

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC LẠC HỒNG**



TRẦN THANH PHƯƠNG

**MỘT SỐ KỸ THUẬT NHẬN DẠNG BIỂU HIỆN
KHUÔN MẶT PHỤC VỤ ĐÁNH GIÁ SỰ TẬP TRUNG
CỦA NGƯỜI HỌC**

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

Đồng Nai, Năm 2023

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC LẠC HỒNG**



TRẦN THANH PHƯƠNG

**MỘT SỐ KỸ THUẬT NHẬN DẠNG BIỂU HIỆN
KHUÔN MẶT PHỤC VỤ ĐÁNH GIÁ SỰ TẬP TRUNG
CỦA NGƯỜI HỌC**

Chuyên ngành : KHOA HỌC MÁY TÍNH

Mã số : 9480101

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

PGS.TS. ĐỖ NĂNG TOÀN

TS. LÂM THÀNH HIỂN

Đồng Nai, Năm 2023

LỜI CAM ĐOAN

Tôi tên là: Trần Thanh Phương

Sinh ngày: 13/11/1983

Nơi sinh: Đồng Nai

Là nghiên cứu sinh khóa 2015 thuộc chuyên ngành Khoa học máy tính của Trường Đại học Lạc Hồng.

Tôi xin cam đoan về tính chất trung thực của những kết quả nghiên cứu trong luận án tiến sĩ với tên “Một số kỹ thuật nhận dạng biểu hiện khuôn mặt phục vụ đánh giá sự tập trung của người học” được thực hiện tại Trường Đại học Lạc Hồng. Đây là sản phẩm với sự nỗ lực nghiên cứu của nghiên cứu sinh dưới sự hướng dẫn của Thầy PGS.TS. Đỗ Năng Toàn và Thầy TS. Lâm Thành Hiền. Luận án được thực hiện dựa trên nền tảng lý thuyết và khảo sát thực tế, hoàn toàn không sao chép nội dung từ bất kỳ những luận án, nguồn tài liệu hoặc công trình nghiên cứu khác.

Các phương pháp trình bày trong Luận án được phát triển dựa trên yêu cầu cụ thể của ứng dụng và kết quả nghiên cứu là trung thực, chưa được công bố dưới bất kỳ hình thức nào trước khi bảo vệ và công nhận bởi “Hội đồng đánh giá luận án Tiến sĩ” của Trường.

Nghiên cứu sinh

Trần Thanh Phương

LỜI CẢM ƠN

Tác giả xin gửi lời cảm ơn chân thành đến Thầy PGS.TS. Đỗ Năng Toàn – Viện Công nghệ thông tin và Thầy TS. Lâm Thành Hiền – Trường Đại học Lạc Hồng. Thầy đã nhiệt tình hướng dẫn, động viên tinh thần và giành nhiều thời gian quý báu để hỗ trợ tác giả trong quá trình tham dự khóa học và hoàn thành luận án.

Bên cạnh đó, tác giả xin gửi lời cảm ơn đến lãnh đạo Trường Đại học Lạc Hồng, lãnh đạo Trung tâm Thông tin Tư liệu và lãnh đạo Khoa Sau đại học, đã tạo điều kiện giúp đỡ tận tình tác giả trong quá trình hoàn thiện luận án.

Cuối cùng, tác giả xin cảm ơn gia đình đã luôn động viên tác giả trong quá trình học tập và hoàn thiện luận án.

Trân trọng,

Trần Thanh Phương

TÓM TẮT LUẬN ÁN

Phân tích học tập tự động đang trở thành một chủ đề quan trọng trong cộng đồng giáo dục, trong đó yêu cầu các hệ thống hiệu quả để giám sát quá trình học tập của người học và phản hồi những thông tin hữu ích kịp thời cho giáo viên. Những tiến bộ gần đây về cảm biến thị giác và kỹ thuật thị giác máy tính cho phép phân tích tự động hành vi và trạng thái cảm xúc của người học theo nhiều mức độ khác nhau. Các trạng thái cảm xúc của người học như thích thú, mệt mỏi, bối rối v.v. được xác định tự động từ những biểu hiện trên khuôn mặt và mức độ tập trung được tính toán từ các dấu hiệu thị giác khác nhau như nhìn khuôn mặt, trạng thái đóng/mở của mắt và tư thế hình thể. Hiểu được hành vi của người học giúp cơ sở quản lý giáo dục cải thiện về chính sách học vụ, chương trình đào tạo, môi trường học tập, trang thiết bị v.v. Bên cạnh đó, cũng giúp giáo viên cập nhật giáo trình, bài giảng và phương pháp sư phạm. Từ đó mang lại hiệu quả cho người học, đồng thời ngăn chặn sự sao lãng, chán nản cũng như tình trạng bỏ học.

Ý tưởng của Luận án: là sử dụng một camera để thu thập dữ liệu hành vi học tập của người học. Sau đó, sử dụng một số kỹ thuật trong xử lý ảnh để trích xuất và xử lý các khung hình trong video. Từ các khung hình này, bước tiếp theo Luận án lựa chọn và xác định các đặc điểm có ý nghĩa để đánh giá mức độ tập trung của người học một cách hiệu quả như trạng thái đóng/mở của mắt, biểu hiện trên khuôn mặt, tư thế hình thể v.v. thông qua các kỹ thuật dựa trên nền tảng thị giác máy tính.

Như đã trình bày ở trên, các đặc điểm có thể đánh giá mức độ tập trung của người học bao gồm: trạng thái đóng/mở của mắt, biểu hiện trên khuôn mặt, tư thế hình thể. Tuy nhiên, Luận án lựa chọn kỹ thuật đánh giá dựa trên trạng thái đóng/mở của mắt và các biểu hiện trên khuôn mặt để thực hiện. Lý do, hai kỹ thuật này mặc dù đã có nhiều công trình nghiên cứu tuy nhiên vẫn còn tồn tại những hạn chế nhất định khi triển khai vào môi trường thực tế. Do đó, Luận án góp phần giải quyết một phần trong những hạn chế này. Cụ thể, Luận án cải tiến kỹ thuật phát hiện trạng thái đóng/mở của mắt và phân rã biểu cảm của người học thành các thành phần cơ bản nhằm giúp cải thiện độ chính xác cho bài toán đánh giá. Mặt khác, vấn đề khan hiếm về bộ dữ liệu huấn luyện cảm xúc hỗn hợp của người học cũng là một yếu tố mà Luận án quan tâm nghiên cứu.

Kết quả Luận án đạt được là

(1) Công trình đánh giá mức độ tập trung của người học dựa trên trạng thái đóng/mở của mắt. Quy trình đánh giá được xem xét và phân tích trên cả quá trình học tập của người học. Kỹ thuật đánh giá trên từng thời điểm riêng lẻ, sau đó tổng hợp lại và tính tỷ lệ của mỗi trạng thái so với cả quá trình.

(2) Kỹ thuật phân rã biểu cảm thành phần cơ bản nhằm giải quyết vấn đề khan hiếm dữ liệu huấn luyện. Trên thực tế, trong quá trình học tập, người học luôn tiếp nhận những thông tin từ giáo viên, bạn bè và môi trường v.v., nên cảm xúc xảy ra ở mỗi thời điểm khác nhau, có thể là hỗn hợp hoặc đơn lẻ, mà dữ liệu huấn luyện bị hạn chế về cảm xúc hỗn hợp. Do đó, cần phân rã để nâng cao độ chính xác cho mô hình nhận dạng cũng như hiểu rõ hơn về chi tiết cảm xúc của người học.

(3) Kỹ thuật cải tiến phát hiện trạng thái đóng/mở của mắt. Kỹ thuật này phù hợp cho mọi đối tượng với ngưỡng được xác định linh động mà không phải phụ thuộc vào một ngưỡng cố định như các công trình trước đó đã công bố. Thực nghiệm cho thấy kỹ thuật cải tiến tỏ ra hiệu quả và phù hợp để giải quyết bài toán của luận án.

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
TÓM TẮT LUẬN ÁN	iii
MỤC LỤC.....	v
DANH MỤC TỪ VIẾT TẮT.....	viii
DANH MỤC HÌNH ẢNH	x
DANH MỤC BẢNG BIỂU	xii
CHƯƠNG 1 MỞ ĐẦU	1
1.1 Phát biểu vấn đề	1
1.2 Bối cảnh và thách thức	2
1.3 Mục tiêu nghiên cứu	4
1.4 Đối tượng, Phạm vi, Phương pháp	5
1.5 Đóng góp của luận án.....	8
1.6 Bố cục của luận án.....	8
CHƯƠNG 2 TỔNG QUAN VỀ ĐÁNH GIÁ SỰ TẬP TRUNG CỦA NGƯỜI HỌC VÀ BÀI TOÁN NHẬN DẠNG BIỂU CẢM KHUÔN MẶT	10
2.1 Tổng quan về đánh giá sự tập trung của người học	10
2.1.1 Ý nghĩa sự tập trung.....	10
2.1.2 Các yếu tố đánh giá sự tập trung.....	10
2.1.3 Phương pháp đánh giá sự tập trung	11
2.1.4 Hệ thống đánh giá sự tập trung của người học	14
2.2 Bài toán nhận dạng biểu cảm khuôn mặt	20
2.2.1 Biểu cảm khuôn mặt	21
2.2.2 Nhận dạng biểu cảm khuôn mặt.....	26
2.2.3 Một số cách tiếp cận trong đánh giá sự tập trung của người học dựa vào biểu cảm khuôn mặt.....	41
2.3 Một số vấn đề cơ sở lý thuyết	42
2.3.1 Nhận dạng mẫu	42

2.3.2	Các bước cơ bản trong xây dựng hệ nhận dạng.....	42
2.3.3	Phương pháp đánh giá.....	45
2.3.4	Kỹ thuật phát hiện khuôn mặt.....	46
2.3.5	Kỹ thuật xử lý ảnh số.....	48
2.3.6	Thuật toán GentleBoost.....	49
2.4	Bộ cơ sở dữ liệu dùng trong nghiên cứu.....	52
2.4.1	Dữ liệu về biểu cảm khuôn mặt.....	52
2.4.2	Dữ liệu phát hiện trạng thái đóng/mở mắt.....	54
2.5	Kết chương.....	55
CHƯƠNG 3 NHẬN DẠNG BIỂU CẢM THÀNH PHẦN KHUÔN MẶT ...		56
3.1	Đánh giá sự tập trung dựa trên biểu hiện trên khuôn mặt.....	56
3.1.1	Kiến trúc mô hình.....	56
3.1.2	Đánh giá thực nghiệm.....	58
3.1.3	Phân loại tập trung.....	61
3.2	Kỹ thuật phân rã biểu cảm thành phần cơ bản.....	62
3.2.1	Giới thiệu vấn đề.....	62
3.2.2	Phương pháp thực hiện.....	66
3.2.3	Thuật toán phân rã.....	67
3.2.4	Kết quả thực nghiệm.....	68
3.3	Kết chương.....	69
CHƯƠNG 4 NHẬN DẠNG BIỂU CẢM HÀNH VI KHUÔN MẶT		71
4.1	Phát hiện mắt người trong ảnh.....	71
4.2	Đánh giá sự tập trung dựa trên trạng thái mắt.....	72
4.2.1	Cơ sở lý thuyết của kỹ thuật đánh giá.....	72
4.2.2	Đánh giá sự tập trung dựa vào mức độ buồn ngủ.....	75
4.2.3	Thuật toán đánh giá sự tập trung.....	77
4.2.4	Độ phức tạp của thuật toán.....	79
4.2.5	Thử nghiệm.....	80
4.3	Kỹ thuật phát hiện trạng thái đóng/mở mắt.....	81
4.3.1	Giới thiệu.....	81

4.3.2	Cơ sở lý thuyết của phương pháp cần cải tiến	82
4.3.3	Cải tiến kỹ thuật phát hiện trạng thái đóng/mở mắt dựa vào biến thiên tỷ lệ khung hình.....	83
4.4	Kết chương	89
CHƯƠNG 5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....		90
DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ		92
TÀI LIỆU THAM KHẢO.....		93
PHỤ LỤC 1 HỆ MÃ HÓA CÁC CỬ ĐỘNG TRÊN KHUÔN MẶT.....		103
PHỤ LỤC 2 MÃ CHƯƠNG TRÌNH.....		108

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng việt
AAM	Active Appearance Model	Mô hình ngoại quan
AEMS	Automatic Engagement Management System	Hệ thống quản lý tương tác tự động
AF	Activate Function	Hàm kích hoạt ứng dụng cho mô hình mạng học sâu
AU	Action Unit	Đơn vị hành động
CLAHE	Contrast-limited adaptive histogram equalization	Cân bằng biểu đồ thích ứng có giới hạn tương phản
CNN	Convolution Neural Network	Mạng nơ-ron tích chập
DA	Data Augmentation	Tăng cường dữ liệu
DBSCAN	Density-Based Spatial Clustering of Applications with Noise	Phân cụm dựa trên mật độ không gian với các dữ liệu có nhiễu
DT	Decision Tree	Cây quyết định
EAR	Eye Aspect Ratio	Tỷ lệ hình dạng của mắt
ECG	Electro Encephalo Graphy	Điện tâm đồ
EEG	Electro Encephalo Graphy	Điện não đồ
FACS	Facial Action Coding System	Hệ thống mã hóa hành động trên khuôn mặt
FER	Facial Expression Recognition	Nhận dạng biểu cảm khuôn mặt
FERC-2013	Facial Expression Recognition Challenge	Cơ sở dữ liệu chuẩn được tổ chức cho cuộc thi “Thử thách nhận dạng cảm xúc trên khuôn mặt năm 2013”
FFT	Fast Fourier Transform	Biến đổi Fourier
HCI	Human Computer Interface	Giao diện người - máy tính
HE	Histogram Equalization	Biểu đồ cân bằng
HMM	Hidden Markov Model	Mô hình Markov ẩn

Từ viết tắt	Tiếng Anh	Tiếng việt
ITS	Intelligent Tutoring Systems	Hệ thống dạy kèm thông minh
JAFFE	Japanese Female Facial Expression	Cơ sở dữ liệu phục vụ cho việc nhận dạng biểu cảm khuôn mặt của phụ nữ Nhật bản
KNN	K-Nearest Neighbors	K-Láng giềng gần nhất
LBP	Local Binary Pattern	Mẫu nhị phân cục bộ
LMS	Learning Management Systems	Hệ thống quản lý học tập
ReLU	Rectified Linear Unit	Hàm đơn vị tuyến tính
RF	Random Forest	Rừng ngẫu nhiên
SoTA	State-of-The-Art	Đề cập đến các mô hình tiên tiến, hiện đại
SVM	Support Vector Machine	Máy vector tựa

DANH MỤC HÌNH ẢNH

Hình 2.1 Phân loại các phương pháp phát hiện sự tham gia [6]	11
Hình 2.2 Trạng thái biểu cảm khuôn mặt cơ bản.....	22
Hình 2.3 Một số ví dụ về AU [66]	23
Hình 2.4 Hình minh họa cường độ hành động.....	24
Hình 2.5 Sơ đồ hệ thống FER thông thường [77]	26
Hình 2.6 Biểu đồ cân bằng [77]	29
Hình 2.7 Ví dụ về tăng cường dữ liệu [77]	31
Hình 2.8 Biểu đồ đặc trưng của Gabor [77].....	32
Hình 2.9 Phương pháp mẫu nhị phân cục bộ [95]	32
Hình 2.10 Các ví dụ về LBP tròn. (a) $r = 1, n = 8$ (b) $r = 2, n = 16$ (c) $r = 2, n = 8$	33
Hình 2.11 Trích xuất biểu đồ LBP từ hình ảnh khuôn mặt [95]	34
Hình 2.12 Đối tượng hình dạng và kết cấu hình ảnh [22].....	35
Hình 2.13 Đặc trưng cơ bản Haar-like	39
Hình 2.14 Đặc trưng mở rộng của Haar-like	39
Hình 2.15 Tính tổng pixel trong vùng ảnh [23]	40
Hình 2.16 Kiến trúc hệ thống đánh mức độ tham gia của người học dựa trên tín hiệu EEG [46]	41
Hình 2.17 Kiến trúc hệ thống đánh giá mức độ tham gia dựa trên thị giác máy tính [101]	42
Hình 2.18 Phân bố số lượng cảm xúc trong tập dữ liệu FEREC-2013	52
Hình 3.1 Kiến trúc mô hình nhận dạng biểu hiện trên khuôn mặt.....	56
Hình 3.2 Một số cảm xúc được phát hiện từ chương trình	59
Hình 3.3 Kết quả huấn luyện từ chương trình.....	59
Hình 3.4 Một số hình ảnh minh họa cảm xúc hỗn hợp [105]	63
Hình 3.5 Biến dạng mô hình khuôn mặt của nhân vật trong dự án Sintel [106] .	63
Hình 3.6 Dữ liệu khuôn mặt trong nghiên cứu của nhóm V. Blanz [107].....	64
Hình 3.7 Quá trình tổng hợp mẫu khuôn mặt [112].....	65
Hình 3.8 Lưu đồ giải thuật phân rã	67

Hình 3.9 Mô hình khuôn mặt 3D trong tập dữ liệu JAFFE	68
Hình 3.10 Tổng hợp một số khuôn mặt ngẫu nhiên.....	69
Hình 3.11 Tiến trình xử lý.....	69
Hình 4.1 Dấu mốc trên khuôn mặt [24]	71
Hình 4.2 Sơ đồ luồng đánh giá sự tập trung.....	72
Hình 4.3 Thời gian mắt mở đến khi mắt đóng hoàn toàn	72
Hình 4.4 Minh họa thời gian mắt mở đến khi mắt đóng hoàn toàn	73
Hình 4.5 Tọa độ và tỷ lệ khung hình của mắt	74
Hình 4.6 Mô hình ước lượng mức độ buồn ngủ	75
Hình 4.7 Mô phỏng trạng thái của quá trình học tập	76
Hình 4.8 Lưu đồ đánh giá sự tập trung của người học.....	77
Hình 4.9 Lưu đồ tính toán thời gian tham gia học tập của người học	78
Hình 4.10 Kết quả ước lượng mức độ buồn ngủ của một sinh viên	80
Hình 4.11 Kết quả đánh giá sự tập trung của một sinh viên	80
Hình 4.12 Minh họa kết quả đánh giá sự tập trung của một sinh viên	81
Hình 4.13 Tọa độ và tỷ lệ khung hình của mắt	82
Hình 4.14 68 điểm đặc trưng trên khuôn mặt	83
Hình 4.15 Biến thiên tỷ lệ của khung hình trong hành vi đóng/mở mắt.....	84
Hình 4.16 Lưu đồ giải thuật phát hiện trạng thái đóng/mở mắt.....	85
Hình 4.17 Kết quả thực nghiệm của kỹ thuật cải tiến trên các tập dữ liệu Talking face, Eyeblink8, HUST_LEBW, ZJU	87
Hình 4.18 Biểu đồ so sánh kỹ thuật cải tiến so với công trình [16].....	88
Hình 4.19 Biểu đồ so sánh kỹ thuật cải tiến so với công trình [115].....	88
Hình 4.20 Biểu đồ so sánh kỹ thuật cải tiến so với công trình [116].....	88

DANH MỤC BẢNG BIỂU

Bảng 2.1 Các công trình nghiên cứu liên quan	14
Bảng 2.2 Ký hiệu diễn giải cường độ hành động.....	23
Bảng 2.3 Cảm xúc và danh mục AU tương ứng [39]	25
Bảng 2.4 Ma trận nhầm lẫn tổng quát.....	45
Bảng 2.5 Bảng phân bố cảm xúc trong tập dữ liệu mẫu FERC-2013.....	52
Bảng 2.6 Một số ảnh mẫu trong tập dữ liệu FERC-2013	53
Bảng 2.7 Một số ảnh mẫu trong tập dữ liệu JAFFE	53
Bảng 2.8 Một số ảnh mẫu trong tập dữ liệu Talking Face, Eyeblink8, HUST-LEBW, ZJU	55
Bảng 2.9 Thông tin về tập dữ liệu Talking Face, Eyeblink8, HUST-LEBW, ZJU	55
Bảng 3.1 Chi tiết mô hình nhận dạng cảm xúc trên khuôn mặt [17]	57
Bảng 3.2 Ma trận nhầm lẫn trên tập dữ liệu kiểm tra của FERC-2013	60
Bảng 3.3 Kết quả chạy thử nghiệm trên tập dữ liệu kiểm tra của FERC-2013 ...	61
Bảng 3.4 Cảm xúc và trọng số tương ứng [19] [20]	62
Bảng 4.1 Sự tương quan giữa mức độ buồn ngủ và hành vi [114]	73
Bảng 4.2 Luật xác định mức độ buồn ngủ	74
Bảng 4.3 Mức độ tập trung	75
Bảng 4.4 Độ phức tạp của kỹ thuật đánh giá sự tập trung của người học dựa trên trạng thái của mắt	79
Bảng 4.5 Độ phức tạp của thuật toán	85
Bảng 4.6 Kết quả thực nghiệm.....	86
Bảng 4.7 Kết quả so sánh với các giải pháp	87
Bảng 0.1 Hệ mã hóa các cử động trên khuôn mặt	103

CHƯƠNG 1

MỞ ĐẦU

1.1 Phát biểu vấn đề

Toàn cầu hóa và sự biến đổi công nghệ đã trở thành xu thế phát triển không ngừng trong nhiều năm qua, đã mở ra một nền kinh tế toàn cầu mới “lấy sức mạnh từ công nghệ, năng lượng từ thông tin và chèo lái bằng kiến thức”. Sự hội nhập của nền kinh tế toàn cầu mới này đã ngầm khẳng định một cách nghiêm túc về tính chất và vai trò của giáo dục. Vì thế, việc học đã trở nên vô cùng quan trọng, đây là chiếc cầu nối tri thức với mỗi người. Nền tảng tri thức đã đổi mới và biến đổi xã hội của chúng ta, đã thay đổi hoàn toàn cách mỗi người suy nghĩ, sống và làm việc. Ngoài ra, tri thức cũng tạo ra môi trường sống tốt đẹp và gắn kết mỗi người gần gũi với nhau hơn. Do đó, có thể khẳng định rằng việc học đóng vai trò quan trọng rất lớn đối với mỗi người, nó là nhân tố quyết định đến sự hòa nhập, tồn tại và phát triển của mỗi người trong xã hội.

Việc học tập của mỗi người cần có thời gian và nỗ lực bền bỉ để hoàn thành. Cùng với quan điểm đó, các nhà khoa học đã đưa ra tầm quan trọng của việc nỗ lực liên tục để hoàn thành nhiệm vụ học tập [1]. Có nhiều yếu tố ảnh hưởng đến kết quả học tập của người học như: môi trường giáo dục, chương trình học, giờ học, cơ sở hạ tầng, các vấn đề về tài chính [2] [3] và một vấn đề cực kỳ quan trọng khác đó chính là hành vi của người học [4] [5]. Do đó, trong bất kỳ hệ thống giáo dục nào, sự tham gia và tập trung của người học là chìa khóa để thành công trong học tập.

Để đánh giá mức độ tham gia của người học, theo phương pháp truyền thống, giáo viên thực hiện các cuộc khảo sát để thu thập thông tin một cách thủ công bằng cách phỏng vấn trực tiếp người học hoặc để người học tự báo cáo riêng mức độ của họ về sự thích thú hay nhàm chán thông qua biểu mẫu khảo sát được thiết lập sẵn [6]. Các phương thức này được can thiệp bởi người và chỉ phù hợp với quy mô lớp học nhỏ, vì tiêu tốn rất nhiều thời gian và công sức.

Ngày nay, công nghệ đã trở thành một công cụ thiết yếu để hỗ trợ người học và giáo viên với những trải nghiệm giáo dục hiệu quả hơn. Bên cạnh đó, hệ thống

quản lý học tập (Learning Management Systems - LMS) cũng ra đời và đóng một vai trò quan trọng trong các mô hình giảng dạy giáo dục đại học [7]. Điều này đã mở ra hướng nghiên cứu mới liên quan đến việc phân tích hành vi của người học trong hệ thống LMS, nhằm tìm kiếm các mẫu học tập để cải thiện phương pháp giảng dạy của giáo viên cũng như quá trình học tập cho người học.

Mặt khác, tác động mạnh mẽ của đại dịch COVID-19 đã gây ảnh hưởng đến các hoạt động trong cuộc sống của mỗi người trong đó có giáo dục. Điều này buộc các hệ thống giáo dục phải chuyển từ hình thức học truyền thống sang hình thức học trực tuyến [8] [9] [10]. Cho đến thời điểm hiện tại, tuy đại dịch đã được kiểm soát và không còn bùng phát mạnh, nhưng nhiều trường đại học vẫn đang kết hợp cả truyền thống và trực tuyến (Blending Learning). Điều này đã tạo ra sự phong phú không ngừng về dữ liệu dạy-học trong môi trường trực tuyến. Một vấn đề quan trọng trong học tập trực tuyến là sự tương tác của người học với các hoạt động giáo dục của họ, vì lo ngại về tỷ lệ bỏ học cao trong các khóa học trực tuyến [6]. Tuy nhiên, hầu hết các hệ thống học tập hiện tại không có các chỉ báo / bộ phát hiện mức độ tương tác [10]. Do đó, để hiểu rõ bản chất của dữ liệu, cụ thể là phương pháp giảng dạy và hành vi học tập thì việc phân tích và đánh giá sự tập trung của người học đã trở thành một trong những chỉ số quan trọng để đo lường chất lượng giảng dạy của giáo viên và hiệu suất của người học.

Ngoài ra, theo quan điểm của tác giả, một hệ thống đánh giá mức độ tập trung của người học có thể giúp nhà quản lý giáo dục hiểu rõ được hành vi của người học. Từ đó giúp họ cải thiện môi trường học tập, cơ sở vật chất, chương trình đào tạo và tạo điều kiện tốt cho người học để ngăn ngừa tình trạng như: chán nản, sao lãng hay là bỏ học. Mặt khác, thông qua hành vi của người học, giúp giáo viên cải tiến giáo trình, bài giảng và phương pháp giảng dạy nhằm nâng cao hiệu suất cho người học. Do đó, để góp phần nâng cao chất lượng dạy-học, tác giả chọn nghiên cứu đề tài “***Một số kỹ thuật nhận dạng biểu hiện khuôn mặt phục vụ đánh giá sự tập trung của người học***” để thực hiện luận án này.

1.2 Bối cảnh và thách thức

Vấn đề tìm kiếm các mẫu học tập để xác định mức độ tham gia của người học đang được quan tâm trong những năm gần đây. Một số cách tiếp cận và phép đo để đánh giá mức độ tham gia của người học được giới thiệu trong các tài liệu.

Tự báo cáo là cách tiếp cận truyền thống như: khảo sát và làm bài kiểm tra để theo dõi các hành động / hành vi của người học. Hai cách tiếp cận này không thuận tiện và thiếu khách quan vì người học có thể trả lời không trung thực trong các cuộc khảo sát và có thể bị can thiệp bởi người [11]. Mặt khác, phương pháp đánh giá này không bao quát trong mọi bối cảnh vì nó chỉ phù hợp trong không gian lớp học nhỏ, tiêu tốn nhiều thời gian, công sức. Đặc biệt các phương pháp này gặp nhiều khó khăn khi mở rộng.

Một thách thức khác đối với hệ thống giáo dục là từ khi đại dịch COVID-19 bùng phát, các hệ thống về quản lý học tập (LMS) và học trực tuyến đã trở thành một phần thiết yếu của hệ thống giáo dục ngày nay. Các khóa học và chương trình được hỗ trợ bởi công nghệ tiếp tục phát triển về quy mô [12]. Việc áp dụng học tập trực tuyến đã mang lại nhiều thuận lợi như: cho phép cung cấp các nguồn tài nguyên học tập vượt qua các rào cản về địa lý và thời gian, cung cấp đa phương tiện, tương tác phong phú và tạo điều kiện tốt cho người học [13]. Tuy nhiên, học trực tuyến đòi hỏi một số tiêu chí nhất định để thành công. Chẳng hạn như thái độ người học, ý thức tham gia tốt và giáo viên giám sát được tổng thể người học trong lớp học ảo. Điều này phù hợp với số lượng người học trong lớp học nhỏ, nhưng với số lượng lớn thì giáo viên rất khó giám sát tất cả người học trong cùng một giao diện màn hình. Do đó, hiểu được hành vi của người học là rất cần thiết trong bối cảnh này và là một thách thức đáng kể trong môi trường giáo dục hiện đại.

Với sự thành công của phương pháp thị giác máy tính, các phương pháp tự động đang được chú ý trong những năm gần đây và rất hữu ích trong việc phân tích hành vi của người học; các mô hình có thể đánh giá mức độ tập trung trong một số trường hợp dựa trên các video trực tiếp hoặc được quay trước về các hoạt động giáo dục. Các công trình tiêu biểu trong lĩnh vực này đã được công bố [5] [14] [15] [16] [4] [17] [18] và đã giải quyết được phần nào cho bài toán đánh giá mức độ tham gia của người học dựa vào các đặc trưng về mắt và biểu hiện trên khuôn mặt. Tuy nhiên, kết quả của các công trình này chỉ dừng ở mức trong phòng thí nghiệm. Để triển khai ở môi trường thực tế vẫn còn nhiều thách thức cần giải quyết. Chẳng hạn, công trình [18] đánh giá mức độ tập trung của người học dựa trên trạng thái đóng/mở của mắt, kỹ thuật này đang sử dụng một ngưỡng cố định để xác định trạng thái của mắt. Đây là một vấn đề cần được cải tiến vì đặc trưng

mắt của mỗi người là khác nhau, có người mắt to, mắt nhỏ v.v., nên khi dùng một ngưỡng cố định sẽ không phù hợp cho mọi đối tượng. Bên cạnh đó, công trình [19] [20] đánh giá mức độ tập trung của người học dựa trên các biểu hiện trên khuôn mặt. Các cảm xúc cơ bản được phát hiện bằng mô hình mạng học sâu. Kết quả thử nghiệm của mô hình được thực hiện trên tập dữ liệu FERC-2013 [21]. Để mô hình nhận dạng đạt được chất lượng cao thì tập dữ liệu huấn luyện là một trong những yếu tố đóng vai trò rất quan trọng. Tuy nhiên, tập dữ liệu thực nghiệm FERC-2013 chỉ chứa bảy cảm xúc cơ bản, mà cảm xúc của người học có thể là đơn lẻ hoặc hỗn hợp. Do đó, vấn đề khan hiếm dữ liệu về cảm xúc hỗn hợp của người học dẫn đến thiếu chính xác cho các mô hình nhận dạng. Vì vậy, đây cũng là một thách thức mà Luận án quan tâm và nghiên cứu.

Mặt khác, để kết quả đánh giá của người học được khách quan và chi tiết thì quy trình đánh giá cũng là một vấn đề mà Luận án cần xem xét. Vì trong quá trình học tập, có những thời điểm người học rất tập trung hoặc có những thời điểm người học bị sao lãng do tác động từ nhiều tố yếu như: bạn bè, môi trường, v.v. Những dấu hiệu này có thể đan xen lẫn nhau trong suốt quá trình của buổi học và nếu chỉ xem xét ở một thời điểm để kết luận là thiếu chính xác. Vì vậy, quy trình đánh giá sự tập trung của người học là một yếu tố quan trọng để mang lại kết quả đánh giá chi tiết. Từ đó, giáo viên có thể nhận được những phản hồi kịp thời và nắm rõ hơn về tình hình học tập của người học. Điều này nhằm giúp giáo viên cải thiện giáo trình, bài giảng và phương pháp giảng dạy để mang lại hiệu suất cao trong học tập cho người học.

Tóm lại, với những khó khăn và thách thức được thảo luận bên trên mà các công trình khoa học nghiên cứu trước đây còn tồn tại, Luận án tổng hợp và xác định ba vấn đề cần hoàn thiện nhằm nâng cao chất lượng cho bài toán đánh giá sự tập trung của người học như sau: (1) đưa ra quy trình đánh giá mức độ tập trung của người học, (2) giải quyết vấn đề khi dùng một ngưỡng cố định để xác định trạng thái đóng/mở của mắt, (3) giải quyết vấn đề khan hiếm dữ liệu huấn luyện về cảm xúc hỗn hợp của người học.

1.3 Mục tiêu nghiên cứu

Luận án tập trung nghiên cứu và giải quyết bài toán đánh giá sự tập trung của người học. Từ đó, đưa ra một số vấn đề đánh giá có thể giải quyết thông qua hình

ảnh dựa vào các kỹ thuật nhận dạng biểu hiện trên khuôn mặt, thừa kế các kết quả hiện có và cải tiến chúng nhằm nâng cao chất lượng cho bài toán đánh giá. Qua quá trình nghiên cứu và tham khảo các công trình nghiên cứu liên quan, Luận án tổng hợp và đưa ra ba mục tiêu chính nhằm khắc phục những hạn chế còn tồn tại như sau:

Thứ nhất, Luận án nghiên cứu đưa ra phương pháp đánh giá mức độ tham gia của người học dựa trên hành vi đóng/mở của mắt. Quy trình đánh giá được xem xét và phân tích trên cả quá trình học tập của người học. Kỹ thuật đánh giá trên từng thời điểm riêng lẻ, sau đó tổng hợp lại và tính tỷ lệ của mỗi trạng thái so với cả quá trình.

Thứ hai, Luận án nghiên cứu cải tiến kỹ thuật phát hiện trạng thái đóng/mở của mắt với ngưỡng được xác định linh động phù hợp cho mọi đối tượng. Vì đặc trưng của mắt ở mỗi người là khác nhau, có người mắt to, có người mắt nhỏ v.v. nên khi sử dụng một ngưỡng cố định thì không bao quát hết mọi trường hợp.

Thứ ba, Luận án nghiên cứu đề xuất kỹ thuật phân rã biểu cảm thành phần cơ bản. Vì trên thực tế, cảm xúc của người học được biểu diễn hỗn hợp từ các cảm xúc cơ bản mà dữ liệu huấn luyện cho mô hình nhận dạng chỉ chứa bảy cảm xúc đơn lẻ. Do đó, kỹ thuật này xác định rõ chi tiết về cảm xúc của người học nhằm giải quyết vấn đề khan hiếm dữ liệu huấn luyện và được áp dụng trong bài toán đánh giá sự tập trung dựa trên cảm xúc của người học.

1.4 Đối tượng, Phạm vi, Phương pháp

a) Đối tượng nghiên cứu

Các kỹ thuật xử lý video và hình ảnh liên quan đến việc phát hiện và rút trích các đặc trưng trên khuôn mặt của người.

Các kỹ thuật đánh giá sự tham gia dựa trên phân tích ảnh trạng thái mắt, biểu hiện trên khuôn mặt.

b) Phạm vi nghiên cứu

Luận án tập trung vào nền tảng thị giác máy tính để phát triển kỹ thuật đánh giá sự tập trung dựa trên hành vi đóng/mở mắt và biểu hiện trên khuôn mặt của người học.

c) Phương pháp nghiên cứu

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

Để đánh giá mức độ tham gia của người học sử dụng thị giác máy tính. Các kỹ thuật đánh giá được trình bày trong Luận án này bao gồm:

- Kỹ thuật dựa vào trạng thái đóng/mở của mắt;
- Kỹ thuật dựa vào biểu hiện trên khuôn mặt;

Lý do Luận án chọn các kỹ thuật này là vì những tín hiệu của sự mệt mỏi và mất tập trung chủ yếu xuất hiện rõ nét trong mắt người học và những biểu hiện trên khuôn mặt.

Quá trình phát hiện mất tập trung dựa vào trạng thái mắt và biểu hiện trên khuôn mặt của người học bao gồm các vấn đề: phát hiện khuôn mặt (Face Detection), phát hiện mắt (Eye Detection), phát hiện trạng thái đóng/mở của mắt (Eye close/open Detection) và nhận dạng cảm xúc trên khuôn mặt của người học trong ảnh (Facial Expression Recognition - FER).

Để phát hiện khuôn mặt, Luận án sử dụng kỹ thuật tại [22]. Trong kỹ thuật, tác giả đề xuất kỹ thuật phân lớp bằng mô hình cây quyết định được xây dựng dựa trên tiếp cận boosting với thuật toán lựa chọn là GentleBoost. Hiện tại có nhiều kỹ thuật phát hiện khuôn mặt người trong ảnh đã được các nhà khoa học công bố trong nhiều năm qua và đạt được độ chính xác cao [23]. Tuy nhiên, Luận án lựa chọn kỹ thuật đề xuất tại [22] với lý do tư thế của người ngồi học và tư thế của người lái xe gần như tương đồng với nhau, khuôn mặt hướng về trước. Các kỹ thuật phát hiện khuôn mặt khác chỉ khoanh vùng những điểm là khuôn mặt không quan tâm đến hướng của mặt người. Còn đối với kỹ thuật [22], do tính chất về an toàn khi lái xe nên có xem xét thêm yếu tố khuôn mặt hướng thẳng, yếu tố này phù hợp với bối cảnh của Luận án là người học phải hướng thẳng về bục giảng hoặc màn hình

(trường hợp học trực tuyến) mới xác định được mức độ tham gia / tập trung trong lớp học.

