

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

CÔNG NGHỆ THÔNG TIN



NHẬN DẠNG - CSC14006

BÁO CÁO GIỮA KỲ

Đề tài: FACE DETECTION BASED ON SKIN COLOR

THẦY (CÔ) HƯỚNG DẪN

PGS.TS. LÊ HOÀNG THÁI

Thầy. ĐƯƠNG THÁI BẢO

Thầy. TRƯƠNG TÂN KHOA

SINH VIÊN THỰC HIỆN

MSSV	Họ và Tên	Email
22127147	Đỗ Minh Huy	dmhuy22@clc.fitus.edu.vn
22127249	Trần Thanh Long	ttlóng22@clc.fitus.edu.vn
22127322	Lê Phước Phát	lpphat22@clc.fitus.edu.vn
22127330	Nguyễn Đức Phúc	ndphuc22@clc.fitus.edu.vn

THÀNH PHỐ HỒ CHÍ MINH, THÁNG 04, 2025

LỜI NÓI ĐẦU

Thị giác máy tính, điển hình là nhận dạng mẫu, là một lĩnh vực quan trọng mang tính khoa học và công nghệ. Ứng dụng của nó, điển hình là phát hiện và nhận dạng gương mặt dựa trên nhiều yếu tố, mang tính cấp thiết trong các lĩnh vực khác nhau như: y học, vật lý, toán học, tìm kiếm, bảo mật, và rất nhiều lĩnh vực khoa học khác nhau, ...

Phát hiện khuôn mặt là một phần trong lĩnh vực xử lý ảnh, là một vấn đề cơ bản trong ứng dụng thị giác máy tính. Đây được xem là một trong những giai đoạn của hệ thống nhận dạng mặt người ban đầu của hệ thống nhận dạng gương mặt.

Tuy nhiên, việc phát hiện gương mặt bị ảnh hưởng bởi nhiều yếu tố khác nhau theo khía cách chủ quan lẫn khách quan. Về mặt chủ quan, hiệu suất của bài toán thường bị ảnh hưởng bởi chất lượng hình ảnh, cơ chế cũng như kiến trúc của ảnh. Về mặt khách quan, hiệu suất cũng bị ảnh hưởng với màu sắc, ánh sáng ngoại lai, nhiễu, ...

Trong đồ án này, nhóm sẽ trình bày phương pháp loại bỏ đi một số yếu tố ảnh hưởng đến độ chính xác trong phát hiện mặt người. Phương pháp này xem như là một bước tiền xử lý hình ảnh trước khi đưa qua các mô hình thuật toán nhận diện gương mặt. Về phương pháp, kỹ thuật này sẽ dựa vào vùng da mặt đã được phân đoạn để phát hiện gương mặt, loại bỏ các vùng không gian dư thừa trong ảnh.

Mặc dù các thành viên trong nhóm đã rất cố gắng để hoàn thành thật tốt bài tập nghiên cứu cá nhân nhưng chắc chắn không thể tránh khỏi những hạn chế và thiếu sót không mong muốn. Nhóm chúng em mong nhận được sự thông cảm và ý kiến đóng góp của thầy cô trong lớp học phần **NHẬN DẠNG MẪU - CSC14006** để em có thể rút kinh nghiệm và hoàn thiện những bài tập sau tốt hơn.

Nhóm chúng em xin được gửi lời cảm ơn chân thành tới thầy **PGS.TS Lê Hoàng Thái**, thầy **Dương Thái Bảo** và thầy **Trương Tân Khoa** vì đã luôn tận tình chỉ bảo, hướng dẫn, và giải đáp mọi thắc mắc của em trong suốt quá trình thực hiện bài tập cá nhân này.

Tóm lại, thông qua bài tập này, chúng ta có thể nhìn nhận được tập quan trọng của phát hiện và nhận diện gương mặt dựa trên yếu tố vùng da nói riêng và ứng dụng nhận dạng nói chung trong đời sống khoa học - xã hội hiện đại ngày nay.

TP.HCM, mùa xuân 2025.

LỜI CAM ĐOAN

Nhóm thực hiện đồ án **SKIN COLOR IN FACE ANALYSIS** gồm các thành viên trên đều là sinh viên khoa Công nghệ Thông tin Chất lượng cao, thuộc trường Đại học Khoa học Tự nhiên, ĐHQG-HCM. Nhóm cam đoan rằng bài tập nghiên cứu này là do các thành viên trong nhóm tìm hiểu, nghiên cứu và thực hiện dưới sự giám sát và hướng dẫn của các thầy **PGS.TS. Lê Hoàng Thái**, thầy **Dương Thái Bảo**, và thầy **Trương Tấn Khoa**. Các dữ liệu được nêu trong đồ án là hoàn toàn trung thực, phản ánh đúng kết quả mô phỏng thực tế. Tất cả các tài liệu được sử dụng trong nghiên cứu này được các thành viên trong nhóm thu thập bằng cách tự thân và từ các nguồn khác nhau, và những tài liệu này được liệt kê đầy đủ trong phần tài liệu tham khảo. Tất cả đều được trích dẫn đúng đắn. Trong trường hợp có vi phạm bản quyền, các thành viên trong nhóm sẽ chịu trách nhiệm cho hành động đó. Do đó, trường **Đại học Khoa học Tự nhiên, ĐHQG-HCM** không chịu trách nhiệm về bất kỳ vi phạm bản quyền nào được thực hiện trong bài tập nghiên cứu này.

TP.HCM, ngày 06 tháng 04 năm 2025.

Người cam đoan

Nhóm trưởng

LÊ PHƯỚC PHÁT

MỤC LỤC

DANH MỤC HÌNH VẼ	ii
CHƯƠNG 01. GIỚI THIỆU	1
CHƯƠNG 02. TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ LÝ THUYẾT	4
2.1 Tín hiệu màu và phân tích hình ảnh khuôn mặt	4
2.1.1 Tín hiệu màu	4
2.1.2 Quá trình vận dụng màu sắc vào phân tích hình ảnh gương mặt .	5
2.2 Sự hiển thị màu sắc của các máy ảnh màu kỹ thuật số	5
2.2.1 Sự hình thành ảnh màu và nguồn chiếu sáng	5
2.2.2 Hiệu ứng của cân bằng trắng	7
2.3 Phân tách nguồn dữ liệu về da	9
2.4 Mô hình hóa màu da	10
2.4.1 Biểu hiện của các sắc tố da ở các không gian màu khác nhau dưới độ chiếu sáng khác nhau	10
2.4.2 Không gian màu cho da	11
2.4.3 Mô hình hóa màu da và độ chiếu sáng	12
2.4.4 Những mô hình toán học về màu da	14
2.5 Tín hiệu màu cho bài toán phát hiện khuôn mặt	18
2.6 Tín hiệu màu cho bài toán nhận diện khuôn mặt	21
2.7 Mô hình phân đoạn hình ảnh dựa theo ngữ nghĩa	23
2.8 Mô hình phát hiện gương mặt kết hợp học sâu	24
2.8.1 Mô hình MTCNN	24
2.8.2 Mô hình RetinaFace	28
CHƯƠNG 03. PHƯƠNG PHÁP NGHIÊN CỨU	30
3.1 Phương án nghiên cứu ban đầu	30
3.2 Phương án nghiên cứu cải tiến	32

3.2.1	Thu thập dữ liệu	32
3.2.2	Tiền xử lý dữ liệu	35
3.2.3	Fine-tuned mô hình phân đoạn UNet	35
3.2.4	Xác định và hiển thị vùng ROIs da	37
3.2.5	Áp dụng các mô hình pretrained phát hiện gương mặt	38
3.3	Kết luận	38

CHƯƠNG 04. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ **39**

KẾT LUẬN **40**

Kết luận chung	40
Hướng phát triển	40
Kiến nghị và đề xuất	41

TÀI LIỆU THAM KHẢO **42**

DANH MỤC HÌNH VẼ

Hình 2.1 Các giai đoạn khác nhau trong quá trình phân tích hình ảnh khuôn mặt	5
Hình 2.2 Màu sắc của khuôn mặt thay đổi ở các phần khác nhau của trường ánh sáng	7
Hình 2.3 Vùng quang phổ gần tia hồng ngoại cho da mặt	9
Hình 2.4 Camera và các đặc tính của nó xác định vị trí da	13
Hình 2.5 Kết quả của mô hình của Hsu et al.	14
Hình 2.6 Ràng buộc không gian được đề xuất cho mô hình hóa màu da thích ứng	16
Hình 2.7 Việc theo dõi khuôn mặt của Raja et al thất bại và thích ứng với mục tiêu không phải khuôn mặt	17
Hình 2.8 Ràng buộc được đề xuất bởi Raja chọn một tập hợp các điểm ảnh không đại diện cho điểm da	18
Hình 2.9 Ví dụ phát hiện khuôn mặt sử dụng định vị khuôn mặt dựa trên màu sắc	20
Hình 2.10 Ví dụ phát hiện gương mặt sử dụng bộ phát hiện gương mặt dựa trên màu	21
Hình 2.11 Kiến trúc mô hình UNet	23
Hình 2.12 Kiến trúc P-Net	25
Hình 2.13 Kiến trúc R-Net	26
Hình 2.14 Kiến trúc O-Net	27
Hình 3.1 Ví dụ thử nghiệm với phương pháp chuyển không gian màu	31
Hình 3.2 Pipeline xử lý nghiên cứu Face Detection based on Skin Segmentation	32
Hình 3.3 Ảnh minh họa bộ dữ liệu CelebAMask-HQ	33
Hình 3.4 Ảnh minh họa bộ dữ liệu Pratheepon	34
Hình 3.5 Kết quả dự đoán của mô hình	37
Hình 3.6 Ảnh sau khi được xử lý hiển thị vùng ROIs và contours	37
Hình 3.7 Kết quả cuối cùng của pipeline	38

CHƯƠNG 01. GIỚI THIỆU

Trong thị giác máy tính, màu sắc luôn là một đặc trưng thường xuyên được sử dụng bởi tính đơn giản, bởi lẽ các phép toán với màu sắc có thể được cài đặt một cách nhanh chóng và hiệu quả, đồng thời trong môi trường ổn định với độ sáng đồng đều, màu sắc thường không bị tác động bởi sự thay đổi của hình học. Do đó, trong một số trường hợp, chỉ cần màu sắc là đủ để nhận dạng mọi sự vật.

