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

HUỲNH CAO TUẤN

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

LUẬN ÁN TIẾN SĨ KỸ THUẬT

Hà Nội – 2021

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

HUỲNH CAO TUẤN

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

Chuyên ngành : Hệ thống thông tin
Mã số : 9.48.01.04

LUẬN ÁN TIẾN SĨ KỸ THUẬT

NGƯỜI HƯỚNG DẪN KHOA HỌC

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

Hà Nội – 2021

LỜI CẢM ƠN

Xin chân thành cảm ơn Quý thầy/cô Học viện Công nghệ Bưu Chính Viễn Thông đã giúp đỡ tận tình để nghiên cứu sinh hoàn thành Luận án. Xin bày tỏ lòng biết ơn sâu sắc đến hai vị Thầy kính yêu, đó là PGS.TS. Đỗ Năng Toàn và TS. Nguyễn Thanh Bình; những người đã tận tình hướng dẫn, gợi ý và cho nghiên cứu sinh những lời khuyên hết sức bổ ích trong suốt quá trình thực hiện luận án. Bên cạnh đó nghiên cứu sinh xin gửi lời cảm ơn đến Ban Lãnh đạo, những thầy cô quản lý Sau đại học của Học viện Công nghệ Bưu Chính Viễn Thông đã hỗ trợ trong quá trình nghiên cứu sinh nghiên cứu tại Trường. Nghiên cứu sinh xin chân thành cảm ơn Trường Đại học Lạc Hồng là nơi nghiên cứu sinh đã và đang làm việc; qua đây xin trân trọng gửi lời cảm ơn chân thành đến Ban lãnh đạo Nhà Trường vì đã không ngừng quan tâm, khích lệ và tạo điều kiện thuận lợi để nghiên cứu sinh có thời gian hoàn thành nhiệm vụ nghiên cứu và thực hiện Luận án này. Đồng thời, nghiên cứu sinh cũng xin cảm ơn sự hợp tác của các đồng nghiệp ở Viện công nghệ thông tin - Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Qua sự hợp tác này đã giúp nghiên cứu sinh nhìn nhận vấn đề nghiên cứu ở nhiều góc độ khác nhau, hoàn thiện hơn thông qua các buổi thảo luận của nhóm. Bên cạnh đó, nghiên cứu sinh xin cảm ơn sự động viên, chia sẻ của các anh/chị nghiên cứu sinh trong suốt thời gian cùng nhau học tập và nghiên cứu tại Trường. Cuối cùng, con xin chân thành cảm ơn Cha, Mẹ đã luôn động viên trong quá trình học tập.

Trân trọng,

Huỳnh Cao Tuấn

LỜI CAM ĐOAN

Trên cơ sở nghiên cứu lý thuyết, kiến thức thực tiễn và dưới sự hướng dẫn khoa học của PGS.TS. Đỗ Năng Toàn và TS. Nguyễn Thanh Bình.

Tôi xin cam đoan rằng, bản luận án tiến sĩ này là công trình nghiên cứu của bản thân và không sao chép nội dung từ các luận án hay công trình khác.

Các kỹ thuật và những kết quả trong luận án là trung thực, được đề xuất từ những yêu cầu thực tiễn, chưa từng được công bố trước đây.

Nghiên cứu sinh

Huỳnh Cao Tuấn

MỤC LỤC

LỜI CẢM ƠN i
LỜI CAM ĐOAN ii
MỤC LỤC iii
DANH MỤC CÁC TỪ VIẾT TẮT v
DANH MỤC KÝ HIỆU TOÁN HỌC vii
DANH MỤC CÁC BẢNG viii
DANH MỤC HÌNH ix
MỞ ĐẦU 1
1. Sự cần thiết của đề tài 1
2. Đối tượng, Mục tiêu, Phạm vi và Phương pháp nghiên cứu 3
3. Các đóng góp của luận án 5
4. Bố cục luận án 6
Chương 1 7
TỔNG QUAN VỀ TRÍCH CHỌN ĐẶC TRƯNG VÀ BIỂU DIỄN BIỂU CẢM KHUÔN MẶT 7
1.1. Tổng quan về một hệ thống biểu cảm khuôn mặt 7
1.1.1. Khái niệm biểu cảm mặt người 8
1.1.2. Định nghĩa điểm điều khiển 10
1.1.3. Một số vấn đề trong một hệ thống mô phỏng biểu cảm 10
1.2. Nhận diện và trích chọn đặc trưng biểu cảm 12
1.2.1. Một số phương pháp phát hiện khuôn mặt trong ảnh 12
1.2.2. Bài toán trích chọn đặc trưng biểu cảm khuôn mặt 16
1.2.3. Một số cách tiếp cận trích chọn đặc trưng 17
1.3. Biểu diễn biểu cảm khuôn mặt 25
1.3.1. Các khía cạnh hình học của khuôn mặt người 26
1.3.2. Một số vấn đề trong biểu diễn biểu cảm 27
1.3.3. Một số phương pháp tiếp cận giải quyết bài toán 28
1.4. Tổng kết chương và vấn đề nghiên cứu 37
Chương 2 39
KỸ THUẬT TRÍCH CHỌN ĐẶC TRƯNG CỬ CHỈ VÀ ƯỚC LƯỢNG BIỂU CẢM 39
2.1 Bài toán phát hiện khuôn mặt 39
2.1.1 Mô hình khuôn mặt mẫu 39
2.1.2 Phát hiện khuôn mặt 44
2.1.3 Kết quả thử nghiệm 44
2.2 Bài toán trích chọn đặc trưng cử chỉ 46

2.2.1 Loại bỏ nhiễu hình ảnh 48
2.2.2 Làm mịn bảo toàn cạnh của khuôn mặt 49
2.2.3 Hiệu chỉnh ánh sáng và cải thiện đặc trưng khuôn mặt 50
2.2.4 Trích chọn đặc trưng cử chỉ của khuôn mặt 51
2.3 Ước lượng biểu cảm khuôn mặt 61
2.4 Đề xuất kỹ thuật ước lượng biểu cảm khuôn mặt 64
2.4.1 Mô hình ước lượng 64
2.4.2 Đề xuất sử dụng đặc trưng hình dạng 65
2.4.3 Hàm quyết định 66
2.4.4 Thử nghiệm 67
2.5 Tổng kết chương 72
Chương 3 74
BIỂU DIỄN CỬ CHỈ, BIỂU CẢM TRÊN KHUÔN MẶT 3D 74
3.1 Bài toán biểu diễn cử chỉ, biểu cảm khuôn mặt 3D 74
3.2 Kỹ thuật nội suy RBF 76
3.2.1 Ý tưởng tiếp cận của thuật toán 76
3.2.2 Lựa chọn tham số hình dạng 77
3.2.3 Thuật toán nội suy 79
3.3 Nội suy thể hiện các cử chỉ, biểu cảm của khuôn mặt người 3D 82
3.3.1 Ước lượng tập vector đặc trưng 83
3.3.2 Tính toán mô hình đích 85
3.3.3 Tính toán quá trình biến đổi 86
3.4 Xác định tập điểm điều khiển phục vụ nắn chỉnh biến dạng 88
3.4.1 Đặc trưng biến dạng của các điểm 88
3.4.2 Xác định các nhóm tương đồng và chọn điểm điều khiển 89
3.4.3 Phân tích hiệu quả trong biến đổi mô hình 91
3.4.4 Thử nghiệm nắn chỉnh mô hình 92
3.5 Đề xuất xây dựng hệ thống mô phỏng biểu cảm khuôn mặt 3D 97
3.6 Tổng kết chương 98
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 100
DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ 102
TÀI LIỆU THAM KHẢO 103

DANH MỤC CÁC TỪ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng Việt
2D	Two Dimension	Hai chiều
3D	Three Dimension	Ba chiều
AAM	Active Appearance Model	Mô hình định vị điểm điều khiển
AF	Average Filter	Bộ lọc trung bình
AMF	Adaptive Median Filter	Bộ lọc trung vị
ANN	Artificial Neural Network	Mạng thần kinh nhân tạo
ASM	Active Shape Model	Mô hình hình dạng
AUs	Facial Action Units	Đơn vị hành động trên khuôn mặt
BEs	Basic Emotions	Cảm xúc cơ bản
BF	Bilateral Filter	Bộ lọc song phương
CEs	Compound Emotions	Cảm xúc tổng hợp
CLAHE	Contrast-limited adaptive histogram equalization	Cân bằng biểu đồ thích ứng có giới hạn tương phản
CLM	Constrained Local Model	Mô hình cục bộ ràng buộc
ECG	Electrocardiogram	Điện tâm đồ
EEG	Electroencephalogram	Điện não đồ
ELM	Extreme Learning Machine	Máy học cực độ
EMG	Electromyography	Điện cơ
FACS	Facial Action Coding System	Hệ thống mã hóa hành động trên khuôn mặt
FER	Facial Expression Recognition	Nhận dạng biểu cảm khuôn mặt
FLs	Facial Landmarks	Dấu mốc trên khuôn mặt
GF	Gaussian Filter	Bộ lọc Gaussian
HMM	Hidden Marko Model	Mô hình ẩn Marko

Từ viết tắt	Tiếng Anh	Tiếng Việt
HSV	HUE, SATURATION, VALUE	Vùng màu, độ bảo hòa màu, giá trị độ sáng của màu sắc.
JAFFE	Japanese Female Facial Expression	Cơ sở dữ liệu chứa những biểu cảm khuôn mặt phụ nữ ở Nhật Bản
LBP	Local Binary Pattern	Mẫu nhị phân cục bộ
LGBP	Local Gabor Binary Pattern	Mẫu nhị phân Gabor cục bộ
LVQ	Learning Vector Quantization	Lượng tử hóa Vector
MEs	Micro Expressions	Biểu cảm quy mô
NURBS	Non-uniform rational B-spline	Mô hình biểu diễn đường cong và bề mặt
PCA	Principal Component Analysis	Phân tích thành phần chính
RBF	Radial Basis Functions	Hàm cơ sở bán kính
RGB	(RED, GREEN, BLUE)	Hệ màu đỏ, xanh lá, xanh dương
SVM	Support Vector Machine	Máy vector hỗ trợ
TV	Television	Tivi

DANH MỤC KÝ HIỆU TOÁN HỌC

$LBP(.)$	Hàm xác định mẫu nhị phân cục bộ
$WMSE(I, v, w)$	Hàm mục tiêu phân lớp dữ liệu
$B(I)$	Hàm so sánh cường độ điểm ảnh trên ảnh I
α $(.)$	Ký hiệu tham số biến đổi tỉ lệ hình ảnh
β $(.)$	Ký hiệu tham số dịch chuyển
$\delta(.)$	Ký hiệu tham số độ lệch vector cường độ ảnh
Δ	Vector độ lệch
f_{LINE_LINE}	Hàm đặc trưng tính tỉ số độ dài 2 đoạn thẳng
$f_{TRIANGLE_TRIANGLE}$	Hàm đặc trưng tính tỉ số diện tích hai tam giác
$f_{LINELINE_LINELINE}$	Hàm đặc trưng tính tỉ số của tổng độ dài 2 đoạn thẳng
$\varepsilon(.)$	Ký hiệu tham số chỉ ra giá trị min của hàm cơ sở bán kính

DANH MỤC CÁC BẢNG

Bảng 2.1 Một số kết quả ví dụ so sánh cường độ ảnh 43
Bảng 2.2 So sánh kết quả định vị khuôn mặt.....45
Bảng 2.3 Sai số trung bình của các loại biểu cảm.....68
Bảng 2.4 Thống kê thời gian xử lý của các loại biểu cảm 72

DANH MỤC HÌNH

Hình 0.1 Các nhân vật trong phim Avatar 1
Hình 0.2 Biểu cảm trạng thái mặt trong phim hoạt hình 3D.....2
Hình 1.1 Minh hoạ phương pháp sử dụng điểm đánh dấu trong phim Avatar 7
Hình 1.2 Trạng thái biểu cảm khuôn mặt cơ bản [82] 9
Hình 1.3 Mẫu khuôn mặt trong phương pháp định vị [97].....14
Hình 1.4 Kiến trúc mạng UnitBox của nhóm Yu [129].....14
Hình 1.5 Quy trình xử lý của nhóm Yang [124].....15
Hình 1.6 Mô hình của nhóm Bai [14] 15
Hình 1.7 Kiến trúc của FANet [134].....16
Hình 1.8 Cảm xúc và không gian V-A [24], ...17
Hình 1.9 Quy trình nhận dạng biểu cảm khuôn mặt 17
Hình 1.10 Ví dụ về phương pháp mã LBP.....18
Hình 1.11 Ví dụ về toán tử LBP mở rộng với các vùng tròn có bán kính và số điểm ảnh lân cận tương ứng (8,1), (16,2) và (24,3) 19
Hình 1.12 Ví dụ về toán tử LBP mở rộng với các vùng tròn có bán kính và số điểm ảnh lân cận tương ứng (8,1), (12, 2.5) và (16, 4) 19
Hình 1.13 Đặc trưng Haar-like cơ bản.....23
Hình 1.14 Các đặc trưng Haar-like mở rộng.....23
Hình 1.15 Tính giá trị ảnh tích phân tại điểm P có tọa độ (x, y) 24
Hình 1.16 Tính nhanh tổng các giá trị điểm ảnh của vùng D trên ảnh 24
Hình 1.17 Hệ thống điểm của MPEG-4 4 [62] 27
Hình 1.18 Đặc trưng được đưa ra của nhóm Chuang [28].....28
Hình 1.19 Nội suy tuyến tính khuôn mặt từ trạng thái tự nhiên sang trạng thái cười 31
Hình 1.20 Nghiên cứu của Tong-Yee Lee và Po-Hua Huang [109].....35
Hình 1.21 Nghiên cứu của Henry Schäfer và cộng sự [52] 36
Hình 2.1 Quy trình trích chọn đặt trưng cử chỉ khuôn mặt.....39

Hình 2.2 Ví dụ của phép so sánh cường độ 43
Hình 2.3 Sơ đồ module hệ thống.....53
Hình 2.4 Phép biến đổi Đồng dạng 54
Hình 2.5 Mặt nạ kết cấu hình dạng 57
Hình 2.6 Một số kết quả định vị tập điểm điều khiển.....68
Hình 2.7 Các biểu đồ thống kê tỉ lệ chính xác theo ngưỡng chấp nhận 70
Hình 2.8 Biểu đồ thống kê tỉ lệ chính xác theo từng loại biểu cảm.....71
Hình 3.1 Tập điểm của Luxand.....75
Hình 3.2 Điều khiển hoạt động của Khủng Long dựa vào điểm điều khiển.....76
Hình 3.3 Hàm cơ sở bán kính (RBF) 77
Hình 3.4 Đầu vào của thuật toán nội suy dựa vào RBF.....80
Hình 3.5 Đầu vào của thuật toán nội suy dựa vào RBF.....81
Hình 3.6 Kết quả nội suy đối tượng.....82
Hình 3.7 Mô hình mẫu ở trạng thái cân bằng (a), và trạng thái đích (b) 83
Hình 3.8 Lược đồ nội suy biểu cảm.....88
Hình 3.9 Ví dụ dữ liệu mô hình mặt 3D 93
Hình 3.10 Một số mẫu biến thể của đối tượng.....93
Hình 3.11 Một số kết quả gom cụm với nhiều mức K.....94
Hình 3.12 Biểu đồ tương quan giữa sai số với các giá trị K.....94
Hình 3.13 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.....95
Hình 3.14 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.....96
Hình 3.15 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 96
Hình 3.16 Mô hình hệ thống mô phỏng biểu cảm khuôn mặt 97
Hình 3.17 Sơ đồ hệ thống.....98

MỞ ĐẦU

1. Sự cần thiết của đề tài

Gần đây, có nhiều bộ phim không phải do con người đóng, vì trên thực tế khó để có thể tìm được diễn viên có ngoại hình đáp ứng những yêu cầu như những nhân vật trong phim Avatar, hay như con khỉ đột trong phim KingKong… Tuy nhiên, điều gì khiến cho con Khỉ đột hung dữ trong KingKong hay nhân vật Jake và Neytiri trong Avatar có thể giành được nhiều tình cảm từ khán giả đến vậy? Có lẽ một trong những yếu tố quan trong nhất chính là chúng ta cảm thấy đồng cảm với nhân vật 3D thông qua cử chỉ trên khuôn mặt nhân vật.

Hình 0.1 Các nhân vật trong phim Avatar

Nghiên cứu về biểu cảm khuôn mặt nhằm mục đích xây dựng những nhân vật hoạt hình có các cử chỉ, biểu cảm như những nhân vật ngoài thế giới thực nhằm ứng dụng trong những bộ phim hoạt hình (ví dụ: Đi tìm Nemo), hoặc trong các bộ phim thực tại tăng cường, ví dụ như: Avatar, Transformers, Avanger, The Lost of Ring ... và trong xây dựng nhân vật của những trò chơi 3D là một công việc quan trọng và đang được sự quan tâm của nhiều nhà khoa học trong các lĩnh vực khác nhau từ các nghệ sỹ hội họa điêu khắc đến các nhà khoa học nghiên cứu về nhân học cho đến những nhà khoa học trong lĩnh vực công nghệ thông tin. Và 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 để nào đó được đặt ra.

Hình 0.2 Biểu cảm trạng thái mặt trong phim hoạt hình 3D

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… Trạng thái biểu cảm khuôn mặt trong ảnh là một vấn đề quan trọng trong một hệ thống phân tích và tổng hợp dữ liệu khuôn mặt người, đây cũng là một trong những bài toán trọng tâm trong những hướng nghiên cứu về dữ liệu đa phương tiện trong thời gian qua. 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…

Tuy có xuất hiện trong nhiều loại ứng dụng khác nhau như vậy nhưng việc sử dụng biểu cảm khuôn mặt trong những ứng dụng đó có thể có những sự khác biệt lớn tùy vào yêu cầu thực tế, chẳng hạn:

Trong một hệ giám sát hoạt động từ luồng video thu tại những địa điểm không có yêu cầu cao về tính bảo mật như thang máy, hành lang, cửa vào một cơ quan… khi đó cần thiết phải có những chức năng tự động phát hiện toàn bộ những khuôn mặt trên dòng video trong thời gian thực, đặc biệt là những khuôn mặt có sự chuyển động. Trong hệ thống này, đòi hỏi các thuật toán trích rút đặc trưng được thực hiện một cách nhanh chóng. Theo các yêu cầu đó, thông thường các giải thuật được thiết kế sẽ dựa trên một mô hình giám sát đối tượng cùng với việc triển khai các phương pháp phát hiện các chuyển động cục bộ cũng như toàn cục trên khung hình.

Các hệ thống biên tập ảnh, thường giúp cho những người không cần có quá nhiều kiến thức về đồ họa và do đó, chương trình xây dựng nhiều tác vụ trang điểm

và làm đẹp. Chẳng hạn như tinh chỉnh toàn bộ hoặc một phần trong khuôn mặt, hiệu chỉnh một số đặc trưng trên da như làm mịn, mụn nhọn, tàn nhang, giảm thiểu các nếp nhăn, hiệu chỉnh sự chiếu sáng trên vùng mặt, khử hiện trượng mắt đỏ… Để có thể hoàn thành được các tác vụ đó và đảm bảo sự thay đổi vẫn mang cảm giác tự nhiên cho người xem, một yêu cầu gần như bắt buộc là phải xác định được một cách chính xác các điểm đặc trưng cho các thành phần khuôn mặt, chẳng hạn như các điểm góc mắt, các điểm cạnh môi, điểm mũi, đường viền cằm… Do vậy, các giải thuật được thiết kế thường không yêu cầu cao về tốc độ thực hiện mà là yêu cầu độ chính xác thật cao trong việc xác định tập điểm đặc trưng chi tiết trên khuôn mặt và do chương trình cũng không yêu cầu phải tự động hoàn toàn nên một hướng tiếp cận được quan tâm đặt ra là một kịch bản sử dụng bán tự động yêu cầu người dùng chọn một lượng giới hạn các điểm trên khuôn mặt như góc miệng, góc mắt, đỉnh mũi để làm cơ sở cho việc tính toán nốt các điểm đặc trưng chi tiết khác trên khuôn mặt.

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.

Với mục đích kết nối giữa hai hướng, cũng như các nghiên cứu về vấn đề trích chọn đặc trưng khuôn mặt mà không sử dụng điểm đánh dấu và tái thể hiện lại các biểu cảm của khuôn mặt trên mô hình 3D dựa trên các đặc trưng đã thu được. Nhất là, ở Việt Nam hiện nay, chưa có nhiều nghiên cứu về vấn đề này, mặc dù những ứng dụng thực tế đang đòi hỏi có những cách giải quyết cụ thể, chẳng hạn như phần mềm nhận dạng và biểu diễn ngôn ngữ cử chỉ cho người khiếm thính, bài toán phát thành viên ảo, cử chỉ của robot, phim ảnh, quảng cáo... vì sự hấp dẫn của bài toán những ứng dụng thực tế của nó, nghiên cứu sinh đã chọn đề tài nghiên cứu : "*Phát triển các kỹ thuật biểu diễn cử chỉ, trạng thái khuôn mặt 3D dưới sự trợ giúp của camera*".

2. Đối tượng, Mục tiêu, Phạm vi và Phương pháp 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 phần 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…

Trong nghiên cứu này, thao tác quan trọng là phát hiện được tập điểm điều khiển từ ảnh khuôn mặt. Do đó các khuôn mặt cần quan sát được rõ ràng vị trí của các điểm điều khiển, vì vậy phạm vi nghiên cứu của luận án là khuôn mặt thẳng hoặc khá gần với khuôn mặt thẳng có góc nghiêng dưới 30 độ.

Để giải quyết các bài toán đã nêu trên, nghiên cứu sinh đề xuất xây dựng một hệ thống rút trích các cử chỉ, biểu cảm khuôn mặt rồi mô phỏng lại trên 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.

c) Phương pháp nghiên cứu

Phương pháp nghiên cứu được áp dụng là nghiên cứu lý thuyết, những công trình đã công bố và kết hợp với thực nghiệm. Các vấn đề cần giải quyết liên quan đến các giải thuật và lý thuyết xử lý ảnh và đồ họa máy tính được thực hiện trên phần mềm máy tính với đầu vào là các thông tin thu nhận được từ thực tế. Quá trình được tiến hành dựa trên việc tìm hiểu tài liệu, cài đặt thử nghiệm sau đó đánh giá kết quả thực nghiệm và cải tiến nhằm nâng cao chất lượng.

Để chứng minh các ưu điểm của giao thức đề xuất trong luận án, sử dụng phương pháp so sánh, đối chiếu đối với mô hình đã được công bố. Để triển khai các phương pháp nghiên cứu nêu trên, tiến hành thực hiện các bước như sau:

- Tìm hiểu các kết quả nghiên cứu trong các bài toán tương tự. Đánh giá các hướng nghiên cứu đó và các kết quả đạt được và từ đó đề xuất mô hình tốt hơn.
- Dựa trên các mô hình đề xuất: Lựa chọn các mô hình trích chọn và biểu diễn phù hợp. Đồng thời lựa chọn thông số phù hợp, chứng minh được ưu điểm của thuật toán.
- Lập trình để kiểm nghiệm kết quả.
- So sánh kết quả đạt được với các nghiên cứu trước trong cùng điều kiện.

3. 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.

4. 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

Hệ thống biểu cảm khuôn mặt người là một lĩnh vực nghiên cứu của ngành thị giác máy tính và đang nhận được nhiều sự quan tâm của các nhà khoa học trong và ngoài nước. 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.

Hình 1.1 Minh hoạ phương pháp sử dụng điểm đánh dấu trong phim Avatar

- 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.

- Bài toán biểu diễn biểu cảm cho mô hình ảo 3D có đầu vào là trạng thái biểu cảm liên tục, đầu ra là biểu cảm của nhân vật ảo thể hiện trạng thái biểu cảm đó. Việc mô phỏng lại biểu cảm khuôn mặt dựa vào các đặc trưng đã rút trích được thực chất là việc nội suy nhằm tính lại bề mặt 3D của khuôn mặt dựa theo các đặc trưng biểu cảm. Hiện nay có rất nhiều phương pháp nội suy khác nhau như NURBS, RBF, Affine, nội suy dựa vào mạng neural, v.v. 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.

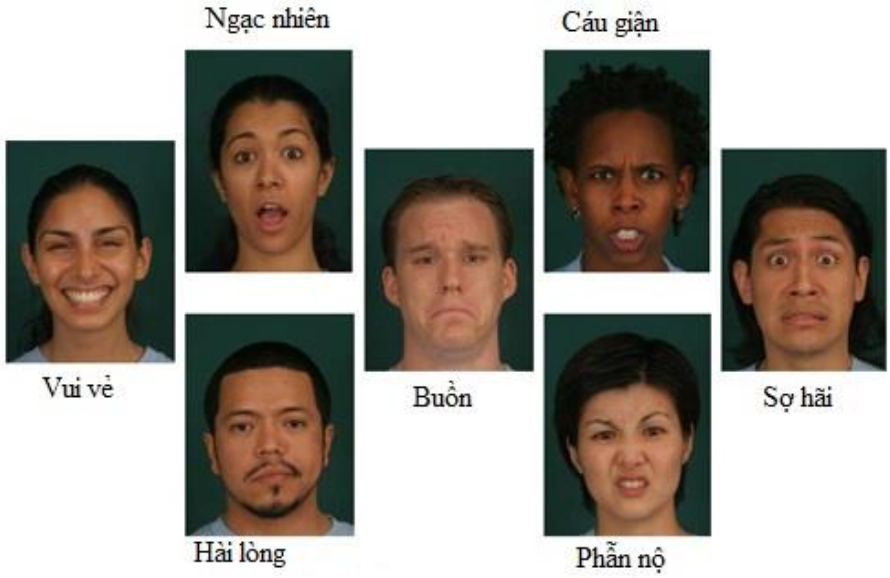
Ekman và Friesen đã mô tả 6 loại biểu hiện tiêu biểu riêng biệt cùng với những nội dung của khuôn mặt người, bao gồm: hạnh phúc, buồn bã, sợ hãi, ghê tởm, ngạc nhiên và tức giận. Những biểu hiện này là phổ biến và độc lập với những điều kiện như dân tộc hay nền văn hóa; Thời gian qua có rất nhiều các nỗ lực nghiên cứu đã được hướng vào mô hình trên khuôn mặt thực tế và biểu hiện trên khuôn mặt hoạt hình: Bickel và các cộng sự [19] trình bày một phương pháp mới để tạo hoạt ảnh thời gian thực cho các biểu hiện trên khuôn mặt có độ chi tiết cao dựa trên sự phân hủy nhiều tỷ lệ của hình học trên khuôn mặt thành chuyển động quy mô lớn và các chi tiết tỷ lệ nhỏ, chẳng hạn như nếp nhăn biểu cảm. Hoạt ảnh kết hợp của Bickel được điều chỉnh để phù hợp với các đặc điểm cụ thể của biến dạng khuôn mặt quy mô lớn và quy mô nhỏ; Ma Thị Châu và cộng sự [3] đã nghiên cứu tái tạo khuôn mặt 3 chiều từ hộp sọ; Phạm Thế Bảo và cộng sự [2] đã trình bày nghiên cứu tổng quan các phương pháp xác định khuôn mặt người; Đỗ Năng Toàn và cộng sự [7] đã đề xuất 02 kỹ thuật liên quan đến mô hình hóa 3D, nâng cao chất lượng mô hình hóa và điều khiển mô hình. Đó là xây dựng mô hình 3D cho các hệ xương, hô hấp và tiêu hóa của cơ thể người; Đỗ Năng Toàn và Nguyễn Văn Huân [5] trình bày nghiên cứu một số kỹ thuật mô phỏng tóc ứng dụng trong thực tại ảo; Trịnh Hiền Anh, Đỗ Năng Toàn cùng cộng sự [1] [57] [6] đã nghiên cứu công nghệ thực tại ảo và chèn đối tượng ảo trong quảng cáo trực tuyến.

1.1.1. Khái niệm biểu cảm mặt người

Biểu cảm là biểu hiện trạng thái hoạt động, tính cách, tình cảm, và tâm lý của một người, tương ứng với một trạng thái của hệ cơ nét mặt nằm dưới vùng da mặt.

Biểu cảm của người nói có tác động lớn đối với người nghe và hiểu đúng biểu cảm có ý nghĩa rất quan trọng, đôi khi mang tính quyết định trong giao tiếp xã hội. Chúng ta thường quan sát biểu cảm của người khác bởi vì đó là dấu hiệu cho thấy hiện tại người đó đang trong tình trạng như thế nào, họ sẽ làm gì... Hiểu được biểu cảm không giúp chúng ta đọc được suy nghĩ nhưng nó có thể giúp chúng ta khẳng định thêm ý nghĩa của lời nói lúc đó [36]. Từ những nhận định trên, biểu cảm khuôn mặt con người trở nên có tính giao tiếp nhất trên cơ thể, cung cấp thêm nhiều thông tin trong quá trình giao tiếp.

Các nhà nghiên cứu tâm lý thì nhìn nhận biểu cảm dựa trên những trạng thái tâm lý của con người và cho rằng nó có cường độ nhất định, tồn tại trong một khoảng thời gian tương đối. Trạng thái tâm lý của một người tại một thời điểm nào đó sẽ có tác động mạnh vào các quyết định hành vi, thao tác, hoạt động của họ. Trạng thái biểu cảm là cách mà con người chia sẻ cảm xúc của mình. Những nghiên cứu trước đây đã cho thấy cử động khuôn mặt đóng vai trò quan trọng trong thể hiện biểu cảm. Các nghiên cứu đã cho thấy sự liên hệ giữa cử động khuôn mặt và trạng thái biểu cảm, điều này được khẳng định trong các công trình nghiên cứu của Darwin năm 1972.



Hình 1.2 Trạng thái biểu cảm khuôn mặt cơ bản [82]

Nhiều nghiên cứu đã thống kê, biểu cảm khuôn mặt có một số loại phổ biến, có tính chất tương đồng. Phương pháp mô phỏng biểu cảm được chia thành hai loại: Phương pháp tĩnh và phương pháp động.

- Phương pháp thể hiện biểu cảm tĩnh [10] [92] là thể hiện trạng thái biểu cảm không liên tục và không tạo ra một quy trình, cơ chế theo thời gian cho việc tạo biểu cảm.
- Phương pháp thể hiện biểu cảm động [79] [132] lưu lại sự thay đổi cường độ biểu cảm theo thời gian và cung cấp một phương thức đồng nhất cho việc tạo biểu cảm trên khuôn mặt, và giải quyết được hạn chế của phương pháp thể hiện biểu cảm tĩnh. Phương pháp này tạo ra biểu cảm liên tục theo cơ chế ánh xạ trực tiếp từ trạng thái biểu cảm, sau đó được mô phỏng trên khuôn mặt. Nhược điểm của nó là sẽ tạo ra biểu cảm không tự nhiên khi có một trạng thái biểu cảm xảy ra với cường độ cao và thời gian dài.

1.1.2. Định nghĩa điểm điều khiển

Một đối tượng ảnh nói chung có thể có nhiều sự biến dạng về mặt hình dạng. Để có thể mô hình hóa được những sự biến dạng này có một cách tiếp cận là sử dụng những điểm điều khiển. Đây là những điểm mốc đặc trưng cho sự biến đổi về hình dạng của đối tượng. Ví dụ, khi con người di chuyển và hành động, những điểm mốc đặc trưng cho sự biến dạng của tư thế người đó là những vị trí khớp; khi ta cử động tay thì những điểm điều khiển có thể được chọn là điểm khuỷu tay, cổ tay… Ta có thể hiểu, đối với một đối tượng ảnh, các điểm điều khiển là những điểm đặc trưng cho sự biến đổi về hình dạng của đối tượng đó. Một đối tượng có bao nhiêu điểm điều khiển, điểm điều khiển có những thông tin gì (chẳng hạn 2D hay 3D) và điểm điều khiển tác động như thế nào đến hình ảnh kết quả phụ thuộc vào bản thân đối tượng và yêu cầu của bài toán.

1.1.3. Một số vấn đề trong một hệ thống mô phỏng biểu cảm

Nếu chúng ta muốn máy tính thực sự thông minh và tương tác với chúng ta một cách tự nhiên thì chúng cần phải có khả năng mô hình hóa, nhận dạng, và thể hiện biểu cảm. Trong lĩnh vực nghiên cứu về nhân vật ảo, biểu cảm nhận được nhiều sự quan tâm bởi ảnh hưởng của nó trong việc tạo các nhân vật ảo tin cậy [23]. Câu hỏi đặt ra là làm thế nào để cung cấp biểu cảm cho nhân vật ảo? Như vậy có hai vấn đề cần quan tâm khi giải quyết bài toán cung cấp biểu cảm cho nhân vật ảo: thứ nhất là cung cấp trạng thái biểu cảm cho nhân vật ảo, thứ hai là cung cấp cơ chế thể hiện biểu cảm cho nhân vật ảo. Nhằm cải tiến sự tương tác giữa người và máy tính, một nhân vật ảo trong máy tính có thể thể hiện biểu cảm khi mà biểu cảm không thực sự tồn tại

bên trong nó. Tuy nhiên, việc này không cung cấp một cơ chế nhất quán cho việc thể hiện biểu cảm, khiến cho nhân vật ảo trở nên khó hiểu và kém thuyết phục. Ngược lại, khi nhân vật ảo đã được cung cấp trạng thái biểu cảm nhưng cơ chế thể hiện biểu cảm không tốt cũng sẽ khiến nhân vật ảo kém tự nhiên. Vì vậy, cách thức hiệu quả nhất đó là sử dụng các kỹ thuật mô hình hóa cho việc cung cấp trạng thái biểu cảm cũng như việc thể hiện biểu cảm cho nhân vật ảo.

Có nhiều hình thức mô phỏng biểu cảm khuôn mặt khác nhau tùy vào cách tiếp cận. Có hướng tiếp cận dựa trên việc nhận dạng một tập giới hạn các biểu cảm rồi mô phỏng tương ứng loại biểu cảm và cường độ, có hướng tiếp cận bằng cách rút trích một số đặc trưng khuôn mặt và ánh xạ qua mô hình khuôn mặt trên máy tính mà không cần quan tâm ý nghĩa của những biểu cảm đó. Ngay trong những hướng tiếp cận trên cũng đã có nhiều mô hình mô phỏng biểu cảm khuôn mặt được đề xuất như: hệ thống phân tán [113], hệ thống liên kết [73], hệ thống dựa trên kế hoạch (plan based system) [46], hệ thống dựa trên luật mờ [38]... Tuy nhiên, có rất ít mô hình giải quyết được một cách đầy đủ và tối ưu các bài toán cài đặt biểu cảm trên máy tính, đó là: sự linh động, độc lập với ứng dụng, biểu cảm cần phải gắn liền với cá tính và trạng thái tâm lý, biểu cảm phải có cường độ và cơ chế phân rã theo thời gian....

Bui và cộng sự [23] đã đề xuất ParleE – là một hệ thống các trạng thái biểu cảm. ParleE xây dựng các sự kiện dựa trên việc học và một giải thuật lập lịch nhằm mô hình hóa cá tính và biểu cảm kèm theo quyết định cách mà khuôn mặt ảo thể hiện biểu cảm thích hợp ở các cường độ khác nhau. Liên quan đến biểu diễn biểu cảm, có hai trường phái nghiên cứu: những người làm việc trong lĩnh vực khoa học máy tính và những họa sỹ.