- Để phát hiện mắt người trong ảnh, Luận án sử dụng các dấu mốc trên khuôn mặt (Facial Landmarks) được trích xuất từ 68 điểm tương ứng với các đặc điểm trên khuôn mặt được ước tính bằng cách sử dụng mô hình phát hiện dấu mốc được đào tạo trước từ thư viện Dlib [24].

- Để xác định trạng thái mắt, Luận án áp dụng kỹ thuật cải tiến dựa trên đặc trưng tỷ lệ khung hình chữ nhật [25]. Thuật toán cải tiến này có khả năng tự động tính toán ngưỡng phù hợp cho mọi đối tượng thay vì phải xác định trước giá trị cố định cho thông số này.

- Cuối cùng để nhận dạng cảm xúc trên khuôn mặt của người học, Luận án sử dụng mô hình mạng [17] và bộ dữ liệu huấn luyện chuẩn FEREC-2013 [21] gồm bảy cảm xúc cơ bản. Trong công trình [17], tác giả xây dựng mô hình dựa trên mô hình VGGNet bằng cách giảm số lớp của nó. Cho đến nay, các mô hình VGG19, Inception_v4 [26] và ResNet [27] là một số mô hình tiên tiến nhất (State-of-the-art - SoTA) cho các tác vụ phân loại hình ảnh dựa trên mạng nơ-ron tích chập (Convolution Neural Network - CNN) được sử dụng cho bài toán FER. Tuy nhiên, Luận án lựa chọn mô hình VGGNet vì thông thường các mô hình SoTA bao gồm một số lớp mạng nơ-ron và hình dạng hình ảnh đầu vào cố định. Ví dụ, VGG19 có 19 lớp với hình dạng hình ảnh đầu vào là $(224 \times 224 \times 3)$. Do đó, việc đào tạo các mô hình SoTA đòi hỏi một tập dữ liệu lớn bao gồm các hình ảnh có độ phân giải cao (ít nhất là $224 \times 224 \times 3$) để đạt được độ chính xác phân loại. Mặt khác, công việc của Luận án cũng hướng đến việc phân tích và đánh giá mức độ tham gia của người học qua các video được ghi nhận trong lớp học bằng máy ghi hình thông thường hoặc là video được ghi nhận các lớp học trực tuyến bằng các phần mềm Microsoft Teams, Google Meet, Zoom v.v., nên có các hạn chế về kích thước và độ phân giải. Bên cạnh đó, tập dữ liệu chuẩn FEREC-2013 [21] dùng để huấn luyện mô hình cho các cảm xúc cơ bản có độ phân giải thấp. Vì vậy, với những hạn chế về kích thước và độ phân giải của hình ảnh mà Luận án dùng mô hình VGGNet để giải quyết cho công việc hiện tại.

1.5 Đóng góp của luận án

(1) Đề xuất kỹ thuật phân rã biểu cảm thành phần cơ bản trên khuôn mặt nhằm giải quyết vấn đề khan hiếm về dữ liệu huấn luyện. Do trong quá trình học, người học luôn luôn tiếp nhận những thông tin từ giáo viên, bạn bè, tác động từ môi trường v.v. Vì vậy, cảm xúc của người học ở mỗi thời điểm xảy ra khác nhau, có thể là hỗn hợp hoặc đơn lẻ nên cảm xúc cũng thể hiện những biến dạng tương ứng. Mặt khác, một trong những yếu tố ảnh hưởng đến chất lượng nhận dạng của mô hình mạng học sâu là tập dữ liệu huấn luyện. Tuy nhiên, trong bộ dữ liệu chuẩn FEREC-2013, chỉ bao gồm bảy cảm xúc cơ bản, mà cảm xúc trong học tập có thể là hỗn hợp. Do đó, Luận án đề xuất kỹ thuật phân rã biểu cảm để nâng cao độ chính xác cho bài toán nhận dạng và để hiểu rõ chi tiết hơn về cảm xúc của người học. Kết quả nghiên cứu được công bố trong công trình [28] và được trình bày chi tiết trong Chương 3 của Luận án

(2) Đưa ra thang đo đánh giá mức độ tham gia của người học dựa trên hành vi đóng/mở của mắt. Quy trình đánh giá được xem xét và phân tích trên cả quá trình học tập của người học. Kỹ thuật đánh giá trên từng thời điểm riêng lẻ, sau đó tổng hợp lại và tính tỷ lệ của mỗi trạng thái so với cả quá trình. Kết quả nghiên cứu được công bố trong công trình [29] và được trình bày chi tiết trong Chương 4 của Luận án.

(3) Do tính chất đặc trưng về mắt của mỗi người là khác nhau, dẫn đến độ cao và độ rộng của mắt ở mỗi người cũng khác nhau. Nên khi sử dụng một ngưỡng cố định để xác định trạng thái đóng/mở của mắt là không phù hợp và thiếu chính xác. Vì vậy, Luận án cải tiến kỹ thuật phát hiện trạng thái đóng/mở của mắt với ngưỡng được xác định một cách linh động dựa trên tỷ lệ giữa chiều cao và chiều rộng của khung hình chữ nhật chứa mắt. Kỹ thuật phù hợp cho mọi đối tượng và nâng cao được độ chính xác khi phát hiện. Kết quả nghiên cứu được công bố trong công trình [25] và được trình bày chi tiết trong Chương 4 của Luận án.

1.6 Bố cục của luận án

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

Chương 1: Là chương mở đầu, trong chương này Luận án giới thiệu khái quát về đề tài, mục tiêu, đối tượng, phạm vi nghiên cứu. Đồng thời nêu tóm tắt những đóng góp mà luận án đạt được.

Chương 2: Là chương tổng quan, trong chương này Luận án trình bày các vấn đề liên quan đến đánh giá sự tập trung nhằm làm nền tảng cho việc xây dựng các kỹ thuật được đề xuất trong các chương sau. Cụ thể, chương này trình bày ý nghĩa của sự tập trung, các yếu tố, phương pháp đánh giá sự tập trung, bài toán về nhận dạng biểu cảm khuôn mặt, các bộ dữ liệu dùng trong nghiên cứu và một số vấn đề về cơ sở lý thuyết trong xử lý ảnh và nhận dạng.

Chương 3: Nhận dạng biểu cảm thành phần khuôn mặt. Trong chương này, Luận án trình bày kỹ thuật đánh giá sự tập trung của người học dựa trên biểu cảm của khuôn mặt. Thông qua kết quả này, Luận án đề xuất một kỹ thuật phân rã biểu cảm thành phần cơ bản trên khuôn mặt của người học. Kỹ thuật này tách riêng biệt biểu cảm của người học thành những thành phần cơ bản, nhằm xác định chi tiết hơn về các cảm xúc của người học. Từ đó cải thiện được độ chính xác cho bài toán đánh giá. Vì trên thực tế các biểu cảm trên khuôn mặt của người học được biểu diễn hỗn hợp từ những cảm xúc cơ bản. Kết quả nghiên cứu của đề xuất này được công bố tại [28].

Chương 4: Nhận dạng biểu cảm hành vi khuôn mặt. Trong chương này, Luận án trình bày kỹ thuật đánh giá sự tập trung của người học dựa trên trạng thái đóng/mở của mắt [29]. Thông qua kết quả nghiên cứu của công trình này, Luận án cải tiến kỹ thuật phát hiện trạng thái đóng/mở của mắt với ngưỡng được xác định linh động, phù hợp cho mọi đối tượng nhằm nâng cao độ chính xác khi phát hiện [25]. Vì đặc trưng của mắt ở mỗi người là khác nhau, có người mắt to, mắt nhỏ v.v. Nên khi sử dụng một ngưỡng cố định để xác định trạng thái đóng/mở của mắt là không phù hợp và thiếu chính xác.

Chương 5: Phân kết luận và hướng phát triển của Luận án trình bày tổng hợp lại kết quả nghiên cứu đã đạt được, đồng thời đề xuất những hướng phát triển tiếp theo của Luận án.

CHƯƠNG 2

TỔNG QUAN VỀ ĐÁNH GIÁ SỰ TẬP TRUNG CỦA NGƯỜI HỌC VÀ BÀI TOÁN NHẬN DẠNG BIỂU CẢM KHUÔN MẶT

2.1 Tổng quan về đánh giá sự tập trung của người học

2.1.1 Ý nghĩa sự tập trung

Sự tập trung là một trong những yếu tố quan trọng tác động đến sự thành công của một nhiệm vụ ngắn hạn hoặc dài hạn. Đó là một cấu trúc phức tạp vì không những được đo lường thông qua quan sát mà còn được đo lường thông qua hành vi của đối tượng. Nó là thước đo đánh giá năng lực và được ứng dụng trong nhiều lĩnh vực bao gồm: Tâm lý học [30], Giao diện người - Máy tính (Human Computer Interface - HCI) [31], Giáo dục [4] [32].

Trong giáo dục, sự tập trung thường được sử dụng để mô tả người học sẵn sàng tham gia vào các hoạt động thường ngày của trường, chẳng hạn như tham gia nghiêm túc trong lớp học, nộp bài tập được yêu cầu và làm theo hướng dẫn của giáo viên trong lớp. Theo [4] [32], có ba yếu tố chính góp phần vào quá trình tập trung của người học bao gồm: hành vi, nhận thức và cảm xúc. Hiểu rõ hành vi của người học trong quá trình tham gia vào lớp học, nó sẽ cung cấp thông tin hữu ích cho môi trường giáo dục như sau:

- (i) **Đối với cơ sở quản lý giáo dục:** Cải thiện chính sách học vụ, môi trường học tập, cơ sở vật chất, chương trình đào tạo và tạo điều kiện tốt cho người học để ngăn ngừa tình trạng như chán nản, sao lãng, bỏ học.
- (ii) **Đối với giáo viên:** cải tiến phương pháp giảng dạy, giáo trình, bài giảng để nâng cao hiệu suất cho người học

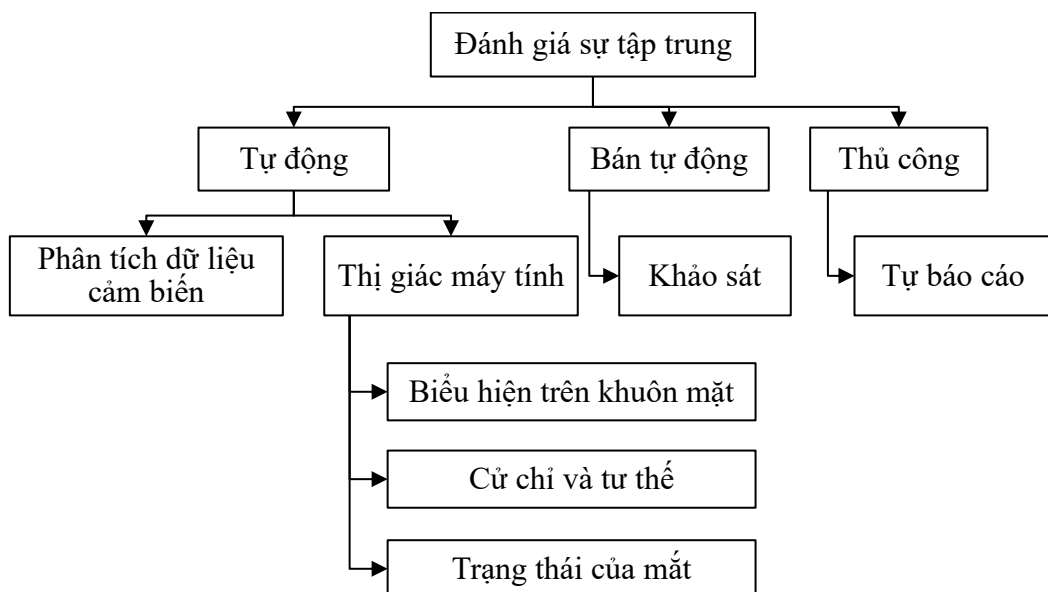
2.1.2 Các yếu tố đánh giá sự tập trung

Các nhà nghiên cứu đã đưa ra những giải pháp để theo dõi và đánh giá sự tập trung của người học khi tham gia học tập. Nhìn chung các giải pháp đều có sự tương đồng và theo [4] [32] các yếu tố nhận biết sự tham gia được các nhà nghiên cứu xác định dựa trên thị giác máy tính gồm có: (1) Hành vi và (2) cảm xúc.

(1) **Về hành vi** dựa trên các hành động ứng xử trong quá trình học tập, tham gia vào các hoạt động xã hội hoặc ngoại khóa. Nó thường được đánh giá dựa trên một loạt các hành vi có thể quan sát được như tham gia lớp học, thời gian làm bài tập về nhà, nỗ lực (ví dụ: thời gian làm nhiệm vụ) và tuân thủ các quy tắc của lớp học. Tham gia hành vi cũng có thể được coi là tham gia học tập và được mô tả các hành động ứng xử trong quá trình học tập như: đóng/mở mắt, ngáp, gãi tay phát biểu, tư thế đầu và tư thế hình thể [33] [34] [35] [36] [37] [38].

(2) **Về cảm xúc** của con người, một nhóm nghiên cứu chủ yếu đề cập đến bảy loại cảm xúc phổ biến: Vui, buồn, sợ hãi, tức giận, ghê tởm, ngạc nhiên và trung lập [32]. Những cảm xúc này có thể liên quan đến các chuyển động của cơ mặt được mô tả trong Hệ thống mã hóa hành động trên khuôn mặt [39], hệ thống này có thể được nhận biết bằng thị giác máy tính. Theo nghiên cứu [40] đã cho thấy độ chính xác thường vượt qua 90% đối với các phân tích phụ thuộc vào đối tượng (khi áp dụng các bộ phân loại khác nhau cho các cá nhân khác nhau) và cao hơn 70% đối với các phương pháp tiếp cận không phụ thuộc vào đối tượng (khi áp dụng cùng một bộ phân loại trên tất cả các cá nhân). Tuy nhiên, những độ chính xác này phụ thuộc vào phương pháp được sử dụng, quy mô của tập huấn luyện. Những cảm xúc hỗn hợp có xu hướng khó xác định, vì chúng ít liên kết với các cơ trên khuôn mặt người và dữ liệu đào tạo kém phong phú.

2.1.3 Phương pháp đánh giá sự tập trung



Hình 2.1 Phân loại các phương pháp phát hiện sự tham gia [6]

Một số nghiên cứu về phát hiện sự tham gia đã được các nhà nghiên cứu công bố và được chia thành thành ba loại chính: tự động, bán tự động và thủ công [6]. Do sự thành công của các giải pháp học máy và học sâu dựa trên thị giác máy tính, các phương pháp tự động đang được chú ý trong những năm gần đây và rất hữu ích trong việc phân tích và dự đoán.

2.1.3.1 Phương pháp thủ công (Manual)

Đánh giá sự tập trung dựa trên việc thu thập kết quả tự báo cáo (Self Reporting) của người học thông qua biểu mẫu khảo sát được thiết lập sẵn mà trong đó người học tự báo cáo riêng mức độ của họ về sự thích thú hay nhàm chán. Tất cả các bảng câu hỏi khảo sát không nhất thiết chỉ ra mức độ tham gia của người học một cách trực tiếp, thay vào đó chúng ngụ ý mức độ tham gia như một biến tiềm ẩn mô tả bằng cách sử dụng phân tích nhân tố [41]. Tự báo cáo được nhiều nhà nghiên cứu quan tâm vì nó dễ thực hiện. Ví dụ, sẽ rất hữu ích khi biết được số lượng người học buồn chán hoặc chán nản trong cuộc khảo sát. Tuy nhiên, hướng tiếp cận này tiêu tốn rất nhiều thời gian và công sức. Bên cạnh đó, kết quả của việc phân tích phụ thuộc vào yếu tố nằm ngoài tầm kiểm soát của nhà nghiên cứu, chẳng hạn như tính trung thực của người học khi tham gia trả lời các câu hỏi trong cuộc khảo sát. Do đó có thể dẫn đến kết quả kém chính xác trong việc phân tích.

2.1.3.2 Phương pháp bán tự động (Semi-Automatic)

Phương pháp bán tự động bao gồm phương pháp khảo sát kiến thức và phương pháp dựa trên sinh lý học. Trong việc khảo sát kiến thức, giáo viên ước tính mức độ tham gia của học sinh bằng cách đánh giá phản ứng của họ đối với các câu hỏi được hỏi trong khi giảng dạy. Ví dụ, thời gian trả lời rất ngắn đối với các câu hỏi đã cho thấy rằng người học không tham gia và chỉ đơn giản là đưa ra các câu trả lời ngẫu nhiên mà không cần nỗ lực. Công việc được thực hiện trong [42] [43] được sử dụng để thực hiện phương pháp này một cách dễ dàng. Mặc dù phương pháp này đã được sử dụng rộng rãi trong các hệ thống dạy kèm thông minh (Intelligent Tutoring Systems - ITS), nhưng không có nhiều ứng dụng của phương pháp này trong các môi trường giáo dục khác, chẳng hạn như trong môi trường học tập trực tuyến [44].

2.1.3.3 Phương pháp tự động (Automatic)

Trích xuất các đặc trưng từ các đặc điểm khác nhau được ghi nhận bởi cảm biến hình ảnh (ví dụ: chuyển động của mắt, nét mặt, cử chỉ và tư thế), cảm biến sinh lý và thần kinh (ví dụ: nhịp tim, điện não đồ, huyết áp). Các phương pháp này trích xuất các đặc trưng một cách tự động và không làm gián đoạn người học trong quá trình phát hiện sự tham gia. Các phương pháp trong danh mục tự động được chia thành hai nhóm, gồm có: phân tích dữ liệu cảm biến và phương pháp dựa trên thị giác máy tính.

(i) Trong phân tích dữ liệu cảm biến (Sensor data analysis), các chỉ số cảm biến sinh lý và thần kinh được sử dụng để đo mức độ tương tác. Trong các tài liệu khoa học thần kinh, sự tham gia thường được đánh giá bằng mức độ kích thích hoặc tỉnh táo. Các phương pháp dựa trên sinh lý, mức độ tham gia của người học được ước tính dựa trên các tín hiệu như: điện não đồ (Electro Encephalo Graphy - EEG), điện tâm đồ (Electro Cardio Gram - ECG) bằng cách sử dụng các thiết bị như dây đeo, cảm biến hoạt động điện da [45] [46]. Tuy nhiên, các biện pháp này yêu cầu các cảm biến chuyên dụng và không thuận tiện để sử dụng, tốn kém và gây khó khăn trong các môi trường giáo dục ngoài đời thực.

(ii) Các phương pháp dựa trên thị giác máy tính (Computer Vision-Based) cung cấp một số cách để đo lường mức độ tương tác của người học bằng cách điều tra các dấu hiệu từ cử chỉ và tư thế, trạng thái của mắt và biểu hiện trên khuôn mặt [4] [32]. Ưu điểm chính của phương pháp dựa trên thị giác máy tính là dễ sử dụng và không làm gián đoạn các hoạt động của người học trong quá trình đánh giá, phân tích các tín hiệu không lời mà không có sự can thiệp của người. Kỹ thuật tính toán hiệu quả và chi phí máy ảnh thấp như: webcam hoặc camera. Tính khả dụng rộng rãi của nó trong điện thoại di động, máy tính bảng, máy tính và thậm chí cả ô tô. Trong nhiều tài liệu gần đây, các công trình đã được đề xuất để phát triển Hệ thống giám sát tham gia tự động (Automatic Engagement Management System - AEMS) trong lĩnh vực giáo dục bằng cách sử dụng các dấu hiệu hành vi và cảm xúc của người học. Trong số đó, các nghiên cứu đề cập đến môi trường học tập điện tử với một học sinh duy nhất trong một khung video [47] [14] [48] [49] [50]. Một số công trình hỗ trợ trong môi trường lớp học truyền thống với nhiều học sinh trong một khung video [15] [51] [5] [52] [32] [53]. Mặc dù các

phương pháp dựa trên thị giác máy tính để phát hiện mức độ tương tác đã được cộng đồng ITS quan tâm, nhưng vẫn còn nhiều việc phải làm trước khi biến các hệ thống tự động này trở nên thực tế trong nhiều môi trường giáo dục. Hơn nữa, những công trình nghiên cứu này đã sử dụng những cảm xúc cơ bản như: hạnh phúc, tức giận, sợ hãi, buồn bã và ngạc nhiên [39] làm tín hiệu cảm xúc để ước tính mức độ tương tác của người học, điều này không phù hợp với môi trường học thuật. Nghiên cứu ban đầu đã chứng minh rằng có một số khác biệt giữa cảm xúc học thuật và cảm xúc cơ bản [54].

2.1.4 Hệ thống đánh giá sự tập trung của người học

Phần này cung cấp một cái nhìn tổng quan về các công trình gần đây đã được sử dụng để giải quyết vấn đề ước tính mức độ tham gia của người học trong môi trường học tập trực tuyến và truyền thống. Như đã thảo luận ở trên, các phương pháp đánh giá sự tham gia của người học được phân thành ba nhóm, đó là: Phương pháp thủ công, bán tự động và tự động. Do sự thành công của các giải pháp học máy và học sâu dựa trên thị giác máy tính, các phương pháp tự động đang được chú ý trong những năm gần đây và đặc biệt hữu ích trong các lớp học cả truyền thống và trực tuyến. Một số kỹ thuật được các nhà khoa học nghiên cứu đề xuất như tư thế đầu, chuyển động của đầu, trạng thái mắt và các biểu hiện trên khuôn mặt của người học [4] [5] [48] [32] [19]. Bảng bên dưới mô tả các công trình được nghiên cứu trong những năm gần đây cũng như giới hạn của phương pháp.

Bảng 2.1 Các công trình nghiên cứu liên quan

Công trình nghiên cứu liên quan và giới hạn của phương pháp	
[53]	Peng và cộng sự đào tạo mô hình sử dụng thuật toán SVM và rừng ngẫu nhiên (RF) dựa trên các tính năng từ khuôn mặt, nhịp tim và thính giác. Họ theo dõi trạng thái tinh thần của người học (chán nản, thất vọng, tập trung và bối rối) trong các buổi học tại lớp. ❖ Giới hạn: Kỹ thuật này cần các thiết bị vật lý để thu thập dữ liệu của người học như Apple watch và Air pods, những thiết bị này đòi hỏi chi phí cao để triển khai hệ thống với số lượng quy mô lớn.

Công trình nghiên cứu liên quan và giới hạn của phương pháp	
[32]	Pieter Vanneste và cộng sự đề xuất một kỹ thuật để đánh giá mức độ tham gia của người học trong môi trường lớp học bằng cách nhận biết các hành vi của họ như giơ tay và ghi chép. Một mô hình học sâu đã được đào tạo để nhận ra những hành vi như vậy, và nó đạt được độ phản hồi (Recall) và độ chính xác (Precision) lần lượt là 63% và 45%. ❖ Giới hạn: Trong nghiên cứu này, Tác giả không xem xét các tín hiệu khác mà chỉ sử dụng các tín hiệu hành vi như giơ tay phát biểu và ghi chép của người học để ước tính mức độ tương tác của họ.
[52]	Luo và cộng sự đào tạo một mô hình 3D sử dụng thuật toán rừng ngẫu nhiên (RF) dựa trên tư thế đầu, nét mặt và sử dụng thiết bị thông minh để thu thập dữ liệu của người học trong lớp để ước tính mức độ tập trung. Mô hình đạt được độ chính xác phân loại là 87,5%. ❖ Giới hạn: Nghiên cứu cần các thiết bị vật lý để thu thập dữ liệu tương tác. Do đó, tốn kém chi phí và nó chỉ sử dụng cảm xúc trên khuôn mặt để ước tính mức độ chú ý của người học.
[5]	Zheng và cộng sự đã thiết kế một khung giám sát hành vi của người học như giơ tay, đứng và ngủ trong môi trường lớp học. Họ đã đào tạo mô hình bằng thuật toán phát hiện đối tượng R-CNN được cải tiến để nhận ra các hành vi nói trên với độ chính xác là 57,6%. ❖ Giới hạn: Nghiên cứu này chỉ hoạt động để phát hiện các hành vi của người học, nó không thể được sử dụng để dự đoán mức độ tham gia dựa trên các tín hiệu cảm xúc trong học tập.
[51]	Ashwin và Guddeti phát triển một hệ thống dựa trên mạng nơ-ron tích chập (CNN) để phân tích các tín hiệu không lời và phân loại mức độ tương tác của người học thành một trong bốn cấp độ: “Không tham gia”, “Tham gia thấp”, “Tham gia trung bình” và “Tham gia cao”. Hệ thống của họ đã được đào tạo và thử nghiệm với cử chỉ tay và tư thế hình thể trong môi trường lớp học trực tiếp lớn với độ chính xác phân loại mức độ tương tác là 71%.

Công trình nghiên cứu liên quan và giới hạn của phương pháp	
	❖ Giới hạn: Phương pháp này không nhấn mạnh vào phân tích cảm xúc khuôn mặt và hệ thống yêu cầu 2153ms để xử lý một khung hình. Do đó, nó không thể được sử dụng triển khai trong thời gian thực.
[15]	Soloviev phát triển hệ thống dựa trên cây quyết định sử dụng các đặc điểm khuôn mặt như tư thế đầu, nét mặt để phân loại người học tham gia hoặc không tham gia ❖ Giới hạn: Hệ thống được đề xuất chưa được thử nghiệm trong môi trường lớp học nên chưa có kết quả đánh giá hiệu quả của phương pháp.
[55]	Zaletelj và Kosir ước tính sự chú ý của người học trong môi trường lớp học trực tiếp bằng cách sử dụng các tín hiệu không lời. Tác giả phát triển mô hình với các tính năng 2D và 3D được trích xuất từ máy ảnh Kinect One bằng cách sử dụng các thuật toán học máy như cây quyết định và k-láng giềng gần nhất (k-Nearest Neighbor - k-NN). Hệ thống của họ đã đạt được độ chính xác là 75,3%, được đánh giá bằng cách so sánh sự chú ý được dự đoán với dữ liệu được gắn nhãn. ❖ Giới hạn: Do hạn chế về công nghệ với máy ảnh Kinect với tầm nhìn ngắn, nên phân tích chỉ giới hạn ở sáu đối tượng mà không phải toàn bộ lớp học.
[56]	Klein và cộng sự thực hiện quá trình học chuyển giao cho các mức độ chú ý của người học với mô hình AlexNet và sử dụng bộ dữ liệu tự phát triển về các hành động, tư thế phổ biến trong lớp học. Hệ thống phân loại mức độ của người học là quan tâm hoặc không quan tâm. ❖ Giới hạn: Trong nghiên cứu này, tác giả chỉ quan tâm đến các tín hiệu về hành vi của người học để phân tích, thiếu xem xét đến các yếu tố như cảm xúc, trạng thái đóng/mở của mắt. Mặt khác, phân đánh giá thực nghiệm, tác giả sử dụng bộ dữ liệu tự xây dựng nên cũng chưa khái quát hết các trường hợp trong thực tế. Do đó, hiệu quả của phương pháp cần được đánh giá lại.

Công trình nghiên cứu liên quan và giới hạn của phương pháp	
[57]	Thomas và cộng sự phát triển hệ thống dựa trên SVM với tập dữ liệu về ánh mắt, tư thế đầu và nét mặt của người học để phân loại mức độ tương tác hoặc không tương tác. ❖ Giới hạn: Tác giả đã thử nghiệm trên quy mô lớp học nhỏ và kỹ thuật này không hỗ trợ theo dõi tương tác thời gian thực.
[58]	Whitehill và cộng sự đề xuất một hệ thống dựa trên máy học để phân loại mức độ tham gia của người học bằng cách phân tích các biểu hiện trên khuôn mặt và hành vi của người học. Các thí nghiệm của họ kết luận rằng máy vector hỗ trợ (Support Vector Machine – SVM) với các đặc trưng Gabor hoạt động tốt trong việc phân loại các biểu hiện trên khuôn mặt và hành vi của người học với giá trị diện tích dưới đường cong (Area Under the Curve - AUC) đạt được là 0,729.
[59]	Moncaresi và cộng sự đã sử dụng các đặc trưng trên khuôn mặt và nhịp tim để phát hiện mức độ tương tác. Tác giả sử dụng dữ liệu tự báo cáo được thu thập từ người học trong khi thực hiện nhiệm vụ để chú thích tập dữ liệu video của họ thành hai loại: tương tác, không tương tác và tác giả sử dụng công cụ học máy WEKA để phân loại. ❖ Giới hạn: Trong nghiên cứu này, tác giả có sử dụng thiết bị để thu thập dữ liệu nhịp tim của người học để đánh giá. Do đó, phương pháp này rất bất tiện và tốn kém chi phí để triển khai trong môi trường thực tế với quy mô lớn.
[60]	Gupta và cộng sự đề xuất và giới thiệu DAiSEE, là tập dữ liệu phân loại video đa nhãn. Tập dữ liệu chứa các nhãn của bốn trạng thái cảm xúc liên quan đến tương tác của người dùng: tương tác, thất vọng, bối rối và buồn chán. Và có bốn cấp độ nhãn trong mỗi danh mục tương tác, đó là rất thấp, thấp, cao và rất cao. Quan điểm của tác giả chủ yếu về khả năng sử dụng của tập dữ liệu và sử dụng một số cấu trúc mạng cổ điển của CNN để phân tích tính thực tiễn của tập dữ liệu. ❖ Giới hạn: Trong nghiên cứu này, tác giả tập trung xây dựng bộ dữ liệu, không đi sâu vào kỹ thuật, nên không có kết quả so sánh mức độ hiệu quả của kỹ thuật so với các công trình khác.

Công trình nghiên cứu liên quan và giới hạn của phương pháp	
[61]	Bosch và cộng sự sử dụng các thuật toán và thị giác máy tính để phát hiện mức độ tham gia của người học từ nét mặt và chuyển động cơ thể. Tác giả xây dựng các mô hình học máy khác nhau như SVM, cây quyết định (Decision Tree) cho mục đích này. Độ chính xác của mỗi trạng thái phân loại được đo lường theo giá trị AUC: buồn chán (0,61), bối rối (0,65), thích thú (0,87), tham gia (0,68) và thất vọng (0,63). ❖ Giới hạn: Hạn chế mà tác giả đề cập trong nghiên cứu này gồm: (1) Hạn chế về dữ liệu đào tạo do khó thu thập trong môi trường tự nhiên; (2) Sự phân bố của các trạng thái phụ thuộc vào giao diện sử dụng. Do trong nghiên cứu này, giao diện xây dựng dựa trên trò chơi, điều này có thể làm tăng mức độ tương tác và trạng thái tích cực trong việc thu thập dữ liệu để đánh giá. Ngoài ra, tác giả cũng đề cập đến vấn đề đa sắc tộc và nền văn hóa cần được xem xét và nghiên cứu trong tương lai.
[62]	Krithika và cộng sự phát triển một hệ thống theo dõi cảm xúc và cung cấp phản hồi theo thời gian thực về mức độ tập trung của người học. Họ sử dụng những cảm xúc như: phấn khích, buồn chán, ngáp, buồn ngủ, kiểu chuyển động bất thường của đầu và mắt để dự đoán mức độ tập trung. Tương tự, [14] [49] ước tính mức độ tương tác của người học trong thời gian thực bằng cách phân tích các nét mặt cơ bản của người học. Họ đã đào tạo một mô hình nhận dạng cảm xúc dựa trên mạng nơ-ron tích chập (CNN) với bộ dữ liệu FEREC-2013 [21] và đạt được độ chính xác lần lượt là 70% và 62%. ❖ Giới hạn: Trong nghiên cứu này, tác giả chỉ thực nghiệm và đánh giá cho 05 đối tượng. Do đó, tính hiệu quả của phương pháp cần được xem xét và đánh giá lại trên tập dữ liệu lớn.
[19]	Sharma và cộng sự trình bày cách tiếp cận của hệ thống để phát hiện mức độ tham gia của người học dựa trên hình ảnh. Tác giả áp dụng thuật toán Viola & Jones [23] để phát hiện khuôn mặt của người học. Sau đó, tác giả dựa trên chuyển động của đầu và mắt, vùng mắt được đưa vào mạng nơ-ron tích chập (CNN) như một bộ phân loại nhị phân để dự đoán trạng thái

Công trình nghiên cứu liên quan và giới hạn của phương pháp	
	chú ý của người học gồm có "phân tâm" hoặc "tập trung". Và tác giả đã tạo ra một mô hình CNN thứ hai với tên gọi là Mini-Xception ghi nhận cảm xúc thể hiện trên khuôn mặt của người học tại mỗi thời điểm. Cuối cùng, tác giả sử dụng bảng câu hỏi để thu thập ảnh hưởng của các biểu hiện khác nhau đến mức độ tương tác và ấn định các trọng số để đánh mức độ tương tác. Do đó, mức độ tham gia của người học được dự đoán qua các biểu hiện trên khuôn mặt dựa trên các trọng số khác nhau. ❖ Giới hạn: Trong phần đánh giá, tác giả thử nghiệm với 15 sinh viên. Các sinh viên được yêu cầu xem bài giảng thời lượng khoảng ~2 phút cho một chủ đề. Phần cuối của video, một bài kiểm tra cho tất cả sinh viên khoảng 3 phút gồm có 5 câu trắc nghiệm. Sinh viên dễ dàng trả lời nếu xem video bài giảng một cách chăm chú. Do đó, trong phần nghiên cứu này, tác giả đánh giá phương pháp dựa trên dữ liệu mà họ tự thu thập. Vì vậy, tính hiệu quả của phương pháp chưa được trình bày trên các tập dữ liệu khác.
[63]	Nezami và cộng sự đã tạo cơ sở dữ liệu về mức độ tương tác. Tác giả đã định lượng và đặc trưng hóa mức độ tương tác bằng cách sử dụng biểu cảm khuôn mặt được trích xuất từ hình ảnh. Đầu tiên, tác giả sử dụng một mạng nơ-ron giả lập trong quá trình nhận dạng nét mặt cơ bản, sau đó tác giả áp dụng mô hình tương tự như kiến trúc VGG-B để khởi tạo mô hình nhận dạng tương tác. Trong giai đoạn nhận biết, tác giả đã thiết kế hai mức đo lường khác nhau, đó là hành vi và cảm xúc ❖ Giới hạn: Nghiên cứu đã đạt được mục tiêu của tác giả đề ra. Trong phần thực nghiệm, tác giả thử nghiệm trên bộ dữ liệu FERC-2013 [21], đây là bộ dữ liệu chuẩn và chỉ chứa 7 cảm xúc cơ bản. Tuy nhiên, trong quá trình học, người học có thể tiếp thu nhiều thông tin cùng một lúc từ giáo viên, bạn bè, môi trường, Do đó, cảm xúc trong học tập tại một thời điểm được biểu diễn hỗn hợp từ các cảm xúc cơ bản. Vì vậy, tác giả chỉ phân tích trên cảm xúc cơ bản mà chưa đề cập chi tiết đến cảm xúc hỗn hợp của người học.

Công trình nghiên cứu liên quan và giới hạn của phương pháp	
[18]	Ayat và cộng sự đề xuất một mô hình đo lường mức độ tập trung của người học trong môi trường học tập điện tử. Họ sử dụng các phương pháp lai kết hợp giữa tư thế đầu và phát hiện trạng thái của mắt. Cụ thể, kỹ thuật thu thập thông tin qua webcam về tư thế xoay đầu của người học và kỹ thuật theo dõi ánh mắt để phân biệt hành vi của người học trong các lớp học trực tuyến. Tác giả đánh giá hiệu suất của kỹ thuật với 5 đối tượng gồm 10 video, mỗi video dài khoảng 5 phút. ❖ Giới hạn: Trong kỹ thuật này, tác giả sử dụng một ngưỡng cố định để theo dõi trạng thái của mắt. Điều này không khái quát được hết cho mọi đối tượng vì đặc trưng mắt ở mỗi người là khác nhau.

Bảng 2.1 trình bày kết quả nghiên cứu cũng như giới hạn của kỹ thuật. Nhìn chung các nhà khoa học đánh giá sự tập trung của người học dựa trên các yếu tố biểu cảm, trạng thái mắt và tư thế ngồi học. Các kỹ thuật giải quyết được mục tiêu của tác giả đề ra. Tuy nhiên, mỗi kỹ thuật vẫn còn tồn tại hạn chế của nó, chẳng hạn như: hạn chế về dữ liệu huấn luyện; thiếu xem xét nhiều yếu tố, chỉ xem xét yếu tố đơn lẻ (ví dụ: biểu cảm, tư thế hình thể hoặc mắt); trình theo dõi trạng thái của mắt sử dụng ngưỡng cố định, do đó không khái quát hết cho mọi đối tượng và chưa trình bày cụ thể quy trình đánh giá sự tập trung. Hạn chế khác là phải gắn thêm thiết bị vật lý để thu thập thông tin của người học, điều này khó triển khai trong môi trường thực tế.

Với những khó khăn và hạn chế của những công trình liên quan còn tồn tại. Luận án tổng hợp và tập trung nghiên cứu nhằm giải quyết những vấn đề sau: (i) vấn đề khan hiếm về tập dữ liệu huấn luyện cảm xúc hỗn hợp, (ii) kỹ thuật phát hiện trạng thái của mắt khi phải sử dụng một ngưỡng cố định, (iii) quy trình đánh giá mức độ tập trung của người học.

