

TỔNG QUAN VỀ ƯỚC LƯỢNG MỨC ĐỘ CẢM XÚC CỦA NGƯỜI QUA BIỂU CẢM KHUÔN MẶT VÀ HƯỚNG TIẾP CẬN DỰA TRÊN ẢNH NHIỆT

Nguyễn Phương Nam, Nguyễn Việt Hưng*

Trường Đại học Sư phạm Thành phố Hồ Chí Minh

*Email: hungnv@hcmue.edu.vn

Ngày nhận bài: 09/7/2019; Ngày chấp nhận đăng: 05/9/2019

TÓM TẮT

Cảm xúc là yếu tố quan trọng của con người trong giao tiếp xã hội. Trong giai đoạn công nghệ trí tuệ nhân tạo đang bùng nổ như hiện nay, phát triển hệ thống dự đoán cảm xúc của người qua khuôn mặt giúp ích rất nhiều trong giao tiếp người-máy. Phần lớn nghiên cứu hiện dựa trên nguồn dữ liệu ảnh và video màu thông thường, còn ít công trình dựa trên nguồn dữ liệu ảnh nhiệt. Ảnh nhiệt hiện đang có những đóng góp quan trọng trong lĩnh vực nhận dạng cảm xúc khi khắc phục những nhược điểm của ảnh thường như điều kiện ánh sáng, màu da, khuôn mặt không biểu cảm. Bài nghiên cứu này sẽ trình bày tổng quan các công trình về nhận dạng cảm xúc kết hợp ảnh thường và ảnh nhiệt. Tiếp theo là tổng quan về những nghiên cứu mới về ước lượng mức độ cảm xúc dựa trên ảnh nhiệt. Cuối cùng là xác định những thách thức, định hướng nghiên cứu trong lĩnh vực nhận diện cảm xúc và giao tiếp người-máy.

Từ khóa: Biểu cảm khuôn mặt, phân loại cảm xúc, ước lượng mức độ cảm xúc, ảnh nhiệt;

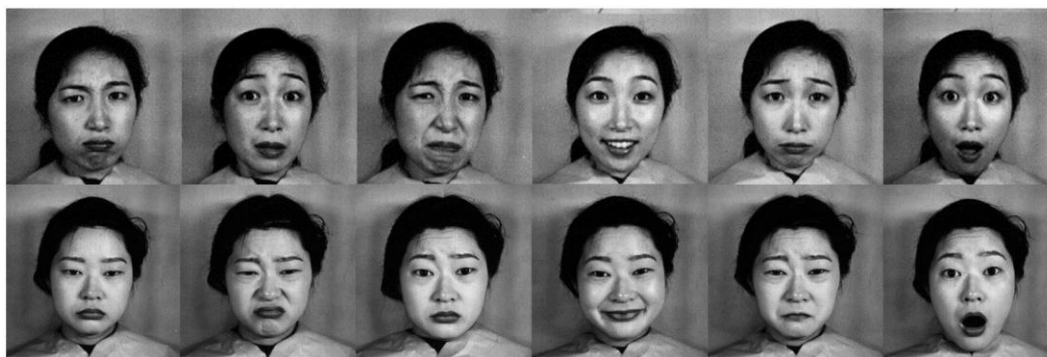
1. GIỚI THIỆU

Trong kỷ nguyên công nghệ số và sự phát triển của các ứng dụng cho công nghiệp 4.0, trao đổi thông tin giữa người và máy đóng một vai trò cực kì quan trọng. Hai nguồn thông tin chính giúp người và máy có thể trao đổi được một cách nhanh chóng và hiệu quả đó là giọng nói và hình ảnh [1-3]. Trong dữ liệu hình ảnh, những biểu hiện trên khuôn mặt chiếm 55% cơ sở để giúp con người có thể hiểu nhau trong quá trình trao đổi và nói chuyện [4]. Bên cạnh đó, rất dễ dàng cho con người có thể hiểu nhau thông qua biểu hiện khuôn mặt, nhưng đó lại là một bài toán khó đối với máy tính và robot. Do vậy, nghiên cứu về biểu hiện trên khuôn mặt được thu hút rất nhiều nghiên cứu và đạt được nhiều kết quả khích lệ [5-10].

Trong những năm gần đây, số lượng nghiên cứu trong lĩnh vực tương tác người-máy (Human-Computer Interaction/HCI) tăng nhanh và đa dạng trong các hướng tiếp cận. Thông thường, sự tương tác diễn ra khi con người thực hiện chỉ lệnh cho máy tính thông qua diễn tả bằng các bộ phận cơ thể khác nhau, trong đó có biểu cảm khuôn mặt. Biểu cảm trên khuôn mặt không chỉ là sự thể hiện đầu tiên trong tương tác xã hội hàng ngày mà còn là khu vực dễ nhận biết nhất trong giao tiếp không lời nói. Nhiều công trình nghiên cứu đã đề xuất các hướng tiếp cận vấn đề nhận dạng biểu cảm tự động qua khuôn mặt [11]. Phần lớn các nghiên cứu dựa trên sự phân loại các cảm xúc cơ bản của Paul Ekman [12-14]. Hình 1 minh họa sáu biểu cảm cơ bản theo nghiên cứu của Ekman, từ trái sang phải được gán nhãn là: giận dữ, ghê tởm, sợ hãi, hạnh phúc, buồn rầu, ngạc nhiên [15]. Hình 2 minh họa sáu biểu cảm của hai người (hàng 1 và hàng 2) trong bộ dữ liệu JAFFE, thứ tự biểu cảm tương tự Hình 1 [16].



Hình 1. Minh họa sáu cảm xúc cơ bản theo nghiên cứu của Ekman [15].



Hình 2. Biểu cảm khuôn mặt của hai người trong bộ dữ liệu JAFFE [16].

Tuy nhiên, hầu hết các phương pháp đề xuất chưa giải quyết triệt để vấn đề điều kiện ánh sáng, vì sự thay đổi cường độ sáng ảnh hưởng đáng kể đến chất lượng của hình ảnh được chụp hay quay bằng camera thông thường. Trong khi đó, camera nhiệt ghi nhận phân bố các vùng nhiệt độ khác nhau của hình ảnh và không phụ thuộc vào điều kiện chiếu sáng. Vì thế, một số nghiên cứu đã tiếp cận thực hiện các giải thuật nhận dạng cảm xúc trên ảnh nhiệt song song với ảnh thông thường như một giải pháp hỗ trợ nâng cao độ chính xác [17–19].