Tuy nhiên, sự khó khăn chính hiện nay trong sử dụng màu sắc cho các ứng dụng thị giác máy tính là các camera thường không thể phân biệt được sự thay đổi của màu sắc do quang phổ của các nguồn ánh sáng khác nhau chiếu vào. Do đó, màu sắc thường nhạy cảm với sự thay đổi của nguồn sáng, điều này xảy ra rất thường xuyên trong môi trường không được kiểm soát. Các sự thay đổi này có thể được gây ra bởi sự thay đổi độ sáng (như bóng của vật thể), hoặc sự thay đổi cường độ ánh sáng (như ánh sáng mặt trời hoặc các nguồn sáng huyền quang từ bóng đèn, ...), hoặc cả hai sự thay đổi trên. Ngoài ra, các camera khác nhau và các mức điều chỉnh khác nhau có thể tạo ra nhiều bức ảnh khác nhau trong mắt con người.

Vào thời điểm hiện tại, vấn đề trên đã có nhiều hướng giải quyết được đề ra nhằm giảm thiểu sự nhạy cảm với sự thay đổi của nguồn sáng một cách không mong muốn. Trong đó bao gồm hai hướng chính sau đây:

- Thông tin màu sắc sẽ được chia làm hai thành phần chính bao gồm cường độ màu sắc và sắc độ. Khi đó, ta sử dụng sắc độ để giảm bớt tác động của sự thay đổi độ sáng, và áp dụng các thuật toán cố định màu sắc nhằm loại bỏ ảnh hưởng của sự thay đổi cường độ ánh sáng. Tuy nhiên, hiệu quả của phương pháp này vẫn còn hạn chế [1].
- Ngoài ra, ta có thể để các mô hình tự thích nghi với sự thay đổi của nguồn sáng. Phương pháp này có thể mang lại những kết quả rất đáng mong đợi được trình bày trong báo cáo này.

Trong thị giác máy tính, việc giảm thiểu sự phụ thuộc vào cường độ ánh sáng là điều thường được ưu tiên. Mục tiêu của chúng ta là phải loại bỏ hoàn toàn ảnh hưởng của màu sắc của nguồn sáng bằng cách xác định một biểu diễn màu chỉ phụ thuộc vào hệ số phản xạ bề mặt. Tuy nhiên, cho đến nay, điều này vẫn chưa đạt được trong lĩnh vực thị giác máy. Hệ thống thị giác của con người vượt trội hơn trong khía cạnh này, vì nhận thức màu sắc của con người phụ thuộc đáng kể vào hệ số phản xạ bề mặt, mặc dù ánh sáng đến mắt là kết quả của sự kết hợp giữa hệ số phản xạ bề mặt, màu sắc của nguồn sáng và cường độ ánh sáng.

Đối với những thành tựu đạt được trong lĩnh vực nhận dạng khuôn mặt, màu sắc thường được sử dụng như bước tiền xử lý đầu tiên để thực hiện các bước nghiên cứu sâu hơn và đòi hỏi tính toán nhiều hơn. Ví dụ như trong việc phát hiện khuôn mặt, thay vì phải xét hết từng vị trí và độ lớn khác nhau trong ảnh để phát hiện, thì với màu sắc, ta có thể tiền xử lý bức ảnh đó, và chỉ lấy những phần có màu sắc giống với màu da.

Trong chương 09 của cuốn sách **Handbook of Face Recognition** [2] của hai tác giả Stan Z. Li và Anil K. Jain hay trong bài báo về **SKIN COLOR IN FACE ANALYSIS** [3] của ba tác giả J. Birgitta Martinkauppi, Abdenour Hadid, và Matti Pietikäinen đã trình bày chi tiết về vai trò của màu sắc trong việc phân tích hình ảnh khuôn mặt như trong việc phát hiện và nhận diện khuôn mặt. Các nội dung ấy sẽ được nhóm chúng em trình bày theo cách hiểu bản thân một cách chi tiết và đầy đủ nhất trong **chương 02. Tổng quan tài liệu và cơ sở lý thuyết**. Nội dung ấy sẽ trình bày các vấn đề như sau:

- Giới thiệu về việc sử dụng màu sắc trong lĩnh vực phân tích ảnh khuôn mặt.
- Giới thiệu về thành phần của màu sắc và ảnh hưởng của từng nguồn sáng lên màu sắc.
- Phân tách các nguồn dữ liệu da.
- Giới thiệu việc mô hình hóa màu da.
- Đánh giá việc sử dụng màu sắc trong phát hiện khuôn mặt.
- Lợi ích của màu sắc trong nhận diện khuôn mặt.
- Mô hình phân đoạn hình ảnh dựa theo ngữ nghĩa
- Mô hình phát hiện khuôn mặt hiện đại kết hợp học sau

Sau khi chúng ta tìm hiểu các nội dung trên, nhóm chúng em sẽ nghiên cứu các phương án tối ưu nhất dựa trên các nghiên cứu hiện đại nhất lúc bấy giờ và đưa ra đề xuất cho đồ án này để có thể hoạt động một cách hiệu quả, được trình bày trong **chương 03. Phương pháp nghiên cứu**.

Sau đó, chúng em sẽ nói về hướng cài đặt mã nguồn cũng như trình bày việc báo cáo đồ án cuối kỳ như thế nào trong phần **chương 04. Thực nghiệm và Kết quả**. Tuy nhiên, do chúng em chỉ mới đề xuất hướng đi nên chúng em chưa có kết quả đầu ra để phân tích cụ thể, mong các thầy có thể bỏ qua thiếu sót này trong bài báo cáo này.

Cuối cùng, nhóm chúng em sẽ đưa ra kết luận cuối cùng về phương pháp nghiên cứu cũng như hướng đi sắp tới của nhóm em trong đồ án **Face Detection based on Skin Color** trong phần **Kết luận và Đề nghị**.

Mong các thầy có thể xem xét và đưa ra những lời khuyên bổ ích cho chủ đề và nội dung đồ án này để nhóm chúng em có thể kịp thời tinh chỉnh cũng như tìm ra hướng nghiên cứu và thực nghiệm hợp lý.

CHƯƠNG 02

TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ LÝ THUYẾT

Trong chương này, nhóm chúng em sẽ trình bày tất cả nội dung mà các tác giả trong bài báo **SKIN COLOR IN FACE ANALYSIS** [3] đã giới thiệu để có thể có cái nhìn tổng quan về vấn đề màu sắc ảnh hưởng đến việc nhận dạng hình ảnh vật thể, điển hình là ứng dụng của nó trong việc phát hiện hình ảnh khuôn mặt như thế nào nhằm trả lời câu hỏi "What is given for the book chapter ?". Đồng thời, nhóm cũng sẽ trình bày các kiến trúc mô hình hiện đại ngày nay được sử dụng cho bài toán phát hiện khuôn mặt dựa vào phân đoạn vùng da (Face Detection based on Skin Segmentation).