2.2 Bài toán nhận dạng biểu cảm khuôn mặt

Biểu hiện trên khuôn mặt là một phương tiện phi ngôn ngữ nhằm phản ánh thông điệp trong giao tiếp của mỗi người. Nhiều nghiên cứu cho thấy rằng 55% thông điệp liên quan đến cảm xúc và thái độ được biểu hiện trên khuôn mặt, 7% thông điệp được truyền đạt bằng từ ngữ, phần còn lại phụ thuộc vào cách ngôn ngữ

được sử dụng (cách mà các từ nói). Cảm xúc có thể dễ dàng nhận biết thông qua những biểu hiện trên khuôn mặt và máy tính có khả năng tương tác với mỗi người tương tự như cách mỗi người tương tác với nhau. Điều này giải thích lý do tại sao nhận dạng cảm xúc trên khuôn mặt ngày càng thu hút được sự quan tâm trong nhiều lĩnh vực. Các nhà khoa học đã chỉ ra rằng cảm xúc trên khuôn mặt có thể được nhận biết ở mọi người, không phân biệt chủng tộc, giới tính và độ tuổi.

2.2.1 Biểu cảm khuôn mặt

2.2.1.1 Giới thiệu

Biểu cảm của người nói có ảnh hưởng to lớn đến người tiếp nhận, vì vậy hiểu đúng biểu cảm có ý nghĩa vô cùng quan trọng, có thể thể hiện yếu tố quyết định trong giao tiếp xã hội. Chúng ta thường quan sát biểu cảm của người nói bởi vì đó là tín hiệu cho thấy hiện tại người đó đang trong trạng thái như thế nào và họ sẽ làm gì v.v. 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 [64].

Ngoài yếu tố trên khuôn mặt, biểu cảm cũng được thể hiện thông qua ngôn ngữ hình thể hoặc giọng điệu để truyền đạt một cảm xúc hoàn chỉnh và chính xác đến người tiếp nhận. Giọng điệu không chỉ là “lớp vỏ” bên ngoài của lời nói, mà còn là cách để chúng ta thể hiện tâm trạng và cảm xúc của thông điệp. Do đó, việc kết hợp yếu tố lời thoại đã nâng cao tính tương tác và thực tế của mô phỏng về biểu cảm [65]. Đồng thời đã tạo ra một góc nhìn độc đáo về cảm xúc và tâm trạng của người học. Việc xác định được sự tương quan giữa cử động và lời thoại có thể giúp chúng ta hiểu rõ hơn về mối quan hệ giữa tâm trạng và hành vi học tập của người học.

Từ những yếu tố trên, biểu cảm của người thể hiện tính giao tiếp nhất trên cơ thể và cung cấp nhiều thông tin hữu ích trong quá trình giao tiếp.

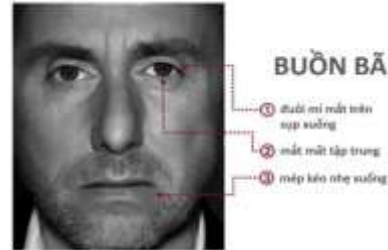
Các chuyên gia nghiên cứu tâm lý những biểu hiện dựa trên tình trạng tâm lý của con người và tin rằng chúng có một mức độ cụ thể và tồn tại trong một khoảng thời gian nhất định. Trạng thái tâm lý của một người tại một thời điểm cụ thể có thể có tác động lớn đến sự quyết định, hành vi, và hoạt động của họ. Trạng thái biểu cảm là cách mà con người truyền tải 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 tương quan giữa cử động khuôn mặt và

trạng thái biểu cảm, điều này được minh chứng qua những công trình nghiên cứu của Darwin năm 1972.

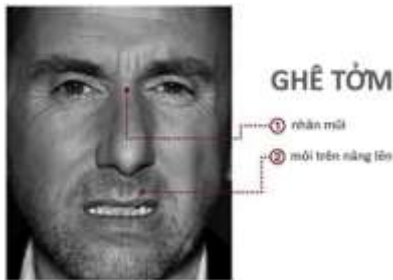
Nghiên cứu của Ekman [39] chỉ ra rằng có bảy biểu cảm khuôn mặt nhất quán, tương ứng với sáu cảm xúc cơ bản là vui, buồn, giận dữ, ngạc nhiên, ghê tởm, sợ hãi và một cảm xúc trung lập.



(1) Cảm xúc vui vẻ, hạnh phúc



(2) Cảm xúc buồn bã, chán nản



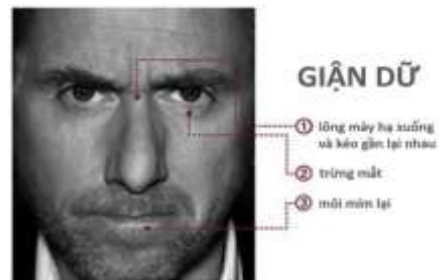
(3) Cảm xúc ghê tởm



(4) Cảm xúc sợ hãi



(5) Cảm xúc ngạc nhiên



(6) Cảm xúc giận dữ

(Nguồn tham khảo tại: <https://blog.trginternational.com/>)

Hình 2.2 Trạng thái biểu cảm khuôn mặt cơ bản

2.2.1.2 Đơn vị hành động trên khuôn mặt

Một cách toàn diện để phân tích các bộ phận trên khuôn mặt là Hệ thống mã hóa hành động trên khuôn mặt (Facial Action Coding System - FACS) [39]. FACS sử dụng các chuyển động của cơ mặt còn được gọi là đơn vị hành động (Action Unit - AU) để thiết kế thước đo lý thuyết về các cảm xúc. AU có thể xảy ra đơn lẻ hoặc kết hợp.

Theo [39] có sáu mươi bốn đơn vị hành động được mã hóa các chuyển động trên khuôn mặt khi nét mặt tạo ra một cảm xúc cụ thể. Trong phụ lục 1 của Luận án trình bày chi tiết các đơn vị cử động của FACS. Hình 2.3 minh họa một số đơn vị biểu hiện trên khuôn mặt.



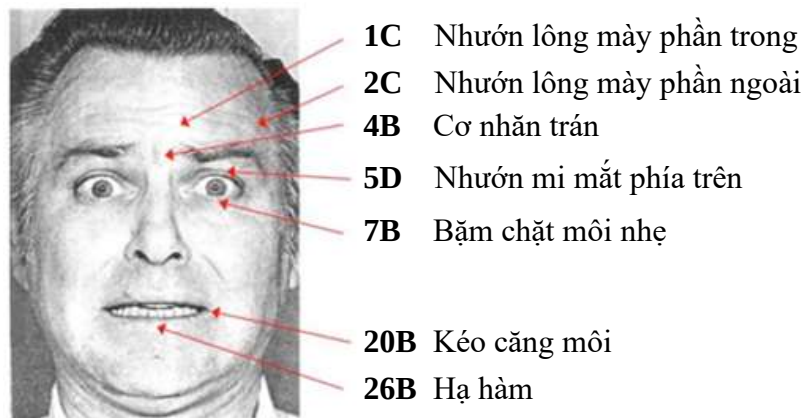
Hình 2.3 Một số ví dụ về AU [66]

Điểm cường độ: các cường độ của FACS được chú thích bằng cách thêm các chữ cái A – E (tương ứng với cường độ tối thiểu - tối đa) vào số đơn vị hành động (ví dụ: AU1A là hành động yếu nhất của AU1 và AU1E là hành động tối đa có thể đối với cá nhân).

Bảng 2.2 Ký hiệu diễn giải cường độ hành động

Ký hiệu	Diễn giải
A	Hành động yếu
B	Biểu hiện nhẹ
C	Biểu hiện đáng chú ý
D	Hành động mạnh (cực kỳ)
E	Tối đa

Ngoài ra, có các từ bổ nghĩa khác có trong mã FACS cho các biểu hiện cảm xúc, chẳng hạn như "R" đại diện cho một hành động xảy ra ở phía bên phải của khuôn mặt và "L" cho các hành động xảy ra ở bên trái.



Hình 2.4 Hình minh họa cường độ hành động

2.2.1.3 Hệ thống mã hóa đơn vị hành động trên khuôn mặt – FACS

Để nắm bắt được một cách chính xác sự đa dạng và phức tạp của biểu cảm khuôn mặt, các nhà nghiên cứu nhận thấy rằng cần thiết phát triển những tiêu chuẩn mã hóa khách quan. FACS là một trong những hệ thống mã hóa biểu đạt khuôn mặt được sử dụng phổ biến nhất trong các nghiên cứu khoa học về hành vi.

FACS được phát triển bởi Friesen và Ekman [39] với mục đích xác định tất cả các cử động trên khuôn mặt có thể quan sát được bằng mắt; nó tạo ra một chú thích rõ ràng và súc tích cho việc hoạt hóa các cơ của một diễn đạt trên khuôn mặt. FACS liên quan tới việc xác định những cơ mặt khác nhau, hoặc là riêng biệt, hoặc là kết hợp tạo ra những sự thay đổi về hành vi khuôn mặt. Những biến đổi trên khuôn mặt kết hợp cùng với một hay nhiều cơ bên dưới tạo nên sự biến đổi đó được gọi là các đơn vị cử động (AU); mỗi đơn vị cử động là một cử động cơ bản và được mô tả là tập hợp của một hoặc một số cơ trên khuôn mặt. FACS là một danh sách bao gồm 64 đơn vị cử động như vậy.

Các cử chỉ trên khuôn mặt (hoạt động của cơ mặt) là phương tiện trực tiếp giúp mỗi người truyền tải cảm xúc của mình [67] và đại diện cho các tín hiệu giọng nói có thể nhìn thấy. Do đó, các hệ thống phân tích tự động đối với những thay đổi tinh tế trên khuôn mặt dường như có một vị trí tự nhiên trong các hệ thống thị giác khác nhau bao gồm những công cụ tự động để nghiên cứu tâm lý, phân tích giọng nói, phân tích cảm xúc và giao diện người dùng tri giác.

Bảng 2.3 Cảm xúc và danh mục AU tương ứng [39]

Cảm xúc	Action Unit	Đặc trưng khuôn mặt
Vui	AU12	Kéo khóe môi (Lip Corner Puller)
	AU25	Tách môi trên và dưới (Lips Part)
Buồn	AU1	Nhướn mày trong (Inner Brow Raiser)
	AU4	Hạ lông mày (Brow Lowerer)
	AU15	Nén khóe môi (Lip Corner Depressor)
Sợ hãi	AU1	Nhướn mày trong (Inner Brow Raiser)
	AU2	Nhướn mày ngoài (Outer Brow Raiser)
	AU4	Hạ lông mày (Brow Lowerer)
	AU5	Nhướn mi trên (Upper Lid Raiser)
	AU20	Kéo căng môi (Lip Stretcher)
	AU25	Tách môi trên và dưới (Lips Part)
	AU26	Hạ hàm (Jaw Drop)
Ghê tởm	AU15	Nén khóe môi (Lip Corner Depressor)
Ngạc nhiên	AU1	Nhướn mày trong (Inner Brow Raiser)
	AU2	Nhướn mày ngoài (Outer Brow Raiser)
	AU5	Nhướn mi trên (Upper Lid Raiser)
	AU25	Tách môi trên và dưới (Lips Part)
	AU26	Hạ hàm (Jaw Drop)
Giận dữ	AU4	Hạ lông mày (Brow Lowerer)
	AU5	Nhướn mi trên (Upper Lid Raiser)
	AU7	Căng mí mắt (Lid Tightener)

Liên quan đến mối quan hệ giữa cảm xúc và cử động khuôn mặt, mỗi đơn vị cử động mã hóa các cử động cơ bản của một hay một nhóm cơ thường được quan sát thấy khi tạo ra biểu cảm khuôn mặt thể hiện cảm xúc. Đơn vị cử động này thường được thấy ở biểu cảm của những cảm xúc buồn, sợ, và giận. Ví dụ, AU4 mã hóa sự co của hai cơ có tên tiếng Anh là Corrugator supercilii và Depressor supercilii, đây là những cơ mà khi co sẽ khiến lông mày hạ thấp. Như vậy, FACS cung cấp một ngôn ngữ khách quan và toàn diện cho việc mô tả các biểu cảm khuôn mặt và gắn kết chúng trở lại với những ý nghĩa đã được biết đến từ các nghiên cứu thuộc khoa học hành vi. FACS được ứng dụng rất phổ biến trong lĩnh vực tạo chuyển động cho khuôn mặt. Các biểu cảm thể hiện cảm xúc trên khuôn mặt được xác định dựa trên các đơn vị cử động của FACS.

2.2.2 Nhận dạng biểu cảm khuôn mặt

Khi một người thể hiện cảm xúc thì luôn tồn tại một số đặc trưng nhất định trên khuôn mặt bất kể độ tuổi, vị trí địa lý hay điều kiện sống, v.v. Dựa trên đặc trưng này, ta có thể rút ra những đặc điểm quan trọng của cảm xúc để mô hình hóa và “dạy” cho máy tính hiểu được cảm xúc đó.

Trong những năm gần đây, với sự phát triển nhanh chóng và mạnh mẽ của trí tuệ nhân tạo, những bài toán về nhận dạng tự động đã nhận được sự nghiên cứu chuyên sâu và quan tâm rộng rãi như: tâm lý học, thị giác máy tính và nhận dạng khuôn mẫu. Nhận dạng biểu hiện trên khuôn mặt có các ứng dụng rộng rãi trong nhiều lĩnh vực, bao gồm hệ thống hỗ trợ lái xe nâng cao [68] [69], thực tế tăng cường [70], tương tác giữa người và máy tính [71] [72], thực tế ảo [73], giáo dục [74] và giải trí [75].

Có hai cách tiếp cận chính trong phân tích nét mặt: cách tiếp cận dựa trên hình học và cách tiếp cận dựa trên ngoại hình [76]. Các đặc điểm hình học trên khuôn mặt thể hiện hình thái và vị trí của từng bộ phận trên khuôn mặt (bao gồm miệng, mắt, lông mày và mũi). Các thành phần trên khuôn mặt hoặc những nét đặc trưng trên khuôn mặt được trích xuất để tạo thành một vector đặc trưng đại diện cho dạng hình học của khuôn mặt. Một trong những phương pháp thành công nhất dựa trên hình học là Mô hình ngoại quan tích cực (Active Appearance Model - AAM). Với các phương pháp dựa trên ngoại hình, các bộ lọc hình ảnh, chẳng hạn như Gabor wavelet và Mẫu nhị phân cục bộ (LBP) được áp dụng cho toàn bộ khuôn mặt hoặc các vùng cụ thể trong ảnh khuôn mặt để trích xuất một vector đặc trưng.



Hình 2.5 Sơ đồ hệ thống FER thông thường [77]

2.2.2.1 Tiền xử lý

Tiền xử lý là một trong những công đoạn quan trọng nhất, không chỉ trong các hệ thống FER mà còn đối với các hệ thống dựa trên Học máy khác. Phân tích dữ liệu thô mà không có bất kỳ sự sàng lọc nào có thể dẫn đến kết quả không mong đợi. Do vậy, việc xử lý hình ảnh nhằm cải thiện độ rõ nét của các đối tượng trong hình ảnh là việc làm quan trọng trước khi tiến hành việc rút trích các đặc trưng cần thiết. Nhìn chung, việc rút trích đặc trưng cử chỉ của khuôn mặt thường được thực hiện theo các trình tự sau:

a) Loại bỏ nhiễu ảnh

Thông thường các hình ảnh đều chứa nhiễu theo một mức độ nào đó dẫn đến làm giảm chất lượng hình ảnh như độ sắc 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 thực hiện phân tích sâu hơn trên ảnh. Có nhiều phương pháp loại bỏ nhiễu, ví dụ: Bộ lọc Wiener, FFT và Biến đổi Wavelet, Bộ lọc trung vị, Tính trung bình hình ảnh, Bộ lọc sigma, v.v.

Trung bình hình ảnh và Bộ lọc trung vị: Tính toán trung bình ảnh là cách nhanh nhất và dễ dàng để loại bỏ nhiễu và phương pháp này được sử dụng phổ biến 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 cửa sổ lọc (kích thước 3×3) quét qua lần lượt từng điểm của ảnh đầu vào. Tại vị trí mỗi điểm ảnh lấy giá trị của các điểm ảnh tương ứng trong vùng 3×3 của ảnh gốc “lấp” vào ma trận lọc. Sau đó sắp xếp các điểm ảnh trong cửa sổ này theo thứ tự (tăng dần hoặc giảm dần). Cuối cùng, gán điểm ảnh nằm chính giữa (Trung vị) của dãy giá trị điểm ảnh đã được sắp xếp ở trên cho giá trị điểm ảnh đầu ra. Cả hai kỹ thuật này đều làm mịn hình ảnh. Tuy nhiên, nó có nhược điểm là làm giảm thông tin cạnh. Như đã mô tả ở trên, điều này rất quan trọng trong việc phát hiện các đặc trưng của ảnh.

Bộ lọc Sigma: Trong [78] đã thiết kế Bộ lọc Sigma nhằm giải quyết những vấn đề còn hạn chế của Bộ lọc trung vị và trung bình. Ý tưởng của Bộ lọc Sigma là trung bình của 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 và không có bất kỳ hiện tượng mờ hay suy giảm cạnh đáng kể nào. Nó cũng có một lợi thế khác là tốc độ xử lý rất nhanh. Vì vậy, Bộ lọc Sigma được xem là phù hợp nhất để sử dụng trong nghiên cứu này.

Bộ lọc Wiener: Một cách tiếp cận khác để lọc nhiễu là Bộ lọc Wiener [79] [80]. 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 tính toán giá trị trung bình cục bộ và phương sai xung quanh mỗi pixel. Phương pháp này hoạt động tốt đối với hình ảnh có nhiễu nhiễu, nhưng hoạt động kém hiệu quả đối với hình ảnh ít nhiễu. Ngoài ra, nó cũng đưa ra một số điểm làm mờ và khó khăn về mặt tính toán nên nó không thể được sử dụng trong các ứng dụng thời gian thực.

FFT và Biến đổi Wavelet: Là hai kỹ thuật quan trọng trong xử lý tín hiệu và phân tích tín hiệu [81].

Phân tích FFT là một công nghệ rất quan trọng trong xử lý tín hiệu, xử lý hình ảnh và hoạt động bằng cách tách 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.

Biến đổi Wavelet dùng để phân tích tín hiệu nhiễu trong hệ thống truyền thông, xử lý hình ảnh để trích xuất các đặc trưng, nén dữ liệu và phân loại bất thường trong hình ảnh y khoa. Sử dụng hàm sóng Wavelet để phân rã tín hiệu thành các thành phần ở các tỷ lệ thời gian khác nhau. Điều này cho phép xác định sự biến đổi của tín hiệu không chỉ ở các tần số khác nhau mà còn ở các khoảng thời gian khác nhau.

b) Làm mịn bảo toàn cạnh của khuôn mặt

Làm mịn trong xử lý ảnh thường là cần thiết. Bằng cách làm mịn hình ảnh, người ta có thể nắm bắt các mẫu có liên quan trong khi lọc nhiễu. Theo cách đó, làm mịn có thể mang lại sự mạnh mẽ cho dữ liệu sẽ được phân tích. Có một số cách để làm mịn hình ảnh, nhưng những cách phổ biến nhất được các nhà nghiên cứu quan tâm là thông qua bộ lọc song phương [82] hoặc bộ lọc Gaussian [83].

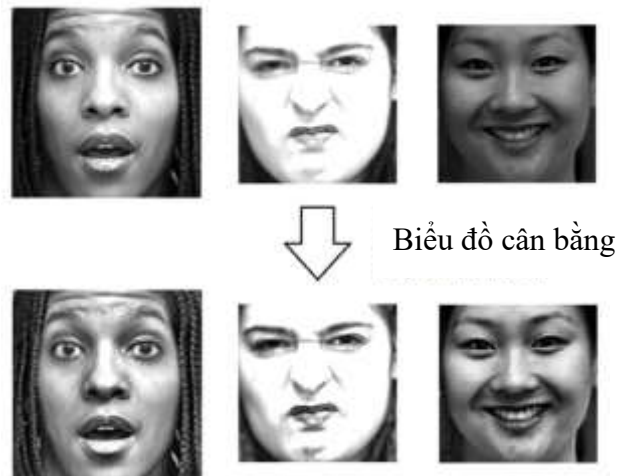
Bộ lọc song phương có hiệu quả trong việc loại bỏ nhiễu trong khi vẫn giữ cho các cạnh sắc nét, vì nó sử dụng hàm Gaussian để chỉ làm mịn các pixel lân cận và hàm cường độ Gaussian để làm mịn các pixel có cường độ tương tự như pixel trung tâm. Bằng cách đó, nó có thể bảo toàn các cạnh, vì chúng thường có các biến thể cường độ cao. Bộ lọc Gaussian có hiệu quả trong việc loại bỏ nhiễu khỏi ảnh. Nó lấy vùng lân cận xung quanh pixel trung tâm và tìm mức trung bình có trọng số Gaussian của nó.

Trong một hệ thống FER thông thường, bộ lọc song phương có ưu thế hơn khi làm mịn bề mặt, vì nó giữ cho các cạnh không bị mờ. Các cạnh trên khuôn mặt thường phân định các đặc điểm trên khuôn mặt như mắt, mũi, lông mày và miệng. Làm mịn các đặc điểm quan trọng này cho hệ thống FER có thể thu được kết quả nhận dạng không mong muốn.

c) Hiệu chỉnh ánh sáng và cải thiện đặc trưng của khuôn mặt

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 trong sáng hơn trong môi trường ánh sáng rực rỡ, trong khi những người da sáng hơn trong 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, [84] đề 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. Các thuật toán nâng cao tính năng hiện có đã được xem xét và được tóm tắt trong các đoạn sau

Biểu đồ cân bằng (Histogram Equalization - HE): HE biểu thị sự phân bố cường độ của một hình ảnh. Do các điều kiện ánh sáng tiềm ẩn khác nhau mà các khuôn mặt được trích xuất có thể xuất hiện, nên độ chính xác của nhận dạng có thể sẽ bị ảnh hưởng. Có các thuật toán dựa trên biểu đồ trong thị giác máy tính có thể kéo dài sự phân bố này, nghĩa là các vùng bị phơi sáng quá mức hoặc thiếu sáng trên khuôn mặt sẽ có cường độ đồng nhất. Phương pháp này có ưu điểm là cải thiện độ tương phản trong ảnh, làm nổi bật các đặc điểm trên khuôn mặt và giảm nhiễu do các điều kiện ánh sáng khác nhau gây ra. Trong thuật toán nhận diện khuôn mặt của Rowley [85], 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.



Hình 2.6 Biểu đồ cân bằng [77]

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 [78]. Điều này có nghĩa là điều chỉnh gamma là một cách hiệu quả để điều chỉnh độ sáng. Cả Martinkauppi [86] và Sterring [87] 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.

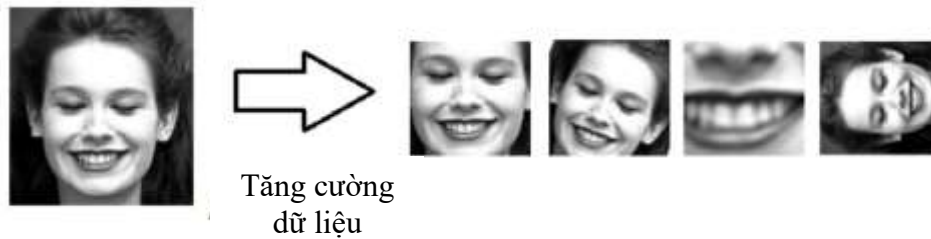
Tăng cường bằng phép toán logic Morph: Phép toán logic Morph được thiết kế để phân biệt các đối tượng ở tiền cảnh với hậu cảnh, bằng cách đo lường sự khác biệt về độ phơi sáng và độ tương phản trên hình ảnh. Thuật toán này xem xét giá trị pixel hiện tại và khu vực xung quanh của nó (được gọi là phần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

Cân bằng biểu đồ thích ứng có giới hạn tương phản: Pizer [88] đã giới thiệu Contrast-limited adaptive histogram equalization (CLAHE) nhằm nâng cao các đặc trưng trong hình ảnh dựa trên các vùng nhỏ hơn, thay vì nâng cao các tính năng theo biểu đồ của toàn bộ hình ảnh. Điều này đã được cải thiện hơn bởi Zuiderveld [89]. 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.

d) Tăng cường dữ liệu (Data Augmentation - DA)

Cơ sở dữ liệu nhận dạng cảm xúc thường nhỏ, đây là điều không mong muốn đối với các bộ phân loại. Đào tạo trên dữ liệu nhỏ có thể dẫn đến trạng bị quá mức, đây là một vấn đề phổ biến trong các mô hình Học máy. Một mô hình phù hợp hơn khi nó phân loại chính xác dữ liệu được sử dụng để huấn luyện, nhưng độ chính

xác của nó giảm đáng kể khi phân loại dữ liệu bên ngoài tập huấn luyện (khái quát hóa kém). Do đó, việc trang bị quá mức có thể được phát hiện khi huấn luyện mô hình: độ chính xác trên dữ liệu huấn luyện cao nhưng độ chính xác trên dữ liệu xác thực lại thấp hơn đáng kể. Một cách để giải quyết vấn đề của việc sử dụng cơ sở dữ liệu nhỏ để đào tạo là thông qua DA [90].



Hình 2.7 Ví dụ về tăng cường dữ liệu [77]

DA là một kỹ thuật cho phép tăng dữ liệu bằng cách sửa đổi nó một cách hợp lý. Những sửa đổi này có thể là cắt xén, lật, xoay, thu phóng, thay đổi kích thước, thay đổi độ sáng, dịch chuyển, v.v. Điều quan trọng là phải theo dõi những sửa đổi này; nếu không, phương pháp này có thể thay đổi ý nghĩa đúng đắn của các mẫu huấn luyện và gây nhầm lẫn cho bộ phân loại.

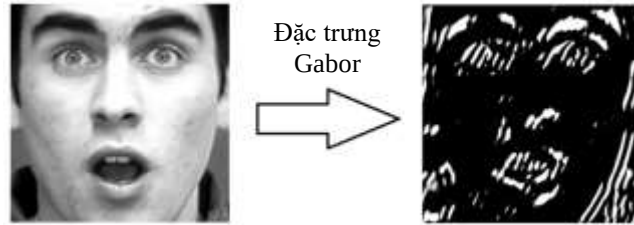
2.2.2.2 Một số cách tiếp cận trích chọn đặc trưng

Đây là quy trình để trích xuất dữ liệu hoặc thông tin hữu ích từ hình ảnh, ví dụ: giá trị, vector 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 Appearance Model (AAM), trích xuất đặc trưng Haar-like, v.v.

a) Trích xuất đặc trưng Gabor

Trích xuất đặc trưng dựa trên Gabor wavelets có những lợi thế đáng kể. Trong [91], 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. Yu và cộng sự [92] 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) [93] cung cấp một vector 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.

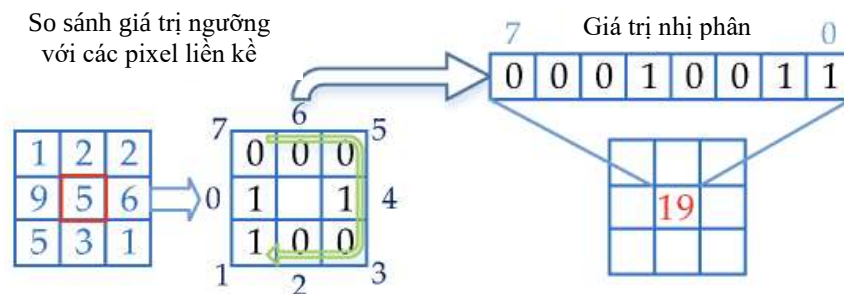


Hình 2.8 Biểu đồ đặc trưng của Gabor [77]

b) Mẫu nhị phân cục bộ (Local Binary Pattern - LBP)

Phương pháp LBP dựa trên vùng lân cận cục bộ của các pixel thường được sử dụng trong các ứng dụng nhận dạng mẫu. Một trong những điểm mạnh của LBP là nó không bị ảnh hưởng quá nhiều bởi sự thay đổi ánh sáng.

LBP [94] tính toán độ sáng liên hệ giữa mỗi pixel và những điểm lân cận của nó trong ảnh. Giá trị pixel ở giữa các khối trong hình ảnh có kích thước 3×3 được đặt làm giá trị ngưỡng và giá trị này được so sánh với các pixel liền kề riêng lẻ. Trong trường hợp khi giá trị pixel liền kề thấp hơn giá trị ngưỡng, giá trị mới của pixel liền kề được đặt là “0”, ngược lại được đặt là “1”. Sau khi quá trình này hoàn tất đối với 8 pixel liền kề, các pixel liền kề được lấy tuần tự theo hướng đã xác định (theo chiều kim đồng hồ hoặc ngược chiều kim đồng hồ) và thu được một số nhị phân gồm 8 chữ số. Số thập phân tương đương của số nhị phân này trở thành mã LBP của pixel trung tâm. Một hình ảnh LBP thu được bằng cách xử lý tất cả các pixel trong hình ảnh. Quá trình tạo ngưỡng được thể hiện như Hình 2.9.



Hình 2.9 Phương pháp mẫu nhị phân cục bộ [95]

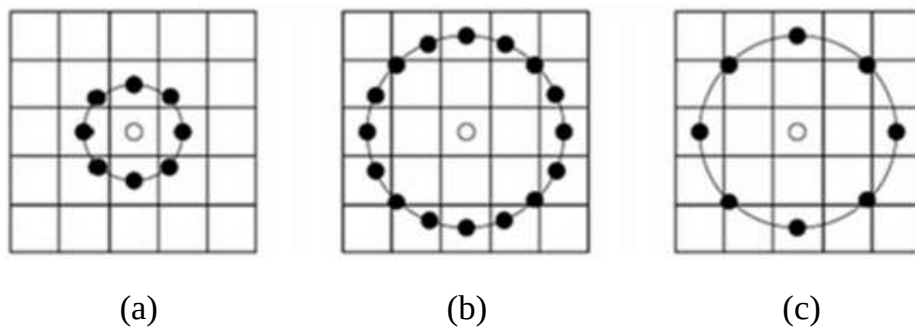
Gọi (x_m, y_m) là tọa độ của pixel ở trung tâm của hình ảnh, d_m là giá trị của pixel ở trung tâm và d_p là giá trị của pixel liền kề được xử lý, mã LBP được biểu thị như trong phương trình (2.1)

$$LBP(x_m, y_m) = \sum_{p=0}^{p-1} 2^p \times f(d_p - d_m) \quad (2.1)$$

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

Trong các hệ thống FER, LBP có khả năng làm nổi bật các đặc điểm khuôn mặt có liên quan để nhận dạng cảm xúc, chẳng hạn như lông mày, mắt, mũi và miệng. Tuy nhiên, nó có nhược điểm là nhạy cảm với nhiễu: vì là phương pháp dựa trên sự khác biệt về cường độ nên nó bị nhiễu ảnh khi xử lý một vùng có cường độ gần như đồng nhất.

Các nhà nghiên cứu quan sát thấy rằng phương pháp LBP ban đầu được xử lý với các khối 3×3 cố định đã thành công trong việc phân loại kết cấu nhưng phương pháp này không cho kết quả thành công ở các tỷ lệ khác nhau. Trong nghiên cứu của Ahonen và cộng sự [96] đã cải tiến của LBP và còn được gọi là LBP tròn, làm cho phương pháp hiện tại ổn định hơn với hai tham số mới, bán kính (r) và số pixel (n), để khắc phục vấn đề về tỷ lệ. Các ví dụ về vùng lân cận trong LBP hình tròn được minh họa như Hình 2.10.



Hình 2.10 Các ví dụ về LBP tròn. (a) $r = 1, n = 8$ (b) $r = 2, n = 16$ (c) $r = 2, n = 8$.

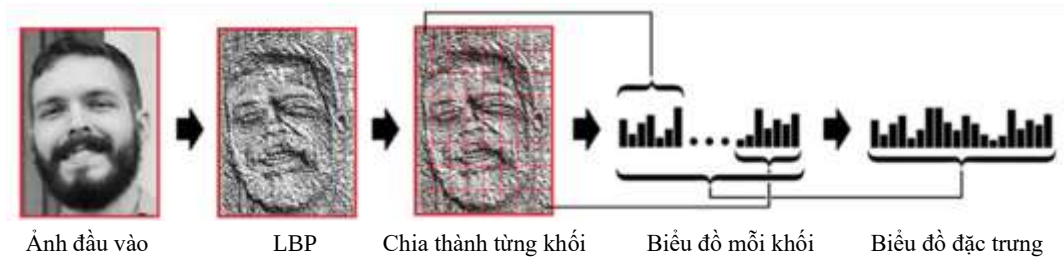
Xét rằng bán kính LBP là r , số pixel liền kề LBP là n , pixel liền kề được xử lý là p , tọa độ của pixel trung tâm là (x_m, y_m) , tọa độ của pixel liền kề được xử lý là (x_p, y_p) và $p \in n$, và tọa độ của pixel liền kề cho LBP tròn có thể được tìm thấy như trong Công thức (2.2)

$$\begin{aligned}x_p &= x_m + r \cos \frac{2\pi p}{n} \\y_p &= y_m + r \sin \frac{2\pi p}{n}\end{aligned}\tag{2.2}$$

Trong trường hợp tọa độ tìm được theo cách này không tương ứng với một pixel nhất định trên ảnh, thì có thể tìm tọa độ bằng cách áp dụng phép nội suy song tuyến tính như trong Công thức (2.3)

$$f(x, y) = \begin{bmatrix} 1-x & x \end{bmatrix} \begin{bmatrix} f(0,0) & f(0,1) \\ f(1,0) & f(1,1) \end{bmatrix} \begin{bmatrix} 1-y \\ y \end{bmatrix}\tag{2.3}$$

Sau khi tìm mã thập phân LBP cho các pixel trong mỗi khối và hình thành hình ảnh LBP mới, biểu đồ LBP cho từng khối trong hình ảnh mới được hình thành. Bằng cách thêm giá trị của từng pixel trong khối đang được xử lý, sẽ thu được biểu đồ cho khối cụ thể đó. Quá trình này được lặp lại cho tất cả các khối trong hình ảnh. Hình ảnh LBP mẫu và biểu đồ kết hợp của hình ảnh này được hiển thị trong Hình 2.11.



Hình 2.11 Trích xuất biểu đồ LBP từ hình ảnh khuôn mặt [95]

So với Gabor wavelet, hệ thống LBP yêu cầu không gian lưu trữ ít hơn nhưng hiệu quả tính toán cao hơn. Tuy nhiên, LBP không hiệu quả trên các hình ảnh có nhiễu. Kỹ thuật này có thể mất một số thông tin đặc trưng hữu ích vì nó chỉ xem xét các pixel đặc trưng của trung tâm ảnh và các vùng lân cận bỏ qua sự khác biệt về biên độ.

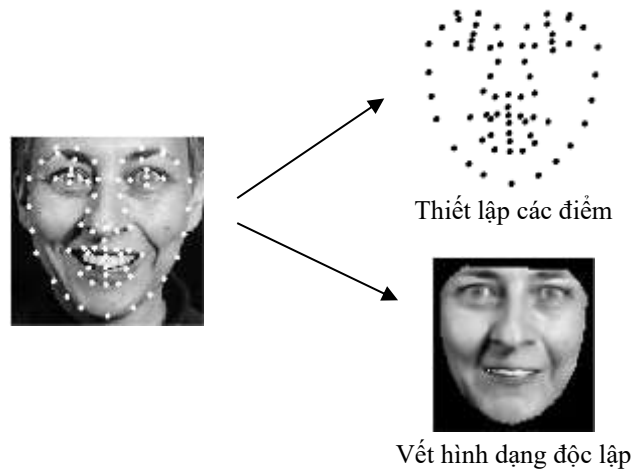
c) Active Appearance Model

Active Appearance Model là một thuật toán tối ưu trong lĩnh vực thị giác máy tính. Thuật toán này dùng để tối ưu một mô hình thống kê hình ảnh của đối tượng vào một hình 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 bộ điểm điều khiển này là một bộ các tham số mô hình thống kê đã được ước lượng mà từ đó 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 ở trong ảnh thử nghiệm.

Mô hình thống kê toán học của hình dạng được dùng để mô tả hình dạng của đối tượng trong ảnh. Đối tượng hình học được biểu diễn bởi n điểm trong không gian d chiều, và ta thể hiện nó bằng một vector $n \cdot d$ 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 như sau: $x = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)^T$.

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 bằng phương pháp phân tích thống kê. 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, một công cụ toán học tuyến tính rất mạnh cho phép thực hiện việc chiếu một mẫu và khôi phục mẫu từ hình chiếu [22].



Hình 2.12 Đối tượng hình dạng và kết cấu hình ảnh [22]

Sau khi đã mô hình hóa được 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, như được thể hiện trong Hình 2.12.