Cảm xúc là đối tượng nghiên cứu với những tranh luận kéo dài xuyên thế kỷ. Từ triết học tinh thần cổ điển thời Hy Lạp cổ đại đến lý thuyết tâm lý hiện đại, định nghĩa về cảm xúc và những đặc trưng của nó không đồng nhất với mỗi mục đích, lĩnh vực nghiên cứu khác nhau. Mối quan hệ giữa biểu cảm khuôn mặt và cảm xúc đã được trình bày và phân tích trong nhiều nghiên cứu đương đại. Cảm xúc thường được nhận định do sự tác động của hệ thần kinh tự chủ, nên dẫn đến những sự biến đổi tự nhiên các cơ trên khuôn mặt [20–23]. Luận điểm này được dẫn chứng với những người khiếm thị khi họ vẫn thể hiện biểu cảm hài lòng hay không hài lòng qua nét mặt, dù họ có thể chưa bao giờ thấy một biểu cảm khuôn mặt trên thực tế hay qua hình ảnh trước đó [24]. Như vậy, một người có thể phỏng đoán cảm xúc của người khác thông qua quan sát khuôn mặt của họ, cũng như nhiều nghiên cứu khoa học hiện nay đang chứng minh cho ý tưởng rằng khuôn mặt là một chỉ báo của cảm xúc. Tuy nhiên, một vấn đề đang tranh luận là khả năng nhận dạng cảm xúc qua khuôn mặt là do cơ chế sinh học tự nhiên hay suy luận từ diễn biến của môi trường xung quanh. Trong bài báo này, cảm xúc được xem là một trải nghiệm chủ quan, được tạo ra bởi sự thay đổi trong cách cơ thể phản ứng, bao gồm biểu cảm khuôn mặt và phản ứng sinh lý.

Nhận dạng cảm xúc là nghiên cứu cơ bản nhưng trong thực tế tâm lý và cảm xúc thay đổi thường xuyên tùy thuộc môi trường, hoàn cảnh, trải nghiệm sống và khả năng quản trị cảm xúc của mỗi người. Vì vậy, ước lượng mức độ thể hiện cảm xúc là nghiên cứu gắn với bản chất tự nhiên của con người. Các công trình nghiên cứu gần đây đã bước đầu cố gắng ước lượng mức độ của cảm xúc, hoặc suy luận độ lớn của một cảm xúc tại thời điểm xác định. Và ảnh nhiệt cũng được xem là một giải pháp hỗ trợ hướng nghiên cứu này.

Đóng góp của bài nghiên cứu gồm: (1) Trình bày tổng quan về nhận dạng cảm xúc của người dựa trên phân tích biểu cảm khuôn mặt; (2) Phân tích những kết quả đạt được trong hướng nghiên cứu ước lượng mức độ của biểu cảm khuôn mặt; (3) Đề xuất hướng nghiên cứu kết hợp với ảnh nhiệt là định hướng nghiên cứu tiềm năng, và sự cần thiết xây dựng bộ dữ liệu ảnh nhiệt chuẩn trong nhận dạng cảm xúc.

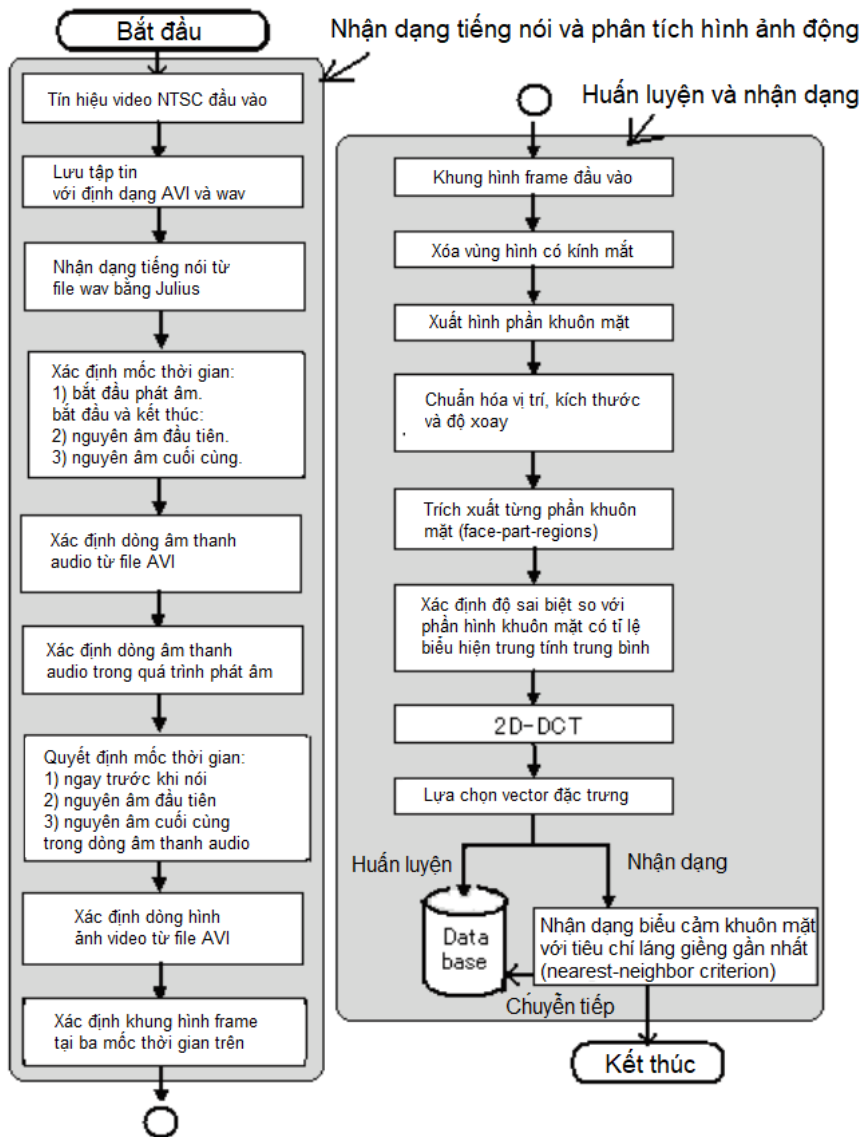
Phần còn lại của bài báo được tổ chức như sau: phần 2, giới thiệu các công trình liên quan đến các nghiên cứu phân loại, nhận dạng biểu cảm khuôn mặt; phần 3, mô tả các hướng tiếp cận chính hiện nay về ước lượng mức độ cảm xúc; phần 4, đưa ra kết luận và hướng phát triển.

