

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG**  
**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**TÓM TẮT LUẬN ÁN**

**PHÁT TRIỂN CÁC KỸ THUẬT BIỂU DIỄN  
BIỂU CẢM KHUÔN MẶT 3D**

**NCS: HUỖNH CAO TUẤN**

**NGƯỜI HƯỚNG DẪN KHOA HỌC**

- 1. PGS. TS. Đỗ Năng Toàn**
- 2. TS. Nguyễn Thanh Bình**

Hà Nội – 2021

## MỞ ĐẦU

Nghiên cứu về mặt người là một trong những hướng nghiên cứu được rất nhiều sự quan tâm của các nhà khoa học trên thế giới cũng như trong nước. Các nghiên cứu trong lĩnh vực này được chia thành hai hướng chính: Thứ nhất, liên quan đến thị giác máy, phát hiện nhận dạng mặt người và các trạng thái biểu cảm của mặt người. Thứ hai, liên quan đến mô phỏng, biểu diễn mô hình mặt người cùng với các trạng thái khác nhau.

Các ứng dụng của biểu diễn biểu cảm khuôn mặt là rất đa dạng như: máy tính tương tác với người sử dụng như một người bình thường với đầy đủ các cảm xúc vui buồn giận dữ vv... Một số ứng dụng trong cuộc sống có liên quan đến biểu cảm khuôn mặt có thể kể đến: Hệ thống giám sát và nhận dạng khuôn mặt, hệ thống tìm ảnh dựa trên nội dung, xây dựng các khuôn mặt nhân vật đóng thế trong điện ảnh, các hệ thống kiểm soát vào ra, các hệ thống xác thực bằng sinh trắc học khuôn mặt...

Mặc dù đã có rất nhiều công trình nghiên cứu về biểu cảm và thể hiện biểu cảm khuôn mặt nhưng chưa có nghiên cứu nào thực sự hoàn thiện và phù hợp cho nhiều lớp bài toán khác nhau, mỗi hướng tiếp cận chủ yếu được đưa ra để giải quyết một vấn đề nào đó được đặt ra.

### **Đối tượng, Mục tiêu nghiên cứu**

#### *a) Đối tượng nghiên cứu*

Đối tượng khuôn mặt người 2D trong xử lý ảnh và mô hình mặt người 3D trong thực tại ảo, Các kỹ thuật xử lý ảnh và xử lý video liên quan đến phát hiện và trích chọn các điểm đặc trưng trên khuôn mặt người. Các kỹ thuật nội suy liên quan đến thể hiện các cử chỉ, trạng thái biểu cảm của mô hình mặt người 3D

#### *b) Mục tiêu và phạm vi nghiên cứu*

Bài toán đặt ra trong nghiên cứu này là làm sao để tái thể hiện lại biểu cảm hoặc cử chỉ của một người thật lên một mô hình khuôn mặt 3D trên máy tính. Bài toán này có 2 giai đoạn chính: Thứ nhất là thu nhận và trích chọn các đặc trưng biểu cảm hoặc cử chỉ trên khuôn mặt; Thứ hai là từ những thông tin rút trích được ở giai đoạn một, sẽ là đầu vào của giai đoạn hai để tái thể hiện lại trên mô hình mặt người 3D.

Mục tiêu của luận án là tập trung vào nghiên cứu phát triển các kỹ thuật xử lý ảnh, xử lý video, trích rút, ghi nhận các đặc trưng thể hiện các cử chỉ, trạng thái biểu cảm của mặt người trong video. Các kỹ thuật nội suy nhằm điều khiển mô hình 3D thể hiện các trạng thái biểu cảm của khuôn mặt. Cụ thể là các kỹ thuật phân đoạn

video, kỹ thuật rút trích các đặc trưng cử chỉ trên khuôn mặt, Mô hình định vị điểm điều khiển Active Appearance Model (AAM), kỹ thuật nội suy Hàm cơ sở bán kính RBF và tham khảo một số kỹ thuật liên quan khác...

Để giải quyết các bài toán đã nêu trên, luận án đề xuất xây dựng một hệ thống rút trích các cử chỉ, biểu cảm khuôn mặt rồi mô phỏng lại trên máy tính. Hệ thống biểu diễn biểu cảm mà luận án nghiên cứu có đầu vào là trạng thái cảm xúc liên tục, đầu ra là biểu cảm của khuôn mặt ảo thể hiện trạng thái cảm xúc đó.

### **Các đóng góp của luận án**

- Lựa chọn kỹ thuật phát hiện khuôn mặt trong ảnh dựa trên ý tưởng cơ bản là thực hiện phân lớp nhị phân với mỗi vùng ảnh quan tâm kết hợp với kỹ thuật phân đoạn video dựa trên trừ ảnh đặc trưng. Kết quả nghiên cứu được công bố trong công trình TCTN3 và được trình bày trong Chương 2 của luận án.

- Đề xuất kỹ thuật định vị các điểm điều khiển dựa trên Mô hình ngoại hình chủ động (AAM) và ước lượng biểu cảm khuôn mặt đã được định vị tập điểm điều khiển. Thay vì thống kê và lựa chọn rời rạc một số đặc trưng hình dạng bằng tay, nghiên cứu sinh tiến hành tự động lựa chọn các đặc trưng hình học một cách ngẫu nhiên và tổ chức dưới mô hình cây quyết định để thực hiện ước lượng các biểu cảm khuôn mặt. Kết quả nghiên cứu được công bố trong các công trình TCQT1, TCTN1 và được trình bày trong Chương 2 của luận án.

- Đề xuất sử dụng 3 loại đặc trưng hình dạng là LINE\_LINE, TRIANGLE\_TRIANGLE, LINELINE\_LINELINE. Lý do chọn 3 loại đặc trưng hình dạng này là vì các đặc trưng trên khuôn mặt tuy khác nhau nhưng luôn có một mối liên hệ gắn liền nhau tạo thành các đường hay tam giác đi theo một mô hình tương tự nhau. Ví dụ vị trí của mắt mũi miệng là luôn tương quan với nhau. Kết quả nghiên cứu được công bố trong các công trình TCQT1, TCTN1 và được trình bày trong Chương 2 của luận án.

- Đề xuất kỹ thuật tự động xác định tập điểm điều khiển để phục vụ nắn chỉnh biến dạng mô hình dựa trên việc phân tích một tập các mô hình quan sát của đối tượng quan tâm. Kết hợp với thuật toán nắn chỉnh biến dạng và một tập các mô hình biến thể để phục vụ đánh giá chất lượng nắn chỉnh. Kết quả nghiên cứu được công bố trong các công trình TCQT2, TCQT3, HNTN1 và được trình bày trong Chương 3 của luận án.

### **Bố cục luận án**

Bố cục của Luận án bao gồm những phần như sau:

Phần Mở đầu: Giới thiệu khái quát về ý nghĩa của nội dung nghiên cứu, giới thiệu bài toán và cách giải quyết vấn đề. Đồng thời nêu khái quát những đóng góp của luận án.

Chương 1: Là chương tổng quan, trong chương này luận án sẽ trình bày các khái niệm về biểu cảm và một số vấn đề cơ bản trong hệ thống biểu cảm khuôn mặt. Đồng thời trình bày một số hướng nghiên cứu, cách tiếp cận và các phương pháp được sử dụng để giải quyết 2 vấn đề: Một là bài toán trích chọn và ước lượng biểu cảm khuôn mặt người. Hai là bài toán biểu diễn biểu cảm khuôn mặt người.

Chương 2: Kỹ thuật trích chọn đặc trưng cử chỉ và ước lượng biểu cảm : Trong chương này, luận án trình bày 3 bài toán: một là phát hiện khuôn mặt người trong ảnh; hai là trích chọn đặc trưng cử chỉ của khuôn mặt và cuối cùng là bài toán ước lượng biểu cảm khuôn mặt. Chương này luận án cũng đề xuất kỹ thuật phát hiện mặt người, kế thừa Mô hình AAM cải tiến và bổ sung thuật toán định vị các điểm điều khiển để rút trích cử chỉ và đề xuất kỹ thuật ước lượng biểu cảm khuôn mặt.

Chương 3: Biểu diễn cử chỉ, biểu cảm trên khuôn mặt 3D: Trong chương này, luận án giới thiệu chi tiết bài toán mô phỏng cử chỉ khuôn mặt, các phương pháp tiếp cận biểu diễn khuôn mặt 3D, các kỹ thuật nội suy và xác định tập điểm điều khiển để phục vụ quá trình mô phỏng cử chỉ, biểu cảm.

Phần kết luận và hướng phát triển của Luận án sẽ trình bày tổng kết lại những nội dung nghiên cứu đã đạt được của Luận án, đồng thời đề xuất các hướng nghiên cứu tiếp theo của Luận án.

## **Chương 1**

# **TỔNG QUAN VỀ TRÍCH CHỌN ĐẶC TRƯNG VÀ BIỂU DIỄN BIỂU CẢM KHUÔN MẶT**

### **1.1 Tổng quan về một hệ thống biểu cảm khuôn mặt**

Các nghiên cứu trong lĩnh vực này được chia thành hai hướng chính: Một là, hướng liên quan đến phát hiện nhận dạng mặt người và các trạng thái biểu cảm của mặt người. Hai là, hướng liên quan đến biểu diễn mô hình mặt người cùng với các trạng thái khác nhau.