2.1 Tín hiệu màu và phân tích hình ảnh khuôn mặt

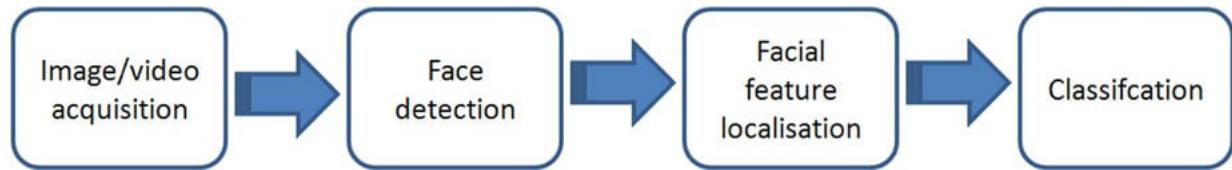
2.1.1 Tín hiệu màu

Các đặc tính của mặt người luôn là một vấn đề khó, phức tạp cho việc phân tích hình ảnh khuôn mặt. Điều trên được chứng minh qua việc khuôn mặt là một thực thể động và không cứng rắn, chính những đặc tính trên là một điều khó có thể giải quyết một cách tối ưu được. Đồng thời, hình dạng của khuôn mặt cũng có thể bị thay đổi theo độ tuổi, theo cảm xúc, theo tư thế, theo nguồn sáng ngoại lai, hay theo cách trang điểm của từng cá nhân. Do đó, việc phân tích này cần rất nhiều chi phí tính toán cho sự phức tạp của các yếu tố trên tác động đến việc phân tích khuôn mặt. Chính vì vậy, chúng ta cần một vài tín hiệu gì đó, chẳng hạn như là tín hiệu về màu sắc hay tín hiệu chuyển động của khuôn mặt, để có thể nhận biết khuôn mặt một cách hiệu quả cũng như nhằm giảm thiểu chi phí tính toán hết mức có thể. Chính nhờ những tín hiệu bổ sung đã là một thước đo thể hiện độ tin cậy vào những kết quả phân tích khuôn mặt hiện nay bởi lẽ càng có nhiều tín hiệu bổ sung hỗ trợ cho việc phân tích khuôn mặt thì càng có nhiều sự tin cậy về những kết quả phân tích đó. Tuy nhiên, khi chúng ta đã có thể xử lý tín hiệu màu một cách toàn diện, chính tín hiệu này có thể giúp giảm thiểu vùng tìm kiếm hình ảnh khuôn mặt bằng cách tiền xử lý hình ảnh đầu vào và chỉ chọn những vùng giống với màu da nhất. Chính vì vậy, điều này không ngạc nhiên rằng màu sắc của da được sử dụng một cách phổ biến trong ứng dụng phát hiện khuôn mặt trong hình ảnh. Cũng như vậy, trong ứng dụng nhận dạng gương mặt, tín hiệu màu vẫn có thể như là tín hiệu cấp thấp trong việc xác định ước tính ranh giới, hình dạng, hay kích thước của những đặc tính gương mặt. Như vậy, việc sử dụng màu sắc có rất nhiều ưu điểm trong việc phát hiện gương mặt về chi phí tính toán, về sự biến đổi các đặc tính sinh học của gương mặt do việc xoay ảnh hay việc rút gọn tỉ lệ ảnh.

Tuy nhiên, việc sử dụng tín hiệu màu trong việc phát hiện hay phân tích hình ảnh gương mặt vẫn còn nhiều bất cập, điển hình là sự hạn chế của nó trong việc nhạy cảm

với sự thay đổi của nhiều nguồn sáng khác nhau bởi lẽ khi màu sắc của da được đặt dưới các nguồn chiếu sáng khác nhau thì sẽ bị thay đổi tinh chất màu da dẫn đến việc sử dụng màu sắc để khoanh vùng da sẽ gặp nhiều khó khăn và có thể đạt được kết quả không như mong đợi.

2.1.2 Quá trình vận dụng màu sắc vào phân tích hình ảnh gương mặt



Hình 2.1 Các giai đoạn khác nhau trong quá trình phân tích hình ảnh khuôn mặt

Sơ đồ 2.1 trình bày các giai đoạn phân tích hình ảnh gương mặt như thế nào. Chúng ta có thể thấy tín hiệu màu tham gia vào rất nhiều giai đoạn trong phân tích hình ảnh gương mặt.

Trong giai đoạn đầu tiên của quá trình trên, những hình ảnh màu (hoặc chuỗi video) được thu thập và tiền xử lý. Quá trình tiền xử lý này có thể bao gồm quá trình hiệu chỉnh gamma, chuyển không gian màu, ... Thông thường, tốt hơn hết là nên loại bỏ càng nhiều sự phụ thuộc vào cường độ ánh sáng càng tốt.

Tiếp theo đó, chúng ta sử dụng màu da trong việc phát hiện gương mặt để chọn ra những vùng giống màu da nhất. Sau đó, chúng ta áp dụng các quá trình tinh chỉnh đơn giản để phân biệt vùng da mặt với các vùng da khác như tay, gối, ... Điều này góp phần cho máy phát hiện khuôn mặt với tốc độ nhanh hơn nhiều khi xét đến tín hiệu màu sắc. Tuy nhiên, có một vài đặc điểm gương mặt, như mắt, thường có vùng da tối hơn các vùng da xung quanh, điều này sẽ tạo nên các lỗ hỏng xuất hiện trên vùng da mặt khi gán nhãn các điểm pixel vùng da. Những quan sát này thường có thể được dễ dàng nhìn thấy khi phát hiện đặc điểm gương mặt trong ảnh màu.

2.2 Sự hiển thị màu sắc của các máy ảnh màu kỹ thuật số

Sau khi chúng ta đã tìm hiểu tín hiệu màu có vai trò như thế nào trong việc phát hiện hay nhận dạng gương mặt được nêu cụ thể trong phần trên, chúng ta sẽ đào sâu về các kỹ thuật xác định tín hiệu màu cũng như các tín hiệu màu được ứng dụng như thế nào trong các máy ảnh màu.

2.2.1 Sự hình thành ảnh màu và nguồn chiếu sáng

Máy ảnh màu thường tái tạo các khung cảnh với ba kênh màu chính lần lượt là kênh màu đỏ (R), kênh màu lục (G), và kênh màu lam (B). Các thành phần màu này thường

được đặt tên theo phạm vi quang phổ mà camera nhận được phản hồi lại. Các bộ lọc quang phổ thường hoạt động trong dãy bước sóng quang phổ khả kiến từ 400nm đến 700nm. Tất nhiên, các lựa chọn bộ lọc khác nhau sẽ ảnh hưởng đến bộ descriptors có thể thu được và rất có thể tạo ra các giá trị khác nhau cho cùng một đầu vào.

Bản thân các bộ descriptors này thường được thu bởi các bộ lọc tín hiệu màu $C(\lambda)$ với các bộ lọc quang phổ thích hợp và tích hợp vào trong cách tín hiệu được chọn lọc. Tín hiệu màu là sự phân bố phổ của bức xạ điện từ, là ánh sáng từ nguồn sáng, ánh sáng phản xạ từ bề mặt hoặc sự kết hợp của các yếu tố này. Tín hiệu màu này có sự tương đồng trong sự phản hồi bởi thị giác của con người.

Dưới đây là một mô hình đơn giản thể hiện đầu ra của camera với cân bằng trắng:

$$D = \frac{\int \eta_D(\lambda) I_p(\lambda) S(\lambda) d\lambda}{\int \eta_D(\lambda) I_c(\lambda) d\lambda} \quad (2.1)$$

Trong đó, D là sự phản hồi của các kênh màu R, G, và B; λ là bước sóng, p là nguồn sáng chủ đạo, và c là sự hiệu chuẩn nguồn sáng, η là độ phản ứng phổ của bộ lọc phổ cụ thể, I là sự phân bố công suất quang phổ của nguồn sáng (SPD), và S là độ phản xạ quang phổ của bề mặt.

Tử số của phương trình 2.1 mô tả cấu trúc hình ảnh như một sự tổng hợp của độ nhạy máy ảnh, độ chiếu sáng SPD và độ phản xạ trên phạm vi bước sóng. Chính vì vậy, với mỗi điểm ảnh (pixel) của ảnh được hình thành, giá trị đầu ra phụ thuộc vào độ chiếu sáng, độ phản xạ và độ nhạy cảm của máy ảnh. Đây là sự biểu diễn đơn giản của sự hình thành hình ảnh nhưng có thể được sử dụng như là một ước tính lý thuyết đơn giản của sự phản ứng của máy ảnh đối với ánh sáng đầu vào. Mẫu số của phương trình 2.1 mô hình hóa sự cân bằng trắng (white balance). Cân bằng trắng là sự điều chỉnh độ khuếch đại của máy ảnh vì thế các phản ứng của máy ảnh đến màu trắng (hoặc màu xám sáng) là như nhau trên mọi kênh màu. Điều này nhằm loại bỏ hoặc giảm thiểu tác động của nguồn sáng không trung tính như ánh sáng vàng, xanh, hay hồng lên màu sắc gốc của hình ảnh. Quá trình này làm cho các đối tượng trắng (hoặc các màu trung tính) trong ảnh hiển thị đúng màu sắc tự nhiên của chúng, bất kể nguồn sáng ban đầu có ảnh hưởng ra sao.

Phương trình 2.1 có thể được sử dụng để mô phỏng hiệu ứng chiếu sáng.

- Khi độ sáng chủ đạo và độ sáng hiệu chuẩn như nhau, ảnh đầu ra sẽ được gọi là hình ảnh chuẩn hoặc hình ảnh đã qua hiệu chuẩn và các màu sắc trong ảnh đây sẽ được gọi là màu sắc chuẩn.

- Khi độ sáng chủ đạo và độ sáng hiệu chuẩn khác nhau, ảnh đầu ra trong trường hợp này là ảnh không chuẩn.

Vấn đề mô hình hóa có nhiều vấn đề, còn vấn đề chuẩn hóa có thể được chứng minh về mặt lý thuyết, điều này đã được tác giả trong bài báo trên chứng minh một cách rõ ràng.

2.2.2 *Hiệu ứng của cân bằng trắng*

Cân bằng trắng là một trong những tác nhân quan trọng ảnh hưởng đến chất lượng hình ảnh. Tác nhân cân bằng trắng phụ thuộc vào độ chiếu sáng. Có rất nhiều ảnh kỹ thuật số được đặt dưới điều kiện hiệu chuẩn hoặc rất gần điều kiện chuẩn nhằm tránh sự biến dạng màu sắc, bởi lẽ việc biến dạng màu sắc này có thể được quan sát một cách dễ dàng và được xem như là một hiện tượng gây sự khó chịu cho người xem do nó làm phá vỡ cấu trúc hình ảnh ban đầu. Điều này là hoàn toàn đúng đắn cho những màu sắc cụ thể mà con người có thể nhớ rất rõ, và chúng sẽ được xem là màu trí nhớ. Một trong những màu trí nhớ này là tông màu da.

