

ỨNG DỤNG MÔ HÌNH KẾT HỢP TRONG NHẬN DẠNG CẢM XÚC APPLIED MULTIMODAL FOR EMOTION RECOGNITION

Nguyễn Thị Khánh Hồng, Võ Thị Hương, Lê Hữu Duy
Trường Đại học Sư phạm Kỹ thuật – Đại học Đà Nẵng, Việt Nam

Ngày toà soạn nhận bài 12/11/2019, ngày phản biện đánh giá 21/11/2019, ngày chấp nhận đăng 9/12/2019.

TÓM TẮT

Ngày nay, nhận dạng và phân loại cảm xúc ngày càng chi tiết và chính xác hơn nhờ vào sự phát triển của các lĩnh vực như điện tử, cảm biến hay kỹ thuật máy tính. Các phương pháp nhận dạng cảm xúc được nghiên cứu hiện nay sử dụng nhiều phương thức thu thập dữ liệu khác nhau và một trong những phương thức mang lại hiệu quả là sử dụng cảm biến y sinh. Phương thức này có thể cung cấp thông tin về các yếu tố sinh học chính xác, bền vững ít bị ảnh hưởng bởi tác nhân gây nhiễu từ bên ngoài. Bài báo này mô tả phương pháp phân loại và đánh giá cảm xúc dựa trên mô hình kết hợp sử dụng tín hiệu cảm biến y sinh và tín hiệu video trên cơ sở học máy. Chúng tôi sẽ mô tả thiết kế của hệ thống thu thập tín hiệu y sinh, qui trình thu thập thông tin và hệ thống xử lý được dùng để nhận dạng các thuộc tính của cảm xúc. Nghiên cứu này đạt được hiệu suất nhận dạng với tỉ lệ chính xác là 83,2%.

Từ khóa: Nhận dạng cảm xúc; cảm biến y sinh; máy học; mạng Nơ ron tích chập; máy học vectơ hỗ trợ.

ABSTRACT

Detecting and classifying emotions has currently become an important item of research and life. The more detailed and accurate emotion recognition system is due to the development of various fields such as electronics, sensors or computer engineering. Emotion recognition methods are studied using different data collection methods and one of the most popular and effective methods is physical – bio sensors. Physical – bio sensor based approaches can provide more accurate, sustainable biological information with external influences and interferences, especially when we compared with other approaches such as image processing, video processing. In this paper, a method of classifying and assessing emotions based on a combination of signals collected from physical – bio sensors, video collection and machine learning are supposed. Specifically, we will describe the platform of a physical – bio signal collection system, the process of collecting information and the information processing system used to identify how emotional behavior is characterized. We have also shown that a combination modals of physical – bio sensor acquisition systems, video processing based on machine learning methods can provide identification performance with an accuracy of 83.2%

Keywords: Emotion Recognition; physical – bio sensor; machine learning; CNN; SVM.

1. TỔNG QUAN

Bài toán phân loại, đánh giá và nhận dạng cảm xúc của con người đang là một lĩnh vực nhận được nhiều sự quan tâm trong các nghiên cứu hiện nay. Bài toán nhận dạng cảm xúc được thực hiện dựa vào các thông tin đầu vào là các tính hiệu y sinh như nhịp tim,

nồng độ oxy, huyết áp, hay các tín hiệu được thu thập từ hình ảnh, video, hay từ các đánh giá của bác sĩ, chuyên gia [1]. Kết quả sau khi xử lý các thông tin được thu thập là các kết luận về tình trạng và trạng thái cảm xúc như vui, buồn, tức giận... Bài toán nhận dạng và phân loại cảm xúc là một bài toán phức tạp vì nhiều yếu tố khác nhau và phụ thuộc

vào các thông tin đầu vào [2]. Ngoài ra sự đánh giá nhận dạng cảm xúc hầu hết dựa trên các suy luận logic từ các thông tin đầu vào, tuy nhiên cảm xúc lại hoạt động theo yếu tố phi logic [3]. Ví dụ như một người có cảm xúc cực kỳ tốt, cực kỳ thoải mái, vẫn không thể miêu tả được định nghĩa về cảm xúc tốt một cách đồng nhất với người khác. Vì vậy bài toán nhận dạng cảm xúc vẫn là một bài toán thách thức.