Trong lĩnh vực khoa học máy tính thì nghiên cứu, thử nghiệm và triển khai những phương pháp mô hình hóa, định vị và mô phỏng khuôn mặt người dựa trên nhiều tiếp cận khác nhau, chẳng hạn phân tích những vùng bất biến trong ảnh đồng thời phân biệt với những dữ liệu khác để xác định vị trí và phạm vi mặt người trong một bức ảnh hoặc trong luồng video. Những kết quả này là một phần quan trọng giúp thiết kế những hệ thống bám sát điểm đặc trưng khuôn mặt. Nhiều công trình khác xem xét vấn đề theo hướng kết hợp dữ liệu ảnh với những nghiên cứu giải phẫu cơ thể người, tâm lý học hay nhân chủng học. Những nghiên cứu này tận dụng một số loại mặt mẫu có sẵn để làm cơ sở cho những đánh giá phân tích khuôn mặt hoặc mô phỏng lại chuyển động của cơ, mô với nhiều mức độ khác nhau. Ngoài ra còn hướng

tiếp cận từ ứng dụng thực tế, như tương tác giữa người và máy bằng cử chỉ trên khuôn mặt, nhận dạng biểu cảm khuôn mặt, biểu diễn khuôn mặt từ lời thoại.

Đối với hoạ sĩ, những người làm công việc mang tính chất kỹ năng nhiều hơn như: tạo ra các hoạt cảnh khuôn mặt chất lượng cao cho cho phim ảnh, cho những trò chơi, những chương trình giải trí trên ti vi, quảng cáo… Những người làm công việc này sẽ không quan tâm đến việc nghiên cứu đề xuất một phương pháp mới mà chủ yếu là họ dựa vào kinh nghiệm, sự rèn luyện kỹ năng, thao tác cùng với những công cụ tốt nhất mà họ có. Chẳng hạn như sử dụng phần mềm mô hình hóa 3D giúp họ làm việc hiệu quả.

1.2. Nhận diện và trích chọn đặc trưng biểu cảm

Biểu hiện trên khuôn mặt là một phương tiện phi ngôn ngữ để thể hiện ý định trong giao tiếp của con người. Nhiều nghiên cứu cho thấy rằng 55% thông điệp liên quan đến cảm xúc và thái độ là ở biểu hiện trên khuôn mặt, 7% trong số đó là các từ được nói, phần còn lại là ngôn ngữ (cách mà các từ nói).

Trong những năm gần đây, với sự phát triển nhanh chóng của trí tuệ nhân tạo, các bài toán nhận dạng tự động nhận được sự quan tâm rộng rãi và nghiên cứu chuyên sâu như: tâm lý học, thị giác máy tính và nhận dạng khuôn mặt. Nhận dạng biểu cảm trên khuôn mặt có các ứng dụng rộng rãi trong nhiều lĩnh vực, bao gồm tương tác giữa người và máy tính [30]; [16], thực tế ảo [17], thực tế tăng cường [13] [25], hệ thống hỗ trợ lái xe nâng cao [12] [60], giáo dục và giải trí.

Nhiều loại dữ liệu khác nhau có thể được sử dụng làm đầu vào cho bài toán. Hình ảnh mặt người là dữ liệu đầu vào chủ đạo và nhiều thách thức, vì nó có thể cung cấp thông tin dồi dào cho nghiên cứu nhận dạng biểu cảm. Bên cạnh hình ảnh khuôn mặt được chụp bằng máy ảnh còn có các tín hiệu về sinh lý [61] cho bài toán nhận dạng, ví dụ: điện cơ (EMG), điện tâm đồ (ECG), điện não đồ (EEG), cũng có thể được sử dụng làm nguồn dữ liệu phụ cho một số ứng dụng nhận dạng trong thế giới thực.

1.2.1. Một số phương pháp phát hiện khuôn mặt trong ảnh

Việc phát hiện khuôn mặt người trong ảnh đã được nghiên cứu nhiều và có nhiều phương pháp khác nhau [48] [123] [55] [53]; chẳng hạn như phương pháp phát hiện và định vị khuôn mặt từ ảnh có nền phức tạp nhằm tìm ra các cạnh, sau đó loại bỏ bớt

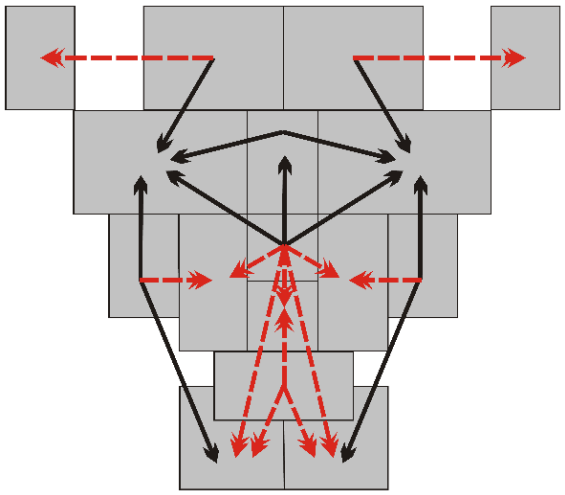
và nhóm các cạnh lại sao cho chỉ còn lại một biên bao quanh khuôn mặt phân biệt vùng đầu và nền [54]. Độ chính xác của giải thuật này đạt đến 80% với 48 ảnh có nền phức tạp.

Nghiên cứu của Cootes [59] dùng các đốm và vạch sọc để phát hiện khuôn mặt. Tiếp cận này dùng thuật toán Laplace để biến đổi hình ảnh và xác định các đốm, sau đó tìm các hình tam giác ứng với với các thành phần của khuôn mặt. Họ dùng các vạch sọc để ghi nhận nét mặt, đồng thời dùng hai hình tam giác để mô tả quan hệ giữa các đốm, 2 đốm tối và 3 đốm sáng để tả 2 mắt, 2 gò má và mũi. Khuôn mặt được phát hiện nếu các vạch sọc bao quanh các thành phần.

Hay các nghiên cứu của Guggisberg [47] đã dùng phương pháp định vị các đặc trưng khuôn mặt cho ảnh xám. Ý tưởng là: sử dụng các bộ lọc để làm nổi các biên, tiếp theo là làm nổi bật các vùng có hình dạng rõ ràng. Lúc này Historgram của ảnh có một đỉnh nổi bật, dựa vào đỉnh này, xác định các giá trị ngưỡng phù hợp để tạo ra 2 ảnh nhị phân tương ứng. Vùng ứng viên của khuôn mặt là vùng có các phần liên hệ giữa hai ảnh nhị phân này. Nhược điểm của phương pháp này là dùng các phép toán hình thái học nên khó có thể kết hợp với các đặc trưng khác để định vị một khuôn mặt.

Vào năm 2013, Li và Chung [71] đã đề xuất một cách tiếp cận để định vị khuôn mặt trong ảnh có nền phức tạp bằng cách sử dụng 5 đặc trưng là 2 lỗ mũi, 2 mắt, đoạn giữa mũi và môi để thể hiện một khuôn mặt chuẩn. Sử dụng mô hình Gauss để mô hình tập các khoảng cách, và với mỗi loại đặc trưng thì ta tính khoảng cách giữa chúng. Vì các đặc trưng không thể xuất hiện ngẫu nhiên nên ta dùng mô hình xác suất để định vị chúng qua khoảng cách.

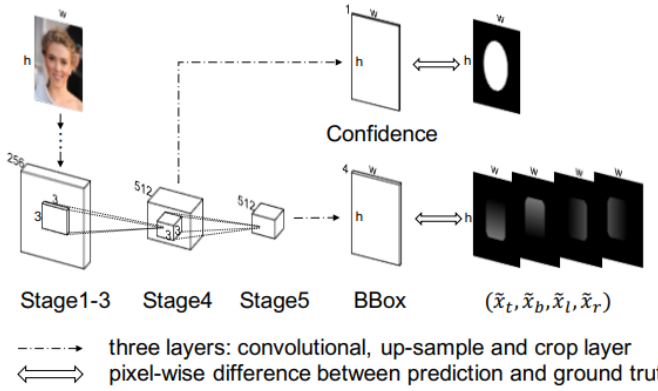
Còn Su và Zheng [104] đã dựa vào sự thay đổi mức độ sáng của các vùng khác nhau của khuôn mặt (như hai má, hai mắt và trán). Sau đó, xác định các cặp tỷ số của một số vùng cho ta một lượng bất biến khá rõ rang [131]. Sự thay đổi độ sáng của các vùng trên khuôn mặt tạo ra các cặp quan hệ sáng-tối giữa các vùng nhỏ đều được ghi nhận lại. Một khuôn mặt được định vị nếu nó thỏa mãn tất cả các cặp sáng-tối và mối quan hệ tương quan giữa các cặp đó với nhau.



Hình 1.3 Mẫu khuôn mặt trong phương pháp định vị [97]

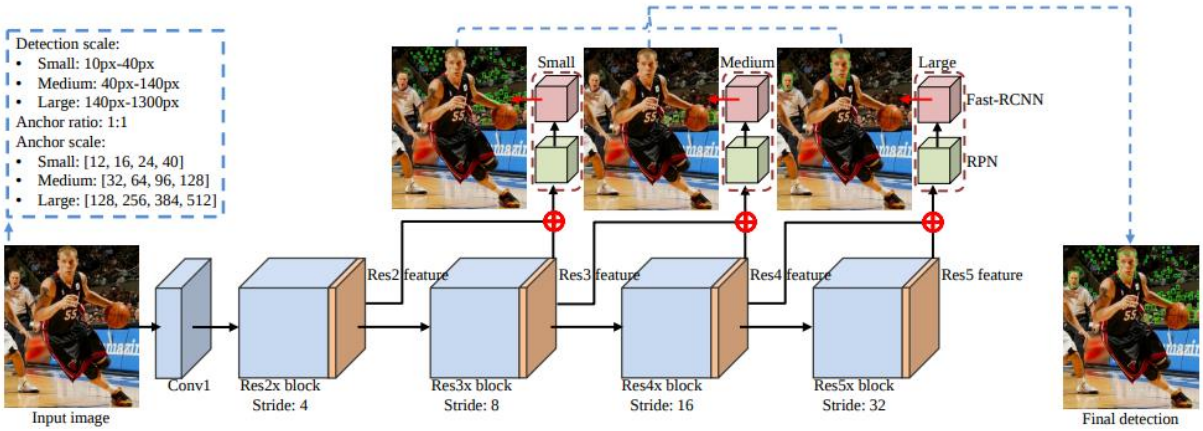
Error! Reference source not found. là một mẫu khuôn mặt có kích thước theo ỷ lệ 14x16 pixel được sử dụng trong phương pháp định vị khuôn mặt của Schneiderman và Kanade [98], gồm 16 vùng trên khuôn mặt và 23 quan hệ được thể hiện thông qua các mũi tên. Có 11 quan hệ thiết yếu (mũi tên đen) và 12 quan hệ xác thực (mũi tên đỏ). Wu và Toàn cùng cộng sự [123] [107] sử dụng đặc trưng Haar trong thuật toán AdaBoost để xác định một thành phần nằm trong một vùng chứa nó.

Trong vài năm gần đây, cũng có một số công trình ứng dụng kỹ thuật CNN vào trong bài toán phát hiện khuôn mặt. Chẳng hạn sử dụng CNN theo tiếp cận hồi quy hộp bao, nhóm Yu [129] đã đề xuất một hàm lỗi dùng độ đo IoU để hồi quy trực tiếp hộp bao để hứa hẹn mang lại sự hiệu quả đối với đối 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.



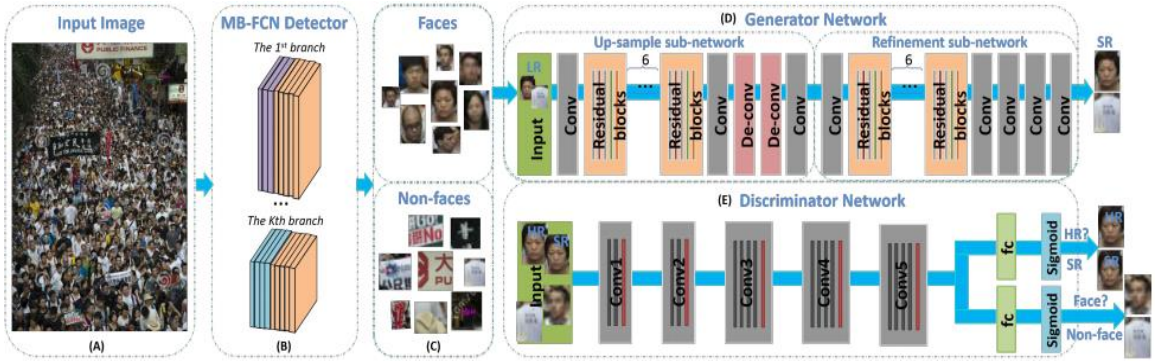
Hình 1.4 Kiến trúc mạng UnitBox của nhóm Yu [129]

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.



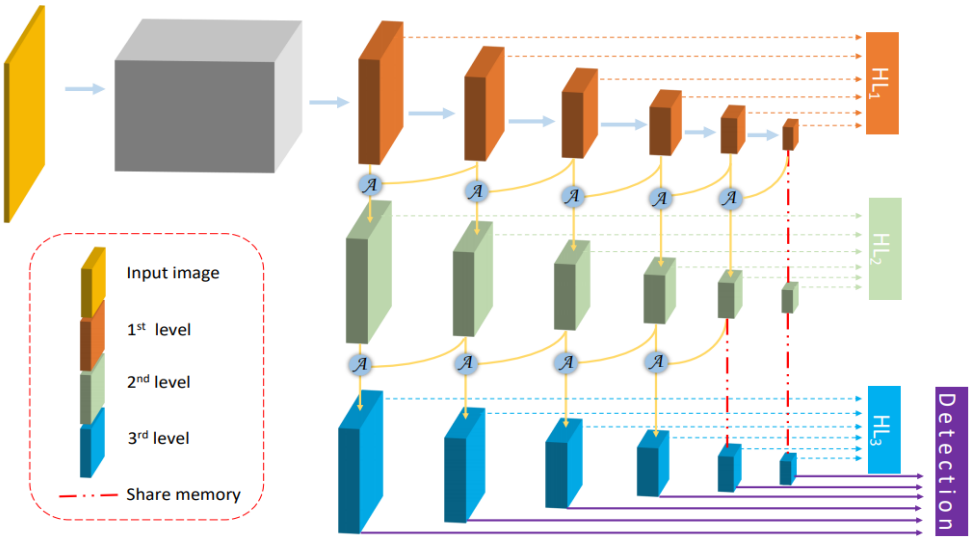
Hình 1.5 Quy trình xử lý của nhóm Yang [124]

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.



Hình 1.6 Mô hình của nhóm Bai [14]

Thêm vào đó, các kỹ thuật phát hiện đối tượng tổng quát dựa trên CNN cũng đưa đến một số ý tưởng trong phát hiện khuôn mặt. Chẳng hạn các công trình Face R-CNN [115], Face R-FCN [116] và FDNet [130] đã áp dụng kỹ thuật Faster R-CNN [94] và R-FCN [31] với nhiều chiến lược cụ thể để xử lý cho bài toán phát hiện khuôn mặt.. FaceBoxes [133] thiết kế một mô hình phát hiện khuôn mặt thời gian thực chạy trên CPU dựa trên kỹ thuật SSD [75]. S3FD [133] cũng đưa ra vài chiến lược khác nhau trên cơ sở SSD [75] để làm giảm bớt vấn đề gặp phải đối với các khuôn mặt nhỏ trong ảnh [74]. FANet [134] thì xây dựng một bộ dò khuôn mặt lấy cảm hứng từ FPN [72] với ý tưởng chính là khai thác các tính năng đa tỷ lệ vốn có của một mạng nơ ron phức hợp đơn lẻ bằng cách tổng hợp các đặc điểm ngữ nghĩa cấp cao hơn với các tỷ lệ khác nhau.

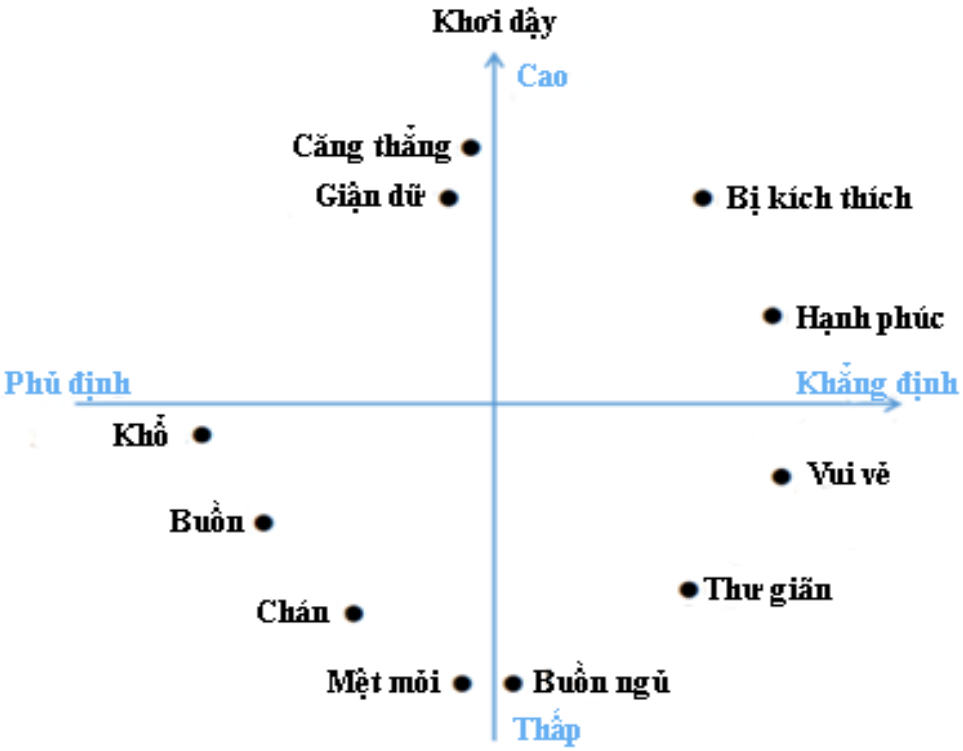


Hình 1.7 Kiến trúc của FANet [134]

1.2.2. Bài toán trích chọn đặc trưng biểu cảm khuôn mặt

Theo [106], công trình nghiên cứu về nhận dạng biểu cảm trên khuôn mặt thì đơn vị hành động (Action Units - AU) và không gian Valence – Arousal (V-A space hay không gian V-A) là hai mô hình phổ biến. Không gian V-A là một mô hình phổ quát được sử dụng rộng rãi trong các nhiệm vụ nhận dạng cảm xúc liên tục của các tín hiệu âm thanh, hình ảnh và sinh lý. Như thể hiện trong Hình 1.8 , không gian V-A xác định các loại cảm xúc theo giá trị của các chiều cảm xúc. Các AU mã hóa các chuyển động cơ bản của cơ mặt và sự kết hợp của các AU có thể được sử dụng cho

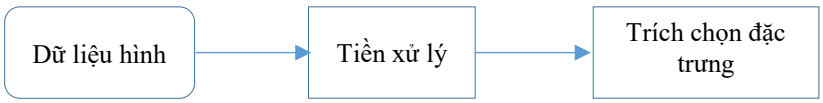
nhận dạng biểu cảm. Trong [24], một nền tảng được đề xuất áp dụng AUs để ước tính cường độ V–A.



Hình 1.8 Cảm xúc và không gian V-A [24],

Trong ba thập kỷ qua, nhiều nghiên cứu đã được tiến hành trong lĩnh vực nhận dạng đối tượng. Tuy nhiên, việc phát hiện khuôn mặt người vẫn là một nhiệm vụ khó khăn và nhiều thử thách. Điều này chủ yếu là do số lượng lớn các đặc điểm khác nhau trên khuôn mặt của mỗi người. Những đặc điểm đó có thể dựa trên giới tính, tuổi tác và dân tộc.

Bài toán trích chọn đặc trưng này được mô tả gồm hai bước chính:



Hình 1.9 Quy trình nhận dạng biểu cảm khuôn mặt

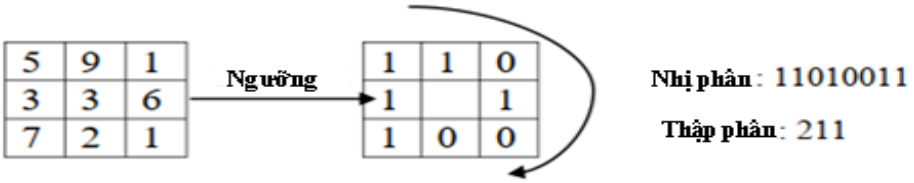
1.2.3. Một số cách tiếp cận trích chọn đặc trưng

Đây là quy trình để trích xuất dữ liệu hoặc thông tin hữu ích từ hình ảnh, ví dụ: giá trị, vectơ và ký hiệu. Các phương pháp trích xuất đặc trưng được sử dụng rộng rãi

cho bài toán nhận dạng chủ yếu gồm: trích xuất đặc trưng Gabor, Local Binary Pattern (LBP), Active Shape Model (ASM), trích xuất đặc trưng Haar-like, v.v.

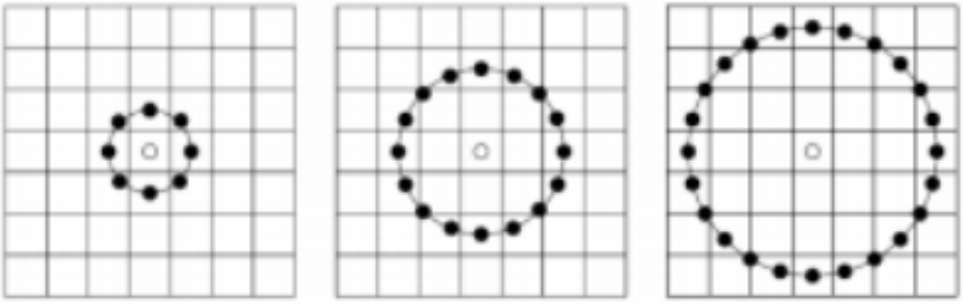
1.2.3.1. Mẫu nhị phân cục bộ (LBP)

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.

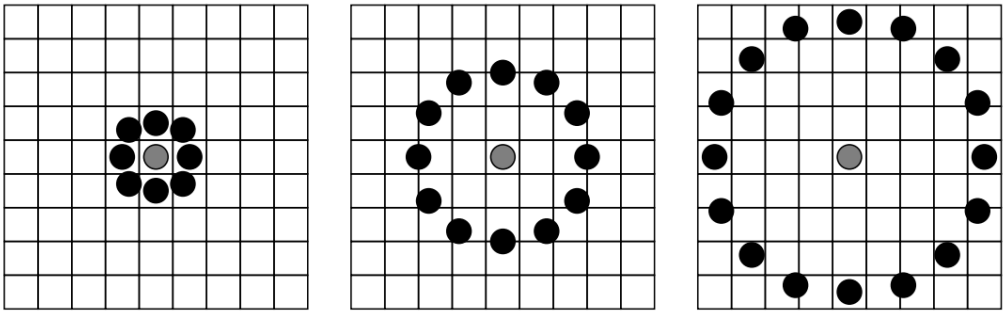


Hình 1.10 Ví dụ về phương pháp mã LBP

Một hạn chế của toán tử LBP cơ bản đó là kích thước của vùng ảnh là 3x3 quá nhỏ để mô tả hết đặc trưng nổi trội của những vùng ảnh có cấu trúc quy mô lớn. Để đáp ứng được ở các quy mô cấu trúc khác nhau thì toán tử sau tổng quát hóa để áp dụng cho các vùng ảnh lân cận của nó có kích thước khác nhau. Mỗi một vùng ảnh cơ sở được định nghĩa là một tập hợp các điểm lấy mẫu cách đều nhau trên một vòng tròn được căn giữa tại pixel được gán nhãn và các điểm lấy mẫu không rơi vào trong các pixel được nội suy bằng phép nội suy song tuyến tính. Do đó cho phép áp dụng cho bất kỳ bán kính và số điểm lấy mẫu bất kỳ trong vùng lân cận. Dưới đây cho thấy một số ví dụ về toán tử LBP mở rộng, biểu thị một lân cận các điểm lấy mẫu P phía trong một đường tròn bán kính R.



Hình 1.11 Ví dụ về toán tử LBP mở rộng với các vùng tròn có bán kính và số điểm ảnh lân cận tương ứng (8,1), (16,2) và (24,3)



Hình 1.12 Ví dụ về toán tử LBP mở rộng với các vùng tròn có bán kính và số điểm ảnh lân cận tương ứng (8,1), (12, 2.5) và (16, 4)

Xét một hình ảnh đơn sắc, nếu tọa độ của pixel trung tâm là (x_c, y_c) thì tọa độ của điểm P lân cận (x_p, y_p) trên cạnh của hình tròn có bán kính R có thể được tính như sau:

$$x_p = x_c + R\cos(2\pi p/P)$$
$$y_p = y_c + R\sin(2\pi p/P) \quad (1.1)$$

Nếu giá trị mức xám của pixel trung tâm là gc và giá trị mức xám của điểm lân cận với nó là g_p với p = 0,…, P-1. Khi đó kết cấu cục bộ T của vùng lận cận với điểm pixel trung tâm được định nghĩa như sau

$$T \approx t(g_c, g_0, g_1, \dots, g_{P-1}) \quad (1.2)$$

Mỗi giá trị thu được của các điểm này có thể được mô tả cấu trúc theo các cách khác nhau. Điều này được thực hiện bằng cách trừ giá trị pixel trung tâm bởi giá trị

của các điểm trên vòng tròn. Theo cách này thì cấu trúc cục bộ được biểu diễn dưới dạng phân phối chung của giá trị của pixel trung tâm và các các pixel khác

$$T \approx t(g_c, g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c) \quad (1.3)$$

Vì t(gc) mô tả độ chói tổng thể của hình ảnh, không liên quan đến kết cấu hình ảnh cục bộ, nó không cung cấp thông tin hữu ích cho phân tích kết cấu. Do đó phần lớn thông tin về đặc điểm của kết cấu trong phân phối chung ban đầu được lưu giữ trong phân phối chênh lệch (Ojala et al. 200)

$$T \approx t(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c) \quad (1.4)$$

Điều này có nghĩa là pixel lân cận (xung quanh) có giá trị xám cao khi so sánh với pixel ở trung tâm. Trong trường hợp đó, giá trị được gán là 1 nếu không được gán là 0

$$T \approx s(g_0 - g_c, g_1 - g_c, \dots, g_{P-1} - g_c) \quad (1.5)$$

Trong đó

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (1.6)$$

Tại bước cuối để tạo LBP cho pixel (xc, yc) một trọng số nhị phân 2p được gán cho mỗi dấu s(gp – gc) như sau

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \quad (1.7)$$

Một mẫu nhị phân cục bộ được gọi là đồng nhất nếu mẫu nhị phân chứa tối đa hai lần chuyển tiếp 0-1 hoặc 1-0. Trong thực tế, điều này có nghĩa là một mô hình LBP thống nhất là không có lần chuyển tiếp hoặc hai lần chuyển tiếp giữa bit 0 và 1. Chỉ một lần chuyển tiếp là không thể vì một chuỗi nhị phân ở đây cần được coi là vòng kín. (Ví dụ: Hai mẫu có số lần chuyển đổi là không, ví dụ tám bit, là 00000000 và 11111111. Ví dụ về các mẫu đồng nhất với tám bit và hai lần chuyển đổi là 00011100 và 11100001; và 01010100 (6 lần chuyển đổi) thì không là mẫu LBP đồng nhất).

Đối với các mẫu có hai lần chuyển đổi bit 0 và 1 là có thể kết hợp P(P-1). Với các mẫu đồng nhất có điểm lấy mẫu là P và bán kính R thì khái niệm $LBP_{P,R}^{u2}$ được sử dụng.

Chỉ sử dụng các mẫu LBP thống nhất có hai lợi ích. Đầu tiên là nó tiết kiệm bộ nhớ. Đối với các mẫu LBP không đồng nhất thì có thể phải 2P kết hợp. Với $LBP_{P,R}^{u2}$ chỉ có P(P-1) + 3 mẫu kết hợp. Ví dụ, số lượng mẫu có thể kết hợp cho vùng lân cận của 16 pixel là 65536 đối với mẫu LBP cơ bản và là 243 đối với mẫu $LBP_{P,R}^{u2}$ (LBP đồng nhất). Lợi ích thứ hai là mẫu LBP đồng nhất chỉ phát hiện các kết cấu cục bộ quan trọng như các điểm, đường cuối, cạnh và góc.

Phương pháp LBP áp dụng trên các hình ảnh để nhận dạng khuôn mặt, sẽ trích xuất các đặc trưng có thể được sử dụng để so khớp sự giống nhau giữa các hình ảnh này. Với mỗi pixel của một ảnh thì giá trị mã LBP được tính toán. Sự xuất hiện của từng mẫu trong ảnh được lần lượt xử lý. Biểu đồ (histogram) của những mẫu này còn được gọi là nhãn (label), tạo thành một vector đặc trưng và do đó là đại diện cho kết cấu của hình ảnh, biểu đồ này sau đó có thể được sử dụng để so sánh độ tương tự giữa các hình ảnh bằng cách tính khoảng cách giữa các biểu đồ.

1.2.3.2. ASM / AAM

Active Shape Model (ASM) được đề xuất trong [27] 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 thử nghiệm. Trong AAM, đối tượng được mô hình hóa bởi một tập điểm mô tả hình dạng và kết cấu hình ảnh của nó, kết cấu hình ảnh được lấy từ những giá trị cường độ ảnh trong vùng được giới hạn bởi tập điểm điều khiển.

Mô hình thống kê của đối tượng có thể mô tả được những biến thể về hình dạng và những biến thể về kết cấu hình ảnh cũng như mối tương quan giữa chúng. Các vấn đề nổi bật được quan tâm trong phương pháp này là việc xây dựng mô hình thống kê cho đối tượng ảnh và thiết kế thuật toán tối ưu để thực hiện tìm kiếm. Xây dựng mô hình thống kê toán học cho đối tượng bao gồm: xây dựng mô hình thống kê toán học cho hình dạng, mô hình thống kê toán học cho kết cấu hình ảnh và kết hợp hai mô hình đó để ra được mô hình cho đối tượng.

Mô hình hình dạng của đối tượng được biểu diễn bởi một tập hợp các điểm điều khiển có thứ tự. Vấn đề cần thiết trong xây dựng mô hình là phải lựa chọn được ra những điểm điều khiển phù hợp với loại đối tượng cần được mô hình hóa. Khi xây dựng tập dữ liệu huấn luyện, trong đó bao gồm một tập ảnh của đối tượng cùng với thông tin đánh dấu tập điểm điều khiển mô tả thể hiện của đối tượng được quan tâm tương ứng trong các ảnh. Đối tượng hình học chính là tập hợp các điểm điều khiển mô tả hình dạng của đối tượng có đánh thứ tự, được biểu diễn bởi n điểm trong không gian d chiều, và được thể hiện bằng một vector nd chiều với các giá trị trong vector chính là giá trị tọa độ cụ thể trên từng trục không gian của từng điểm. Ví dụ, trong ảnh 2 chiều, ta có thể mô tả n điểm điều khiển, $\{(x_i, y_i)\}$, thành vector 2n thành phần: $x = (x_1, \dots, x_n, y_1, \dots, y_n)^T$

Sau khi thể hiện lại các đối tượng hình dạng thành các vector tương ứng, thực hiện mô hình hóa trên dữ liệu là tập các vector này. Bước đầu là thực hiện chuẩn hóa các đối tượng hình dạng này vào chung một không gian tọa độ nhằm chuẩn hóa các đối tượng hình dạng để tổng các khoảng cách của mỗi đối tượng hình dạng tới kỳ vọng của chúng $D = (\sum_{i=1}^{n} |x_i - \bar{x}|^2)$ đạt cực tiểu trên PCA

Sau khi đã mô hình hóa được các đối tượng hình dạng, thực hiện tiếp quá trình mô hình hóa kết cấu hình ảnh của các đối tượng được quan tâm trong ảnh. Kết cấu hình ảnh của đối tượng được hiểu là các giá trị cường độ ảnh được giới hạn trong vùng ảnh được bao bởi đối tượng hình dạng tương ứng. Dữ liệu đối tượng hình dạng và kết cấu hình ảnh của đối tượng được quan tâm trong các mẫu có thể được biểu diễn dưới dạng các tham số [68]. Vấn đề tiếp theo là thực hiện mô hình hóa tổng quát cho các dữ liệu này và tương tự phương pháp được lựa chọn là phân tích thành phần chính PCA.

Giải thuật tìm kiếm tối ưu được sử dụng trong AAM được thiết kế cho phép tự động ước lượng các tham số mô hình cái mà có thể tổng hợp ra được ảnh mẫu gần

nhất có thể với ảnh mục tiêu đầu vào nhằm mục đích cực tiểu sự sai lệch giữa ảnh mẫu đầu vào và ảnh tổng hợp được theo mô hình đã được lựa chọn.

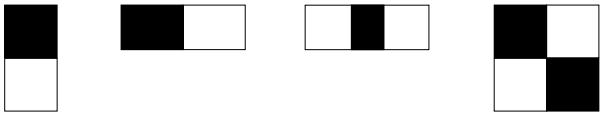
1.2.3.3. Trích xuất đặc trưng Gabor

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.

Gabor wavelet không nhạy cảm với cường độ chiếu sáng. Tuy nhiên, Gabor wavelet có thể tiêu tốn nhiều bộ nhớ vì nó thường hoạt động trên các đặc trưng toàn cục.