- Để giải quyết bài toán trích chọn các đặc trưng biểu cảm hoặc cử chỉ trên khuôn mặt có 2 hướng tiếp cận chính:

- Hướng thứ nhất là trích chọn các đặc trưng biểu cảm khuôn mặt dựa trên các điểm đánh dấu. Với hướng này người ta cũng chia thành nhiều hướng con khác tùy vào cách lựa chọn loại điểm đánh dấu, hay số lượng camera quan sát một hoặc nhiều camera.
- Hướng nghiên cứu thứ hai đang được tập trung nghiên cứu nhiều trong thời gian gần đây là trích chọn đặc trưng mà không sử dụng các điểm đánh dấu. Với hướng nghiên cứu này có thể có một số cách tiếp cận như sử dụng các bộ học để đoán nhận biểu cảm khuôn mặt trên ảnh từ đó tính được các đặc trưng biểu cảm, hoặc sử dụng mô hình AAM (Active Appearance Model) để nội suy hình dạng của khuôn mặt từ đó trích rút được các đặc trưng biểu cảm của khuôn mặt.

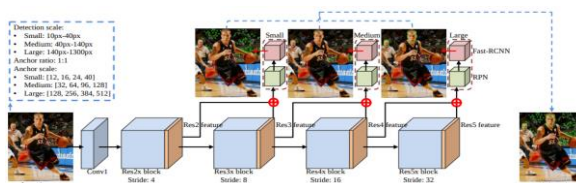
## 1.2 Phát hiện khuôn mặt trong ảnh

Trong vài năm gần đây, nhóm Yu [129] ứng dụng kỹ thuật CNN, đã đề xuất một hàm lỗi dùng độ đo IoU để hồi quy trực tiếp hộp bao mang lại sự hiệu quả đối với các đối tượng khuôn mặt có sự biến động về hình dạng và kích thước.

Theo một cách nhìn khác, nhóm Zhang [134] đã kết hợp các mạng CNN xếp tầng để xử lý đa chức năng kết hợp giữa phát hiện khuôn mặt và phát hiện các điểm đặc trưng khuôn mặt.

Quan tâm đến các khuôn mặt ở nhiều mức tỷ lệ, nhóm của Yang [124] mô hình hóa khuôn mặt với các tỷ lệ khác nhau thông qua một tập hợp gồm các mạng CNN có cấu trúc khác nhau.

Cũng tương tự, nhóm Hu [56] quan tâm đến các khuôn mặt nhỏ trong ảnh. Theo đó, các tác giả đã huấn luyện một số bộ phát hiện khác nhau để cho các khuôn mặt nhỏ ở nhiều mức tỉ lệ.

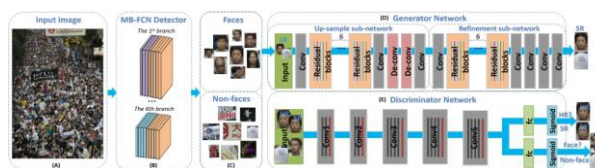


*Quy trình xử lý của nhóm Yang [124]*

Nhóm của Bai [14] sử dụng cơ sở mô hình GAN để phát hiện các khuôn mặt mờ nhỏ thông qua việc sinh dữ liệu rõ ràng hơn. Một vấn đề khác cũng có liên quan đến mô hình GAN đó là các khuôn mặt giả.

Nhóm Zhengzhe [78] đã tiến hành một nghiên cứu thực nghiệm về khuôn mặt giả và khuôn mặt thật và tập trung vào đặc điểm kết cấu toàn cục của ảnh. Các tác giả

cũng đề xuất Gram-Net để xử lý kết cấu hình ảnh toàn cầu nhằm phát hiện ảnh khuôn mặt giả mạo.



Mô hình của nhóm Bai [14]

### 1.3 Trích chọn đặc trưng biểu cảm

#### 1.3.1 Mẫu nhị phân cục bộ (LBP)

Thuật toán LBP [88] [8] là một phương pháp không tham số, nghĩa là LBP tóm tắt các cấu trúc hình ảnh cục bộ một cách hiệu quả bằng cách so sánh từng cấu trúc pixel với các pixel lân cận của nó, ban đầu sẽ gán nhãn cho các pixel của một vùng ảnh với những số thập phân. Mẫu nhị phân cục bộ được mô tả như sau: Mỗi pixel được so sánh với 8 pixel xung quanh nó trong vùng ảnh có kích thước 3x3 pixel bằng cách trừ đi giá trị của pixel trung tâm; các giá trị âm được mã thành giá trị 0, các giá trị khác được mã thành giá trị 1; mỗi số nhị phân có được bằng cách nối tất cả các mã này theo chiều kim đồng hồ bắt đầu từ phía trên bên trái tương ứng với các số thập phân được gán nhãn trước đó. Chuỗi số nhị phân được tạo thành đó gọi là mẫu nhị phân cục bộ.

#### 1.3.2 ASM / AAM

Active Shape Model (ASM) được đề xuất trong [9] dựa trên các mô hình thống kê và được dùng để trích xuất các điểm đặc trưng trên đường viền. Kỹ thuật này chủ yếu sử dụng mô hình toàn cục để so khớp với hình dạng ban đầu của khuôn mặt người, sau đó thiết lập một mô hình kết cấu cục bộ để điểm đặc trưng trên đường viền đạt được độ chính xác hơn.

Active Appearance Model (AAM) [59] được phát triển dựa trên cơ sở ASM là một thuật toán phổ biến tối ưu mô hình thống kê đối tượng của ảnh đầu vào, kết quả của quá trình tối ưu là một bộ điểm điều khiển thể hiện cấu trúc của đối tượng đã được học với các tọa độ tương ứng với thể hiện của đối tượng trong ảnh đầu vào, cùng với nó là một bộ các tham số mô hình thống kê đã được ước lượng từ những tham số mô hình này, ta có thể dễ dàng tái cấu trúc cả về hình dạng cũng như kết cấu hình ảnh của đối tượng tương ứng một cách tương đối với thể hiện của đối tượng ở trong ảnh.

### 1.3.3 Trích xuất đặc trưng Gabor

Trích xuất đặc trưng dựa trên Gabor wavelets có những lợi thế đáng kể. Trong [85], một tập hợp các bộ lọc Gabor, với đa hướng và đa độ phân giải, được sử dụng để mã hình ảnh biểu cảm khuôn mặt. Boughida Adil và cộng sự [18] sử dụng tuyến tính và phi tuyến tính tổ hợp các thuật toán mới trên cơ sở của Gabor. Gabor-mean-DWT (Discrete Wavelet Transform) [83] cung cấp một vectơ đặc trưng nhỏ gọn hơn so với biểu thức dựa trên Gabor hiện có để giảm bớt vấn đề về chiều.

### 1.3.4 Trích xuất đặc trưng Haar-like

Đặc trưng Haar-like [29] công bố bao gồm 4 đặc trưng cơ bản để xác định mặt người. Đặc trưng Haar-like được tạo thành bằng việc kết hợp các hình chữ nhật màu đen và trắng với nhau theo một trật tự, kích thước nào đó. Mỗi đặc trưng Haar-like là sự kết hợp của hai hay ba hình chữ nhật “đen” hay “trắng”.

Để chọn các đặc trưng Haar-like dùng cho việc thiết lập ngưỡng, tạo bộ phân loại thì Viola và Jones sử dụng phương pháp máy học gọi là AdaBoost. AdaBoost sẽ kết hợp các bộ phân loại yếu tạo thành một bộ phân loại mạnh. Với những bộ phân loại yếu thì việc đưa ra câu trả lời chính xác thì chỉ hơn việc đoán ngẫu nhiên một chút, trong khi đó bộ phân loại mạnh có thể đưa ra kết quả chính xác cao hơn nhiều.

## 1.4 Một số cách tiếp cận biểu diễn biểu cảm khuôn mặt

Parker [90] xây dựng một mô hình mô phỏng cơ sử dụng cơ để mô phỏng các biểu thức với các vector cơ. Vị trí của điểm trên khuôn mặt được cập nhật dựa trên các chức năng đặc biệt. Sau đó, rất nhiều nhà nghiên cứu [125] [42] dành cho mô hình cơ bắp để sản xuất hình ảnh động trên khuôn mặt.

Chuang [28] thiết kế một hệ thống tự động phát hiện các đặc điểm hình dạng chính và trọng số tương ứng để định hướng điều khiển mô hình mục tiêu. Joshi [63] đưa ra ý tưởng phân chia sử dụng các hình khối hỗn hợp. Lewis [70] trình bày một cách tiếp cận thao tác trực tiếp lên mô hình hình dạng hỗn hợp sử dụng chuyển động nghịch đảo. Liu [76] sử dụng một mô hình tối ưu trong tự động phát hiện ra tham số của một mô hình phi tuyến mô tả hình dạng khuôn mặt. Yano [125] tập hợp được một tập hợp các thông số chuyển động của hệ thống cơ và áp dụng chúng trên các mô hình mục tiêu để tạo ra các biểu hiện tương tự. You [128] xây dựng một mô hình toán học dựa theo các tính chất vật lý của biến dạng da và sử dụng các hình dạng khuôn mặt mới được tổng hợp trên cơ sở tương tác các lực tại các điểm điều khiển. Bickel [19] đã xây dựng ánh xạ biến đổi mô hình thông qua một bộ thừa thớt những điểm điều khiển tự định nghĩa.