Tuy gặp nhiều khó khăn và thách thức, bài toán phân loại và nhận dạng cảm xúc được tập trung nghiên cứu vì nhiều lý do khác nhau:

- Bằng việc hiểu được cảm xúc của con người, các hệ thống thông tin hiện nay có thể tăng mức độ tương tác với con người, tạo trải nghiệm người dùng hoàn toàn mới. Ví dụ: Hệ thống âm thanh có thể giảm nhạc, giảm âm lượng trong thời gian con người đang căng thẳng tốt độ hoặc sợ hãi, hoặc gợi ý những bộ phim phù hợp theo tâm trạng người xem. Máy tính và hệ thống giải trí cũng có thể nhận dạng được phản ứng của người sử dụng như thích thú, hay khó chịu khi một nội dung giải trí được trình chiếu ngẫu nhiên. Như vậy, hệ thống giải trí và thông tin dường như trở thành người bạn đồng hành hơn là những công cụ phục vụ cho con người hằng ngày [4].
- Đánh giá và nhận dạng cảm xúc còn có thể được sử dụng như một phương thức giám sát sinh học, nhằm bổ sung để tự đánh giá, giám sát các trạng thái cảm xúc của con người. Kết quả của việc giám sát sinh học có nhiều lợi ích chẳng hạn như một phương pháp để cải thiện khả năng giao tiếp, hoặc đánh giá các hành vi cảm xúc có thể gây những tác động không tích cực lên con người và xã hội. Từ đó có thể đưa ra những phản ứng và phương pháp phù hợp để hạn chế, hoặc cải thiện nó [5].

Như đã trình bày ở trên, có nhiều phương pháp được dùng để thu thập các thông tin tín hiệu cảm xúc như tín hiệu y sinh (nhịp tim, huyết áp,...) hay các yếu tố như nét mặt, nhịp điệu lời nói, cử chỉ,... hay các đánh

giá khách quan từ chuyên gia, bác sĩ [6]. Mỗi phương pháp thu thập các thông tin tín hiệu cảm xúc có những ưu nhược điểm khác nhau. Và trong bài báo này, chúng tôi sẽ tập trung vào phương pháp hiện nay đang được tập trung nghiên cứu. Đó là phương pháp sử dụng các loại cảm biến sinh học để thu thập các thông tin y sinh như: Điện cơ đồ (Electromyography-EMG); Điện tâm đồ (Electrocardiogram-EKG) và Độ điện dẫn của da (Electrodermal activity- EDA). Phương pháp sử dụng cảm biến y sinh học có nhiều ưu điểm nổi trội so với các phương pháp khác. Vì thế các thiết bị thu thập tín hiệu sử dụng cảm biến y sinh có thể được nhỏ gọn hơn nhờ vào công nghệ tích hợp như các thiết bị đeo trên người, thậm chí là trang sức, tạo cho người sử dụng cảm giác riêng tư hơn các phương pháp khác, khi người sử dụng phải chịu “giám sát” bởi camera hoặc các thiết bị ghi âm, ghi hình khác. Tuy nhiên, để tăng độ chính xác đối với dữ liệu thu thập được trong nhận dạng cảm xúc, bài báo đã sử dụng thêm kỹ thuật nhận dạng cảm xúc bằng hình ảnh và đề xuất mô hình kết hợp [7] [8].

Mục tiêu của chúng tôi là thu thập các tín hiệu y sinh của con người trong các điều kiện khác nhau của cuộc sống thực để tự động nhận dạng cảm xúc. Chúng tôi đề xuất một phương pháp để nhận dạng cảm xúc dựa trên mô hình kết hợp trong đó sử dụng các tín hiệu y sinh và tín hiệu video. Trong phần 2 của bài báo này là những trích dẫn các phương pháp nghiên cứu nhằm tăng tỷ lệ nhận dạng cảm xúc dựa trên mô hình kết hợp. Tiếp đến phần 3, sẽ trình bày và giải thích tất cả các bước của phương pháp được đề xuất. Trong phần 4, so sánh giữa kết quả thu được là tỷ lệ nhận dạng cảm xúc dựa trên mô hình đơn và mô hình kết hợp cũng như với các mô hình nghiên cứu khác. Cuối cùng, kết luận và những nghiên cứu trong tương lai được mô tả trong phần 5.