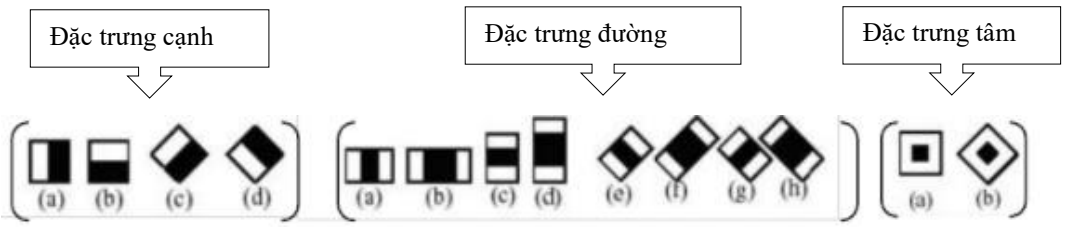
1.2.3.4. Trích xuất đặc trưng Haar-like

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". Hình dưới đây mô tả 4 đặc trưng Haar-like cơ bản như sau



Hình 1.13 Đặc trưng Haar-like cơ bản

Để sử dụng các đặc trưng này vào việc phát hiện khuôn mặt người, 4 đặc trưng Haar-like cơ bản được mở rộng bởi [101] và được chia thành nhóm các đặc trưng cạnh, đặc trưng đường và đặc trưng xung quanh tâm



Hình 1.14 Các đặc trưng Haar-like mở rộng

Giá trị của đặc trưng Haar-like là sự chênh lệch giữa tổng các điểm ảnh của các vùng màu đen và các vùng màu trắng [66]. Để tính các giá trị của đặc trưng Haar-like ta phải tính tổng của các vùng pixel trên ảnh. Nhưng để tính toán các giá trị của đặc trưng Haar-like cho tất cả các vị trí trên ảnh đòi hỏi chi phí tính toán khá lớn. Do đó, [114] [29] đã đưa ra khái niệm ảnh tích phân (Integral Image) để tính toán nhanh giá trị của các đặc trưng Haar-like. Integral Image là một mảng hai chiều có kích thước bằng với kích thước của ảnh cần tính giá trị đặc trưng Haar-like, với mỗi phần tử của mảng này được tính bằng tổng của các điểm ảnh phía trên và bên trái của nó

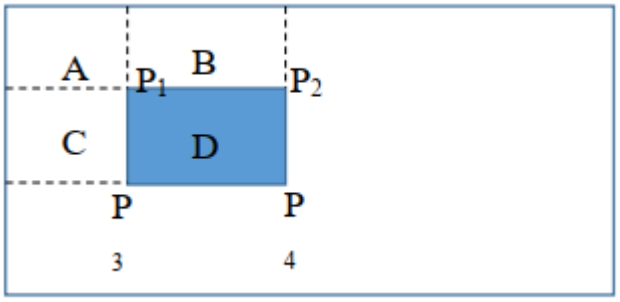
Hình 1.15 Tính giá trị ảnh tích phân tại điểm P có tọa độ (x, y)

Giá trị tích phân tại điểm P có tọa độ (x, y) được tính như sau

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y') \qquad (1.8)$$

Sau khi đã tính được ảnh tích phân, việc tính tổng các giá trị mức xám của một vùng ảnh bất kỳ nào đó trên ảnh ta thực hiện theo cách sau. Ví dụ tính giá trị của vùng D trong hình dưới như sau

D = A + B + C + D – (A+B) – (A + C) + A



Hình 1.16 Tính nhanh tổng các giá trị điểm ảnh của vùng D trên ảnh

Với A + B + C + D chính là giá trị tại điểm P4 trên ảnh tích phân, A + B là giá trị tại điểm P2, A + C là giá trị tại điểm P3 và A là giá trị tại điểm P1. Vì vậy, ta có thể tính lại biểu thức tính D ở trên như sau:

$$D = (x4, y4) – (x2, y2) – (x3, y3) + (x1, y1) \quad (1.9)$$

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.

1.3. Biểu diễn biểu cảm khuôn mặt

Mô phỏng biểu cảm khuôn mặt là bài toán tái tạo lại phần đầu của người thật với những biểu hiện trên khuôn mặt lên các mô hình 3D. Đó là một trong những vấn đề khó khăn nhất trong đồ họa máy tính. Mặc dù đã có nhiều sự tiến bộ đạt được trong việc thu thập, trích rút các đặc trưng biểu cảm khuôn mặt, cũng như các kỹ thuật biểu diễn sự biến đổi các nét biểu cảm lên các mô hình, tuy nhiên đây luôn là một vấn đề khó và phức tạp không chỉ với những người mới tiếp cận mà nó còn là vấn đề phức tạp với những chuyên gia trong lĩnh vực tạo và thể hiện biểu cảm khuôn mặt. Vì vậy, một hệ thống trực quan, dễ dàng và hiệu quả để tổng hợp biểu hiện trên khuôn mặt luôn là một ứng dụng hữu ích trong một loạt các ứng dụng của các ngành công nghiệp như phim, trò chơi video và hội thảo trực tuyến vv...

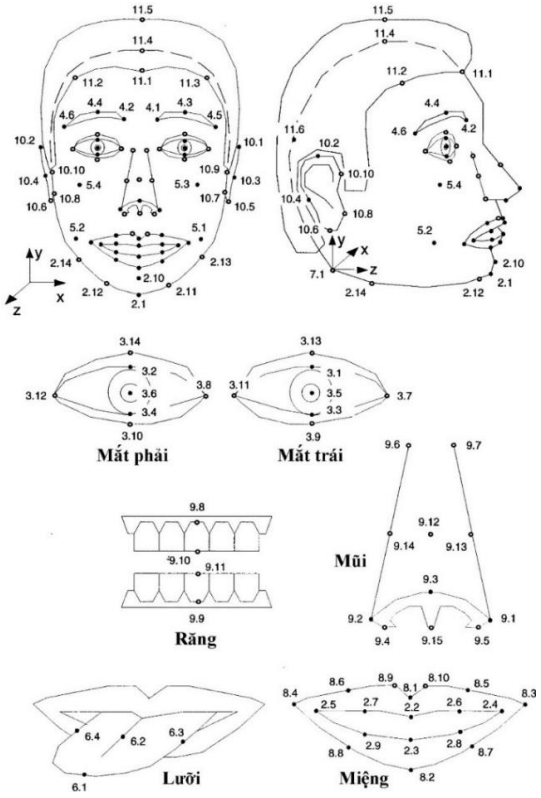
Hiện nay, có nhiều cách tiếp cận khác nhau trong việc biểu diễn khuôn mặt 3D; chẳng hạn, máy quét thường số hóa khuôn mặt 3D dưới dạng các đám mây điểm 3D dày đặc; tuy nhiên, kỹ thuật này tạo ra kết quả không trực quan vì mỗi phép đo bề mặt cho kết quả là một điểm 3D. Biểu diễn đám mây điểm có khả năng biểu diễn hoàn toàn các bề mặt 3D dạng tự do mở hoặc đóng (đặc trưng bởi hình dạng không đều). Trong một số hệ thống nhận dạng khuôn mặt, biểu diễn đám mây điểm 3D có thể được sử dụng trong các mô-đun khác nhau của hệ thống, cụ thể là tiền xử lý, trích xuất tính năng và đối sánh [105]; Tuy nhiên, 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.

Tập điểm điều khiển của một đối tượng là tập những điểm đặc trưng phân biệt có tính nhất quán trên nhiều quan sát khác nhau chúng thường là những điểm có tính dao động với cường độ cao. Ví dụ đối tượng quan sát là cái miệng trên khuôn mặt, các điểm đặc trưng có thể là những điểm khoé miệng hay hình dạng môi trên và môi dưới – những đặc điểm có thể dễ dàng xác định. Tính chất của tập điểm điều khiển trên một loại đối tượng cụ thể yêu cầu đến những tri thức về bản thân đối tượng trong bài toán sử dụng, do đó việc lựa chọn thường dựa trên tri thức chuyên gia, là một người có kiến thức vững chắc về đối tượng được lựa chọn.

1.3.1. Các khía cạnh hình học của khuôn mặt người

Việc nâng cao khả năng nhận dạng khuôn mặt người 3D đòi hỏi việc xác định hiệu quả các khía cạnh hình học đặc trưng cho khuôn mặt người. Tuy nhiên, đây là một yêu cầu khá khó khăn bởi vì khuôn mặt con người dễ bị thay đổi do quá trình lão hóa, giảm/ tăng cân, các tác động thay đổi diện mạo khuôn mặt (phẫu thuật thẩm mỹ, các biến dạng do tai nạn, v.v…). Những biến đổi này đặt ra thách thức đối với nhận dạng khuôn mặt 3D. Điều này là do các biến thể hình dạng 3D của khuôn mặt người giữa các cá nhân khác nhau là rất nhỏ về mặt thống kê. Trên thực tế, một số biểu hiện trên khuôn mặt, ngoài những thay đổi về hình học, có thể gây ra những thay đổi về cấu trúc liên kết đối với khuôn mặt 3D, chẳng hạn như những thay đổi liên quan đến việc mở miệng. Tuy nhiên, khuôn mặt người không phải tự nhiên dễ dàng thay đổi; đặc biệt là sự liên quan đến nét mặt vì cấu trúc giải phẫu của khuôn mặt không thay đổi. Yếu tố này làm cho việc mô hình hóa các biểu cảm khuôn mặt hoặc trích xuất các đặc điểm bất biến của biểu cảm để nhận dạng khuôn mặt 3D. Ngoài ra còn có một số vùng trên khuôn mặt như mũi và trán (được gọi là vùng bán cận) ít bị ảnh hưởng bởi các biểu hiện trên khuôn mặt. Một số vùng trên khuôn mặt vẫn có thể ít bị biến dạng hơn một số vùng khác, tùy thuộc vào biểu cảm của khuôn mặt. Do vậy, chúng ta có thể phát triển của các phương pháp tiếp cận nhận dạng khuôn mặt 3D với các biểu hiện trên khuôn mặt.

Chuẩn MPEG-4 [62] đặc tả 84 điểm đặc trưng trên khuôn mặt cùng với các tham số hoạt hóa. Các tham số này tương ứng với các cử động trên khuôn mặt và tạo ra sự biến dạng của mô hình mặt so với trạng thái trung tính. Việc mô phỏng quá trình biến dạng của mô hình với các tham số hoạt hóa và giá trị thời gian sẽ sinh ra chuỗi hoạt cảnh của khuôn mặt. Các điểm đặc trưng ứng với các vị trí trên khuôn mặt như vị trí mắt, mũi, miệng, v.v..



Hình 1.17 Hệ thống điểm của MPEG-4 4 [62]

1.3.2. Một số vấn đề trong biểu diễn biểu cảm

Có những khó khăn nhất định khi biểu diễn biểu cảm của khuôn mặt như sau:

- Lưới 3D yêu cầu nhiều bộ nhớ và lưu trữ hơn. Còn việc ánh xạ để cấu tạo khuôn mặt 3D có thể được thực hiện đơn giản hơn lưới 3D bởi vì kết quả được tạo ra giống với hình ảnh cấp độ xám 2D; do đó, chúng ta có thể sử dụng các kỹ thuật phân tích, xử lý hình ảnh 2D và thị giác máy tính, đặc biệt là kỹ thuật lọc và phân rã nhân. Tuy nhiên, việc xoay các khuôn mặt 3D có thể dẫn đến việc các phần của khuôn mặt bị chồng lấn lên nhau.

- Nhiều biểu cảm khác nhau như: Vui, buồn, ngạc nhiên, sợ hãi,…

- Những chi tiết không phải là đặc trưng của khuôn mặt như: mắt kính, râu…

- Khuôn mặt bị che khuất bởi một số đối tượng khác

- Chất lượng về hình ảnh

- Hướng của khuôn mặt nhìn từ trên xuống, nhìn nghiêng, nhìn thẳng. Và trong cùng một ảnh có thể có nhiều khuôn mặt ở những tư thế khác nhau.

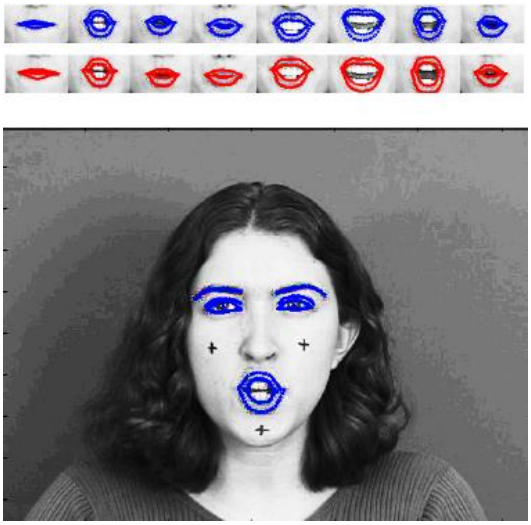
- Trong cùng một ảnh các khuôn mặt có kích thước khác nhau như: xa, gần, …

- Nhiều khuôn mặt có vùng da dính lẫn nhau.

Các khó khăn trên chứng tỏ rằng bất cứ phương pháp giải quyết (thuật toán) bài toán xác định khuôn mặt người và trích xuất các đặc trưng trên khuôn mặt sẽ không thể tránh khỏi một số khiếm khuyết nhất định.

1.3.3. Một số phương pháp tiếp cận giải quyết bài toán

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. Khó khăn của phương pháp này là mô hình rất phức tạp và không thể tái sử dụng lại trên khuôn mặt khác. Các hàm cơ sở xuyên tâm - xấp xỉ hàm [22] thường được triển khai để có được sự chuyển động của các điểm trên mô hình mục tiêu.



Hình 1.18 Đặc trưng được đưa ra của nhóm Chuang [28]

Đã có rất nhiều nghiên cứu tập trung vào những vấn đề được đưa ra trong một hệ thống mô phỏng biểu cảm khuôn mặt trong môi trường thực tại ảo, cụ thể hơn là vấn đề trích rút đặc trưng biểu cảm trên khuôn mặt thực tế và biểu diễn lại nó trong khuôn mặt ảo trên chương trình máy tính. Để mô tả được những hiệu ứng sẽ thể hiện trên mô hình khuôn mặt mục tiêu, nét mặt được phân tích với những đặc điểm chi tiết được lựa chọn trong một mô hình có trọng số mô tả đặc điểm vị trí của điểm quan trọng trên khuôn mặt. Hình dạng pha trộn [100] là một phương pháp được sử dụng

rộng rãi, trong đó mô hình thực hiện nội suy từ nhiều hình dạng đã được lựa chọn để có được những hình tổng hợp mong muốn cho mô hình mục tiêu. 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. Một vấn đề trong việc xây dựng mô hình mô tả hình dạng khuôn mặt là các biến dạng hình học được chi phối bởi các cơ được thiết kế sẵn và bề mặt mô. Trên cơ sở đó, 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 thưa thớt những điểm điều khiển tự định nghĩa.

1.3.3.1. Nội suy tuyến tính

Trong toán học, nội suy tuyến tính [20] là phương pháp đường cong phù hợp sử dụng đa thức tuyến tính để xây dựng các điểm dữ liệu mới trong phạm vi của một tập rời rạc các điểm dữ liệu đã biết.

Giả sử chúng ta có hai điểm đã biết tọa độ (x_0, y_0) và (x_1, y_1) thì nội suy tuyến tính chính là đường thẳng giữa hai điểm này. Cho giá trị x trong khoảng (x_0, x_1) thì giá trị y dọc theo đường thẳng được đưa ra từ phương trình

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0} \tag{1.10}$$

$$y = y_0 + (y_1 - y_0)\frac{x - x_0}{x_1 - x_0} \tag{1.11}$$

Hay

$$y = y_0\left(1 - \frac{x - x_0}{x_1 - x_0}\right) + y_1\left(\frac{x - x_0}{x_1 - x_0}\right) \tag{1.12}$$

Đây là công thức nội suy tuyến tính trong khoảng (x_0, x_1).

Nội suy tuyến tính là phương pháp đơn giản nhất để nhận được các giá trị tại các vị trí giữa các điểm dữ liệu. Các điểm dữ liệu này được kết nối một cách đơn giản bởi một đoạn thẳng. Mỗi đoạn thẳng (giới hạn bởi hai điểm dữ liệu) có thể được nội suy một cách độc lập. Khi đó hàm nội suy giữa hai điểm có thể được viết như sau:

$$y = y_0(1 - t) + y_1 t \tag{1.13}$$

Tham số *t* xác định nơi để ước tính giá trị trên đường nội suy, nó là *0* tại điểm đầu tiên và là *1* tại điểm thứ hai. Các giá trị nội suy giữa hai điểm *t* nằm trong khoảng giữa *0* và *1*.

Giả sử rằng khuôn mặt 3D được biểu diễn bởi một đa giác thích hợp. Ý tưởng của kỹ thuật này tương tự như phương pháp tạo ra chuyển động của các phim hoạt hình thông thường. Các phim hoạt hình xác định chuyển động mong muốn bất kỳ bằng cách tạo ra một loạt các bản vẽ các chuyển động chính, sau đó người tạo chuyển động sẽ tạo ra một loạt các khung hình (frame) trung gian giữa các chuyển động của các bản vẽ chính. Đối với việc hoạt hóa trạng thái khuôn mặt trên máy tính thì các bản vẽ chính ở đây được thay thế bởi các tập tin dữ liệu mô tả các trạng thái khuôn mặt khác nhau.

Các trạng thái khuôn mặt khác nhau này bao gồm các trạng thái như trạng thái khuôn mặt tự nhiên, trạng thái cười, trạng thái tức giận, ngạc nhiên, v.v.

Các dữ liệu trong các tập tin mô tả trạng thái khuôn mặt là các vị trí ba chiều của các điểm xác định da đa giác sử dụng để đại diện cho khuôn mặt hoặc vị trí ba chiều của các điểm tại các vùng đặt trong cơ bản nhất của khuôn mặt mà ở đó trạng thái khuôn mặt được thể hiện.

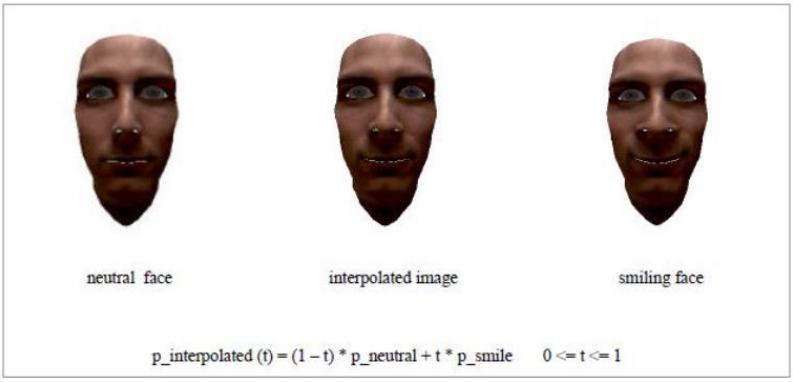
Các khung hình trung gian được xác định bằng cách sử dụng phép nội suy tuyến tính.

$$P_interpolated\ (t) = (1\text{-}t)^* P_prev + t^* P_next \qquad 0 <= t <= 1 \tag{1.14}$$

Trong đó:

P_interpolated: vị trí các điểm trong frame trung gian
P_prev: vị trí các điểm trong frame trạng thái khuôn mặt nguồn.

P_next: vị trí các điểm trong frame trạng thái khuôn mặt đích
t: Tham số nội suy (*0 <= t <= 1)*



Hình 1.19 Nội suy tuyến tính khuôn mặt từ trạng thái tự nhiên sang trạng thái cười

Tuy nhiên, khuôn mặt luôn được chi phối bởi các quy luật vật lý, chuyển động của nó không phải là tuyến tính mà có xu hướng tăng tốc và giảm tốc. Do đó, một hàm nội suy Cosin hoặc các biến thể khác có thể được sử dụng để cung cấp khả năng tăng tốc và giảm tốc độ thể hiện trạng thái khi bắt đầu và kết thúc một hoạt cảnh.

- **Kết hợp giữa kỹ thuật nội suy tuyến tính và kỹ thuật Morphing:**

Morphing trạng thái khuôn mặt 3D nghĩa là tạo ra các chuyển tiếp liên tục và thực tế giữa các biểu cảm khác nhau của khuôn mặt. Phương pháp này đạt được các hiệu ứng bằng cách morphing giữa các mô hình mặt tương ứng. Trình tự morphing 3D có thể đặt được bằng cách sử dụng kỹ thuật nội suy tuyến tính đơn giản giữa các tọa độ hình học của các đỉnh tương ứng trong mỗi lưới của hai lưới khuôn mặt. Cùng với quá trình nội suy hình học thì các texture của mô hình khuôn mặt cũng cần được blend từ các textures có liên quan. Khi morphing hai biểu cảm khác nhau của cùng một mô hình mặt thì mô hình khuôn mặt trung gian được tạo ra bởi phép nội suy hình học. Sau đó, texture của mô hình khuôn mặt trung gian này trực tiếp được tạo ra từ khuôn mặt tự nhiên bằng cách thiết lập tương ứng giữa mô đa giác của mô hình khuôn mặt tự nhiên và mô hình biểu cảm trung gian cho mỗi dòng quét với mỗi pixel. Với morphing giữa hai trạng thái biểu cảm của bất kỳ hai mô hình khuôn mặt nào thì texture trung gian được tạo ra bằng cách sử dụng nội suy tuyến tính của nguồn tương ứng và tam giác đích cho từng dòng quét.

- **Kết hợp giữa kỹ thuật nội suy tuyến tính và hệ mã hóa hành động mặt:**

FACS được giới thiệu bởi Ekman và Friesen năm 1976 để đánh giá và mô tả các hành động mặt bằng cách kiểm tra tất cả các cơ mặt. Hiện nay, nó được coi là một

chuẩn để hiển thị sự xuất hiện khuôn mặt kích thích bởi những thay đổi trên từng cơ mặt. Nó xuất phát từ việc phân tích giải phẫu khuôn mặt bằng cách mô tả các hoạt động cơ mặt của con người. Nguyên tắc làm việc của FACS trong việc tìm hiểu hành vi của khuôn mặt được dựa trên các hành động mặt. FACS chia khuôn mặt con người thành 46 đơn vị hành động (AUs). Mỗi đơn vị hiện thân của một hành động cơ riêng biệt hoặc một nhóm các cơ đặc trong cho một vị trí duy nhất trên khuôn mặt. Các đơn vị hành động mặt được xây dựng sao cho phù hợp với các hành động nơi mà mỗi đơn vị hành động mặt có thể liên quan đến nhiều cơ.

Kỹ thuật nội suy tuyến tính có một số ưu điểm và nhược điểm trong việc hoạt hóa trạng thái khuôn mặt như sau:

Ưu điểm:

- Đơn giản để hiểu và thực thi.
- Tốc độ hoạt hóa nhanh
- Không yêu cầu tính toán chuyên sâu.
- Dễ dàng tạo ra các hoạt ảnh trên khuôn mặt với kết quả được dự đoán cao

Nhược điểm:

- Bị giới hạn bởi một số lượng các biểu cảm chính nhất định.
- Nếu sự chuyển đổi giữa hai key-frame trạng thái khuôn mặt là không tuyến tính cũng không phù hợp với bất kỳ hàm nào đang tồn tại thì sẽ khó khăn trong việc tính toán vị trí của các trạng thái khuôn mặt trung gian.
- Khó khăn trong việc tạo ra sự kết hợp động của các chuyển động mặt độc lập
- Không hiệu quả đối với các nét mặt có sự biến đổi phức tạp

1.3.3.2. Nội suy song tuyến.

Trong toán học, nội suy song tuyến [65] (Bilinear interpolation) là mở rộng của phép nội suy tuyến tính cho các hàm nội suy hai biến (ví dụ: x và y) trong một lưới phẳng 2D. Ý tưởng của phương pháp này là nội suy tuyến tính theo hướng đầu tiên, và sau đó nội suy tuyến tính một lần nữa theo hướng khác. Hay nói cách khác, nội suy song tuyến tính là phương pháp được sử dụng để tính trọng số khoảng cách trung bình của bốn điểm gần nhất để ước lượng giá trị điểm mới được tạo ra.

Ý tưởng của phương pháp nội suy song tuyến là sử dụng bốn khung hình nội suy [49] thay vì sử dụng hai khung hình trong phép nội suy tuyến tính. Các khung hình (hay keyframe) này đạt được bằng cách kết hợp hai dữ liệu hình dạng của khuôn mặt và hai dữ liệu biểu cảm trên khuôn mặt.

Ưu điểm:

- Nhanh, dễ dàng tạo ra các hoạt ảnh trên khuôn mặt
- Do sử dụng bốn khung hình chính có liên quan chứ không phải hai, nên kỹ thuật nội suy song tuyến tạo ra các biểu hiện khuôn mặt đa dạng khác nhau hơn so với nội suy tuyến tính. Nội suy song tuyến, khi kết hợp với mô phỏng morphing hình ảnh sẽ tạo ra một loạt các thay đổi biểu cảm khuôn mặt thực tế.

Nhược điểm:

Cũng giống như các kỹ thuật nội suy khác, kỹ thuật nội suy song tuyến cũng có các nhược điểm như:

- Khó khăn trong việc tạo ra sự kết hợp động của các chuyển động mặt độc lập
- Không hiệu quả đối với các nét mặt có sự biến đổi phức tạp
- Chỉ phù hợp trong các bài toán nhỏ, sử dụng một vài khung hình cơ bản để sinh ra một tập ảnh nhỏ.

1.3.3.3. Nội suy dựa trên hàm cơ sở bán kính

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 [125] [58] [42]. Có rất nhiều phương pháp như NURBS [4], RBF, Affine, nội suy dựa vào mạng neural, v.v... Trong đó, Wang và Ierapetritou [117] và Dias [34] 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.

Giả sử rằng giá trị của một hàm vô hướng là đại diện cho sự biến đổi của n điểm rời rạc khác biệt x_i trong không gian R^3. Khi đó RBF nội suy cho hàm F trong không gian R^3. Hàm này có dạng tổng của n xấp xỉ bởi một hàm cơ sở bán kính $g(r_i)$: *[0,*

rMax] → [0,1] trong đó r_i là khoảng cách giữa các điểm $p = (x, y, z)$ được xấp xỉ bởi các điểm $p_i = (x_i, y_i, z_i)$

Ý tưởng của kỹ thuật này là khuôn mặt 3D sẽ được đặc trong bởi một tập điểm gọi là điểm điều khiển, việc biểu diễn trạng thái khuôn mặt 3D sẽ dựa trên việc điều chỉnh các điểm điều khiển của mô hình khuôn mặt ban đầu thành mô hình khuôn mặt đích (mô hình trạng thái khuôn mặt 3D mà chúng ta muốn biểu diễn. Ví dụ trạng thái vui, buồn, ngạc nhiên, …). Bắt đầu chúng ta sẽ tính các hệ số biến đổi dựa vào sự biến đổi của tập điểm điều khiển từ mô hình khuôn mặt ban đầu tới mô hình khuôn mặt đích, sau đó việc tính toán lại dữ liệu của khuôn mặt sẽ được thực hiện bởi nội suy dựa trên hàm cơ sở bán kính với các giá trị vừa tính được.

***Ưu điểm*:**

- Kỹ thuật nội suy này sử dụng các điểm điều khiển để nội suy ra các trạng thái biểu cảm
- khuôn mặt, do đó chúng ta có thể dễ dàng thiết kế các điểm điều khiển này tại các vùng đặc trong có sự biến đổi rõ rệt giữa các trạng thái biểu cảm sao cho phù hợp với khuôn mặt và phù hợp với các vị trí phức tạp có sự biến đổi trên khuôn mặt.
- Trong kỹ thuật nội suy này việc thêm các điểm điều khiển mới không làm ảnh hưởng nhiều tới thuật toán.
- Kỹ thuật này không làm thay đổi được cấu trúc lưới bề mặt của khuôn mặt

Nhược điểm:

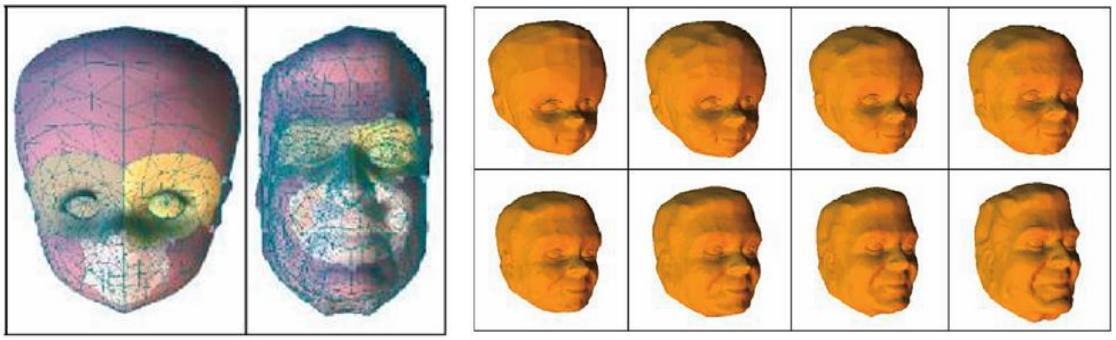
- Khả năng tính toán chậm, khó cài đặt.
- Nếu số điểm điều khiển tăng lên thì độ phức tạp thời gian tính toán của thuật toán cũng tăng lên khá nhanh.

1.3.3.4. Nội suy địa phương

Tiếp cận nội suy địa phương thực hiện dựa trên cơ sở phân chia bề mặt của đối tượng 3D thành nhiều vùng riêng biệt, có thể có ràng buộc giữa các vùng với nhau, mỗi một vùng sẽ được áp dụng thống nhất một phép nội suy, nội suy cục bộ từng vùng và làm mịn vùng biên. Việc phân vùng cho đối tượng cũng là một pha rất quan trọng trong hướng tiếp cận này, phân vùng có chính xác, tối tưu thì nội suy mới hiệu quả. Phân vùng đối tượng thường được thực hiện dựa trên hai phương pháp chính dó

là phân vùng cứng và phân vùng động. Phân vùng cứng tức là chúng ta sẽ phân vùng đối tượng thành các vùng riêng biệt và cố định trên đối tượng, phân vùng động tức là chúng ta sẽ thực hiện tính toán vùng ảnh hưởng tùy ý mỗi khi thực hiện nội suy. Một số nghiên cứu thuộc cách tiếp cận này có thể kể đến như:

Khi nghiên cứu, giải quyết bài toán tạo sự biến đổi cho hoạt hình khuôn mặt, năm 2003 Tong-Yee Lee và Po-Hua Huang [109] đã đề xuất một phương pháp biến đổi mô hình khuôn mặt 3D nguồn thành khuôn mặt 3D đích dựa vào nội suy biến dạng, phương pháp mà hai ông đưa ra là phân chia khuôn mặt thành các vùng cứng, sau đó áp dụng các phép nội suy, biến đổi lên các vùng đó nhằm làm tăng độ chính xác và thời gian tính toán.



a) Đầu vào b) Kết quả đầu ra

Hình 1.20 Nghiên cứu của Tong-Yee Lee và Po-Hua Huang [109]

Vào năm 2005, Yong Joo Kim và cộng sự đã khái niệm "3D Warp Brush" [126]. Trong nghiên cứu của mình, mỗi một lần thực hiện một biến đổi trên mô hình họ sẽ tính toán vùng ảnh hưởng tại một điểm (giả sử là điểm con trỏ chuột) lên các vùng xung quanh và thực hiện biến đổi vùng theo điểm ban đầu.

Một nghiên cứu khác, năm 2014 Henry Schäfer và cộng sự đã nghiên cứu về việc nội suy đối tượng 3D bằng cách thực hiện phân vùng động và tính toán trên GPU, kết quả nghiên cứu cho thấy phương pháp này khác hiệu quả và có tính ứng dụng cao [52]

Hình 1.21 Nghiên cứu của Henry Schäfer và cộng sự [52]

Về cơ bản, các bước thực hiện chính của phương pháp này bao gồm:

Bước 1: Phân vùng bề mặt hình học của đối tượng (Phân vùng được thực hiện một cách tự động hoặc thủ công, tùy thuộc vào đối tượng).

Bước 2: Lựa chọn hàm biến đổi cho các vùng (Việc lựa chọn hàm biến đổi phụ thuộc rất nhiều vào yêu cầu biến đổi đặt ra, thông thường với những biến đổi đơn giản người ta hay chọn các phép biển đổi affine như phương pháp dựa vào Hệ tọa độ Barycentric. Còn với những biến đổi phức tạp thường người ta hay chọn các phép biến đổi bậc cao).

Bước 3: Tính toán dữ liệu trong vùng theo hàm biến đổi và làm trơn vùng biên (Sau khi đã lựa chọn phương pháp biến đổi cho từng vùng, với mỗi một kỹ thuật được lựa chọn thì việc tính toán lại khác nhau tùy thuộc vào từng kỹ thuật).

1.3.3.5. Biến đổi 3D cứng (Rigid)

Các phép biến đổi 3D cứng [87] (cụ thể là phép quay, phép tịnh tiến hoặc kết hợp cả hai) là các trường hợp đặc biệt của phép biến đổi tuyến tính. Cho vectơ $\bar{v}$ đại diện cho một điểm 3D được tăng thêm một điểm, $\bar{v} = [x\ y\ z\ 1]^T$, vectơ $\bar{v}$ biểu thị sự biến đổi tuyến tính của $\bar{v}$ và vectơ $t = \left[t_x\ t_y\ t_z\ \right]^T$ biểu thị phép tịnh tiến (dịch chuyển của các điểm trên bề mặt theo các phương x, y và z). Phép biến đổi tuyến tính được biểu diễn trong phương trình sau:

$$\bar{v}' = A\bar{v} = \begin{bmatrix} M & t \\ O^T & 1 \end{bmatrix} \bar{v} \qquad (1.15)$$

trong đó O là vectơ không, O = [0 0 0]; ma trận A là ma trận vuông (4 × 4) và M là ma trận biến đổi có kích thước 3 × 3. Việc sử dụng ma trận 4 × 4 làm cho nó có thể

biểu diễn một chuỗi các phép biến đổi tuyến tính thông qua phép nhân ma trận, $A_t = A_1 A_2 \dots A_n$, trong đó A_i biểu thị các phép biến đổi cứng liên tiếp để tạo thành một phép biến đổi cứng toàn phần được biểu thị bằng A_t.

Trong trường hợp tổng quát, khi M và t nhận giá trị ngẫu nhiên thì phép biến đổi đó là phép biến đổi affine. Việc áp dụng phép biến đổi affine trên khuôn mặt 3D có thể thay đổi hình dạng của nó (bao gồm tỷ lệ và độ cắt), hướng của nó (do các phép quay) và vị trí của nó (do các phép tịnh tiến). Tuy nhiên, độ song song giữa các đường (do các điểm trên khuôn mặt tạo thành) vẫn được bảo toàn. Trường hợp đặc biệt, khi M là một ma trận trực chuẩn ($M^T = M^{-1}$ hoặc tương đương $I = MM^T$) và thỏa mãn det(M) = 1 thì ta sẽ có một phép biến đổi cứng. Trong trường hợp này, ma trận M được gọi là ma trận quay và được ký hiệu là R, và đối với phép biến đổi cứng như vậy thì chỉ có hướng và vị trí của khuôn mặt có thể thay đổi trong khi hình dạng khuôn mặt được giữ nguyên. Khi ma trận M bằng với ma trận đồng dạng I thì không có phép quay nhưng có phép tịnh tiến theo t.

1.4. 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.

Luận án tập trung vào hai vấn đề sau:

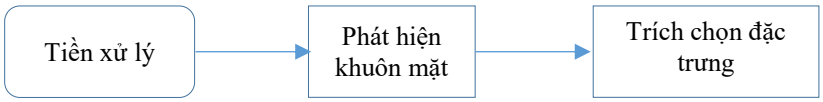
Thứ nhất, nghiên cứu và đề xuất các kỹ thuật trích chọn đặc trưng cử chỉ và ước lượng biểu cảm khuôn mặt. Thuật toán AAM được lựa chọn đẻ nghiên cứu và cải tiến vì phương pháp này dùng mô hình hình dạng và tập trung vào tập điểm điều khiển ảnh hưởng tới biểu cảm của khuôn mặt, như vậy nó phù hợp với bài toán ước lượng biểu cảm mà mục tiêu đặt ra.

Thứ hai, nghiên cứu và đề xuất cải tiến kỹ thuật biểu diễn biểu cảm khuôn mặt bằng phương pháp nội suy RBF. Vì nó có đầu vào là các điểm điều khiển rút trích được từ AAM và đồng thời phương pháp này cũng có nhiều ưu điểm như : chúng ta có thể dễ dàng thiết kế các điểm điều khiển này tại các vùng đặc trưng có sự biến đổi

rõ rệt giữa các trạng thái biểu cảm, việc thêm các điểm điều khiển mới không làm ảnh hưởng nhiều tới thuật toán và kỹ thuật này không làm thay đổi cấu trúc lưới bề mặt của khuôn mặt.

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. Đây là một khâu quan trọng trong một hệ thống phân tích biểu cảm khuôn mặt, trong đó, 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.