2. NHẬN DẠNG BIỂU CẢM KHUÔN MẶT

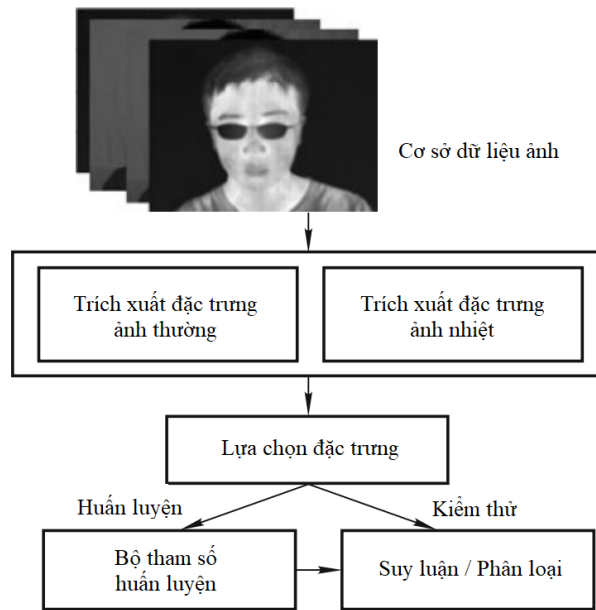
Phần nhiều những công trình gần đây nghiên cứu nhận dạng biểu cảm khuôn mặt người dựa trên phân tích ảnh chụp (visible image) và phim video thông thường. Tóm tắt bao quát về các phương pháp phân tích trên được nêu trong [10, 11, 25]. Vì thế, trong phần này của bài nghiên cứu chỉ trình bày cô đọng tổng quan về nhận dạng biểu cảm khuôn mặt dựa trên ảnh nhiệt (thermal image), và sự kết hợp đa dữ liệu với ảnh thông thường.

Yoshitomi và cộng sự đề xuất phân loại năm cảm xúc (giận dữ, hạnh phúc, bình thường, buồn rầu, và ngạc nhiên) dựa trên sự kết hợp ở mức quyết định các dữ liệu tiếng nói, ảnh thường và ảnh nhiệt [17]. Đầu tiên, nhóm tác giả thực hiện rút trích đặc trưng trên dữ liệu ảnh thường và ảnh nhiệt bằng phương pháp biến đổi cosin rời rạc (DCT), sau đó đưa vào huấn luyện bởi hai mô hình mạng neron để phân loại. Bên cạnh đó, mô hình thống kê Markov ẩn (HMM) được áp dụng với dữ liệu tiếng nói để phát hiện các cảm xúc. Cuối cùng, kết quả nhận dạng cảm xúc là kết hợp của ba kết quả phân loại trên với bộ trọng số đơn giản xác định trước. Thực nghiệm được tiến hành trên bộ dữ liệu có biểu cảm được sắp đặt (posed facial expression).

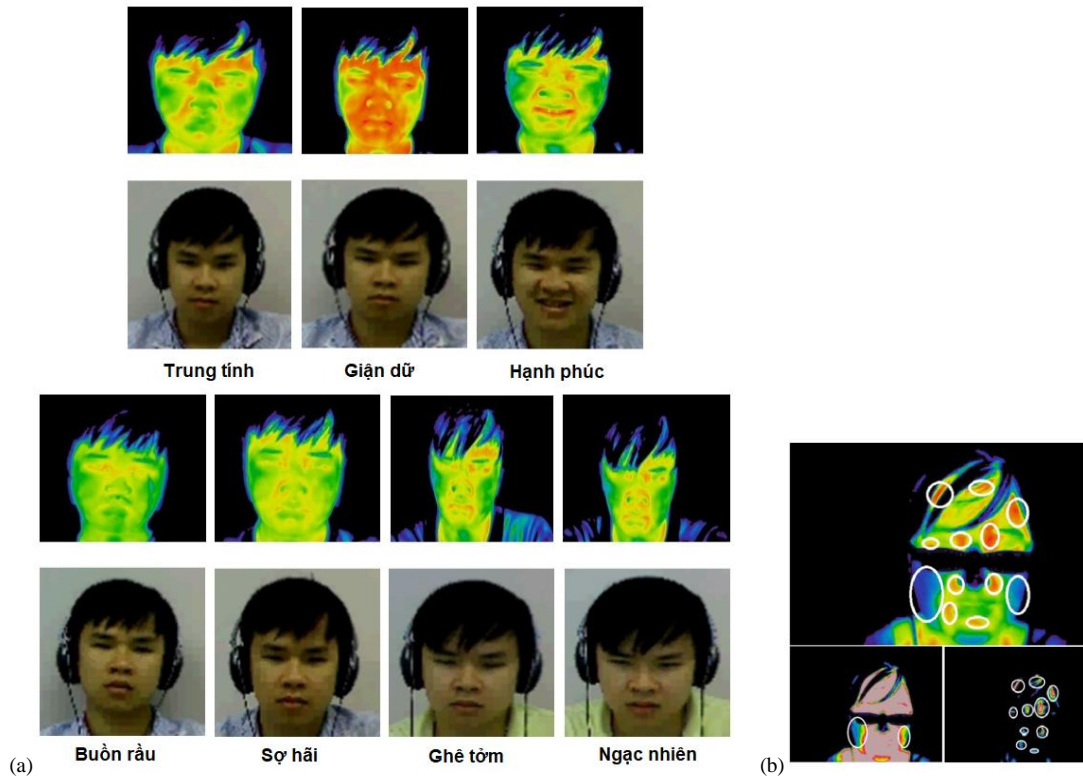
Nguyen.H và cộng sự đề xuất phương pháp kết hợp ảnh nhìn thấy và chuỗi ảnh nhiệt để ước lượng bảy cảm xúc cơ bản [19]. Vùng quan tâm (t-ROIs), được định nghĩa là vùng có biên độ tăng giảm nhiệt độ cao khi cảm xúc thay đổi, được rút trích từ chuỗi ảnh nhiệt. Đối với ảnh thường, phép biến đổi Wavelet (Wavelet Transform) được áp dụng để trích xuất đặc trưng. Nhóm tác giả thực hiện hai cách thức kết hợp: sau giai đoạn rút trích đặc trưng và ở giai đoạn quyết định phân loại. Các phương pháp được sử dụng khi thực hiện kết hợp tại mức quyết định gồm: phân tích thành phần chính (PCA), phương pháp không gian riêng dựa trên các đặc trưng của lớp (EMC) và kết hợp PCA-EMC. Thực nghiệm được tiến hành trên bộ dữ liệu ảnh nhiệt-cảm xúc Kotani (KTFE) cho kết quả nhận dạng cảm xúc tốt hơn so với sử dụng một kiểu dữ liệu [8].