2. TỔNG QUAN TÌNH HÌNH NGHIÊN CỨU VỀ MÔ HÌNH NHẬN DẠNG CẢM XÚC

Một hệ thống nhận dạng cảm xúc nhìn chung đều dựa trên ba bước cơ bản:

- Thu nhận các tín hiệu,
- Tính năng khai thác các thuộc tính,
- Việc nhận dạng cảm xúc.

Một số nghiên cứu đã tập trung vào việc nhận dạng cảm xúc bằng biểu cảm khuôn mặt, biểu cảm giọng nói hoặc tín hiệu sinh học [9], [10], [11]. Tuy nhiên, có ít nghiên cứu tập trung vào đề xuất mô hình kết hợp trong nhận dạng cảm xúc [12].

Ngoài ra, cách tiếp cận theo hướng sử dụng mô hình kết hợp không chỉ giúp tăng tỷ lệ nhận dạng mà còn tăng thêm độ tin cậy cho hệ thống khi có sự kết hợp đa dạng từ nhiều nguồn dữ liệu khác nhau trong các điều kiện môi trường khác nhau [13].

Về lý thuyết, có ba phương pháp [14] để hợp nhất các tín hiệu từ các cảm biến khác nhau:

- Mô hình kết hợp các tín hiệu y sinh: được thực hiện trực tiếp trên dữ liệu thô từ mỗi cảm biến tín hiệu y sinh; nó chỉ có thể được áp dụng khi các tín hiệu y sinh này có bản chất tương tự và có cùng độ phân giải tín hiệu. Do đó, kỹ thuật này hiếm khi được sử dụng vì khó hợp nhất các tín hiệu với nhau và dễ bị nhiễu do phụ thuộc vào độ nhạy các cảm biến.
- Mô hình kết hợp các thuộc tính [15]: phương pháp này thường được sử dụng nhằm mục đích hình thành một vector đa thức từ các vector thuộc tính được trích xuất từ mỗi cảm biến. Phương pháp này có lợi thế là chỉ cần một giai đoạn học tập duy nhất và kết quả của vector thuộc tính là vector đa thức.
- Mô hình kết hợp các thuộc tính và mức độ ra quyết định nhận dạng [16]: sau khi được phân loại riêng biệt từ mỗi tín hiệu cảm biến, đây là cách hợp nhất các quyết định nhận dạng khác nhau để có được kết quả nhận dạng cảm xúc cuối cùng.

Trong phạm vi bài báo này, chúng tôi trình bày một phương pháp nhận dạng cảm xúc theo mô hình kết hợp các thuộc tính và tự động nhận dạng dựa trên sự kết hợp của các mức độ ra quyết định nhận dạng như nói

trên và gọi chung là mô hình kết hợp trong nhận dạng cảm xúc.

3. PHƯƠNG PHÁP NGHIÊN CỨU

Trong phần này, chúng tôi đề xuất mô hình kết hợp trong nhận dạng cảm xúc. Nghiên cứu này gồm có 2 phần chính là:

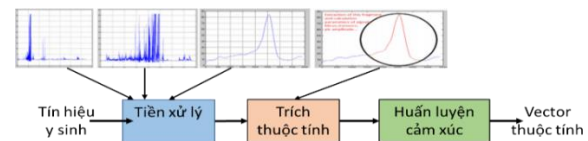
- Phần huấn luyện cảm xúc.
- Phần nhận dạng cảm xúc.

3.1 Phần huấn luyện cảm xúc

Phần huấn luyện cảm xúc này có 3 khối xử lý chính như mô tả ở Hình 1.

- Khối tiền xử lý.
- Khối trích thuộc tính.
- Khối huấn luyện dữ liệu.

Kết quả của phần huấn luyện cảm xúc sẽ được sử dụng để tự động nhận dạng được cảm xúc trong phần nhận dạng cảm xúc.



