio, Visual, and Spontaneous Expressions”, *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol.31, no. 1, pp. 39-58, 2009.
- [5] M.M.Khan, R.D.Ward, M.Ingleby, “Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature”, *ACM Transactions on Applied Perception*, vol. 6, pp. 122, 2009.
- [6] M.Pantic, J.M. Rothkrantz, “Automatic Analysis of Facial Expressions: The State of the Art”, *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol.22, no. 12, pp. 1424-1445, 2000.
- [7] A.Corneanu, M.Oliu, F. Cohn, S.Escalera, “Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-related Applications”, pp. 1-20, 2016.
- [8] E.Sariyanidi, H.Gunes, A.Cavallaro, “Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition”, *Pattern*

Analysis and Machine Intelligence, IEEE Transactions, vol.37, no. 6, pp. 1113-1133, 2015.

[9] C.M.Naveen Kumar, G. Shivakumar: “A survey on human emotion analysis using thermal imaging and physiological variables”, *International Journal Of Current Engineering And Scientific Research (Ijcesr)*, Vol -4, Issue-4, pp. 122-126, 2017.

[10] J.A. Russell, “Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies”, *Psychological Bulletin*, vol. 115, no. 1, pp. 102-141, Jan. 1994.

[11] P.Ekman, “Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique”, *Psychological Bulletin*, vol. 115, no. 2, pp. 268-287, Mar. 1994.

[12] E.Izard, “Innate and universal facial expressions: Evidence from developmental and cross-cultural research”, *Psychological Bulletin*, vol. 115, no. 2, pp.288-299, Mar 1994.

[13] B.Scassellati, “Eye finding via face detection for a foveated, active vision system”, in *Proc. 15th Nat. Conf. Artificial Intelligence*, pp. 969 - 976, 1998.

[14] X.Ya, L.Guang-Yuan, “A Method of Emotion Recognition Based on ECG Signal” in *proc. of International Conference on Computational Intelligence and Natural Computing*, pp. 202-205, 2009.

[15] C.Puri, “Stress Cam: Non-contact Measurement of Users' Emotional States through Thermal Imaging”, in *proc of Int. Con. on Human Computer Interaction*, pp. 1725–1728, 2005.

[16] H. Xu and K. N. K. Plataniotis, “Affect Recognition Using EEG Signal”, *IEEE 14th International Workshop on Multimedia Signal Processing*, pp. 299-304, 2012.

- [17] P.Ekman, V.Friesen, “Facial action coding system: a technique for the measurement of facial movement”, *Rivista Di Psichiatria* 47 (2), pp. 126–138, 1978.
- [18] H.C.Chi, S.Y.Oh, “Realtime Facial Expression Recognition Using Active Appearance Model and Multilayer Perceptron”, in *Proceedings of the International Joint Conference SICE-ICASE, Busan, Korea*, pp. 5924–5927, October 2006.
- [19] A.Asthana, J.Saragih, M.Wagner, R.Goecke, “Evaluating AAM Fitting Methods for Facial Expression Recognition”, in *Proceedings of the International Conference on Affective Computing and Intelligent Interaction, Amsterdam, The Netherlands*, pp. 1–8, September 2009.
- [20] N.Sebe, M.S.Lew, Y.Sun, I.Cohen, T.Gevers, T.S.Huang, “Authentic facial expression analysis”. *Image Vision Comput*, pp.1856–1863, 2007.
- [21] Y.Chang, C.Hu, R. Feris, M.Turk, “Manifold based analysis of facial expression”, *Image Vision Comput*, pp.605–614, 2006.
- [22] S.Moore, R.Bowden, “Local binary patterns for multi-view facial expression recognition”, *Comput. Vision Image Underst*, pp.541–558, 2011.
- [23] D.Ghimire, J.Lee, “Histogram of orientation gradient feature-based facial expression classification using bagging with extreme learning machine”. *Adv. Sci. Lett*, pp.156–161,2012.
- [24] A.Dhall, A.Asthana, R.Goecke, T.Gedeon, “Emotion Recognition Using PHOG and LPQ Features”. In *Proceedings of the IEEE International Conference on Face and Gesture Recognition and Workshop, Santa Barbara, CA, USA*, pp. 878–883, 2011.
- [25] T.Jabid, Md.H.Kabir, O.Chae, “Robust facial expression recognition based on local directional pattern”. *ETRI J*, pp.784–794, 2010.