Hình 3. Lưu đồ mô hình phương pháp đề xuất của Yoshitomi và cộng sự [17].



Hình 4. Mô hình đề xuất kết hợp ảnh thường và ảnh nhiệt [18]



Hình 5. (a) Ví dụ trong bộ cơ sở dữ liệu KTFE với bảy cảm xúc và hai loại ảnh của một người [8].
 (b) Ví dụ Vùng quan tâm trên ảnh nhiệt (t-ROIs) [19].

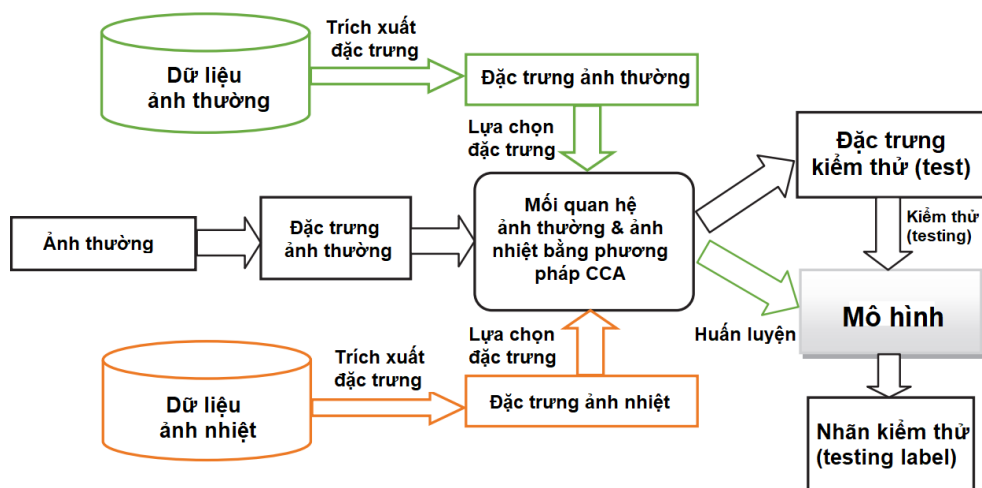
Liu và Wang phân tích dữ liệu chuỗi nhiệt độ phân bố trên khuôn mặt, tính toán thống kê và đưa ra biểu đồ khác biệt về nhiệt độ. Bên cạnh đó, nhóm tác giả sử dụng mô hình Markov ẩn (HMM) để nhận dạng các cảm xúc hạnh phúc, ghê tởm và sợ hãi với tỷ lệ chính xác tương

ứng là: 68,11%, 57,14% và 52,30%. Nhóm tác giả sử dụng các mẫu từ cơ sở dữ liệu NVIE (A Natural Visible and Infrared facial Expression Database) để đánh giá kết quả [6].

Shi và cộng sự đề xuất phương pháp nhận dạng cảm xúc từ hình ảnh nhìn thấy với sự trợ giúp của hình ảnh nhiệt [26]. Hình ảnh nhiệt được sử dụng trong quá trình huấn luyện. Nhóm tác giả xây dựng một không gian đặc trưng mới cho hình ảnh nhìn thấy bằng cách sử dụng phân tích tương quan chuẩn (CCA) với sự trợ giúp của hình ảnh nhiệt và dùng thuật toán SVM để phân loại cảm xúc. Phương pháp đề xuất này sử dụng hình ảnh nhiệt như thông tin đặc quyền. Điều này tối đa hóa tác động của hình ảnh nhiệt và nhận ra biểu hiện trong các tình huống thực tế mà không làm tăng chi phí thiết bị.

Wang và cộng sự đề ra phương pháp dựa trên tính chất của thuật toán SVM+ để phân loại cảm xúc [27]. Mục đích của nhóm tác giả nhằm giải quyết yêu cầu thực tế là chi phí cho camera nhiệt rất tốn kém. Theo đó, quá trình huấn luyện được thực hiện với cả dữ liệu ảnh thường và ảnh nhiệt, trong khi việc kiểm thử chỉ thực hiện với dữ liệu ảnh thường. Tuy kết quả phân loại chưa cao nhưng đây là một hướng phát triển khá mới và hứa hẹn.

Nguyen. Thu và cộng sự với hướng tiếp cận tương tự như Nguyen.H và cộng sự nhưng sử dụng không gian ảnh xám để cải tiến phương pháp tìm t-ROIs [19, 28]. Nhóm tác giả tiến hành thực nghiệm trên bộ dữ liệu KTFE với t-ROIs mới và các phương pháp kết hợp như PCA, EMC, PCA-EMC đã cải thiện và nâng cao độ chính xác trong việc nhận dạng và phân loại cảm xúc [8].



Hình 6. Mô hình phương pháp đề xuất kết hợp ảnh thường và ảnh nhiệt của Shi và cộng sự [26].

Những công trình nghiên cứu trên đây đã phần nào chứng minh phương thức kết hợp hình ảnh nhìn thấy thông thường và hình ảnh nhiệt giúp tăng tỷ lệ chính xác cho nhận dạng biểu cảm khuôn mặt và cảm xúc.

3. ƯỚC LƯỢNG MỨC ĐỘ CẢM XÚC

Mục đích của việc phân tích biểu cảm khuôn mặt là nhận dạng theo thời gian thực sáu biểu cảm cơ bản khi đưa vào một ảnh hay một chuỗi ảnh khuôn mặt. Các công trình nêu trên đã đạt những kết quả hứa hẹn trong các điều kiện thực nghiệm nhất định. Tuy nhiên, việc phân loại biểu cảm này chưa đủ để hiểu sâu hơn về cảm xúc con người một cách tự nhiên. Đôi khi các ứng dụng trong thế giới thực cần biết thêm những biểu hiện đó được thể hiện ở mức độ nào: thoáng qua, vừa phải hay quá mức [15]. Một số nghiên cứu tâm lý đã chỉ ra bên

cạnh những biểu cảm đơn thuần có thể thấy rõ, những thay đổi nét mặt tự nhiên theo thời gian cũng rất quan trọng trong việc giải mã ý nghĩa cảm xúc tại một thời điểm [29]. Hiểu một cách đơn giản, những thay đổi đó chính là sự thay đổi mức độ cảm xúc theo thời gian của khuôn mặt người.

Ước lượng mức độ biểu cảm có nhiều ứng dụng tiềm năng trong các lĩnh vực khác nhau như tương tác người-máy, theo dõi tình trạng bệnh nhân, giám sát an ninh và trong lĩnh vực thương mại, giải trí. Ví dụ, mức độ biểu cảm giúp một robot thông minh nhận biết cảm xúc của người sử dụng để có thể phản ứng phù hợp hoặc thậm chí dự đoán được mệnh lệnh sắp được đưa ra.