Hình 1. Những khối cơ bản của phần huấn luyện cảm xúc

3.1.1. Khối tiền xử lý

Đây là khối thu thập tín hiệu y sinh thô đầu vào và được lưu trữ trong cơ sở dữ liệu UTE-EMOTICA của nhóm nghiên cứu Trường Đại học Sư phạm Kỹ thuật – Đại học Đà Nẵng. Chúng tôi tách tín hiệu thô này theo chu kỳ của mỗi loại tín hiệu cảm xúc. Tiếp đến, tín hiệu này sẽ được lọc nhiễu bằng cách sử dụng bộ lọc Hanning.

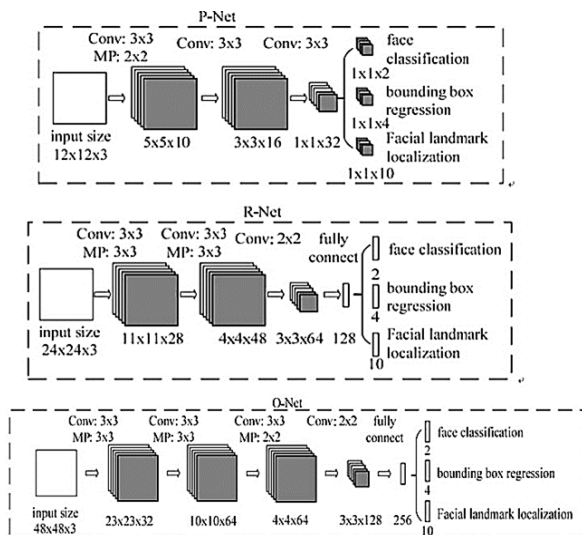
3.1.2. Khối trích thuộc tính

Tín hiệu sau khi qua khối tiền xử lý sẽ tiếp tục được xác định giá trị đỉnh của tín hiệu bằng cách tính toán độ dốc của tín hiệu và sau đó tìm thấy các dấu hiệu thay đổi của độ dốc. Để phát hiện và tách một đỉnh, phương pháp được sử dụng ở đây là tính toán một vector thuộc tính bao gồm năm thuộc tính đó là: giá trị trung bình, phương sai, giá trị trung bình của tín hiệu được lọc, phương sai của tín hiệu được lọc và biên độ của đỉnh.

Đối với tín hiệu video, quá trình trích thuộc tính cảm xúc dựa trên thông tin khuôn mặt bao gồm 3 bước chính. Bước đầu tiên là bước nhận dạng khuôn mặt từ ảnh đầu vào trích từ tín hiệu video đã quay, sau đó thực hiện các phép biến đổi và cân chỉnh khuôn mặt sử dụng phương pháp Facial Landmark ở bước 2. Và cuối cùng đưa qua một mô hình để trích xuất dữ liệu.

Hai bước đầu tiên là nhận dạng khuôn mặt và cân chỉnh khuôn mặt đóng vai trò quan trọng cho bước cuối cùng.

Ở bước 1, chúng tôi sử dụng mô hình MTCNN (Multi-Task Cascaded Convolutional Neural Network) [17] là mô hình cho kết quả tốt nhất trong bài toán nhận dạng khuôn mặt. Mô hình MTCNN bao gồm 3 mạng CNN là: P-Net, R-Net và O-Net như trong Hình 2. Kết quả đầu ra bao gồm hình chữ nhật (còn gọi là bounding box) chứa vị trí khuôn mặt và 10 điểm nhận dạng (còn gọi là landmark) của khuôn mặt. Chúng tôi chỉ sử dụng kết quả bounding box cho bước tiếp theo.



Hình 2. Mạng Nơ ron tích chập đa nhiệm

Ở bước tiếp theo, sau khi nhận dạng khuôn mặt, cân chỉnh khuôn mặt được sử dụng trước khi dữ liệu được đi qua bước trích thuộc tính ở bước cuối cùng. Hệ thống trích xuất thông tin trên khuôn mặt phụ thuộc rất lớn vào góc nhìn của người đối với camera, vì thế bước cân chỉnh khuôn mặt này rất quan trọng [18] là thực hiện cân chỉnh khuôn

mặt như nhìn trực diện vào camera. Bước này sẽ trích ra 194 điểm landmark trên vị trí khuôn mặt, nhưng chỉ chọn ra 10 điểm landmark. Tuy nhiên, chúng tôi không sử dụng kết quả landmark được tạo ra bởi mạng MTCNN.