Ngoài ra nghiên cứu sinh còn khảo sát một số phương pháp như:

- Nội suy tuyến tính
  - Kết hợp giữa kỹ thuật nội suy tuyến tính và kỹ thuật Morphing:
  - Kết hợp giữa kỹ thuật nội suy tuyến tính và hệ mã hóa hành động mặt:
- Nội suy song tuyến.
- Nội suy dựa trên hàm cơ sở bán kính
- Nội suy địa phương
- Biến đổi 3D cứng (Rigid)

### 1.5 Tổng kết chương và vấn đề nghiên cứu

Trong chương này, luận án đã trình bày tổng quan về một hệ thống rút trích đặc trưng, cử chỉ khuôn mặt và biểu diễn biểu cảm khuôn mặt cùng với các vấn đề cơ bản liên quan. Bên cạnh đó luận án cũng giới thiệu một số tiếp cận trong việc rút trích đặc trưng và biểu diễn biểu cảm. Đặc biệt là trình bày các hướng tiếp cận hiện có để giải quyết bài toán rút trích đặc trưng cử chỉ, biểu cảm và bài toán biểu diễn biểu cảm khuôn mặt. Đây là hai bài toán quan trọng và được ứng dụng nhiều trong thực tế.

## Chương 2

### KỸ THUẬT TRÍCH CHỌN ĐẶC TRƯNG CỬ CHỈ VÀ ƯỚC LƯỢNG BIỂU CẢM

Trích chọn đặc trưng biểu cảm khuôn mặt là một kỹ thuật tính toán và ước lượng các thông số cho các mô hình toán học nhằm phân tích và tổng hợp biểu cảm khuôn mặt. Bài toán này nhận đầu vào là một ảnh khuôn mặt và đầu ra là một mô hình toán học của biểu cảm khuôn mặt đã được tham số đầy đủ nhằm mục đích làm dữ liệu đầu vào cho bài toán nhận dạng hoặc mô phỏng biểu cảm khuôn mặt trên ảnh.

#### 2.1 Bài toán phát hiện khuôn mặt

Quy trình phát hiện mặt người gồm hai bước chính: Thứ nhất, từ một tập các ảnh khuôn mặt và các ảnh không phải khuôn mặt, ta xây dựng mô hình cho hệ thống học để tạo ra một bộ dữ liệu khuôn mặt mẫu. Thứ hai, khi tiến hành tìm kiếm khuôn mặt thì ta sẽ so sánh các đặc điểm trên ảnh đầu vào với các đặc điểm trong mô hình đã được xây dựng.

### 2.1.1 Mô hình khuôn mặt mẫu

Mỗi cây được xây dựng dựa trên tập dữ liệu huấn luyện như sau:

$$\{(I_s, v_s, w_s): s = 1, 2, \dots, S\} \quad (2.1)$$

Trong đó,  $v_s$  là nhãn đúng của ảnh  $I_s$ , và  $w_s$  là trọng số tương ứng.

Cụ thể, trong trường hợp này, vấn đề bài toán đưa ra là thực hiện phân lớp nhị phân, các nhãn có giá trị tương ứng là +1 và -1. Giá trị trọng số  $w_s$  cho phép đánh dấu mức độ quan trọng khác nhau của mỗi mẫu đầu vào trong tập huấn luyện. Quá trình xây dựng cây được thực hiện tại từng nút trên cơ sở lựa chọn bộ phân lớp nhị phân con có khả năng phân lớp tốt nhất bộ dữ liệu huấn luyện, tức là đạt giá trị cực tiểu cho hàm mục tiêu.

Nói cách khác, tại mỗi nút trong cây, ta xét duyệt các bộ phân lớp nhị phân con có thể, tùy từng dạng phân lớp mà ta học hoặc lựa chọn tham số, hoặc thậm chí là không có tham số. Như vậy, từ tập dữ liệu ban đầu, tại mỗi bước học nút khi xây dựng cây, tập dữ liệu huấn luyện tương ứng được chia đôi. Cơ sở của thuật toán đề xuất nằm ở việc học từng nút.

### 2.1.2 Kết quả thử nghiệm

*Bảng 2.1 So sánh kết quả định vị khuôn mặt*

	<i>Haar Adaboost</i>	<i>PP đề xuất</i>
Số mẫu không phát hiện được	7/450	27/450
Số mẫu phát hiện nhầm	16	7
Thời gian xử lý trung bình ảnh (giây/ảnh)	0,109713	0,018982

Trong 450 ảnh đó, thuật toán phát hiện được 423 ảnh có khuôn mặt người. Bằng việc kiểm tra thủ công 423 ảnh được rút ra, Luận án phát hiện có 7 ảnh không có khuôn mặt người; tức là thuật toán phát hiện đúng 416 ảnh trong 423 ảnh đó.

Khả năng phát hiện của thuật toán như sau:

$$\text{Precision} = 416 / 423 = 0,9834 \text{ (hay đạt } 98,34\%);$$

$$\text{Recall} = 416 / 450 = 0,9244 \text{ (hay đạt } 92,44\%).$$

Như vậy, kết quả 92,44% cho thấy phương pháp được đề xuất cho độ chính xác phát hiện khuôn mặt khá tốt. Mặc dù thuật toán Haar Adaboost đạt giá trị recall 94,89% tốt hơn so với 92,44% của phương pháp đề xuất trong Luận án nhưng bù lại

thì nó có tốc độ xử lý nhanh hơn, đây là cơ sở để có thể phát triển những ứng dụng xử lý khuôn mặt yêu cầu thời gian thực trên dữ liệu video.

## 2.2 Bài toán trích chọn đặc trưng cử chỉ

Hầu hết các phương pháp phân tích cử chỉ khuôn mặt tự động trong chuỗi hình ảnh khuôn mặt đều cố gắng nhận ra một tập hợp các biểu hiện cảm xúc nguyên mẫu trên khuôn mặt, tức là hạnh phúc, buồn bã, sợ hãi, ngạc nhiên, tức giận và ghê tởm [89]. Tuy nhiên, trong cuộc sống hàng ngày, những biểu hiện nguyên mẫu như vậy xảy ra không thường xuyên; cảm xúc được thể hiện thường xuyên hơn bằng những thay đổi tinh tế ở một hoặc vài đặc điểm riêng biệt trên khuôn mặt, chẳng hạn như nhướn mày ngạc nhiên [25].

## 2.3 Đề xuất kỹ thuật ước lượng biểu cảm khuôn mặt

Việc ước lượng biểu cảm khuôn mặt được trình bày trong đây cũng lấy ý tưởng từ việc lựa chọn rồi rọc một số đặc trưng hình dạng tương tự như trong [28] trên cơ sở khuôn mặt đã được định vị tập điểm điều khiển thông qua AAM. Nhận định ban đầu của nghiên cứu sinh cũng tương tự, khi tập điểm điều khiển của khuôn mặt bị biến đổi do các biểu cảm, khả năng tồn tại những thông số đặc trưng có sự biến đổi phân biệt. Thay vì thống kê và lựa chọn bằng tay như trong [28], nghiên cứu sinh tiến hành tự động lựa chọn các đặc trưng hình học một cách ngẫu nhiên và tổ chức dưới mô hình cây quyết định để thực hiện ước lượng các biểu cảm khuôn mặt.

### 2.3.1 Mô hình ước lượng

Mô hình ước lượng được xây dựng trên tập dữ liệu có cấu trúc:

$$\{(I_s, v_s, w_s) : s = 1, 2, \dots, S\}$$

trong đó,  $v_s$  là nhãn đúng của mẫu  $I_s$ , và  $w_s$  là trọng số. Cụ thể, trong trường hợp này, các nhãn có giá trị tương ứng là trọng số biểu cảm và nằm trong đoạn  $[0, 1]$ . Giá trị trọng số  $w_s$  cho phép đánh dấu mức độ quan trọng khác nhau của mỗi mẫu đầu vào trong tập huấn luyện. Quá trình xây dựng cây được thực hiện tại từng nút trên cơ sở lựa chọn một hàm quyết định có khả năng phân lớp tốt nhất bộ dữ liệu huấn luyện, tức là đạt giá trị cực tiểu cho hàm mục tiêu; cụ thể là thực hiện tìm sai số bình phương nhỏ nhất ứng với việc phân chia tập huấn luyện được xác định tại nút đó.

Hàm mục tiêu có dạng: (2.9)

$$WMSE(I, v, w) = \sum_{(I, v, w) \in C_0} w \cdot (v - \bar{v}_0)^2 + \sum_{(I, v, w) \in C_1} w \cdot (v - \bar{v}_1)^2$$

trong đó:

Tham số  $\bar{v}_0$  và  $\bar{v}_1$  là trung bình các giá trị nhãn trong  $C_0$  và  $C_1$ .

Tham số  $C_0$  và  $C_1$  là hai cụm của tập huấn luyện tương ứng với kết quả phân chia của hai giá trị 0 và 1.

Nói cách khác, tại mỗi nút trong cây, ta xét duyệt các hàm quyết định có thể được xây dựng từ các đặc trưng hình dạng và tìm hàm phân chia tốt nhất theo nghĩa cực tiểu hàm mục tiêu. Như vậy, từ tập dữ liệu ban đầu, tại mỗi bước học nút khi xây dựng cây, tập dữ liệu huấn luyện tương ứng được chia đôi. Cơ sở của thuật toán đề xuất nằm ở việc học từng nút.