Quá trình lấy ra được dữ liệu kết cấu hình ảnh này được thực hiện qua hai bước: tam giác hóa và lấy dữ liệu thông qua phép biến đổi Piece-wise Affine. Mỗi mẫu kết cấu hình ảnh lấy được đều có chứa ít nhiều những thông tin 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, cần thực hiện chuẩn hóa các mẫu kết cấu hình ảnh. 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 Luận án 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 [22]. 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:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix} \quad (2.4)$$

Trong đó: W_s là ma trận đường chéo thể hiện trọng số của mỗi tham số hình dạng tương ứng; b_s mô tả các thông tin cho hình dạng đối tượng là tọa độ các điểm điều khiển; b_g thể hiện thông tin về kết cấu hình ảnh đối tượng là cường độ của vùng ảnh của đối tượng;

Nói chung, các kết quả của việc tổng hợp mẫu và kết quả của các giải thuật tìm kiếm dựa trên mô hình này sẽ tương đối nhạy cảm với việc lựa chọn W_s .

Thực hiện tính toán dựa trên PCA, ta có được mô hình như sau: $b = P_c c$, với P_c là các vector riêng và c là vector tham số biểu diễn hình ảnh, nó thể hiện cả hai thông tin về hình dạng và kết cấu hình ảnh. Do tính chất tuyến tính của công thức sẽ cho phép ta biểu diễn trực tiếp từ c :

$$x = \bar{x} + P_s W_s^{-1} P_{cs} c, \quad g = \bar{g} + P_g P_{cg} c \quad (2.5)$$

P_{cs} và P_{cg} trong phương trình này là các ma trận hiệp phương sai. Ma trận P_{cs} cho biết mức độ sai lệch giữa trạng thái thực tế và dự báo, trong khi ma trận P_{cg} cho biết mức độ sai lệch giữa hàm chuyển đổi dự báo và hàm chuyển đổi thực tế.

Trong đó $P_c = \begin{pmatrix} P_{cs} \\ P_{cg} \end{pmatrix}$, c là vector điều khiển sự xuất hiện cả hình dạng và kết cấu của hình ảnh. Biểu thức sau cho phép khôi phục lại tham số xuất hiện c từ mẫu đã cho: $c = P_c^T b$

Hoặc có thể biểu diễn ngắn gọn

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \quad (2.6)$$

Trong đó: $Q_s = P_s W_s^{-1} P_{cs}$

$$Q_g = P_g P_{cg} \quad (2.7)$$

Như vậy, với 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 kết cấu hình ảnh của khuôn mặt;
- 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 vector hình dạng tương ứng.

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. 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:

$$\delta I = I_i - I_m \quad (2.8)$$

Trong đó:

I_i là vector biểu diễn cường độ ảnh của đối tượng trong ảnh đầu vào;

I_m là vector biểu diễn cường độ ảnh của hình 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:

$$\begin{aligned} x &= \bar{x} + Q_s c \\ g &= \bar{g} + Q_g c \end{aligned} \quad (2.9)$$

Trên những cơ sở đó, ta xây dựng thuật toán dựa trên độ đo:

$$E(p) = r^T r \quad (2.10)$$

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;

p là tham số cần được ước lượng $p^T = (c^T \mid t^T \mid u^T)$ với c là tham số của mô hình kết hợp, $t = (s_x, s_y, t_x, t_y)^T$ là tham số dịch chuyển biểu diễn vị trí của tập điểm điều khiển trong khung hình là tham số biến đổi kết cấu hình ảnh;

$$g_{im} = T_u(\bar{g} + p_g b_g) = (1 + u_1)(\bar{g} + p_g b_g) + u_2 1 \quad (2.11)$$

Thực hiện khai triển Taylor một lần ta được $r(p + \delta p) = r(p) + \frac{\partial r}{\partial p} \delta p$ với phần tử thứ i của ma trận $\frac{\partial r}{\partial p}$ là $\frac{dr_i}{dp_i}$

Ma trận Jacobi $\frac{\partial r}{\partial p}$ được ước lượng trước từ một tập mẫu. Các thành phần của ma trận được ∂p 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 được xác định trước tương ứng

Thuật toán tối ưu như sau:

Tính trước:

- Bước 1. Tính ma trận R
- Bước 2. Khởi tạo giá trị ban đầu cho x
- Bước 3. Khởi tạo mảng tham số $K = \{1, 0.5, 0.25, 0.125, 0.0625\}$

Các bước lặp:

- Bước 1. Tính vector độ lệch r
- Bước 2. Tính $E = r^T r$
- Bước 3. Tính $\partial p = Rr(p)$
- Bước 4. Với mỗi $k \in K$, cập nhật
- Bước 5. Tính r'
- Bước 6. Tính $E' = r'^T r'$
- Bước 7. 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.

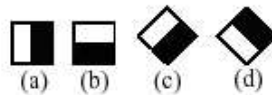
d) Trích xuất đặc trưng Haar-like

Đặc trưng Haar-like [23] là sự kết hợp của các đặc điểm cạnh, tuyến tính, tâm và đường chéo. Vì chức năng trích xuất đặc trưng này đều tuân theo các bước do Alfred Haar đề xuất nên chúng được gọi là Haar-like. Các mẫu đặc trưng được chia thành hai vùng hình chữ nhật, nghĩa là trắng và đen, và giá trị đặc trưng của mẫu được định nghĩa là sự khác biệt giữa giá trị của pixel hình chữ nhật màu trắng và pixel hình chữ nhật màu đen. Giá trị riêng Haar phản ánh sự thay đổi thang độ xám của hình ảnh.

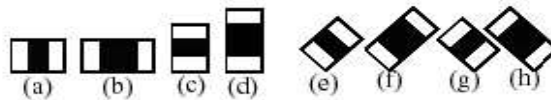


Hình 2.13 Đặc trưng cơ bản Haar-like

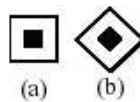
Để sử dụng các đặc trưng này vào việc xác định khuôn mặt người, 4 đặc trưng Haar-Like cơ bản được mở rộng ra và được chia làm 3 tập đặc trưng sau:



(i) Đặc trưng cạnh – edge features



(ii) Đặc trưng đường – line features



(iii) Đặc trưng xung quanh tâm - center-surround features

Hình 2.14 Đặc trưng mở rộng của Haar-like

Với các đặc trưng trên, ta có thể tính được các giá trị của đặc trưng Haar-Like là sự chênh lệch giữa tổng các pixel của vùng đen và vùng trắng như sau:

$$f(x) = \text{Tổng}_{\text{vùng đen}}(\text{Các mức xám của pixel}) - \text{Tổng}_{\text{vùng trắng}}(\text{Các mức xám của pixel})$$

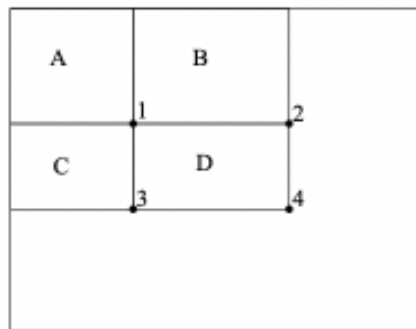
Sử dụng giá trị này so sánh với các giá trị của pixel thô, các đặc trưng Haar-like có thể tăng / giảm sự thay đổi in-class / out-of-class (bên trong hay bên ngoài lớp khuôn mặt người), do đó sẽ làm cho bộ phân loại dễ hơn.

Như vậy, để tính các giá trị của đặc trưng Haar-like, ta phải tính tổng của các vùng pixel trên ảnh. Nhưng để tính các giá trị của các đặc trưng Haar-like cho tất cả các vị trí trên ảnh đòi hỏi chi phí tính toán khá lớn, không đáp ứng được cho các hệ thống đòi hỏi tính run-time. Do đó [23] đưa ra thêm một khái niệm là *Integral Image*, là một mảng hai chiều với kích thước bằng với kích thước của hình ảnh cần tính đặc trưng Haar-like. Với mỗi phần tử của mảng này được tính bằng cách tính tổng của điểm ảnh phía trên (dòng – 1) và bên trái cột (cột – 1) của nó. Bắt đầu từ vị trí trên, bên trái đến vị trí dưới phải của ảnh. Việc tính toán này chỉ là phép cộng số nguyên, do đó tốc độ thực hiện rất nhanh.

$$p(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (2.12)$$

Sau khi đã tính được *Integral Image*, việc tính tổng các giá trị mức xám của một vùng bất kỳ nào đó trên ảnh thực hiện theo cách sau

Tổng giá trị mức xám của vùng D như trong Hình 2.15 được tính toán với bốn tham chiếu mảng. Giá trị *Integral Image* tại vị trí 1 là tổng các pixel trong hình chữ nhật A, tại vị trí 2 là A + B, tại vị trí 3 là A + C và tại vị trí 4 là A + B + C + D. Tổng bên trong vùng D được tính là $4 + 1 - (2 + 3)$



Hình 2.15 Tính tổng pixel trong vùng ảnh [23]

2.2.2.3 Phân loại / Hồi quy

Một mô hình phân loại chịu trách nhiệm dự đoán một nhãn nhất định cho một hình ảnh hoặc tính năng đầu vào. Một mô hình hồi quy có nhiệm vụ xác định mối quan hệ giữa một biến phụ thuộc với các biến độc lập. Các thuật toán phân loại và hồi quy được sử dụng nhiều nhất như: Mạng nơ-ron tích chập (CNN), Máy Vector hỗ trợ (Support Vector Machine - SVM), K-láng giềng gần nhất (K-Nearest

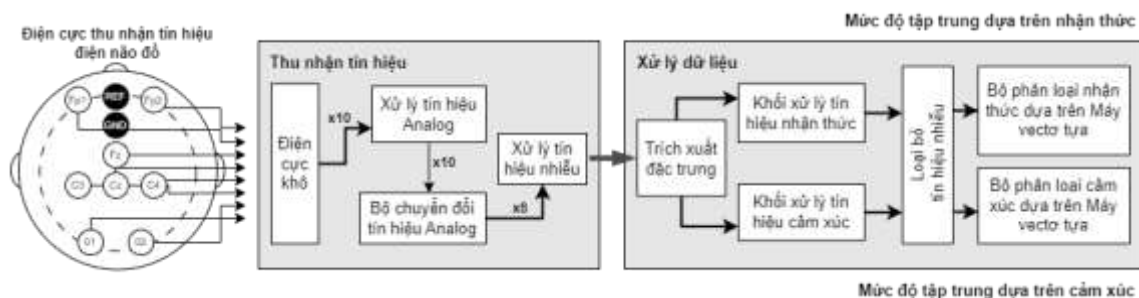
Neighbor - KNN) [97], Cây quyết định (Decision Tree - DT) [98], Rừng ngẫu nhiên (Random Forest - RF) [99], Mô hình Markov ẩn (Hidden Markov Model – HMM) [100]

2.2.3 Một số cách tiếp cận trong đánh giá sự tập trung của người học dựa vào biểu cảm khuôn mặt

Trong lĩnh vực giáo dục, phân tích cảm xúc của người học đã thu hút được sự quan tâm lớn trong nhiều thập kỷ qua vì nó gắn liền với hiệu quả học tập và kết quả giáo dục. Việc đánh giá trạng thái tham gia của người học dựa trên cảm xúc của người học sẽ giúp ích rất nhiều cho giáo viên. Do đó, việc tự động nhận dạng trạng thái cảm xúc của người học trong quá trình học tập là điều đáng mong đợi. Phân tích nét mặt là một trong những phương pháp đánh giá tự động các trạng thái cảm xúc của người học trong hoạt động giáo dục một cách hiệu quả. Nhiều nhà khoa học đã đưa ra nhiều phương pháp để đánh giá sự tập trung của người học dựa trên cảm xúc, gồm có 2 hướng tiếp cận: xâm nhập và không xâm nhập.

Phương pháp xâm nhập: người học được gắn trực tiếp các thiết bị để đo các tín hiệu sinh lý như điện não đồ (EEG) [46]

Kiến trúc của hệ thống được đề xuất được mô tả trong Hình 2.16. Tám Điện cực Active Dry Electrodes thu tín hiệu Điện não đồ (EEG) trực tiếp từ da đầu. Sau khi tải nạp, các tín hiệu liên tục (hay tín hiệu Analog) được điều chỉnh bởi giao diện người dùng. Tiếp theo, các tín hiệu này được số hóa bằng bộ chuyển đổi kỹ thuật số (Analog Digital Converter - ADC). Sau đó, các tín hiệu được gửi bằng đường truyền Bluetooth không dây đến giai đoạn xử lý dữ liệu (Data Processing). Ở đây, đặc trưng phù hợp được trích xuất bởi bộ lọc (Filter Bank). Hai bộ phân loại Máy vectơ hỗ trợ (SVM) đã được huấn luyện nhận mảng đặc trưng từ hai khối xử lý Mẫu không gian chung (Common Spatial Pattern - CSP) để phát hiện nhận thức và cảm xúc tương ứng [46].



Hình 2.16 Kiến trúc hệ thống đánh mức độ tham gia của người học dựa trên tín hiệu EEG [46]

không gian Euclide n chiều R^n ; (2) Sau khi dạng được mã hóa, chúng ta cần lựa chọn các dấu hiệu đặc trưng điển hình, loại bỏ các dấu hiệu phụ để giảm bớt kích thước của dạng cũng như mức độ phức tạp trong tính toán; (3) Đánh giá các thông số đặc trưng để có thể xác định dạng của đối tượng trước khi tiến hành so khớp để nhận dạng chính xác.

Một hệ nhận dạng được cấu thành từ hai bộ phận chính: (1) Khối cảm biến, là một bộ phận thiết bị dùng để chuyển các đặc trưng vật lý thành một bộ dấu hiệu đặc trưng cho dạng; tức là dùng để mã hóa dạng; và (2) Khối phân lớp, là thiết bị dùng để tiếp nhận các dạng đã được mã hóa từ khối cảm biến phân hoạch thành từng lớp xác định cùng với những dạng mẫu tương ứng được lưu trữ vào bộ nhớ nhận dạng mới và xếp lớp các lớp mới.

2.3.2.1 Các khái niệm cơ sở

a) Không gian biểu diễn dạng

“Không gian biểu diễn dạng” (Space of Representation) X được định nghĩa là tập hợp các dạng có thể có (có thể là hữu hạn hoặc vô hạn) của đối tượng. Giả sử có m dạng và dạng thứ i ($i = 1, \dots, m$), được ký hiệu là A_i được biểu diễn bằng một bộ n dấu hiệu đặc trưng $A_i = (x_1, x_2, \dots, x_n)$; trong đó, x_j ($j = 1, \dots, n$) thể hiện một đặc trưng j của A_i . Khi đó, không gian X được biểu diễn như sau:

$$X = \{A_1, A_2, \dots, A_m\} \quad (2.13)$$

b) Không gian diễn dịch

Tập hợp các tên gọi của các dạng được xác định trong không gian biểu diễn dạng X được gọi là “không gian diễn dịch” (Space of Interpretation), ký hiệu là Ω . Giả sử có m dạng được xác định trong X ; đặt T_i ($i=1, \dots, m$) là tên gọi của dạng thứ i . Khi đó, không gian diễn dịch n được biểu diễn như sau: $\Omega = \{T_1, T_2, \dots, T_m\}$.

Từ hai khái niệm “Không gian biểu diễn dạng” và “Không gian diễn dịch”, ta có thể biểu diễn toán học của bài toán nhận dạng như sau: $f: X \rightarrow \Omega$

Trong đó, ánh xạ f là tập hợp các quy tắc, tiêu chuẩn để xác định một phần tử thuộc X ứng với một phần tử Ω .

c) Xác suất tiên nghiệm và Mật độ phân phối xác suất

Giả sử tập tất cả các lớp $B \subset \mathcal{R}^n$ đã được phân hoạch thành k lớp xác định

$B_1, B_2, \dots, B_k$ và chúng ta cần phân lớp cho một dạng x nào đó. Khi đó, xác suất xuất hiện dạng x trong tập dạng B được ký hiệu là $p(B)$; và xác suất xuất hiện x trong lớp B_i , ký hiệu là $p(B_i)$, được gọi là “xác suất tiên nghiệm” của B_i ($i=1, \dots, k$) (tức là xác suất đã được biết trước). Trong trường hợp đặc biệt, nếu các lớp có đồng khả năng xuất hiện thì xác suất tiên nghiệm được xác định bằng $p(B_i) = 1/k \forall i \in [1, k]$.

Mật độ phân phối xác suất dạng x vào lớp B_i là một xác suất có điều kiện để phân lớp dạng x đúng vào lớp B_i và được ký hiệu là $p(B_i / x)$.

2.3.2.2 Các phương pháp học trong nhận dạng

Trong việc lựa chọn để biểu diễn dạng, dạng có thể được xác định theo cách định lượng (mô hình tham số) hay định tính (mô hình cấu trúc). Khi dạng đã được xác định, quá trình nhận dạng chuyển sang giai đoạn thứ hai - giai đoạn học (*Learning*). Học là giai đoạn cung cấp tri thức cho hệ thống. Mục đích học nhằm cải thiện, điều chỉnh việc phân loại tập dạng thành các lớp. Nhận dạng là tìm ra quy luật và các thuật toán để có thể gán đối tượng vào một lớp; hay nói cách khác, nhận dạng là xác định cho đối tượng một tên.

a) Học có giám sát (Supervised learning)

Đây là kỹ thuật phân loại nhờ kiến thức biết trước được thể hiện qua một thư viện các mẫu chuẩn. Mẫu cần nhận dạng được đem so sánh với mẫu chuẩn để xem nó thuộc loại nào. Vấn đề chủ yếu là thiết kế một hệ thống để có thể đối sánh đối tượng với mẫu chuẩn và quyết định đặt cho chúng vào một lớp. Việc đối sánh này nhờ vào các thủ tục ra quyết định dựa trên một công cụ gọi là hàm phân lớp hoặc hàm ra quyết định. Các phương pháp thuộc nhóm này gồm có: Hệ thống vector hỗ trợ (Support Vector Machine - SVM); K láng giềng gần nhất (K Nearest Neighbours - KNN); Xác suất Bayes cổ điển (Naive Bayes - NB); Cây quyết định (Decision Tree - DT); Mạng nơron (Neural Network); Vector trọng tâm (Centroid-base vector); hay tuyến tính bình phương nhỏ nhất (Linear Least Square Fit), v.v.

b) Học không có giám sát (Unsupervised learning)

Kỹ thuật này phải tự định ra các lớp khác nhau và xác định các tham số đặc trưng cho từng lớp; do đó, kỹ thuật này gặp khó khăn hơn so với kỹ thuật học có giám sát bởi vì chúng ta không biết trước số lớp cũng như các đặc trưng của các lớp này. Kỹ thuật này được tiến hành bằng cách mọi cách gộp nhóm có thể và chọn lựa cách tốt nhất. Bắt đầu từ tập dữ liệu, nhiều thủ tục xử lý khác nhau nhằm phân

lớp và nâng cấp dần để đạt được một phương án phân loại tốt nhất. Một số phương pháp thuộc nhóm học này gồm có K-means, HAC (Hierarchical Agglomerative Clustering), DBSCAN, SOM (Self-Organizing Map) v.v...

c) Học bán giám sát (Semi-supervised learning)

Kỹ thuật này kết hợp những ưu điểm của hai phương pháp học trước (học có giám sát và học không có giám sát) nhằm tăng độ chính xác cho quá trình nhận dạng; tức là học dựa trên cơ sở dữ liệu trong thư viện mẫu chuẩn và những mẫu chưa biết trước. Theo đó, kỹ thuật này dùng một thuật toán để học các dạng trong mẫu chuẩn, sau đó tiến hành định nhãn cho một số dạng mới sau quá trình có lựa chọn một cách hợp lý, có đánh giá chất lượng công việc hay độ chính xác. Tiếp theo, chuyển những dạng mới có độ tin cậy cao theo một ngưỡng qui định và được gắn nhãn này vào thư viện mẫu chuẩn để làm cơ sở cho việc đối sánh, nhận dạng tiếp theo.

2.3.3 Phương pháp đánh giá

Để đánh giá hiệu suất mô hình thực nghiệm thì cần dùng công thức như là độ chính xác (*Precision*), độ phản hồi (*Recall*) và điểm số *F1* (*F1-Score*). Giá trị tính toán cho các thước đo này dựa vào ma trận nhầm lẫn mà mô hình nhận dạng được.

Bảng 2.4 Ma trận nhầm lẫn tổng quát

		Mô hình dự đoán	
		Tích cực (Positive)	Tiêu cực (Negative)
Giá trị thực tế	Tích cực Positive	True Positive (TP) Mô hình dự đoán đúng	False Negative (FN) Mô hình dự đoán nhầm giá trị đúng thành sai
	Tiêu cực Negative	False Positive (FP) Mô hình dự đoán nhầm giá trị sai thành đúng	True Negative (TN) Mô hình dự đoán đúng

Độ chính xác (*Precision*) – là mức độ gần của các phép đo, có giá trị gần với 1 khi kết quả là một tập phân loại tốt. Precision là 1 chỉ khi tử số và mẫu số bằng nhau ($TP = TP + FP$), điều này cũng có nghĩa là FP bằng 0. Khi giá trị của FP tăng dần đến mẫu số lớn hơn tử số cũng đồng nghĩa với việc độ chính xác giảm. Công thức tính *Precision* được định nghĩa như (2.14)

$$Precision = \frac{TP}{TP + FP} \quad (2.14)$$

Độ phản hồi (*Recall*) – càng gần với 1 sẽ cho một phân loại tốt. Recall là 1 chỉ khi tử số và mẫu số bằng nhau ($TP = TP + FN$), điều này cũng có nghĩa là FN bằng 0. Khi giá trị FN tăng dẫn đến giá trị của Recall giảm. Chỉ số này nhằm đánh giá mức độ tổng quát hóa mô hình tìm được và được xác định theo công thức (2.15)

$$Recall = \frac{TP}{TP + FN} \quad (2.15)$$

Như vậy, hai thước đo *Precision* và *Recall* càng cao thì kết quả phân loại càng tốt. Tuy nhiên trong thực tế, nếu ta điều chỉnh mô hình để tăng *Recall* quá mức có thể dẫn đến *Precision* giảm và ngược lại, điều chỉnh mô hình để tăng *Precision* có thể làm giảm *Recall*. Như vậy để cân bằng hai giá trị này, ta dùng đến *F1-Score* và được xác định theo công thức (2.16)

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (2.16)$$

2.3.4 Kỹ thuật phát hiện khuôn mặt

Việc phát hiện khuôn mặt được thực hiện dựa trên ý tưởng cơ bản là phân lớp nhị phân với mỗi vùng hình quan tâm, từ đó kết luận vùng ảnh đó là khuôn mặt hay không. Việc phát hiện khuôn mặt được tiến hành thông qua hai pha chính. Pha thứ nhất là pha học; tức là từ một tập các ảnh khuôn mặt và các ảnh không phải khuôn mặt để tạo ra một bộ dữ liệu học mô hình khuôn mặt mẫu. Pha thứ hai là pha phát hiện; việc phát hiện khuôn mặt là việc dò tìm xem một vùng bất kỳ trong ảnh có thỏa mãn các đặc điểm của mô hình đã được xây dựng hay không [22].

a) Mô hình nhận dạng khuôn mặt mẫu: Mỗi cây được xây dựng dựa trên tập dữ liệu huấn luyện có cấu trúc như sau:

$$\{(I_s, v_s, w_s): s = 1, 2, \dots, S\} \quad (2.17)$$

Trong đó, v_s là giá trị nhãn đúng của ảnh I_s và w_s là trọng số tương ứng. 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 như sau:

$$WMSE(I, v, w) = \sum_{(I, v, w) \in C_0} w \cdot (v - \bar{v}_0)^2 + \sum_{(I, v, w) \in C_1} w \cdot (v - \bar{v}_1)^2 \quad (2.18)$$

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 v_0 và v_1 là trung bình các giá trị nhãn trong C_0 và C_1 . Thuật toán huấn luyện của kỹ thuật tham khảo ở Phụ lục 2, chỉ mục 1.

b) Phép so sánh cường độ điểm ảnh

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

$$B(I: l_1 l_2) = \begin{cases} 0, & I(l_1) \leq I(l_2) \\ 1, & otherwise \end{cases} \quad (2.19)$$

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 $[-1, +1] \cdot [-1, +1]$. 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.

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.

c) Giá trị trung bình cục bộ

$$B(I: R(x, y, w, h), \delta) = \begin{cases} 1, & \frac{1}{w * h} \sum_{x \leq i < x+w} \sum_{y \leq j < y+h} I(i, j) < \delta \\ x, & otherwise \end{cases} \quad (2.20)$$

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 .

Kết quả thử nghiệm được tác giả công bố là đạt 92,44% cho thấy phương pháp được đề xuất cho độ chính xác phát hiện khuôn mặt khá tốt. Tốc độ xử lý nhanh, đây là kỹ thuật phù hợp phát triển những ứng dụng xử lý khuôn mặt yêu cầu thời gian thực trên dữ liệu video.

2.3.5 Kỹ thuật xử lý ảnh số

Trước đây, khi mà công nghệ kỹ thuật số chưa phát triển thì các kỹ thuật xử lý ảnh chủ yếu được sử dụng để tạo cảm giác về sự gia tăng chất lượng ảnh quang học trong mắt người quan sát. Với sự phát triển mạnh mẽ trong lĩnh vực công nghệ thông tin ngày nay, công nghệ xử lý ảnh được ứng dụng một cách rộng rãi trong hầu hết mọi lĩnh vực khoa học kỹ thuật số. Chẳng hạn như: (1) trong y học, công nghệ xử lý ảnh giúp biến đổi hình ảnh được tạo ra từ nguồn bức xạ X-ray hay nguồn bức xạ siêu âm thành hình ảnh quang học trên bề mặt film X-quang hoặc trực tiếp trên bề mặt màn hình hiển thị; hay là hình ảnh của các cơ quan chức năng của con người có thể được xử lý theo hướng nâng cao độ tương phản, lọc, tách các thành phần cần thiết trong kỹ thuật chụp cắt lớp hoặc tạo ra hình ảnh trong không gian ba chiều trong kỹ thuật siêu âm hiện đại; (2) Trong lĩnh vực địa chất thì công nghệ xử lý ảnh giúp nhà khoa học xác định cấu trúc bề mặt trái đất thông qua hình ảnh nhận được từ vệ tinh. Và với kỹ thuật làm nổi đường biên (edge detection) và khôi phục hình ảnh (image restoration), công nghệ xử lý ảnh còn cho phép nâng cao chất lượng ảnh vệ tinh và tạo ra các bản đồ địa hình 3D với độ chính xác cao. Hay trong ngành khí tượng học, việc dự báo thời tiết có thể được thực hiện một cách chính xác hơn thông qua công nghệ xử lý hình ảnh nhận được từ hệ thống vệ tinh theo dõi thời tiết nhằm nâng cao chất lượng và ghép hình để tạo ra ảnh bề mặt trái đất trên một vùng rộng lớn; (3) Đặc biệt, công nghệ xử lý ảnh số được sử dụng rộng rãi trong lĩnh vực hình sự, an ninh và các hệ thống bảo mật hoặc kiểm soát đối tượng; chẳng hạn như ứng dụng xử lý ảnh trong công tác nhận dạng vân tay, nhận dạng khuôn mặt để nhanh chóng phát hiện các đối tượng nghi vấn, nâng cao hiệu quả hệ thống bảo mật cá nhân cũng như kiểm soát ra vào, phòng chống tội phạm quốc tế, sự lan truyền dịch bệnh qua hình ảnh thân nhiệt của người đi qua các thiết bị đo ở các sân bay quốc tế, v.v..., và (4) Đặc biệt, kỹ thuật xử lý ảnh cũng được ứng dụng triệt để trong kỹ thuật quân sự thông qua việc tự động nhận dạng mục tiêu quân sự, hay trong các lĩnh vực khác như máy nhìn công nghiệp, các hệ

thống điều khiển tự động, nén ảnh tĩnh, ảnh động để lưu trữ và truyền trong mạng viễn thông, v.v...

Trong những năm gần đây, với sự phát triển của công nghệ kỹ thuật số, hình ảnh kỹ thuật số đã trở nên ngày càng phổ biến trong đời sống xã hội; theo đó, việc thu nhận ảnh và lưu trữ ảnh trên máy tính trở nên ngày càng đơn giản. Và việc xử lý ảnh kỹ thuật số trở thành một lĩnh vực được nhiều học giả quan tâm và nghiên cứu. Trong khoa học kỹ thuật thị giác máy tính, xử lý ảnh số là một trong những mảng quan trọng nhất bởi vì nó là tiền đề cho nhiều nghiên cứu ứng dụng trong thực tiễn. Việc xử lý ảnh số cần đảm bảo hai nhiệm vụ cơ bản: (1) nâng cao chất lượng thông tin hình ảnh, và (2) xử lý số liệu cung cấp cho các quá trình khác trong đó có việc ứng dụng thị giác vào điều khiển.

Quá trình xử lý ảnh bắt đầu từ việc thu nhận ảnh nguồn từ các thiết bị thu nhận ảnh dạng số hoặc tương tự được gửi đến máy tính. Tuy nhiên, ảnh nguồn cần được lưu trữ ở định dạng thích hợp để có thể sử dụng các thuật toán tương ứng tác động lên dữ liệu ảnh nhằm thay đổi cấu trúc ảnh phù hợp với các ứng dụng khác nhau. Do đó, việc chọn định dạng của hình ảnh nguồn là bước quan trọng đầu tiên trong quá trình xử lý ảnh số nhằm cho ra kết quả mong muốn, cụ thể là một ảnh “đã được xử lý” hoặc một kết luận nào đó dựa trên hình ảnh được xử lý.

2.3.6 Thuật toán GentleBoost

Thuật toán Gentleboost được phát triển dựa trên nền tảng Adaboost. Tư tưởng của thuật toán Adaboost là kết hợp các bộ phân loại yếu thành một bộ phân loại mạnh. Friedman [102] đề xuất thuật toán Gentle AdaBoost (viết tắt là GentleBoost) và thuật toán này có nhiều ưu điểm như đơn giản, ổn định và cho kết quả phân loại tốt trong nhiều ứng dụng. Phương pháp GentleBoost được mô tả tóm tắt như sau:

Cho tập dữ liệu huấn luyện bao gồm M mẫu: $(x_1, y_1), \dots, (x_M, y_M)$ với x_i là vectơ các đặc trưng và y_i là nhãn phân loại nhận giá trị $y_i = +1$ hoặc $y_i = -1$. Bộ phân loại mạnh $F(x)$ được tạo thành bằng cách tổ hợp tuyến tính các bộ phân loại yếu:

$$F(x) = \sum_{k=1}^n f_k(x) \quad (2.21)$$

trong đó: x là vectơ đặc trưng đầu vào, $f_k(x)$ là bộ phân loại yếu có khả năng dự đoán nhãn phân loại cho vectơ đầu vào x , n là số vòng lặp. Kết quả phân loại cuối cùng được tạo ra bằng cách tính $sign[F(x)]$.

Thuật toán bao gồm n vòng lặp. Tại vòng lặp thứ k , các mẫu huấn luyện được đánh trọng số lại sao cho những mẫu bị phân loại sai trong vòng lặp trước nhận được trọng số cao hơn; và do vậy, được bộ phân loại chú ý hơn. Bộ phân loại $f_k(x)$ được huấn luyện trên dữ liệu có trọng số trong vòng thứ k . Thuật toán GentleBoost được thể hiện như sau:

Đầu vào:

- Cho tập dữ liệu huấn luyện bao gồm M mẫu: $(x_1, y_1), \dots, (x_M, y_M)$ với x_i là vectơ các đặc trưng và y_i là nhãn phân loại nhận giá trị $y_i = +1$ hoặc $y_i = -1$.

- n : Số vòng lặp

Đầu ra:

$$sign[F(x)] = sign\left[\sum_{k=1}^n f_k(x)\right] \quad (2.22)$$

Các bước thực hiện:

Bước 1. Khởi tạo các trọng số $w_i = 1/M$, ($i = 1, \dots, M$), w_i là trọng số của mẫu huấn luyện thứ i ;

Khởi tạo $F(x) = 0$;

Bước 2. Lặp với $k = 1, 2, \dots, n$

Bước 2.1. Huấn luyện $f_k(x)$ sử dụng dữ liệu huấn luyện có trọng số;

Bước 2.2. Cập nhật $F(x) \leftarrow F(x) + f_k(x)$;

Bước 2.3. Cập nhật $w_i \leftarrow w_i e^{-y_i f_k(x_i)}$; ($i=1,2,\dots,M$)

Bước 3. Trả về bộ phân loại

$$sign[F(x)] = sign\left[\sum_{k=1}^n f_k(x)\right] \quad (2.23)$$

Cụ thể, tại bước 2.1 của mỗi vòng lặp, thuật toán lựa chọn $f_k(x)$ sao cho sai số phân loại nhỏ nhất:

$$\varepsilon = \sum_{i=1}^M w_i (y_i - f_k(x_i))^2 \quad (2.24)$$

Để tìm được bộ phân loại cho phép cực tiểu hoá ε , cần xây dựng bộ phân loại yếu $f_k(x)$ cho phép cực tiểu hoá bình phương lỗi phân loại có tính tới trọng số. Có nhiều phương pháp xây dựng bộ phân loại yếu khác nhau. Ở đây, bộ phân loại yếu được sử dụng là gốc quyết định. Gốc quyết định là phiên bản đơn giản của cây quyết định với một nút duy nhất. Gốc quyết định lựa chọn một đặc trưng của mẫu huấn luyện, sau đó tùy thuộc vào giá trị của đặc trưng để gán cho nhãn giá trị +1 hay -1.

$$f_k(x) = a^{\text{TM}}(x^f > t) + b^{\text{TM}}(x^f \leq t) \quad (2.25)$$

trong đó $\delta(e) = 1$ nếu e đúng và $\delta(e) = 0$ nếu ngược lại, t là một giá trị ngưỡng, a và b là tham số, x^f là giá trị đặc trưng thứ f của vector x . Trong trường hợp dữ liệu đánh giá chỉ bao gồm giá trị 1 và 0 hoặc +1 và -1, có thể chọn ngưỡng $t = 0$. Như vậy, ngoài việc phân loại, gốc quyết định còn thực hiện trích chọn đặc trưng, mỗi gốc chỉ chọn một đặc trưng duy nhất.

Quá trình huấn luyện để chọn ra gốc tốt nhất được thực hiện bằng cách thử tất cả đặc trưng f . Với mỗi giá trị của f , giá trị tối ưu của a và b được ước lượng theo kỹ thuật bình phương tối thiểu (*Least Square Estimation*) mà bản chất là tính giá trị tham số tại điểm có đạo hàm bằng 0.

$$a = \frac{\sum_i w_i y_i \delta(x^f > 0)}{\sum_i w_i \delta(x^f > 0)} \quad b = \frac{\sum_i w_i y_i \delta(x^f \leq 0)}{\sum_i w_i \delta(x^f \leq 0)} \quad (2.26)$$

Giá trị f , a và b tính được cho sai số ε nhỏ nhất được chọn để tạo ra bộ phân loại $f_k(x)$ cho vòng lặp thứ k . Bộ phân loại yếu $f_k(x)$ sau đó được cộng thêm vào bộ phân loại mạnh $F(x)$ tại bước 2.2.

Tại bước 2.3, thuật toán tiến hành cập nhật lại trọng số $w_i \leftarrow w_i e^{-y_i f_k(x_i)}$; Các mẫu phân loại sai có $y_i f_k(x) < 0$ được GentleBoost tăng trọng số; các mẫu phân loại đúng có $y_i f_k(x) > 0$ bị giảm trọng số. Với cách làm này, thuật toán khiến bộ phân loại chú ý hơn tới những mẫu hiện đang bị phân loại sai.

2.4 Bộ cơ sở dữ liệu dùng trong nghiên cứu

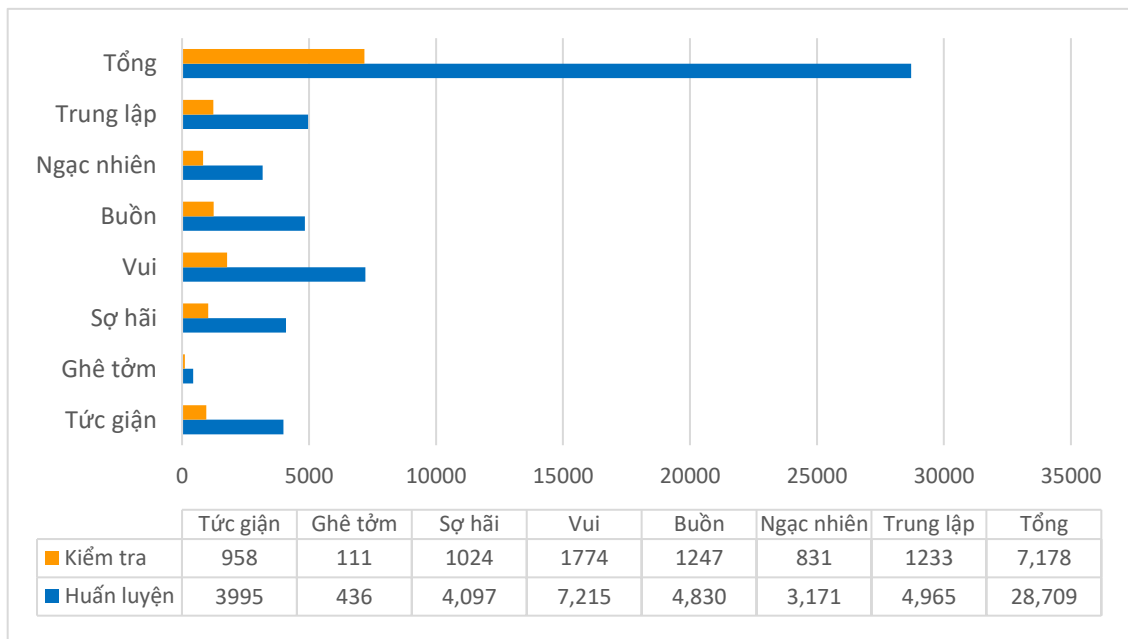
2.4.1 Dữ liệu về biểu cảm khuôn mặt

a. FER-2013 (Facial Expression Recognition Challenge 2013)

Tập dữ liệu FER-2013 [21] được giới thiệu tại hội nghị quốc tế vào năm 2013 về học máy (ICML) và là tập dữ liệu chuẩn được dùng trong cuộc thi về thách thức nhận dạng biểu cảm trên khuôn mặt. Trong tập dữ liệu này bao gồm các hình ảnh thang độ xám, có kích thước 48×48 pixel và có hơn 35.000 ảnh. Trong đó: Tập dữ liệu dùng để huấn luyện là 28.709 ảnh và dùng cho việc kiểm tra là 7.178 ảnh, cho 7 loại biểu cảm khác nhau trên khuôn mặt người.