- [26] I.Kotsia, I.Buciu, I.Pitas, “An analysis of facial expression recognition under partial facial image occlusion”, *Image Vision Comput*, pp.1052–1067, 2008.
- [27] D.T.Lin, “Facial expression classification using PCA and hierarchical radial basic function network”. *J. Inf. Sci. Eng*, pp.1033–1046, 2006.
- [28] Z.Wang, Q.Ruan, “Facial Expression Recognition Based Orthogonal Local Fisher Discriminant Analysis”, *In Proceedings of the International Conference on Signal Processing (ICSP), Beijing, China*, pp. 1358–1361, 2010.
- [29] T.J.Phillips, “High performance thermal imaging technology”, *Advanced Semiconductor Magazine*, vol. 15, no. 7, pp. 32-36, 2002.
- [30] S. Ioannou, V. Gallese, A. Merla, “Thermal infrared imaging in psychophysiology: potentialities and limits”, *Psychophysiology* 51 pp.951–963, 2014.
- [31] L.Trujillo, G.Olague, R.Hammoud, B.Hernandez. “Automatic feature localization in thermal images for facial expression recognition”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, CVPR Workshops*, p. 14, 2005.
- [32] A. Di Giacinto, M. Brunetti, G. Sepede, A. Ferretti, A. Merla, “Thermal signature of fear conditioning in mild post traumatic stress disorder”, *Neuroscience* 266, pp.216–223, 2014.
- [33] B.Hernández, G.Olague, R.Hammoud, L.Trujillo, E. Romero. “Visual learning of texture descriptors for facial expression recognition in thermal imagery”, *Computer Vision and Image Understanding*, vol. 106, pp. 258269 , 2007.
- [34] Z.Liu and S.Wang. “Emotion recognition using Hidden Markov Model from facial temperature sequence”, *LNCS 6975*, pp.240-247, 2011.

- [35] H. Nguyen, K. Kotani, F. Chen, B.Le, “Fusion of Visible Images and Thermal Image Sequences for Automated Facial Emotion Estimation”, *Journal of Mobile Multimedia*, Vol. 10, No. 3&4, pp.294-308, 2014.
- [36] B.R.Nhan and T.Chau.: “Classifying affective states using thermal infrared imaging of the human face”, *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 979987, 2010.
- [37] S.He, S.Wang, W.Lan, H.Fu, Q.Ji, “Facial Expression Recognition using Deep Boltzmann Machine from Thermal Infrared Images”, *Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp.239-244, 2013.
- [38] A.Basu, A.Routray, S.Shit, A.K.Deb, “Human Emotion Recognition from Facial Thermal Image based on Fused Statistical Feature and Multi-Class SVM”, *IEEE INDICON*, pp.1-5, 2015.
- [39] <http://www.vcipl.okstate.edu/otcbvs/bench/>.
- [40] <http://www.equinoxsensors.com/>.
- [41] S.Wang, Z.Liu, S.Lv, Y.Lv, G.Wu, P. Peng, F.Chen, X.Wang, “A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference”, *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 682-691, 2010.
- [42] H.Nguyen, K.Kotani, F.Chen, B.Le, “A thermal facial emotion database and its analysis”, *PSIVT*, 2013.
- [43] A. Patil, J. Deore, Ph.D, “Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) based Face Recognition”, *NCACC*, pp.1-5, 2014.
- [44] R.L. Kristensen, editor, “Emotion recognition in blurred images with local features and machine learning”, 2014.

- [45] D.T.LIN, “Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network”, *Journal of Information Science and Engineering*, vol. 22, pp. 1033-1046 , 2006.
- [46] T.Kurozumi, Y.Shinza, Y.Kenmochi, K.Kotani, “Facial Individuality and Expression Analysis by Eigenspace Method Based on Class Features or Multiple Discriminant Analysis”, *ICIP*, 1999.
- [47] X.Gao, Y.Su, X.Li, D.Tao, “A review of active appearance models”, *IEEE Transactions on Systems, Man and Cybernetics C: Applications and Reviews*, vol. 40, no. 2, pp. 145–158, 2010.
- [48] E.Vural, M.Bartlett, G.Littlewort, M.Cetin, A.Ercil, J.Movellan, “Discrimination of moderate and acute drowsiness based on spontaneous facial expressions”, *In ICPR*, 2010.
- [49] F.D.L.Torre and J.F.Cohn, “Facial expression analysis”, *In Visual Analysis of Humans, Springer*, pp. 377-409, 2011.
- [50] V.W.Zue and J.R.Glass, “Conversational interfaces: Advances and challenges”, *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1166-1180, 2002.
- [51] T.Nguyen, K.Tran, H.Nguyen, “Towards Thermal Region of Interest for Human Emotion Estimation”, *10th International Conference on Knowledge and Systems Engineering, KSE 2018*, pp.1-6, 2018.

PHỤ LỤC