### 2.3.2 Đề xuất sử dụng đặc trưng hình dạng

Trên cơ sở mô hình cây được mô tả như trên. Trong luận án này, nghiên cứu sinh đề xuất sử dụng 3 loại đặc trưng hình dạng là *LINE\_LINE*, *TRIANGLE\_TRIANGLE*, *LINELINE\_LINELINE*. Lý do chọn 3 loại đặc trưng hình dạng này là vì các đặc trưng trên khuôn mặt tuy khác nhau nhưng luôn có một mối liên hệ gắn liền nhau tạo thành các đường hay tam giác đi theo một mô hình tương tự nhau. Ví dụ vị trí của mắt mũi miệng là luôn tương quan với nhau:

#### a. *LINE\_LINE*

Đây là đặc trưng được tính bằng tỉ số độ dài hai đoạn thẳng, do vậy sử dụng tham số đầu vào là 4 điểm.

$$f_{LINE\_LINE}(p_1, p_2, p_3, p_4) = \frac{d(p_1, p_2)}{d(p_3, p_4)} \quad (2.10)$$

Trong đó,  $p_1, p_2, p_3, p_4$  là 4 điểm đầu vào,  $d(p_i, p_j)$  là độ dài đoạn thẳng tạo bởi 2 điểm  $p_i, p_j$ .

#### b. *TRIANGLE\_TRIANGLE*

Đây là đặc trưng được tính bằng tỉ số diện tích hai tam giác, do vậy sử dụng tham số đầu vào là 6 điểm.

$$f_{TRIANGLE\_TRIANGLE}(p_1, p_2, p_3, p_4, p_5, p_6) = \frac{S(p_1, p_2, p_3)}{S(p_4, p_5, p_6)} \quad (2.11)$$

Trong đó,  $p_1, p_2, p_3, p_4, p_5, p_6$  là 6 điểm đầu vào,  $S(p_i, p_j, p_k)$  là diện tích tam giác được tạo bởi 3 đỉnh  $p_i, p_j, p_k$ .

#### c. *LINELINE\_LINELINE*

Đây là đặc trưng được tính bằng tỉ số của tổng độ dài hai đoạn thẳng, do vậy sử dụng tham số đầu vào là 8 điểm

$$f_{LINELINE\_LINELINE}(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8) = \frac{d(p_1, p_2) + d(p_3, p_4)}{d(p_5, p_6) + d(p_7, p_8)} \quad (2.12)$$

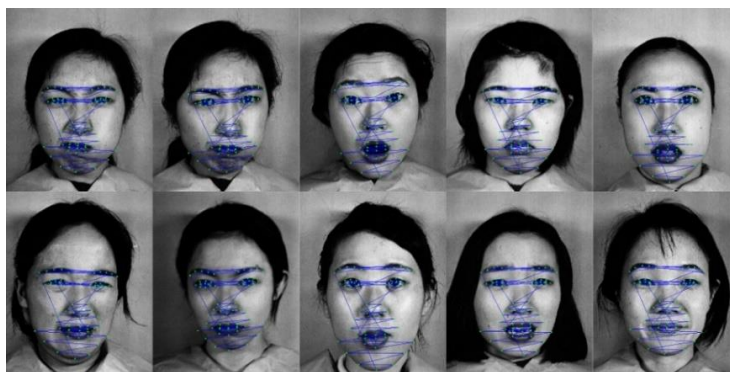
Trong đó,  $p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8$  là 8 điểm đầu vào,  $d(p_i, p_j)$  là độ dài đoạn thẳng tạo bởi 2 điểm  $p_i, p_j$ .

### 2.3.3 Thử nghiệm

Cơ sở dữ liệu được lựa chọn để thử nghiệm là cơ sở dữ liệu JAFFE. Cơ sở dữ liệu JAFFE chứa 213 ảnh của 6 biểu cảm khuôn mặt khác nhau bao gồm: vui, buồn, ngạc nhiên, tức giận, phần nộ, sợ hãi, được thu nhận trên 10 người mẫu nữ Nhật Bản. Cơ sở dữ liệu đã được xây dựng bởi Michael Lyons, Miyuki Kamachi, và Jiro Gyoba cùng với trợ lý nghiên cứu Reiko Kubota. Các bức ảnh được chụp tại Khoa Tâm lý học tại Đại học Kyushu. Các ảnh trong cơ sở dữ liệu JAFFE được đánh giá kèm tỉ lệ của 6 trạng thái cảm xúc. Những đánh giá này được thực hiện bởi bảng cảm quan của 60 sinh viên Nhật Bản với mức cao nhất là 5, thấp nhất là 1.

Cơ sở dữ liệu JAFFE cũng đã được sử dụng nhiều trong các bài toán nhận dạng biểu cảm khuôn mặt, chẳng hạn nhóm Bashyal và Venayagamoorthy [102] đã đạt kết quả chính xác 90,2% với việc trích rút đặc trưng bằng bộ lọc Gabor kết hợp với kỹ thuật learning vector quantization (LVQ); nhóm Oliveira [95] đạt độ chính xác 94% với kỹ thuật lựa chọn đặc trưng 2DPCA kết hợp SVM.

Trong thực nghiệm của nghiên cứu sinh, tại bước đầu, các ảnh trong cơ sở dữ liệu được thực hiện định vị tập điểm điều khiển làm cơ sở để tính toán các đặc trưng hình dạng.



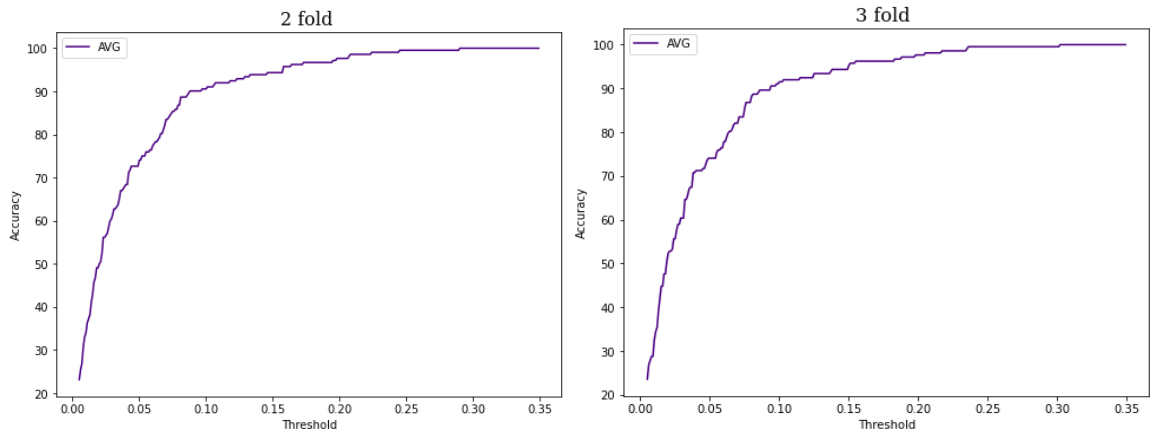
Hình 2.1 Một số kết quả định vị tập điểm điều khiển

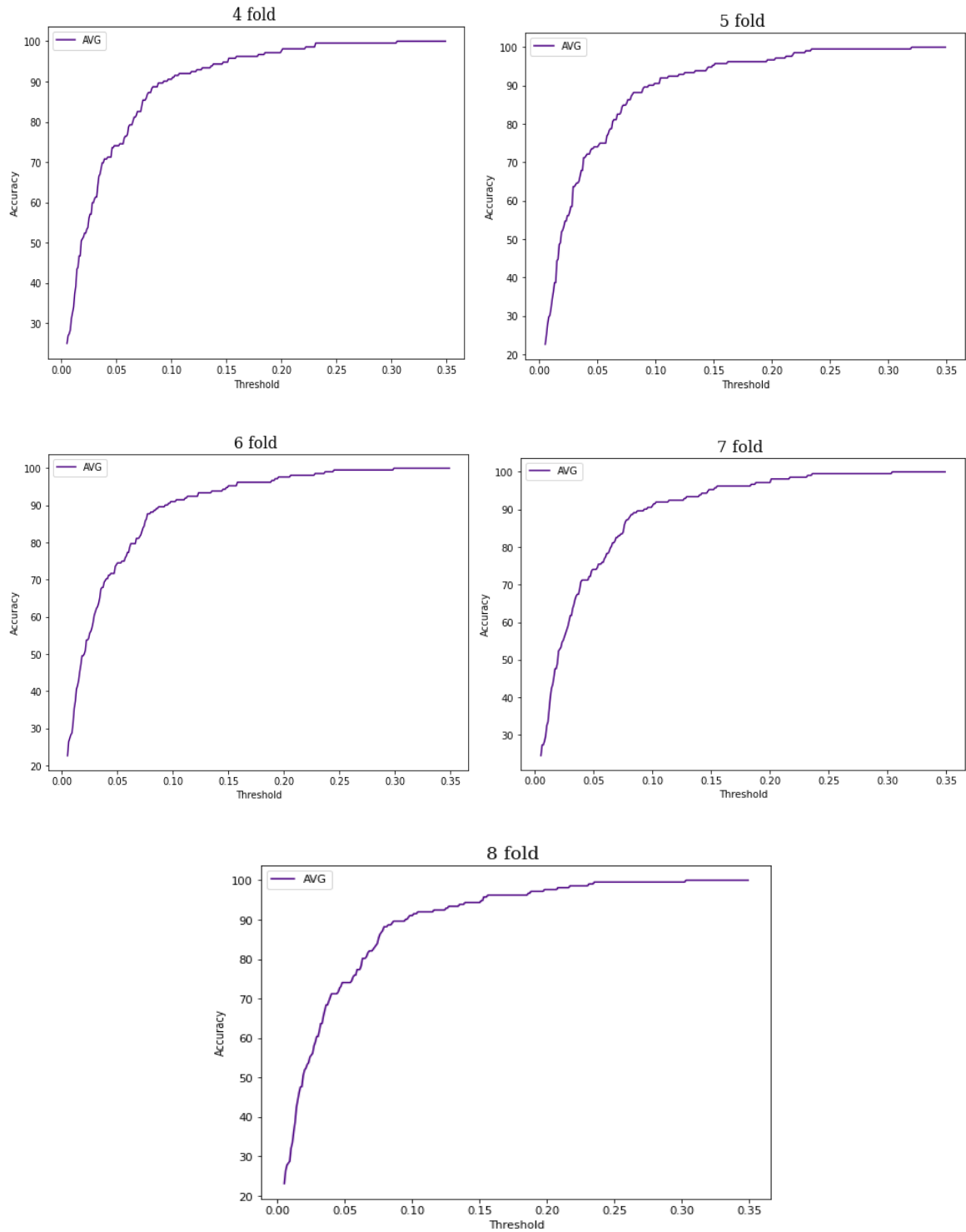
Trên cơ sở đã có tập các điểm điều khiển cho mỗi ảnh trong cơ sở dữ liệu, tác giả tiến hành thử nghiệm theo cách thức kiểm chứng chéo trên cơ sở dữ liệu với các mức phân đoạn từ 2 đến 8. Các cây nhị phân xây dựng được giới hạn chiều sâu bằng 8. Các thử nghiệm được tiến hành trên máy tính Core i7-4790 3.6GHz 8GB RAM dùng môi trường phát triển Visual Studio C++ 2015. Trên cơ sở xây dựng bộ ước lượng và thực nghiệm với mỗi phân đoạn, ta được sai số trung bình, được tính bằng trung bình bình phương độ lệch, trên mỗi loại biểu cảm như sau:

*Bảng 2.2 Sai số trung bình của các loại biểu cảm*

Biểu cảm	2 fold	3 fold	4 fold	5 fold	6 fold	7 fold	8 fold
<b>HAP</b> (Hạnh phúc)	0.075648	0.075657	0.075761	0.075889	0.075896	0.075651	0.075659
<b>SAD</b> (Buồn)	0.04838	0.048428	0.048552	0.048898	0.048755	0.048566	0.048371
<b>SUR</b> (Ngạc nhiên)	0.085872	0.085553	0.085907	0.086064	0.086282	0.08658	0.085854
<b>ANG</b> (Giận dữ)	0.058989	0.059137	0.0597	0.059162	0.059336	0.059349	0.059686
<b>DIS</b> (Ghê tởm)	0.078655	0.078796	0.079557	0.079028	0.078821	0.078919	.079176
<b>FEA</b> (Sợ hãi)	0.038529	0.038167	0.038288	0.038275	0.0385	0.038231	0.0381

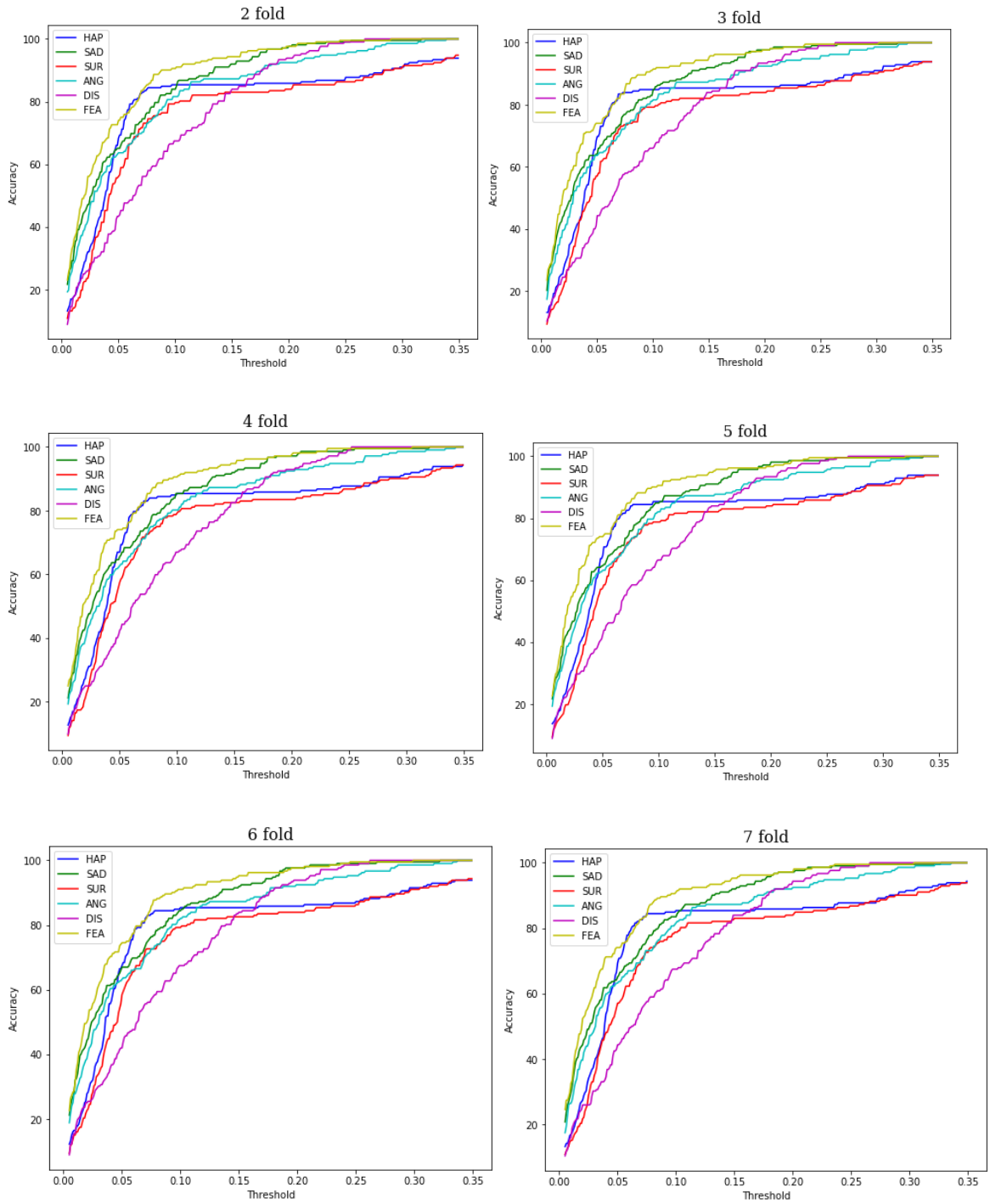
Ngoài ra, tác giả cũng thực hiện đánh giá thêm độ chính xác trên từng mẫu khi thực hiện so sánh sai số độ lệch với một giá trị ngưỡng truyền vào. Trên cơ sở đó thực hiện xây dựng lược đồ thống kê tương quan giữa tỉ lệ chính xác và ngưỡng chấp nhận.

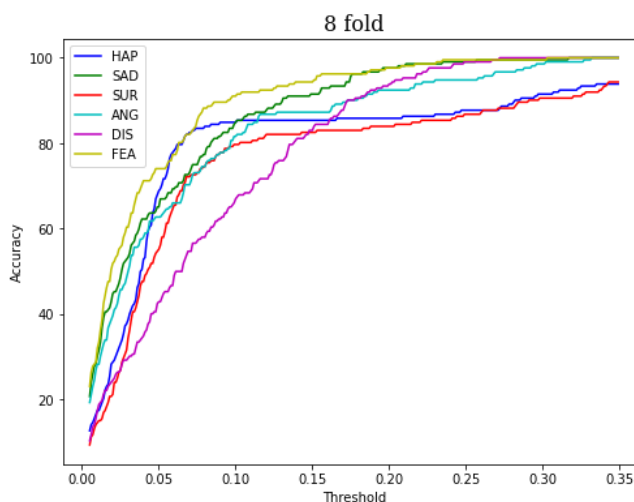




*Hình 2.2 Các biểu đồ thống kê tỉ lệ chính xác theo ngưỡng chấp nhận*

Từ những biểu đồ trên ta có thể nhận thấy độ chính xác đạt đến mức xấp xỉ 95% chỉ từ mức ngưỡng khoảng 0,14. Ta có thể quan sát thêm khi xét riêng từng loại biểu cảm





Hình 2.3 Biểu đồ thống kê tỉ lệ chính xác theo từng loại biểu cảm

Ngoài ra, một điểm mạnh của phương pháp đó là sử dụng những phép tính toán khá đơn giản, cụ thể là dựa vào các phép tính hình học trong đặc trưng LINE\_LINE, TRIANGLE\_TRIANGLE và LINELINE\_LINELINE. Điều này cho phép thực hiện nhanh chóng thao tác ước lượng biểu cảm dựa trên tập điểm điều khiển đã có. Thống kê thời gian xử lý như sau (tính theo mili giây):

Bảng 2.3 Thống kê thời gian xử lý của các loại biểu cảm

Biểu cảm	2 fold	3 fold	4 fold	5 fold	6 fold	7 fold	8 fold
HAP	0.0123819 292452829 9	0.0134987 405660377 36	0.0183626 886792452 94	0.0220075 660377358 62	0.0224209 952830188 6	0.0148654 764150943 31	0.0086954 009433962 3
SAD	0.0121416 226415094 2	0.0134867 264150943 37	0.0185748 018867924 57	0.0218773 207547169 8	0.0223673 820754717 24	0.0143458 962264150 87	0.0082268 537735849 1
SUR	0.0122020 849056603 49	0.0136303 537735849 1	0.0182861 698113207 83	0.0214464 339622641 44	0.0216531 509433962 3	0.0136693 679245282 89	0.0081529 481132075 5
ANG	0.0118784 292452830 07	0.0136477 830188679 3	0.0183371 933962264 3	0.0213981 509433962 48	0.0214275 801886792 47	0.0137592 547169811 28	0.0082657 783018868 02
DIS	0.0123028 915094339 29	0.0137767 169811320 81	0.0183680 566037736	0.0212772 830188679 43	0.0221149 858490565 82	0.0138357 075471698 03	0.0084845 424528301 93
FEA	0.0121655 849056603 75	0.0135605 188679245 3	0.0180350 943396226 47	0.0204382 216981132 3	0.0210759 009433962 1	0.0137390 518867924 47	0.0080590 094339622 71

Từ bảng số liệu thời gian như trên, không tính thời gian xác định tập điểm điều khiển khuôn mặt, ta có thể nhận thấy thao tác ước lượng biểu cảm có thời gian thực hiện rất nhanh, chỉ cỡ xấp xỉ 0,015 ms cho một biểu cảm, như vậy để ước lượng đầy đủ 6 loại biểu cảm sẽ mất xấp xỉ 0,9 ms. Đây sẽ là một thế mạnh cho những nghiên cứu tiếp sau và có nhiều triển vọng khi muốn triển khai trên môi trường có năng lực tính toán hạn chế, chẳng hạn như trên các hệ thống nhúng.