Hình 2.1 Quy trình trích chọn đặt trưng cử chỉ khuôn mặt

2.1 Bài toán phát hiện khuôn mặt

Kỹ thuật được đề xuất dựa trên ý tưởng cơ bản là thực hiện phân lớp nhị phân mỗi vùng ảnh, từ đó kết luận vùng ảnh đó là khuôn mặt hay không. Bằng cách sử dụng cây quyết định, tại mỗi nút ta xây dựng một bộ phân lớp nhị phân con có nhiều dạng khác nhau, mục đích là để lựa chọn cách thức phân lớp đơn giản và nhanh chóng. Tại một số nhánh đầu tiên có thể nhanh chóng loại nền, đồng thời cho phép việc nghiên cứu thử nghiệm với những cách thức phân lớp khác nhau tại mỗi nút. Việc kết luận được thực hiện bằng một phép so sánh các bộ phân lớp nhị phân, một khuôn mặt được chấp nhận nếu nó được khớp với tất cả các bộ phân lớp nhị phân trong chuỗi.

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:

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

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; 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:

$$WMSE(I, v, w) = \sum_{(I,v,w)\in C_0} w \cdot (v - \bar{v}_0)^2 + \sum_{(I,v,w)\in C_1} w \cdot (v - \bar{v}_1)^2 \tag{2.2}$$

trong đó: C_0 và C_1 là hai cụm của tập huấn luyện tương ứng là kết quả phân chia của hai giá trị 0 và 1. Các tham số $\bar{v}_0$ và $\bar{v}_1$ là trung bình các giá trị nhãn trong C_0 và C_1.

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. Thuật giải học cây được mô tả như sau:

Thuật giải 2.1: Học khuôn mặt mẫu

Input:

$U = \{(I_s, v_s, w_s): s = 1,2,\dots,S\}$ *// tập dữ liệu huấn luyện*

Output:

$T = \{N_0, N_1...\}$ *// là tập các nút*

Process:

1. $T := \emptyset$
2. *Idx*0 = {0, 1, 2, ..., *S*-1} *// các trọng số tương ứng với các mẫu trong dữ liệu huấn luyện. Sử dụng 1 ngăn xếp để chứa lần lượt từng nút. Mỗi phần tử trong ngăn xếp là một bộ gồm nút và tập trọng số các mẫu dùng để*

học nút đó.

3. *Stack* := Ø *// Đưa nút gốc vào ngăn xếp*

4. **push** (*Stack*, (N_0, Idx_0))

5. **while** (*Stack* ≠ Ø) *// lấy phần tử ở đầu ngăn xếp*

6. { N_i, Idx_i} := **pop**(*Stack*) *// Nếu nút đã đạt đủ độ cao, không thực hiện tính toán chia đôi nữa*

7. **if** (N_i.*level* >= MAX_DEPTH)

8. **Continue**

9. **else**

10. *min_err* := MAX_VAL

11. *best_bincls* := **null**

12. **for all** *bincls* **of** BCS

13. *e* := WMSE(*bincls*, *U*, Idx_i)

14. **if** (*e* < *minerr*)

15. *best_bincls* := *bincls*

16. *min_err* := *e*

17. **endif**

18. **endfor**

19. **setupNode** (N_i, *best_bincls*, *U*, Idx_i) *// Chia đôi tập chỉ số các mẫu của nút hiện tại để làm dữ liệu huấn luyện cho hai nút con*

20. { Idx_{i*2+1}, Idx_{i*2+2}} := **SplitDataSet**(*U*, Idx_i, *best_bincls*) *// Tiếp tục đưa vào ngăn xếp hai nút con cùng với tập chỉ số các mẫu tương ứng*

21. **push**(*Stack*, (N_{i*2+1}, Idx_{i*2+1}))

22. **push**(*Stack*, (N_{i*2+2}, Idx_{i*2+2}))

23. **endif**

24. **endwhile**

Trong đó: Giá trị MAX_DEPTH là tham số giới hạn chiều sâu của cây sẽ học, đây là tham số đặt bằng tay, hiện tại đặt mặc định là 5; MAX_VAL là giá trị khởi tạo cho sai số, dùng để tìm bộ phân lớp con có giá trị sai số nhỏ nhất.

Theo đó, với mỗi kết quả phân lớp trên tập học của nút, ta nhận được giá trị hàm mục tiêu, và kết quả lựa chọn cho nút là bộ tương ứng đạt giá trị nhỏ nhất cho hàm mục tiêu. Như vậy, ta có thể phân tích, thời gian học cây bằng tổng thời gian học từng mức của cây. Thời gian học mỗi mức của cây là tổng thời gian học từng nút và tổng số lượng dữ liệu huấn luyện của các nút tại mỗi mức bằng tập dữ liệu học ban đầu. Thời gian học một nút bằng tổng thời gian học từng bộ phân lớp con có thể. Trong quá trình tối ưu hóa cho một nút trong cây, vì không gian các bộ phân lớp có thể phải xét là khá lớn, bởi vậy, một tập con được sinh ngẫu nhiên đã được sử dụng thay thế. Do đó, các cây được xây dựng kết hợp dựa trên tiếp cận boosting với lựa chọn là GentleBoost.

Trên cơ sở cấu trúc phân lớp được xây dựng như vậy, việc còn lại là thiết kế các bộ phân lớp nhị phân con.

a) Phép so sánh cường độ điểm ảnh

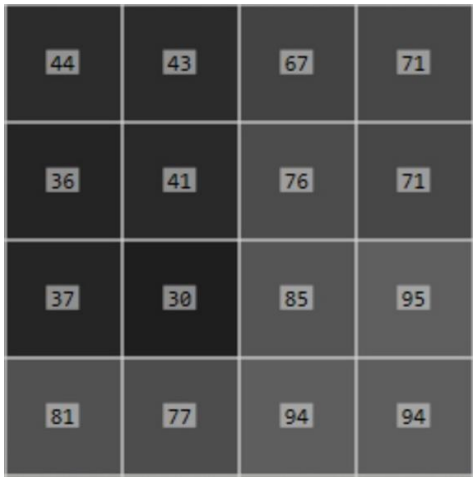
Một phép so sánh cường độ điểm ảnh trên ảnh I được định nghĩa như sau:

$$B(I: l_1, l_2) = \begin{cases} 0, & I(l_1) \le I(l_2) \\ 1, & otherwise \end{cases} \tag{2.3}$$

trong đó: $I(l_i)$ là giá trị cường độ điểm ảnh trong ảnh I tại vị trí l_i. Trong kỹ thuật này, tọa độ l_1 và l_2 được xác định trong phạm vi $[-1,+1]\times[-1,+1]$. Các giá trị tọa độ l_1 và l_2 được xác định như vậy cho phép việc triển khai xác định tập các vị trí không phụ thuộc vào kích thước ảnh mẫu. Do đó, với mỗi vùng ảnh cần xác định, giá trị tọa độ cần được thay đổi theo tỉ lệ phù hợp.

Để hình dung phép so sánh cường độ ảnh, ta xét ví dụ như trong **Error! eference source not found.**; theo đó, ảnh được xét là ảnh xám, tức là các giá trị nằm trong đoạn [0,255]. Cụ thể, trong trường hợp này, ta xét vị trí trên tọa độ ảnh. Một số kết quả so sánh cường độ được chỉ ra trong Bảng 2.1. So sánh cường độ điểm ảnh là một trong những cách phân lớp đơn giản nhất và không cần đến tham số. Việc tính toán nó còn đơn giản hơn đặc trưng Haar, chưa kể đến nó không yêu cầu đến dữ liệu được tính trước như trường hợp ảnh tích phân cho đặc trưng Haar. Về mặt mở rộng,

nó có khả năng tùy biến cho các góc nghiêng khác nhau của đối tượng một cách dễ dàng chỉ với một phép biến đổi tọa độ trong không gian hai chiều đối với 2 vị trí cần so sánh.



Hình 2.2 Ví dụ của phép so sánh cường độ

Để xét duyệt tại mỗi nút, một tập con các phép so sánh được sử dụng. Theo đó, hai vị trí trong phép so sánh được sinh ngẫu nhiên theo phân phối đều trong phạm vi [-1, +1]×[-1, +1]. Nếu cây cần xây dựng có chiều sâu *D*, ta phải thử *B* phép so sánh tại mỗi nút và tập huấn luyện chứa *S* mẫu, khi đó thời gian huấn luyện cây là *O(D·B·S)*.

Bảng 2.1 Một số kết quả ví dụ so sánh cường độ ảnh

l_1	l_2	$I(l_1)$	$I(l_2)$	$B(I:l_1,l_2)$
(0, 0)	(1, 0)	44	43	1
(0, 0)	(2, 0)	44	67	0
(0, 0)	(0, 1)	44	36	1
(0, 0)	(1, 1)	44	41	1
(0, 0)	(2, 1)	44	76	0
(2, 0)	(3, 0)	67	71	0
(2, 1)	(3, 1)	76	71	1
(1, 2)	(2, 3)	30	94	0
(0, 3)	(3, 1)	81	71	1

(1, 1)	(3, 3)	41	94	0

b) Sử dụng giá trị trung bình cục bộ

$$B(I:R(x,y,w,h),\delta) = \begin{cases} 1, & \frac{1}{w*h}\sum_{x\le i<x+w}\sum_{y\le j<y+h} I(i,j) < \delta \\ 0, & otherwise \end{cases} \quad (2.4)$$

trong đó: *I(i,j)* là giá trị cường độ điểm ảnh trong ảnh *I* tại vị trí *(i,j)*. Kết quả phân lớp dựa trên việc so sánh giá trị trung bình các điểm ảnh trong vùng hình chữ nhật *R(x,y,w,h)* với giá trị ngưỡng δ.

Việc tính giá trị trung bình cục bộ vốn cần xét duyệt các giá trị trong vùng, tuy nhiên bằng cách kế thừa sử dụng kỹ thuật ảnh tích phân, việc tính toán có thời gian là *O(1)*. Tiếp đến, việc sử dụng giá trị trung bình cục bộ cần đến một tham số ngưỡng δ. Tham số này được lựa chọn trong quá trình học. Về cơ bản, các vị trí ngưỡng được xét duyệt tương ứng với vị trí sát cạnh của từng giá trị tính được từ các mẫu trên trục số. Tương tự như trên, nếu cây cần xây dựng có chiều sâu *D*, ta phải thử *B* phép so sánh tại mỗi nút và tập huấn luyện chứa *S* mẫu, khi đó thời gian huấn luyện cây là $O(D{\cdot}B{\cdot}S^2)$.

2.1.2 Phát hiện khuôn mặt

Không xét đến các thông tin có thể biết trước nhằm giảm thiểu không gian tìm kiếm, chẳng hạn như vùng ảnh nền, ảnh chuyển động trong ngữ cảnh dùng video, ảnh đầu vào được thực hiện xét duyệt tại mỗi vị trí và kích thước có thể. Như đề xuất đã nêu ở phần trên, một khuôn mặt được chấp nhận nếu nó khớp với mô hình đã xây dựng, trong đó tại mỗi nút là một bộ phân lớp nhị phân con và phương pháp lựa chọn là phép so sánh giá trị trung bình cục bộ của điểm ảnh.

Xét việc tính toán trên từng cây và giả sử rằng cây đang xét có chiều sâu *D*. Khi đó, thời gian phân lớp với cây là *O(D·K)* với *K* là thời gian thực hiện tính toán tại bộ phân lớp nhị phân con. Trong các trường hợp bộ phân lớp nhị phân con được sử dụng là so sánh điểm ảnh và đánh giá giá trị trung bình cục bộ, thời gian tính toán đều là *O(1)*; do đó, thời gian phân lớp với cây là *O(D)*.

2.1.3 Kết quả thử nghiệm

Việc thử nghiệm phương pháp được đề xuất trong Luận án này được tiến hành

theo 2 bước: học và chạy thử nghiệm phát hiện. Pha học được tiến hành với cơ sở dữ liệu khuôn mặt GENKI-SZSL nằm trong bộ MPLab GENKI tại Đại học California, San Diego. Bộ GENKI-SZSL bao gồm 3.500 ảnh khuôn mặt. Việc huấn luyện cũng sử dụng Bộ dữ liệu Face negative gồm 3.019 ảnh được cung cấp trong một hướng dẫn sử dụng bộ Haartraining của OpenCV.

Việc thử nghiệm được tiến hành với cơ sở dữ liệu khuôn mặt của Viện Công nghệ California, bao gồm 450 ảnh khuôn mặt, có độ phân giải 896x592 và định dạng jpeg. Các khuôn mặt trong ảnh được thu thập dưới nhiều điều kiện khác nhau, về ánh sáng, biểu cảm mặt và nền. Phương pháp cũng được so sánh với kết quả phát hiện sử dụng thuật toán Haar Adaboost được cung cấp bởi bộ thư viện OpenCV, cụ thể là sử dụng mẫu haarcascade frontalface alt_tree với các tham số mặc định. Nghiên cứu sinh sử dụng mô hình đã huấn luyện để tìm kiếm và định vị khuôn mặt. Sau đó, sử dụng lớp Haar Feature-based Cascade Classifiers từ thư viện OpenCV để rút trích các đặc trưng Haar của tất cả các ảnh khuôn mặt (không phải ảnh gốc) và lưu vào cơ sở dữ liệu Haar. Sau 5 lần kiểm thử thì tính trung bình cộng để xác định tổng giá trị lỗi.

Về mặt lý thuyết, phương pháp đề xuất và haar adaboost đều dùng tiếp cận boosting. Tuy nhiên, việc xây dựng các bộ phân lớp yếu thì có những tiếp cận khác nhau. Trong haar adaboost, các bộ phân lớp yếu được xây dựng từ các đặc trưng haar; trong phương pháp đề xuất các bộ phân lớp yếu được tổ chức bằng mô hình cây quyết định và có khả năng tùy biến việc thiết kế các bộ phân lớp nhị phân con. Một bộ phân lớp nhị phân con được sử dụng là phép so sánh cường độ điểm ảnh cũng đã đưa ra phân tích so sánh với đặc trưng Haar. Cụ thể: so sánh cường độ điểm ảnh là một trong những cách phân lớp đơn giản nhất và không cần đến tham số. Việc tính toán nó còn đơn giản hơn đặc trưng Haar, chưa kể đến nó không yêu cầu đến dữ liệu được tính trước như trường hợp ảnh tích phân cho đặc trưng Haar. Về mặt mở rộng, nó có khả năng tùy biến cho các góc nghiêng khác nhau của đối tượng một cách dễ dàng chỉ với một phép biến đổi tọa độ trong không gian hai chiều đối với 2 vị trí cần so sánh. Kết quả thử nghiệm được tóm tắt như trong Bảng 2.2.

Bảng 2.2 So sánh kết quả định vị khuôn mặt

	Haar Adaboost	***PP đề xuất***
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, nghiên cứu sinh 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:

Precision = 416 / 423 = 0,9834 (hay đạt 98,34%);

Recall = 416 / 450 = 0,9244 (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

Các cử chỉ trên khuôn mặt (hoạt động của cơ mặt) điều chỉnh tương tác xã hội: chúng đại diện cho các tín hiệu giọng nói có thể nhìn thấy và chúng chỉ ra trọng tâm chú ý hiện tại của chúng ta (ví dụ: một người hoặc những gì đã được nói) là quan trọng, hài hước hay khó chịu đối với chúng ta. Chúng là phương tiện trực tiếp, tự nhiên ưu việt để con người truyền đạt cảm xúc của mình [80]. Do đó, các phân tích tự động về những thay đổi tinh tế trên khuôn mặt dường như có một vị trí quan trọng trong các hệ thống thị giác khác nhau bao gồm các công cụ tự động để nghiên cứu tâm lý, đọc môi, phân tích giọng nói hai phương thức, tính toán cảm xúc, tổng hợp khuôn mặt, giọng nói trực quan và giao diện người dùng tri giác. Do đó, trong những năm gần đây, việc tự động hóa phân tích cử chỉ trên khuôn mặt đã được quan tâm rất nhiều.

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 nhướng mày ngạc nhiên [80]. Để phát hiện sự tinh tế như vậy của cảm xúc con người,

cần phải tự động nhận dạng các cử chỉ trên khuôn mặt (tức là những thay đổi chi tiết trong biểu hiện trên khuôn mặt).

Từ một số phương pháp nhận dạng cử chỉ khuôn mặt dựa trên hoạt động cơ mặt có thể quan sát được bằng mắt thường, hệ thống FACS được sử dụng phổ biến nhất trong nghiên cứu tâm lý. Theo xu hướng này, tất cả các phương pháp hiện có để phân tích cử chỉ khuôn mặt tự động, bao gồm cả phương pháp được đề xuất ở đây, diễn giải thông tin hiển thị trên khuôn mặt dưới dạng đơn vị hành động trên khuôn mặt (AU) của hệ thống FACS [89] [43].

Tuy nhiên, không có hệ thống tự động nào có khả năng mã hóa toàn bộ phạm vi bắt chước khuôn mặt, tức là không có hệ thống nào có khả năng nhận ra tất cả 44 AU tạo ra những thay đổi trong hiển thị trên khuôn mặt. Từ các công trình trước đây về nhận dạng cử chỉ khuôn mặt tự động từ chuỗi hình ảnh khuôn mặt, phương pháp được trình bày trong nghiên cứu của Pentland hoạt động tốt nhất ở khía cạnh này: nó mã hóa 16 AU xảy ra một mình hoặc kết hợp trong chuỗi hình ảnh khuôn mặt xem trực diện.

Tuy nhiên, ngay cả khi một hệ thống nhận dạng các hoạt động cơ mặt có thể xảy ra, việc giải thích cảm xúc của các dấu hiệu trên khuôn mặt sẽ không phải là một nhiệm vụ tầm thường. Mục tiêu này thực hiện khó khăn bởi các sắc thái phong phú của trạng thái tình cảm / suy nghĩ mà con người nhận ra trong một biểu hiện trên khuôn mặt. Một yếu tố khó khăn chính khác là một cử chỉ khuôn mặt thể hiện có thể dễ bị hiểu sai nếu không tính đến sự hiện diện của dữ liệu giọng nói trực quan. Ví dụ, người nói có thể cau mày để nhấn mạnh mức độ khó của vấn đề đang được thảo luận và người nghe có thể biểu thị rằng anh ta không hiểu vấn đề đang được đề cập. Tuy nhiên, cho đến nay, các hệ thống phân tích thông tin khuôn mặt tự động thường không thực hiện việc giải thích dữ liệu cảm nhận do người dùng lập và hầu như tất cả các phương pháp phân tích cử chỉ khuôn mặt tránh được việc giải quyết các câu hỏi liên quan đến việc đối tượng được quan sát có đang nói hay không. Sau đó rất dễ thực hiện nếu người ta có thể giới hạn ngữ cảnh. Ví dụ, nếu bạn biết rằng ngoại trừ đối tượng được quan sát không có người nào khác trong khu vực, thì việc mím môi có thể biểu thị một dấu hiệu trên khuôn mặt đang buồn chán hoặc đang suy nghĩ chứ không phải là tín hiệu giọng nói dễ thấy. Tuy nhiên, khi nghiên cứu hướng tới việc tạo điều kiện tổng quát hơn cho các hội nghị truyền hình, truy cập ảo vào các trang web Internet, v.v., các tác giả sẽ phải đối mặt trực tiếp với vấn đề phân biệt các cử

chỉ khuôn mặt đặc trưng cho việc phát âm giọng nói với những thái độ hoặc ảnh hưởng của tín hiệu. Do đó, một hệ thống nhận dạng về việc đối tượng được quan sát có đang nói hay không và kiến thức về cử chỉ khuôn mặt tạo thành tín hiệu giọng nói đều cần thiết cho một diễn giải cảm xúc của các dấu hiệu trên khuôn mặt.

Trích chọn đặc trưng là một phần quan trọng để nhận dạng đối tượng và các đối tượng có đặc điểm rõ ràng, cố định có thể được phân loại dễ dàng. Do vậy, việc xử lý hình ảnh để cải thiện độ rõ nét của các đối tượng trong hình ảnh là việc làm quan trọng trước khi thực hiện việc trích chọn các đặc trưng mong muốn. Nhìn chung, việc trích chọn đặc trưng cử chỉ của khuôn mặt thường được thực hiện theo quy trình sau:

2.2.1 Loại bỏ nhiễu hình ảnh

Các hình ảnh đều chứa nhiễu ở một mức độ nào đó, làm giảm chất lượng hình ảnh và độ nét của cạnh và các đặc trưng của ảnh. Tuy nhiên, có thể giảm nhiễu xuống đến một giới hạn có thể chấp nhận được trước khi tiến hành phân tích sâu hơn trên ảnh. Có nhiều phương pháp loại bỏ nhiễu, ví dụ: tính trung bình hình ảnh, Bộ lọc trung vị, Bộ lọc sigma, Bộ lọc Wiener, FFT và Biến đổi Wavelet, v.v... Cụ thể như sau:

a) Trung bình hình ảnh và Bộ lọc trung vị

Tính trung bình ảnh là cách đơn giản và nhanh chóng để loại bỏ nhiễu và được sử dụng rộng rãi trong xử lý ảnh kỹ thuật số. Trong thuật toán này, giá trị của mỗi pixel đầu ra được xác định từ việc lấy trung bình giá trị pixel của các vùng lân cận của nó. Còn Bộ lọc trung vị sử dụng một thuật toán tương tự, trong đó giá trị của mỗi pixel đầu ra được xác định từ giá trị pixel trung bình của các vùng lân cận của nó. Tuy nhiên, cả hai kỹ thuật này đều dẫn đến việc làm mịn hình ảnh, điều này có nhược điểm là nó cũng làm mờ hình ảnh và làm giảm thông tin cạnh, như đã mô tả ở trên, rất quan trọng để phát hiện các đặc trưng.

b) Bộ lọc Sigma

Lee [69] đã thiết kế Bộ lọc Sigma để khắc phục các vấn đề của Bộ lọc trung vị và trung bình. Ý tưởng của Bộ lọc Sigma chỉ bao gồm trung bình các giá trị màu xám trong một cửa sổ khác với giá trị màu xám của pixel trung tâm không quá một tham số cố định - được gọi là giá trị "Sigma". Bộ lọc này có ưu điểm là làm mịn hình ảnh, không có bất kỳ hiện tượng nhòe hay suy giảm cạnh đáng kể nào. Nó cũng có một lợi thế nữa là tốc độ xử lý rất nhanh. Vì vậy, Bộ lọc Sigma được coi là thích hợp nhất để

sử dụng trong nghiên cứu này.

c) Bộ lọc Wiener

Một cách tiếp cận cổ điển để lọc nhiễu là Bộ lọc Wiener [120] [21]. Bộ lọc Wiener sử dụng phương pháp thích ứng theo pixel dựa trên thông tin thu thập từ vùng lân cận cục bộ xung quanh mỗi pixel. Nó sử dụng thông tin này để ước tính giá trị trung bình cục bộ và phương sai xung quanh mỗi pixel. Tuy nhiên, một vấn đề với phương pháp này là nó tương đối thành công đối với hình ảnh có nhiều nhiễu, nhưng không hoạt động tốt đối với hình ảnh ít nhiễu. Ngoài ra, nó cũng giới thiệu một số điểm làm mờ và phức tạp về mặt tính toán nên nó không thể được sử dụng trong các ứng dụng thời gian thực.

d) FFT và Biến đổi Wavelet

Phân tích FFT là một công nghệ rất quan trọng trong xử lý tín hiệu và xử lý hình ảnh và hoạt động bằng cách tách một hình ảnh thành các miền tần số không gian khác nhau của nó. Bằng cách tách thành tần số cao và tần số thấp, nhiễu được loại bỏ dễ dàng. Một thuật toán được phát triển bởi Kovesi cung cấp một ví dụ điển hình về việc loại bỏ nhiễu ảnh bằng cách sử dụng Bộ lọc FFT. Thuật toán này đã được sử dụng trong nghiên cứu này để chiết xuất các đặc điểm trên khuôn mặt. Một quá trình tương tự được gọi là biến đổi wavelet [11].

2.2.2 Làm mịn bảo toàn cạnh của khuôn mặt

Như đã thảo luận ở trên, lọc trung bình và lọc trung vị sẽ làm mờ hình ảnh và làm giảm thông tin về cạnh và rất khó để tìm ra giá trị sigma tối ưu để sử dụng trong lọc Sigma. Các nhà khoa học đã cố gắng tạo ra một thuật toán có thể làm mịn hình ảnh mà không bị nhòe hoặc suy giảm cạnh. Cách tiếp cận của họ là xuất ra giá trị của từng pixel riêng lẻ bằng cách sử dụng các phép tính dựa trên các pixel xung quanh. Phương pháp này đã được cải tiến, đầu tiên là bởi [86] và Tomita [108], những người đã đề xuất sử dụng một số mặt nạ hình chữ nhật có chứa một số pixel tập hợp làm cơ sở cho các tính toán. Mức xám trung bình của mặt nạ đồng nhất sẽ được gán cho mỗi pixel làm giá trị đầu ra. Điều này đồng nghĩa với việc hình ảnh mịn, giảm nhiễu nhưng hình ảnh không bị nhòe hay viền giảm chất lượng đáng kể. Thuật toán này đã được sử dụng trong nghiên cứu này để hỗ trợ việc tạo ra các đường biểu diễn để thể hiện các đặc điểm trên khuôn mặt. Nhìn chung, các thuật toán khác nhau được mô tả ở trên có các ứng dụng khác nhau và mỗi thuật toán đều có ưu điểm và nhược điểm.

Đối với hình ảnh kích thước lớn, lọc trung bình và lọc trung vị là một kỹ thuật hiệu quả. FFT và biến đổi wavelet là một thuật toán tốt để tăng cường các tính năng cạnh đặc biệt, trong khi Bộ lọc Sigma phổ biến trong xử lý ảnh.

2.2.3 Hiệu chỉnh ánh sáng và cải thiện đặc trưng khuôn mặt

Như đã đề cập ở trên, sự khác biệt về màu da và điều kiện ánh sáng có thể khiến việc phân tích biểu hiện trên khuôn mặt trở nên khó khăn hơn. Những người da sẫm màu trông sáng hơn trong môi trường ánh sáng rực rỡ, trong khi những người da sáng hơn trông tối hơn trong môi trường ánh sáng kém. Những vấn đề này không thể được khắc phục bằng cách chỉ sử dụng các kỹ thuật liên quan đến phân tích không gian màu RGB hoặc HSV. Vì vậy, [121] để làm cho các đặc trưng rõ ràng hơn để trích chọn bất kể điều kiện ánh sáng và màu da, người ta đã quyết định rằng nên sử dụng hình ảnh mức xám thay vì hình ảnh màu. Là một phần của nghiên cứu này, các thuật toán nâng cao tính năng hiện có đã được xem xét và được tóm tắt trong các đoạn sau.

a) Điều chỉnh gamma hình ảnh

Kỹ thuật này nâng cao hình ảnh bằng cách điều chỉnh giá trị gamma theo cách đơn giản - kỹ thuật tương tự có sẵn trong nhiều phần mềm xử lý hình ảnh hiện nay. Việc điều chỉnh gamma thay đổi mối quan hệ giữa màu đen và trắng. Giá trị mặc định cho gamma là 1, có nghĩa là tuyến tính giữa màu đen và trắng. Nếu gamma nhỏ hơn 1, ánh xạ có trọng số đối với giá trị đầu ra cao hơn (màu trắng) và nếu gamma lớn hơn 1, ánh xạ có trọng số đối với giá trị đầu ra thấp hơn [69]. Điều này có nghĩa là điều chỉnh gamma là một cách hiệu quả để điều chỉnh độ sáng. Cả Martinkauppi [81] và Sterring [103] đều chỉ ra rằng hiệu chỉnh gamma có thể được áp dụng để điều chỉnh biểu đồ hình ảnh giúp cải thiện các đặc điểm trên khuôn mặt. Tuy nhiên, nhược điểm đáng kể của quá trình này là việc điều chỉnh gamma cần thiết để tối ưu hóa hình ảnh giữa các ảnh khác nhau là khác nhau. Do vậy, nghiên cứu này không sử dụng phương pháp điều chỉnh Gamma hình ảnh.

b) Biểu đồ cân bằng

Biểu đồ cân bằng là thuật ngữ mô tả quá trình mà một thuật toán tính trung bình biểu đồ của một hình ảnh. Thuật toán này có thể điều chỉnh biểu đồ theo nhiều cách khác nhau. Ví dụ, cũng như tính trung bình, độ tương phản của hình ảnh có thể được tăng cường bằng cách kéo giãn biểu đồ hoặc nó có thể được điều chỉnh để gần như khớp với một biểu đồ cụ thể.

Biểu đồ cung cấp thông tin về độ phơi sáng của hình ảnh. Nếu biểu đồ chỉ ra rằng có một số lượng lớn các điểm ảnh tối thì hình ảnh có thể bị thiếu sáng. Nếu có quá nhiều các điểm ảnh sáng thì có lẽ nó đã bị phơi sáng quá mức. Trong thuật toán nhận diện khuôn mặt của Rowley [96], một thuật toán hiệu chỉnh ánh sáng được áp dụng, sau đó cân bằng biểu đồ được áp dụng để nâng cao các đặc điểm khuôn mặt.

c) Tăng cường bằng phép toán logic Morph

Phép toán logic Morph được thiết kế để phân biệt các đối tượng ở tiền cảnh với hậu cảnh, bằng cách đo lường sự khác biệt về độ phơi sáng và độ tương phản trên hình ảnh. Thuật toán này xem xét giá trị pixel hiện tại và khu vực xung quanh của nó (được gọi là phần tử cấu trúc). Tương tự như kỹ thuật điều chỉnh gamma, kỹ thuật này chỉ phù hợp với từng hình ảnh riêng lẻ.

d) Cân bằng biểu đồ thích ứng có giới hạn tương phản (CLAHE)

Thay vì nâng cao các tính năng theo biểu đồ của toàn bộ hình ảnh, Pizer [91] đã giới thiệu Contrast-limited adaptive histogram equalization (CLAHE) nhằm nâng cao các đặc trưng trong hình ảnh dựa trên các vùng nhỏ hơn. Điều này đã được cải thiện hơn bởi Zuiderveld [135]. CLAHE là một thuật toán tương tự với thuật toán cân bằng biểu đồ, nhưng nó cung cấp các tính năng nâng cao tốt hơn. Ý tưởng cơ bản là chia hình ảnh gốc thành nhiều ô hoặc nhiều vùng. Thay vì tính toán biểu đồ chung của toàn bộ hình ảnh, CLAHE chỉ tính toán biểu đồ cho từng khu vực. Nó cho độ tương phản tốt hơn và kết quả chính xác hơn bằng cách nâng cao độ tương phản của ảnh là ước tính nền ảnh và trừ ảnh khỏi ảnh nền.

2.2.4 Trích chọn đặc trưng cử chỉ của khuôn mặt

Một số nghiên cứu đã được hoàn thành về nhận dạng nét mặt dựa trên video để kiểm tra hiệu suất của các kỹ thuật khác nhau được sử dụng trong bối cảnh này cho các cơ sở dữ liệu khác nhau. Chew [26] trình bày hệ thống FER dựa trên phương pháp CLM (mô hình cục bộ bị ràng buộc) và trích xuất tính năng LBP bằng cách sử dụng bộ dữ liệu GEMEP-FERA. Họ kết luận rằng LBP có độ chính xác thấp hơn khi điều kiện chiếu sáng được cố định. Bên cạnh đó, Baltrusaitis [15] đề xuất một hệ thống phát hiện các đặc trưng cử chỉ trên khuôn mặt. Mục đích thứ hai nhằm xác định tỷ lệ phân loại trung bình cho các trạng thái cảm xúc khác nhau bằng cách sử dụng điểm, bộ lọc Gabor và bộ phân loại SVM (Support Vectors Machine). Ngoài ra, việc sử dụng thuật toán LBP để trích xuất đặc trưng và SVM cho bước phân loại là khá khó

khăn. Điều này cũng đã được khẳng định bởi Ghimire [44], trong đó dùng phương pháp nhận dạng nét mặt bằng thuật toán Adaboost, các nét mặt được nhận dạng bằng ELM sau đó hỗ trợ máy vectơ SVM phân loại.

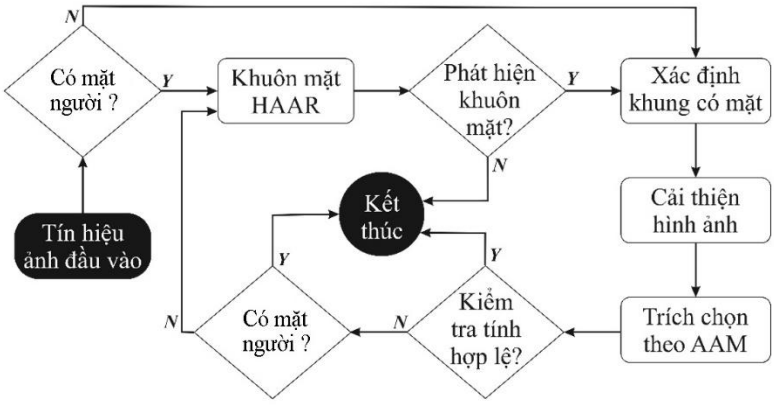
Valstar [112] đề xuất sự kết hợp của hai loại mô tả để giải quyết những thách thức của nhận dạng cảm xúc khuôn mặt: Mô hình nhị phân Gabor cục bộ (LGBP) và Mô hình ngoại hình chủ động (AAM) tự động phát hiện chuyển động trên khuôn mặt. Theo thử nghiệm của họ, sự kết hợp của biểu đồ AAM và LGBP làm tăng độ chính xác của phân loại. Các nghiên cứu khác đã tập trung vào các vấn đề tương tự [122]. Chính vì vậy mà trong nghiên cứu này, nghiên cứu sinh kế thừa Mô hình AAM cải tiến này và bổ sung thuật toán định vị các điểm điều khiển. Cụ thể như sau:

AAM là một thuật toán trong lĩnh vực thị giác máy nhằm mục tiêu tối ưu một mô hình thống kê hình ảnh thể hiện của đối tượng vào một ảnh đầu vào mới, kết quả của quá trình tối ưu là một bộ điểm điều khiển thể hiện cấu trúc của đối tượng đã được học với các tọa độ tương ứng với thể hiện trong ảnh đầu vào của đối tượng cùng với nó là một bộ các tham số mô hình thống kê đã được ước lượng mà 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 thử nghiệm.

Mô hình thuật toán được đề xuất bởi Cootes [59]. Cách tiếp cận này trở thành một chủ đề rất được thu hút đối với các nhà nghiên cứu thị giác máy cũng như các nhà phát triển phần mềm bởi vì nền tảng lý thuyết chắc chắn của thuật toán, sự mềm dẻo và tính linh động trong việc thiết kế và chỉnh sửa thuật toán theo những yêu cầu cụ thể cũng như vì khả năng ứng dụng thực tế đầy hứa hẹn trong những hệ thống phần mềm rất thiết có thể hình dung được rõ ràng trong thực tế cuộc sống. Và thực sự, hướng tiếp cận này đã được sử dụng rộng rãi trong việc lĩnh vực xử lý ảnh y tế và trong các hướng nghiên cứu về so khớp và giám sát khuôn mặt.