Như tác giả đã đề cập, con người rất nhạy cảm đối với những biến dạng trong tông màu da. Tông màu da ở đây thực chất là màu sắc về da chính xác hoặc có thể chấp nhận được theo cảm nhận của con người. Màu da ám chỉ tất cả các ảnh RGBs mà máy ảnh có thể cảm nhận được như da dưới các điều kiện chiếu sáng khác nhau. Chúng ta cũng cần lưu ý rằng con người và máy ảnh có thể cảm nhận các màu da khác nhau.

Đối với các máy ảnh, việc cân bằng trắng có thể được thực hiện một cách tự động hoặc một cách thủ công. Về cách lựa chọn thủ công, người dùng có thể chọn tùy chọn tốt nhất cho độ chiếu sáng chủ đạo, trong khi đó tùy chọn tự động có thể cung cấp những cài đặt từ chương trình. Tuy nhiên, điều này không luôn khả thi cho việc lựa chọn hoặc tính toán các yếu tố cân bằng trắng. Điều này đặc biệt đúng khi ánh sáng thay đổi không đồng đều, có thể gây ra những thay đổi màu sắc mạnh hơn.

Đối với một khung ảnh, thường có nhiều hơn một nguồn sáng đầu vào trên cảnh. Nếu như một vật thể chịu nhiều sự ảnh hưởng của các nguồn sáng với các độ chiếu sáng khác nhau SPDs thì việc cân bằng trắng có thể không chính xác cho toàn bộ hình ảnh.



Hình 2.2 Màu sắc của khuôn mặt thay đổi ở các phần khác nhau của trường ánh sáng

Trong hình ảnh 2.2, gương mặt của người trong ảnh được đặt trong một trường ánh sáng không đồng bộ. Chúng ta có thể dễ dàng thấy được phần bên trái chịu sự chiếu sáng của đèn huỳnh quang ở trần nhà và ở phần này đã được máy ảnh cân bằng độ chiếu sáng huỳnh quang, và chúng ta có thể thấy được màu da thật. Tuy nhiên, ở phần bên phải ảnh, gương mặt này chịu ảnh hưởng của ánh sáng ban ngày từ cửa sổ, và điều này đã gây ra sự biến đổi màu sắc, nên chúng ta có thể thấy được màu hơi xanh của gương mặt. Điều này chứng tỏ rằng màu sắc bị méo mó vì cân bằng trắng một phần. Sự méo mó này thay đổi ở các mức độ khác nhau tùy thuộc vào trường chiếu sáng. Chúng ta có thể dễ dàng chú ý rằng hình ảnh có thể thường gặp các trường chiếu sáng không đồng nhất, nhưng chúng thường ít khi được xem xét trong các ứng dụng phát hiện hoặc nhận dạng gương mặt.

Tất nhiên, chúng ta cũng có thể áp dụng một số kỹ thuật hiệu chỉnh màu sắc để cải thiện chất lượng. Tuy nhiên, việc cân bằng trắng không hiệu quả có thể gây mất thông tin, thường rất khó để khắc phục đúng cách.

2.2.2.1 *Những ảnh và màu tiêu chuẩn*

Mặc dù một hình ảnh có thể được chụp trong điều kiện chuẩn (canonical condition), nhưng điều đó không đảm bảo rằng các vật thể xuất hiện với cùng màu sắc dưới các nguồn sáng chuẩn khác nhau. Màu trắng, xám, và đen thường xuất hiện khá giống nhau dưới các nguồn sáng khác nhau, nhưng tất nhiên điều này vẫn có những hạn chế nhất định. Nếu ánh sáng không có phổ bao phủ toàn bộ các thành phần RGB, thì không thể tái tạo đầy đủ các thành phần màu.

Máy ảnh chỉ có thể tái tạo các màu achromatic (trắng, xám, đen) một cách nhất quán dưới các nguồn sáng khác nhau nếu nó được cân bằng trắng theo nguồn sáng hiện tại. Tuy nhiên, màu da có thể được thay đổi đáng kể khi chụp dưới các điều kiện ánh sáng khác nhau.

Thực nghiệm cho thấy rằng màu da có thể được tái tạo với độ chính xác cao bằng cách sử dụng chỉ ba vector cơ sở do tính chất phản xạ quang phổ của các sắc tố trong da (melanin, carotene và hemoglobin). Tuy nhiên, máy ảnh khác nhau có thể tạo ra những biến đổi lớn về màu da.

Đối với màu trắng trong điều kiện chuẩn, các giá trị RGB của nó nên bằng nhau dưới các nguồn sáng khác nhau, miễn là chúng không quá cực đoan. Máy ảnh có thể tái tạo bề mặt trắng chính xác trong nhiều điều kiện ánh sáng, nhưng vẫn có những giới hạn vật lý như gain control (kiểm soát độ lợi). Nếu máy ảnh có phản hồi tuyến tính trong một phạm vi tín hiệu đầu vào nhất định, thì các vật thể màu xám trong phạm vi đó sẽ được tái tạo dưới dạng màu xám.

2.2.2.2 *Những ảnh và màu không tiêu chuẩn*

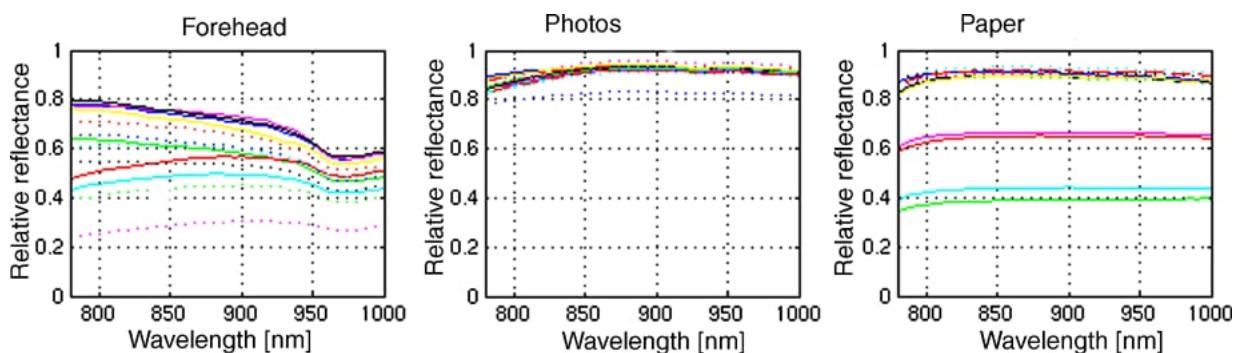
Nếu như ảnh không được chụp dưới ánh sáng được hiệu chỉnh bởi camera, thì màu sắc sẽ còn biến dạng nhiều hơn nữa. Sự biến dạng ở đây tức là sự thay đổi về màu sắc.

Màu sắc của da có xu hướng thay đổi theo cách mà màu của nguồn sáng thay đổi. Chẳng hạn như, khi ánh sáng có thành phần đỏ nhiều hơn sẽ khiến cho màu sắc của da trở nên đỏ hơn, và ánh sáng có thành phần xanh nhiều hơn sẽ khiến cho màu sắc của da trở nên xanh hơn... Tất nhiên trong một nguồn sáng có quang phổ tăng đột biến sẽ khiến cho màu sắc càng trở nên biến dạng hơn, bởi vì camera có phạm vi phản hồi động là giới hạn, màu sắc bị biến dạng cũng có thể do độ bão hòa hay độ phơi sáng kém.

2.3 Phân tách nguồn dữ liệu về da

Có rất nhiều vật liệu, như mực hay thuốc nhuộm, được sử dụng nhằm mô phỏng về bề ngoài của làn da. Một vài nghiên cứu đã được thực nghiệm săn để kiểm tra xem sự mô phỏng này có hoạt động hiệu quả như thế nào và làm thế nào để phân biệt được giữa da thật và da nhân tạo được mô phỏng.

Dữ liệu của da có thể được thu thập từ nhiều nguồn khác nhau như dữ liệu về gương mặt thật, ảnh hoặc từ bản in ấn. Những nguồn này thường không thể được xác định từ dữ liệu RGB thông thường, nên chúng ta cần dữ liệu quang phổ.



Hình 2.3 Vùng quang phổ gần tia hồng ngoại cho da mặt

Một vùng quang phổ cần được xem xét là vùng quang phổ gần tia hồng ngoại. Hình 2.3 chỉ ra rằng vùng quang phổ gần tia hồng ngoại cho da gương mặt thật (trái), da mặt từ ảnh (giữa) và da mặt từ giấy in của ba màu da khác nhau.

Vùng quang phổ từ ảnh và giấy là phẳng và khác biệt so với quang phổ của khuôn mặt thật. Do đó, tỷ lệ đơn giản giữa hai kênh có thể được sử dụng để tách da thật ra khỏi các nguồn khác. Các mức độ khác nhau trong vùng quang phổ thực giữa các màu da bắt đầu giảm dần theo mức độ của bước sóng. Các nhóm nước da có thể được tách biệt trong quang phổ của giấy in, nhưng điều này là ngược lại đối với ảnh.