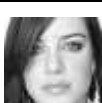
Bảng 2.5 Bảng phân bố cảm xúc trong tập dữ liệu mẫu FER-2013

Cảm xúc	Số lượng ảnh	Huấn luyện	Kiểm tra	Tỉ lệ (%)
Tức giận	4.953	3.995	958	13.81
Ghê tởm	547	436	111	1.52
Sợ hãi	5.121	4.097	1.024	14.27
Vui	8.989	7.215	1.774	25.05
Buồn	6.077	4.830	1.247	16.93
Ngạc nhiên	4.002	3.171	831	11.15
Trung lập	6.198	4.965	1.233	17.27
Tổng	35.887	28.709	7.178	100



Hình 2.18 Phân bố số lượng cảm xúc trong tập dữ liệu FER-2013

Bảng 2.6 Một số ảnh mẫu trong tập dữ liệu FEREC-2013

Vui						
Tức giận						
Buồn						
Sợ hãi						
Ghê tởm						
Ngạc nhiên						
Trung lập						

b. JAFFE (Japanese Female Facial Expression)

Bộ dữ liệu JAFFE đã đóng góp quan trọng cho lĩnh vực nhận dạng biểu hiện khuôn mặt; đã được sử dụng trong nhiều nghiên cứu và thử nghiệm để kiểm tra hiệu suất của các thuật toán nhận dạng và phân loại cảm xúc.

Bảng 2.7 Một số ảnh mẫu trong tập dữ liệu JAFFE

Tức giận	Ghê tởm	Sợ hãi	Vui	Buồn	Ngạc nhiên	Trung lập
						
						
						

JAFFE được xây dựng bởi Michael Lyons, Miyuki Kamachi và Jiro Gyoba cùng với trợ lý nghiên cứu Reiko Kubota [103]. Những bức ảnh được chụp tại Khoa Tâm lý của Đại học Kyushu. Trong bộ dữ liệu có chứa 213 ảnh, gồm có sáu cảm xúc cơ bản: vui, buồn, ngạc nhiên, tức giận, ghê tởm, sợ hãi và một cảm xúc trung lập. Hình ảnh mức xám có kích thước 256×256 và được chụp trong môi trường kiểm tra ở các điều kiện ánh sáng khác nhau để tái hiện các biểu hiện khuôn mặt một cách thực tế.

2.4.2 Dữ liệu phát hiện trạng thái đóng/mở mắt

a. Talking face

Tập dữ liệu Talking Face bao gồm 5000 khung hình được lấy từ video của một người tham gia vào cuộc trò chuyện. Khung hình được chụp với tốc độ 25 khung hình/giây với độ phân giải 720×576

b. Eyeblink8

Bộ dữ liệu Eyeblink8 chứa 8 video với 4 đối tượng (1 người đeo kính). Video được quay trong môi trường gia đình. Mọi người đang ngồi trước máy quay và chủ yếu hành động tự phát với các mô phỏng khuôn mặt trực tiếp, tương tự như tập dữ liệu Talking Face. Có tổng cộng 408 lần nhấp nháy mắt trên 70.992 khung hình với độ phân giải 640×480 . Cả hai bộ dữ liệu đều được các tác giả gắn nhãn và chú thích về hành vi mở và đóng của mắt ở mỗi khung hình.

c. HUST-LEBW

HUST_LEBW là một cơ sở dữ liệu đa phương thức được thu thập trong tự nhiên. Nó bao gồm 673 mẫu video clip nhấp mắt tương ứng với 172 đối tượng. Mỗi mẫu clip bao gồm 13 khung hình và bao gồm một mẫu nhấp nháy hoặc mẫu không nhấp nháy. Có tổng cộng 381 lần nhấp nháy mắt trên 8.749 khung hình với cả hai độ phân giải là 1280×720 và 1456×600 . Bộ dữ liệu này rất phong phú về tư thế, sự thay đổi ánh sáng và các yếu tố tự nhiên khác, được xuất bản vào năm 2019 bởi [104] cho mục đích nghiên cứu về nhấp mắt trong tự nhiên.

d. ZJU

Tập dữ liệu ZJU chứa 80 video ngắn với 20 đối tượng bao gồm cả đeo kính và không đeo kính. Video được quay trong nhà với tốc độ 30 khung hình/giây. Có tổng cộng 264 lần chớp mắt trên 10876 khung hình với độ phân giải 320×240 .

Bảng 2.8 Một số ảnh mẫu trong tập dữ liệu Talking Face, Eyeblink8, HUST-LEBW, ZJU



Bảng 2.9 Thông tin về tập dữ liệu Talking Face, Eyeblink8, HUST-LEBW, ZJU

Tập dữ liệu	Độ dài	Độ phân giải	Đối tượng
Talking face	5,000 frames	720 × 576	1
Eyeblink8	70,992 frames	640 × 480	4
HUST-LEBW	8,749 frames	1280 × 720 1456 × 600	172
ZJU	10,867 frames	320 × 240	20

2.5 Kết chương

Chương này trình bày tổng quan về các vấn đề liên quan đến việc phát hiện sự tham gia của người học; cụ thể, chương này trình bày ý nghĩa sự tập trung trong giáo dục, các yếu tố và phương pháp đánh giá sự tập trung. Ngoài ra, chương này cũng trình bày các công trình nghiên cứu liên quan, bài toán về nhận dạng biểu cảm và một số vấn đề lý thuyết cơ bản như: lý thuyết nhận dạng (gồm có các khái niệm cơ sở và các bước cơ bản trong việc xây dựng hệ nhận dạng, các tiêu chuẩn và phương pháp học trong nhận dạng) và kỹ thuật xử lý ảnh số. Những vấn đề được trình bày trong chương này làm nền tảng cho việc xây dựng kỹ thuật đánh giá mức độ tập trung được đề xuất trong các chương sau; chẳng hạn như kỹ thuật đánh giá sự tập trung của người học dựa trên thành phần của khuôn mặt được trình bày ở Chương 3. Hay, Chương 4 trình bày kỹ thuật đánh giá sự tập trung của người học dựa trên hành vi của khuôn mặt.

CHƯƠNG 3

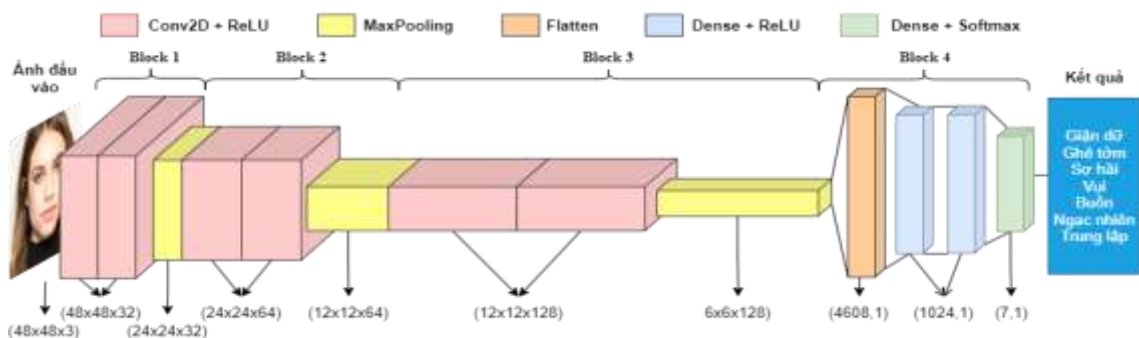
NHẬN DẠNG BIỂU CẢM THÀNH PHẦN KHUÔN MẶT

Từ các vấn đề cơ bản trong Chương 2, trong Chương này Luận án trình bày kỹ thuật đánh giá mức độ tập trung của người học dựa trên các biểu hiện trên khuôn mặt. Thông qua kết quả nghiên cứu của kỹ thuật này, Luận án đề xuất kỹ thuật phân rã biểu cảm thành phần cơ bản của người học nhằm nâng cao chất lượng cho mô hình nhận dạng, đồng thời hiểu rõ hơn về chi tiết cảm xúc của người học. Vì trên thực tế, biểu cảm của người học được biểu diễn hỗn hợp từ các cảm xúc cơ bản. Ngoài ra, đối với kỹ thuật này, Luận án có sử dụng bài toán nhận diện khuôn mặt làm cơ sở lý thuyết cho các công đoạn tiếp theo.

3.1 Đánh giá sự tập trung dựa trên biểu hiện trên khuôn mặt

3.1.1 Kiến trúc mô hình

Nhận dạng biểu cảm khuôn mặt (FER) là một trong những nhiệm vụ thuộc phân loại hình ảnh, phân loại trạng thái tình cảm của người thông qua các đặc điểm trên khuôn mặt. FER đã trở thành một trong những vấn đề thú vị nhất trong cộng đồng nghiên cứu về lĩnh vực thị giác máy tính. Trong những năm gần đây, nhiều công trình đã được thực hiện để xác định trạng thái cảm xúc của con người từ các nguồn video hoặc khung ảnh [77]. Theo [17], tác giả xây dựng mô hình dựa trên mô hình VGGNet bằng cách giảm số lớp của nó. Kiến trúc của mô hình được mô tả như Hình 3.1



Hình 3.1 Kiến trúc mô hình nhận dạng biểu hiện trên khuôn mặt

Như thể hiện trong Hình 3.1, kiến trúc mô hình được đề xuất bao gồm nhiều lớp tích chập, hàm gộp tối đa, một lớp làm phẳng theo sau là hai lớp ẩn được kết nối dày đặc và một lớp softmax. Phần khai thác tính năng được chia thành ba khối.

Mỗi khối bao gồm hai lớp tích chập liên tiếp theo sau là một lớp gộp tối đa. Khối đầu tiên bắt đầu bằng cách đọc một hình ảnh có kích thước $48 \times 48 \times 3$ làm đầu vào lớp tích chập đầu tiên của nó. Lớp tích chập đầu tiên bao gồm 32 hạt nhân có kích thước 3×3 mỗi hạt. Các hạt nhân này được biến đổi trên hình ảnh đầu vào một cách độc lập để tạo ra 32 bản đồ đặc trưng phi tuyến tính. Bản đồ đối tượng địa lý phi tuyến tính được tạo bằng cách áp dụng một hàm kích hoạt phi tuyến tính được tính toán từ mỗi trường tiếp nhận trong quá trình tích chập. Trong trường hợp này ReLU được sử dụng làm chức năng kích hoạt phi tuyến tính.

Bảng 3.1 Chi tiết mô hình nhận dạng cảm xúc trên khuôn mặt [17]

CNN block	CNN Layer	Number of kernels	Input size	Output size
Block 1	Conv2D Layer 1 + Input	32	(48, 48, 3)	(48, 48, 32)
	BatchNormalization	--	(48, 48, 32)	(48, 48, 32)
	Conv2D Layer 2	32	(48, 48, 32)	(48, 48, 32)
	BatchNormalization	--	(48, 48, 32)	(48, 48, 32)
	MaxPooling2D Layer 1	--	(48, 48, 32)	(24, 24, 32)
	Dropout Layer 1	--	(24, 24, 32)	(24, 24, 32)
Block 2	Conv2D Layer 3	64	(24, 24, 32)	(24, 24, 64)
	BatchNormalization	--	(24, 24, 64)	(24, 24, 64)
	Conv2D Layer 4	64	(24, 24, 64)	(24, 24, 64)
	BatchNormalization	--	(24, 24, 64)	(24, 24, 64)
	MaxPooling2D Layer 2	--	(24, 24, 64)	(12, 12, 64)
	Dropout Layer 2	--	(12, 12, 64)	(12, 12, 64)
Block 3	Conv2D Layer 5	128	(12, 12, 64)	(12, 12, 128)
	BatchNormalization	--	(12, 12, 128)	(12, 12, 128)
	Conv2D Layer 6	128	(12, 12, 128)	(12, 12, 128)
	BatchNormalization	--	(12, 12, 128)	(12, 12, 128)
	MaxPooling2D Layer 3	--	(12, 12, 128)	(6, 6, 128)
	Dropout Layer 3	--	(6, 6, 128)	(6, 6, 128)
Block 4	Flatten Layer	--	(6, 6, 128)	(4608)
	Dense Layer 1	--	(4608)	(1024)
	BatchNormalization	--	(1024)	(1024)
	Dropout Layer 4	--	(1024)	(1024)
	Dense Layer 2	--	(1024)	(1024)
	BatchNormalization	--	(1024)	(1024)
	Dropout Layer 4	--	(1024)	(1024)
	Dense Layer 3 + Softmax	--	(1024)	(7)

Hơn nữa, 32 bản đồ đối tượng địa lý phi tuyến tính này được chuyển tiếp làm đầu vào cho lớp tích chập thứ hai. Lớp này lặp lại quá trình tương tự như lớp tích chập đầu tiên và tạo ra 32 bản đồ đối tượng phi tuyến tính được chập đôi dưới dạng đầu ra. Mục đích của phép toán tích chập kép là để cải thiện tính phi tuyến tính của bản đồ đối tượng. Nó giúp mô hình tìm ra các bản đồ tính năng phân biệt giữa các lớp mạnh mẽ. Tiếp theo, lớp gộp tối đa, hoạt động trên từng bản đồ đối tượng một cách độc lập để tạo ra một bộ sưu tập mới gồm 32 bản đồ đối tượng được lấy mẫu xuống. Tốc độ lấy mẫu giảm phụ thuộc vào hai tham số, kích thước hạt nhân và bước trượt. Vì kích thước hạt nhân và khoảng cách được đặt thành 2×2 , nên các bản đồ đối tượng địa lý được gộp chung sẽ được lấy mẫu xuống còn một nửa kích thước ban đầu của nó. Sau khi hoàn thành Block 1, mô hình nhận được 32 bản đồ đặc trưng với kích thước 24×24 mỗi bản đồ.

Các bản đồ tính năng tổng hợp này được chuyển làm đầu vào cho khối thứ hai. Quá trình làm việc của cả ba khối trong phần phân loại là giống nhau, ngoại trừ số lượng hạt nhân được sử dụng trong mỗi khối. Khối thứ hai tạo ra 64 bản đồ đặc trưng có kích thước 12×12 mỗi bản đồ. Tiếp tục, khối thứ ba chấp nhận 64 bản đồ tính năng này làm đầu vào và tạo ra 128 bản đồ tính năng có kích thước 6×6 mỗi bản đồ. Lớp làm phẳng đóng vai trò là lớp đầu vào khối phân loại của mô hình. Khối phân loại chứa một mạng nơ-ron được kết nối đầy đủ với hai lớp ẩn có kích thước 1024 mỗi lớp và một lớp đầu ra. Lớp đầu ra có bảy tế bào thần kinh (cho bảy lớp cảm xúc) và sử dụng hàm softmax để lấy điểm xác suất. Ngoài ra, mô hình được bổ sung thêm các lớp chuẩn hóa hàng loạt và Drop-out được sử dụng để tính toán nhanh hơn và để tránh các vấn đề quá mức tương ứng.

3.1.2 Đánh giá thực nghiệm

3.1.2.1 Cơ sở dữ liệu thực nghiệm

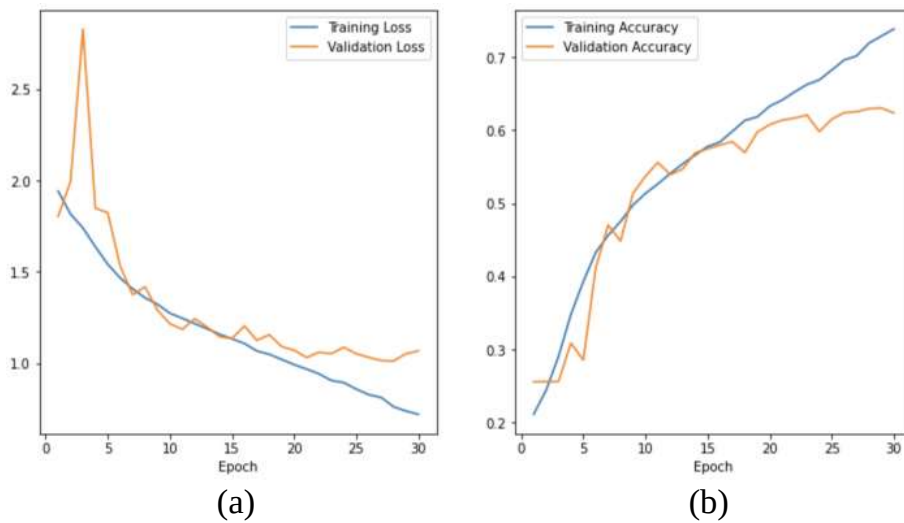
Để đánh giá nhận dạng cảm xúc trên khuôn mặt với độ chính xác cao, chúng ta cần có bộ dữ liệu chuẩn và được nhiều công trình nghiên cứu sử dụng. Chất lượng bộ dữ liệu rất quan trọng sẽ làm ảnh hưởng đến kết quả cho quá trình thực nghiệm và đánh giá các phương pháp nhận dạng cảm xúc trên khuôn mặt. Trong phần thực nghiệm này, Luận án sử dụng bộ cơ sở dữ liệu chuẩn FEREC-2013 và được trình bày chi tiết trong Chương 2.

3.1.2.2 Phương pháp đánh giá và kết quả thực nghiệm

Luận án thực nghiệm trên bộ dữ liệu chuẩn FER-2013. Để đánh giá hiệu suất mô hình thực nghiệm thì cần dùng công thức như là độ chính xác (*Precision*), độ phản hồi (*Recall*) và điểm số *F1* (*F1-Score*). Giá trị tính toán cho các thước đo này dựa vào ma trận nhầm lẫn mà mô hình nhận dạng được. Các công thức tính toán được trình bày chi tiết trong Chương 2



Hình 3.2 Một số cảm xúc được phát hiện từ chương trình



(a) Đồ thị biểu diễn mất mát dữ liệu của tập huấn luyện và tập đánh giá;

(b) Đồ thị biểu diễn độ chính xác của tập huấn luyện và tập đánh giá

Hình 3.3 Kết quả huấn luyện từ chương trình.

Bảng 3.2 Ma trận nhầm lẫn trên tập dữ liệu kiểm tra của FERC-2013

	Tức giận	Ghê tởm	Sợ hãi	Vui	Buồn	Ngạc nhiên	Trung lập
Tức giận	555	14	90	44	154	17	84
Ghê tởm	26	62	8	2	10	2	1
Sợ hãi	153	5	395	47	235	92	97
Vui	33	1	33	1540	53	35	79
Buồn	134	4	110	65	711	15	208
Ngạc nhiên	20	4	72	49	21	648	17
Trung lập	68	3	54	97	222	21	768

Trong bảng trên, tiêu đề cột đại diện cho nhãn dự đoán, tiêu đề dòng đại diện cho nhãn thực tế. Các giá trị của đường chéo chính thể hiện kết quả nhận dạng đúng từ mô hình tương ứng với mỗi lớp. Các giá trị còn lại thể hiện kết quả nhận nhầm từ biểu cảm này sang biểu cảm khác.

Ví dụ, xem xét dòng 1 và cột 1 từ Bảng 3.2 thì trong tập dữ liệu kiểm tra chứa 958 hình với biểu cảm Tức giận. Vậy theo kết quả từ bảng trên, mô hình nhận dạng đúng những tám hình có biểu cảm Tức giận là 555. Trong đó, nhận dạng sai từ biểu cảm Tức giận sang các biểu cảm khác tương ứng là Ghê tởm: 14, Sợ hãi: 90, Vui: 44, Buồn: 154, Ngạc nhiên: 17, Trung lập: 84 và nhận dạng sai từ các biểu cảm khác sang biểu cảm Tức giận là Ghê tởm: 26, Sợ hãi: 153, Vui: 33, Buồn: 134, Ngạc nhiên: 20, Trung lập: 68.

Vậy các giá trị độ đo Precision, Recall và F1-Score của biểu cảm Angry được tính như sau:

$$Precision_{Tức\ giận} = \frac{555}{(555 + 26 + 153 + 33 + 134 + 20 + 68)} = 0.5611$$

$$Recall_{Tức\ giận} = \frac{555}{(555 + 14 + 90 + 44 + 154 + 17 + 84)} = 0.5793 \quad (3.1)$$

$$F1 - Score_{Tức\ giận} = \frac{2 \times (0.5611 \times 0.5793)}{(0.5611 + 0.5793)} = 0.5701$$

Tương tự với các cảm xúc còn lại, kết quả được trình bày trong Bảng 3.3 thể hiện các giá trị độ đo mà mô hình nhận dạng đạt được trong tập dữ liệu kiểm tra.

Bảng 3.3 Kết quả chạy thử nghiệm trên tập dữ liệu kiểm tra của FERC-2013

	Precision	Recall	F1-score
Tức giận	0.56	0.58	0.57
Ghê tởm	0.67	0.56	0.61
Sợ hãi	0.52	0.39	0.44
Hạnh phúc, Vui mừng	0.84	0.87	0.85
Buồn	0.51	0.57	0.54
Ngạc nhiên	0.78	0.78	0.78
Trung lập	0.61	0.62	0.62

3.1.3 Phân loại tập trung

a) Chỉ số tập trung – Concentration Index (CI)

Sự tham gia của người học có thể được đo lường bằng cách phân tích trạng thái tình cảm của người học (như tâm trạng và cảm xúc) dựa trên các biểu hiện trên khuôn mặt để ước tính mức độ tham gia tổng thể của người học. Chỉ số đo lường mức độ tập trung được thể hiện như công thức sau:

$$CI = DEP \times EW \quad (3.2)$$

Trong đó:

- **DEP** (Dominant Emotions Probability): là giá trị thể hiện cảm xúc trên khuôn mặt được dự đoán từ mô hình nhận dạng (CNN).

- **EW** (Emotion Weight): Trọng số cảm xúc, là giá trị mô tả mức độ trạng thái cảm xúc cụ thể phản ánh mức độ tập trung của người học tại thời điểm nhất định. Giá trị này nằm trong khoảng từ 0 đến 1, được trình bày bởi [19] [20].

Tác giả khảo sát 30 sinh viên bằng cách trình chiếu video với những câu hỏi đã được chuẩn bị trước và dữ liệu về cảm xúc trên khuôn mặt của sinh viên khi tham gia khảo sát được ghi lại. Sau đó các sinh viên được nhóm lại dựa trên cảm xúc chính mà họ thể hiện; ví dụ: nếu một sinh viên thể hiện phần lớn thời gian (hơn 50% thời lượng video) cho một biểu cảm trung lập, thì sinh viên đó được đưa vào nhóm cảm xúc trung lập. Các sinh viên được phân chia thành bảy nhóm cảm xúc theo biểu cảm khuôn mặt thu được trong video. Điểm trung bình đạt được trong bài khảo sát của mỗi cảm xúc được tính toán và trình bày như Bảng 3.4. Do đó, trọng số của mỗi cảm xúc ở một mức độ nào đó phản ánh mối quan hệ giữa mức độ tập trung của một sinh viên và trạng thái cảm xúc của họ.

Bảng 3.4 Cảm xúc và trọng số tương ứng [19] [20]

STT	Cảm xúc	Trọng số (EW)	CI
1	Trung lập	0.9	$(DEP \times 0.9) \times 100$
2	Hạnh phúc, vui mừng	0.6	$(DEP \times 0.6) \times 100$
3	Ngạc nhiên	0.6	$(DEP \times 0.5) \times 100$
4	Buồn	0.3	$(DEP \times 0.3) \times 100$
5	Ghê tởm	0.2	$(DEP \times 0.2) \times 100$
6	Tức giận	0.25	$(DEP \times 0.25) \times 100$
7	Sợ hãi	0.3	$(DEP \times 0.3) \times 100$

b) Phân loại

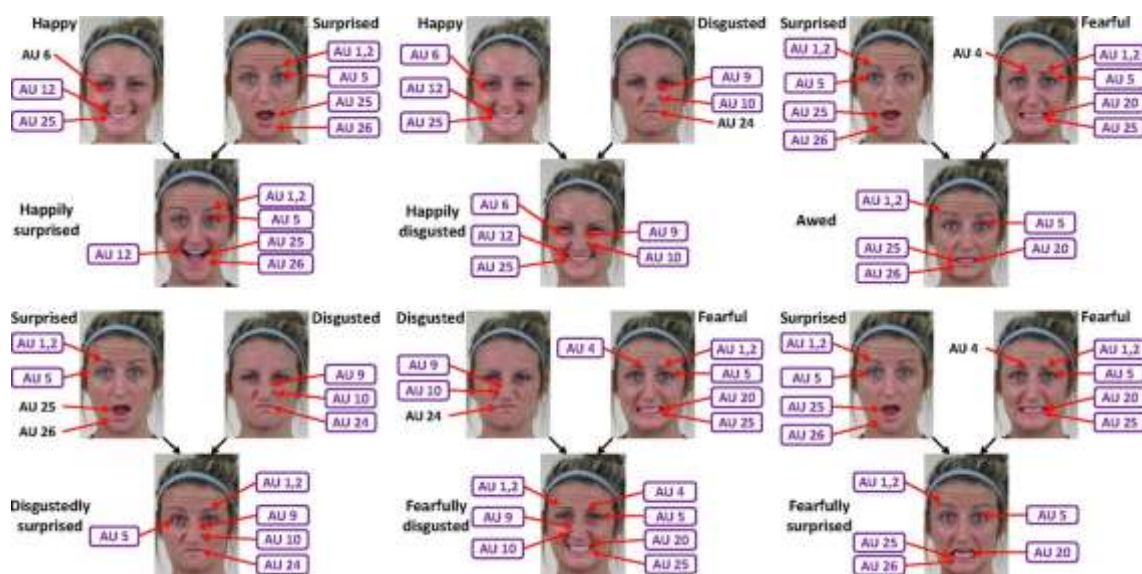
Theo công trình [19] [20], sự tham gia của người học được chia thành 3 loại: Không tập trung, tập trung thấp, tập trung cao được định nghĩa là mức độ gắn kết cảm xúc của người học khi nghe giảng và được phân loại thành ba mức tương ứng với các cảm xúc cơ bản sau:

$$f(CI) = \begin{cases} \text{Không tập trung, } CI < 20\% \\ \text{Tập trung thấp, } CI \in [20\%, 50\%] \\ \text{Tập trung cao, } CI \in (50\%, 100\%] \end{cases} \quad (3.3)$$

3.2 Kỹ thuật phân rã biểu cảm thành phần cơ bản

3.2.1 Giới thiệu vấn đề

Như đã thảo luận ở Chương 1, một trong những yếu tố quan trọng ảnh hưởng đến chất lượng của mô hình nhận dạng cảm xúc trên khuôn mặt là bộ dữ liệu huấn luyện. Tuy nhiên, trong bộ dữ liệu chuẩn FEREC-2013 [21], chỉ bao gồm các cảm xúc cơ bản, mà cảm xúc trong học tập có thể là hỗn hợp do đó cần phân rã biểu cảm thành phần cơ bản để nâng cao độ chính xác cho bài toán nhận dạng và để hiểu rõ chi tiết hơn về cảm xúc của người học. Do đó, phân rã biểu cảm là một bài toán quan trọng trong việc nghiên cứu và phát triển ứng dụng để đánh giá mức độ tham gia của người học dựa trên cảm xúc. Trong quá trình học, người học luôn tiếp nhận những thông tin từ giáo viên, bạn bè. Vì vậy, cảm xúc của người học ở mỗi thời điểm xảy ra khác nhau, có thể là hỗn hợp hoặc đơn lẻ nên cảm xúc cũng thể hiện những biến dạng tương ứng.



Hình 3.4 Một số hình ảnh minh họa cảm xúc hỗn hợp [105]

Trong thực tế, những biến dạng mà chúng ta có thể thấy hàng ngày như độ dẫn của sợi mì ăn liền khi nhúng vào nước nóng cũng chứa đựng các mô hình toán học của chính nó. Về bản chất, các biểu cảm trên khuôn mặt người thay đổi được xem như là biến dạng một khối vật thể 3D. Do đó, có thể hiểu là bao gồm một tập hợp các điểm trong không gian ba chiều với các quan hệ về cạnh cũng như bề mặt giữa các điểm đó. Ngoài ra, tùy từng trường hợp nghiên cứu và ứng dụng cụ thể mà mô hình 3D có thể thêm các tính năng khác nhau.

Theo đó, biến dạng mô hình 3D thực chất là các quy tắc thay đổi thông tin đó trên một mô hình. Trong phần lớn các trường hợp nghiên cứu và ứng dụng, biến dạng mô hình 3D thường được xem xét dưới dạng độ biến thiên tương đối của các điểm với nhau trong mô hình 3D. Ví dụ: khi xem xét việc chuyển đổi mô hình khuôn mặt 3D từ mô hình biểu cảm trung lập sang mô hình biểu cảm cười. Rõ ràng là số điểm cũng như quan hệ về cạnh và bề mặt giữa các điểm là cố định, chỉ là có sự thay đổi tọa độ của các điểm với nhau, chẳng hạn khi cười thì các điểm khóe miệng giãn ra và di chuyển lên trên (Hình 3.5).



Hình 3.5 Biến dạng mô hình khuôn mặt của nhân vật trong dự án Sintel [106]

Nhóm của V. Blanz [107] đã nghiên cứu một mô hình thống kê khuôn mặt được xây dựng từ dữ liệu quét laser 3D trên máy Cyberware 3030ps. Kết quả nhóm cho phép thực hiện các phép biến đổi mô hình sao cho phù hợp nhất với khuôn mặt trên một hình ảnh đầu vào. Như vậy, sau quá trình đó, các tham số mô hình sẽ tương ứng đại diện cho khuôn mặt đó. Nghiên cứu này cho phép phân tích sâu dữ liệu khuôn mặt trên mô hình 3D như phân tích cảm xúc cũng như nhận dạng khuôn mặt (Hình 3.6).

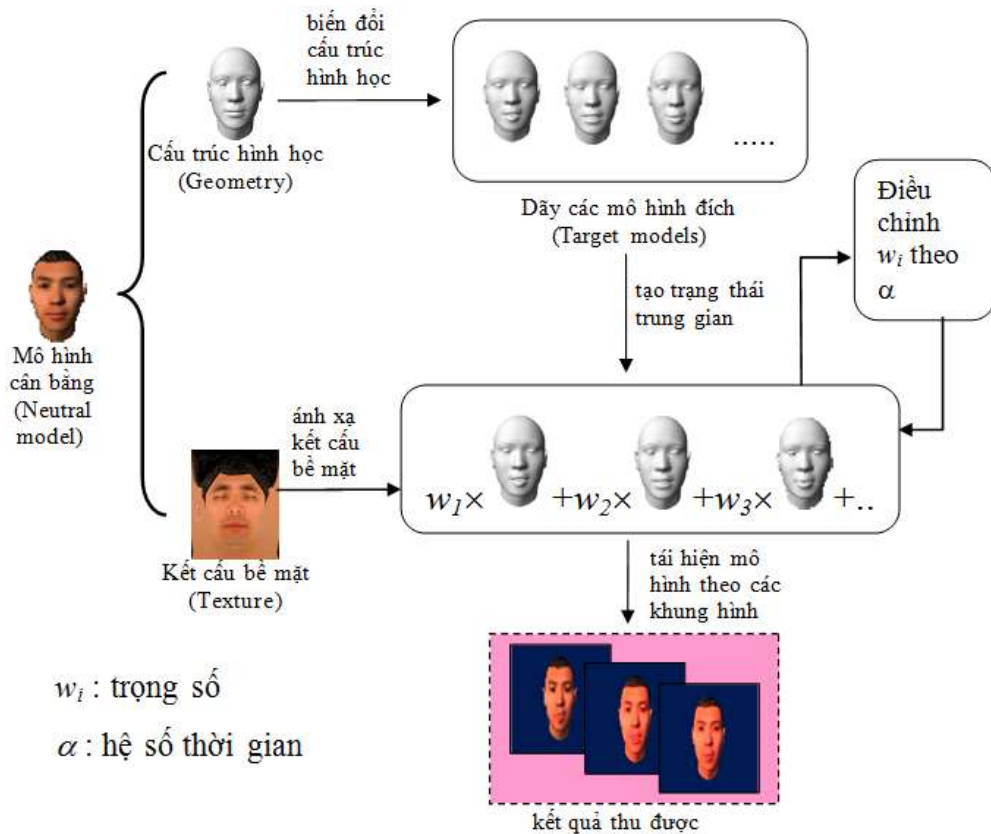


Hình 3.6 Dữ liệu khuôn mặt trong nghiên cứu của nhóm V. Blanz [107]

Weiyue Wang và cộng sự [108] đã đề xuất một kỹ thuật tạo biến thể của mô hình 3D, theo mô hình đích (được biểu diễn dưới dạng hình ảnh 2D hoặc 3D), tác giả thực hiện biến dạng lưới nguồn sao cho giống với mô hình đích nhất có thể. Trong mô hình biến dạng này, tác giả giữ cố định cấu trúc liên kết tam giác của lưới nguồn và chỉ cập nhật các vị trí đỉnh.

Groueix và cộng sự [109] đề xuất Mạng biến dạng hình dạng, một giải pháp toàn diện, tất cả trong một để so khớp hình dạng dựa trên mẫu. Mạng biến dạng hình dạng học cách biến dạng hình mẫu để căn chỉnh với hình dạng được quan sát đầu vào. Đưa ra hai hình dạng đầu vào, chúng căn chỉnh mẫu với cả hai đầu vào và thu được bản đồ cuối cùng giữa các đầu vào bằng cách đọc các tương ứng từ mẫu. Dương và cộng sự. đề xuất Foldingnet [110] làm biến dạng lưới 2D thành

đám mây điểm 3D trong khi vẫn lưu giữ thông tin địa phương. Huỳnh Cao Tuấn và cộng sự [111] đã đề xuất thuật toán cải tiến tự động xác định các điểm chính và phân cụm các biến dạng tương tự bằng kỹ thuật hàm cơ bản hướng tâm (RBF) để cải thiện biến dạng của khuôn mặt 3D. Một nghiên cứu khác liên quan đến vấn đề tổng hợp một mô hình khuôn mặt từ một tập các mô hình bề mặt cơ bản [112]. Trong nghiên cứu này, để thực hiện phép nội suy biểu diễn các cử chỉ, trạng thái biểu cảm của khuôn mặt người, nhóm nghiên cứu đã sử dụng bộ mô hình mẫu bao gồm mô hình khuôn mặt trung tính và mô hình thể hiện trạng thái khuôn mặt theo các biểu cảm cụ thể. Mô hình kết quả được tạo ra bằng cách tổng hợp tuyến tính của các mô hình cơ bản đó (Hình 3.7).



Hình 3.7 Quá trình tổng hợp mẫu khuôn mặt [112]

Ý tưởng nghiên cứu biến dạng mô hình của kỹ thuật được đặt ra theo cách nghiên cứu ngược lại của nghiên cứu [112]. Theo đó, Luận án đề xuất một kỹ thuật phân tách có điều kiện các biến thể mô hình 3D dựa trên một tập hợp các quan sát 3D nhất định của một đối tượng, theo một tập hợp các trọng số biến dạng đầu vào. Kết quả của kỹ thuật đề xuất cho phép tạo ra một biến thể mô hình 3D mới của đối tượng một cách đơn giản và có thể quan sát trực quan.

Cụ thể, kỹ thuật được đề xuất sẽ phân tích nhân trọng số của tập dữ liệu mô hình 3D cho từng loại biến dạng nhất định. Đầu ra của kỹ thuật là một tập hợp các mô hình cơ bản tương ứng với các loại biến dạng khác nhau

3.2.2 Phương pháp thực hiện

3.2.2.1 Mô hình tổng hợp 3D

Các biến thể của mô hình 3D được thể hiện dưới dạng sau:

$$s = m + \sum_{i=0}^n \alpha_i \times e_i \quad (3.4)$$

Trong đó:

- s là biến thể của mô hình được tổng hợp;
- m là model pattern, có thể hiểu là model cân bằng giữa các biến dạng;
- e_i là mô hình offset tương ứng với biến dạng i ;
- α_i là trọng lượng ứng với biến dạng i ;
- n là số loại biến dạng của mô hình.