Trong AAM, đối tượng quan tâm được mô hình hóa bởi một tập điểm mô tả hình dạng và kết cấu hình ảnh của nó, kết cấu hình ảnh được lấy chính là mẫu của những giá trị cường độ ảnh trong những vùng được giới hạn bởi tập điểm điều khiển. Mô hình thống kê của đối tượng đảm bảo có thể mô tả được những biến thể về hình dạng và những biến thể về kết cấu hình ảnh cũng như mối tương quan giữa chúng. Các vấn đề nổi bật được quan tâm trong phương pháp này là việc xây dựng mô hình thống kê cho đối tượng ảnh và việc thiết kế thuật toán tối ưu để thực hiện tìm kiếm.

Ở đây, việc xây dựng mô hình thống kê toán học cho đối tượng bao gồm việc xây dựng mô hình thống kê toán học cho hình dạng, mô hình mô hình thống kê toán học cho kết cấu hình ảnh và việc kết hợp hai mô hình đó để ra được mô hình cho đối tượng. Module trích rút đặc trưng được mô tả trong tài liệu này bao gồm hai bước là phát hiện khuôn mặt và định vị tập điểm điều khiển. Mô hình chi tiết về module được thể hiện:



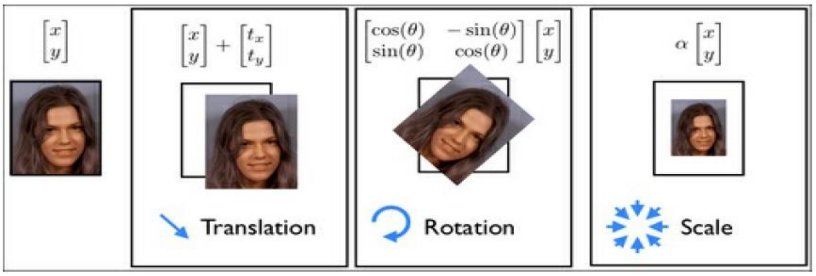
Hình 2.3 Sơ đồ module hệ thống

Module được thiết kế trên cơ bản dựa vào tiếp cận giám sát. Việc sử dụng hay không thao tác phát hiện khuôn mặt là dựa trên cờ trạng thái có thực hiện phát hiện khuôn mặt hay không. Cờ trạng thái này khi khởi tạo hệ thống được đặt là true. Nếu phát hiện được khuôn mặt hợp lệ cờ trạng thái sẽ được đặt ra là false; cờ trạng thái này sẽ được đặt lại là true nếu ở bước sau kiểm tra hình dạng kết quả không hợp lệ. Ở các lượt sử dụng tiếp theo vị trí khuôn mặt sẽ được lấy trực tiếp từ vị trí cũ kết hợp với một vài thao tác ước lượng đơn giản dựa trên vận tốc chuyển động của khuôn mặt trên khung hình. Sau khi đã xác định được vị trí của khuôn mặt trong khung hình, hệ thống sẽ tiếp tục tính toán vùng ảnh sẽ sử dụng để giới hạn dữ liệu tìm kiếm tập điểm điều khiển thể hiện hình dạng của khuôn mặt. Vùng ảnh này sẽ được chuẩn hóa và tăng cường chất lượng cho kết cấu hình ảnh của vùng ảnh. Đây chính là bước chuẩn bị dữ liệu thuật toán AAM được sử dụng để định vị đối tượng hình dạng của khuôn mặt trong khung hình. Sau bước AAM, hệ thống tiến hành kiểm tra lại đối tượng hình dạng đã tính được để một bước nữa kiểm tra sự hợp lệ. Nếu thỏa mãn, đối tượng hình dạng được trả ra làm kết quả module, nếu không thỏa mãn, hệ thống kiểm tra tiếp xem vị trí khuôn mặt được dùng là kết quả từ tác vụ giám sát khuôn mặt; nếu trong tình huống thứ hai, hệ thống sẽ quay lại bước phát hiện khuôn mặt, nếu không thì

thông báo không trích rút được đặc trưng hình dạng khuôn mặt.

a. Mô hình hình dạng

Mô hình thống kê toán học của hình dạng ở đây được dùng để mô tả hình dạng của đối tượng trong ảnh. Mô hình được biểu diễn bởi một tập hợp có thứ tự các điểm điều khiển, thông thường là các điểm 2 chiều hoặc 3 chiều. Mô hình hình dạng mà các tác giả xây dựng cần thiết để có thể thể hiện được các biến thể của đối tượng ảnh về mặt trong những điều kiện thu nhận khác nhau, nói cách khác, mô hình này cần có đặc điểm là bất biến với các phép biến đổi hình học. Trong công trình đã đề xuất, nghiên cứu sinh đã sử dụng phép biến đổi tương đương bao gồm phép dịch chuyển, phép quay và phép thay đổi tỉ lệ kích thước.



Hình 2.4 Phép biến đổi Đồng dạng

Mục tiêu chính của việc xây dựng mô hình toán học được đề cập ở trên chính là sẽ cho phép hệ thống có khả năng phân tích một mẫu hình dạng mới theo để tính toán được các tham số cụ thể tương ứng với mô hình đã xây dựng đồng thời từ mô hình đã được xây dựng có thể tổng hợp ra được một thể hiện hình dạng tương tự với mẫu hình dạng đầu vào. Phương pháp được xây dựng trên lý thuyết là tổng quát cho cả những mô hình trên không gian n chiều, còn trong thực nghiệm của các tác giả cũng như của những người nghiên cứu tiếp nối, phương pháp thường được hiện thực với dữ liệu là những mô hình 2 chiều cùng với phép biến đổi Similarity, có thể coi đây là một trong những thể hiện đơn giản và dễ triển khai nhất khi viết chương trình.

Một vấn đề quan trọng trong xây dựng mô hình chính là việc cần thiết phải lựa chọn được ra những điểm điều khiển phù hợp với loại đối tượng cần được mô hình hóa. Thông thường việc lựa chọn tập điểm phù hợp là dựa vào tri thức của những chuyên gia, là những người đã có nghiên cứu nhất định về bản chất của đối tượng ta quan tâm, ví dụ như khi nghiên cứu về ảnh chụp cắt lớp trong y tế, việc xây dựng những bộ điểm thể hiện những đối tượng đặc thù như những đoạn xương, mấu khớp,

não, tim… cần đến những người am hiểu về giải phẫu học và những bác sĩ chuyên khoa, những người biết được đâu là những vị trí đặc thù cần phải được đặc biệt lưu ý. Khi xây dựng tập dữ liệu huấn luyện, tập dữ liệu bao gồm một tập ảnh của đối tượng cùng với thông tin đánh dấu tập điểm điều khiển mô tả thể hiện của đối tượng được quan tâm tương ứng trong các ảnh, các chuyên gia sẽ sử dụng một công cụ tin học để thực hiện đánh dấu trên một chuỗi ảnh chính là các ảnh cần để xây dựng cơ sở dữ liệu. Trong thực tế, ngoài cách làm thủ công hoàn toàn này, còn có những hướng tiếp cận khác, có thể tự động hoặc bán tự động, chẳng hạn như khi ta đã có một bộ mô hình đã huấn luyện tương đối tốt, ta sẽ sử dụng nó để phát hiện các điểm điều khiển trong ảnh mới, có thể kèm lựa chọn là một chuyên gia sẽ tiến hành kiểm định kết quả và thực hiện những hiệu chỉnh cho những kết quả không đủ chính xác nếu cần thiết.

Nếu đối tượng hình học, chính là tập hợp có đánh thứ tự của các điểm điều khiển mô tả hình dạng của đối tượng ảnh được quan tâm, được biểu diễn bởi n điểm trong không gian d chiều, ta thể hiện nó bằng một vector nd chiều với các giá trị trong vector chính giá trị tọa độ cụ thể trên từng trục không gian của từng điểm. Ví dụ, trong ảnh 2 chiều, ta có thể mô tả n điểm điều khiển, $\{(x_i, y_i)\}$, thành vector 2n thành phần như sau:

$$\mathrm{x} = (x_1, \ldots, x_n, y_1, \ldots, y_n)^T \tag{2.5}$$

Sau khi thể hiện lại các đối tượng hình dạng thành các vector tương ứng, ta thực hiện mô hình hóa trên dữ liệu là tập các vector này. Bước đầu cần chuẩn hóa các đối tượng hình dạng này vào chung một không gian tọa độ, các tác giả sử dụng phương pháp Procrustes Analysis. Phương pháp này thực hiện chuẩn hóa các đối tượng hình dạng để tổng các khoảng cách của mỗi đối tượng hình dạng tới kỳ vọng của chúng ($D = \sum|x_i - \bar{x}|^2$) đạt cực tiểu. Quá trình thực hiện được thực hiện theo từng bước như sau:

Bước 1: Dịch chuyển mỗi mẫu về tọa độ tâm

Bước 2: Lấy kỳ vọng hiện tại làm ước lượng khởi đầu và thay đổi tỉ lệ để để nó có độ dài bằng 1

Bước 3: Ghi nhận ước lượng đầu tiên là $\bar{x}_0$.

Bước 4: Thực hiện chuẩn hóa tất cả các mẫu về ước lượng hiện tại của kỳ vọng

Bước 5: Tính lại kỳ vọng hiện tại

Bước 6: Thực thi những ràng buộc trên ước lượng hiện tại của kỳ vọng bằng cách chuẩn hóa nó theo $\bar{x}_0$ và thay đổi tỉ lệ để để nó có độ dài bằng 1

Bước 7: Nếu ước lượng hiện tại không thay đổi so với trạng thái trước, thuật toán kết thúc. Nếu không quay lại bước 4

Trong các bước trên có đề cập đến việc chuẩn hóa một đối tượng hình dạng theo một đối tượng hình dạng khác, cụ thể trong trường hợp này, quá trình đó được tính như sau: cho hai đối tượng hình dạng x và x', mỗi đối tượng đều chuẩn hóa về tọa độ tâm, cần thực hiện việc chọn một tỉ lệ co giãn s và một góc quay θ để cực tiểu tổng khoảng cách giữa các điểm của x sau khi biến đổi với tương ứng các điểm trong x', cụ thể trong trường hợp này là dùng phép biến đổi Similarity, cần cực tiểu giá trị biểu thức $|sAx - x'|$. Phép biến đổi có dạng như sau:

$$T\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}a & -b\\b & a\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} + \begin{pmatrix}t_x\\t_y\end{pmatrix} \tag{2.6}$$

$$a = (x.x')/|x|^2$$

$$b = \left(\sum_{i=1}^{n}(x_i y_i' - y_i x_i')\right)/|x|^2 \tag{2.7}$$

Trong đó $s^2 = a^2 + b^2$ và $\theta = tan^{-1}(b/a)$

Sau bước chuẩn hóa, thực hiện tiếp quá trình mô hình hóa các biến dạng của hình dạng. Phương pháp được lựa chọn là phân tích thành phần chính PCA, đây là một công cụ toán học tuyến tính rất mạnh cho phép thực hiện việc chiếu một mẫu và khôi phục mẫu từ hình chiếu. Giả sử ta có một tập các mẫu hình dạng đã chuẩn hóa $\{x_i\}$, i=1..s, các bước thực hiện:

Tính kỳ vọng của dữ liệu

$$\bar{x} = \frac{1}{s}\sum_{i=1}^{s} x_i \tag{2.7}$$

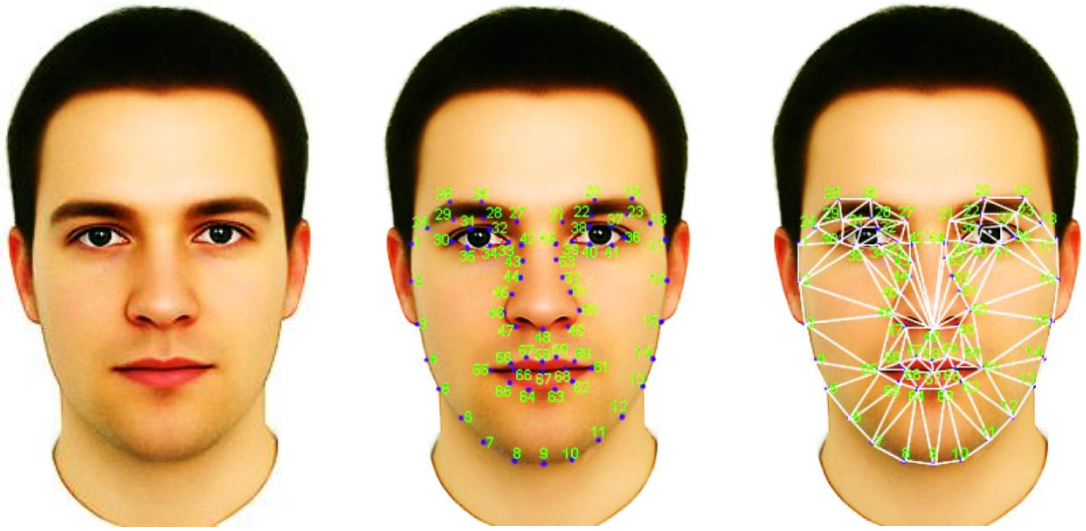
Tính tổng các khoảng cách của mỗi đối tượng

$$S = \frac{1}{s-1}\sum_{i=1}^{s}(x_i - \bar{x})(x_i - \bar{x})^T \quad (2.8)$$

Tính các vector riêng và giá trị riêng tương ứng, thực hiện sắp xếp theo thứ tự giảm dần của các giá trị riêng.

b. Mô hình kết cấu hình ảnh

Sau khi đã mô hình hóa được các đối tượng hình dạng, ta thực hiện tiếp quá trình mô hình hóa kết cấu hình ảnh của các đối tượng được quan tâm trong ảnh. Ở đây, kết cấu hình ảnh của đối tượng được hiểu là các giá trị cường độ ảnh được giới hạn trong vùng ảnh được bao bởi đối tượng hình dạng tương ứng, xem **Error! eference source not found.**.



Hình 2.5 Mặt nạ kết cấu hình dạng

Lưới tam giác về mặt bản chất là 1 loại mối quan hệ giữa các điểm điều khiển. Về mặt ý nghĩa, 1 tam giác sẽ phản ánh 1 vùng texture đủ nhỏ trên khuôn mặt. Tập tam giác được dựng thủ công trên một đối tượng hình học tiêu chuẩn đã được chọn trước cụ thể là theo phép biến đổi affine. Phép biến đổi này được gọi là Piece-wise Affine.

Quá trình lấy ra được dữ liệu kết cấu hình ảnh này được các tác giả thực hiện qua hai bước. Đầu tiên là thực hiện quá trình tam giác hóa trên tập điểm điều khiển để lấy ra được một tập các vùng ảnh và bước tiếp theo là thực hiện quá trình biến đổi hình học từng phần trên ảnh tương ứng với từng tam giác đã được tính toán trước đó.

Mỗi mẫu kết cấu hình ảnh lấy được đều có chứa ít nhiều những biến dạng gây ra bởi sự chiếu sáng khi thu nhận ảnh. Để cực tiểu những biến đổi gây ra bởi sự chiếu sáng toàn cục, các tác giả thực hiện chuẩn hóa các mẫu kết cấu hình ảnh bằng cách áp dụng hai tham số sau: tham số biến đổi tỉ lệ α và tham số dịch chuyển β theo công thức sau:

$$g = (g_{im} - \beta 1)/\alpha \tag{2.9}$$

Các giá trị của α và β được lựa chọn tối ưu để khớp các vector với kỳ vọng đã được chuẩn hóa. Giả sử $\bar{g}$ là kỳ vọng của các dữ liệu đã được chuẩn hóa về tổng bằng 0 và phương sai bằng 1, các giá trị α và β được chọn để chuẩn hóa g_{im} được tính toán như sau:

$$\alpha = g_{im}.\bar{g} \qquad \beta = (g_{im}.1)/n \tag{2.10}$$

trong đó n là độ dài của vector kết cấu hình ảnh.

Quá trình tính toán ra được kỳ vọng chuẩn hóa là một quá trình lặp, tương tự như việc chuẩn hóa các đối tượng hình dạng đã nêu ở trên, qua mỗi bước thì giá trị của kỳ vọng được tính lại và thuật toán dừng khi giá trị của kỳ vọng không thay đổi. Tương tự sau đó các tác giả thực hiện mô hình hóa các dữ liệu kết cấu hình ảnh đã được chuẩn hóa bằng phương pháp phân tích thành phần chính PCA. Bằng cách này, các mẫu kết cấu hình ảnh có thể được biểu diễn bởi mô hình tuyến tính như sau:

$$g = \bar{g} + P_g b_g \tag{2.11}$$

trong đó $\bar{g}$ là vector kỳ vọng đã được chuẩn hóa của các mẫu kết cấu hình ảnh, P_g là cơ sở không gian con được xây dựng và b_g là hình chiếu tương ứng của mẫu kết cấu được biểu diễn.

Một cách tổng quát, kết cấu hình ảnh của đối tượng trong ảnh có thể được tái tạo bởi các tham số mô hình b_g cái chính là hình chiếu của mẫu kết cấu, cùng các tham số chuẩn hóa α và β như sau:

$$g_{im} = T_u(\bar{g} + P_g b_g) = (1 + u_1)(\bar{g} + P_g b_g) + u_2 1 \tag{2.12}$$

trong đó $u = (\alpha - 1, \beta)^T$

c. Mô hình kết hợp

Dữ liệu đối tượng hình dạng và kết cấu hình ảnh của đối tượng được quan tâm trong các mẫu có thể được biểu diễn tóm tắt qua hai tham số b_s và b_g. Vấn đề tiếp theo của nhóm tác giả là thực hiện mô hình hóa tổng quát cho các dữ liệu này và tương tự phương pháp được lựa chọn là phân tích thành phần chính PCA. Do có thể tồn tại những mối tương quan giữa những biến thể của hình dạng và kết cấu hình ảnh của đối tượng được quan tâm, các mẫu được dùng làm đầu vào cho quá trình học PCA được biểu diễn như sau:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix} \tag{2.13}$$

trong đó W_s là ma trận đường chéo thể hiện trọng số của mỗi tham số hình dạng tương ứng.

Vấn đề lựa chọn các tham số này được hiểu như sau. Các giá trị thành phần trong b_s và b_g thể hiện những thông tin rất khác nhau, cụ thể là b_s mô tả các thông tin cho hình dạng đối tượng chính là tọa độ các điểm điều khiển và b_g thể hiện thông tin về kết cấu hình ảnh đối tượng mà chính là cường độ của vùng ảnh của đối tượng. Trong trường hợp này, việc tính toán được thực hiện khá đơn giản: $W_s = rI$ trong đó r^2 là tỉ số giữa tổng các giá trị riêng của tập kết cấu hình ảnh và tổng các giá trị riêng của tập đối tượng hình dạng. Nói chung, các kết quả của việc tổng hợp mẫu và kết quả của cac giải thuật tìm kiếm dựa trên mô hình này sẽ tương đối nhạy cảm với với việc lựa chọn W_s.

Thực hiện tính toán PCA, ta có được mô hình sau: $\mathbf{b} = \mathbf{P}_c\mathbf{c}$

trong đó $\mathbf{P}_c$ là các vector riêng và $\mathbf{c}$ là vector tham số biểu hiện hình ảnh, cái sẽ thể hiện cả hai thông tin về hình dạng và kết cấu hình ảnh. Chú ý rằng do tính chất tuyến tính của công thức sẽ cho phép ta biểu diễn trực tiếp từ $\mathbf{c}$:

$$x = \bar{x} + P_s W_s^{-1} P_c c \qquad g = \bar{g} + P_g P_{cg} c \tag{2.14}$$

trong đó $P_c = \begin{pmatrix} P_{cs} \\ P_{cg} \end{pmatrix}$

Hoặc ta có thể biểu diễn ngắn gọn $\begin{matrix} x = \bar{x} + Q_s c \\ g = \bar{g} + Q_g c \end{matrix}$ (2.15)

trong đó

$$Q_s = P_s W_s^{-1} P_{cs}$$

$$Q_g = P_g P_{cg}$$

Như vậy, bằng quá trình trên, một ảnh mẫu có thể được tổng hợp bởi vector đặc trưng **c** bằng quá trình sau: từ **c** sinh ra mẫu đối tượng hình dạng và mẫu kết cấu hình ảnh, thực hiện quá trình biến đổi hình học để dán các dữ liệu về cường độ ảnh trong mẫu kết cấu hình ảnh lên vùng ảnh được tạo bởi đối tượng hình dạng.

d. Vấn đề tìm kiếm tối ưu

Trong mô hình AAM mà nghiên cứu sinh đề xuất, thuật toán tìm kiếm được thiết kế để tự động ước lượng các tham số mô hình và tổng hợp ra được ảnh mẫu gần nhất có thể với ảnh mục tiêu đầu vào. Về mặt lý thuyết, quá trình tối ưu nhằm mục đích cực tiểu sự sai lệch giữa ảnh mẫu đầu vào và ảnh tổng hợp được theo mô hình đã được đề cập ở trên. Vector độ lệch được định nghĩa như sau:

$$\delta I = I_i - I_m \tag{2.16}$$

trong đó I_i là vector biểu hiện cường độ ảnh của đối tượng trong ảnh đầu vào, I_m là vector biểu hiện cường độ ảnh của ảnh tổng hợp được từ mô hình. Để xác định được tham số tối ưu cho mô hình, ta cần cực tiểu độ lớn của vector độ lệch $\Delta = |\delta I|^2$

Như đã nói ở trên, vector tham số mô hình **c** được sử dụng để tái tạo lại đối tượng hình dạng và kết cấu hình ảnh theo công thức sau:

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \tag{2.17}$$

Dựa trên những cơ sở đó, các tác giả xây dựng thuật toán dựa trên độ đo $E(p) = r^T r$ trong đó $r(p) = g_s - g_m$ là độ lệch giữa kết cấu hình ảnh sinh ra từ mô hình và kết cấu hình ảnh lấy từ ảnh đầu vào. Tham số P là tham số cần được ước lượng, $P^T = (c^T | t^T | u^T)$, trong đó **c** là tham số của mô hình kết hợp, **t** là tham số dịch chuyển biểu thị vị trí của tập điểm điều khiển trong khung hình, **t** = $(s_x, s_y, t_x, t_y)^T$, **u** là tham số biến đổi kết cấu hình ảnh, cụ thể $g_{im} = T_u(g) = (u_1 + 1)g_{im} + u_2 1$.

Thực hiện khai triển Taylor một lần cho biểu thức độ lêch ta được

$$r(p + \delta p) = r(p) + \frac{\partial r}{\partial p}\delta p \tag{2.18}$$

Bằng việc tối thiểu hóa đại lượng |r(p + δp)|^2, ta nhận được

$$\delta p = Rr(p) = -\left(\left(\frac{\partial r^T}{\partial p}\frac{\partial r}{\partial p}\right)^{-1}\frac{\partial r^T}{\partial p}\right)r(p) \text{ theo [99]} \quad (2.19)$$

Ma trận Jacobi $\frac{\partial r}{\partial p}$ được ước lượng trước từ một tập mẫu. Các thành phần của ma trận được ước lượng bằng cách tính toán một lượng lớn các độ lệch hàm số tương ứng với các độ lệch đối số được truyền vào. Các độ lệch đối số được lấy ngẫu nhiên trong một khoảng cho trước tương ứng.

Có thể hình dung tóm tắt thuật toán như sau:

Tính trước:

- Tính ma trận R
- Khởi tạo giá trị ban đầu cho χ
- Khởi tạo mảng tham số $K = \{1, 0.5, 0.25, 0.125, 0.0625\}$

Các bước lặp:

- Tính vector độ lệch r
- Tính $E = r^T r$
- Tính $\partial p = Rr(p)$
- Với mỗi $k \in K$
- Cập nhật $p' = p + k * \square p$
- Tính r'
- Tính $E' = r'^T r'$
- Nếu $E' < E$, cập nhật $p = p'$, cập nhật độ thay đổi giá trị lỗi hiện tại và sang bước lặp tiếp, nếu không, tiếp tục thử với các k khác.

2.3 Ước lượng biểu cảm khuôn mặt

Việc trích chọn đặc trưng cử chỉ của khuôn mặt theo thời gian thực được thực hiện bằng cách kết hợp một số tác vụ. Nó liên quan đến việc phát hiện khuôn mặt chính xác, trích chọn các đặc điểm khuôn mặt, biểu diễn các đặc điểm khuôn mặt và đào tạo một bộ phân loại mạnh mẽ. Nhiệm vụ thách thức nhất là phát hiện các đặc

điểm trên khuôn mặt và nhận dạng các đặc điểm này. Có một số lượng lớn các biểu cảm có thể được tạo ra bởi khuôn mặt con người. Để đơn giản hóa nhiệm vụ, nghiên cứu này đã được giới hạn trong sáu biểu cảm cơ bản trên khuôn mặt - vui, buồn, ghê tởm, ngạc nhiên, tức giận và tự nhiên (tức là không biểu cảm). Nghiên cứu trước đây trong lĩnh vực phát hiện đặc điểm khuôn mặt được xem xét dưới đây.

Biểu hiện trên khuôn mặt là kết quả của một hoặc nhiều chuyển động hoặc vị trí của các cơ trên khuôn mặt và theo thuyết tiến hóa của Darwin, các biểu hiện trên khuôn mặt là cố hữu, thay vì học được [32]. Những chuyển động này truyền tải trạng thái cảm xúc của cá nhân cho người quan sát và là một hình thức giao tiếp phi ngôn ngữ. Chúng là một phương tiện quan trọng để truyền đạt thông tin xã hội giữa con người, nhưng cũng xuất hiện ở hầu hết các loài động vật có vú khác và một số loài động vật khác.

Các nhà nghiên cứu đã thử nhiều phương pháp khác nhau để nhận ra và phân loại các biểu hiện trên khuôn mặt [40]. Nghiên cứu của Ekman chỉ ra rằng sáu biểu thức cơ bản có tính chất thống nhất. Sau đó, Ekman và Friesen [37] đã phát triển Hệ thống mã hóa hành động trên khuôn mặt (FACS), để phân loại mọi biểu hiện trên khuôn mặt con người có thể hình dung được; Essa và Pentland [39] đã cố gắng phân loại các đơn vị hành động trên khuôn mặt này. Đây là tiêu chuẩn phổ biến nhất hiện nay được sử dụng để phân loại một cách có hệ thống các biểu hiện thể chất của cảm xúc, và nó đã được chứng minh là hữu ích cho cả các nhà tâm lý học và các nhà hoạt hình. Có một số kỹ thuật xử lý hình ảnh hiện có mà các nhà nghiên cứu đã cố gắng sử dụng hoặc sửa đổi để tự động phân tích các biểu hiện trên khuôn mặt con người, bao gồm Mô hình Marko Ẩn, ANN và SVM. Ví dụ, Yoshitomi [127], đã cố gắng nhận ra các nét mặt bằng cách sử dụng ANN (ANN). Kumar và Poggio [67] đã thử sử dụng SVM (SVM) để phân loại các đặc điểm trên khuôn mặt và đã cố gắng nhận dạng nét mặt bằng Mô hình Hidden Marko (HMM). Kết quả của tất cả các thí nghiệm và thử nghiệm này chỉ ra rằng, trong khi các kỹ thuật hiện có này có thể phát hiện và nhận dạng chính xác nhiều loại đối tượng, chúng không phù hợp với việc phát hiện các biểu hiện trên khuôn mặt. Những thất bại của các hệ thống xử lý hình ảnh chung này đã khiến các nhà nghiên cứu cố gắng tạo ra các kỹ thuật đặc biệt để phát hiện và nhận dạng các biểu hiện trên khuôn mặt.

Một trong những vấn đề lớn khi cố gắng nhận dạng nét mặt bằng Mô hình Marko Ẩn, ANN hoặc SVM là số lượng lớn các kích thước tính năng được tạo ra trong quá

trình xử lý hình ảnh, do đó đòi hỏi một lượng lớn sức mạnh tính toán. Do đó, nhiều nhóm nghiên cứu đã cố gắng theo nhiều cách khác nhau để giảm số lượng thứ nguyên đối tượng được tạo ra bởi nhiệm vụ này. Các thuật toán trích xuất các đặc điểm khuôn mặt này có thể được phân loại rộng rãi thành các nhóm sau: cải thiện các đặc điểm trên khuôn mặt, thông tin hình học các đặc điểm khuôn mặt, thông tin màu sắc của các đặc điểm khuôn mặt, phương pháp lập bản đồ siêu mặt phẳng và các mô hình thống kê. Việc chiết xuất các đặc điểm trên khuôn mặt bằng phương pháp được gọi là hình ảnh khảm. Trong nghiên cứu của họ, hình ảnh được phân tách qua các mạng khác nhau và giá trị màu xám cho mỗi mạng được tính toán. Sau đó, nó sử dụng các đặc điểm đã biết của khuôn mặt, mắt và mũi để định vị những đặc điểm này trong ảnh. Phương pháp này yêu cầu một lượng tính toán đáng kể.

Turkan [111] đã giới thiệu phép chiếu ngang và dọc của các đặc điểm trên khuôn mặt, trong đó phân tách các đặc điểm trên khuôn mặt trong sơ đồ chiếu. Điều này đã được cải thiện bởi Feng và Yuen với Hàm chiếu phương sai của họ. Thuật toán này không đòi hỏi nhiều sức mạnh tính toán, nhưng nó thất bại hoàn toàn khi phát hiện bằng mắt không thành công. Ngoài ra, độ chính xác của nó giảm đáng kể khi các mục tiêu ở trong nền phức tạp hoặc ánh sáng không cân bằng.

Reisfeld [93] đã giới thiệu phương pháp biến đổi đối xứng tổng quát (GST) dựa trên sự đối xứng của các đặc điểm trên khuôn mặt và hình dạng của các đặc điểm trên khuôn mặt. Phương pháp này có thể được sử dụng cho các biểu thức và điều kiện ánh sáng khác nhau nhưng đòi hỏi sức mạnh tính toán rất lớn. Phương pháp này cũng không hoạt động khi phát hiện mắt không thành công vì nó sử dụng vị trí của mắt cho đối xứng tâm. Một kỹ thuật xử lý hình ảnh tổng quát được gọi là mô hình đường viền chủ động (còn được gọi là con rắn) được giới thiệu bởi Kass [64]. Điều này đã được Liu và Sclaroff [77] áp dụng cụ thể vào nhận dạng khuôn mặt, những người đã đưa ra các mô hình tham số để xác định vị trí mắt và miệng trong ảnh. Phương pháp này có một hạn chế tương tự như mô hình đường viền chủ động, độ chính xác phụ thuộc vào các điểm kiểm soát ban đầu.

Một cải tiến hơn nữa là một kỹ thuật dựa trên mô hình đường nét chủ động và được gọi là Mô hình ngoại hình chủ động (AAM) được giới thiệu bởi Cootes và cộng sự [59], tập trung sâu hơn vào việc chiết xuất khuôn mặt đặc trưng. Kỹ thuật này đã được sử dụng trong một số bài báo phân tích nét mặt gần đây. Thuật toán AAM đã được cải tiến thêm bởi Torre và cộng sự [110]. Wang [119] đã giới thiệu Căn chỉnh

đối tượng không cứng nhắc với Mẫu không khớp Dựa trên Tìm kiếm cục bộ hoàn chỉnh. Datcu và Rothkrantz [33] đã giới thiệu phân tích nét mặt đối với hình ảnh tĩnh và video. Wang [118] đã cải thiện hơn nữa việc đào tạo hệ thống đặc điểm khuôn mặt AAM bằng cách sử dụng SVM. Tuy nhiên, mặc dù, tất cả những cải tiến đáng kể này, thuật toán AAM vẫn có những hạn chế của nó. Mặc dù AAM là một công cụ tốt để phát hiện và định vị các đặc điểm khuôn mặt của mọi người, nhưng nói chung nó không đủ chính xác để phát hiện các biểu hiện trên khuôn mặt. Tuy nhiên, việc đào tạo một hệ thống AAM rất khó và tốn thời gian vì nó đòi hỏi nhiều điểm khác nhau (chẳng hạn như mắt và miệng) được đánh dấu bằng tay trên hình ảnh đào tạo. Hơn nữa, các nhà nghiên cứu vẫn cần cải thiện bước nhấn mạnh các đặc điểm trên khuôn mặt để giúp SVM phân loại biểu cảm trên khuôn mặt.

2.4 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 rời rạc 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. Theo đó, cây được tổ chức với mỗi nút là một hàm quyết định được học với một đặc trưng hình dạng. Với một ảnh khuôn mặt sau khi được định vị tập điểm điều khiển sẽ được đưa vào cây quyết định và xử lý. Khi đến nút lá, giá trị ước lượng biểu cảm đầu ra sẽ được xác định đơn giản là giá trị trung bình các mẫu học của nút lá.

Tập điểm định vị được xác định trên ảnh đầu vào bằng cách sử dụng thuật toán AAM. Từ vị trí khuôn mặt được phát hiện, tập điểm được khởi tạo ban đầu. Sau đó tại mỗi bước lặp, các vị trí sẽ được tối ưu dần đến khi được kết quả cuối cùng.

2.4.1 Mô hình ước lượng

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

$$\{(I_s, v_s, w_s): s = 1,2,\dots,S\} \tag{2.20}$$

trong đó, v_s là nhãn đúng của mẫu Is, 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:

$$WMSE(I, v, w) = \sum_{(I,v,w)\in C_0} w \cdot (v - \bar{v}_0)^2 + \sum_{(I,v,w)\in C_1} w \cdot (v - \bar{v}_1)^2 \qquad (2.21)$$

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.

Điều kiện để các hàm mục tiêu có được nghiệm tối ưu là hàm tồn tại đạo hàm bậc 2 dương

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.

2.4.2 Đề xuất sử dụng đặc trưng hình dạng

Trên cơ sở mô hình cây được mô tả như trên, 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.

$$f_{LINE_LINE}(p_1, p_2, p_3, p_4) = \frac{d(p_1,p_2)}{d(p_3,p_4)} \quad (2.22)$$

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.

$$f_{TRIANGLE_TRIANGLE}(p_1, p_2, p_3, p_4, p_5, p_6) = \frac{S(p_1,p_2,p_3)}{S(p_4,p_5,p_6)} \quad (2.23)$$

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

$$f_{LINELINE_LINELINE}(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8) = \frac{d(p_1,p_2)+d(p_3,p_4)}{d(p_5,p_6)+d(p_7,p_8)} \quad (2.24)$$

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.

2.4.3 Hàm quyết định

Hàm quyết định được xây dựng trên cơ sở một đặc trưng hình dạng và quyết định được đưa ra trên cơ sở so sánh giá trị đặc trưng với một ngưỡng. Ngưỡng của hàm sẽ được tính toán dựa trên tập dữ liệu đầu vào tại mỗi nút. Khi xây dựng một hàm quyết định tại mỗi nút, một tập các đặc trưng hình dạng sẽ được sinh ngẫu nhiên và tương ứng sẽ là một tập các hàm quyết định. Thông qua việc ước lượng ngưỡng cho mỗi hàm quyết định trong tập, hàm quyết định có sai số nhỏ nhất sẽ được lựa chọn.