## **2.4 Tổng kết chương**

Chương này, trình bày 3 bài toán : (i) phát hiện khuôn mặt người trong ảnh; (ii) trích chọn đặc trưng cử chỉ của khuôn mặt và (iii) ước lượng biểu cảm khuôn mặt.

Tác giả cũng đã tiến hành thử nghiệm phát hiện khuôn mặt người trong ảnh và đạt kết quả chính xác ở mức 92,44%; điều này cho thấy phương pháp lựa chọn cho độ chính xác phát hiện khuôn mặt khá tốt và ưu điểm của nó là có tốc độ xử lý nhanh hơn.

Chương này cũng đã trình bày những nét cơ bản của bài toán trích chọn đặc trưng cử chỉ. Trong đó, việc xác định các điểm điều khiển là một bước quan trọng để rút trích cử chỉ. Thuật toán AAM được chọn sử dụng trong luận án tỏ ra hiệu quả nhất khi dữ liệu thu nhận được trong những điều kiện ánh sáng tốt, trong điều kiện ánh sáng và nền phức tạp thì thuật toán AAM không hiệu quả. Do đó, tác giả đề xuất một kỹ thuật định vị điểm điều khiển có ràng buộc dựa trên mô hình chất liệu mặt. Thử nghiệm cho thấy kết quả có độ chính xác cao, đặc biệt là trong trường hợp màu da khác nhau, ánh sáng thay đổi. Qua thực nghiệm tác giả lựa chọn ngưỡng 0,13 là khá tốt khi kiểm chứng lại bằng mắt thường, theo đó, độ chính xác đạt khoảng 94,31%.

## **Chương 3**

### **BIỂU DIỄN CỬ CHỈ, BIỂU CẢM TRÊN KHUÔN MẶT 3D**

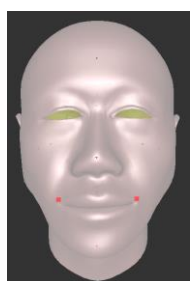
#### **3.1 Bài toán biểu diễn cử chỉ, biểu cảm khuôn mặt 3D**

Nhiều nhà nghiên cứu đã quan tâm đến việc chuyển đổi từ biểu diễn đám mây điểm sang các biểu diễn hoàn chỉnh khác như lưới 3D bởi vì biểu diễn lưới 3D cho phép thực hiện các thao tác linh hoạt và hiệu quả trên các bề mặt. Bên cạnh đó, lưới 3D còn cho phép lưu trữ thông tin cục bộ được lập chỉ mục và tính toán trước của bề mặt. Ví dụ, một khi lưới biểu diễn được tính toán, nó cho phép phát triển một vùng hiệu quả hơn so với biểu diễn đám mây điểm. Ngoài ra, sự biến dạng của lưới 3D linh hoạt hơn các đám mây điểm.

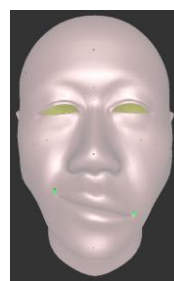
### 3.2 Kỹ thuật nội suy RBF

Việc mô phỏng lại biểu cảm khuôn mặt dựa vào các đặc trưng thực chất là việc nội suy nhằm tính lại bề mặt 3D của khuôn mặt dựa theo các đặc trưng biểu cảm. Tuy nhiên việc lựa chọn phương pháp nội suy nào cho phù hợp nhất với bài toán mô phỏng biểu cảm khuôn mặt là một vấn đề cần nghiên cứu. Có rất nhiều phương pháp như NURBS, RBF, Affine, nội suy dựa vào mạng neural, v.v... Trong đó, Wang và Ierapetritou và Dias đã chỉ ra rằng RBF cho kết quả có độ chính xác cao và thường được sử dụng trong lĩnh vực khoa học máy tính. Do đó, kỹ thuật nội suy RBF được sử dụng trong nghiên cứu này.

Ý tưởng của kỹ thuật nội suy mô hình 3D dựa vào hàm cơ sở bán kính là đối tượng sẽ được mô tả đặc trưng bởi một tập điểm gọi là điểm điều khiển, việc nội suy đối tượng sẽ chính là nội suy các điểm điều khiển.



a) Đối tượng trước nội suy



b) Đối tượng sau nội suy

### 3.3 Xác định tập điểm điều khiển phục vụ nắn chỉnh biến dạng

#### 3.3.1 Xác định các nhóm tương đồng và chọn điểm điều khiển

Từ tập biến dạng  $M$ , ta thực hiện tính các đặc trưng biến dạng của từng điểm khuôn mặt và thu được một tập đặc trưng, mỗi đặc trưng sẽ tương ứng thể hiện các tọa độ khác nhau của một điểm mô hình mặt 3D. Ta sẽ thực hiện gom cụm trên tập đặc trưng này để tìm ra các nhóm điểm có sự biến dạng tương đồng với nhau trong tập dữ liệu đầu vào  $M$ . Kỹ thuật gom cụm được xây dựng dựa trên thuật toán K-means, được thiết kế dựa trên dữ liệu là các đặc trưng biến dạng của điểm khuôn mặt 3D và điểm tâm của mỗi nhóm được chọn là một điểm của mô hình.

Từ kết quả của thuật toán ta có được các nhóm tương đồng về biến dạng, mỗi nhóm ta chọn những điểm khuôn mặt biến dạng nhiều nhất và ít nhất để đưa vào tập điểm điều khiển. Mức độ biến dạng được đánh giá dựa trên một mô hình tham chiếu. Như vậy mỗi điểm của mô hình khuôn mặt tham chiếu sẽ tương ứng với một đặc trưng biến dạng và ta xây dựng vec tơ độ lệch bằng khoảng cách của từng điểm trong đặc trưng biến dạng tới điểm tương ứng trên mô hình tham chiếu. Giá trị đánh giá sự

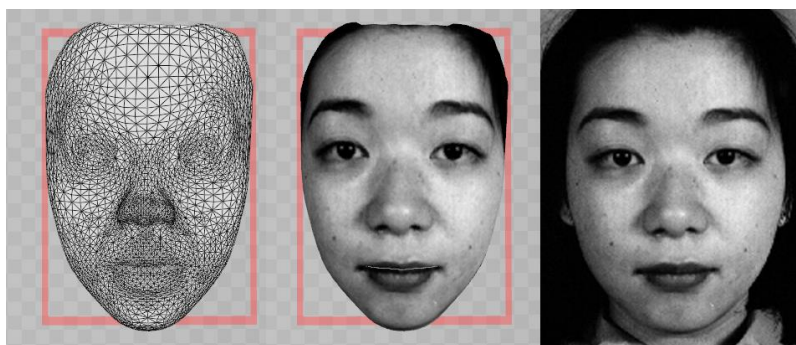
biến đổi của mỗi điểm tương ứng sẽ được tính dựa vào vec tơ độ lệch đó.

Mục tiêu của việc xác định tập điểm điều khiển là để phục vụ nắn chỉnh biến dạng. Hoặc nói cách khác, tiêu chí để đánh giá tập điểm điều khiển có tốt hay không là dựa trên kết quả sử dụng nó trong một thuật toán nắn chỉnh biến dạng cụ thể. Trong trường hợp này, để chọn được một tập điểm điều khiển tốt, ta cần kết hợp thêm một thuật toán nắn chỉnh biến dạng và một tập các mô hình biến thể để phục vụ đánh giá chất lượng nắn chỉnh

### 3.3.2 Thử nghiệm nắn chỉnh mô hình

Quá trình thử nghiệm được tiến hành theo kịch bản như sau: đầu vào cần chuẩn bị 1 kỹ thuật nắn chỉnh đối tượng dựa trên điểm điều khiển, 1 đối tượng 3D làm tham chiếu, 1 tập đối tượng 3D các biến thể để tính toán chọn điểm điều khiển và 1 tập đối tượng 3D các biến thể để đánh giá chất lượng tập điểm điều khiển chọn được; trên cơ sở đó tiến hành thử nghiệm với nhiều K khác nhau, với mỗi giá trị ta thực hiện đánh giá sai số. Kỹ thuật nắn chỉnh được sử dụng trong thử nghiệm là kỹ thuật nắn chỉnh đối tượng 3D sử dụng hàm cơ sở bán kính (RBF – Radial Basic Function), đây là kỹ thuật cũng đã được sử dụng trong luận án phục vụ biến đổi mô hình 3D khuôn mặt người theo biểu cảm.