Phương trình trên được giả định là đại diện cho bất kỳ biến thể nào của mô hình. Như vậy, để tổng hợp một mô hình, chúng ta cần có dữ liệu ở vế phải của công thức (3.4). Trong đó tập các mô hình cơ bản bao gồm m và e_i là dữ liệu được xác định trước, α_i là trọng số biến dạng.

3.2.2.2 Bài toán bình phương tối thiểu

Trong bài toán này, mục tiêu là ước lượng một tập hợp các mô hình cơ bản, nghĩa là tính toán m và e_i . Chúng ta cần tập dữ liệu đầu vào của các biến thể s và các giá trị trọng số α_i tương ứng với từng biến thể. Tập dữ liệu đầu vào được mô hình hóa theo công thức trên dưới dạng ma trận như sau:

$$S = M + A \times E \quad (3.5)$$

Trong đó:

- S là tập các mẫu đầu vào, mỗi mẫu được coi là một vectơ $1 \times L$, tập S gồm N mẫu, S là ma trận $N \times L$.
- M là một ma trận $N \times L$, trong mô hình này m được sao chép N lần và được xếp trên mỗi dòng.
- A là trọng số biến dạng $N \times n$ của ma trận.
- E là một ma trận $n \times L$ là các mô hình bù.

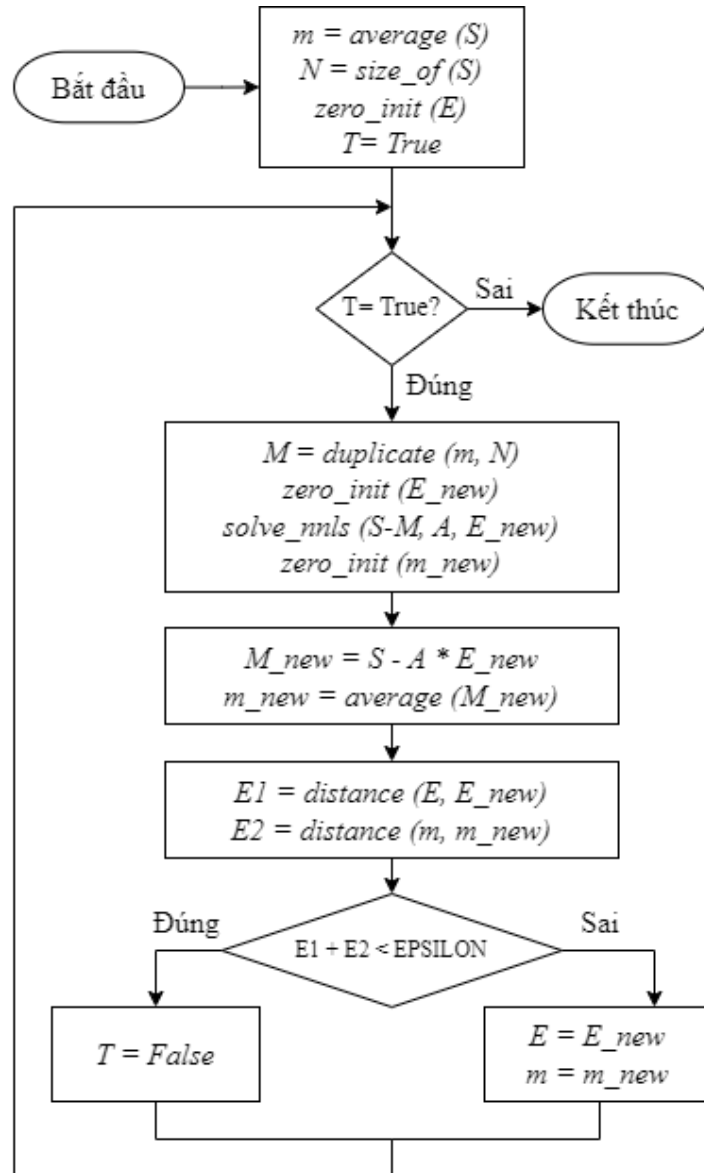
Nếu M là giả định, công thức (3.6) sẽ trả về:

$$S - M = A \times E \quad (3.6)$$

Do, các giá trị trọng lượng biến dạng α_i là các giá trị không âm nên để tìm E , nghiệm của phương trình (3.6) được chuyển thành bình phương tối thiểu [113].

3.2.3 Thuật toán phân rã

Đầu vào S, A
Đầu ra m, E
Xử lý Xem Hình 3.8



Hình 3.8 Lưu đồ giải thuật phân rã

Việc ước lượng tập của mô hình cơ sở được thực hiện thông qua một quy trình lặp tối ưu. Ở bước ban đầu, mô hình mẫu m được khởi tạo bằng cách lấy trung bình các mô hình biến thể trong tập dữ liệu đầu vào S và các mô hình bổ sung được khởi tạo bằng 0. Tại mỗi lần lặp lại, phép tính lại được thực hiện lần lượt với hai giá trị.

Đầu tiên, các mô hình m được nhân đôi để thu được ma trận M và Luận án giải quyết bằng phương pháp bình phương nhỏ nhất không âm để có được tập hợp các mô hình bổ sung mới E_{new} . Sau đó, tập hợp mô hình bổ sung mới E_{new} được sử dụng để tính toán lại mô hình mới m_{new} bằng cách tính toán ma trận M_{new} và lấy trung bình cộng để thu được m_{new} .

Sau khi tính giá trị E_{new} và m_{new} tại mỗi bước lặp, Luận án thực hiện tính sai lệch của từng bước cập nhật trên 2 giá trị là sai lệch $E1$ của mô hình bù và sai số lệch $E2$ của mô hình mẫu. Nếu tổng các độ lệch đủ nhỏ, nghĩa là quá trình lặp không còn cập nhật mô hình bổ sung và mô hình mẫu, thì thuật toán kết thúc.

3.2.4 Kết quả thực nghiệm

Trong chương trình thực nghiệm, phương pháp bình phương tối thiểu được lấy từ thư viện Eigen chuyên về tính toán đại số tuyến tính. Kỹ thuật được thử nghiệm trên tập dữ liệu JAFFE trong đó đối tượng 3D là một mô hình khuôn mặt bao gồm 6736 bề mặt và 3448 đỉnh được tạo tương ứng với từng mẫu trong cơ sở dữ liệu.



Hình 3.9 Mô hình khuôn mặt 3D trong tập dữ liệu JAFFE

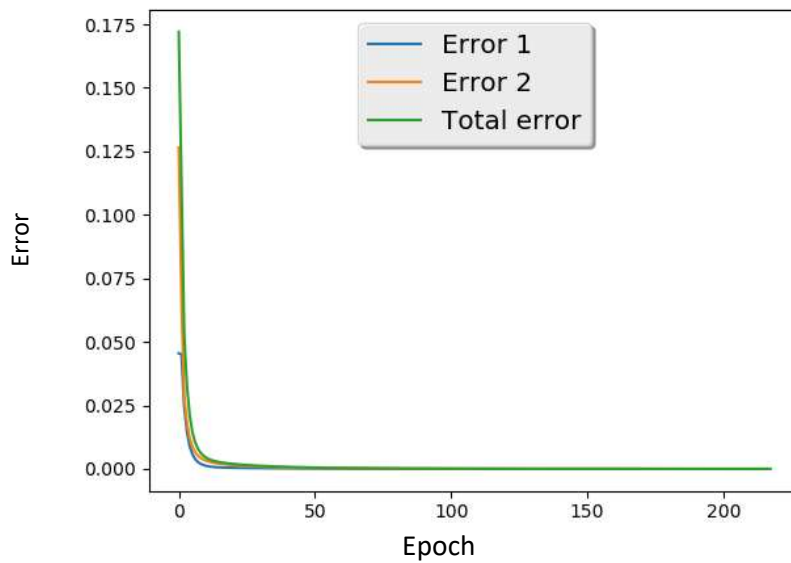
Do đó, thuật toán được áp dụng để phân tách mô hình khuôn mặt 3D theo các tham số cảm xúc. Trong trường hợp này, mỗi độ lệch mẫu tương ứng với một biểu thức cụ thể trong tập dữ liệu đầu vào.

Để giảm ảnh hưởng của một yếu tố khác, thử nghiệm đã được tiến hành trên dữ liệu của một người cụ thể. Như vậy, trong phép thử, đầu vào là tập hợp các mô

hình 3D của khuôn mặt với các trọng số tương ứng các mức độ của từng loại biểu cảm và đầu ra là mô hình khuôn mặt người và mẫu bù cho từng biểu cảm cụ thể.



Hình 3.10 Tổng hợp một số khuôn mặt ngẫu nhiên



Hình 3.11 Tiến trình xử lý

Trong quá trình thử nghiệm, cả mô hình mẫu và mô hình bổ sung đều hội tụ và được hiển thị với sai số lệch tiệm cận bằng 0 sau hơn 200 vòng lặp. Các kết quả sau đó được áp dụng để tổng hợp ngẫu nhiên mô hình khuôn mặt. Tại bước tổng hợp, mỗi giá trị trong tập trọng số biểu thức được lấy ngẫu nhiên trong đoạn $[0, 1]$.

3.3 Kết chương

Đối với kỹ thuật đánh giá dựa trên biểu hiện trên khuôn mặt, để mô hình đạt được độ chính xác cao cần có tập dữ liệu huấn luyện chuẩn về các cảm xúc của người học. Tuy nhiên, tập dữ liệu huấn luyện mẫu về cảm xúc trên khuôn mặt chỉ

chứa các cảm xúc cơ bản [21]. Trên thực tế, cảm xúc của người học được tạo hỗn hợp từ các cảm xúc cơ bản [17]. Do đó, khan hiếm dữ liệu cũng là một trong những yếu tố ảnh hưởng đến chất lượng của mô hình nhận dạng. Để giải quyết vấn đề hạn chế này, Luận án đề xuất một kỹ thuật phân tách có điều kiện các biến thể mô hình 3D dựa trên một tập hợp các quan sát 3D nhất định của một đối tượng. Thuật toán hội tụ và kết quả thu được cũng đã chứng minh tính hiệu quả khi áp dụng để thực hiện cho bài toán phân rã biểu cảm của người học. Qua đó, giúp nâng cao độ chính xác cho mô hình nhận dạng biểu cảm khuôn mặt để phục vụ cho việc đánh giá mức độ tham gia của người học dựa trên nét mặt.

CHƯƠNG 4

NHẬN DẠNG BIỂU CẢM HÀNH VI KHUÔN MẶT

Trong Chương này, Luận án trình bày kỹ thuật đánh giá sự tập trung của người học dựa trên hành vi của khuôn mặt được thể hiện qua trạng thái đóng/mở của mắt. Như đã được đề cập ở Chương 1, khi sử dụng một ngưỡng cố định để phát hiện trạng thái của mắt đã gây ra những hạn chế cho kỹ thuật đánh giá. Điều này sẽ gây ra những giới hạn nhất định khi triển khai vào môi trường thực tế. Do đó, ngoài phần trình bày kỹ thuật đánh giá sự tập trung của người học dựa trên hành vi đóng/mở của mắt, Luận án còn trình bày phương pháp cải tiến phát hiện trạng thái đóng/mở của mắt nhằm khắc phục những điểm yếu còn tồn tại và nâng cao hiệu quả cho kỹ thuật đánh giá sự tập trung mà Luận án đã sử dụng.

4.1 Phát hiện mắt người trong ảnh

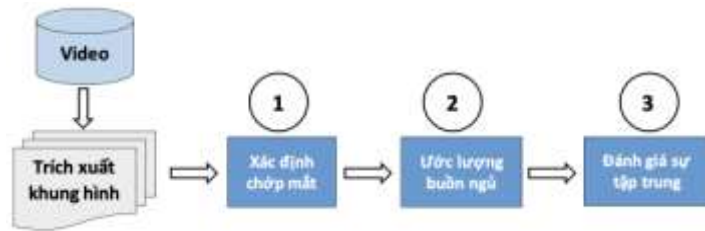
Sau khi nhận diện khuôn mặt, các mốc trên khuôn mặt được trích xuất, vị trí của 68 tọa độ tương ứng với các đặc điểm trên khuôn mặt được ước tính bằng cách sử dụng mô hình phát hiện điểm mốc được đào tạo trước từ thư viện Dlib [24]. Nhận dạng mức độ tập trung dựa trên trạng thái đóng/mở của mắt không yêu cầu tất cả các điểm mốc trên khuôn mặt. Do đó, Theo Hình 4.1 mắt trái có thể trích xuất thông qua các tọa độ từ 36 đến 41 và mắt phải có thể trích xuất từ tọa độ 42 đến 47. Tiếp theo, công việc theo dõi trạng thái của mắt được tính toán dựa trên 12 điểm mốc này và được trình bày chi tiết trong phần tiếp theo.



Hình 4.1 Dấu mốc trên khuôn mặt [24]

4.2 Đánh giá sự tập trung dựa trên trạng thái mắt

Để đánh giá sự tập trung của người học, phương pháp thực hiện ba nhiệm vụ trọng tâm: (i) xác định hành vi chớp mắt của người học trong video; (ii) ước lượng mức độ buồn ngủ của người học dựa trên kết quả của (i) và Bảng 4.1; (iii) đánh giá sự tập trung của người học dựa vào kết quả từ (ii) và kết hợp mức độ ở Bảng 4.2 để ra quyết định.



Hình 4.2 Sơ đồ luồng đánh giá sự tập trung

Trong công việc (1) của Hình 4.2 bao gồm có hai bài toán liên quan như: phát hiện khuôn mặt và phát hiện mắt của người học trong ảnh đã được trình bày ở các phần trên.

4.2.1 Cơ sở lý thuyết của kỹ thuật đánh giá

Mức độ buồn ngủ được ước lượng dựa trên các yếu tố khác nhau, chẳng hạn như tỷ lệ khung hình và thời lượng đóng của mắt. Ứng với mỗi mức độ sẽ có những biểu hiện, hành vi và thời lượng mắt đóng tương ứng. Ngoài ra yếu tố mắt đóng nhanh hay chậm cũng được quan tâm xem xét. Đối với người thức giấc/tỉnh táo thì mắt đóng và mở rất nhanh. Đối với người trong tình trạng mệt mỏi, buồn ngủ thì mắt có xu hướng đóng chậm [114].



Hình 4.3 Thời gian mắt mở đến khi mắt đóng hoàn toàn

Trong Hình 4.3 dựa trên dữ liệu Talking Face¹ với máy ảnh ghi hình là 24fps, trạng thái mắt đóng hoàn toàn xảy ra tại khung hình thứ tư. Do đó thời gian đóng mắt hoàn toàn là:

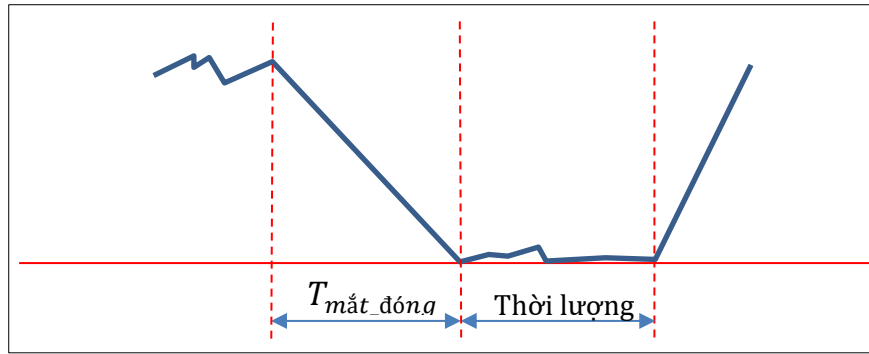
¹ https://personalpages.manchester.ac.uk/staff/timothy.f.cootes/data/talking_face/talking_face.html

$$T_{\text{mắt_đóng}} = \frac{N_{\text{khung_hình}} \times 1000}{fps} \quad (4.1)$$

Trong đó:

- $N_{\text{khung_hình}}$ là số lượng khung hình tính từ khung có trạng thái mắt mở đến khung có trạng thái mắt đóng hoàn toàn (trường hợp này là 4)
- fps (frames per second) là số lượng khung hình trên mỗi giây, giá trị này thay đổi tùy vào máy ghi hình (trong trường hợp này là 24)

Nên theo (4.1) giá trị $T_{\text{mắt_đóng}} = (4 \times 1000) / 24 \approx 170 \text{ ms}$



Hình 4.4 Minh họa thời gian mắt mở đến khi mắt đóng hoàn toàn

Trước khi thực hiện ước lượng mức độ buồn ngủ, bước tiền xử lý của kỹ thuật là loại bỏ những đoạn video có mắt đóng nhanh ($T_{\text{mắt_đóng}} \leq 170\text{ms}$) xem công thức (4.1) và thời lượng mắt đóng ngắn (Thời lượng $< 1\text{s}$), những đoạn này hành vi diễn ra nhanh không ảnh hưởng đến quá trình tham gia nên được xem xét là tập trung. Kỹ thuật chỉ quan tâm đến những đoạn video có mắt đóng chậm ($T_{\text{mắt_đóng}} > 170\text{ms}$) và thời lượng mắt đóng dài (Thời lượng $\geq 1\text{s}$) (xem Bảng 4.1)

Bảng 4.1 Sự tương quan giữa mức độ buồn ngủ và hành vi [114]

Thức giấc/Tỉnh táo (Không buồn ngủ): Thời lượng mắt đóng $< 1\text{s}$ Người thức giấc/tỉnh táo có các hành vi và biểu cảm như: Sắc mặt bình thường; đóng/mở mắt nhanh; các cử động/cử chỉ của cơ thể xảy ra không thường xuyên
Hơi buồn ngủ: Thời lượng mắt đóng $1\text{-}2\text{s}$ Khi người trở nên hơi buồn ngủ, các hành vi khác nhau có thể được thể hiện. Những hành vi này, được gọi là cách cư xử, có thể bao gồm dụi mắt hoặc mắt, gãi, sắc mặt giảm và di chuyển không yên trên ghế. Những hành động này có thể được coi là biện pháp đối phó với cơn buồn ngủ. Chúng xảy ra trong các

giai đoạn trung gian của buồn ngủ. Không phải tất cả các cá nhân đều thể hiện cách cư xử trong các giai đoạn trung gian. Một số cá nhân có vẻ trầm lắng hơn, đóng mắt chậm hơn, sắc mặt của họ có thể giảm, thiếu hoạt động rõ ràng, họ có thể nhìn chăm chăm vào một vị trí cố định.

Ngủ gật: Thời lượng mắt đóng 2-3s hoặc lâu hơn

Khi người trở nên ngủ gật, thường xảy ra tình trạng mí mắt khép lâu hơn. Điều này thường đi kèm với sự di chuyển của mắt lên hoặc sang một bên. Mắt không tập trung thích hợp, thiếu linh hoạt, sắc mặt giảm, thiếu hoạt động rõ ràng, có thể có các chuyển động bị cô lập (hoặc ngắt quãng) lớn.

Bắt đầu giấc ngủ: Thời lượng mắt đóng $\geq 4s$

Người trong tình trạng này thường có biểu hiện mí mắt đóng kéo dài (4 giây trở lên) và thời gian thiếu hoạt động kéo dài tương tự. Có thể có những chuyển động bị ngắt quãng lớn.

Dựa vào mức độ và hành vi ở Bảng 4.1, Luận án xây dựng luật xác định mức độ buồn ngủ như sau:

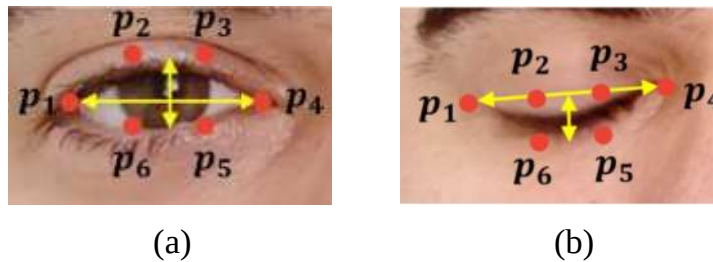
Bảng 4.2 Luật xác định mức độ buồn ngủ

Mức độ	Điều kiện
Thức giấc/Tỉnh táo	$EAR > 0.2$
Hơi buồn ngủ	$EAR < 0.2$ AND $Thời lượng \in [1s, 2s)$
Ngủ gật	$EAR < 0.2$ AND $Thời lượng \in [2s, 4s)$
Bắt đầu giấc ngủ	$EAR < 0.2$ AND $Thời lượng \geq 4s$

Trong đó, EAR [16] là mức độ đóng/mở của mắt được tính theo tỷ lệ giữa chiều cao và chiều rộng của mắt như sau:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2 \times \|p_1 - p_4\|} \quad (4.2)$$

Trong đó: $p_1, p_2, p_3, p_4, p_5, p_6$ là các tọa độ của mắt



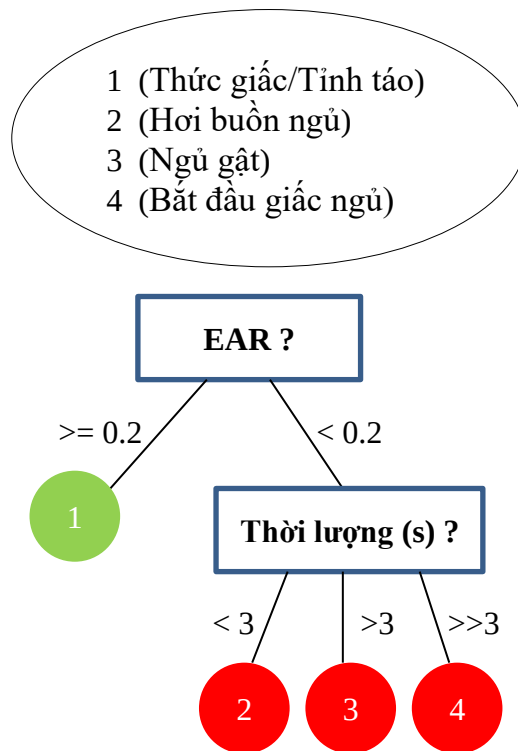
(a) Tỷ lệ khung hình của mắt khi mở; (b) Tỷ lệ khung hình của mắt khi đóng

Hình 4.5 Tọa độ và tỷ lệ khung hình của mắt

Sau khi tính toán được giá trị EAR , theo [16], để xác định mắt đóng hay mở, tác giả so sánh với một ngưỡng ($threshold = 0.2$) như sau:

$$f(EAR) = \begin{cases} \text{Mắt đóng}, & EAR < 0.2 \\ \text{Mắt mở}, & \text{Ngược lại} \end{cases} \quad (4.3)$$

Dựa vào Bảng 4.2, Luận án định nghĩa 4 lớp thể hiện mức độ buồn ngủ theo cấp độ từ 1 đến 4 và hai thuộc tính cần xem xét là EAR và Thời lượng (s). mô hình quyết định mức độ buồn ngủ như sau:



Hình 4.6 Mô hình ước lượng mức độ buồn ngủ

4.2.2 Đánh giá sự tập trung dựa vào mức độ buồn ngủ

Dựa vào bảng Bảng 4.3 luận án xây dựng thang độ đánh giá sự tập trung cho hệ thống gồm 3 mức độ

Bảng 4.3 Mức độ tập trung

STT	Mức độ buồn ngủ	Mức độ tập trung
1	Thức giấc/Tỉnh táo	Tập trung cao
2	Hơi buồn ngủ	Tập trung thấp
3	Ngủ gật	Mất tập trung
	Bắt đầu giấc ngủ	

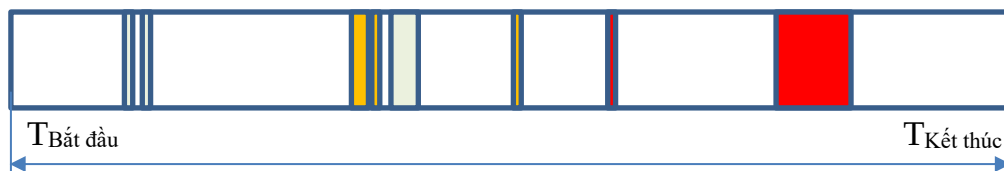
Do đây là hệ thống phân tích đánh giá sự tập trung của người học, nên nhiệm vụ cần được thực hiện cả một quá trình, có nghĩa là hệ thống phải xem xét và đánh giá người học từ lúc bắt đầu video cho đến kết thúc video. Không giống như các hệ thống khác như phát hiện ngủ gật trong lúc lái xe chẳng hạn khi phát hiện tài xế mất tập trung, hệ thống cảnh báo và đưa ra kết luận tại thời điểm đó. Sở dĩ Luận án đặt ra điều kiện này bởi các lý do:

- (i) Giả sử tại khung hình thứ i người học có thể mất tập trung nhưng sang đến khung hình thứ k thì người học tập trung trở lại. Do đó, đánh giá sự tập trung của người học dựa vào khung hình hiện tại là không phù hợp
- (ii) Trong quá trình học tập, người học có thể có nhiều mức độ tập trung ở những thời điểm khác nhau. Vì vậy, dựa vào kết quả phân tích hiện tại để đánh giá cả quá trình học tập là không phù hợp.
- (iii) Người học tập trung trong suốt quá trình học tập, nhưng tại khung hình cuối cùng có hành vi đóng mắt, hoặc trường hợp ngược lại. Do đó, hệ thống đánh giá dựa vào khung hình cuối cùng là thiếu chính xác.

Chính vì vậy, Luận án đề xuất mô hình đánh giá sự tập trung bằng cách tính tổng thời gian của mỗi trạng thái, sau đó lập tỷ lệ mỗi trạng thái so với tổng thời lượng của video. Trạng thái nào chiếm tỷ lệ cao nhất thì sẽ quyết định mức độ tham gia của người học trong suốt quá trình học tập. Đây là phương pháp đề xuất mang tính chủ quan tuy nhiên kết quả từ hệ thống đưa ra rất phù hợp và dễ dàng thực hiện được.

Cách đánh giá được trình bày chi tiết như sau:

Giả sử quá trình tham gia học tập của người học được mô phỏng như Hình 4.7 gồm có nhiều trạng thái và xảy ra ở những thời điểm khác nhau



Hình 4.7 Mô phỏng trạng thái của quá trình học tập

Trong đó

- $T_{\text{Bắt đầu}}$: Thời gian bắt đầu học
 - $T_{\text{Kết thúc}}$: Thời gian kết thúc học
- | | |
|--|------------------------|
| | 1 - Thức giấc/Tỉnh táo |
| | 2 - Hơi buồn ngủ |
| | 3 - Ngủ gật |
| | 4 - Bắt đầu giấc ngủ |

Gọi t_1 là tổng thời gian của mức độ 1 và t_2, t_3, t_4 là tổng thời gian tương ứng cho các mức độ còn lại.

Gọi r_1, r_2, r_3, r_4 là các tỷ lệ tương ứng của các mức độ t_1, t_2, t_3, t_4

Tỷ lệ của các mức độ được xác định:

$$\begin{aligned} r_1 &= \frac{t_1}{T_{\text{Kết thúc}} - T_{\text{Bắt đầu}}} & r_2 &= \frac{t_2}{T_{\text{Kết thúc}} - T_{\text{Bắt đầu}}} \\ r_3 &= \frac{t_3}{T_{\text{Kết thúc}} - T_{\text{Bắt đầu}}} & r_4 &= \frac{t_4}{T_{\text{Kết thúc}} - T_{\text{Bắt đầu}}} \end{aligned} \quad (4.4)$$

Mức độ tập trung được xác định như sau:

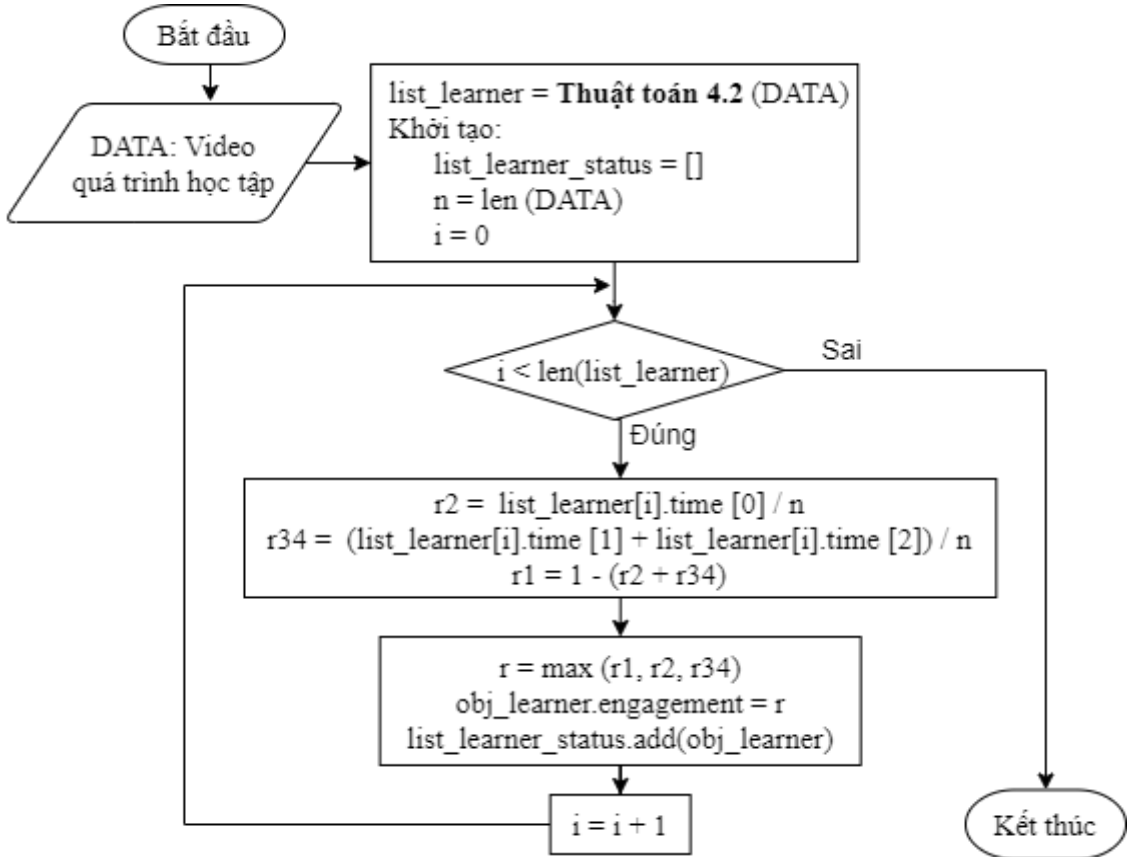
$$x = \text{MAX}(r_1, r_2, r_3 + r_4) = \begin{cases} \text{Tập trung cao, } x = r_1 \\ \text{Tập trung thấp, } x = r_2 \\ \text{Không tập trung, } x = r_3 + r_4 \end{cases} \quad (4.5)$$

4.2.3 Thuật toán đánh giá sự tập trung

a) **Thuật toán 4.1:** Đánh giá sự tập trung của người học (Xem phụ lục 2)

Đầu vào: Video quá trình tham gia học tập của người học

Đầu ra: Mức độ tập trung của những người học

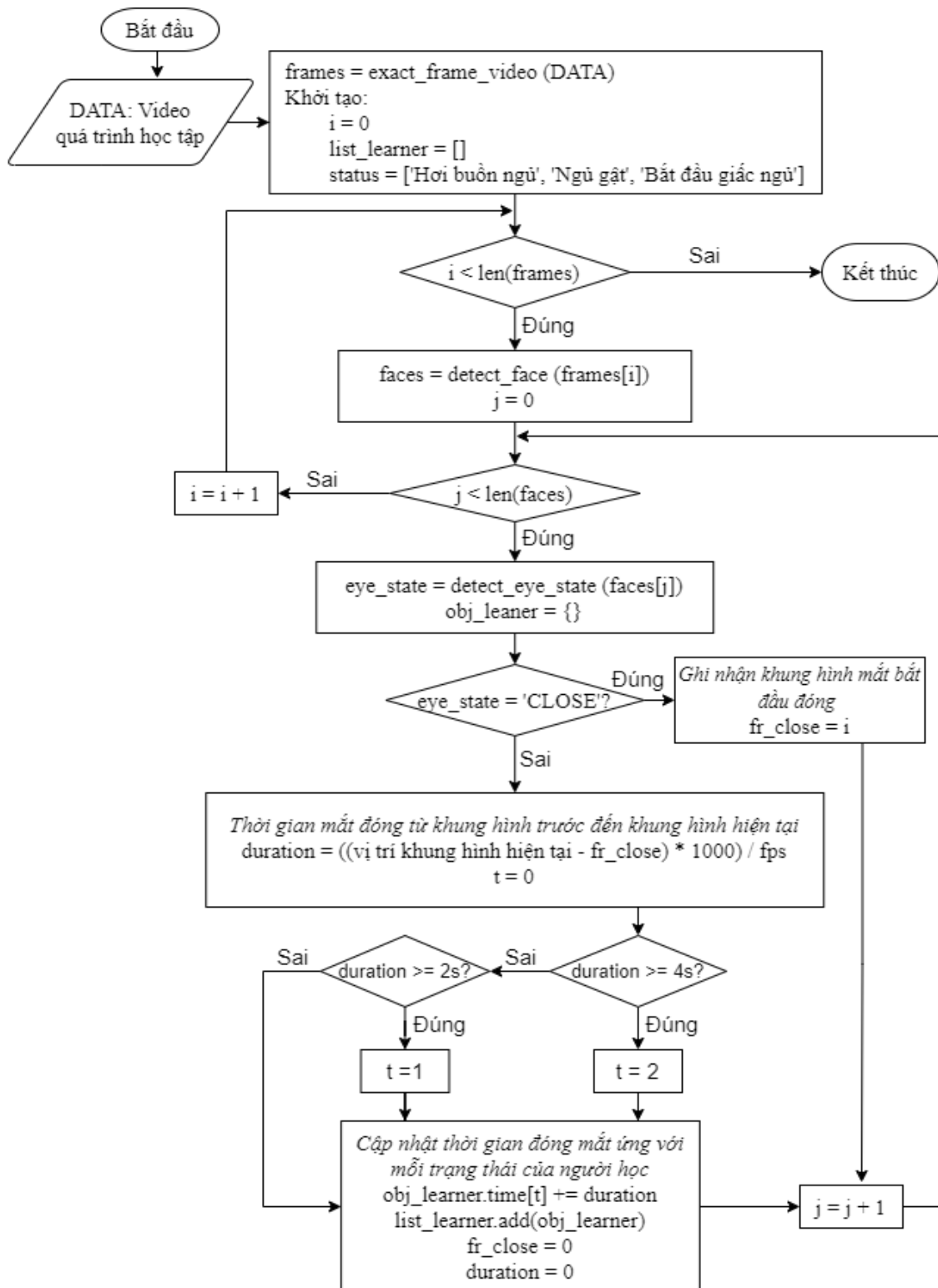


Hình 4.8 Lưu đồ đánh giá sự tập trung của người học

b) **Thuật toán 4.2:** Tính toán thời gian tham gia học tập của người học

Đầu vào: Video quá trình học tập

Đầu ra: Thời gian tham gia học tập của những người học ở từng mức độ



Hình 4.9 Lưu đồ tính toán thời gian tham gia học tập của người học

4.2.4 Độ phức tạp của thuật toán

Thời gian thực hiện thuật toán được xác định, bao gồm trích xuất khung hình từ video, xác định hành vi mở / nhắm mắt của các đối tượng trong khung hình đó và các công thức xác định thời lượng mắt đóng, cập nhật thời lượng mắt đóng cho mỗi trạng thái tương ứng như sau:

Gọi: l_v là thời lượng của video

n_o là số đối tượng trong lớp học

Bảng 4.4 Độ phức tạp của kỹ thuật đánh giá sự tập trung của người học dựa trên trạng thái của mắt

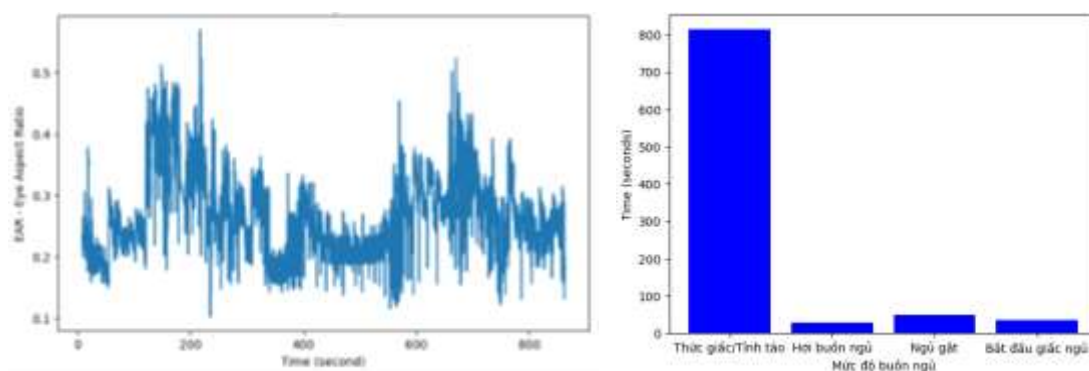
STT	Bước xử lý	Thời gian thực hiện
Thuật toán 4.2		
1	Trích xuất các khung hình	$O(l_v)$
2	Phát hiện khuôn mặt của mỗi đối tượng	$O(n_o)$
3	Xác định trạng thái đóng mở của mắt	$O(1)$
	Tính toán thời gian đóng mắt của người học từ khung hình có mắt đóng đến khung hình hiện tại	$O(1)$
	Xác định status dựa vào thời lượng đóng mắt	$O(1)$
	Cập nhật lại thời gian của người học tương ứng với mỗi trạng thái	$O(1)$
4	Tổng cộng (1)	$O(l_v \times n_o)$
Thuật toán 4.1		
5	Tính toán thời gian tham gia học tập của người học ở mỗi cấp độ (Thuật toán 4.2)	$O(l_v \times n_o)$
6	Đánh giá mức độ tham gia của người học	$O(n_o)$
7	Tổng cộng (2)	$O(l_v \times n_o) + O(n_o)$
8	Tổng cộng	$O(l_v \times n_o) + O(n_o)$

Độ phức tạp của thuật toán phụ thuộc vào độ dài của video và số đối tượng trong lớp học. Trường hợp tốt nhất là đánh giá một sinh viên trong trường hợp học trực tuyến, nghĩa là lúc này $n_o = 1$. Khi đó, độ phức tạp của thuật toán là $O(l_v \times 1) + O(1) = O(l_v)$. Hiện tại, Tác giả quan tâm đến yếu tố chính xác. Vì vậy, hướng phát triển trong tương lai, Tác giả tích hợp thêm công nghệ xử lý tính toán chẳng hạn như GPU nhằm cải thiện tốc độ xử lý cho kỹ thuật.