48x48x3 Input
2x(Conv 3x3x64, stride =1) Maxpooling 2x2, stride =2
3x(Conv 3x3x128, stride =1) Maxpooling 2x2, stride =2
3x(Conv 3x3x256, stride =1) Maxpooling 2x2, stride =2
3x(Conv 3x3x512, stride =1) Maxpooling 2x2, stride =2
Flatten FullConnected(128)

Hình 3. Kiến trúc mạng CNN trích thuộc tính cảm xúc

Ở bước cuối cùng, vector 128 chiều được trích ra từ dữ liệu đã được cân chỉnh và làm thuộc tính của dữ liệu khuôn mặt. Để trích thuộc tính một mạng CNN 6 lớp được sử dụng và đã được huấn luyện với cơ sở dữ liệu về cảm xúc Kaggle như Hình 3.

3.1.3. Khối nhận dạng cảm xúc

Nghiên cứu này được thực hiện thu thập dữ liệu dựa trên 7 loại cảm xúc được gán nhãn là (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). Chúng tôi chọn 280 mẫu tín hiệu của 7 loại cảm xúc nêu trên trong cơ sở dữ liệu UTE-EMOTICA.

Tại từng frame, 5 thuộc tính được trích ra từ 1 tín hiệu y sinh. Như vậy có tổng cộng 15 thuộc tính được trích xuất từ 3 loại dữ liệu y sinh. Kết hợp với thuộc tính 128 chiều từ dữ liệu video, chúng tôi có vector 143 chiều và sẽ được sử dụng để huấn luyện hay nhận dạng.

Sau khi trích thuộc tính, giải thuật SVM (Support Vector Machine) tuyến tính được sử dụng để huấn luyện mô hình. SVM được sử dụng rộng rãi trong bài toán nhận dạng và phân loại.

Phương pháp SVM phân các lớp dữ liệu với mục tiêu là tìm mặt phân cách sao cho

biên (còn gọi là margin) tìm được là lớn nhất, đồng nghĩa với việc các điểm dữ liệu an toàn nhất so với mặt phân cách [19] [20].

3.2 Phân nhận dạng cảm xúc

Trong bước nhận dạng, vector thuộc tính được trích xuất từ các dữ liệu tín hiệu y sinh đầu vào và tín hiệu video được trích xuất. Những tín hiệu này đã được phân loại bằng mô hình SVM và đã được học trước đó. Kết quả trả lại của quá trình phân loại là nhãn của trạng thái cảm xúc tương ứng.

4. KẾT QUẢ NGHIÊN CỨU

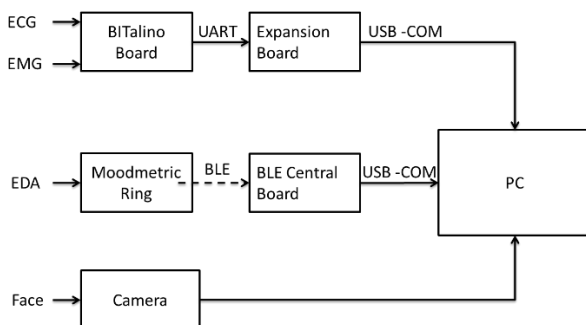
4.1 Cơ sở dữ liệu UTE-EMOTICA

Trong cơ sở dữ liệu này, chúng tôi đã xây dựng như sau:

- Sử dụng 3 cảm biến y sinh là: EMG, ECG, EDA và các tín hiệu video ghi lại khuôn mặt người bằng camera.
- 7 cảm xúc được đo là "Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral".
- Độ dài của mỗi mẫu tín hiệu cảm xúc được thu thập trong năm phút mỗi ngày của mỗi người.
- Tiến hành đo trên 20 người, độ tuổi là 21-23, giới tính nam.
- Thời gian tiến hành đo liên tục trong 6 ngày (từ thứ Hai đến thứ Bảy).