Màu da của ma-nơ-canhh cũng được ưa chuộng, nhưng điều này rõ ràng là khác so với da thật. Giáo sư Kim cùng với những người cộng sự đã nghiên cứu những sự khác

nhau giữa những mặt nạ da giả với da mặt thật. Họ kết luận rằng các bước sóng 685nm và 850nm có thể được sử dụng để phân biệt chúng.

2.4 Mô hình hóa màu da

Mô hình hóa màu da là việc thống kê về các tông màu da có thể có. Để tạo ra một mô hình như vậy, trước tiên ta phải chọn không gian màu mà mô hình có thể thuộc vào, sau đó áp dụng các mô hình học để mô tả các màu da có thể có và cuối cùng là dữ liệu mà mô hình được xác định. Hiệu suất của mô hình phụ thuộc vào tất cả các yếu tố này và là sự đánh đổi giữa tính tổng quát của mô hình và độ chính xác cho một hình ảnh nhất định.

Các phương pháp phát hiện da đã được so sánh trong một số nghiên cứu sử dụng dữ liệu khác nhau. Các nghiên cứu khác, có thể là do tính tối ưu của mô hình phụ thuộc vào mục đích, dữ liệu, vật liệu và các tham số mô hình của nó.

2.4.1 Biểu hiện của các sắc tố da ở các không gian màu khác nhau dưới độ chiếu sáng khác nhau

Đối với việc xử lý dữ liệu về da, không gian màu cũng có ảnh hưởng đến việc nhận diện da. Không phải tất cả không gian màu đều giống nhau: các không gian màu này có thể được ánh xạ đến những giá trị RGB khác nhau, và những giá trị RGB này sẽ được dùng để phân biệt những màu sắc cụ thể. Thậm chí một hỗn hợp của những không gian màu cũng có thể được sử dụng, ít nhất là đối với hình ảnh chuẩn hoặc gần chuẩn.

Cũng như được đề cập trước đó, việc chuyển đổi không gian màu không thể loại bỏ được sự thay đổi sắc độ do việc chiếu sáng hoặc những ảnh hưởng được gây ra bởi nhiều. Trong thực tế, nhiều có thể gây hại cho các giá trị RGB thấp hoặc gần ngưỡng. Sự điều chỉnh ánh sáng hoặc thiếu ánh sáng có thể có một ảnh hưởng to lớn đến sắc tố da có thể có. Nếu không có bộ điều khiển hoặc bộ khuếch đại ánh sáng tự động, một kênh màu có thể có những giá trị thấp hoặc thậm chí những giá trị này có thể bị cắt giảm. Chính vì vậy, những màu sắc của da đã được nghiên cứu dưới những sự chiếu sáng khác nhau.

Như chúng ta đã biết, các tọa độ không gian của RGB thường hướng đến thiết bị, nhưng chúng có thể được chuyển đổi thành những không gian hướng đến thị giác con người như không gian XYZ hoặc CIE Lab. Một sự chuyển đổi đúng đắn từ không gian máy tính (thiết bị) sang không gian thị giác con người đòi hỏi một ma trận biến đổi phụ thuộc vào sự chiếu sáng (an illumination-dependent transform matrix), và ma trận này cũng bao gồm hiệu ứng của những đặc tính thiết bị. Tất nhiên, chúng ta đã có những ma trận biến đổi tổng quát. Không có bất kỳ ma trận biến đổi làm giảm đi hiệu ứng thay đổi ánh sáng bởi vì những ma trận biến đổi này đã ảnh hưởng sẵn đến tọa độ không gian RGB.

Nhiều không gian màu hướng đến thiết bị có thể được phân loại dựa trên phương pháp chuyển đổi, được chia làm hai nhóm:

- Nhóm 1: không gian màu sử dụng biến đổi tuyến tính từ RGB
- Nhóm 2: những kết quả của việc biến đổi không gian màu thu được thông qua phép biến đổi phi tuyến.

Để hiểu rõ hơn về vấn đề trên đang bàn luận, chúng ta sẽ đi qua các ví dụ về việc biến đổi các không gian màu.

- Các phép biến đổi tuyến tính (linear transforms) dựa trên những không gian màu là: I1I2I3, YES, YIQ, YUV, YCrCb.
- Những phép biến đổi phi tuyến (nonlinear transforms) là: NCC rgb, rgb được điều chỉnh (rgb modified), natural logarithm ln-chromaticity, P1P2, l1l2l3, những tỷ lệ giữa các kênh màu (G/R, B/R, và B/G), HSV, HSL, ab điều chỉnh (modified ab), TLS và Yuv.

Sự chồng lấp giữa những màu da khác nhau thay đổi theo không gian màu. Sự chồng lấp giữa hai màu da (mặt tái nhợt, xanh xao và mặt da vàng) đã được so sánh trong những không gian màu khác nhau và thông qua những máy ảnh khác nhau: tỷ lệ chồng chéo giữa chúng khá cao trong tất cả không gian màu (dao động từ 50% đến 75%) khi sử dụng các hình ảnh chuẩn khác nhau. Khi sử dụng những hình ảnh tiêu chuẩn và không tiêu chuẩn, sự chồng lấp giữa những màu khác nhau vẫn tăng lên do có nhiều màu sắc hơn rơi vào vùng đó. Tuy nhiên, khi so sánh dữ liệu về da từ những máy ảnh khác nhau, tỷ lệ chồng chéo giữa dữ liệu da từ RGBs là nhỏ hơn và phụ thuộc vào những máy ảnh được sử dụng trong việc so sánh này. Chính vì vậy, chúng ta có thể khẳng định một điều rằng những không gian màu và máy ảnh được sử dụng có ảnh hưởng đến việc phân loại da và hơn thế nữa cho việc nhận dạng khuôn mặt.

2.4.2 Không gian màu cho da

Nhìn nhận chung, đã có một vài không gian màu được đề xuất cho mô hình về màu da, nhưng cho đến hiện nay, chưa có không gian màu nào được chứng minh là có tính vượt trội. Tuy nhiên, điều này thường như thể hiện rằng những không gian màu mà cường độ của nó không được phân biệt rõ ràng khỏi sắc độ là tương tự như những không gian tọa độ RGBs.

Sự phân biệt này có thể được đánh giá bằng việc sử dụng sự tuyến tính (linear) hoặc dữ liệu RGB được tuyến tính hóa: không gian tọa độ RGB sẽ được biến đổi thành không gian màu sử dụng các phép thay thế sau: $R \rightarrow cR$, $G \rightarrow cG$, và $B \rightarrow cB$, trong đó

đại lượng c mô tả sự thay đổi chuẩn về mức độ của cường độ màu. Nếu hệ số c không bị triệt tiêu đối với các mô tả sắc độ, sự phân biệt này sẽ không hoàn chỉnh.

Tọa độ màu chuẩn hóa (Normalized color coordinates – NCC) thường được sử dụng trong những mô hình, và khi dùng chúng, chúng ta có thể phân tách một cách rõ ràng sắc độ và cường độ màu. Để tránh những thay đổi về cường độ, chỉ có những tọa độ về sắc độ được sử dụng. Điều này có nghĩa là cường độ sẽ là hằng số và sắc độ sẽ là biến số. Những không gian màu khác nhau được so sánh trong khía cạnh về hiệu suất và khả năng chuyển đổi của mô hình. Hiệu suất của NCC và CIE xy vượt trội hơn so với một vài mô hình về màu da khác. Điều này cũng đã được chứng minh rằng tọa độ màu chuẩn hóa - NCC có một khả năng phân biệt tốt.

Một màu có thể được xác định duy nhất trong cường độ của nó và hai tọa độ sắc độ vì $r + g + b = 1$. Những tọa độ về sắc độ trong không gian màu NCC được định nghĩa như sau:

$$r = \frac{R}{R+G+B} \quad (2.2)$$

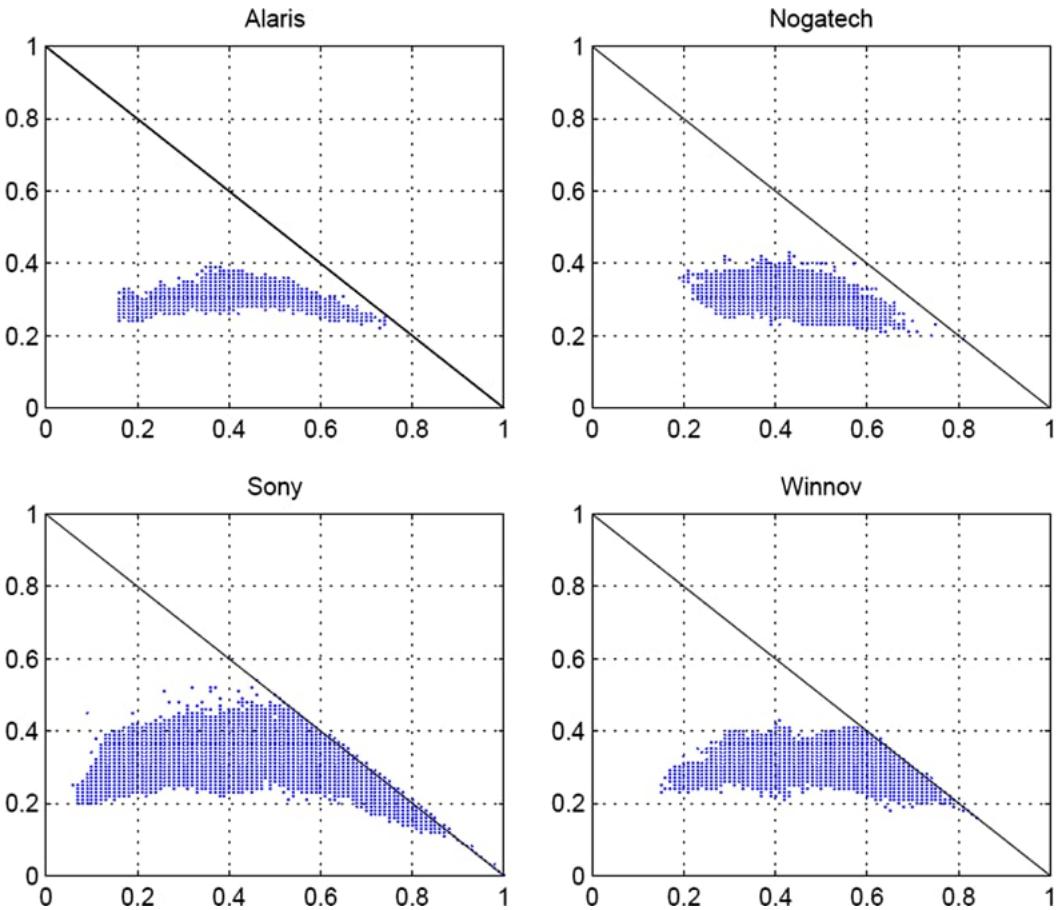
$$g = \frac{G}{R+G+B} \quad (2.3)$$

Cường độ này sẽ bị triệt tiêu khỏi tọa độ sắc độ (chromaticity coordinates) bởi vì cường độ này được tính toán bằng cách lấy giá trị vector biểu diễn của một kênh màu chia cho tổng tất cả các vector biểu diễn (cường độ) tại mỗi điểm pixel.

Việc mô hình hóa có thể được hoàn thành bằng cách chỉ sử dụng các tọa độ không gian về sắc độ để giảm thiểu ảnh hưởng của những sự thay đổi cường độ chiếu sáng, những thay đổi này khá phổ biến trong những videos và ảnh. Một số mô hình bao gồm cường độ (intensity) nhưng cần nhiều dữ liệu hơn để xây dựng mô hình và chi phí tính toán tăng lên do một thành phần bên thứ ba.

2.4.3 Mô hình hóa màu da và độ chiếu sáng

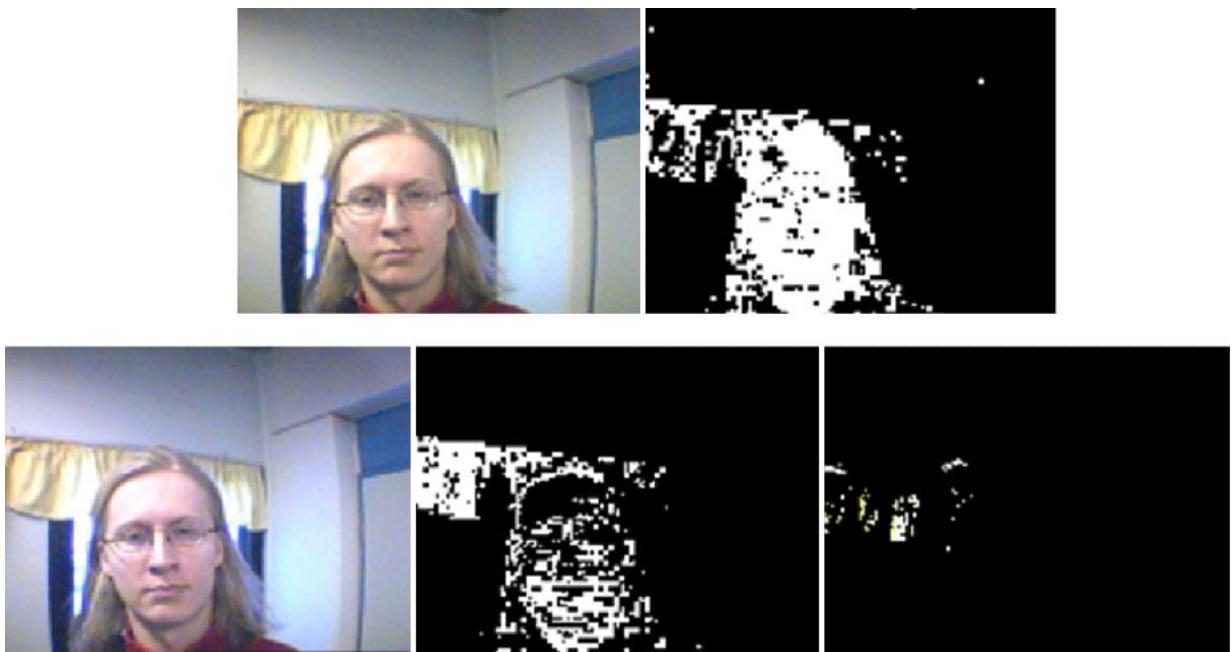
Phần 2.2 chứng minh rằng độ sáng ảnh hưởng đến màu da trong cả hình ảnh chuẩn (canonical images) và không chuẩn (non-canonical). Hơn nữa, sự phụ thuộc này là đặc thù của máy ảnh: cảm biến máy ảnh và quá trình xử lý hình ảnh bên trong của máy ảnh ảnh hưởng đến quá trình tạo màu, và do đó ảnh hưởng đến kết quả cuối cùng. Do đó, việc tạo ra một mô hình tổng quát là rất khó.



Hình 2.4 Camera và các đặc tính của nó xác định vị trí da

Nhiều thuật toán phát hiện khuôn mặt giả sử rằng hình ảnh được chụp trong điều kiện chuẩn hoặc gần chuẩn. Đối với nhiều tập dữ liệu, điều này là đúng. Một ví dụ về loại tập dữ liệu hình ảnh này là một tập ảnh cá nhân. Khi độ sáng thay đổi, các phương pháp tiếp cận trước đây có nguy cơ thất bại cao. Tất nhiên, hình ảnh có thể được hiệu chỉnh màu hoặc thuật toán giữ nguyên màu, nhưng đôi khi điều này có thể dẫn đến tình trạng biến dạng màu thậm chí còn nghiêm trọng hơn.

Hsu et al. [4] đã đề xuất phương pháp tiếp cận dựa trên hiệu chỉnh màu sắc: màu sắc trong hình ảnh được hiệu chỉnh để da xuất hiện theo tông màu da và sau phân đoạn này, hình ảnh sử dụng mô hình màu da. Hiệu chỉnh màu sắc dựa trên một điểm ảnh có giá trị độ sáng cao được cho là thuộc về một vật thể màu trắng. Các điểm ảnh này được sử dụng để tính hệ số hiệu chỉnh được áp dụng cho hình ảnh. Phương pháp tiếp cận này có thể không thành công vì nhiều lý do như mất dữ liệu do bão hòa hoặc nếu điểm ảnh có độ sáng cao thuộc về một vật thể không phải màu trắng. Trường hợp sau được minh họa trong hình ảnh 2.5.



Hình 2.5 Kết quả của mô hình của Hsu et al.

Trong đó, hàng trên hiển thị kết quả phân đoạn màu sử dụng mô hình Hsu et al. mà không có phần hiệu chỉnh màu. Hàng dưới hiển thị phân đoạn với phương pháp hiệu chỉnh màu của phương pháp. Hiệu chỉnh màu không thành công vì rèm màu vàng có giá trị độ sáng cao nhất và được coi là vật thể màu trắng.

Đối với một mô hình da tổng quát hơn, người ta nên sử dụng kiến thức về các thay đổi về độ sáng, hiệu chuẩn và cài đặt máy ảnh như trong phương pháp tiếp cận dựa trên vị trí da. Nhược điểm của mô hình này là không cụ thể như các mô hình chuẩn—nhiều tông màu hơn. Do đó, nhiều đối tượng không phải da sẽ được đề xuất là da. Vì bản thân màu sắc hiếm khi đủ để xác định mục tiêu có phải là da hay không nên các ứng viên khuôn mặt sẽ được xử lý thêm.

2.4.4 Những mô hình toán học về màu da

Mô hình màu da là một vùng được xác định toán học trong không gian màu hoặc một phương pháp tiếp cận thống kê, trong đó một xác suất thuộc về da được gán cho các tông màu. Mô hình có thể là cố định hoặc thích nghi, và trong trường hợp sau, việc cập nhật mô hình sẽ phụ thuộc vào việc nó được áp dụng trên từng ảnh hay các khung hình video.

Phương pháp dựa trên vùng (area-based approach) sử dụng một ràng buộc không gian trong không gian màu để xác định các vùng có thể là da. Hình dạng của ràng buộc này có thể là các ngưỡng đơn giản về màu da hoặc hàm phức tạp hơn. Thông thường, không có ngưỡng cụ thể nào được áp dụng, vì các màu nằm trong vùng xác định sẽ được coi là da. Các mô hình này thường giả định rằng da đã thuộc vùng xác định hoặc có thể

được hiệu chỉnh để thỏa. Một ngoại lệ là vùng da mà khi các sự thay đổi về ánh sáng cũng được đưa vào mô hình.