4.2.5 Thử nghiệm

Luận án thực nghiệm kiểm tra các mục tiêu đã đặt ra như: ước lượng mức độ buồn ngủ dựa trên hành vi đóng/mở mắt và cuối cùng là đánh giá sự tập trung dựa trên mức độ buồn ngủ. Video được ghi nhận tại hai lớp khác nhau trong giờ ra chơi ngành kỹ thuật tại trường Đại học Lạc Hồng, độ dài thử nghiệm của mỗi video là khoảng 10 - 15 phút.

Trường hợp 1



(a) trạng thái đóng/mở mắt

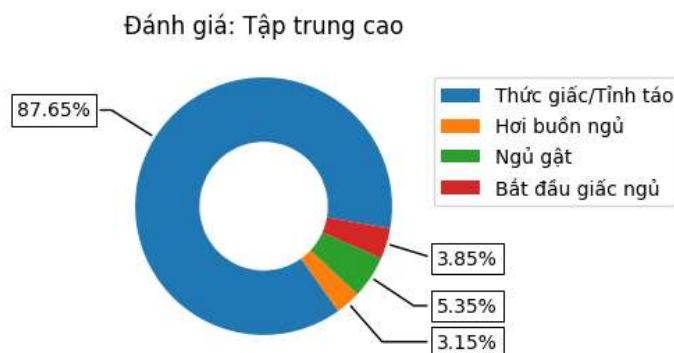
(b) ước lượng mức độ buồn ngủ

Hình 4.10 Kết quả ước lượng mức độ buồn ngủ của một sinh viên

Trong đó các chỉ số được hệ thống đo chi tiết từ Hình 4.10 như sau:

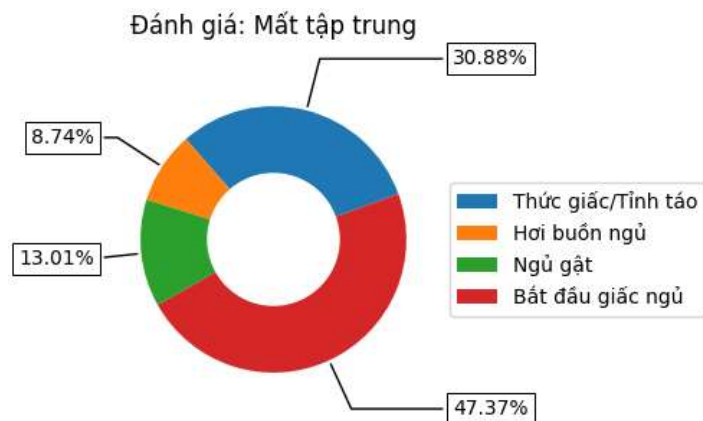
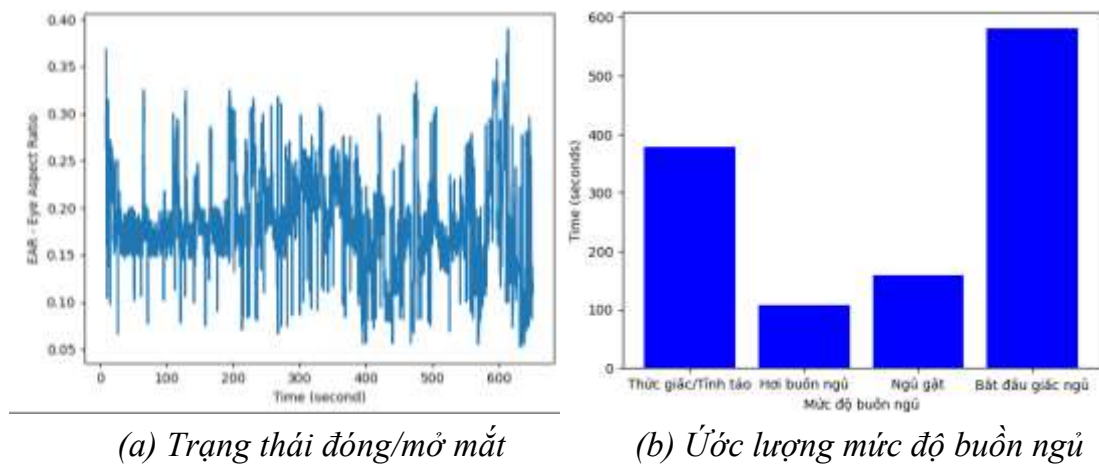
- Thức giấc/tỉnh táo: 813.99s chiếm 87.65%
- Hơi buồn ngủ: 29.22s chiếm 3.15%
- Ngủ gật: 49.65s chiếm 5.35%
- Bắt đầu giấc ngủ: 35.78s chiếm 3.85%

Tỷ lệ đạt cao nhất của các mức độ là 87.65%. Do đó, dựa vào Bảng 4.3 hệ thống kết luận đối tượng tập trung cao trong giờ học xem Hình 4.11



Hình 4.11 Kết quả đánh giá sự tập trung của một sinh viên

Trường hợp 2



(c) Đánh giá sự tập trung dựa trên mức độ buồn ngủ

Hình 4.12 Minh họa kết quả đánh giá sự tập trung của một sinh viên

4.3 Kỹ thuật phát hiện trạng thái đóng/mở mắt

4.3.1 Giới thiệu

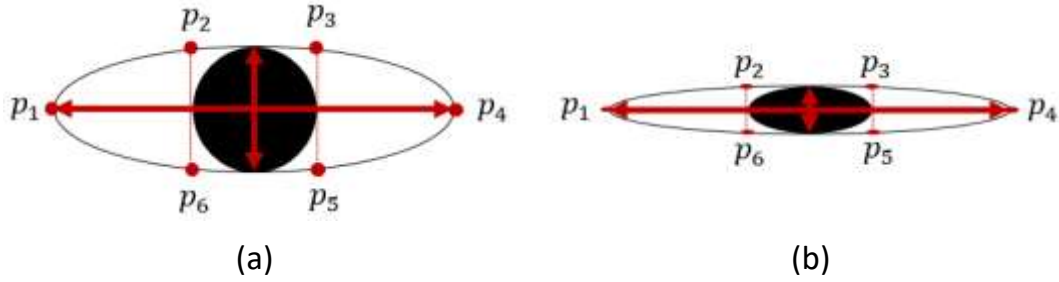
Để phát hiện trạng thái đóng/mở của mắt, kỹ thuật cần thực hiện các công việc liên quan như sau: phát hiện khuôn mặt người trong ảnh, phát hiện vùng mắt. Để phát hiện khuôn mặt người, Luận án sử dụng kỹ thuật được đề xuất bởi [22]; để phát hiện mắt người trong ảnh, Luận án sử dụng các dấu mốc trên khuôn mặt (Facial Landmarks) được trích xuất từ 68 điểm tương ứng với các đặc điểm trên khuôn mặt được ước tính bằng cách sử dụng mô hình phát hiện dấu mốc được đào tạo trước từ thư viện Dlib [24].

Kỹ thuật này thực hiện hiệu quả, phù hợp với các ứng dụng yêu cầu chạy thời gian thực. Tuy nhiên, như đã thảo luận ở các phần trước, kỹ thuật này vẫn còn tồn tại hạn chế cần được khắc phục.

4.3.2 Cơ sở lý thuyết của phương pháp cần cải tiến

Như đã trình bày trong phần 4.2, kỹ thuật đánh giá mức độ tham gia của người học dựa trên trạng thái đóng/mở của mắt. Trong đó, việc xác định trạng thái của mắt được tính toán dựa trên tỷ lệ giữa chiều cao và chiều rộng của mắt.

Để tính toán tỷ lệ khung hình của mắt, theo Hình 4.13, mắt của người được biểu thị bằng một tập hợp 6 điểm, bắt đầu bằng góc trái của mắt và được đánh dấu theo chiều kim đồng hồ. Mỗi điểm đại diện cho 1 tọa độ trong không gian 2D với $p_1(x_1, y_1)$ và tương ứng cho các điểm còn lại từ p_2 đến p_6 .



(a) Tỷ lệ khung hình của mắt khi mở; (b) Tỷ lệ khung hình của mắt khi đóng

Hình 4.13 Tọa độ và tỷ lệ khung hình của mắt

Chiều rộng của mắt là khoảng cách giữa điểm p_1 và p_4 được xác định như công thức (4.6) theo Euclid.

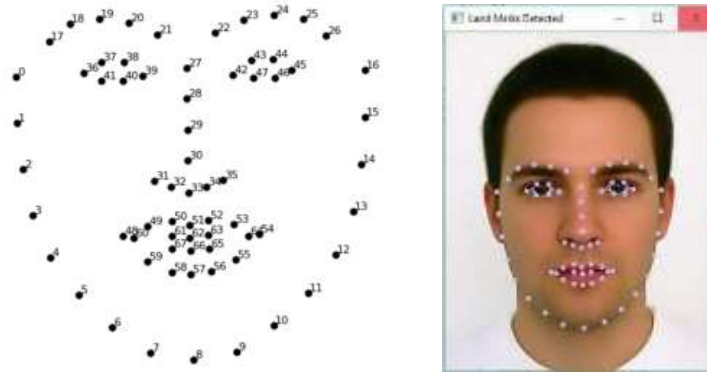
$$D(p_1, p_4) = \sqrt{(p_1.x_1 - p_4.x_4)^2 + (p_1.y_1 - p_4.y_4)^2} \quad (4.6)$$

Chiều cao của mắt là khoảng cách giữa các cặp điểm (p_2, p_6) và (p_3, p_5) được xác định theo Euclid tương tự công thức (4.6). Khoảng cách từ p_1 đến p_4 luôn luôn là hằng số, trong khi chiều cao sẽ thay đổi tùy thuộc vào độ mở hoặc nhắm của mắt. Để phát hiện sự nhấp nháy bằng cách tính tỷ lệ giữa chiều cao trung bình và chiều dài của mắt. Tỷ lệ này sẽ gần như không đổi khi mắt mở và nó sẽ nhanh chóng giảm xuống 0 khi xảy ra mắt đóng. Công thức tỷ lệ khung hình của mắt (ký hiệu: *EAR* - Eye Aspect Ratio) được xác định như sau:

$$EAR = \frac{D(p_2, p_6) + D(p_3, p_5)}{2 \times D(p_1, p_4)} \quad (4.7)$$

Trong đó: $p_1, p_2, p_3, p_4, p_5, p_6$ đại diện cho các tọa độ tương ứng trên mắt

- Với mắt trái bao gồm các tọa độ từ p_{36} đến p_{41} , mắt phải bao gồm các tọa độ từ p_{42} đến p_{47} tính theo chiều kim đồng hồ (xem Hình 4.14)



Hình 4.14 68 điểm đặc trưng trên khuôn mặt

Ưu điểm của phương pháp này là dễ cài đặt, không cần đào tạo trước về dữ liệu mắt đóng hay mở. Kỹ thuật phù hợp với các ứng dụng yêu cầu thời gian thực. Tuy nhiên, kỹ thuật này vẫn còn hạn chế là phải so sánh với một ngưỡng cố định. Vấn đề này không khái quát được hết tất cả đối tượng, vì đặc trưng mắt của mỗi người là khác nhau (người có mắt to, người có mắt nhỏ, ...) nên khi so sánh với ngưỡng cố định dẫn đến thiếu chính xác trong vài trường hợp khi phát hiện. Do đó, Luận án đề xuất một cải tiến cho phương pháp này là xác định ngưỡng tự động tùy vào đặc trưng mắt của mỗi đối tượng.

4.3.3 Cải tiến kỹ thuật phát hiện trạng thái đóng/mở mắt dựa vào biến thiên tỷ lệ khung hình

Do kích thước của đôi mắt không giống nhau giữa người này với người kia, vì vậy một người có đôi mắt to và một người có đôi mắt nhỏ thì sẽ không có cùng chiều cao và các giá trị tọa độ. Điều này dẫn đến sự thiếu chính xác trong việc xác định hành vi đóng/mở mắt khi so sánh giá trị EAR với ngưỡng cố định (EAR nằm trong ngưỡng nào đó thì mắt sẽ mở hoặc đóng). Để cải tiến điều này, luận án đưa ra giải pháp xác định hành vi đóng/mở mắt linh động dựa trên độ biến thiên của khung hình chứa mắt ứng với mỗi cá nhân như sau:

Khung hình chứa mắt ở trạng thái bình thường

$$EAR_{MAX} = \frac{D(p_2, p_6)_{max} + D(p_3, p_5)_{max}}{2 \times D(p_1, p_4)} \quad (4.8)$$

Khung hình chứa mắt hiện hành

$$EAR_{CURRENT} = \frac{D(p_2, p_6)_{current} + D(p_3, p_5)_{current}}{2 \times D(p_1, p_4)} \quad (4.9)$$

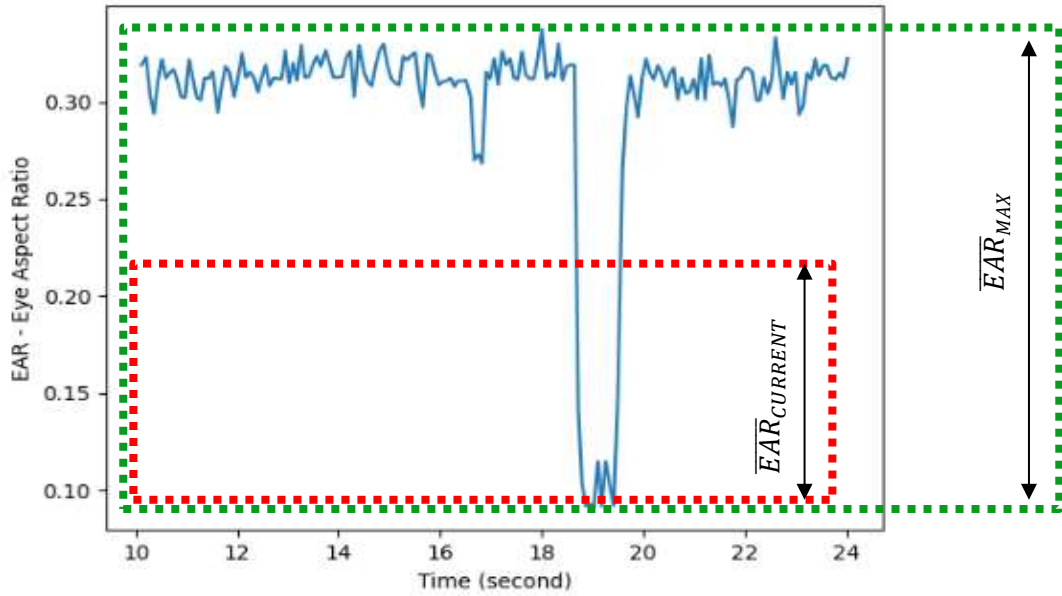
Giả sử hành vi đóng/mở mắt đối với người là xảy ra đồng thời ở cả hai mắt, nên tỷ lệ khung hình chứa mắt ở trạng thái bình thường và trạng thái hiện hành được xác định lần lượt như sau:

$$\begin{aligned}\overline{EAR}_{MAX} &= \frac{EAR_{MAX_L} + EAR_{MAX_R}}{2} \\ \overline{EAR}_{CURRENT} &= \frac{EAR_{CURRENT_L} + EAR_{CURRENT_R}}{2}\end{aligned}\quad (4.10)$$

Trong đó: EAR_{MAX_L} là tỷ lệ khung hình bình thường của mắt trái
 EAR_{MAX_R} là tỷ lệ khung hình bình thường của mắt phải
 $EAR_{CURRENT_L}$ là tỷ lệ khung hình hiện hành của mắt trái
 $EAR_{CURRENT_R}$ là tỷ lệ khung hình hiện hành của mắt phải

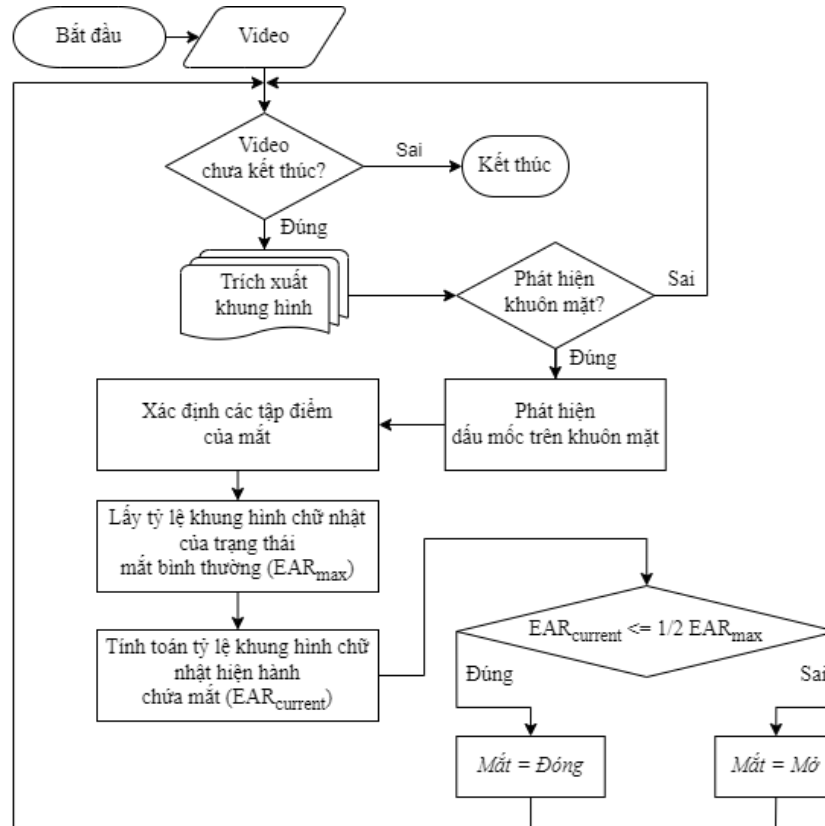
Cuối cùng, để ra quyết định mắt đóng hay mở, kỹ thuật so sánh giá trị $\overline{EAR}_{MAX}$ với giá trị $\overline{EAR}_{CURRENT}$ như sau

$$x = \overline{EAR}_{CURRENT} - \frac{\overline{EAR}_{MAX}}{2} = \begin{cases} \text{EYE CLOSE, } x < 0 \\ \text{EYE OPEN, } x \geq 0 \end{cases} \quad (4.11)$$



Hình 4.15 Biến thiên tỷ lệ của khung hình trong hành vi đóng/mở mắt

4.3.3.1 Lưu đồ giải thuật của Blink detection+



Hình 4.16 Lưu đồ giải thuật phát hiện trạng thái đóng/mở mắt

4.3.3.2 Độ phức tạp của thuật toán

Thời gian thực hiện thuật toán được xác định, bao gồm trích xuất khung hình từ video, xác định hành vi mở / nhắm mắt của các đối tượng trong khung hình đó và công thức tính tỷ lệ của mắt được thể hiện như sau:

Gọi: l_v là thời lượng của video

n_o là số đối tượng trong mỗi khung hình

Bảng 4.5 Độ phức tạp của thuật toán

STT	Bước xử lý	Thời gian
1	Trích xuất các khung hình	$O(l_v)$
2	Phát hiện khuôn mặt của mỗi đối tượng	$O(n_o)$
3	Trích xuất điểm đặc trưng trên khuôn mặt	$O(1)$
	Xác định tọa độ các điểm thuộc về mắt	$O(1)$
	Tính toán khung hình hiện tại chứa mắt	$O(1)$
	So sánh khung hình hiện tại với khung hình cực đại chứa mắt	$O(1)$
4	Tổng cộng	$O(l_v \times n_o)$

4.3.3.3 Đánh giá thử nghiệm

Phương pháp Blink detection+ so với phương pháp [16], [115] và [116] dựa trên bốn tập dữ liệu: **Talking Face**, **Eyeblink8**, **HUST-LEBW** và **ZJU** (những tập dữ liệu này được trình bày chi tiết trong Chương 2). Kết quả thực nghiệm của kỹ thuật được trình bày trong Bảng 4.6.

Hiệu suất của kỹ thuật được đánh giá bằng các công thức sau:

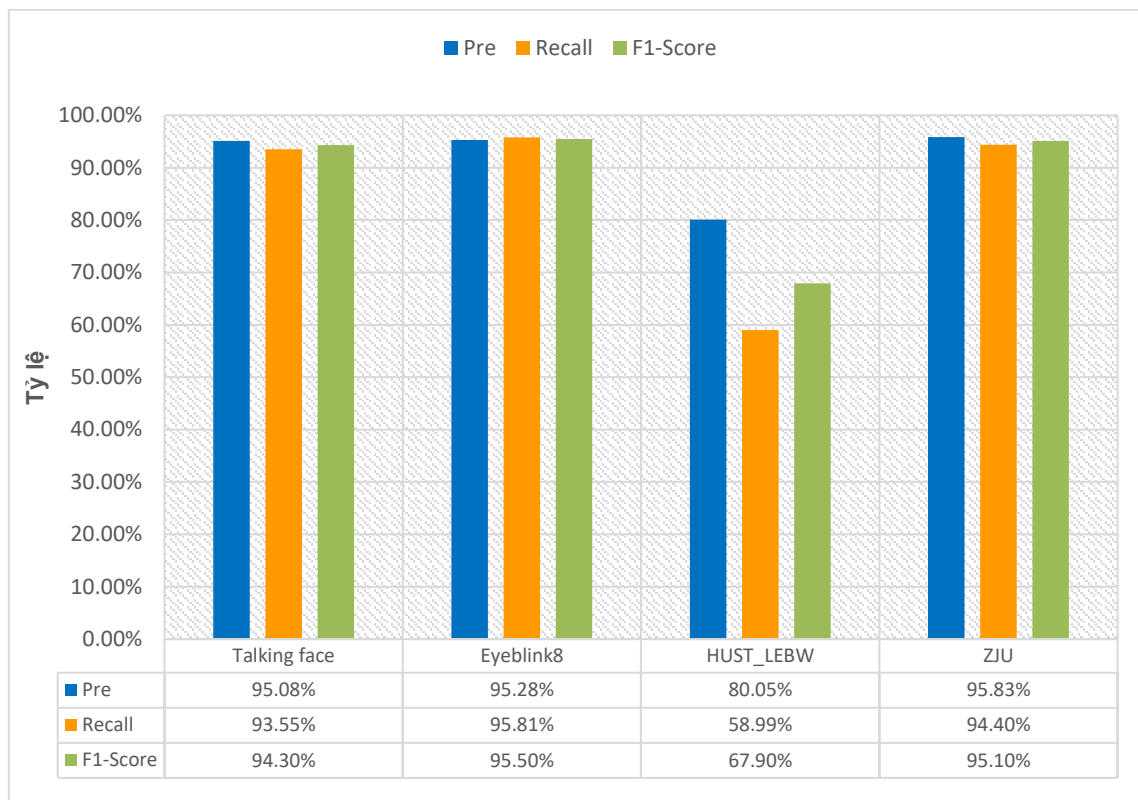
$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 Recall &= \frac{TP}{TP + FN} \\
 F1 - Score &= \frac{2 \times (Precision \times Recall)}{Precision + Recall}
 \end{aligned} \tag{4.12}$$

Bảng 4.6 Kết quả thực nghiệm

Tập dữ liệu	SA	EC	TP	FP	FN	Pre	Recall	F1-Score
Talking face	1	61	58	3	4	0.9508	0.9355	0.9430
Eyeblink8	1	38	35	3	2	0.9211	0.9459	0.9330
	2	88	86	2	2	0.9773	0.9773	0.9770
	3	65	63	2	1	0.9692	0.9844	0.9770
	4	31	30	1	2	0.9677	0.9375	0.9520
	5	30	28	2	2	0.9333	0.9333	0.9330
	6	41	39	2	1	0.9512	0.9750	0.9630
	7	72	70	2	3	0.9722	0.9589	0.9660
	8	43	40	3	2	0.9302	0.9524	0.9410
						0.9528	0.9581	0.9550
HUST_LEBW	673	381	305	76	212	0.8005	0.5899	0.6790
ZJU	80	264	253	11	15	0.9583	0.9440	0.9510

Trong đó:

- **SA**: Mẫu
- **EC**: Số lượng mắt đóng
- **Pre**: Precision

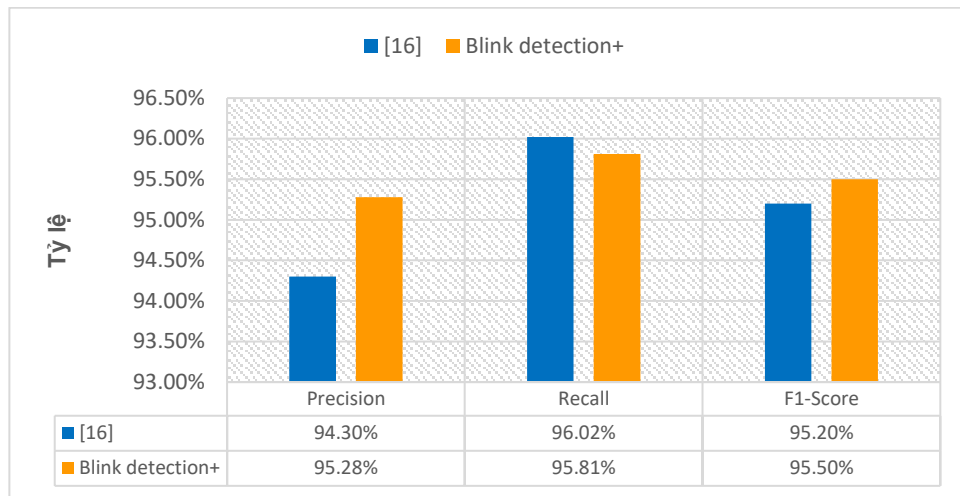


Hình 4.17 Kết quả thực nghiệm của kỹ thuật cải tiến trên các tập dữ liệu Talking face, Eyeblink8, HUST_LEBW, ZJU

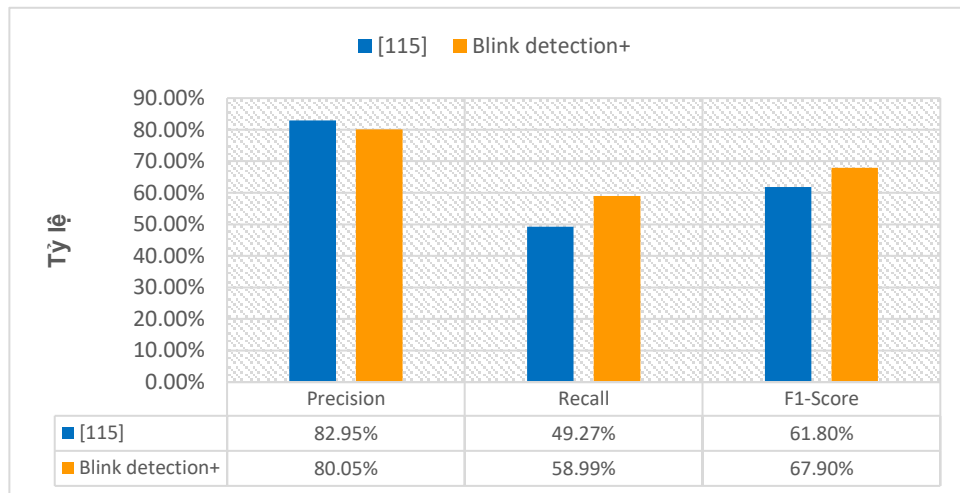
Biểu đồ cho thấy, Blink detection+ thực hiện hiệu quả trên tập dữ liệu Talking face, Eyeblink8, ZJU và kém hiệu quả trên tập dữ liệu HUST_LEBW. Do tập dữ liệu này chứa nhiều thách thức trong môi trường tự nhiên như đeo kính râm, hình ảnh nghiêng, v.v.

Bảng 4.7 Kết quả so sánh với các giải pháp

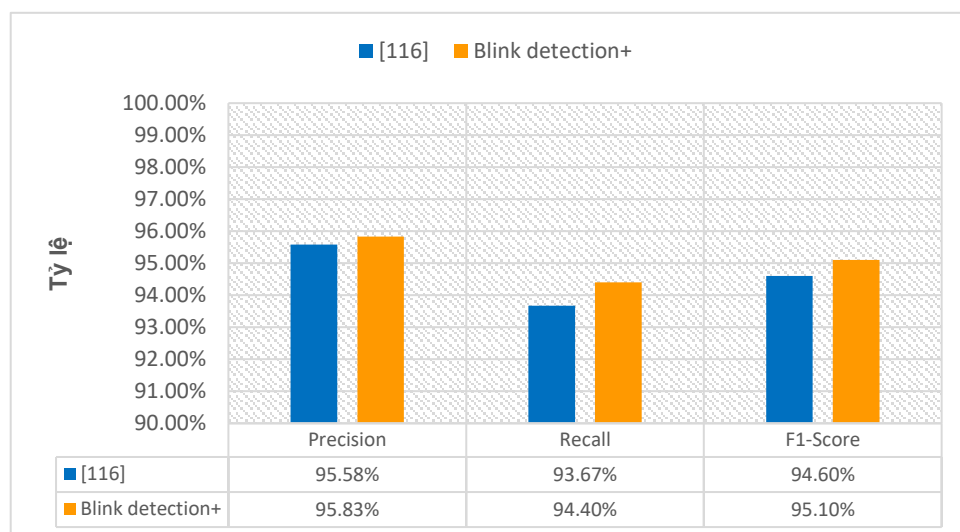
Giải pháp	Tập dữ liệu	Precision	Recall	F1-Score
[16]	Talking face	-	-	
Blink detection+	Talking face	0.9508	0.9355	0.9430
[16]	Eyeblink8	0.9430	0.9602	0.9520
Blink detection+	Eyeblink8	0.9528	0.9581	0.9550
[115]	HUST_LEBW	0.8295	0.4927	0.6180
Blink detection+	HUST_LEBW	0.8005	0.5899	0.6790
[116]	ZJU	0.9558	0.9367	0.9460
Blink detection+	ZJU	0.9583	0.9440	0.9510



Hình 4.18 Biểu đồ so sánh kỹ thuật cải tiến so với công trình [16]



Hình 4.19 Biểu đồ so sánh kỹ thuật cải tiến so với công trình [115]



Hình 4.20 Biểu đồ so sánh kỹ thuật cải tiến so với công trình [116]

Thông qua thực nghiệm phương pháp đề xuất thực hiện có độ chính xác cao hơn so với các phương pháp đề xuất trước đó. Phương pháp cho thấy rằng dựa trên dấu mốc đủ để ước tính chính xác mức độ đóng/mở của mắt. Tuy nhiên phương pháp này vẫn còn tồn tại khuyết điểm và thực hiện kém hiệu quả trong tình huống khi mà đối tượng nghiêng đầu qua trái hoặc qua phải nhiều, điều này làm cho phương pháp không nhận dạng được hoàn toàn khuôn mặt nên đã bỏ qua một vài trường hợp khi phát hiện.

4.4 Kết chương

Phương pháp đánh giá mức độ tham gia của người học dựa trên trạng thái đóng/mở của mắt phù hợp để triển khai trong cả môi trường học truyền thống và trực tuyến. Tuy nhiên, phương pháp này vẫn còn hạn chế là hiệu chỉnh mắt thích hợp để nhận được độ chính xác [117], tức là hạn chế chuyển động của người tham gia để duy trì trong phạm vi của trình theo dõi mắt điều này là không khả thi trong môi trường giáo dục ngoài đời thực, và để xác định hành vi đóng/mở của mắt phải dựa trên một ngưỡng cố định [16] [18]. Đây là điểm hạn chế của phương pháp này vì đặc trưng mắt của mỗi người khác nhau.

Vì vậy, trong phần này của Luận án, đề xuất một kỹ thuật phát hiện trạng thái đóng/mở của mắt dựa trên tỷ lệ khung hình chữ nhật chứa mắt. Kỹ thuật linh động xác định ngưỡng khi xảy ra đóng mắt phù hợp với mọi đối tượng. Kỹ thuật đã được cài đặt thử nghiệm và kết quả cho thấy điểm số *F1-Score* tốt hơn với các phương pháp còn lại và phù hợp để áp dụng cho việc giải quyết một phần công việc trong luận án.

CHƯƠNG 5

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Luận án đã hoàn thành được mục tiêu nghiên cứu đã đề ra nhằm mục đích xây dựng một hệ thống hỗ trợ giáo viên phân tích và đánh giá sự tập trung của người học. Luận án tập trung vào hành vi và biểu cảm của người học để đánh giá. Hệ thống hoạt động như một trợ lý cho quá trình ra quyết định một cách tự động. Luận án đã hoàn thành việc xây dựng hệ thống, trực quan hóa dữ liệu và đã đạt được kết quả nghiên cứu như sau:

- (i) Đưa ra phương pháp đánh giá mức độ tham gia của người học dựa trên hành vi đóng/mở mắt. Quy trình đánh giá sự tập trung được xem xét và phân tích trên cả quá trình học tập của người học. Trước tiên, luận án đánh giá trên từng thời điểm riêng lẻ, sau đó tổng hợp lại và tính tỷ lệ của mỗi trạng thái so với cả quá trình (vì trong quá trình học tập, người học có nhiều mức độ tham gia ở những thời điểm khác nhau).
- (ii) Cải tiến kỹ thuật phát hiện trạng thái đóng/mở của mắt với ngưỡng được xác định linh động phù hợp cho mọi đối tượng. Do tính chất đặc trưng về mắt của mỗi người là khác nhau, dẫn đến độ cao và độ rộng của mắt ở mỗi người cũng khác nhau. Nên khi sử dụng một ngưỡng cố định để xác định sẽ không bao quát hết mọi trường hợp.
- (iii) Kỹ thuật phân rã biểu cảm thành phần cơ bản[28] nhằm giải quyết vấn đề về khan hiếm dữ liệu huấn luyện. Do trong quá trình học, người học luôn luôn tiếp nhận những thông tin từ giáo viên, bạn bè, tác động từ môi trường v.v. Vì vậy, cảm xúc của người học ở mỗi thời điểm xảy ra khác nhau, có thể là hỗn hợp hoặc đơn lẻ nên cảm xúc cũng thể hiện những biến dạng tương ứng.

Hướng phát triển

Mặc dù luận án đã đề xuất kết quả nghiên cứu, tuy nhiên vẫn còn tiềm ẩn một số hạn chế. Do vấn đề giám sát hành vi của người học được đặt ra với nhiều yêu cầu khắt khe và chặt chẽ, việc đánh giá cần được xem xét và phân tích trong suốt quá trình học tập.

- (i) Hạn chế đầu tiên của luận án bao gồm việc thiếu theo dõi các thông tin hữu ích khác, chẳng hạn như hướng đầu, giờ tay phát biểu, ngáp... Nên kỹ thuật chưa bao quát được các tình huống trong học tập
- (ii) Thứ hai là không thể nhận dạng tất cả các điểm mốc trên khuôn mặt cho tất cả người học vì lý do bị che khuất một phần hoặc do họ xoay đầu hướng khác, không hướng về giáo viên, nên hệ thống không có đủ dữ liệu để phân tích trong những trường hợp như vậy, do đó dẫn đến thiếu chính xác trong kết quả.

Từ những nhược điểm trên, tác giả đề xuất ý tưởng kết hợp nhiều yếu tố mà không phải phụ thuộc vào một phương pháp cố định, từ đó có cơ sở để ra quyết định tốt hơn chẳng hạn kết hợp việc đóng mở mắt để phát hiện buồn ngủ, hướng đầu để phát hiện tư thế ngồi học và biểu hiện trên khuôn mặt để đánh giá. Vì mỗi phương pháp chạy đơn lẻ có khả năng bị giới hạn trong một số trường hợp khi phân tích dữ liệu.