Thuật toán tính ngưỡng được mô tả như sau:

Thuật giải 2.2 Tính toán Ngưỡng

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

Output: threshold

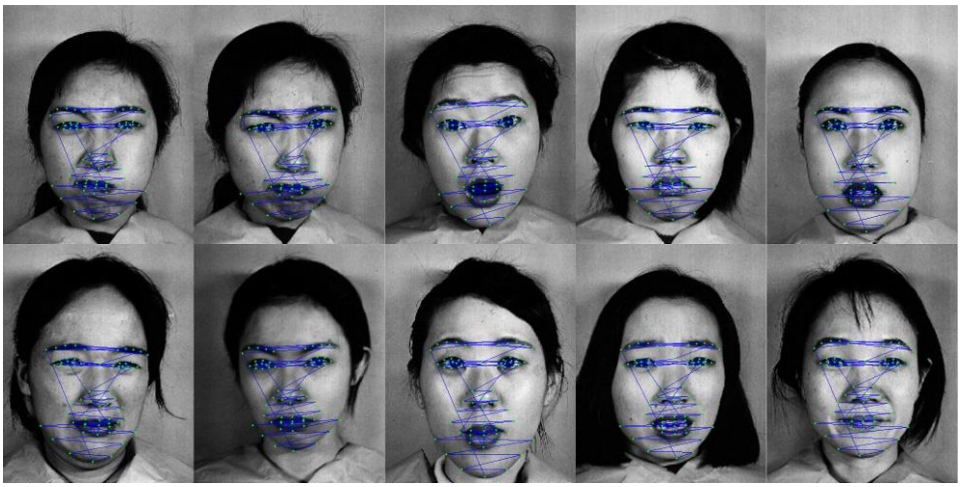
```
begin
    for each I_s
        feat_s := calc_feat(I_s);
    endfor
    sort(feat_s);
    best_threshold := 1;
    threshold := 1;
    min_error := WMSE(U);
    for i:=1 to S
        threshold := (feat_{i-1}+ feat_i)/2;
        error:= WMSE(U);
        if (error<min_error)
            min_error := error;
            best_threshold := threshold;
        endif
    endfor
    threshold := best_threshold;
end
```

2.4.4 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, phẫn 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 bằng 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 2.6 Một số kết quả định vị tập điểm điều khiển

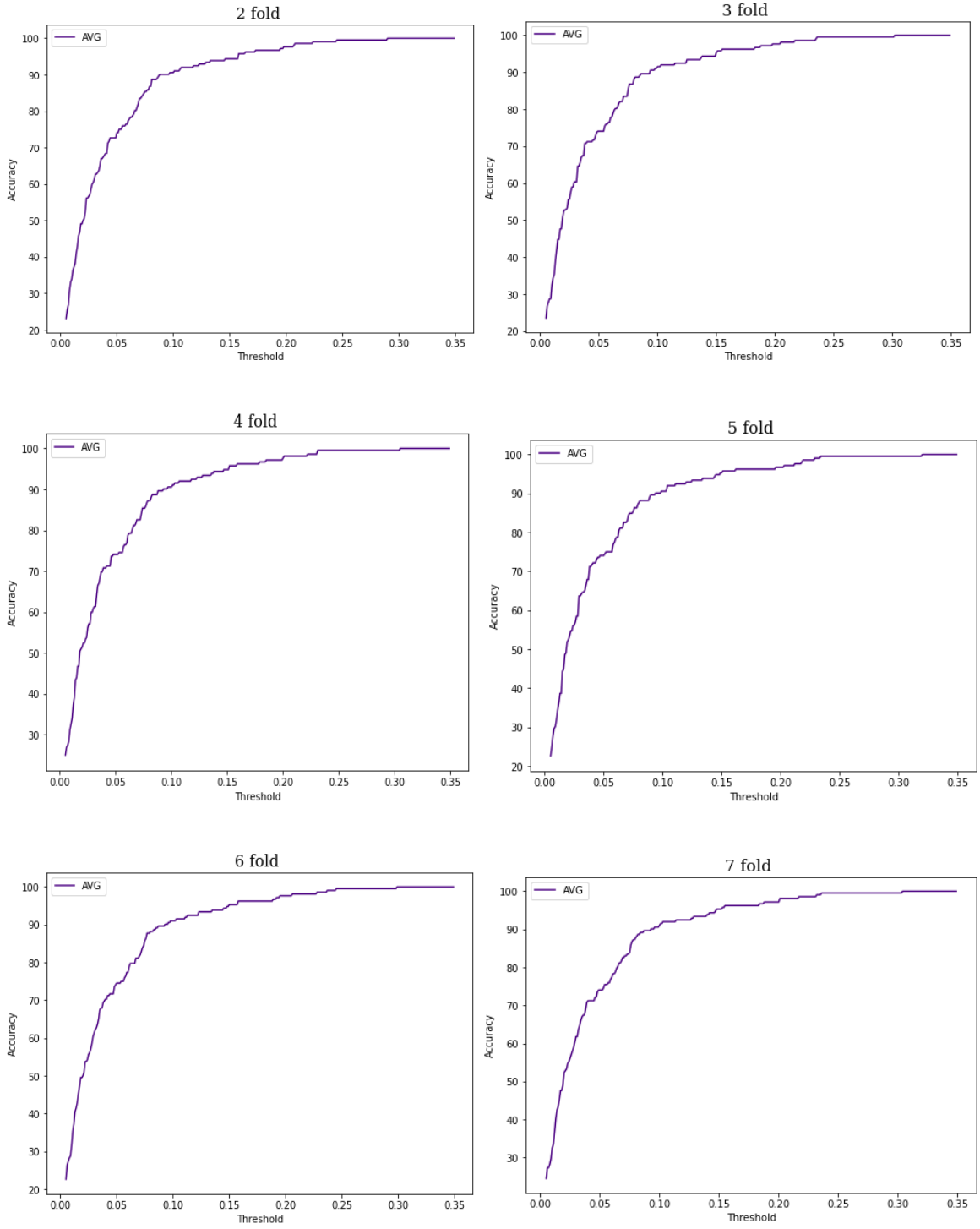
Trên cơ sở đã có tập các điểm điều khiển cho mỗi ảnh trong cơ sở dữ liệu, tác giả tiến hành thử nghiệm theo cách thức kiểm chứng chéo trên cơ sở dữ liệu với các mức phân đoạn từ 2 đến 8. Các cây nhị phân xây dựng được giới hạn chiều sâu bằng 8. Các thử nghiệm được tiến hành trên máy tính Core i7-4790 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:

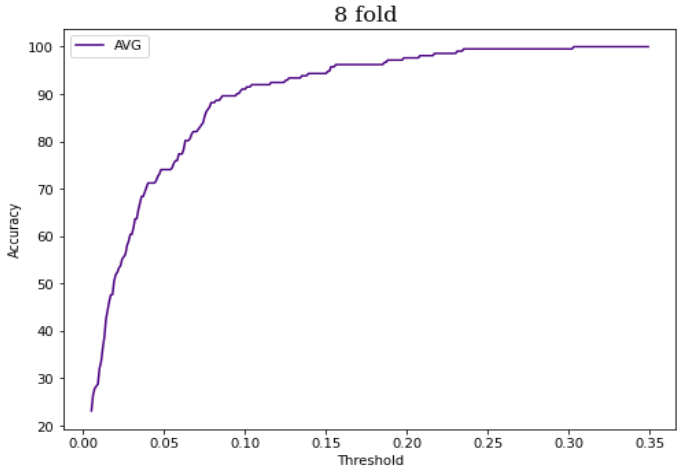
Bảng 2.3 Sai số trung bình của các loại biểu cảm

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

FEA (Sợ hãi)	0.038529	0.038167	0.038288	0.038275	0.0385	0.038231	0.0381

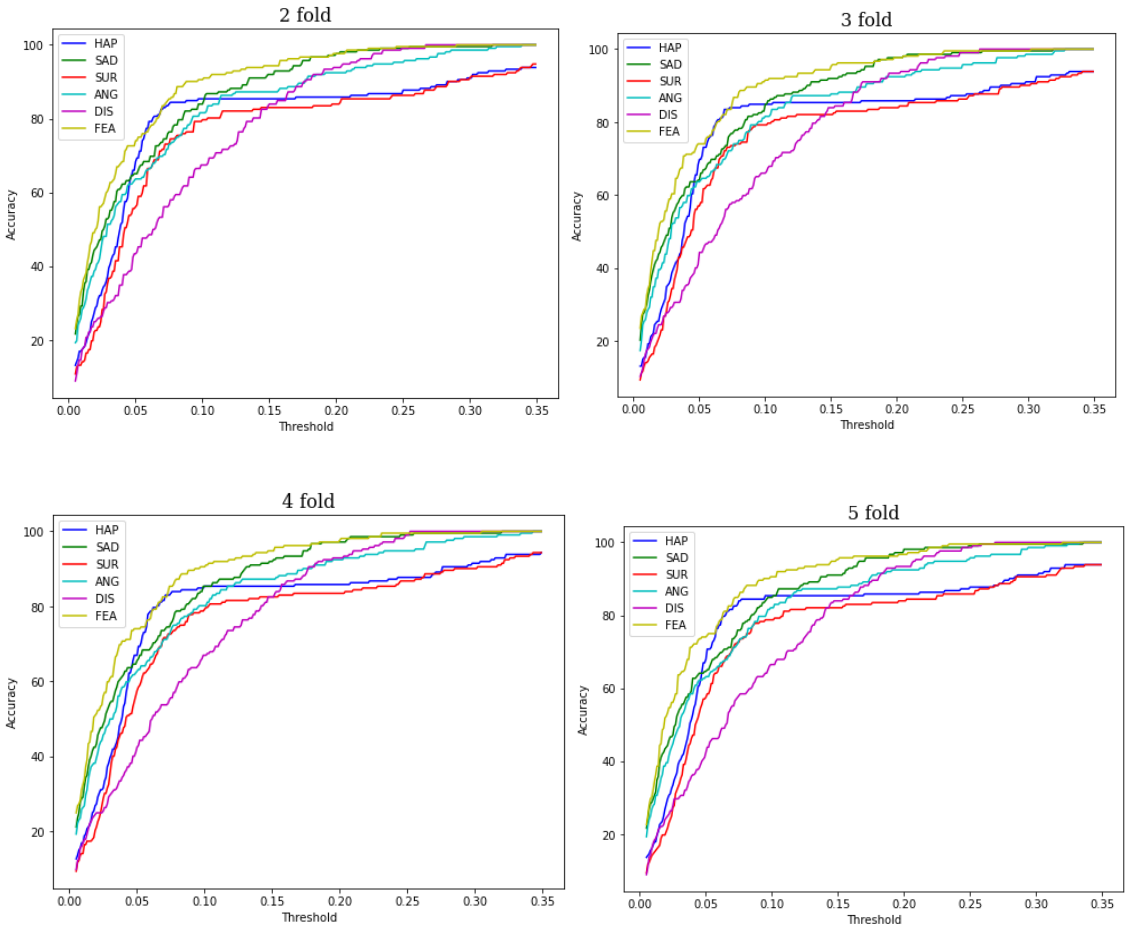
Ngoài ra, tác giả cũng thực hiện đánh giá thêm độ chính xác trên từng mẫu khi thực hiện so sánh sai số độ lệch với một giá trị ngưỡng truyền vào. Trên cơ sở đó thực hiện xây dựng lược đồ thống kê tương quan giữa tỉ lệ chính xác và ngưỡng chấp nhận.

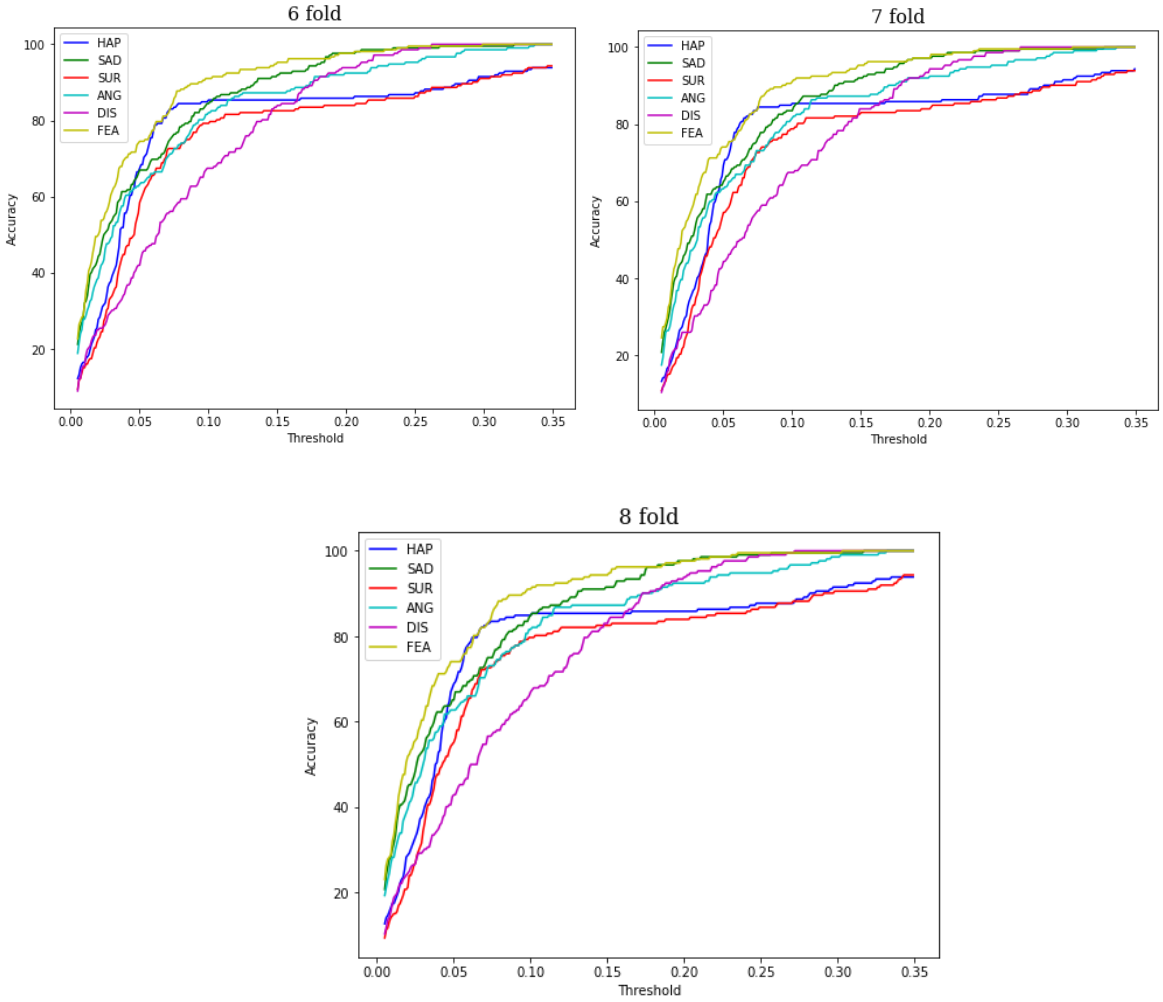




Hình 2.7 Các biểu đồ thống kê tỉ lệ chính xác theo ngưỡng chấp nhận

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 2.8 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 2.4 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	0.0123819292452829	0.013498740566037736	0.018362688679245294	0.022007566037735862	0.022420995283018886	0.014865476415094331	0.008695400943396223
SAD	0.0121416226415094 2	0.013486726415094337	0.018574801886792457	0.021877320754716 98	0.022367382075471724	0.014345896226415087	0.008226853773584 91
SUR	0.012202084905660349	0.013630353773584 91	0.018286169811320783	0.021446433962264144	0.021653150943396 23	0.013669367924528289	0.008152948113207 55
ANG	0.011878429245283007	0.013647783018867 93	0.018337193396226 43	0.021398150943396248	0.021427580188679247	0.013759254716981128	0.008265778301886802
DIS	0.012302891509433929	0.013776716981132081	0.018368056603773 6	0.021277283018867943	0.022114985849056582	0.013835707547169803	0.008484542452830193
FEA	0.012165584905660375	0.013560518867924 53	0.018035094339622647	0.020438221698113 23	0.021075900943396 21	0.013739051886792447	0.008059009433962271

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.

2.5 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.

Phát hiện được khôn mặt người chính là bước đầu tiên quan trọng của một hệ thống rút trích cử chỉ và biểu diễn biểu cảm. Vậy nên trong chương này, nghiên cứu sinh đã trình bày một số phương pháp tiếp cận và lựa chọn một kỹ thuật phát hiện khuôn mặt người 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 có khuôn mặt hay không. Nghiên cứu sinh cũng đã tiến hành thử nghiệm và đạt kết quả chính xác ở mức

92,44%; điều này cho thấy phương pháp được đề xuất cho độ chính xác phát hiện khuôn mặt khá tốt và ưu điểm của nó là có tốc độ xử lý nhanh hơn. Kết quả nghiên cứu đã được công bố trong công trình TCTN3.

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, nghiên cứu sinh đề xuất một kỹ thuật định vị điểm điều khiển có ràng buộc dựa trên mô hình chất liệu mặt. Thực nghiệm cho thấy kết quả có độ chính xác cao, đặc biệt là trong trường hợp màu da khác nhau, ánh sáng thay đổi.

Phần đóng góp quan trọng của chương 2 là nghiên cứu sinh đã đề xuất kỹ thuật ước lượng biểu cảm khuôn mặt dựa trên một số đặc trưng hình dạng. Trong đó cải tiến việc tự động lựa chọn các đặc trưng hình học một cách ngẫu nhiên và tổ chức dưới dạng mô hình cây quyết định để thực hiện ước lượng các biểu cảm khuôn mặt thay thế cho cách lựa chọn bằng tay một số đặt trưng. Trong đề xuất này, 3 loại đặc trưng hình dạng được sử dụng và từ đó xây dựng hàm quyết định. Thông qua kết quả thực nghiệm trên cơ sở dữ liệu JAFFE nghiên cứu sinh nhận thấy độ chính xác đạt đến mức xấp xỉ 95% chỉ từ mức ngưỡng khoảng 0,14. Kết quả nghiên cứu được công bố trong các công trình TCQT1, TCTN1.

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

Khả năng biến đổi của mô hình ba chiều có thể sẽ rất khác biệt với những chuyển động mà ta rút trích được từ biểu cảm khuôn mặt. Vì vậy, một trong những hướng tiếp cận điều khiển mô hình 3D thường được sử dụng là tiếp cận dựa vào điểm điều khiển. Phương pháp này xây dựng ánh xạ tương ứng giữa các tập điểm trên mô hình ba chiều với tập điểm thu nhận được từ giai đoạn trích chọn đặc trưng. Bước kế tiếp là biến đổi mô hình ba chiều với dữ liệu chuyển động đã được ước lượng từ tập điểm đặc trưng đã được tính toán. Giả sử mô hình 3D gồm một tập rất lớn các điểm, một cách tiếp cận thường được dùng là coi chuyển động của mỗi điểm thành phần có thể được ước lượng từ một tập giới hạn các điểm điều khiển, khi chúng ta biến đổi tập điểm này thì dữ liệu của đối tượng sẽ biến đổi theo, đây là cơ sở để biến đổi toàn bộ hình dạng của khuôn mặt. Cách tiếp cận này được thực hiện qua hai giai đoạn: thứ nhất là xác định tập điểm điều khiển và giá trị biến đổi, thứ hai là nắn chỉnh mô hình 3D để thể hiện biến đổi bằng cách tìm hàm biến đổi từ sự biến đổi của tập điểm điều khiển, sau đó áp dụng hàm biến đổi này lên toàn bộ đối tượng để thu được kết quả.

Bên cạnh đó, khuôn mặt người có tính đối xứng qua một mặt phẳng chia khuôn mặt thành hai nửa trái và phải. Sự cân xứng trên khuôn mặt không chỉ có tác dụng đối với những khuôn mặt có biểu cảm trung tính mà còn đối với hầu hết các biểu hiện trên khuôn mặt, đặc biệt là những biểu hiện tự nhiên thể hiện trạng thái cảm xúc hoặc do nói chuyện. Tính đối xứng của khuôn mặt đã được nhiều hệ thống nhận dạng khuôn mặt 3D sử dụng cho một số tác vụ như ước lượng tư thế và chỉnh sửa khuôn mặt 3D, nhờ đó họ ước tính mặt phẳng đối xứng của khuôn mặt và xác định các điểm tham chiếu trên mặt cắt đối xứng (tức là giao điểm giữa mặt phẳng đối xứng và mặt 3D). Nó cũng được sử dụng để trích xuất đặc điểm hoặc giảm kích thước của dữ liệu trên khuôn mặt [45] [51]. Tính đối xứng đó còn được ứng dụng trong việc nội suy các lỗ trong dữ liệu khuôn mặt thô (để ước tính tốt hơn dữ liệu bị thiếu), địa phương hóa các điểm khuyết (để có độ chính xác và chắc chắn hơn), phát hiện khuôn mặt (để nâng cao hiệu quả phát hiện) và bình thường hóa độ chiếu sáng của kết cấu khuôn

mặt 3D, v.v... Tuy nhiên, sự bất đối xứng của khuôn mặt người cũng có thể quan trọng đối với nhận dạng khuôn mặt vì nó có thể là một người cụ thể.

Hơn thế nữa, trên khuôn mặt người còn có các điểm mốc, có thể được phát hiện ngay cả khi khuôn mặt bị biến dạng. Các điểm mốc điển hình bao gồm khóe mắt, điểm giữa hai mắt, đầu mũi và hai góc dưới, điểm cằm xa nhất và khóe miệng. Đôi khi, cần thiết lập sự tương ứng điểm - điểm giữa hai hoặc nhiều lần quét khuôn mặt của cùng một người hoặc của những người khác nhau, như trường hợp của một số phương pháp tiếp cận mô hình biểu cảm và nhận dạng khuôn mặt bất biến. Tuy nhiên, đối với các khuôn mặt có thể biến dạng, việc thiết lập các tương ứng như vậy khó khăn hơn. Vấn đề có thể được xác định bằng cách thiết lập sự tương ứng giữa các điểm mốc của các lần quét mặt trước. Các điểm mốc cũng được sử dụng làm vị trí tiêu chuẩn mà từ đó các đặc trưng cục bộ có thể được trích xuất để nhận dạng khuôn mặt 3D. Các tính năng bổ sung có thể được trích xuất từ các vị trí tương đối của chúng để tăng cường khả năng nhận dạng.

Công ty Luxand sử dụng 66 điểm đặc trưng tương ứng với tọa độ của các thành phần khuôn mặt như mắt, viền mắt, lông mày, viền môi, mũi, má và cằm. Bộ công cụ này đã được sử dụng trong các ứng dụng như điều khiển vào ra, giám sát an ninh, xây dựng hoạt cảnh v.v..

Hình 3.1 Tập điểm của Luxand

Những tập điểm được đưa ra bởi các chuyên gia như vậy đã được ứng dụng thành công trong nhiều nghiên cứu và ứng dụng về biến đổi nắn chỉnh mô hình. Trong luận án này, nghiên cứu sinh đặt vấn đề theo một khía cạnh khác, đó là tự động lựa chọn tập điểm điều khiển dựa trên những mẫu quan sát về sự biến đổi của đối tượng. Nghiên cứu sinh nhận xét rằng khi một mô hình có sự biến đổi thì việc biến đổi của

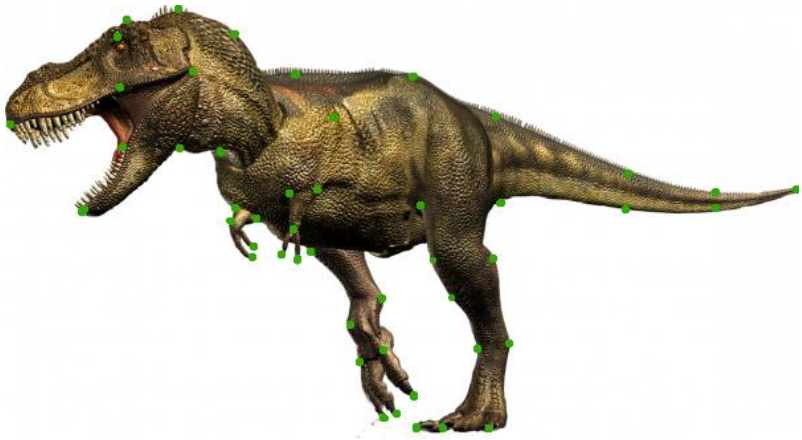
các điểm có sự phụ thuộc lẫn nhau, có những điểm biến đổi nhiều, có điểm biến đổi ít. Một cách hình dung từ thực tế: khi ta túm vào da và kéo lên, các điểm có sự biến đổi mạnh nhất là các điểm ta tóm vào, điểm có sự biến đổi ít sẽ là những điểm nằm ở xa vị trí kéo. Như vậy có thể coi rằng những điểm biến đổi mạnh nhất sẽ kéo những điểm khác lên còn những điểm ít biến đổi nhất sẽ giữ những điểm khác lại. Tiếp cận của nghiên cứu sinh được xây dựng dựa trên nhận xét này, đó là tìm kiếm những tập điểm có sự biến đổi tương đồng và từ đó xác định các điểm biến đổi mạnh nhất, yếu nhất để cho vào tập điểm điều khiển.

3.2 Kỹ thuật nội suy RBF

3.2.1 Ý tưởng tiếp cận của thuật toán

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.

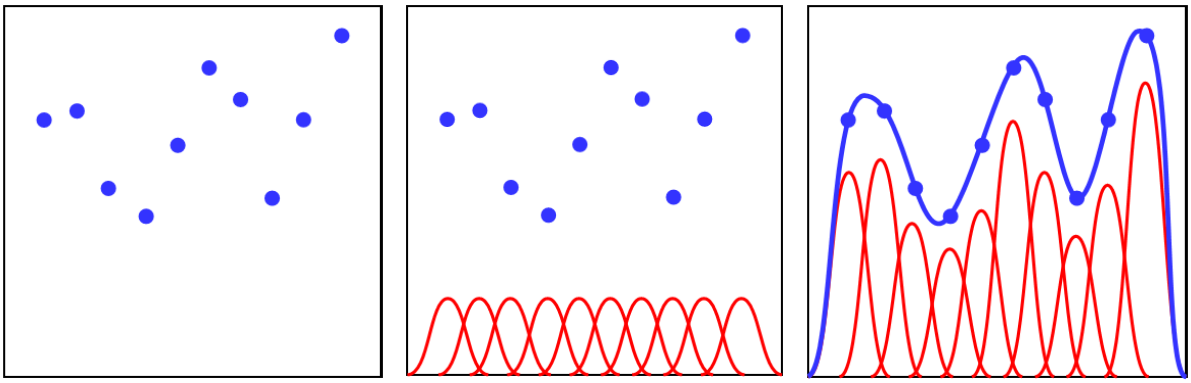
Trước tiên, các hệ số biến đổi sẽ được tính toán dựa vào sự thay đổi của tập điểm điều khiển, sau đó dữ liệu của đối tượng sẽ được tính toán lại bởi phương pháp nội suy dựa trên hàm cơ sở bán kính với các hệ số là các giá trị biến đổi vừa tính được. Lấy ví dụ là điều khiển một con Khủng Long, chúng ta sẽ xây dựng một tập điểm điều khiển là các điểm màu xanh và việc điều khiển hoạt động của nó được thực hiện trên tập điểm này thay cho việc điều khiển toàn bộ mô hình.



Hình 3.2 Điều khiển hoạt động của Khủng Long dựa vào điểm điều khiển

Ý tưởng của việc xây dựng Hàm cơ sở bán kính (RBF) lần đầu tiên được R.L. Hardy đưa ra dựa trên việc đặt các câu hỏi, "Cho một tập các dữ liệu nằm rải rác thưa

thớt $\{f_j\}_{j=1}^N$ tại các vị trí nút $\{x_j\}_{j=1}^N \subset \mathbb{R}^d$, có thể sử dụng một phép nội suy để tính toán lại đầy đủ các điểm chưa biết trên bề mặt?" [50]. Hardy xây dựng hàm nội suy một biến từ một hàm cơ sở bằng cách loại bỏ phép chiếu trực giao cho bất kỳ tập hợp các nút phân tán nào. Mặc dù không có điều kiện của điểm kỳ dị trong phép nội suy, nhưng nó có thể được dự báo sớm. Năm 1986 thì RBFs được đảm bảo là không có điểm kỳ dị [84] đã làm gia tăng hiệu quả của RBFs. Nghiên cứu của M.J.D. Powell tại Đại học Cambridge cũng đóng một vai trò quan trọng của sự phát triển RBFs.



a) Điểm điều khiển b) Hàm lựa chọn c) Hàm RBF

Hình 3.3 Hàm cơ sở bán kính (RBF)

Từng phần RBFs đại diện cho một bước nhảy trong một đạo hàm nào đó và có thể dẫn đến hội tụ. Mặt khác, việc nội suy với nhiều RBFs, như $\sqrt{1+(\varepsilon r)^2}, exp(-(\varepsilon r)^2)$, và $1/(1+(\varepsilon r)^2)$ sẽ dẫn đến quang phổ hội tụ. Driscoll và Fornberg [35] chỉ ra rằng độ min của RBFs phụ thuộc vào một tham số ε, trong 1-D, trong giới hạn của $\varepsilon \to 0$ kỹ thuật RBF tái tạo kỹ thuật pseudo-spectral (PS) nếu các nút được đặt phù hợp. Tương tự, trên bề mặt quả cầu, Fornberg và Piret [41] cho thấy RBFs tạo ra sóng hình cầu trong giới hạn của $\varepsilon \to 0$. Nghĩa là kéo dài một không gian tương ứng với bất kỳ một tập nút rời rạc nào.

3.2.2 Lựa chọn tham số hình dạng

Giả sử rằng giá trị của một hàm vô hướng $F: \mathbb{R}^3 \to \mathbb{R}$ là đại diện cho sự biến đổi của n điểm rời rạc khác biệt x_i trong không gian $\mathbb{R}^3$. Khi đó RBF cung cấp một phương thức để tạo ra một phép nội suy trơn cho hàm F trong không gian $\mathbb{R}^3$ và được viết dưới dạng tổng của n xấp xỉ bởi một hàm cơ sở bán kính. $g(r_i): \mathbb{R} \to \mathbb{R}$ với r_i là

khoảng cách giữa các điểm $p = (x,y,z)$ được xấp xỉ tương ứng $p_i = (x_i, y_i, z_i)$:

$$F(p) = \sum_{i=1}^{n} a_i g(||\boldsymbol{x} - \boldsymbol{x}_i||) + c_0 + c_1 x + c_2 y + c_3 z, \boldsymbol{x} = (x, y, z) \qquad (3.1)$$

Trong đó:

a_i : là các hệ số vô hướng.

c_0 đến c_3 : là đa thức bậc một mô tả một biến đổi affine mà không được thực hiện bởi hàm cơ sở bán kính.

Từ (3.1) cho thấy: với *n* điểm, để biết được giá trị $F(x_i, y_i, z_i) = F_i$ chúng ta phải mở rộng thêm bốn giá trị *n+1, n+2, n+3, n+4*, lý do là chúng ta có thêm bốn tham số c_i là các hệ số của đa thức.

Khi đó phương trình tuyến tính của (3.1) có dạng:

$$\boldsymbol{GA} = \boldsymbol{F} \qquad (3.2)$$

Trong đó:

F = $(F_1, F_2, \ldots, F_n, 0,0,0,0)^T$

A = $(a_1, a_2, \ldots, a_n, c_0, c_1, c_2, c_3)^T$

G là một ma trận cấp $(n + 4)$ x $(n + 4)$:

$$\boldsymbol{G} = \begin{bmatrix} g_{11} & g_{12} & \bullet & \bullet & \bullet & g_{1n} & 1 & x_1 & y_1 & z_1 \\ g_{21} & g_{22} & \bullet & \bullet & \bullet & g_{2n} & 1 & x_2 & y_2 & z_2 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ g_{n1} & g_{n2} & \bullet & \bullet & \bullet & g_{nn} & 1 & x_n & y_n & z_n \\ 1 & 1 & \bullet & \bullet & \bullet & 1 & 0 & 0 & 0 & 0 \\ x_1 & x_2 & \bullet & \bullet & \bullet & x_n & 0 & 0 & 0 & 0 \\ y_1 & y_2 & \bullet & \bullet & \bullet & y_n & 0 & 0 & 0 & 0 \\ z_1 & z_2 & \bullet & \bullet & \bullet & z_n & 0 & 0 & 0 & 0 \end{bmatrix}$$

Với: $g_{ij} = g(||\boldsymbol{x}_i - \boldsymbol{x}_j||)$, ta có thể lựa chọn các hàm g_{ij}, các hàm được lựa chọn là x – biharmonic ($\mathbb{R}^3$), x^{2n+1} – polyharmonic ($\mathbb{R}^3$) và $\sqrt{x^2 + c^2}$ *(với c là một hằng số)* - multiquadric ($\mathbb{R}^3$):

$$g(x) = \begin{cases} x \\ x^{2n+1} \\ \sqrt{x^2 + c^2} \end{cases} \tag{3.3}$$

Gải hệ phương trình (3.2) cho chúng ta các hệ số của **A** chính là các hệ số a_i của phương trình (3.1).

3.2.3 Thuật toán nội suy

Bài toán của chúng ta là: bề mặt hình học của đối tượng sẽ được đặc trưng bởi một tập điểm gọi là điểm điều khiển, khi chúng ta thay đổi các giá trị của các điểm điều khiển này thì đồng thời bề mặt hình học của đối tượng cũng thay đổi theo.

Giả sử rằng sự biến dạng được biết đến với n vị trí x_i trong không gian $\mathbb{R}^3$ và rằng thông tin này được đại diện bởi một vector mô tả 3D u_i rời rạc của hình học đối tượng đã được đặt ở vị trí x_i trong bề mặt đối tượng ban đầu, tức là bề mặt chưa được biến dạng. Chúng ta có thể xem các vị trí x_i như các điểm điều khiển đã được chuyển đến vị trí $x_i + u_i$. Phương pháp nội suy RBF bây giờ có thể được sử dụng cho các nội suy chuyển vị cho các vị trí khác. Để giải quyết bài toán này, RBF đưa ra hai bước: thứ nhất là xác định các giá trị biến đổi của tập điểm điều khiển và tính ra các tham số **A;** thứ hai là tính toán các giá trị biến đổi của các điểm dựa vào vector **A** và ma trận khoảng cách **G'**.

Dùng các ký hiệu: $\boldsymbol{x}_i = (x_i, y_i, z_i)$ và $\boldsymbol{u}_i = (u_i^x, u_i^y, u_i^z)$

Ba hệ thống tuyến tính được thiết lập như trên cho phép chúng ta tính các dịch chuyển u dựa vào các tham số **A** tính được ở phần trước:

$$\begin{aligned} G'A_x &= (u_1^x, u_2^x, \dots, u_m^x)^{\mathrm{T}} \\ G'A_y &= (u_1^y, u_2^y, \dots, u_m^y)^{\mathrm{T}} \\ G'A_z &= (u_1^z, u_2^z, \dots, u_m^z)^{\mathrm{T}} \end{aligned} \tag{3.4}$$

Với m là số điểm trên bề mặt, và **G'** là ma trận được xây dựng là *(m*n+4)* tương tự như ma trận **G** trong công thức (2).

Quy trình thực hiện:

Thuật giải 3.1 Nội suy RBF

Đầu vào: Bề mặt hình học của đối tượng 3D và tập điểm điều khiển

Đầu ra: Bề mặt hình học của đối tượng 3D sau khi nội suy

Thuật toán:

Bước 0: Bắt đầu

Bước 1: Xác định giá trị biến đổi của tập điểm điều khiển.