Bên cạnh các công trình nghiên cứu về phân loại cảm xúc cơ bản thì còn ít các công trình nghiên cứu về mức độ cảm xúc. Hiện nay chủ yếu được chia thành hai nhóm dựa trên bộ cơ sở dữ liệu thực nghiệm: dữ liệu có gán nhãn mức độ và không có gán nhãn.

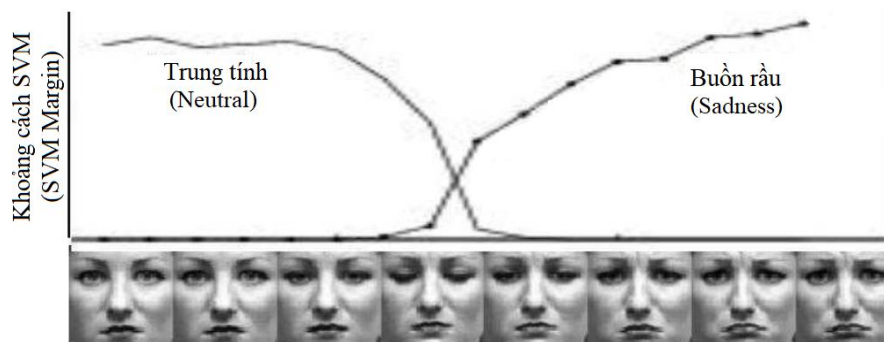


Hình 7. Ảnh tuần tự của biểu hiện ngạc nhiên từ mức độ thấp tới cao [15].

3.1. Dữ liệu mức độ cảm xúc không gán nhãn

Với nhóm dữ liệu không gán nhãn, các tác giả chủ yếu mở rộng việc phân loại biểu cảm khuôn mặt để ước lượng mức độ cảm xúc. Littlewort và cộng sự áp dụng thuật toán SVM nhận diện biểu cảm và mức độ cảm xúc được xác định là khoảng cách từ biểu cảm đó đến siêu phẳng (hyperplane) tìm được [30]. Chang và cộng sự sử dụng mô hình huấn luyện đặc trưng tách biệt (discriminative manifold learning) xác định thành phần khuôn mặt và ước lượng mức độ dựa trên khoảng cách của biểu cảm đến thành phần đặc trưng đó [31]. Yoshitomi và cộng sự thực hiện trích xuất từng khung ảnh nhiệt từ chuỗi ảnh động mỗi 0.1 giây, từ đó xác định vector đặc trưng là vùng quan tâm của phần miệng và hàm bằng phép biến đổi cosine rời rạc hai chiều (2D-DCT) cho mỗi vùng ảnh có kích thước 8×8 pixel [32]. Vector đặc trưng có 15 giá trị và mức độ biểu cảm khuôn mặt được xác định là khoảng cách giữa vector đặc trưng của biểu cảm trung tính (neutral) với các biểu cảm còn lại.

Tuy nhiên, các nghiên cứu về ước lượng mức độ này chỉ thực hiện trên dữ liệu được gán nhãn theo biểu cảm (không gán nhãn mức độ), đây chính là có điểm hạn chế trong việc học các luật phân loại. Bên cạnh đó, phương pháp xác định mức độ được ước lượng dựa trên tỉ lệ khoảng cách giữa mẫu quan sát và đường biên phân loại (classification boundary) trong không gian đặc trưng là không đủ để phản ánh biểu cảm hay cảm xúc đang ở mức nào nên kết quả thực nghiệm không cao.



Hình 8. Một kết quả nhận diện bằng SVM cho cảm xúc bình thường và buồn rầu của Chang và cộng sự [31].

3.2. Dữ liệu mức độ cảm xúc có gán nhãn

Với nhóm dữ liệu có gán nhãn. Nhãn dữ liệu được biểu diễn dưới hai dạng: mức xếp hạng có giá trị liên tục và các mức rời rạc. Một số phương pháp hồi quy phi tuyến (nonlinear /ordinal regression) được áp dụng trên nhóm dữ liệu này và đã cho kết quả ước lượng cao hơn so với nhóm dữ liệu không gán nhãn.

Nhằm ước tính giá trị liên tục của nhãn, các nghiên cứu thường cố gắng đưa ra một hàm số để tính giá trị của mức độ dựa theo kết quả của bài toán phân loại, nhận dạng cảm xúc. Liao và cộng sự đề xuất hướng huấn luyện nhằm xây dựng các nhân (kernel) áp dụng cùng phương pháp máy học kernel (kernel machine) để ước lượng mức độ biểu cảm khuôn mặt [33]. Tác giả thiết lập quá trình huấn luyện như một bài toán học đa kernel (Multi Kernel Learning-MKL), trong đó kỹ thuật ước lượng di chuyển (optical flow) trong nội tại đối tượng đang xét được sắp đặt để giảm thiểu độ lệch giữa các lớp đối tượng với nhau. Kernel được học có thể đo lường sự đồng dạng giữa hình ảnh khuôn mặt người cần xác định với khuôn mặt có biểu cảm, nhờ đó việc phân tích biểu cảm của khuôn mặt thuận lợi hơn. Mục tiêu của nhóm tác giả là bằng phương pháp MKL, tìm ra kernel cho một biểu cảm có cường độ tương tự nhau, và có giá trị lớn hơn so với kernel của các biểu cảm hoặc cường độ biểu cảm khác.

Song và cộng sự nhận thấy biểu cảm khuôn mặt là mơ hồ (fuzzy), khó xác định trong việc phân loại và xác định mức độ [34]. Phân loại một cách đơn giản biểu cảm khuôn mặt thành các loại cảm xúc cơ bản khác nhau là chưa đầy đủ. Vì thế, nhóm tác giả đề xuất một hệ thống nhận dạng biểu cảm với sự ước lượng tỷ lệ và mức độ. Hệ thống này sử dụng đặc trưng kết cấu bề mặt (texture) được trích xuất bằng mô hình xuất hiện tích cực (AAM) và hệ thống mã hóa hành động mặt (FACS) để nhận diện một cách hỗn hợp biểu cảm khuôn mặt. Trong phân nhận dạng, một hệ thống được phát triển nhằm tính toán giá trị tỷ lệ hỗn hợp của biểu cảm, sau đó sử dụng sáu mô hình mạng neural lan truyền ngược phân lớp một ảnh đưa vào thuộc sáu biểu cảm cơ bản cùng một giá trị đánh giá mức độ. Kết quả cuối cùng là kết hợp các giá trị trên với bộ trọng số xác định.