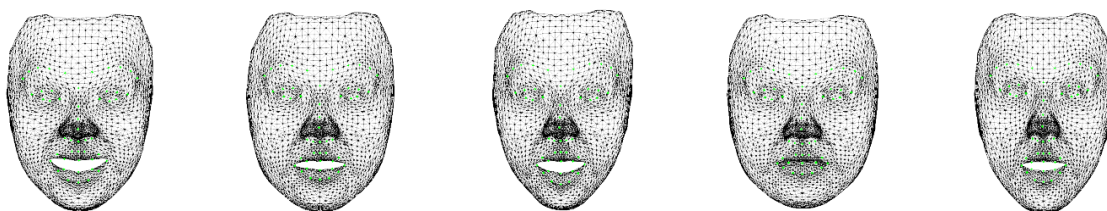
Đối tượng 3D sử dụng là mô hình khuôn mặt có 3448 đỉnh và 6736 bề mặt được tạo ra tương ứng với các mẫu trong cơ sở dữ liệu JAFFE



Hình 3.1 Ví dụ dữ liệu mô hình mặt 3D

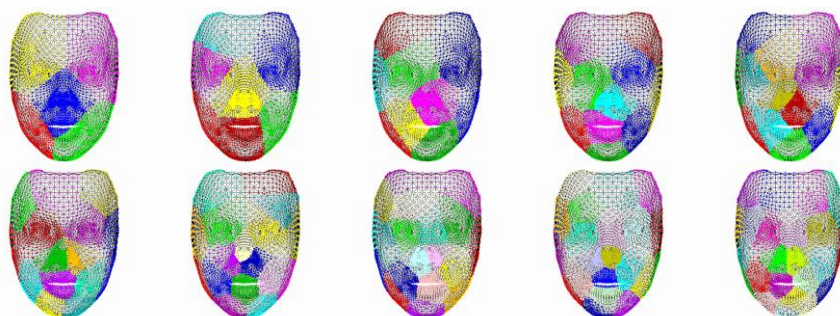
Dựa trên tập dữ liệu như vậy, việc thử nghiệm được tiến hành theo 3 trường hợp khác nhau: biến đổi giữa những người khác nhau với các biểu cảm khác nhau, biến đổi giữa những người khác nhau trên cùng biểu cảm trung tính và biến đổi của một người với các biểu cảm khác nhau.

Với trường hợp biến đổi giữa những người khác nhau với các biểu cảm khác nhau, hai tập biến thể phân biệt phục vụ thử nghiệm có được bằng cách chia đôi ngẫu nhiên tập dữ liệu các mẫu mô hình.



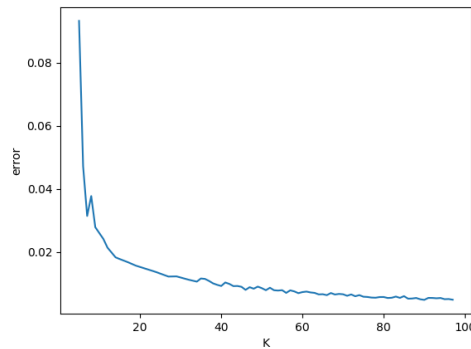
*Hình 3.2 Một số mẫu biến thể của đối tượng*

Trên tập biến thể phục vụ tính toán chọn điểm điều khiển, tiến hành tính toán tập đặc trưng biến dạng của các điểm, trên cơ sở đó thử nghiệm gom cụm với nhiều giá trị  $K$  khác nhau để xác định từng vùng có sự biến đổi tương đồng.



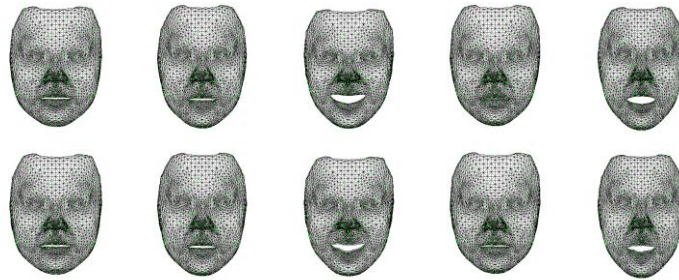
*Hình 3.3 Một số kết quả gom cụm với nhiều mức  $K$  (Với mỗi hình, các điểm cùng một cụm được tô cùng màu)*

Trên cơ sở từng cụm đã được tính toán với mỗi  $K$ , thực hiện tính toán tập điểm điều khiển và áp dụng tập điểm điều khiển để biến đổi mô hình 3D tham chiếu theo từng mẫu trong tập các biến thể phục vụ đánh giá chất lượng kết quả. Mỗi mô hình sau khi nắn chỉnh bằng kỹ thuật RBF sẽ được so sánh với mô hình mục tiêu để tính sai số. Vì bản chất dữ liệu trong mô hình 3D và đặc trưng biến dạng của điểm đều là một tập điểm 3D, sai số giữa mô hình sau khi nắn chỉnh và mô hình mục tiêu cũng được tính thông qua hàm khoảng cách giữa hai đặc trưng biến dạng. Giá trị sai số của tập điểm điều khiển trên tập biến thể phục vụ đánh giá sẽ được tính bằng trung bình sai số của từng mẫu được xét.



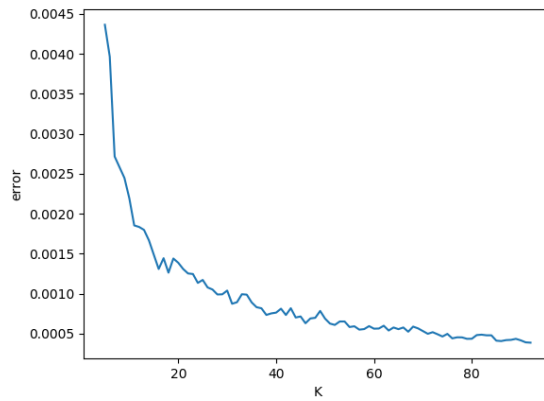
Hình 3.4 Biểu đồ tương quan giữa sai số với các giá trị  $K$

Để nhận thấy, khi tăng giá trị  $K$ , tức là tương ứng tăng dần kích thước của tập điểm điều khiển, độ chính xác của việc biến đổi cũng tăng. Đây cũng là cơ sở để ta có thể chọn một tập điểm điều khiển có kích thước hợp lý trong phạm vi sai số cho phép cũng như khả năng áp dụng kỹ thuật với các dạng biến thể trên các loại đối tượng khác nhau.

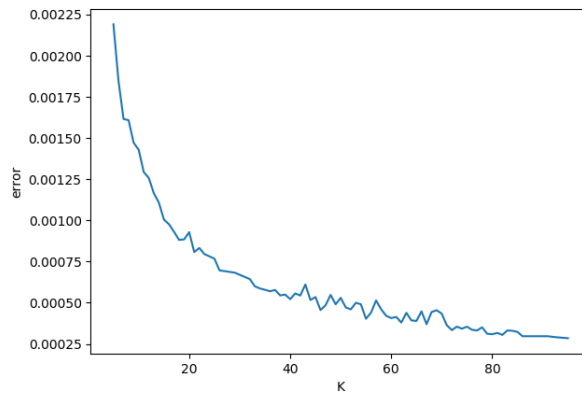


Hình 3.5 Một số hình ảnh kết quả nắn chỉnh: hàng 1 là các mô hình mục tiêu, hàng 2 là tương ứng các mô hình được nắn chỉnh từ mô hình 3D tham chiếu

Với hai trường hợp còn lại, cụ thể là biến đổi giữa những người khác nhau trên cùng biểu cảm trung tính và biến đổi của một người với các biểu cảm khác nhau, các bước tiến hành tính toán được thực hiện tương tự, chỉ khác ở việc tại hai tập biến thể phục vụ tính toán chọn điểm điều khiển và đánh giá chất lượng tập điểm điều khiển chọn được. Với trường hợp đánh giá giữa những người khác nhau trên cùng biểu cảm trung tính, các mẫu mô hình tương ứng với biểu cảm trung tính được chọn ra. Hai tập biến thể được tạo ra bằng cách tổng hợp tuyến tính giữa một số mô hình trong tập được chọn một cách ngẫu nhiên. Việc phân tích trong trường hợp này hướng đến tiêu chí biến đổi mô hình người này sang người khác. Tương tự với trường hợp biến đổi của một người với các biểu cảm khác nhau, với một người cụ thể, tập được chọn là các mô hình tương ứng với người đó và việc phân tích trong trường hợp này hướng đến tiêu chí biến đổi các trạng thái biểu cảm trên một người cụ thể.



Hình 3.6 Biểu đồ tương quan giữa sai số với các giá trị  $K$  với trường hợp biến đổi các biểu cảm khác nhau của người mẫu KA



Hình 3.7 Biểu đồ tương quan giữa sai số với các giá trị  $K$  với trường hợp biến đổi giữa những người khác nhau trên cùng biểu cảm trung tính

Ta dễ dàng nhận thấy, trong cả 3 trường hợp thử nghiệm, khi giá trị  $k$  tăng lên, sai số đều dần tiệm cận 0, điều này cũng rất rõ ràng với ý tưởng của việc biến đổi mô hình dựa trên tập điểm điều khiển. Tuy nhiên, sai số trong trường hợp biến đổi giữa những người khác nhau trên cùng biểu cảm trung tính nhỏ hơn nhiều so với hai trường hợp còn lại, điều này cũng phản ánh bản chất của phép nội suy RBF là biến đổi toàn cục, việc biến đổi này phù hợp hơn so với trường hợp biến đổi giữa các biểu cảm khác nhau của khuôn mặt người – những biến đổi thường chỉ mang tính chất cục bộ trên một số thành phần khuôn mặt như miệng, mắt.