Trong quá trình chọn lọc từ các mẫu tín hiệu thu được, chúng tôi đã chọn ra 280 mẫu có chất lượng tốt để thực hiện việc huấn luyện và nhận dạng cảm xúc.

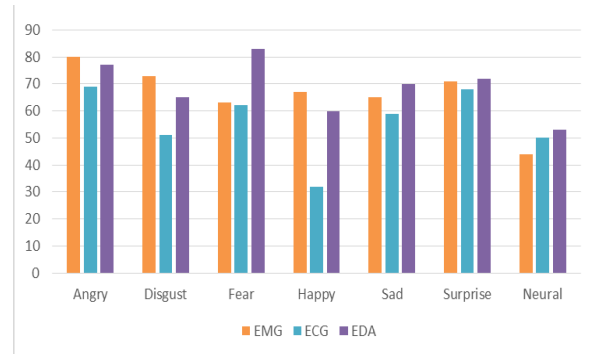
Hệ thống phần cứng sử dụng để thực đo lấy số liệu được thực hiện theo sơ đồ sau:



Hình 4. Sơ đồ hệ thống phần cứng để thu thập dữ liệu.

4.2 Kết quả đạt được

Các kết quả được mô tả trên biểu đồ bên dưới (Hình 5) bằng cách sử dụng các mô hình đơn của từng loại cảm xúc. Cách tiếp cận này cho phép có tỷ lệ nhận dạng trung bình là 63,52%.



Hình 5. Tỷ lệ nhận dạng tín hiệu cảm xúc của mô hình đơn

Từ phân tích dữ liệu trong Hình 5 cho thấy một số cảm xúc nhất định được nhận dạng tốt hơn với mỗi loại cảm biến y sinh nhất định so với các loại cảm biến khác. Thật vậy, phương thức sử dụng EMG cho phép nhận dạng tốt hơn với các cảm xúc "Angry" và "Disgust", trong khi phương thức sử dụng ECG nhận dạng tốt hơn các cảm xúc "Disgust" và "Surprise". Phương thức sử dụng EDA lại cho phép nhận dạng tốt hơn những cảm xúc "Angry" và "Fear". Những đặc điểm này là rất quan trọng bởi điều này sẽ giúp chúng ta lựa chọn phương thức nhận dạng cảm xúc phù hợp nhằm mục đích nhận dạng và phân biệt cảm xúc hiệu quả hơn.

Bên cạnh đó, để cải thiện tỷ lệ nhận dạng tín hiệu cảm xúc trung bình so với mô hình đơn. Chúng tôi đề xuất 2 mô hình là:

Mô hình 1: có hai tín hiệu y sinh học (EMG và ECG) và tín hiệu video trích xuất khuôn mặt.

Mô hình 2: có ba tín hiệu y sinh học: EMG, ECG và EDA và tín hiệu video trích xuất khuôn mặt.

Sau đó, để tăng tỷ lệ nhận dạng cảm xúc chúng tôi áp dụng mô hình kết hợp (như được trình bày trong phần 2). Thật vậy, cách

tiếp cận sử dụng mô hình kết hợp này cho phép có tỷ lệ nhận dạng trung bình là 74,3% với mô hình 1 và 83,2% với mô hình 2. Đây là một sự cải thiện đáng kể về tỷ lệ nhận dạng so với phương pháp không theo mô hình kết hợp có tỷ lệ nhận dạng trung bình chỉ đạt 63,52 %. Kết quả so sánh hai mô hình này được biểu diễn chi tiết hơn trong Bảng 1.