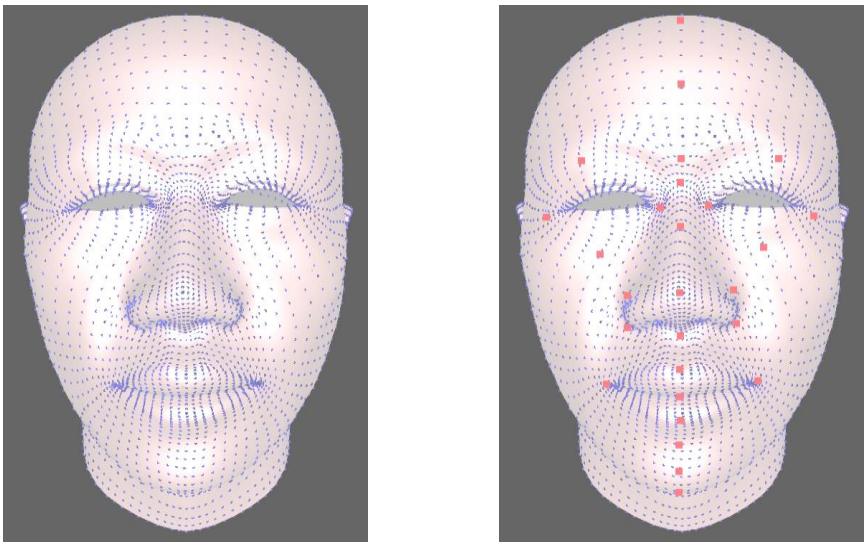
Bước 2: Tính $\mathbf{A} = (a_1, a_2, \ldots, a_M, c_0, c_1, c_2, c_3)$ theo (3.2).

Bước 3: Nội suy

Bước 4: Kết thúc

Mô tả chi tiết các bước như sau:

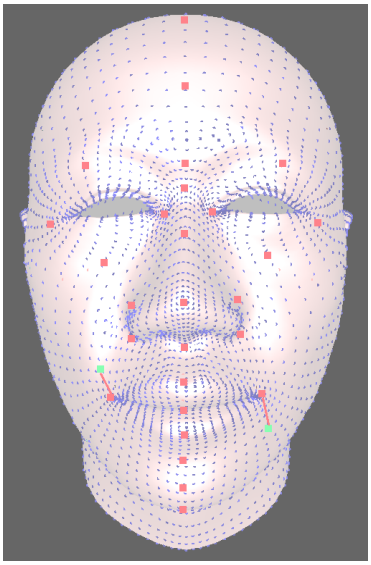
Bước 0: Chuẩn bị các tham số đầu vào cho thuật toán: Đối tượng 3D (quan tâm đến hình học của đối tượng), thông tin về tập điểm điều khiển (tập điểm này có thể là biết trước, hoặc cũng có thể khai báo).



a) Thông tin bề mặt đối tượng b) Thông tin điểm điều khiển

Hình 3.4 Đầu vào của thuật toán nội suy dựa vào RBF

Bước 1: Chúng ta xác định giá trị biến đổi của tập điểm điều khiển. Giả sử rằng các điểm sẽ có sự dịch chuyển từ vị trí màu đỏ sang vị trí màu xanh.



Hình 3.5 Đầu vào của thuật toán nội suy dựa vào RBF

Bước 2: Để tính tham số $\boldsymbol{A} = (a_1, a_2, \dots, a_M, c_0, c_1, c_2, c_3)$.

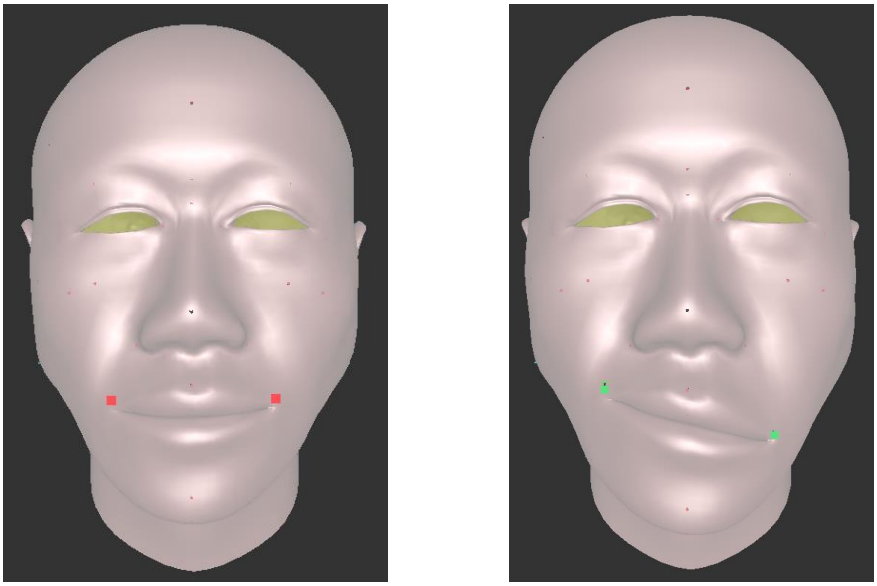
Sau khi có được giá trị của tập điểm điều khiển mới, ta tiến hành áp dụng công thức (3.2): Chúng ta xây dựng được ma trận **G** cho tập điểm điều khiển, và lựa chọn hàm $g(x)$. Trong trường hợp thử nghiệm, tôi lựa chọn hàm

$g(x) = \sqrt{x^2 + c^2}$ *(với c là một hằng số)* – multiquadric ($\mathbb{R}^3$). (3.5)

Tiếp theo là xác định giá trị của vector $\boldsymbol{F} = (F_1, F_2, \dots, F_M, 0,0,0,0)$ là độ dịch chuyển của điểm điều khiển từ giá trị ban đầu sang giá trị mới. Và khi đó vector tham số **A** được xác định.

Bước 3: Nội suy là quá trình tính lại các giá trị tọa độ của các điểm trên mô hình. Giá trị của các điểm này được tính dựa vào công thức (3.4) như sau: với mỗi một điểm, chúng ta xẽ xây dựng được ma trận **G'** kết hợp với tham số **A** tính được ở bước 2 theo công thức (3.4) ta tính được giá trị của các $F = (u_1, u_2, \dots, u_m)$.

Bước 4: Kết thúc nội suy, chúng ta thu được hình dạng mới của đối tượng.



a) Đối tượng trước nội suy b) Đối tượng sau nội suy

Hình 3.6 Kết quả nội suy đối tượng

3.3 Nội suy thể hiện các cử chỉ, biểu cảm của khuôn mặt người 3D

Toàn bộ quá trình nội suy tạo thể hiện các cử chỉ, biểu cảm của khuôn mặt người có thể được tóm tắt như sau: Đầu tiên là bước chuẩn bị dữ liệu. Một tập các mô hình mẫu bao gồm một mô hình cân bằng và các mô hình biểu diễn trạng thái mặt theo các trạng thái được xác định trước. Trên cơ sở đó, ta sẽ ước lượng được một tập các vector đặc trưng làm dữ liệu đầu vào cho phép biến đổi một mô hình cân bằng sang một mô hình biểu diễn biểu cảm tương ứng. Mỗi tập vector biến đổi này có thể được lưu trữ vào cơ sở dữ liệu của hệ thống để phục vụ cho mục đích sử dụng lại. Bước tiếp theo là thực hiện việc tính toán mô phỏng biến đổi cho một mô hình cân bằng đầu vào.

Việc biến đổi nội suy được thực hiện trên cơ sở mô hình cân bằng - là một mô hình biểu diễn trạng thái cân bằng của mặt người với các đặc điểm như sau:

- Ánh mắt theo hướng trục Z
- Tất cả cơ mặt đều thư giãn
- Các mí mắt tiếp tuyến với tròng đen
- Đồng tử (con ngươi) bằng 1/3 đường kính của tròng đen
- Các môi khép lại
- Đường giữa hai môi nằm ngang và có cùng độ cao với khoé môi.

- Miệng đóng và răng trên chạm vào răng dưới
- Lưỡi bằng, nằm ngang với đầu lưỡi chạm vào đường biên giữa răng trên và răng dưới

3.3.1 Ước lượng tập vector đặc trưng

Giả sử, cần tính tập vector biến đổi cho biểu cảm khuôn mặt tương ứng với khi phát âm tiết "a". Như vậy, trước tiên ta cần lấy ra cặp mô hình mẫu ở trạng thái cân bằng, và mô hình biểu cảm theo âm tiết "a". Tập các vector biến đổi sẽ được tính toán như sau. Với mỗi một trong bốn bộ phận trên của mô hình, mỗi điểm trên bộ phận đó sẽ có một vector biến đổi tương ứng. Vector này có gốc tại điểm trên mô hình cân bằng, và ngọn là điểm tương ứng trên mô hình đích. Tuy nhiên, ta cũng nhận thấy rằng, thực tế, chỉ có một số điểm trên mô hình được dịch chuyển. Như vậy, việc lưu trữ toàn bộ vector biến đổi cho tất cả các điểm sẽ trở nên lãng phí và không cần thiết. Do đó, ta sẽ chỉ cần lưu các vector cho các điểm có dịch chuyển. Bên cạnh các toạ độ x, y, z của vector, ta cần lưu trữ thêm chỉ số của điểm trong mô hình ứng với vector đó để có thể xác định điểm cần dịch chuyển khi thực hiện biến đổi các mô hình sau này.

(a) (b)

Hình 3.7 Mô hình mẫu ở trạng thái cân bằng (a), và trạng thái đích (b)

Giả sử đã có *N* là tập điểm của mô hình cân bằng và *T* tập điểm của mô hình đích.

$$N = (x_{11}, y_{11}, z_{11}, \dots x_{1n_1}, y_{1n_1}, z_{1n_1}, \dots x_{mn_m}, y_{mn_m}, z_{mn_m}) \in \Re^{3\times m\times n}$$

$$T = (x_{11}, y_{11}, z_{11}, \dots x_{1n_1}, y_{1n_1}, z_{1n_1}, \dots x_{mn_m}, y_{mn_m}, z_{mn_m}) \in \Re^{3\times m\times n}$$

Ta cần tính ra tập vector V để biến đổi từ N sang T, cùng với tập chỉ số I tương ứng. Công thức tính như sau:

$$\begin{cases} V_{ik} = T_{ij} - N_{ij} \\ I_{ik} = j \end{cases} \text{ với } \begin{cases} j \in [1, n_i] / N_{ij} \neq T_{ij} \\ k \in [1, p_i] \end{cases} \quad (3.6)$$

Trong đó:

m: số bộ phận của mô hình chịu sự biến đổi.

n_i: số điểm trên bộ phận thứ i.

p_i: số vector biến đổi của bộ phận thứ i.

Dưới đây là thuật toán để tính vector biến đổi cho một phép biến đổi từ mô hình cân bằng sang một mô hình ứng với một biểu cảm đặc trưng nào đó.

Thuật giải 3.2 Tập hợp các vector biến dạng

Input:

NeutralModel : Mô hình cân bằng.

TargetModel: Mô hình được biến đổi theo biểu cảm đặc trưng.

Output:

vector: Tập vector biến đổi theo một biểu cảm đặc trưng nào đó.

vec_index: Tập các chỉ số của điểm trên mô hình có vector biến đổi tương ứng trong tập vector.

For each part ∈(Skin, LowerTeeth, UpperTeeth, Tongue)
Begin
{ Tính tập vector biến đổi cho bộ phận part }
$num_of_vector_{part}$ = 0
For i = 1 → num_of_point(part)
Begin
{ Tính vector biến đổi tại điểm i }
If ($x_{i(NeutralModel)} \neq x_{i(TargetModel)}$
and $y_{i(NeutralModel)} \neq y_{i(TargetModel)}$
and $z_{i(NeutralModel)} \neq z_{i(TargetModel)}$
)
Begin

```
                vector_i = (x_i(TargetModel) - x_i(NeutralModel) ,
                        y_i(TargetModel) - y_i(NeutralModel) ,
                        z_i(TargetModel) - z_i(NeutralModel) )
                vec_index num_of_vector  = i
                increase(num_of_vector)
            End
        End
    End
```

Tập các vector này sẽ được lưu lại. Với mỗi biểu cảm định trước, ta sẽ có một vector biến đổi tương ứng.

3.3.2 Tính toán mô hình đích

Đầu tiên, ta nạp vào mô hình cân bằng cần biến đổi. Sau đó, đọc các vector từ file lưu trữ tập vector cho biểu cảm tương ứng. Tiếp theo, các bộ phận *Da mặt, Hàm răng dưới, Hàm răng trên* và *Lưỡi* của mô hình đích sẽ được biến đổi bằng các dịch chuyển các điểm theo các vector biến đổi tương ứng. Lưu ý là chỉ có một số điểm được lưu trữ vector biến đổi mới được dịch chuyển.

Bây giờ, giả sử ta đã có *N* là tập điểm của mô hình cân bằng, *V* tập vector biến đổi, và *I* là tập chỉ số vector.

$$N = (x_{11}, y_{11}, z_{11}, \dots x_{1n_1}, y_{1n_1}, z_{1n_1}, \dots x_{mn_m}, y_{mn_m}, z_{mn_m}) \in \Re^{3\times m\times n}$$
$$V = (x_{11}, y_{11}, z_{11}, \dots x_{1n_1}, y_{1n_1}, z_{1n_1}, \dots x_{mp_m}, y_{mn_m}, z_{mp_m}) \in \Re^{3\times m\times p} \quad (3.7)$$
$$I = \{k \in [1, n_i] / T_{ik} \neq N_{ik}\}$$

Trong đó:

m: số bộ phận của mô hình chịu sự biến đổi.

n_i: Số điểm trên bộ phận thứ *i*.

Ta cần tính ra tập điểm *T* của mô hình đích. Công thức tính như sau

$$\begin{cases} j = I_{ik} \\ T_{ij} = N_{ij} - N_{ik} \end{cases} \quad \text{với} \quad \begin{cases} i \in [1, m] \\ k \in [1, n_i] / N_{ik} \neq T_{ik} \\ j \in [1, n_i] \end{cases} \tag{3.8}$$

Dưới đây là thuật toán để biến đổi mô hình theo mỗi biểu cảm đặc trưng.

Thuật giải 3.3 Tính toán mô hình mục tiêu

Input:

NeutralModel: Mô hình cân bằng.

vector: Tập vector biến đổi theo một biểu cảm đặc trưng nào đó.

vec_index: Tập các chỉ số của điểm trên mô hình có vector biến đổi tương ứng trong tập vector.

Output:

TargetModel: Mô hình theo biểu cảm đặc trưng nào đó.

For each part ∈(Skin, LowerTeeth, UpperTeeth, Tongue)
Begin
{ Biến đổi bộ phận part }
For i = 1 → num_of_vector(part)
Begin
{ Dịch chuyển điểm trên mô hình ứng với vector biến đổi thứ. }
$point_{vec_index\ (i)}$ (TargetModel) ← $point_{vec_index\ (i)}$ (NeutralModel) + $vector_i$
End
End

3.3.3 Tính toán quá trình biến đổi

Để nội suy biến đổi giữa các mô hình để biểu thị quá trình biểu cảm hay thực hiện cử chỉ, ta cần sinh ra các trạng thái chuyển tiếp và hiển thị chúng liên tục trên các khung hình. Quá trình đó thực hiện như sau:

- Sinh ra một mô hình hình học (geometry) trung gian, cụ thể trong mô hình là các lưới đa giác.
- Sinh ra cấu trúc bề mặt (texture) trung gian, và ánh xạ với mô hình hình học trung gian được tạo.

- Hiển thị mô hình vào trong mỗi khung hình.

a. Thiết lập phép biến đổi ra mô hình trung gian

Giả sử ta cần thực hiện biến đổi giữa hai đối tượng hình $S_i$có dạng hình học là G_i và kết cấu bề mặt là T_i, và S_j có dạng hình học là G_j và kết cấu bề mặt là T_j. Ta ký hiệu: $S = <G,T>$.

Với giả thiết đã có $T_t = T_i = T_j$, do đó, chỉ cần thực hiện biến đổi giữa G_i và G_j.

$$G_i = (x_1, y_1, z_1, \dots x_n, y_n, z_n) \in \Re^{3\times n}$$
$$G_j = (x_1, y_1, z_1, \dots x_n, y_n, z_n) \in \Re^{3\times n} \quad (3.9)$$

Quá trình biến đổi phát sinh ra các trạng thái chuyển tiếp $S_t = <G_t, T_t>$, giữa S_i và S_j. Các trạng thái này được quy định bởi một hệ số thời gian $t \in [0,1]$. Hình trung gian được sinh theo công thức nội suy sau:

$$G_t = (1-t)G_i + tG_j \text{ với } 0 \le t \le 1 \quad (3.10)$$

Như vậy, khi thay đổi giá trị t thì sẽ sinh ra các trạng thái chuyển tiếp khác nhau. Cụ thể, khi t càng gần với 0 thì trạng thái chuyển tiếp càng giống với S_i ; ngược lại, t càng tiến về 1 thì trạng thái chuyển tiếp càng giống với S_j.

Bây giờ, giả sử phép biến đổi được thực hiện giữa nhiều hơn hai hình, ví dụ n hình $S_1 ... S_n$. Việc phát sinh các hình trung gian cũng được thực hiện tương tự.

Ta gọi w_i là trọng số đối với hình S_i. Các w_i được ràng buộc để có tổng là 1. Một dạng hình học trung gian giữa các hình G_i với trọng số w_i được tính theo công thức sau:

$$G_t = \sum_{i=1}^{n} w_i G_i \text{ với } \sum_{i=1}^{n} w_i = 1 \quad (3.11)$$

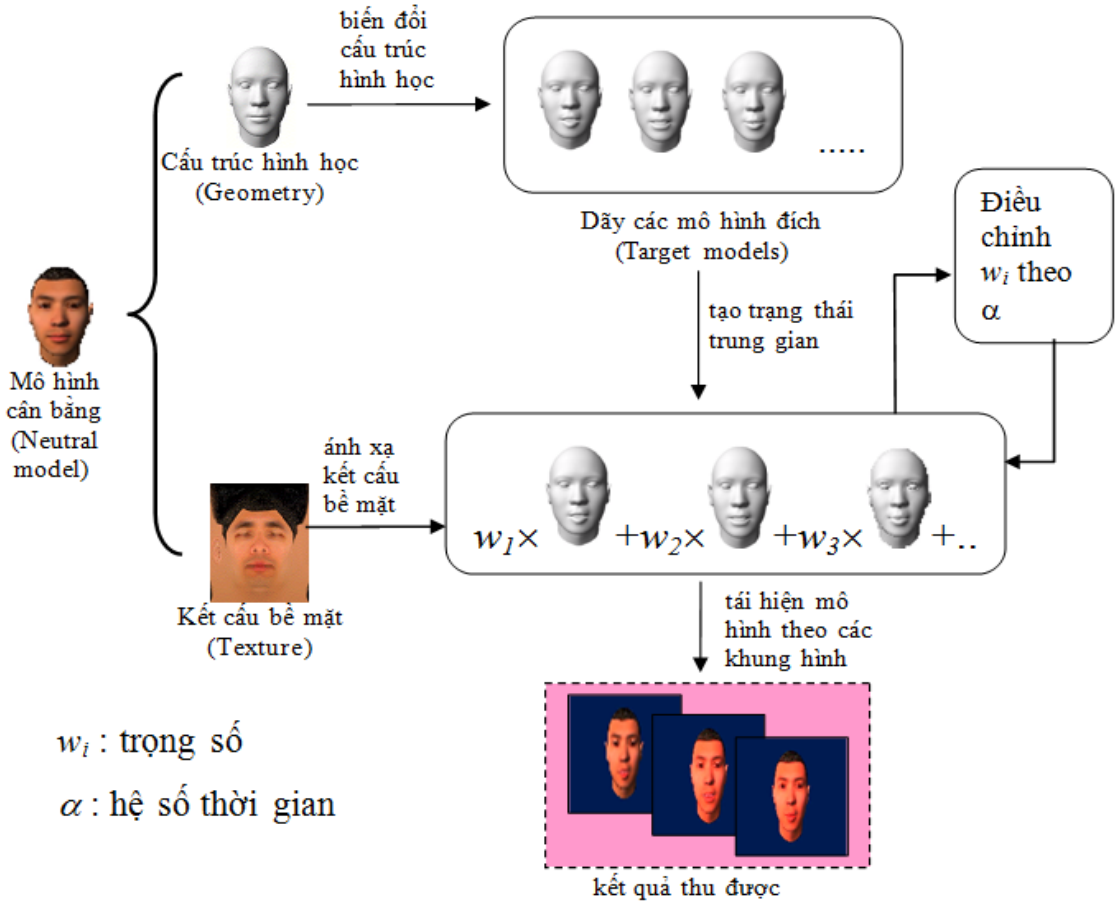
Tương tự, việc thay đổi trọng số w_i sẽ cho ta các hình trung gian khác nhau. Cụ thể là khi w_i càng lớn thì S_t sẽ càng giống với S_i.

b. Điều chỉnh theo thời gian

Mỗi hình S_i có một trọng số w_i kèm theo. Trọng số này sẽ được thay đổi theo thời gian. Theo đó, các w sẽ được điều chỉnh theo công thức dưới đây.

$$w_{j=1..n} = \begin{cases} 1-\alpha, j = i \\ \alpha \quad , j = i+1 \\ 0 \quad , j \neq i \text{ và } j \neq i+1 \end{cases} \tag{3.12}$$

Trong đó, $\alpha \in [0,1]$, là một hệ số biến đổi theo thời gian giữa mỗi cặp khung hình cơ bản.



Hình 3.8 Lược đồ nội suy biểu cảm

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

3.4.1 Đặc trưng biến dạng của các điểm

Việc xác định tập điểm điều khiển của mô hình phục vụ nắn chỉnh biến dạng được thực hiện dựa trên việc phân tích một tập các mô hình quan sát của đối tượng quan tâm. Các mô hình này có số lượng điểm và các quan hệ cạnh giống nhau. Giả sử mô hình A của một đối tượng 3D gồm N điểm, và được biểu diễn như sau:

$A = \{a_0, a_1, .., a_{N-1}\}$, trong đó a_i là điểm thứ i của mô hình A, $a_i \in R^3$

Ta thu thập được M quan sát cho đối tượng đó, tức là tương ứng với M mô hình 3D khác nhau, tập quan sát các điểm đó có thể được biểu diễn là:

$S = \{a_{ij} \mid i=0..N-1, j=0..M-1\}$,

Trong đó a_{ij} là điểm thứ i của mẫu quan sát thứ j của đối tượng 3D ta đang quan tâm, $a_{ij} \in R^3$.

Như vậy điểm thứ i của đối tượng 3D sẽ có M thể hiện tương ứng với M quan sát. Ta gọi tập $T_i = \{a_{i0}, a_{i1},..,a_{iM-1}\}$ là quỹ đạo biến đổi của điểm i trong tập S. Như vậy, T_i sẽ tương ứng thể hiện các biến thể của điểm i trong tập quan sát S của đối tượng 3D quan tâm.Việc xác định các điểm có sự biến đổi mạnh và yếu sẽ được thực hiện thông qua việc tính toán trên các quỹ đạo biến đổi của từng điểm trong mô hình.

3.4.2 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 [9] – một thuật toán được sử dụng khá phổ biến trong các bài toán gom cụm. Thuật toán 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. Để gom cụm, ta cần đến một hàm khoảng cách giữa hai đặc trưng:

Thuật giải 3.4 Chỉ số tính năng biến dạng

Input: 2 quỹ đạo biến đổi T_1, T_2
Output: khoảng cách d

```
Begin
    d := 0
    foreach i in [0, |T1|)
        d = d + distance(T1[i], T2[i])
    endfor
    d := d/|T1|
End
```

Tại mỗi bước lặp của vòng lặp tại dòng 3, ta cần tính lại tâm của các nhóm điểm. Để bảo đảm tâm là một điểm của mô hình khuôn mặt, từ tâm đã tính được, với mỗi nhóm điểm, ta chọn 1 điểm làm tâm theo tiêu chí gần nhất với điểm trung bình của nhóm điểm với khoảng cách được sử dụng là khoảng cách giữa 2 quỹ đạo từ hàm Chỉ số tính năng biến dạng

Thuật toán gom cụm các nhóm tương đồng như sau:

Thuật giải 3.5 Tính năng biến dạng cụm

Input: Tập các đặc trưng T, số cụm K
Output: Tập đại diện delegates, tập cụm clusters
Variables: Biến tạm lưu các cụm tmpClusters, biến tạm lưu các đại diện tmpDelegates

```
Begin
    delegates := select_random(T, K)
    resize(tmpClusters, K)
    foreach l in [0, MAX_LOOP)
        foreach k in [0, K)
            tmpClusters[k] := ∅
        endfor
        foreach i in [0, |T|)
            cid = get_nearest_cluster(T, i, delegates)
            tmpClusters[cid] := tmpClusters[k] ∪ {i}
        endfor
        tmpDelegates := calc_delegates (T, tmpClusters)
        eps := 0
        foreach k in [0, K)
            eps := eps + euclidDistPointTrajectory(
                        delegates[k], tmpDelegates[k])
        endfor
        eps := eps/K
        if eps < MIN_EPS
            break
        endif
        delegates := tmpDelegates
```

```
        cluster := tmpCluster
    endfor
End
```

Từ kết quả của thuật toán ClusterDeformationFeature, ta có được các nhóm tường 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.

3.4.3 Phân tích hiệu quả trong biến đổi mô hình

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. Thuật toán chung như sau:

Thuật giải 3.6 Phân tích biến đổi

Input: Mô hình tham chiếu R, tập quỹ đạo biến đổi T, tập mô hình đánh giá M

Output: tập điểm điều khiển P

Variables: danh sách các giá trị K ứng cử cho gom cụm KC

Biến tạm lưu các cụm `clusters`

Biến tạm lưu các đại diện `delegates`

```
Begin
    KC := init_K_candidate()
    minErr := MAX_VAL
    foreach K in KC
        Tính_năng_biến_dạng_cụm(T, K, delegates, clusters)
        tmpP := select_controller_points(T₁, clusters, R)
```

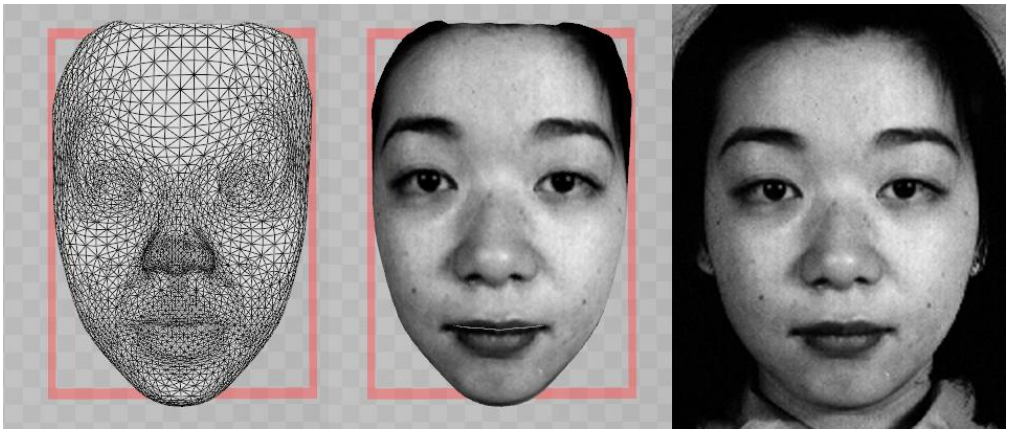
```
            err := 0
            foreach i in [0, |M|)
                  err := err + err_morph(M[i], R, tmpP)
            endfor
            err := err/|M|
            if err < miner
                  minerr := err
                  P := tmpP
            endif
      endfor
End
```

Trong thuật toán, hàm errMorph thực hiện đánh giá sai số giữa mô hình mặt đánh giá M[i] và mô hình mặt được biến đổi từ mô hình mặt tham chiếu R dựa vào tập điểm điều khiển được cho

3.4.4 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 [125] sử dụng hàm cơ sở bán kính [22] [42] (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.

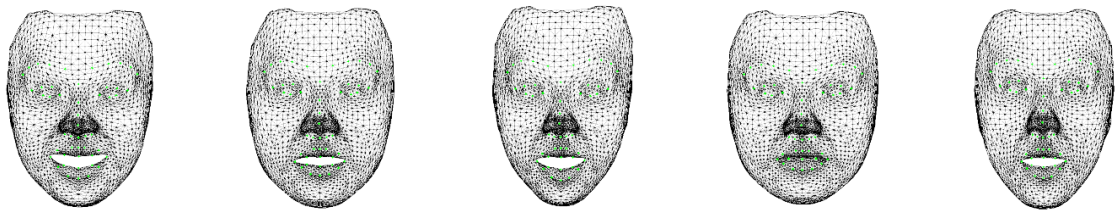
Đối tượng 3D sử dụng là mô hình khuôn mặt có 3448 đỉnh và 6736 bề mặt được tạo ra tương ứng với các mẫu trong cơ sở dữ liệu JAFFE [85]. JAFFE chứa 213 ảnh của 6 biểu cảm khuôn mặt khác nhau bao gồm: vui, buồn, ngạc nhiên, tức giận, phẫn nộ, sợ hãi, được thu nhận trên 10 người mẫu nữ Nhật Bản. Cơ sở dữ 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 bằng cảm quan của 60 sinh viên Nhật Bản với mức cao nhất là 5, thấp nhất là 1.



Hình 3.9 Ví dụ dữ liệu mô hình mặt 3D

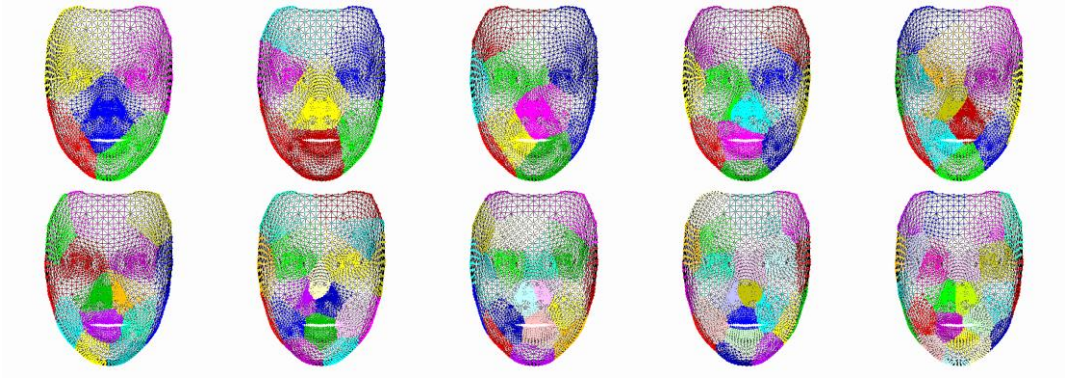
Dựa trên tập dữ liệu như vậy, việc thử nghiệm được tiến hành theo 3 trường hợp khác nhau: biến đổi giữa những người khác nhau với các biểu cảm khác nhau, biến đổi giữa những người khác nhau trên cùng biểu cảm trung tính và biến đổi của một người với các biểu cảm khác nhau.

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.



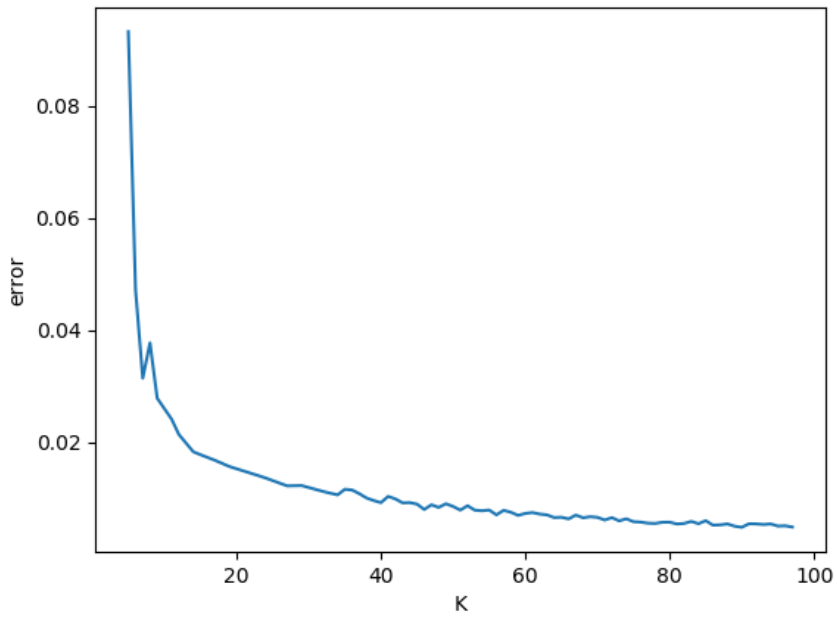
Hình 3.10 Một số mẫu biến thể của đối tượng

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.



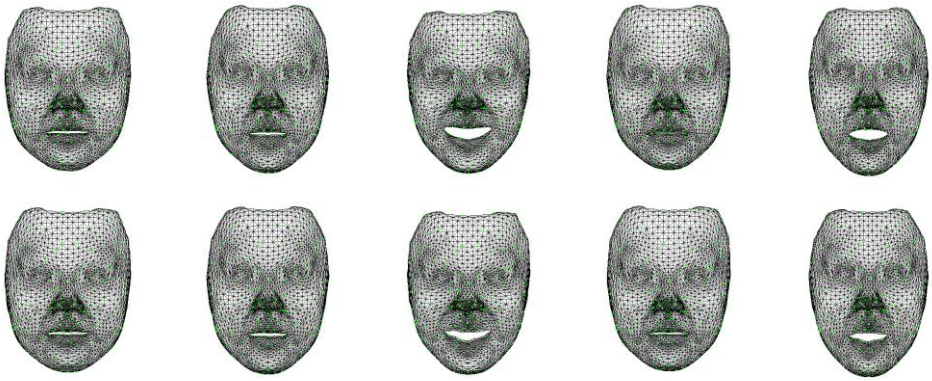
Hình 3.11 Một số kết quả gom cụm với nhiều mức K (Với mỗi hình, các điểm cùng một cụm được tô cùng màu)

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.



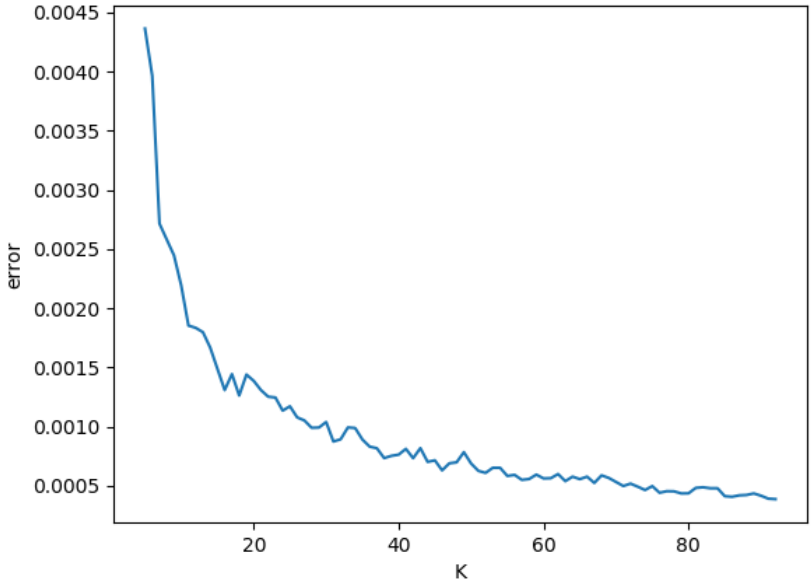
Hình 3.12 Biểu đồ tương quan giữa sai số với các giá trị K

Dễ 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.