### 3.4 Tổng kết chương

Chương này, đã trình bày bài toán và một số cách tiếp cận để biểu diễn cử chỉ, biểu cảm khuôn mặt. Và đề xuất một mô hình mô phỏng biểu cảm khuôn mặt 3D, trong đó đã chỉ ra rằng việc mô phỏng lại biểu cảm khuôn mặt dựa vào các đặc trưng thực chất là việc nội suy nhằm tính lại bề mặt 3D của khuôn mặt dựa theo các đặc

trung biểu cảm.

Kỹ thuật nội suy đối tượng 3D dựa vào hàm cơ sở bán kính (RBF) được lựa chọn sử dụng trong chương này. RBF cho kết quả có độ chính xác cao và thường được sử dụng trong lĩnh vực khoa học máy tính. Từ đó luận án xây dựng hệ thống nội suy thể hiện các cử chỉ, biểu cảm của khuôn mặt người 3D

Đóng góp chính của chương này là đề xuất kỹ thuật xác định tập điểm điều khiển phục vụ nắn chỉnh biến dạng mô hình. Trong phần này, tác giả đã trình bày các bước xác định quỹ đạo biến đổi của các điểm, sau đó tiến hành gom cụm quỹ đạo biến đổi rồi từ đó với mỗi cụm sẽ chọn điểm có sự biến đổi mạnh nhất và yếu nhất làm điểm điều khiển. Qua thử nghiệm với dữ liệu giả lập mặt cầu 3D hay với dữ liệu mô hình mặt tương ứng với ảnh chụp người thật cho thấy sự tương quan giữa kích thước của tập điểm điều khiển và độ chính xác của việc biến đổi nghĩa là khi tăng giá trị  $K$ , tức là tương ứng tăng dần kích thước của tập điểm điều khiển, độ chính xác của việc biến đổi cũng tăng dần. Đây cũng là cơ sở để có thể chọn một tập điểm điều khiển có kích thước hợp lý trong phạm vi sai số cho phép cũng như khả năng áp dụng kỹ thuật với các dạng biến thể trên các loại đối tượng khác nhau.

## KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Luận án đã hoàn thành được mục tiêu nghiên cứu được đặt ra ban đầu. Trình bày các hướng tiếp cận hiện có để giải quyết bài toán rút trích đặc trưng cử chỉ, biểu cảm và bài toán biểu diễn biểu cảm khuôn mặt. Đây là hai bài toán được ứng dụng nhiều trong thực tế.

Luận án đã chọn 1 kỹ thuật và đề xuất 2 kết quả nghiên cứu như sau:

Thứ nhất, chọn kỹ thuật phát hiện khuôn mặt trong ảnh dựa trên ý tưởng cơ bản là thực hiện phân lớp nhị phân với mỗi vùng ảnh quan tâm, từ đó kết luận vùng ảnh đó là khuôn mặt hay không. Ý tưởng cải tiến là kết hợp với kỹ thuật phân đoạn video dựa trên trừ ảnh đặc trưng. Kết quả cho thấy tốc độ phát hiện mặt người chuyển động được cải thiện. Kết quả nghiên cứu được công bố trong công trình TCTN3.

Thứ hai, luận án đề xuất kỹ thuật ước lượng biểu cảm khuôn mặt đã được định vị tập điểm điều khiển thông qua thuật toán AAM. Ý tưởng chính của đề xuất là sử dụng 3 loại đặc trưng hình dạng sau đó xây dựng hàm quyết định dựa trên cơ sở so sánh giá trị đặc trưng với một ngưỡng. Thông qua kết quả thực nghiệm trên cơ sở dữ

liệu JAFFE tác giả lựa chọn ngưỡng 0.13 là khá tốt khi kiểm chứng lại bằng mắt thường, theo đó, độ chính xác đạt khoảng 94,31%. Ngoài ra để tiện đánh giá, tác giả đã xây dựng lược đồ thống kê tỉ lệ chính xác theo ngưỡng chấp nhận.

Thứ ba, luận án đề xuất kỹ thuật xác định tập điểm điều khiển phục vụ nắn chỉnh biến dạng mô hình. Ý tưởng là dựa trên việc phân tích một tập các mô hình quan sát của đối tượng quan tâm rồi sau đó gom cụm và lựa chọn 2 điểm điều khiển trong cụm với tiêu chí ít biến đổi nhất và biến đổi nhiều nhất. Sau đó kết hợp với thuật toán nắn chỉnh biến dạng và một tập các mô hình biến thể để phục vụ đánh giá chất lượng nắn chỉnh. Kết quả thử nghiệm với dữ liệu giả lập mặt cầu 3D hay với dữ liệu mô hình mặt tương ứng với ảnh chụp người thật cho thấy sự tương quan giữa kích thước của tập điểm điều khiển và độ chính xác của việc biến đổi. Đây cũng là cơ sở để có thể chọn một tập điểm điều khiển có kích thước hợp lý trong phạm vi sai số cho phép cũng như khả năng áp dụng kỹ thuật với các dạng biến thể trên các loại đối tượng khác nhau.

Hướng phát triển:

Mặc dù luận án đã đề xuất được 3 kết quả nghiên cứu, tuy nhiên công đoạn mô phỏng lại những biểu cảm sau khi ước lượng được lên mô hình khuôn mặt 3D một cách chính xác và hợp lý trong ngữ cảnh thì vẫn còn bỏ ngỏ.

(a) Ví dụ điển hình như khi một người miệng đang cười nhưng chân mày và mắt có yếu tố bẩm sinh là sụp xuống thì hệ thống có thể ước lượng thành vừa vui và vừa buồn, hay là khi người đó nổi giận nhưng lại thể hiện nhiều qua màu sắc của da mặt (ửng đỏ) chứ ko hẳn là qua cử chỉ trên khuôn mặt và như vậy việc ước lượng trạng thái biểu cảm sẽ không còn chính xác nữa dẫn đến mô phỏng trạng thái sẽ bị sai.

(b) Hiện tại hệ thống trải qua nhiều công đoạn xử lý hình ảnh từ khâu phát hiện khuôn mặt rồi mới đến xác định tập điểm điều khiển sau đó mới ước lượng đặc trưng rồi cuối cùng mới đến nội suy mô phỏng biểu cảm. Như vậy, khi xử lý rời rạc sẽ làm chậm toàn bộ hệ thống.

Từ những nhược điểm trên, tác giả đề xuất ý tưởng cải tiến hệ thống bằng cách xây dựng chức năng học mẫu trạng thái cân bằng của đối tượng cần theo dõi để sau đó khi cử chỉ thay đổi thì ước lượng được chính xác hơn và sẽ khắc phục được nhược điểm (a). Còn để tăng tốc độ xử lý khắc phục được nhược điểm (b) thì ý tưởng là sẽ rút gọn các công đoạn của hệ thống, nghĩa là sẽ xây dựng các mô hình để học mẫu và phát hiện ra trực tiếp các đặc trưng của khuôn mặt trên toàn bộ khung ảnh và tính toán các góc nghiêng, vị trí tương quan giữa các đặc trưng để đi trực tiếp đến kết luận trạng thái biểu cảm.

## DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ

### QUỐC TẾ

TCQT1. **Huỳnh Cao Tuan**, Do Nang Toan, Lam Thanh Hien, Thanh-Lam Nguyen (2019), “An innovative approach to automatically identify control point set for model deformation rectification”, *International Journal of Advanced and Applied Sciences*”, 6(8), Pages: 45-52.

TCQT2. Do Nang Toan, **Huỳnh Cao Tuan**, Ha Manh Toan (2018), “A novel selection model of random features for the estimation of facial expression”, *International Journal of Advanced and Applied Sciences*, 5(6), Pages: 56-60

TCQT3. **Huỳnh Cao Tuan**, Do Nang Toan, Lam Thanh Hien (2021), “Automatic selection of key points for 3D-face deformation”, *Journal of Advances in Information Technology*.

### TRONG NƯỚC

TCTN1. **Huỳnh Cao Tuấn**, Đỗ Năng Toàn, Nguyễn Thanh Bình, Lâm Thành Hiền (2018), “Một kỹ thuật nhận dạng biểu cảm khuôn mặt dựa trên mô hình chất liệu”, *Tạp chí khoa học công nghệ thông tin và truyền thông*, số 4 (CS.01).

TCTN2. Lâm Thành Hiền, **Huỳnh Cao Tuấn**, Nguyễn Trọng Vinh (2014), “Phân đoạn video dựa trên kỹ thuật trừ ảnh đặc trưng”, *Tạp chí Khoa học và Công nghệ* 52 (1B) 150-162.

TCTN3. **Huỳnh Cao Tuấn**, Đỗ Sĩ Trường, Nguyễn Thanh Bình, Lâm Thành Hiền, (2020) “Phát hiện khuôn mặt người trong ảnh bằng kỹ thuật phân lớp nhị phân”, *Tạp chí khoa học công nghệ thông tin và truyền thông*. Vol 1 No 4A

HNTN1. **Huỳnh Cao Tuấn**, Đỗ Năng Toàn, Hà Mạnh Toàn (2018) – “Một kỹ thuật xác định tập điểm điều khiển phục vụ nắn chỉnh biến dạng mô hình”, *Kỷ yếu Hội nghị KH-CN Quốc gia lần thứ XI về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR)*; Hà Nội, ngày 09-10/8, DOI: 10.15625/vap.2018.00041.