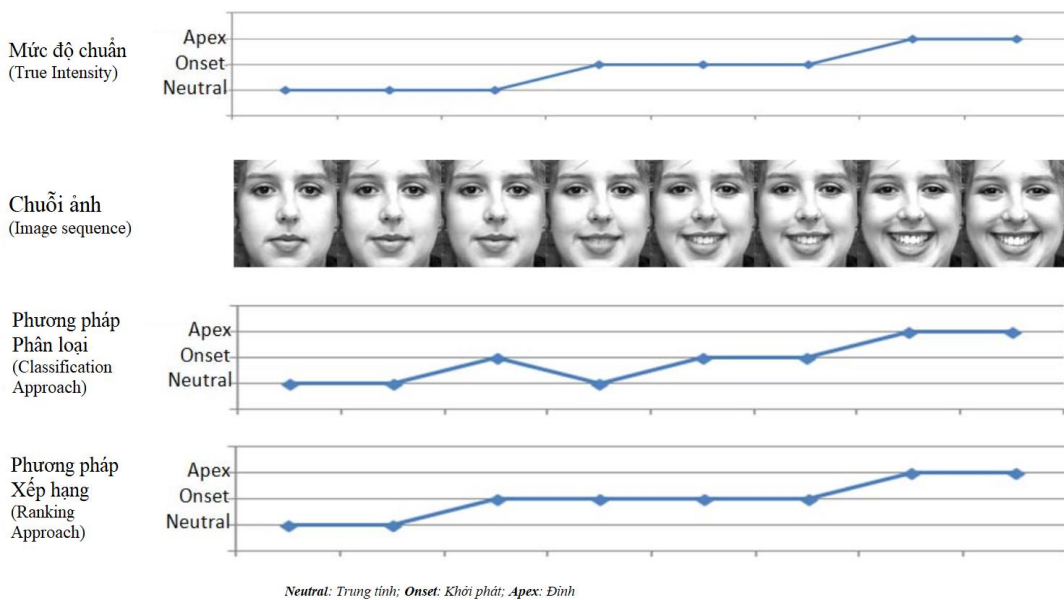
Với mức xếp hạng rời rạc, một số nhà nghiên cứu cho rằng mức độ biểu cảm có được sự ước lượng gần với bản chất tự nhiên của con người hơn so với giá trị cách đánh giá bằng giá trị mức độ liên tục.

Kim và cộng sự nhận định rằng việc gán nhãn mức độ cảm xúc khuôn mặt trong một đoạn phim có giá trị sắp xếp trên một thang đo có thứ tự (ordinal scale) thay đổi theo thời gian, ví dụ: thấp, trung bình và cao [35]. Nhóm tác giả đề xuất một mô hình xếp hạng linh hoạt, theo đó tín hiệu cường độ tại mỗi thời điểm là một nhãn trên thang đo với các mốc thời gian tương đối có các điều kiện ràng buộc linh hoạt (smoothness constraints). Mô hình mới

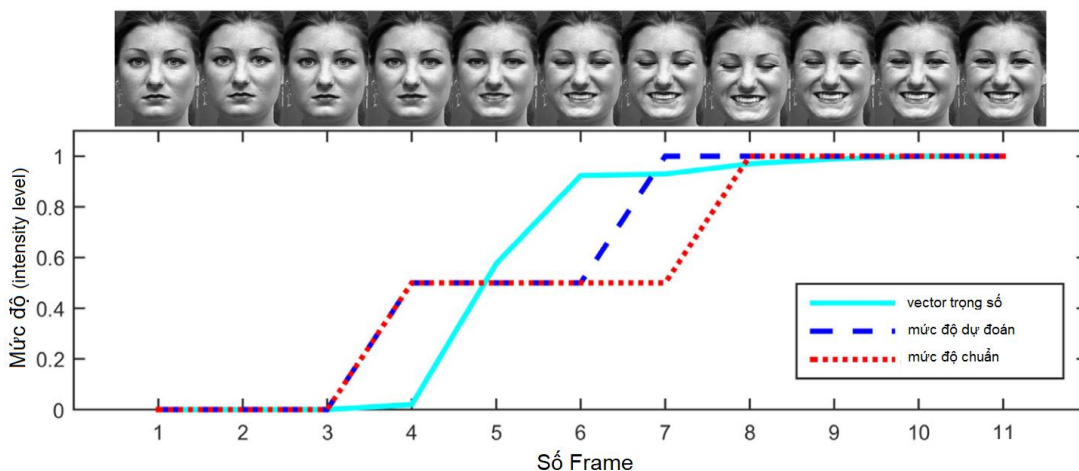
này là sự mở rộng của phương pháp hồi quy thứ tự tĩnh (static ordinal regression) cho dữ liệu có cấu trúc tương tự như mô hình trường điều kiện ngẫu nhiên (Conditional Random Field-CRF) thường sử dụng cho bài toán dự đoán phân lớp có cấu trúc. Kết quả dự đoán của mô hình xếp hạng linh hoạt trên cho thấy sự cải tiến đáng kể so với mô hình CRFs chuẩn, vì CRFs chuẩn không thành công trong việc đánh giá mối quan hệ thứ tự giữa các nhân được dự đoán.

Delannoy và cộng sự đề xuất mô hình với ba mức độ để ước lượng biểu cảm (thấp, trung bình, và cao) sử dụng thuật toán giảm chiều dữ liệu phi tuyến (Locally Linear Embedding) [36]. Tương tự, Chang và cộng sự xử lý việc ước lượng mức độ bằng đề xuất xếp hạng các ảnh đơn đầu vào thành ba mức độ cảm xúc (thấp, trung bình, cao) [15]. Nhóm tác giả áp dụng thuật toán RED-SVM để xây dựng bộ khung ước lượng mức độ cảm xúc có ba mức, trích xuất vector đặc trưng bằng cách kết hợp hình khuôn mặt chính diện và đặc trưng phân tán (phép biến đổi phân tán – scattering transforms), sau đó giảm chiều dữ liệu bằng thuật toán PCA. Nhờ được huấn luyện trên bộ dữ liệu có gán nhãn nên mức độ biểu cảm khuôn mặt được nhận diện và kiểm chứng với độ lỗi thấp so với một số phương pháp trước đây.