Bên cạnh đó, dựa trên kết quả ban đầu đạt được từ công trình [65]. Trong tương lai, Luận án mong muốn xem xét thêm yếu tố quỹ đạo chuyển động của đầu dựa trên lời thoại nhằm đánh giá sự tương quan giữa cử động và nội dung. Điều này có thể giúp chúng ta hiểu rõ hơn về mối quan hệ giữa tâm trạng và hành vi học tập của người học. Qua đó, luận án cũng hướng đến việc mô phỏng chân thực về cách mà người học tham gia trong quá trình học tập. Bằng cách theo dõi và tái tạo lại các cử động. Nghiên cứu này cho thấy khả năng tiềm ẩn trong việc đo lường sự tập trung dựa trên các dấu hiệu về hình thái cơ thể. Điều này có thể cung cấp một cái nhìn sâu hơn về tình trạng tâm lý và tình hình của người học trong thời gian học tập.

Ngoài ra, Luận án này còn có hai điểm cần hoàn thiện như sau: (1) Do các phương pháp được các tác giả khác công bố không thử nghiệm trên dữ liệu chuẩn mà thử nghiệm trên các dữ liệu riêng rẽ được thu thập trong nghiên cứu của họ cho nên việc so sánh tính hiệu quả của các phương pháp được đề xuất trong Luận án này với những phương pháp khác gặp khó khăn; (2) quy mô và dữ liệu thử nghiệm còn “*khá khiêm tốn*” nên tính hiệu quả của phương pháp được đề xuất cần được xem xét thêm.

DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ

QUỐC TẾ

- HTQT1. **Tran Thanh Phuong**, Lam Thanh Hien, Do Nang Toan, Ngo Duc Vinh (2022), “An Eye Blink detection technique in video surveillance based on Eye Aspect Ratio”, *Proceedings of International Conference on Advanced Communication Technology (ICACT)*, 13-16 Feb. 2022, Pyeong Chang, Korea, pp534-538
- TCQT1. **Tran Thanh Phuong**, Lam Thanh Hien, Ngo Duc Vinh, Ha Manh Toan, Do Nang Toan (2023), “An algorithm for decomposing variations of a 3D model”, *International Journal of Electrical and Computer Engineering (IJECE)*.

TRONG NƯỚC

- HTTN1. **Trần Thanh Phương**, Lâm Thành Hiền, Đỗ Năng Toàn, Ngô Đức Vĩnh (2021), “Một kỹ thuật phát hiện trạng thái đóng/mở của mắt dựa vào đặc trưng tỷ lệ”, *Kỷ yếu Hội thảo Quốc gia ”Một số vấn đề chọn lọc của CNTT và Truyền thông”*, Thái Nguyên 13-14/12/2021, tr265-269.
- TCTN1. **Trần Thanh Phương**, Trần Văn Lăng, Đỗ Năng Toàn (2021), “A technique of assessing student engagement based on eye close/open behavior”, *TNU Journal of Science and Technology*, Vol 226, No.07, tr262-269.
- TCTN2. **Trần Thanh Phương**, Ngô Đức Vĩnh, Hà Mạnh Toàn, Nông Minh Ngọc, Đỗ Năng Toàn (2023), “Một kỹ thuật mô phỏng cử động của mô hình đầu người 3D theo lời thoại tiếng việt”, *TNU Journal of Science and Technology*, Vol 228, No.15, tr20-28.

TÀI LIỆU THAM KHẢO

- [1] Michael ER Nicholls, Kellie M Loveless, Nicole A Thomas, Tobias Loetscher, and Owen Churches, "Some participants may be better than others: Sustained attention and motivation are higher early in semester," *The Quarterly Journal of Experimental Psychology*, vol. 68, no. 1, pp. 10-18, 2015.
- [2] Anthony F. Grasha, "Teaching with Style: A Practical Guide to Enhancing Learning by Understanding Teaching and Learning Styles," *Alliance Publishers: San Barnadino, CA, USA*, p. 167–174, 2002.
- [3] Michelle Richardson, Charles Abraham, "Modeling antecedents of university students' study behavior and grade point average," *Journal of Applied Social Psychology*, vol. 43, no. 3, p. 626–637, 2013.
- [4] Bui Ngoc Anh, Ngo Tung Son, Phan Truong Lam, Le Phuong Chi, Nguyen Huu Tuan, Nguyen Cong Dat, Nguyen Huu Trung, Muhammad Umar Aftab, Tran Van Dinh, "A Computer-Vision Based Application for Student Behavior Monitoring in Classroom," *Applied sciences*, vol. 9, no. 22, 2019.
- [5] Rui Zheng, Fei Jiang, Ruimin Shen, "Intelligent student behavior analysis system for real classrooms.," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 9244-9248. <https://doi.org/10.1109/ICASSP40776.2020.9053457>, 2020.
- [6] M. Ali Akber Dewan, Mahbub Murshed and Fuhua Lin, "Engagement detection in online learning: a review," *Springer Open*, vol. 6, no. 1, 2019.
- [7] Magdalena Cantabella, Raquel Martínez, España Belén, Ayuso Juan, Antonio Yáñez and Andrés Muñoz, "Analysis of student behavior in learning management systems through a Big Data framework," *Future Generation Computer Systems*, vol. 90, pp. 262-272, 2019.
- [8] Marko Horvat, Tomislav Jaguš, "Emerging opportunities for education in the time of COVID-19: Adaptive e- learning intelligent agent based on assessment of emotion and attention," *Proceedings of the Central European Conference on Information and Intelligent System*, 2020.
- [9] Jessie S. Barrot, Ian I. Llenares & Leo S. del Rosario , "Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines," *Education and Information Technologies*, 2021.
- [10] Shoroog Khenkar and Salma Kammoun Jarraya, "Engagement Detection Based on Analyzing Micro Body Gestures Using 3D CNN," *Computers, Materials & Continua*, vol. 70, pp. 2656-2677, 2022.
- [11] Grafsgaard JF, Fulton RM, Boyer KE, Wiebe EN, "Multimodal analysis of the implicit affective channel in computer mediated textual communication,"

Proceedings of the 14th ACM international conference on Multimodal interaction, 2012.

- [12] M. Saqr, B. Wasson, "COVID-19: lost opportunities and lessons for the future," *Int. J. Health Sci*, 2020.
- [13] Sonsoles López-Pernas, Mohammed Saqr, Aldo Gordillo, Enrique Barra, "A learning analytics perspective on educational escape rooms," *Open Access, Interactive Learning Environments*, Routledge, 2022.
- [14] Sharma , P., Joshi, S., Gautam, S., Filipe, V., & Reis, M . J., "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning.," *CORR*, 190 9.12913. *abs-1909-12913*, p. <http://arxiv.org/abs/1909.12913>, 2019.
- [15] Soloviev, V., "Machine learning approach for student engagement automatic recognition from facial expressions.," *Scientific Publications of the State University of Novi Pazar Series A: Applied Mathematics, Informatics and mechanics*, vol. 10, no. 2, pp. 79-86. <http://doi.org/10.5937/spsunp1802079s>, 2018.
- [16] Soukupova, T. & Cech, J., "Real-Time Eye Blink Detection using Facial Landmarks," *21st Computer Vision Winter Workshop*, 2016.
- [17] Chakradhar Pabba, Praveen Kumar, "An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition," *Expert System Willey*, 2021.
- [18] Ayat Alrawahneh, DR. Suhailan Bin Safei, "A model of video watching concentration level measurement among students using head pose and eye tracking detection," *Journal of Theoretical and Applied Information Technology*, pp. 4305-4315, 2021.
- [19] P. Sharma, S. Joshi, S. Gautam, V. Filipe, and M. J. Reis, "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning," *arXiv preprint arXiv:1909.12913*, 2019.
- [20] Bouhlal Meriem, Habib Benlahmar, Mohamed Amine Naji, Elfilali Sanaa, Kaiss Wijidane, "Determine the Level of Concentration of Students in Real Time from their Facial Expressions," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 159-166, 2022.
- [21] Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.H., et al., "Challenges in representation learning: A report on three machine learning contests. In: *ICONIP*," *Springer*, p. 117–124, 2013.

- [22] Lâm Thành Hiền, Phát triển một số kỹ thuật phát hiện lái xe ngủ gật dựa trên trạng thái mắt và hành vi gật lắc, Luận án tiến sĩ toán học - Viện Hàn Lâm và Khoa học công nghệ Việt Nam, 2017.
- [23] Viola, P.; Jones, M., "Rapid object detection using a boosted cascade of simple features," *In Proceedings of the CVPR, Kauai, HI, USA, 8–14 December 2001*, vol. 1, p. 511–518, 2001.
- [24] Davis E. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, pp. 1755-1758, 2009.
- [25] Tran Thanh Phuong, Lam Thanh Hien, Do Nang Toan, Ngo Duc Vinh, "An Eye Blink detection technique in video surveillance based on Eye Aspect Ratio," *Proceedings of International Conference on Advanced Communication Technology (ICACT)*, pp. 534-538, 2022.
- [26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception resnet and the impact of residual connections on learning," *in Thirty-first AAAI conference on artificial intelligence*, 2017.
- [27] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, "Deep residual learning for image recognition," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.
- [28] Tran Thanh Phuong, Lam Thanh Hien, Ngo Duc Vinh, Ha Manh Toan, Do Nang Toan, "An algorithm for decomposing variations of a 3D model," *International Journal of Electrical and Computer Engineering (IJECE)*, 2023.
- [29] Trần Thanh Phương, Trần Văn Lăng, Đỗ Năng Toàn, "A technique of assessing student engagement based on eye close/open behavior," *TNU Journal of Science and Technology*, vol. 226, no. 07, pp. 262-269, 2021.
- [30] Seligman, Martin E. P.; Csikszentmihalyi, Mihaly, "Positive Psychology: An introduction," *American Psychologist*, pp. 5-14, 2000.
- [31] Heather L. O'Brien, Ido Roll, Andrea Kampen, Nilou Davoud, "Rethinking (Dis)engagement in human-computer interaction," *ELSEVIER*, 2021.
- [32] Pieter Vanneste, José Oramas, Thomas Verelst, Tinne Tuytelaars, Annelies Raes, Fien Depaepe and Wim Van den Noortgate, "Computer Vision and Human Behaviour, Emotion and Cognition Detection: A Use Case on Student Engagement," *MDPI Open Access*, 2021.
- [33] D Dinesh, A Narayanan, K Bijlani, "Student analytics for productive teaching/learning," *International Conference on Information Science (ICIS), Kochi, India*, pp. 97-102, 2016.
- [34] AS Won, JN Bailenson, JH Janssen, "Automatic detection of nonverbal behavior predicts learning in dyadic interactions," *IEEE Trans. Affect Comput*, vol. 5, no. 2, pp. 112-125, 2014.

- [35] Patricia Goldberg, Ömer Sümer, Kathleen Stürmer, Wolfgang Wagner, Richard Göllner, Peter Gerjets, Enkelejda Kasneci, Ulrich Trautwein, "Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction," *Springer*, vol. 33, p. 27–49, 2021.
- [36] Xiaoyang Ma, Min Xu, Yao Dong, and Zhong Sun, "Automatic Student Engagement in Online Learning Environment Based on Neural Turing Machine," *International Journal of Information and Education Technology*, vol. 11, no. 3, 2021.
- [37] T. S. Ashwin, R. M. R. Guddeti, "Unobtrusive Behavioral Analysis of Students in Classroom Environment Using Non-Verbal Cues," *IEEE Access*, vol. 7, 2019.
- [38] Van Acker, B.B.; Bombeke, K.; Durnez, W.; Parmentier, D.D.; Mateus, J.C.; Biondi, A.; Saldien, J.; Vlerick, P., "Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure," *Int. J. Ind. Ergon*, 2020.
- [39] P. Ekman, W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement," *Consulting Psychologists Press*, 1978.
- [40] Tarnowski, P.; Kołodziej, M.; Majkowski, A.; Rak, R.J, "Emotion recognition using facial expressions," *Procedia Comput. Sci*, vol. 108, p. 1175–1184, 2017.
- [41] N. Wixon, D. Allessio, S. Schultz, W. Burleson, I. Arroyo, K. Muldner, B. Woolf, "User Modeling Adaptation and Personalization," *Conference Halifax*, 2016.
- [42] Griol, D., Molina, J. IM., & Caflejas, Z., "Incorporating android conversational agents in m-learning apps," *Expert Systems Wiley*, vol. 34, no. 4, p. e12156. <http://doi.org/10.1111/exsy.12156>, 2017.
- [43] Mogwe, A. W, "Real-time concept feedback in lectures for botho university students," *International Journal of Advanced Computer Science and Application*, pp. 246-251, 2018.
- [44] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, J. Movel, "The faces of engagement: Automatic recognition of student engagement from facial expressions," *IEEE Transactions on Affective Computing*, pp. 86-98, 2014.
- [45] Di Lascio, E., Gashi, S., & Santini, S., "Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1-21. <http://doi.org/10.1145/3264913>, 2018.
- [46] Andrea Apicella, Pasquale Arpaia, Mirco Frosolone, Giovanni Improta, Nicola Moccaldi, Andrea Pollastro, "EEG-based Measurement System for Student Engagement Detection in Learning 4.0," *Research Square*, 2021.

- [47] Ruipérez-Valiente, J. A., Munoz-Merino, P. J., & Delgado Kloos, C., "Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators.," *Expert Systems Wiley*, vol. 35, no. 6, p. e12298. <http://doi.org/10.1111/exsy.12298>, 2018.
- [48] Zhang H., Xiao X., Huang T., Liu S., Xia Y., & Li J., "An novel end-to-end network for automatic student engagement recognition.," *IEEE 9th International Conference on Electronics Information and Emergency Communication (ICE/EC)*, pp. 342 - 345. <https://doi.org/10.1109/ICEIEC.2019.8784507>, 2019.
- [49] Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P. K. D., Dasgupta, N., & Choudhury, P., "Facial emotion detection to assess learner's state of mind in an online learning system.," *Proceedings of the 2020 5th International Conference on Intelligent Information Technology*, pp. 107-115. <https://doi.org/10.1145/3385209.3385231>, 2020.
- [50] Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A., "Application of Deep Learning on Student Engagement in e-learning environments.," *Computers & Electrical Engineering*, pp. 93, 107277. <http://doi.org/10.1016/j.compeleceng.2021.107277>, 2021.
- [51] Ashwin, T. S., & Guddeti, R. M. R., "Unobtrusive behavioral analysis of students in classroom environment using non-verbal cues.," *IEEE Access*, pp. 150693-150709. <http://doi.org/10.1109/access.2019.2947519>, 2019.
- [52] Luo, Z., Jingying, C., Guangshuai, W., & Mengyi, L., "A three-dimensional model of student interest during learning using multimodal fusion with natural sensing technology.," *Interactive Learning Environments*, pp. 1-14. <http://doi.org/10.1080/10494820.2019.1710852>, 2020.
- [53] Peng, S., & Nagao, K., "Recognition of students' mental states in discussion based on multimodal data and its application to educational support.," *IEEE Access*, pp. 18235-18250. <http://doi.org/10.1109/access.2021.3054176>, 2021.
- [54] Pekrun, R., "A social-cognitive control-value theory of achievement emotions. In J. Heckhausen (Ed.)," *Advances in psychology*, pp. 243-163. Elsevier Science. [https://doi.org/10.1016/S0166-4115\(00\)80010-2](https://doi.org/10.1016/S0166-4115(00)80010-2), 2000.
- [55] Zaletelj, J., & Kosir, A., "Predicting students' attention in the classroom from Kinect facial and body features.," *EURASIP Journal on Image and Video Processing*, 2017.
- [56] Klein, R., & Celik, T., "The wits intelligent teaching system: Detecting student engagement during lectures using convolutional neural network," in *IEEE International Conference on Image Processing*, 2017.

- [57] Thomas, C., & Jayagopi, D. B., "Predicting student engagement in classrooms using facial behavioral cues.," in *Proceedings of the 1st ACM Sigchi International Workshop on Multimodal Interaction for Education*, 2017.
- [58] Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., & Movellan, J. R., "The faces of engagement: automatic recognition of student engagement from facial expressions.," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86-98, 2014.
- [59] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Transactions on Affective Computing*, pp. 15-28, 2016.
- [60] A. Gupta, A. D'Cunha, K. Awasthi, and V. Balasubramanian, "Daisee: Towards user engagement recognition in the wild," *arXiv preprint arXiv:1609.01885*, 2016.
- [61] Bosch, N., D'mello, S. K., Ocumpaugh, J., Baker, R. S., & Shute, V., "Using Video to Automatically Detect Learner Affect in Computer-Enabled Classrooms.," *ACM Transactions on Interactive Intelligent Systems*, vol. 6, no. 2, pp. 1-26, 2016.
- [62] Krithika, L. B., & Lakshmi, P. G. G., "Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric.," *Procedia Computer Science*, pp. 767-776, 2016.
- [63] O. M. Nezami, M. Dras, L. Hamey, D. Richards, S. Wan, and C. Paris, "Automatic recognition of student engagement using deep learning and facial expression," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, pp. 273-289, 2019.
- [64] Ekman P., Friesen W. V. , "Unmasking the Face: A Guide To Recognizing Emotions From Facial Clues," *Prentice-Hall, Englewood Cliffs*, 1975.
- [65] Trần Thanh Phương, Ngô Đức Vĩnh, Hà Mạnh Toàn, Nông Minh Ngọc, Đỗ Năng Toàn, "Một kỹ thuật mô phỏng cử động của mô hình đầu người 3D theo lời thoại tiếng việt," *TNU Journal of Science and Technology*, vol. 228, no. 15, pp. 20-28, 2023.
- [66] Lucey P., Cohn J.F., Kanade T., Saragih J., Ambadar Z., Matthews I., "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," *In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, San Francisco, CA, USA, p. 94–101, 2010.
- [67] Russell J. and Fernandez-Dols J., "The psychology of facial expression," *Cambridge University Press*, 1997.

- [68] Assari M.A., Rahmati M., "Driver drowsiness detection using face expression recognition," *In Proceedings of the IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, Malaysia*, p. 337–341, 2011.
- [69] Jabon M., Bailenson J., Pontikakis E., Takayama L., Nass C., "Facial expression analysis for predicting unsafe driving behavior," *IEEE Perv. Comput*, vol. 10, p. 84–95, 2011.
- [70] Chen C.H., Lee I.J., Lin L.Y., "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders," *Res. Dev. Disabil*, vol. 36, p. 396–403, 2015.
- [71] Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz W., Taylor J.G., "Emotion recognition in human-computer interaction," *IEEE Signal Process. Mag*, vol. 18, p. 32–80, 2001.
- [72] Bartlett M.S., Littlewort G., Fasel I., Movellan J.R., "Real Time Face Detection and Facial Expression recognition: Development and Applications to Human Computer Interaction," *In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop, Madison, WI, USA*, vol. 5, p. 53–53, 2003.
- [73] Bekele E., Zheng Z., Swanson A., Crittendon J., Warren Z. Sarkar, "Understanding how adolescents with autism respond to facial expressions in virtual reality environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, p. 711–720, 2013.
- [74] Kapoor A., Burleson W., Picard R.W., "Automatic prediction of frustration," *Int. J. Hum.-Comput. Stud*, vol. 65, p. 724–736, 2007.
- [75] Lankes M., Riegler S., Weiss A., Mirlacher T., Pirker M., Tscheligi M., "Facial expressions as game input with different emotional feedback conditions," *In Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, Yokohama, Japan*, p. 253–256, 2008.
- [76] Nazeerah A. Sheth, Mahesh M. Goyani, "A Comprehensive study of Geometric and Appearance based Facial Expression Recognition Methods," *International Journal of Scientific Research in Science, Engineering and Technology*, vol. 4, no. 2, pp. 163-175, 2018.
- [77] Daniel Canedo and António J. R. Neves, "Facial Expression Recognition Using Computer Vision: A Systematic Review," *Applied Sciences, MDPI*, vol. 9, no. 21, 2019.
- [78] Lee, J. S., "Digital image smoothing and the sigma filter," *Computer Vision, Graphics and Image Processing*, vol. 24, pp. 255--269, 1983.
- [79] Wiener, N., "Extrapolation, Interpolation, and Smoothing of Stationary Time Series," *New York: Wiley*, 1949.

- [80] Brown, R. G., Hwang, Y. C., "Introduction to Random Signals and Applied Kalman Filtering," 3 ed. New York: John Wiley & Sons, 1996.
- [81] Arivazhagan, S., Deivalakshmi, S., Kannan, K., Gajbhiye, B. N., Muralidhar, C., Lukose, N., Sijo N., Subramanian, M. P., "Multi-resolution system for artifact removal and edge enhancement in computerized tomography images," *Pattern Recognition Letters* 28(13), pp. 1769-1780, 2007.
- [82] Tomasi, C.; Manduchi, R., "Bilateral filtering for gray and color images," *ICCV*, 1998.
- [83] Lindenbaum, M.; Fischer, M.; Bruckstein, A. , "On Gabor's contribution to image enhancement," *Pattern Recognit*, pp. 1-8, 1994.
- [84] Wilson CA, Alexander O, Tunwattanapong B, Peers P, Ghosh A, Busch J, Hartholt A, Debevec P, "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, 2011.
- [85] Rowley, H., Baluja, V., Kanade, T., "Neural Network-Based Face Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20(1), pp. 23-38, 1998.
- [86] Martinkauppi, B., "Face colour under varying illumination - analysis and applications," *University of Oulu repository*, 2002.
- [87] Sterring, M., Andersen, H. J., and Granum, E., "Skin Color Detection Under Changing Lighting Condition.," *7-th Symposium on the Intelligent Robotics Systems*, pp. 187-195, 1999.
- [88] Pizer, S. M., Amburn, E. P., Austin, J. D. , Cromartie, R., Geselowitz, R., Greer, T., Romeny, B. T. H., Zimmerman, J. B., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing* , vol. 39(3), pp. 355-368, 1987.
- [89] Zuiderveld, K., "Contrast limited adaptive histogram equalization," *Graphics gems IV, Academic Press Professional, Inc., San Diego, CA*, 1994.
- [90] Hawkins, D.M., "The problem of overfitting.," *J. Chem. Inf. Comput. Sci*, pp. 1-12, 2004.
- [91] Lyons M., Akamatsu S., Kamachi M., Gyoba J., "Coding facial expressions with gabor wavelets," *In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan*, p. 200–205, 1998.
- [92] Yu J., Bhanu B., "Evolutionary feature synthesis for facial expression recognition," *Pattern Recognit. Lett*, vol. 27, p. 1289–1298, 2006.

- [93] Mattela, G., Gupta, S.K., "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, p. 575–580, 2018.
- [94] Ahonen, T., Hadid, A., Pietikäinen, M., "Face recognition with local binary patterns," *In European Conference on Computer Vision*, Springer: Berlin, Germany, p. 469–481, 2004.
- [95] Mustafa Ugur Ucar and Ersin Ozdemir, "Recognizing Students and Detecting Student Engagement with Real-Time Image Processing," *Electronic MDPI*, 2022.
- [96] Ahonen, T.; Hadid, A.; Pietikäinen, M., "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Trans. Pattern Anal. Mach. Intell*, p. 2037–2041, 2006.
- [97] Cover, T.; Hart, P., "Nearest neighbor pattern classification.," *IEEE Trans. Inf. Theory*, pp. 21-27, 1967.
- [98] Quinlan, J.R., "Induction of decision trees.," *Mach. Learn.*, pp. 81-106, 1986.
- [99] Breiman, L., "Random forests," *Mach. Learn.*, pp. 3-52, 2001.
- [100] Eddy, S.R., "Hidden markov models.," *Curr. Opin. Struct. Biol.* , p. 361–365, 1996.
- [101] Archana Sharma, Vibhakar Mansotra, "Deep Learning based Student Emotion Recognition from Facial Expressions in Classrooms," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 8, no. 6, pp. 4691-4699, 2019.
- [102] Friedman J., Hastie T., Tibshirani R., "Additive logistic regression: A statistical view of boosting," *The Annals of Statistics*, pp. 337-374, 2000.
- [103] Lyons M J, Akamatsu S, Kamachi M, et al, "The Japanese female facial expression (JAFFE) database," *in Proceedings of third international conference on automatic face and gesture recognition*, pp. 14-16, 1998.
- [104] Yang Xiao, Zhiguo Cao, Lubin Meng, Zhiwen Fang, Joey Tianyi Zhou and Junsong Yuan Guilei Hu, "Towards Real-Time Eyeblink Detection in the Wild: Dataset, Theory and Practices," *IEEE*, pp. 2194-2208, 2019.
- [105] Shichuan Du, Yong Tao, and Aleix M. Martinez , "Compound facial expressions of emotion," *Computer Sciences PNAS*, 2014.
- [106] [Online]. Available: <https://durian.blender.org/news/sintel-lite/>. [Accessed 11 10 2023].
- [107] Blanz, Volker, Sami Romdhani, and Thomas Vetter, "Face identification across different poses and illuminations with a 3D morphable model," *In Proceedings*

of fifth IEEE international conference on automatic face gesture recognition, pp. 202-207, 2022.

- [108] Weiyue Wang, Duygu Ceylan, Radomir Mech, Ulrich Neumann, "3DN: 3D deformation network," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1038-1046, 2019.
- [109] T. Groueix, M. Fisher, V. G. Kim, B. Russell, and M. Aubry, "3D-coded: 3D correspondences by deep deformation," *in Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 230-246, 2018.
- [110] Y. Yang, C. Feng, Y. Shen, and D. Tian, "Foldingnet: Point cloud auto-encoder via deep grid deformation," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 206-215, 2018.
- [111] Huynh Cao Tuan, Do Nang Toan, and Lam Thanh Hien, "Automatic Selection of Key Points for 3D-Face Deformation," *Journal of Advances in Information Technology*, vol. 13, no. 4, pp. 332-337, 2022.
- [112] "Project of Ministry "Development of 3D face state and gesture representation techniques based on interpolation techniques for the problem of building virtual broadcasters"; Code: B2015-TN06-02".
- [113] A. Kurenkov, J. Ji, A. Garg, V. Mehta, J. Gwak, C. Choy, and S. Savarese, "Deformnet: Free-form deformation network for 3D shape reconstruction from a single image," *In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE*, 2018.
- [114] Wiegand, Douglas M.; McClafferty, Julie A.; McDonald, Shelby E.; Hanowski, Richard J., "Development and evaluation of a naturalistic observer rating of drowsiness protocol : final report," 2009.
- [115] Y. X. Z. C. L. M. Z. F. J. T. Z. a. J. Y. Guilei Hu, "Towards Real-Time Eyeblink Detection in the Wild: Dataset, Theory and Practices," *IEEE*, pp. 2194 - 2208, 2019.
- [116] Fogelton, A., and Benesova, W., "Eye Blink Detection Based on Motion Vectors Analysis," *Comput. Vision Image Understand*, 2016.
- [117] S. Raina, L. Bernard, B. Taylor, S. Kaza, "Using Eye-Tracking to Investigate Content Skipping: A Study on Learning Modules in Cybersecurity," *y (IEEE Conference on Intelligence and Security Informatics, Tucson)*, 2016.

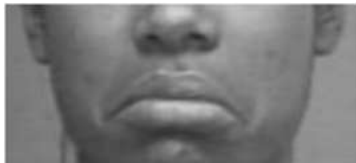
PHỤ LỤC 1

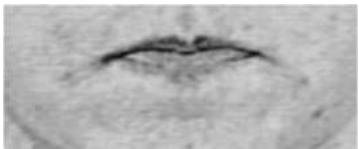
HỆ MÃ HÓA CÁC CỬ ĐỘNG TRÊN KHUÔN MẶT

Hệ thống mã hóa các cử động khuôn mặt (FACS) được đề xuất bởi Paul Ekman và Wallace Friesen [39], mô tả tất cả các cử động cơ bản có thể quan sát được trên khuôn mặt. Hệ thống bao gồm 64 đơn vị cử động (Action Unit, viết tắt là AU), mỗi AU được mô tả là kết hợp của một hoặc một số các cơ trên khuôn mặt. Bảng 0.1 mô tả danh sách các AU của FACS.

Bảng 0.1 Hệ mã hóa các cử động trên khuôn mặt

AU	Mô tả	Chuyển động	Minh họa
1	Nhướn mày trong	Frontalis, pars medialis	
2	Nhướn mày ngoài	Frontalis, pars lateralis	
4	Hạ lông mày	Corrugator supercilii, Depressor supercilii	
5	Nhướn mí trên	Levator palpebrae superioris	
6	Nâng má	Orbicularis oculi, pars orbitalis	
7	Căng mí mắt	Orbicularis oculi, pars palpebralis	
9	Nhăn mũi	Levator labii superioris alaeque nasi	

AU	Mô tả	Chuyển động	Minh họa
10	Nâng môi trên	Levator labii superioris	
11	Làm sâu mũi	Levator anguli oris (a.k.a. Caninus)	
12	Kéo khóe môi	Zygomaticus major	
13	Phồng má	Zygomaticus minor	
14	Má lúm đồng tiền	Má lúm đồng tiền	
15	Nén khóe môi	Depressor anguli oris (a.k.a. Triangularis)	
16	Bặm môi dưới	Depressor labii inferioris	
17	Nâng cằm	Mentalis	

AU	Mô tả	Chuyển động	Minh họa
18	Nhàu môi	Incisivii labii superioris and Incisivii labii inferioris	
20	Kéo căng môi	Risorius w/ platysma	
22	Môi hình phễu	Orbicularis oris	
23	Bặm chặt môi	Orbicularis oris	
24	Ép môi	Orbicularis oris	
25	Tách môi trên và dưới	Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris	
26	Hạ hàm	Masseter, relaxed Temporalis and internal Pterygoid	
27	Căng miệng	Pterygoids, Digastric	

AU	Mô tả	Chuyển động	Minh họa
28	Bặm môi	Orbicularis oris	
41	Rủ mí mắt	Relaxation of Levator palpebrae superioris	
42	Tì hí mắt	Orbicularis oculi	
43	Nhắm mắt	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis	
44	Liếc mắt	Orbicularis oculi, pars palpebralis	
45	Chớp mắt	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis	
46	Nháy mắt	Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis	
51	Quay đầu sang trái		
52	Quay đầu sang phải		
53	Ngước đầu lên		

AU	Mô tả	Chuyển động	Minh họa
54	Đầu cúi xuống		
55	Hướng đầu sang trái		
56	Hướng đầu sang phải		
57	Hướng đầu về phía trước		
58	Hướng đầu về phía sau		
61	Liếc mắt sang trái		
62	Liếc mắt sang phải		
63	Liếc mắt lên trên		
64	Mắt nhìn xuống dưới		

PHỤ LỤC 2

MÃ CHƯƠNG TRÌNH

1. Thuật toán huấn luyện

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

Đầu ra: $T = \{N_0, N_1 \dots\}$ //cây kết quả được mô tả là tập các nút

Xử lý

1. $T := \emptyset$
 2. $Idx_0 = \{0, 1, 2, \dots, S - 1\}$ // tập các chỉ số tương ứng với các mẫu trong tập huấn luyện. Sử dụng 1 ngăn xếp để quản lý việc học lần lượt từng nút. Mỗi phần tử ngăn xếp là một bộ gồm nút và tập chỉ số các mẫu dùng để học nút đó
 3. $Stack := \emptyset$ // Đưa nút gốc vào ngăn xếp
 4. **push**(Stack, (N_0, Idx_0))
 5. **while** (Stack $\neq \emptyset$) // lấy phần tử ở đầu ngăn xếp
 6. $\{N_i, Idx_i\} := \mathbf{pop}$ (Stack) // Nếu nút chưa đạt đủ độ cao
 7. **if** ($N_i.level \geq MAX_DEPTH$) **continue**
 8. **else**
 9. $min_err := MAX_VALUE$
 10. $best_bincls := \mathbf{null}$
 11. **for all** bincls **of** BCS
 12. $e := WMSE(bincls, U, Idx_i)$
 13. **if** ($e < minerr$)
 14. $best_bincls := bincls$
 15. $min_err := e$
 16. **endif**
 17. **endfor**
 18. **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
 19. $\{Idx_{i*2+1}, Idx_{i*2+2}\} := \mathbf{SplitDataSet}(U, Idx_i, best_bincl)$ Tiếp tục đưa vào ngăn xếp hai nút con cùng với tập chỉ số các mẫu tương ứng
 20. **push**(Stack, (N_{i*2+1}, Idx_{i*2+1}))
 21. **push**(Stack, (N_{i*2+2}, Idx_{i*2+2}))
 22. **endif**
 23. **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_VALUE là giá trị khởi tạo cho sai số, dùng để tìm bộ phân lớp con có giá trị sai số nhỏ nhất.

2. Mã chương trình tính thời gian tham gia học tập của người học ở từng mức độ

Mã chương trình 1

Đầu vào: *DATA* – Dữ liệu video về quá trình học tập của người học

Đầu ra: *list_learner* = {*learner_i*{(*status_j*), *j* = [0,2]}, *i* = [0, *n* - 1]} // *n*: number of objects

Xử lý

1. Khởi tạo *status* = { hơi buồn ngủ, ngủ gật, bắt đầu giấc ngủ }
 2. Trích xuất các khung hình trong *DATA*: *frames* = *extract_frame(DATA)*
 3. Lấy số lượng khung hình trên giây từ *DATA*: *fps* = *get_fps(DATA)*
 4. **Lặp** qua từng khung hình (*fr*, *index_fr*) trong các khung hình của video (*frames*)
 5. Phát hiện các khuôn mặt trong khung hình *fr* và gán vào *faces*
 6. **Lặp** qua từng khuôn mặt (*f*, *fid*) trong *faces*
 7. Khởi tạo đối tượng người học (*obj*) có định danh là *fid*
 8. Phát hiện trạng thái của mắt ở khung hình hiện tại của khuôn mặt *f*
 9. **Nếu** trạng thái của mắt ở khung hình hiện tại của khuôn mặt *f* là ‘đóng’
 10. **Nếu** trạng thái của mắt ở khung hình trước là ‘mở’
 11. Cập nhật lại trạng thái của mắt ở khung hình trước là ‘đóng’
 12. Ghi nhận vị trí khung hình mắt bắt đầu đóng là *index_fr*
 13. **Tiếp tục vòng lặp**
 14. **Kết thúc kiểm tra điều kiện**
 15. **Ngược lại** trạng thái của mắt ở khung hình hiện tại là ‘mở’
 16. **Nếu** trạng thái mắt ở khung hình trước là ‘đóng’
 17. Tính toán thời gian đóng mắt từ khung hình trước đến khung hình hiện tại: *close_duration* = ((*index_fr* - *khung hình mắt bắt đầu đóng*) * 1000) / *fps*
 18. Cập nhật lại trạng thái của mắt ở khung hình trước là ‘mở’
 19. Ghi nhận vị trí khung hình mắt bắt đầu đóng về 0
 20. // xác định vị trí trong *status* dựa vào thời gian đóng mắt của người học
 21.
$$index = \begin{cases} 0, & \text{nếu } 1 \leq close_duration < 2 \\ 1, & \text{nếu } 2 \leq close_duration < 4 \\ 2, & \text{nếu } close_duration \geq 4 \end{cases}$$
 22. // Cộng dồn thời gian vào trạng thái tương ứng của người học
 23. *obj.status[index].append(close_duration)*
 24. **Nếu** người học chưa tồn tại trong danh sách *list_learner*
 25. *list_learner.insert(obj)*
 26. **Ngược lại** cập nhật lại thông tin của người học đó *list_learner.update(obj)*
 27. Đặt lại *close_duration* = 0
 28. **Kết thúc kiểm tra điều kiện**
 29. **Kết thúc kiểm tra điều kiện**
 30. **Kết thúc lặp**
 31. **Kết thúc lặp**
-

32. **Trả về** *list_learner* (thời gian tham gia học ứng với mỗi trạng thái của người học)

3. Mã chương trình đánh giá sự tập trung

Mã chương trình 2

Đầu vào: *DATA* – Dữ liệu video về quá trình học tập của người học

Đầu ra: *list_learner_status* = {*user_state*_{*i*}{(*id*, *state*), *i* = [0, *n* – 1]}} // Danh sách trạng thái của người học được biểu diễn bằng tập hợp các bộ (*id*, *state*), với *i* là chỉ số đối tượng trong danh sách và *state* là trạng thái của người học, *n* là số đối tượng trong video

Xử lý

1. Khởi tạo *state* = {Tập trung cao, tập trung thấp, không tập trung}
 2. Tính thời gian tham gia học tập của mỗi đối tượng trong video theo Algorithm 4.2
list_learner = **Mã chương trình 1** (*DATA*)
 3. Đặt *n* là tổng thời lượng của *DATA*
 4. **Lặp** qua từng đối tượng (*obj*) trong danh sách người học *list_learner*
 5. Tính *r2*: Tỷ lệ thời gian tập trung thấp *obj.status*[0] / *n*
 6. Tính *r34*: Tỷ lệ thời gian không tập trung (*obj.status*[1] + *obj.status*[2]) / *n*
 7. Tính *r1*: Tỷ lệ tập trung cao bằng 1 – (*r2* + *r34*)
 8. So sánh và tìm ra giá trị lớn nhất trong *r1*, *r2*, *r34*
 9. Dựa vào giá trị lớn nhất, thêm một bộ mới *obj*{(*id*, *state*)} vào danh sách *list_learner_status* với *state* tương ứng
 10. **Kết thúc lặp**
 11. **Trả về** danh sách người học với trạng thái tương ứng *list_learner_status*
-