Có thể thích ứng mô hình ngay cả với từng ảnh riêng lẻ, mặc dù tính hiệu quả phụ thuộc vào độ chính xác của các giả định đãng sau tiêu chí thích ứng. Phương pháp thích ứng thường sử dụng một mô hình da tổng quát thu được từ một tập hợp ảnh đại diện, sau đó tinh chỉnh để phù hợp với từng ảnh cụ thể. Ví dụ, trong nghiên cứu của Cho et al., giai đoạn tinh chỉnh giả định rằng biểu đồ màu da (skin color histogram) là đơn đỉnh (unimodal) và màu da chủ yếu xuất hiện ở các vùng thực sự là da. Tuy nhiên, phương pháp này có thể thất bại nếu ảnh chứa các vật thể không phải da nhưng có màu giống da chiếm ưu thế, hoặc nếu biểu đồ màu không phải là đơn đỉnh.

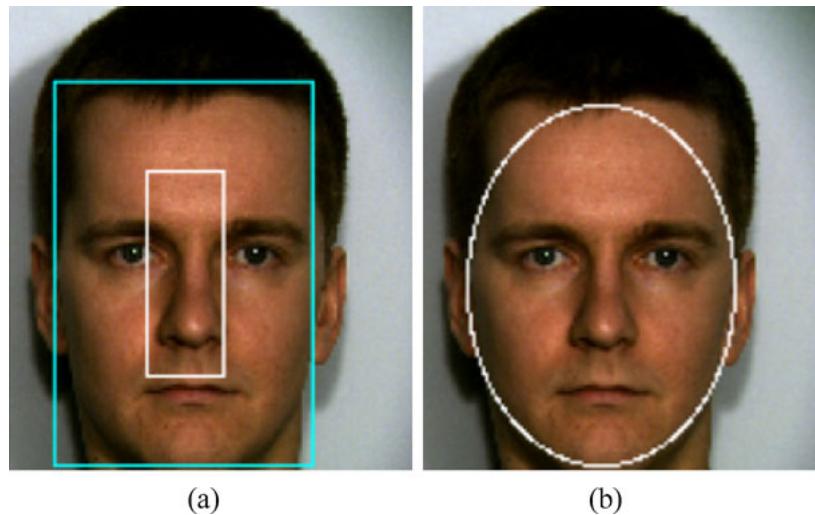
Thách thức của phương pháp dựa trên xác suất (probability-based approach) là làm sao có thể xác định được phân phối xác suất của màu da một cách đáng tin cậy. Điều này đòi hỏi phải thu thập một tập dữ liệu hình ảnh đại diện để xây dựng mô hình. Một ví dụ về mô hình thống kê là của Jones và Rehg, trong đó họ tính toán biểu đồ màu và mô hình Gaussian bằng cách sử dụng hơn 1 tỷ pixel có nhãn. Nhiều mô hình thống kê khác, chẳng hạn như mạng SOM (Self-Organizing Map) hoặc mạng nơ-ron nhân tạo, cũng đã được đề xuất cùng với các bài đánh giá. Ngoài mô hình thống kê, ta cũng cần xác định ngưỡng (threshold) để phân biệt giữa vùng da và vùng không phải da. Tuy nhiên, việc tự động tìm ngưỡng này là rất khó khăn, vì mô hình xác suất được tạo ra có thể không phù hợp với tất cả các hình ảnh.

2.4.4.1 Chuỗi Video

Việc xử lý chuỗi video tương tự như xử lý các hình ảnh đơn lẻ, độc lập. Do đó, phương pháp phát hiện da đã được đề cập trước đó cũng có thể áp dụng cho video. Các mô hình màu da cố định phù hợp với các video mà sự thay đổi ánh sáng là tối thiểu. Tuy nhiên, trong hầu hết các trường hợp, điều kiện này không được đảm bảo, do đó mô hình màu da cần được cập nhật liên tục. Việc thích ứng mô hình (model adaptation) thường dựa vào sự phụ thuộc giữa các khung hình liên tiếp, điều này đúng với hầu hết các video: các khung hình liên tiếp có sự phụ thuộc tuần tự. Điều này có thể được quan sát trong Hình dưới, trong đó sự chồng lấp giữa các sắc độ (chromaticities) của hai khung hình liên tiếp là đáng kể.

Nếu sự thay đổi ánh sáng giữa các khung hình diễn ra chậm (không có sự thay đổi đột ngột về màu sắc của đối tượng) hoặc đối tượng di chuyển trong vùng ánh sáng không đồng nhất một cách chậm rãi, thì mô hình màu da có thể thích nghi với các thay đổi màu sắc. Tuy nhiên, điều này đòi hỏi một số ràng buộc để chọn các pixel sử dụng trong quá trình cập nhật mô hình. Ba cơ chế thích ứng (adaptive schemes) khác nhau đã được đề xuất: Hai phương pháp sử dụng ràng buộc không gian (spatial constraints) (xem hình

2.6) và phương pháp sử dụng skin locus. Mặc dù có sự khác biệt trong cách tiếp cận, nhưng ý tưởng chung của cả ba phương pháp là sử dụng một số ràng buộc để chọn pixel cho quá trình cập nhật mô hình. Ràng buộc không gian (Spatial Constraints): Phương pháp của Raja et al. [39] cập nhật mô hình màu da bằng cách chọn các pixel bên trong vùng khuôn mặt được xác định. Các pixel được lấy từ $1/3$ diện tích vùng khuôn mặt được xác định và $1/3$ diện tích rìa của vùng này. Phương pháp của Yoo and Oh cho rằng vùng xác định nên giống với hình dạng thực tế của khuôn mặt, do đó họ chọn tất cả pixel bên trong vùng khuôn mặt hình elip. Phương pháp Skin Locus: Có hai cách để sử dụng skin locus: Sử dụng toàn bộ locus, hoặc Chỉ sử dụng một phần của locus để chọn các pixel có màu da từ khuôn mặt và vùng lân cận.



Hình 2.6 Ràng buộc không gian được đề xuất cho mô hình hóa màu da thích ứng

Hình 2.6 biểu diễn hình ràng buộc không gian được đề xuất cho mô hình hóa màu da thích ứng, trong đó hình ảnh bên trái cho thấy phương pháp được đề xuất bởi Raja và cộng sự. Hộp bên ngoài cho biết khuôn mặt được định vị trong khi các điểm ảnh bên trong hộp bên trong được sử dụng để cập nhật mô hình. Hình ảnh bên phải cho thấy ràng buộc hình elip của Yoo and Oh.

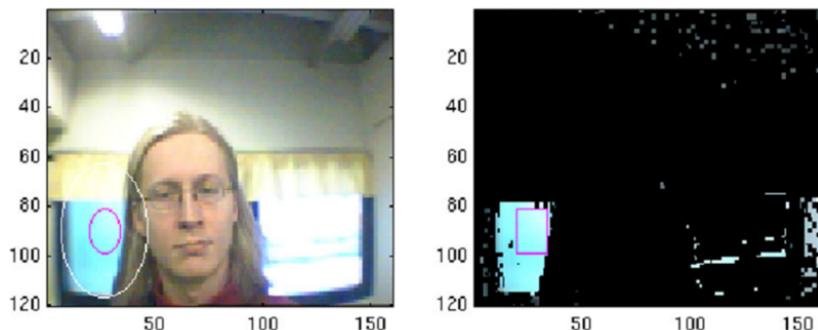
Có nhiều phương pháp có thể áp dụng để cập nhật mô hình màu da, nhưng có lẽ phương pháp phổ biến nhất là phương pháp trung bình động:

$$\tilde{M} = \frac{(1 - \alpha) \times M_t + \alpha \times M_{t-1}}{\max((1 - \alpha) \times M_t + \alpha \times M_{t-1})} \quad (2.4)$$

trong đó \tilde{M} là mô hình mới, M là mô hình, t là số khung hình và α là hệ số trọng số. Thông thường, hệ số trọng số được đặt thành 0.5 để nhấn mạnh như nhau vào mô hình màu da của khung hình hiện tại và trước đó. Phương pháp trung bình động cung cấp sự chuyển đổi mượt mà giữa các mô hình từ các khung hình khác nhau. Nó cũng

làm giảm hiệu ứng nhiễu, có thể thay đổi màu điểm ảnh mà không có bất kỳ sự thay đổi nào ở các yếu tố bên ngoài và do đó gây bất lợi cho các mô hình.

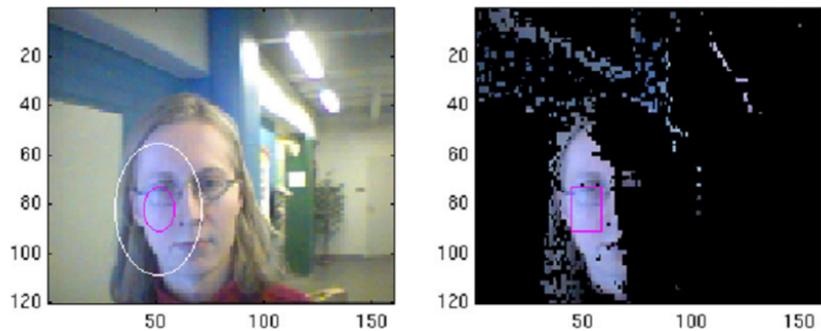
Tuy nhiên, các mô hình ràng buộc không gian đã được chứng minh là rất nhạy cảm với các lỗi định vị, do đó, chúng có thể dễ dàng thích ứng với các đối tượng không phải da. Sự thất bại do các ràng buộc này có thể xảy ra ngay cả khi thay đổi độ sáng khá vừa phải. Trong Hình 2.7, phương pháp của Raja và cộng sự đã thất bại khi theo dõi khuôn mặt trên chuỗi video và mô hình màu da được điều chỉnh cho mục tiêu không phải màu da, như thể hiện trong hình ảnh này.