Kamarol và cộng sự đề xuất khung (framework) nhận dạng biểu cảm khuôn mặt và ước lượng mức độ với yêu cầu tính toán có độ phức tạp thấp [37]. Thuật toán rút trích đặc trưng được xây dựng dựa trên thuật toán k láng giềng gần nhất (kNN) và sơ đồ trọng số xác định, sau đó áp dụng mô hình Markov ẩn để phân loại đoạn video đầu vào kết xuất ra sáu biểu cảm cơ bản. Phân đoạn thời gian, biểu cảm trung tính, điểm bắt đầu và đỉnh của biểu cảm được xác định bằng cách lấy trung bình các thay đổi. Mức được ước lượng dựa trên sự thay đổi của giá trị đỉnh (apex) của biểu cảm đang quan sát so với biểu cảm trung tính. Thuật toán được thực nghiệm phân tích các đối tượng độc lập trên bộ dữ liệu CK (Cohn-Kanade) và BU (Beihang University).



Hình 9. So sánh kết quả ước lượng mức độ của hai phương pháp tiếp cận: phân loại (classification) và xếp hạng (ranking) Chang và cộng sự [15].



Hình 10. So sánh kết quả ước lượng mức độ của hai hướng tiếp cận: phân loại (classification) và xếp hạng (ranking). Ba mức phân biệt tương ứng theo một khoảng thời gian: bình thường (neutral), điểm bắt đầu (onset), điểm đỉnh (apex) của Kamarol và cộng sự [37].

Sabri và cộng sự thực nghiệm mô hình mạng Siamese và Triplet trên dữ liệu ảnh tuần tự nhằm ước lượng mức độ cảm xúc [38]. Phương pháp đề xuất dựa trên mô hình xếp hạng (ranking model) để tìm mối quan hệ thứ tự giữa các dữ liệu ảnh tuần tự và xác định tất cả những thay đổi liên quan đến mức độ bằng cách kết hợp hai mạng trên. Mô hình huấn luyện được học thực hiện việc xác định vị trí (object localization) và những AU (action unit) phân biệt mà không cần đánh dấu khung (bounding box). Mô hình mạng này được đánh giá cho kết quả tốt hơn các mạng CNN trước đây, cũng như riêng mạng Triplet trích xuất, xác định đặc trưng rõ và chính xác hơn trong quá trình huấn luyện. Kết quả được thực nghiệm trên các bộ dữ liệu cảm xúc khác nhau (CK, MUG, MMI, CASME) cho kết quả tốt trong việc ước lượng mức độ và biểu cảm (micro-expression).

4. HẠN CHẾ CỦA PHƯƠNG PHÁP SỬ DỤNG ẢNH NHIỆT

Như đã trình bày, ảnh nhiệt có ưu điểm trong điều kiện ánh sáng yếu, khuôn mặt không thể hiện biểu cảm (poker-face) hay có màu da sẫm nhưng với bài toán nhận dạng và ước lượng cảm xúc vẫn có những hạn chế sau: (1) dữ liệu thông tin nhiệt trên khuôn mặt sẽ không chính xác khi nhiệt độ môi trường thay đổi với khoảng độ lớn vì khi đó nhiệt độ toàn cơ thể người cũng thay đổi; (2) khi đối tượng đeo kính râm, camera hồng ngoại thường không thu được chính xác nhiệt độ ở khu vực mắt dẫn đến làm mất thông tin nhiệt xung quanh khu vực rất hữu ích này; (3) có những vùng trên khuôn mặt mà sự thay đổi nhiệt độ tại đó không tương quan với sự thay đổi của cảm xúc.

Ngoài ra, các biểu hiện ngạc nhiên, giận dữ, ghê tởm...có thể có cùng một màu ảnh nhiệt (biểu hiện màu đỏ) nên sử dụng ảnh nhiệt có thể gây nên sự trùng lặp dễ nhận diện các loại cảm xúc khác nhau. Cảm xúc của con người là trạng thái tâm lý phức tạp rất riêng của con người với hàng triệu biểu hiện sắc thái khác nhau kèm theo đó là hàng triệu các tín hiệu biểu hiện khác nhau trên khuôn mặt, nên việc chỉ nhận diện một số biểu cảm khuôn mặt thông qua ảnh nhiệt thì chưa phản ánh hết các sắc thái biểu cảm. Do đó phương pháp này cần có sự kết hợp của nhiều phương pháp khác như phân tích hành vi tổng thể (dựa vào camera), trải nghiệm tâm lý; kết hợp đặc trưng từ các nguồn dữ liệu như hình ảnh ba chiều, giọng nói, điện não đồ (EEG)...

5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Bài viết đã trình bày tổng quan: (1) về ước lượng cảm xúc của người dựa trên phân tích biểu cảm khuôn mặt, những kết quả đạt được về sự phân loại cảm xúc bằng phương pháp kết hợp hai nguồn dữ liệu ảnh khuôn mặt: ảnh nhìn thấy thông thường và ảnh nhiệt; (2) về hướng nghiên cứu ước lượng mức độ của biểu cảm khuôn mặt hay cảm xúc của người, hướng tiếp cận này được chia thành hai nhóm: nhóm thứ nhất sử dụng mẫu dữ liệu được gán nhãn phân lớp, nhưng thiếu nhãn về mức độ; nhóm thứ hai thực hiện trên mẫu dữ liệu có gán nhãn mức độ biểu cảm.

Qua đó, nghiên cứu này đã phân nào đưa ra một số định hướng và ý tưởng nghiên cứu hứa hẹn trong tương lai. Một, xây dựng một bộ dữ liệu kết hợp với hai mẫu dữ liệu ảnh nhìn thấy và ảnh nhiệt thể hiện cảm xúc tự nhiên, nhằm khắc phục những hạn chế của ảnh thường trong điều kiện ánh sáng yếu. Hai, xây dựng mô hình phân tích, dự đoán xếp hạng ước lượng mức độ cảm xúc dựa trên dữ liệu ảnh nhiệt kết hợp đa kỹ thuật phân tích. Ba, tìm hiểu mối quan hệ tương quan giữa biểu cảm với sự thay đổi nhiệt độ của những khu vực xác định trên khuôn mặt, rộng hơn là mối quan hệ giữa cảm xúc và nhiệt độ nhằm góp phần xây dựng một hệ thống hoàn chỉnh về ước lượng cảm xúc.