Bảng 1. Tỷ lệ nhận dạng của hai mô hình

Loại tín hiệu cảm xúc	Tỷ lệ nhận dạng tín hiệu cảm xúc	
	Mô hình 1	Mô hình 2
Angry	86.02	94.11
Disgust	77.63	89.89
Fear	62.4	73.2
Happy	69.6	80.3
Sad	71.67	77.45
Surprise	80.2	87.4
Neural	72.6	79.89

Các kết quả được trình bày trong Bảng 1, cho thấy tỷ lệ nhận dạng trung bình của từng loại cảm xúc trong số bảy cảm xúc nêu trên với mức thấp nhất là 62,4% thu được cho cảm xúc "Fear" và tối đa cho cảm xúc "Angry" đạt được tỷ lệ nhận dạng là 94,11%. Trong một số các kết quả nghiên cứu khác: tác giả Imen [21] đã sử dụng 4 loại cảm biến y sinh là EMG, RESP (Respiratory - đo nhịp thở), BVP (Blood Volume Pulse – đo thể tích máu), GSR (Galvanic skin response – độ dẫn điện của da) đã đạt được kết quả nhận dạng cảm xúc là 71 % trong nhận dạng 5 loại cảm xúc là “neutral, pleasure, unpleasant, joy, reviviscence”. Bên cạnh đó, trong nghiên cứu của Chaka Koné và đồng tác giả [22] đã sử dụng mô hình kết hợp 4 loại cảm biến EMG, RESP, GSR, BVP đạt được 81.69 % cho nhận dạng 8 loại cảm xúc là "no emotion, anger, hate, grief, platonic love, romantic love, joy and reverence". Nghiên cứu của tác giả Santamaria-Granados và các đồng tác giả sử dụng 3 loại cảm biến kết hợp là ECG và GSR để nhận dạng 2 cảm xúc là “arousal” và “valence” có tỷ lệ nhận dạng là 76%. Bảng 2 sẽ thể hiện sự so sánh với các kết quả nghiên cứu khác trong nhận dạng cảm xúc.

Bảng 1. Độ chính xác các mô hình kết hợp

Phương pháp nghiên cứu	Tỷ lệ nhận dạng tín hiệu cảm xúc (%)
Nghiên cứu của tác giả Santamaria-Granados và đồng tác giả [23]	76
Mô hình kết hợp sử dụng 4 loại tín hiệu y sinh [21]	71
Nghiên cứu của Chaka Koné và đồng tác giả [22]	81.69
Mô hình kết hợp đề xuất (mô hình 2) sử dụng cơ sở dữ liệu UTE-Emotica	83.2

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN CỦA ĐỀ TÀI

Chúng tôi đề xuất một phương pháp cải thiện tỷ lệ nhận dạng cảm xúc trung bình theo mô hình kết hợp dựa vào mức độ ra quyết định trong đó sử dụng các tín hiệu y sinh và tín hiệu video. Những tín hiệu của hai mô hình 1 và 2 này đã được áp dụng để nhận biết 7 cảm xúc cơ bản (Angry, Disgust, Fear, Happy, Sad, Surprise, Neutral). Mô hình đề xuất của chúng tôi đã thực hiện nhận dạng dựa trên thuật toán máy học. Các kết quả so với các nghiên cứu khác nhau thể hiện một sự cải thiện đáng kể trong tỷ lệ nhận dạng cảm xúc trung bình. Trong tương lai, chúng tôi sẽ tăng số lượng mẫu tín hiệu cho cơ sở dữ liệu UTE-EMOTICA sử dụng các loại tín hiệu (gồm tín hiệu y sinh và tín hiệu video). Đồng thời, chúng tôi sẽ cải thiện thuật toán nhận dạng cảm xúc nhằm tăng tỷ lệ nhận dạng cảm xúc trung bình của hệ thống.

LỜI CẢM ƠN

“Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Đại học Đà Nẵng trong đề tài có mã số B2017-ĐN06-03”

“Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Trường Đại học Sư phạm Kỹ thuật – Đại học Đà Nẵng trong đề tài có mã số T2018-06-86”