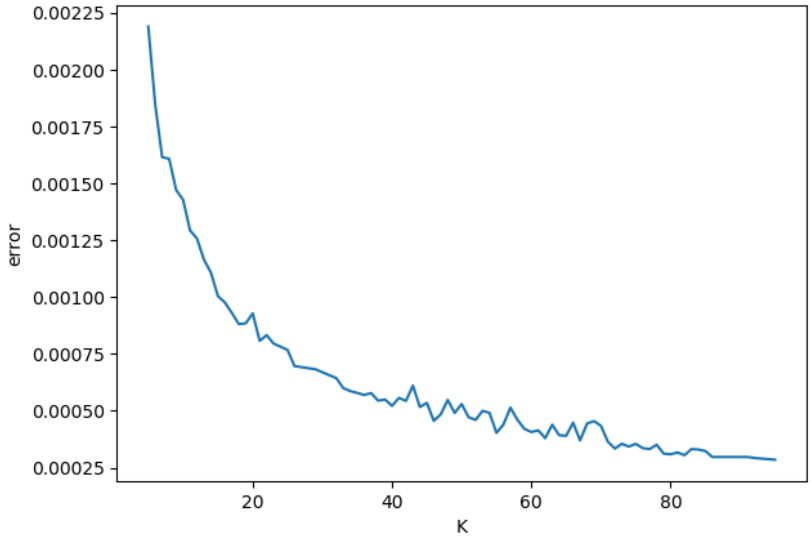


Hình 3.13 Một số hình ảnh kết quả nắn chỉnh: hàng 1 là các mô hình mục tiêu, hàng 2 là tương ứng các mô hình được nắn chỉnh từ mô hình 3D tham chiếu

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 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 trong trường hợp này hướng đến tiêu chí biến đổi các trạng thái biểu cảm trên một người cụ thể.



Hình 3.14 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 3.15 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

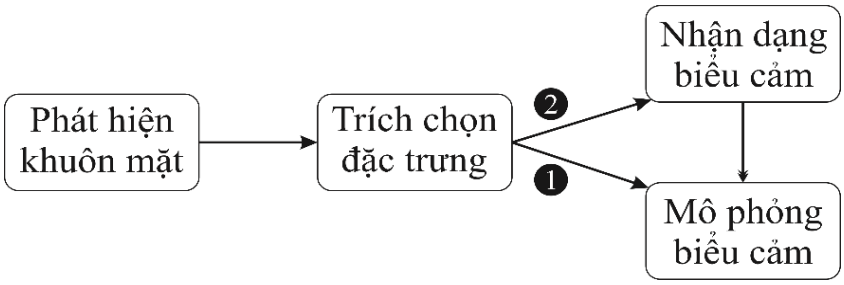
Ta dễ dàng nhận thấy, trong cả 3 trường hợp thử nghiệm, khi giá trị k tăng lên, sai số đều dần tiệm cận 0, điều này cũng rất rõ ràng với ý tưởng của việc biến đổi mô hình dựa trên tập điểm điều khiển. Tuy nhiên, sai số trong trường hợp biến đổi giữa những người khác nhau trên cùng biểu cảm trung tính 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.

3.5 Đề xuất xây dựng hệ thống mô phỏng biểu cảm khuôn mặt 3D

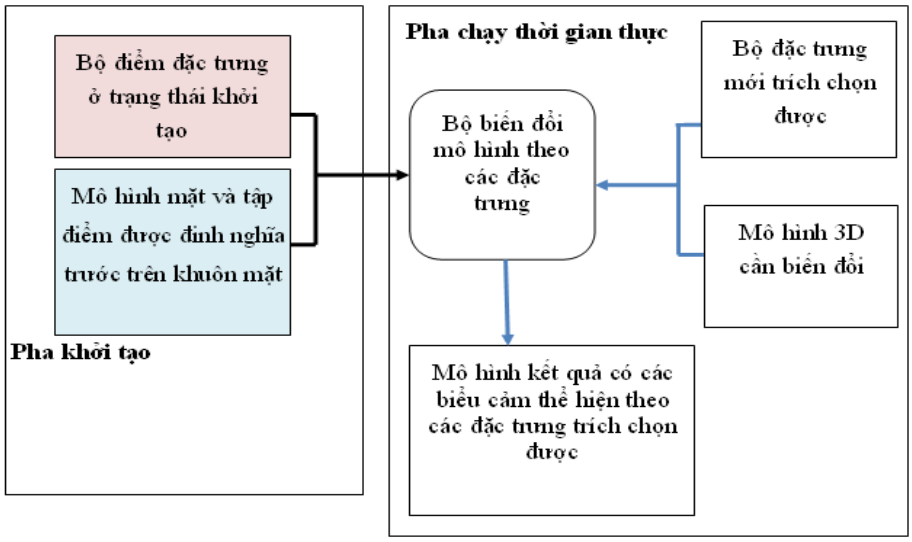
Một hệ thống mô phỏng biểu cảm có thể sẽ có rất nhiều cách thức xây dựng khác nhau tùy vào mục đích của ứng dụng. Có những tiếp cận bằng cách ước lượng một tập hữu hạn các biểu cảm rồi tiến hành mô phỏng tương ứng loại biểu cảm và cường độ nếu có. Có hướng tiếp cận bằng cách rút trích ra một số điểm đánh dấu đặc trưng của khuôn mặt và biến đổi mô hình theo các điểm đặc trưng đó mà không cần quan tâm đến ý nghĩa của những trạng thái cảm xúc đó.

Xét trên vấn đề trọng tâm là mô phỏng biểu cảm khuôn mặt phục vụ bài toán xây dựng mô hình mô phỏng cử chỉ khuôn mặt 3D, nghiên cứu sinh đề xuất hệ thống mô phỏng biểu cảm khuôn mặt được hình dung một cách đơn giản như sau:



Hình 3.16 Mô hình hệ thống mô phỏng biểu cảm khuôn mặt

Từ ảnh đầu vào, vị trí khuôn mặt được định vị để tiến hành xác định một tập điểm điều khiển. Khi đã có tập điểm điều khiển, các hệ thống sẽ thực hiện ước lượng ra các biểu cảm khuôn mặt đã được định nghĩa trước. Từ đó, tiến hành mô phỏng biểu cảm theo kịch bản dựng sẵn hoặc trực tiếp ánh xạ mô phỏng lại các biểu cảm từ những điểm trong tập điểm điều khiển đã thu được. Về cơ bản sơ đồ hệ thống có thể biểu diễn như sau:



Hình 3.17 Sơ đồ hệ thống

Theo như những phân tích đã nêu, ta có thể hình dung quy trình với một vài bước như sau đây. Vì khả năng biến đổi của mô hình ba chiều có thể sẽ rất khác biệt với những chuyển động mà ta ghi nhận được từ camera do vậy cần xây dựng ánh xạ tương ứng giữa các tập điểm trên mô hình ba chiều với tập điểm thu nhận được từ camera. Bước kế tiếp là biến đổi mô hình ba chiều với dữ liệu chuyển động đã được ước lượng từ tập điểm đặc trưng đã được tính toán bởi quá trình đề cập bên trên. Giả sử mô hình ba chiều gồm một tập rất lớn các điểm, một cách tiếp cận thường được dùng là coi chuyển động của mỗi điểm thành phần có thể được ước lượng từ một tập giới hạn các điểm điều khiển, đây là cơ sở để biến đổi toàn bộ hình dạng của khuôn mặt. Từ đó, quá trình mô phỏng biểu cảm khuôn mặt 3D có thể hình dung về mặt kỹ thuật tương ứng là một phương pháp để ánh xạ sự biến đổi các nét biểu cảm trên khuôn mặt từ một tập các biến đổi của tập điểm đặc trưng 3 chiều hoặc một tập các biến đổi của tập điểm đặc trưng 2 chiều trên khuôn mặt thu được từ luồng video 2D đến một mô hình 3D khuôn mặt.

3.6 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 trưng 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 nghiên cứu sinh 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, nghiên cứu sinh 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 quả nghiên cứu đã được công bố trong các công trình TCQT2, TCQT3, HNTN1

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. Sau đó, đề xuất mô hình mô phỏng biểu cảm khuôn mặt 3D theo 2 pha là : Pha thứ nhất là rút trích cử chỉ thông qua tập điểm điều khiển rồi ánh xạ trực tiếp lên mô hình khuôn mặt 3D; Pha thứ hai là định vị tập điểm điều khiển rồi ước lượng biểu cảm sau đó mô phỏng lại biểu cảm.

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 nghiên cứu sinh nhận thấy độ chính xác đạt đến mức xấp xỉ 95% chỉ từ mức ngưỡng khoảng 0,14. Ngoài ra để tiện đánh giá, nghiên cứu sinh đã xây dựng lược đồ thống kê tỉ lệ chính xác theo ngưỡng chấp nhận. Kết quả nghiên cứu được công bố trong các công trình TCQT1, TCTN1

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. Kết quả nghiên cứu được công bố trong các công trình TCQT2, TCQT3, HNTN1.

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 ngõ.

(a) Ví dụ điển hình như khi một người miệng đang cười nhưng chân mài và mắt có yếu tố bẩm sinh là cụp 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 trãi 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, nghiên cứu sinh 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 chi 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 ngọn 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 nghiên, 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. **Huynh 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, **Huynh 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. **Huynh Cao Tuan**, Do Nang Toan, Lam Thanh Hien, "Automatic selection of key points for 3D-face deformation", *Journal of Informatics and Mathematical Sciences, (Mới gửi tháng 12-2020)*

TRONG NƯỚC

TCTN1. **Huỳnh Cao Tuấn**, Đỗ Năng Toàn, Nguyễn Thanh Bình, Lâm Thành Hiển (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 Hiển, **Huỳnh Cao Tuấn**, Nguyễn Trọng Vinh (2014), "Phân đoạn video dựa trên kĩ 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 Hiển, (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*.

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 KHCN 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*.

TÀI LIỆU THAM KHẢO

TÀI LIỆU THAM KHẢO TIẾNG VIỆT

[1] Trịnh Hiền Anh , Đỗ Năng Toàn (2017), Công nghệ thực tại ảo và những bước đi đầu tiên của Việt Nam. *Tạp chí Khoa học và Công nghệ*, 7, tr. 28-30.

[2] Phạm Thế Bảo, Nguyễn Thành Nhựt, Cao Minh Thịnh, Trần Anh Tuấn, Phan Phúc Doãn (2007), Tổng quan các phương pháp xác định khuôn mặt người. *Tạp chí Công nghệ thông tin & Truyền thông*.

[3] Ma Thị Châu, Nguyễn Đình Tư, Đinh Quang Huy (2011), Tái tạo khuôn mặt 3 chiều từ hộp sọ. *Tạp chí Khoa học ĐHQGHN Khoa học Tự nhiên và Công nghệ*, 27, tr. 213-221.

[4] Nguyễn Văn Huân, Trịnh Xuân Hùng, Phạm Bá Mấy, Trần Ngọc Thái (2009), Cải tiến kỹ thuật biểu diễn bề mặt NURBS. *Kỷ yếu Hội thảo Quốc Gia "Một số vấn đề chọn lọc của CNTT và Truyền thông*, tr. 202-213.

[5] Đỗ Năng Toàn, Nguyễn Văn Huân (2010), Một kỹ thuật tăng tốc biểu diễn tóc. *Tạp chí Tin học và Điều khiển học*, 26(4), tr. 332-340.

[6] Hà Mạnh Toàn, Đỗ Năng Toàn, Trịnh Hiền Anh (2017), Một kỹ thuật chèn đối tượng ảo trong quảng cáo trực tuyến. *Kỷ yếu Hội nghị quốc gia Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin – FAIR*, tr. 511-515.

[7] Đỗ Năng Toàn, Nguyễn Văn Huân (2010), Một thuật toán rút gọn bề mặt biểu diễn mô hình 3D. *Tạp chí Khao học và Công nghệ-Viện Khoa học và Công nghệ Việt Nam*, 48, tr. 123-133.

TÀI LIỆU THAM KHẢO TIẾNG ANH

[8] Abdur Rahim, Najmul Hossain, Tanzillah Wahid & Shafiul Azam (2013), Face Recognition using Local Binary Patterns. *Global Journal of Computer Science and Technology*, Vol. 13.

[9] Aina OO, Zhang JJ (2010), Automatic muscle generation for physically-based facial animation. *Proceedingsof the ACM SIGGRAPH 2010 Posters. Los Angeles*, pp. 105.

[10] Albrecht I. (2005), Faces and Hands- Modeling and animating anatomical and photorealistic models with regard to the communicative competence of virtual humans. *PhD thesis, University at des Saarlandes*.

[11] Arivazhagan, S., Deivalakshmi, S., Kannan, K., Gajbhiye, B. N., Muralidhar, C., Lukose, N., Sijo N., Subramanian, M. P. (2007), Multi-resolution system for artifact removal and edge enhancement in computerized tomography images. *Pattern Recognition Letters 28(13)*, pp. 1769-1780.

[12] Assari M.A., Rahmati M. (2011), Driver drowsiness detection using face expression recognition. *In Proceedings of the IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, Malaysia*, pp. 337–341.

[13] Azuma, Ronald T. (1997), A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6.4, pp. 355-385.

[14] Bai Y., Zhang Y., Ding M., and Ghanem B. (2018), Finding tiny faces in the wild with generative adversarial network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21-30

[15] Baltrusaitis T., McDuff D., Banda N., Mahmoud M., Kaliouby R. El, Robinson P. and Picard R. (2011), Real-time inference of mental states from facial expressions and upper body gestures. *Proc. IEEE Int. Conf. Autom. Face Gesture Anal*, pp. 909-914.

[16] Bartlett M.S., Littlewort G., Fasel I., Movellan J.R (2003), Real Time Face Detection and Facial Expression ecognition: Development and Applications to Human Computer Interaction. *In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop, Madison, WI, USA*, 5, pp. 53–53.

[17] Bekele E., Zheng Z., Swanson A., Crittendon J., Warren Z. Sarkar (2013), Understanding how adolescents with autism respond to facial expressions in virtual reality environments. *IEEE Trans. Vis. Comput. Graphics*, 19, pp. 711–720.

[18] Bhanu B, Boughida A, Kouahla M.N, Lafifi Y (2019), A novel approach for facial expression recognition. *Pattern Recognit. Lett*, 27, pp. 1289–1298.

[19] Bickel B, Lang M, Botsch M, Otaduy MA, Gross M (2008), Pose-space animation and transfer of facial details. *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 57-66.

[20] Blu T., Thevenaz P., Unser M. (2004), Linear interpolation revitalized. *IEEE Transactions on Image Processing*, 13, pp. 710 - 719.

[21] Brown, R. G., Hwang, Y. C. (1996), Introduction to Random Signals and Applied Kalman Filtering. *3 ed. New York: John Wiley & Sons*.

[22] Buhmann MD (2003), Radial basis functions: theory and implementations. *Cambridge University*.

[23] Bui T. D., Heylen D., Poel M., and Nijholt A. (2002), Parlee: An adaptive planbased event appraisal model of emotions. *In KI 2002: Advances in Artificial Intelligence*, pp. 129–143.

[24] Chang W.Y., Hsu S.H., Chien J.H. (2017), FATAUVA-Net: An integrated deep learning framework for facial attribute recognition, action unit detection, and valence-arousal estimation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Honolulu, HI, USA*, pp. 17–25.

[25] Chen C.H., Lee I.J., Lin L.Y. (2015), Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. *Res. Dev. Disabil*, 36, pp. 396–403.

[26] Chew S., Lucey P., Lucey S., Saragih J., Cohn J., and Sridharan S. (2011), Person-independent facial expression detection using constrained local models. *Proc. IEEE Facial Expression Recognition and Analysis Challenge*, pp. 915-920.

[27] Chi, Nguyen Ngoc Khanh (2011), Active shape models their training and application. *International University HCMC, Vietnam*.

[28] Chuang E, Bregler C (2002), Performance driven facial animation using blendshape interpolation. *Stanford University*.

[29] Cindykia G.B, Putri W, Arief S.M (2018), Face Detection Using Haar Cascade in Difference Illumination. *IEEE: International Seminar on Application for Technology of Information and Communication*, pp: 4

[30] Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., Taylor J.G. (2001), Emotion recognition in human-computer interaction. *IEEE Signal Process. Mag*, 18, pp. 32–80.

[31] Dai J. Li Y. He K. and Sun J. (2016), R-FCN: object detection via region based fully convolutional networks. *Advances in Neural Information Processing Systems 29*.

[32] Darwin, C. (1872), The expression of the emotions in man and animals. *London: Murray*.

[33] Datcu, D., Rothkrantz, L. J. M. (2007), Facial Expression Recognition in still pictures and videos using Active Appearance Models. *A comparison approach*. pp: 1-6

[34] Dias L., Bhosekar A., Ierapetritou M. (2019), Adaptive Sampling Approaches for Surrogate-Based Optimization. *Proceedings of the 9th International Conference on Foundations of Computer-Aided Process Design*, pp. 377–384.

[35] Driscoll TA, Fornberg B (2002), Interpolation in the limit of increasingly flat radial basis functions. *Comput Math Appl 43*, pp. 413–422.

[36] Ekman P., Friesen W. V. (1975), Unmasking the Face: A Guide To Recognizing Emotions From Facial Clues. *Prentice-Hall, Englewood Cliffs*.

[37] Ekman, P., Friesen, W. V. (1976), Measuring facial movement. *Environmental Psychology and Nonverbal Behavior, 1*, pp. 56-75.

[38] El-Nasr M. S., J. Y., and Ioerger T. R. (2000), FLAME-fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems*, 3(3), pp. 219–257.

[39] Essa, I., Pentland, A. (1998), Coding, Analysis, Interpretation, and Recognition of Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp. 757-763.

[40] Fasel, B., Luettin, J. (2003), Automatic Facial Expression Analysis: A Survey. *Pattern Recognition, 36(1)*, pp. 259-275.

[41] Fornberg B, Piret C (2007), A stable algorithm for flat radial basis functions on a sphere. *SIAM J Sci Comput 30*, pp. 60–80.

[42] Fratarcangeli M (2012), Position, Based facial animation synthesis.. *Computer Animation and Virtual Worlds*, 23(3-4), pp. 457-466.

[43] G. Donato, et al. (1999), Classifying facial actions. *IEEE TPAMI, vol. 21, no. 10*, pp. 974-989.

[44] Ghimire D., Lee J., Li Z. N., Jeong S., Park S. H., and Choi H. S. (2015), Recognition of Facial Expressions Based on Tracking and Selection of Discriminative Geometric Features. *International Journal of Multimedia and Ubiquitous Engineering*, 10(3), pp. 35-44.

[45] Gnanaprakasam C, Sumathi S, Malini RR (2010), Average-half-face in 2D and 3D using wavelets for face recognition. *In: Niola V, Quartieri J, Neri F, et al., editors. Proceedings of the 9th WSEAS International Conference on Signal Processing, SIP'10; 2010 May 2931. Stevens Point, WI: World Scientific and Engineering Academy and Society (WSEAS)*, pp. 107–112.

[46] Gratch J. (2000), Émile: Marshalling passions in training and education. *In M. Gini C. Sierra and J. S. Rosenschein, editors, Proceedings of the Fourth International Conference on Autonomous Agents*, pp. 325–332.

[47] Guggisberg A.G., Mathis J., Schnider A., Hess C.W. (2010), Why do we yawn?. *Neuroscience & Biobehavioral Reviews, 34*, pp. 1267-1276.

[48] Han C.C., Liao H.Y.M., Yu G.J., Chen L.H. (2000), Fast face detection via morphology-based pre-processing. *Pattern Recognition*, 33, pp. 1701-1712.

[49] Han D. (2013), Comparison of Commonly Used Image Interpolation Methods. *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*.

[50] Hardy RL (1971), Multiquadric equations of topograpy and other irregular surfaces. *J Geophy Res 76*, pp. 1905–1915.

[51] Harguess J, Gupta S, Aggarwal J (2008), 3D face recognition with the average-half-face. *19th IEEE International Conference on Pattern Recognition*, pp. 1–4.

[52] Henry S., Benjamin K., Matthias N., Marc S. (2014), Local Painting and Deformation of Meshes on the GPU. *Computer Graphics Forum 2014*.

[53] Hien L.T., Toan D.N. (2016), An algorithm to detect driver's drowsiness based on nodding behavior. *International Journal of Soft Computing, Mathematics and Control*, 5(1), pp. 1-8.

[54] Hien L.T., Toan D.N., Lang T.V. (2015), Detection of human head direction based on facial normal algorithm. *International Journal of Electronics Communication and Computer Engineering*, 6(1), pp. 110-114.

[55] Hien L.T., Toan D.N., Toan H.M. (2015), Detecting Human Face with Ridge-Valley-Normal Model. *International Journal of Computer Science and Software Engineering (IJCSSE)*, 4, pp. 107-113.

[56] Hu P., Ramanan D. (2017), Finding tiny faces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 951-959.

[57] Huan V. N. , Toan N. D. (2010), Vector Fields in Expressing Hairstyles. *Proceedings of the 2nd IEEE International Conference on Advanced Computer Control- IACC 2010, Shenyang 27-29/03/2010, China*, pp. 514-546.

[58] Huang H, Chai J, Tong X, Wu HT (2011), Leveraging motion capture and 3D scanning for high-fidelity facial performance acquisition. *ACM Transactions on Graphics (TOG)*, 30(4), pp. 1-10.

[59] Iqtait M , Mohamad F S, Mamat M (2017), Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 332, pp. 681–685.

[60] Jabon M., Bailenson J., Pontikakis E., Takayama L., Nass C. (2011), Facial expression analysis for predicting unsafe driving behavior. *IEEE Perv. Comput*, 10, pp. 84–95.

[61] Jerritta S.; Murugappan M.; Nagarajan R.; Wan K. (2011), Physiological signals based human emotion recognition: a review. *In Proceedings of the IEEE 7th International Colloquium on Signal Processing and its Applications, Penang, Malaysia*, pp. 410–415.

[62] John Wiley & Sons, S. Pandzic và R. Forchheimer (2002), MPEG-4 Facial Animation. *The standard, implementations and applications*.

[63] Joshi P, Tien WC, Desbrun M, Pighin F (2003), Learning Controls for Blend Shape Based Realistic Facial Animation. *Proceedings of the Eurographics/SIGGRAPH Symposium on Computer Animation*, pp. 187-192.

[64] Kass, M., Witkin, A., Terzopoulos, D. (1987), Snakes: Active Contour Models.. *1st International Conference On Computer Vision*, pp. 259-268.

[65] Keltner D. ; Ekman P. (2000), Facial expression of emotion. *Handbook of Emotions, Guilford Press*, pp. 236-249.

[66] Ki-Yeong P., Sun-Young H. (2014), An improved Haar-like feature for efficient object detection, *Pattern Recognition Letters.* 42(1):148–153.

[67] Kumar, V. P., Poggio T. (2002), Recognizing Expressions by Direct Estimation of the Parameters of a Pixel Morphable Model. *Biologically Motivated Computer Vision*, pp. 519-527.

[68] Lakshmi S.V, M. R. Narasinga. R.D (2018), Deformable Facial Fitting Using Active Appearance Model for Emotion Recognition. *Smart Intelligent Computing and Applications*, pp. 135-144.

[69] Lee, J. S. (1983), Digital image smoothing and the sigma filter. *Computer Vision, Graphics and Image Processing*, 24, pp. 255--269.

[70] Lewis JP, Anjyo K (2010), Direct manipulation blendshapes. *Computer Graphics and Applications*, 30(4), pp. 42-50.

[71] Li G., Chung W.Y. (2013), Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. *Sensors*, 13(12), pp. 16494-16511.

[72] Lin T., Dollar P., Girshick R., He K., Hariharan B., and Belongie S. (2017), Feature pyramid networks for object detection. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[73] Lisetti C. L. (1999), Emotion generation for artificial agents via a hybrid architecture. *In Proceedings of the Autonomous Agents Workshop on EmotionBased Agent Architectures (EBAA'99)*.

[74] Zuoxin Li, Fuqiang Zhou. (2018), FSSD: Feature Fusion Single Shot Multibox Detector. *Computer Vision and Pattern Recognition*.

[75] Liu W. ,Anguelov D. ,Erhan D. ,Szegedy C. ,Reed S., Fu C. , and Berg A., (2016), Single shot multibox detector. *European Conference on Computer Vision, pp 21-37*.

[76] Liu X., Xia S., Fan Y., Wang Z. (2011), Exploring Non-Linear Relationship of Blendshape Facial Animation. *Computer Graphics Forum*, 30(6), pp. 1655-1666.

[77] Liu, L., Sclaroff, S. (1997), Color Region Grouping and Shape Recognition with Deformable Models.. *Boston University Computer Science Technical Report*, pp. 97-019.

[78] Liu, Zhengzhe, Xiaojuan Qi, and Philip HS Torr. (2020), Global texture enhancement for fake face detection in the wild. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[79] Mahardika C., Itimad R., Ahmad H. and Nadz S. (2013), Eye, lip and crying expression for virtual human. *International Journal of Interactive Digital Media*, 1(2).

[80] Manusov V. (2016), Facial Expressions. *The International Encyclopedia of Interpersonal Communication*.

[81] Martinkauppi, B. (2002), Face colour under varying illumination - analysis and applications. *University of Oulu repository*.

[82] Matsumoto D., Sung Hwang H. (2011), Reading facial expressions of emotion. *Psychological Science Agenda*.

[83] Mattela, G., Gupta, S.K. (2018), Facial Expression Recognition Using Gabor-Mean-DWT Feature Extraction Technique. *In Proceedings of the 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 22–23 Februar*, pp. 575–580.

[84] Micchelli CA (1986), Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constr Approx 2*, pp. 11–22.

[85] Michael J. Lyons, Miyuki Kamachi, Jiro Gyoba (2020), Coding Facial Expressions with Gabor Wavelets (IVC Special Issue). *In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan*, pp. 200–205.

[86] Nagao, M. A., Matsuyama, T. M. (1979), Edge Preserving Smoothing.. *Computer Graphics and Image Processing*, 9, pp. 374-407.

[87] Ngo P.; Nicolas P.; Yukiko K.; Hugues T. (2014), Topology-Preserving Rigid Transformation of 2D Digital Images. *IEEE Transactions on Image Processing*, 23.

[88] Ojala, Timo, Matti Pietikainen, and David Harwood (1994), Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. *Proceedings of 12th international conference on pattern recognition.*, Vol. 1. IEEE.

[89] Pantic M.; Rothkrantz L.J.M. (2000), Automatic analysis of facial expressions: The state of the art. *IEEE TPAMI*, 22(12), pp. 1424-1445.

[90] Parke FI (1972), Computer generated animation of faces. *Proceedings of the ACM annual conference*, pp. 451-457.

[91] Pizer, S. M., Amburn, E. P., Austin, J. D. , Cromartie, R., Geselowitz, R., Greer, T., Romeny, B. T. H., Zimmerman, J. B. (1987), Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3), pp. 355-368.

[92] Raouzaiou A., Karpouzis K., and Kollias S. D. (2003), Online gaming and emotion representation. *In N. N. García, J. M. Martínez, and L. Salgado, editors, Volume 2849 of Lecture Notes in Computer Science,* pp. 298– 305.

[93] Reisfeld, D., Wolfson, H., Yeshurun, Y. (1995), Context Free Attentional Operators: The Generalized Symmetry Transform. *Computer Vision, Special Issue on Qualitative Vision*.

[94] Ren S. , He K. Girshick R., and Sun J. (2017), Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp: 1137 - 1149

[95] Renan Fonteles Albuquerque; Paulo D. L. de Oliveira; Arthur P. de S. Braga (2018), Adaptive Fuzzy Learning Vector Quantization (AFLVQ) for Time Series Classification. *Fuzzy Information Processing*, pp. 385-397.

[96] Rowley, H., Baluja, V., Kanade, T. (1998), Neural Network-Based Face Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1), pp. 23-38.

[97] Scassellati B. (1998), Eye finding via face detection for a foevated, active vision system. *Proceedings of 15th National Conference on Artificial Intelligence*, pp. 1146-1151.

[98] Schneiderman H., Kanade T. (1998), Probabilistic modeling of local appearance and spatial relationships for object detection. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 40-50.

[99] Schoenberg IJ (1938), Metric spaces and completely monotone functions. *Ann Math*, 39, pp. 811–841.

[100] Seo Y, Lewis J, Seo J, Anjyo K, Noh J. (2012), Spacetime expression cloning for blendshapes. *ACM Transactions on Graphics*, 31(2), pp. 1-12.

[101] Shaif C.; Soummyo P.C; Tapan K.H (2017), Vehicle detection and counting using haar feature-based classifier. *IEEE: Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, 16-18 Aug.

[102] Shishir Bashyal; Ganesh K.Venayagamoorthy (2008), Recognition of facial expressions using Gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence*, 21, pp. 1056-1064.

[103] Sterring, M., Andersen, H. J., and Granum, E. (1999), Skin Color Detection Under Changing Lighting Condition. *Symposium on the Intelligent Robotics Systems*, pp. 187-195.

[104] Su H., Zheng G. (2008), A partial least squares regression-based fusion model for predicting the trend in drowsiness. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 38(5), pp. 1085-1092.

[105] Tayib S. Jamaludin Z. (2015), An Algorithm to Define Emotions Based on Facial Gestures as Automated Input in Survey Instrument. *American Scientific Publishers Advanced Science Letters*.

[106] Tian Y.I., Kanade T., Cohn J.F (2001), Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell*, 23, pp. 97–115.

[107] Toan H.M., Toan D.N., Hien L.T., Lang T.V. (2012), Modeling the human face and its application for detection of driver drowsiness. *International Journal of Computer Science and Telecommunications*, 3(11), pp. 56-59.

[108] Tomita, F. A., Tsuji, S. (1997), Extraction of Multiple regions by smoothing in selected neighborhoods. *IEEE Trns. Systems, Man and Cybernetics SMC-7*, pp. 107-109.

[109] Tong-Yee L, Po-Hua H. (2003), Fast and intuitive metamorphosis of 3D polyhedral models using SMCC mesh merging scheme. *Visualization and Computer Graphics, IEEE Transactions*, 9(01), pp. 85 – 98.

[110] Torre, F. D. L., Campoy, J., Cohn, J. F., Kanade, T. (2007), Simultaneous registration and clustering for temporal segmentation of facial gestures from video. *Proceedings of the Second International Conference on Computer Vision Theory and Applications*, 2, pp. 110-115.

[111] Turkan M., Pardas M., Enis Cetin A. (2008), Edge projections for eye localization. *Optical Engineering*, 47(4).

[112] Valstar M., Jiang B., Méhu M., Pantic M., and Scherer K. (2011), The first facial expression recognition and analysis challenge. *IEEE Int'l. Conf. Face and Gesture Recognition*.

[113] Velásquez J. D. (1997), Modeling emotions and other motivations in synthetic agents. *In Proceedings of the 14th National Conference on Artificial Intelligence and 9th Innovative Applications of Artificial Intelligence Conference (AAAI-97/IAAI-97)*, pp. 10–15.

[114] Viola, Paul, and Michael J. (2001), Rapid object detection using a boosted cascade of simple features.. *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition.*, Vol. 1. Ieee..

[115] Wang H. ,Li Z. ,Ji X., and Wang Y. (2017), Face r-cnn. *Computer Vision and Pattern Recognition (cs.CV)*.

[116] Wang Y., Ji X., Zhou Z. , Wang H., and Li Z. (2017), Detecting faces using region-based fully convolutional networks. *Computer Vision and Pattern Recognition (cs.CV)*.

[117] Wang Z., Ierapetritou M. (2017), A novel feasibility analysis method for black-box processes using a radial basis function adaptive sampling approach. *AIChE Journal*, 63, pp. 532-550.

[118] Wang, P., Barrett, F., Martin, E., Milonova, M., Gur, R. E., Gur, R. C., Kohler, C., Verma, R (2008), Automated video-based facial expression analysis of neuropsychiatric disorders. *Journal of Neuroscience Methods*, 168(1), pp. 224–238.

[119] Wang, Y., Lucey, S., Cohn, J. (2007), Non-Rigid Object Alignment with a Mismatch Template Based on Exhaustive Local Search. *IEEE Workshop on Non-rigid Registration and Tracking through Learning*.

[120] Wiener, N. (1949), Extrapolation, Interpolation, and Smoothing of Stationary Time Series. *New York: Wiley*.

[121] Wilson CA, Alexander O, Tunwattanapong B, Peers P, Ghosh A, Busch J, Hartholt A, Debevec P (2011), Facial cartography: interactive high-resolution scan correspondence. *Conference: International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2011, Vancouver, BC, Canada, August 7-11, 2011*, pp. 205-214.

[122] Wu T., Butko N. , Ruvolo P. , Whitehill J. , Bartlett M. and Movellan J. (2011), Action unit recognition transfer across datasets. *Proc. IEEE Int. Conf. Autom. Face Gesture Anal*, pp. 889-896.

[123] Wu Y., Ai X. (2008), Face detection in color images using Adaboost algorithm based on skin color information. *Proceedings of the First International Workshop on Knowledge Discovery and Data Mining*, pp. 339-342.

[124] Yang S. ,Xiong Y. ,Loy C. C. , and TangX. (2017), Face detection through scale-friendly deep convolutional networks. *Computer Vision and Pattern Recognition (cs.CV)*.

[125] Yano K., Harada K. (2009), A facial expression parameterization by elastic surface model. *International Journal of Computer Games Technology*, 2009, pp. 1-11.

[126] Yong Joo K.,et al (2005), Yong Joo Kim et al (2005), "3D warp brush: interactive free-form modeling on the responsive workbench. *Proceedings Virtual Reality IEEE*, pp. 279–280.

[127] Yoshitomi, Y., Kim, S., Kawano, T., and Kitazoe, T. (2000), Effect of Sensor Fusion for Recognition of Emotional States Using Voice, Face Image and Thermal Image of Face.. *presented at IEEE International Workshop on Robot and Human Interactive Communication*.

[128] You L., Southern R., Zhang J. (2009), Adaptive physics–inspired facial animation. *Motion in Games*, 5884, pp. 207-218.

[129] Yu J.,Jiang Y.,Wang Z.,Cao Z., and Huang T. S. (2016), Unitbox: An advanced object detection network. *Proceedings of the 24th ACM international conference on Multimedia*, pp: 516–520

[130] Zhang C., Xu X., and Tu D. (2018), Face detection using improved faster RCNN. *Computer Vision and Pattern Recognition (cs.CV)*.

[131] Zhang S. ,Zhu X. ,Lei Z. , Shi H., Wang X., and Li S. (2017), Single shot scale-invariant face detector. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 192-201

[132] Zhang S., Wu Z., Meng H.M., and Cai L. (2010), Facial expression synthesis based on emotion dimensions for affective talking avatar. *In Modeling Machine Emotions for Realizing Intelligence, SIST*, pp. 109–132.

[133] Zhang S., Zhu X., Lei Z., Shi H., Wang X., and Li S. Z. (2017), A CPU real-time face detector with high accuracy. *IEEE International Joint Conference on Biometrics*..

[134] Zhang, Jialiang, et al. (2020), Feature agglomeration networks for single stage face detection. *Neurocomputing*, 380, pp. 180-189.

[135] Zuiderveld, K. (1994), Contrast limited adaptive histogram equalization. *Graphics gems IV, Academic Press Professional, Inc., San Diego, CA*.