TÀI LIỆU THAM KHẢO

1. Bartneck C. - How convincing is mr. data's smile: affective expressions of machines, *User Modeling and User-Adapted Interaction* **11** (4) (2001) 279–295.
2. Books P. - *Flesh and Machines. How Robots Will Change Us.*, Nelson (2002) 1–5.
3. Picard R. - *Affective Computing*, MIT Press (1997).
4. Mehrabian A. - *Nonverbal Communication*, Taylor and Francis (2017).
5. Cohen I., Sebe N., Garg A., Chen L. S., Huang T. S. - Facial expression recognition from video sequences: temporal and static modeling, *Computer Vision and Image Understanding* **91** (2003) 160–187.
6. Liu Z., Wang S. - Emotion recognition using hidden markov models from facial temperature sequence, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2011) 240–247.
7. Martinez B., F. Valstar M., Jiang B., Pantic M. - Automatic analysis of facial actions: A Survey, *IEEE Transactions on Affective Computing* (2017).
8. Nguyen H., Kotani K., Chen F., Le B. - A thermal facial emotion database and its analysis, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2014) 397–408.
9. Corneanu C. A., Simón M. O., Cohn J. F., Guerrero S. E. - Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38** (2016) 1548–1568.
10. Sariyanidi E., Gunes H., Cavallaro A. - Automatic analysis of facial affect: a survey of registration, representation, and recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37** (2015) 1113–1133.
11. Bettadapura V. - Face expression recognition and analysis: The State of the Art, *arXiv preprint arXiv:1203.6722* (2012) 1–27.
12. Ekman P. - Universals and cultural differences in facial expressions of emotion, *Nebraska Symposium on Motivation* **19** (1971) 207–283.

13. Ekman P. - Basic Emotions, in Handbook of cognition and emotion (1999) Vol. 98 45–60.
14. Pantic M., Rothkrantz L. Ü. M. - Automatic Analysis of Facial Expressions: The State of the Art, IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 1424–1445.
15. Chang K. Y., Chen C. S., Hung Y. P. - Intensity rank estimation of facial expressions based on a single image, in Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013 (2013).
16. Lyons M. J., Akamatsu S., Kamachi M., Gyoba J., Budynek J. - The Japanese Female Facial Expression (JAFFE) Database, Proceedings of third international conference on automatic face and gesture recognition (1998) 14–16.
17. Yoshitomi Y. - Facial expression recognition for speaker using thermal image processing and speech recognition system, in International Conference on Applied Computer Science - Proceedings (2010) 182–186.
18. Wang S., He S., Wu Y., He M., Ji Q. - Fusion of visible and thermal images for facial expression recognition, Frontiers of Computer Science **8** (2014) 232–242.
19. Nguyen H., Chen F., Kotani K., Le B. - Human emotion estimation using wavelet transform and T-ROIs for fusion of visible images and thermal image sequences, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2014) 224–235.
20. Kreibig S. D. - Autonomic nervous system activity in emotion: a review, Biological Psychology **84** (2010) 394–421.
21. Ekman P., Levenson R. W., Friesen W. V. - Autonomic nervous system activity distinguishes among emotions, Science **221** (1983) 1208–1210.
22. Ekman P. - Expression and the Nature of Emotion, in Approaches to Emotion (1984) 319–344.
23. Ekman P. - All Emotions Are Basic, in The Nature of Emotion (1994) 15–19.
24. Matsumoto D., Willingham B. - Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals, Journal of Personality and Social Psychology **96** (1) (2009) 1–10.
25. Zeng Z., Pantic M., Roisman G. I., Huang T. S. - A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (1) (2009) 39–58.
26. Shi X., Wang S., Zhu Y. - Expression recognition from visible images with the help of thermal images, in ICMR 2015 - Proceedings of the 2015 ACM International Conference on Multimedia Retrieval (2015) 563–566.
27. Wang S., Pan B., Chen H., Ji Q. - Thermal augmented expression recognition, IEEE Transactions on Cybernetics **48** (7) (2018) 2203–2214.
28. Nguyen T., Tran K., Nguyen H. - Towards thermal region of interest for human emotion estimation, in Proceedings of 2018 10th International Conference on Knowledge and Systems Engineering, KSE 2018 (2018) 152–157.
29. Ambadar Z., Schooler J. W., Conn J. F. - Deciphering the enigmatic face the importance of facial dynamics in interpreting subtle facial expressions, Psychological Science **16** (5) (2005) 403–410.
30. Littlewort G., Bartlett M. S., Fasel I., Susskind J., Movellan J. - Dynamics of Facial Expression Extracted Automatically from Video, Image and Vision Computing **24** (6) (2006) 615–625.
31. Chang W. Y., Chen C. S., Hung Y. P. - Analyzing facial expression by fusing

- manifolds, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2007) 621–630.
32. Yoshitomi Y., Tabuse M., Asada T. - Facial expression recognition using facial expression intensity characteristics of thermal image, *Journal of Robotics, Networking and Artificial Life* **2** (1) (2015) 5–8.
 33. Liao C. Te, Chuang H. J., Lai S. H. - Learning expression kernels for facial expression intensity estimation, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2012) 2217–2220.
 34. Song K. T., Chien S. C. - Facial expression recognition based on mixture of basic expressions and intensities, in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* (2012) 3123–3128.
 35. Kim M., Pavlovic V. - Structured output ordinal regression for dynamic facial emotion intensity prediction, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2010) 649–662.
 36. Delannoy J. R., McDonald J. - Automatic estimation of the dynamics of facial expression using a three-level model of intensity, in *2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008* (2008) 1–6.
 37. Kamarol S. K. A., Jaward M. H., Kälviäinen H., Parkkinen J., Parthiban R. - Joint facial expression recognition and intensity estimation based on weighted votes of image sequences, *Pattern Recognition Letters* **92** (2017) 25–32.
 38. Sabri M., Kurita T. - Facial expression intensity estimation using siamese and triplet networks, *Neurocomputing* **313** (2018) 143–154.

ABSTRACT

HUMAN EMOTION INTENSITY ESTIMATION: A SURVEY OF CATEGORIZATION AND TOWARDS THERMAL IMAGERY

Nguyen Phuong Nam, Nguyen Viet Hung*

Ho Chi Minh City University of Education

*Email: hungnv@hcmue.edu.vn

Emotion is an important factor for social interaction. In the era of the explosion of artificial intelligence, developing a system that can automatically estimate human emotions has attracted enormous interest to researchers. Understanding inner emotion makes human-computer interaction applications be able to join mankind life extensively. Most of human emotion estimation works focus on categorizing emotion based on visible images or videos of facial expression, but only few works address the use of infrared thermal imaging for emotion estimation, which has significant advantages in the field. In addition, we still lack a good survey of human emotion estimation using infrared thermal imaging. Therefore, in this paper, we concentratedly survey about human emotion estimation exploiting both visible and infrared thermal imaging. Firstly, we want to give a review of the most recent studies about human emotion estimation, which include emotion categorization and emotion intensity estimation. Secondly, we identify several challenges in this field and suggest some idea for future research.

Keywords: Facial expression, emotion categorization, emotion estimation, emotion intensity, visible image, thermal image.