TÀI LIỆU THAM KHẢO

- [1] Rangaraj M. Rangayyan, Biomedical Signal Analysis – A Case-Study Approach, *IEEE Press* 2002.
- [2] Barreto A., Heimer M., and Garcia M., Characterization of Photoplethysmographic Blood Volume Pulse Waveforms for Exercise Evaluation, *Proceedings 14th Southern Biomedical Engineering Conference, Shreveport, Louisiana, April, 1995*, pp. 220-223.
- [3] Christie, Israel C, Multivariate Discrimination of Emotion-specific Autonomic Nervous System Activity, *Master Thesis in Science of Psychology, Virginia*.
- [4] Picard R.W., Toward computers that recognize and respond to user emotion, *IBM Systems Journal; Vol 39, Nos 3&4, 2000*.
- [5] Healy J. and Picard R., Digital processing of Affective Signals, *ICASSP 98*.
- [6] Cowie R., Describing the emotional states expressed in speech, *ISCA workshop on speech and emotion, Belfast 2000*.
- [7] Juang B.H & Soong F.K., Hands-free Telecommunications, *HSC 2001*, pp.5-10; Kyoto, Japan.
- [8] Pentland A., Perceptual Intelligence, *Communications of the ACM; Volume 43, Number 3 (2000)*, Pages 35-44.
- [9] E. Monte-Moreno, M. Chetouani, M. Faundez-Zanuy and J. SoleCasals, Maximum likelihood linear programming data fusion for speaker recognition, *Speech Communication*, 51(9):820–830, 2009. 68.
- [10] A. Mahdhaoui and M. Chetouani, Emotional speech classification based on multi view characterization, *IAPR International Conference on Pattern Recognition, ICPR, 2010*. 51.
- [11] Ammar Mahdhaoui, Analyse de Signaux Sociaux pour la Modélisation de l'interaction face à face. *Traitement du signal et de l'image, Université Pierre et Marie Curie - Paris VI, 2010. French*. <tel-00587051>.
- [12] Sebe, Nicu, et al, Bimodal emotion recognition. *Proceedings of the 5th International Conference on Methods and Techniques in Behavioral Research. 2005*.
- [13] Caifeng Shan, Shaogang Gong, Peter W. McOwan, Facial expression recognition based on Local Binary Patterns: A comprehensive study, *Image and Vision Computing* 27 (2009) 803–816.
- [14] E. Monte-Moreno, M. Chetouani, M. Faundez-Zanuy et J. SoleCasals, Maximum likelihood linear programming data fusion for speaker recognition, *Speech Communication*, 51(9):820–830, 2009. 68.
- [15] Hamza Hamdi, Plate-forme multimodale pour la reconnaissance d'émotions via l'analyse de signaux physiologiques: Application à la simulation d'entretiens d'embauche, *Modeling and Simulation. Université d'Angers, 2012. French*. <tel-00997249>.
- [16] R. Sharma, V.I. Pavlovic, and T.S. Huang, Toward multimodal human-computer interface. *Proceedings of the IEEE*, 86(5):853–869, 1998. 29, 30, 32, 167.
- [17] Kaipeng Z., Zhanpeng Z., Zhifeng L., Yu Q., Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks, *IEEE Signal Processing Letters*, vol. 23, pp. 1499-1503 (2019).
- [18] Vahid K., Josephine S., One Millisecond Face Alignment with an Ensemble of Regression Trees, *CVPR IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867-1874 (2014).
- [19] Chen, Jun-Cheng, Vishal M. Patel, and Rama Chellappa., Unconstrained face verification using deep cnn features. *IEEE winter conference on applications of computer vision (WACV). IEEE, 2016*.
- [20] Hsu C.W., Lin C.J., A comparison of methods for multiclass support vector machines, *IEEE Trans Neural Network* 13(2):415–425 (2002).

- [21] Imen Tayari Meftah, Modélisation, détection et annotation des états émotionnels à l'aide d'un espace vectoriel multidimensionnel, *Artificial Intelligence, Université Nice Sophia Antipolis, 2013*. French. <NNT: 2013NICE4017>. <tel-00908233>.
- [22] Koné C., Tayari I.M., Le-Thanh N., Belleudy C., Multimodal Recognition of Emotions Using Physiological Signals with the Method of Decision-Level Fusion, *Healthcare Applications. Inclusive Smart Cities and e-Health. ICOST 2015*. Lecture Notes in Computer Science, vol 9102. Springer.
- [23] Santamaria-Granados, Luz, et al., Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS). *IEEE Access* 7 (2018): 57-67.

Tác giả chịu trách nhiệm bài viết:

Nguyễn Thị Khánh Hồng

Trường Đại học Sư phạm Kỹ thuật – Đại học Đà Nẵng

Email: ntkhong@ute.edu.vn