Hình 2.7 Việc theo dõi khuôn mặt của Raja et al thất bại và thích ứng với mục tiêu không phải khuôn mặt

trong đó, hình ảnh bên trái hiển thị "khuôn mặt được định vị". Hình ảnh bên phải hiển thị các pixel được chọn bởi mô hình màu da hiện tại. Hộp màu đỏ hiển thị các pixel được sử dụng để làm mới mô hình.

Ràng buộc do Raja và cộng sự đề xuất dễ dàng thất bại khi thay đổi trường chiếu sáng không đồng nhất, như minh họa trong hình 2.8. Mô hình được cập nhật bằng cách sử dụng pixel bên trong vị trí định vị và do đó, nó chỉ có thể thích ứng với những thay đổi về độ sáng toàn cục, nhưng không thích ứng với sự thay đổi trường chiếu sáng không đồng nhất. Vị trí định vị chính xác của khuôn mặt không quá nhạy đối với phương pháp tiếp cận dựa trên vị trí da vì các pixel không có màu da có thể bị lọc ra. Các vật thể có màu da lớn được kết nối với khuôn mặt là vấn đề và cần có các tín hiệu khác ngoài màu sắc để giải quyết vấn đề này.



Hình 2.8 Ràng buộc được đề xuất bởi Raja chọn một tập hợp các điểm ảnh không đại diện cho điểm da

2.5 Tín hiệu màu cho bài toán phát hiện khuôn mặt

Màu sắc là một gợi ý hữu ích trong phát hiện khuôn mặt vì giúp thu hẹp vùng tìm kiếm bằng cách chọn các khu vực có màu da. Tuy nhiên, chỉ sử dụng màu da là chưa đủ để phân biệt khuôn mặt với các vật thể khác có màu tương tự (như tay, gỗ, v.v.), do đó cần thêm các bước xử lý.

Tùy thuộc vào độ bền của mô hình da và sự thay đổi trong điều kiện chiếu sáng, người ta có thể nhận thấy hai trường hợp:

- **Trường hợp 1:** Mô hình màu da hoạt động ổn định, cho kết quả đáng tin cậy trong điều kiện ánh sáng và cài đặt máy ảnh nhất định. Khi đó, chỉ cần kiểm tra nhanh đặc điểm khuôn mặt trên các vùng da đã phát hiện.
- **Trường hợp 2:** Mô hình màu da kém chính xác hoặc thất bại, do thay đổi ánh sáng hoặc thiết bị. Khi đó, cần kết hợp phương pháp khác (dựa trên đặc điểm khuôn mặt hoặc hình dạng) để tìm khuôn mặt trong vùng da được phát hiện.

Dù trong trường hợp nào, việc sử dụng màu vẫn giúp tăng tốc độ phát hiện khuôn mặt. Hầu hết các phương pháp dựa trên màu đều xác định vùng da trước, sau đó nhóm chúng bằng phân tích thành phần kết nối. Tiếp theo, xác suất là khuôn mặt được kiểm tra bằng cách tìm đặc điểm như mắt và miệng (vốn khác màu da). Các phương pháp chủ yếu khác nhau ở lựa chọn không gian màu và thiết kế mô hình da.

Một số nghiên cứu tiêu biểu:

- Hsu et al. [4] sử dụng mô hình da hình elip trong không gian YCbCr, kết hợp bù sáng, sau đó phân đoạn vùng da và xác minh bằng bản đồ mắt, miệng. Định vị khuôn mặt gồm hai bước chính:

- **Xác định ứng viên khuôn mặt:** Các pixel có tông màu da được xác định dựa trên mô hình da hình elip trong không gian màu YCbCr. Trước khi phân đoạn,

ảnh được hiệu chỉnh ánh sáng để tăng độ ổn định. Sau đó, các pixel da được phân nhóm thành các vùng kết nối, từ đó trích xuất các ứng viên khuôn mặt.

- Phát hiện đặc điểm khuôn mặt: Để xác minh vùng ứng viên có phải khuôn mặt hay không, hệ thống xây dựng bản đồ mắt, miệng và đường viền khuôn mặt, tận dụng đặc điểm màu sắc của các vùng này khác với màu da. Kết quả thí nghiệm cho thấy phương pháp này hoạt động tốt trên nhiều bộ dữ liệu thử nghiệm, nhưng chưa có nghiên cứu so sánh trực tiếp với các phương pháp khác.

- Garcia & Tziritas [5] áp dụng phân cụm màu trong YCbCr và HSV (thang màu HSL Hue (màu) - Saturation (độ bão hòa) - Lightness (độ sáng tối) thay vì RGB), sau đó phân tích kết cấu bằng biến đổi wavelet để nhận diện khuôn mặt.

- Bước 1: Áp dụng phân cụm và lọc màu trong không gian YCbCr và HSV để trích xuất các vùng da từ ảnh gốc. Quá trình này giúp tạo ra các vùng da có độ phân giải thấp nhưng rõ ràng hơn.

- Bước 2: Kết hợp các vùng da tương đồng bằng một thuật toán hợp nhất lặp, tạo ra danh sách các ứng viên khuôn mặt.

- Bước 3: Kiểm tra ứng viên khuôn mặt dựa trên hình dạng và kích thước, sau đó thực hiện phân tích kết cấu bằng biến đổi wavelet để xác định xem vùng đó có chứa khuôn mặt thực sự hay không.

Kết quả thử nghiệm cho thấy tỷ lệ phát hiện đạt 94.23% trên tập dữ liệu gồm 100 ảnh (104 khuôn mặt). Tuy nhiên, phương pháp này có nhược điểm là tiêu tốn nhiều tài nguyên tính toán do thuật toán phân đoạn phức tạp và phân tích wavelet tốn thời gian.

- Sobottka và Pitas [6] dùng phân đoạn màu HSV, phân tích thành phần kết nối, và mạng nơ-ron để xác minh.

- Bước 1: Sử dụng không gian màu HSV để xác định các vùng da.

- Bước 2: Tiến hành phân tích thành phần kết nối và tính toán hình elip tốt nhất phù hợp với mỗi vùng da.

- Bước 3: Để kiểm tra ứng viên khuôn mặt, hệ thống sử dụng mạng nơ-ron, với đầu vào là 11 giá trị mô-men hình học bậc thấp nhất để xác minh tính hợp lệ của khuôn mặt.

Hệ thống đạt tỷ lệ phát hiện 85% trên tập thử nghiệm gồm 100 ảnh, tuy nhiên nhược điểm là mô hình cần nhiều dữ liệu huấn luyện và không hoạt động tốt trong điều kiện ánh sáng thay đổi mạnh.

- Haiyuan et al. [7] kết hợp mô hình màu da và tóc trong không gian CIE XYZ để khớp với mẫu đầu người.
 - **Bước 1:** Xây dựng hai mô hình mờ riêng biệt mô tả màu da và tóc trong không gian màu CIE XYZ.
 - **Bước 2:** Trích xuất các vùng màu da và màu tóc, sau đó so sánh với các mẫu hình dạng đầu người có sẵn bằng phương pháp so khớp mẫu dựa trên lý thuyết mờ.

Phương pháp này có ưu điểm là có thể xác định khuôn mặt ngay cả khi không có đủ đặc điểm mắt, mũi, miệng, nhưng nhược điểm là phụ thuộc vào mô hình tóc, dễ sai lệch với những kiểu tóc không điển hình.



Hình 2.9 Ví dụ phát hiện khuôn mặt sử dụng định vị khuôn mặt dựa trên màu sắc

- Hadid et al. [8] sử dụng mô hình skin locus, tinh chỉnh bằng các tiêu chí đối xứng khuôn mặt và sắp xếp thành phần kết nối.
 - **Bước 1:** Trích xuất vùng da bằng mô hình skin locus, sau đó tinh chỉnh bằng một số tiêu chí như đối xứng khuôn mặt, sự hiện diện của các đặc điểm khuôn mặt và biến thiên cường độ điểm ảnh.
 - **Bước 2:** Tiến hành kiểm tra từng ứng viên khuôn mặt qua một loạt các bước tinh lọc được sắp xếp theo cấu trúc dạng thác để đảm bảo độ chính xác cao. Kết quả cho thấy hệ thống có thể xử lý tốt các điều kiện khác nhau như kích thước, góc quay, ánh sáng và nền phức tạp.
- Hadid & Pietikänen [9] kết hợp tìm vùng da trước khi quét toàn bộ ảnh, dùng SVM với đặc trưng LBP để tăng độ chính xác.



Hình 2.10 Ví dụ phát hiện gương mặt sử dụng bộ phát hiện gương mặt dựa trên màu

Tuy nhiên, các phương pháp dựa trên màu thường phụ thuộc vào máy ảnh và chưa được đánh giá rộng rãi trên nhiều điều kiện ánh sáng khác nhau. Hầu hết hệ thống phát hiện khuôn mặt hiện nay vẫn dựa trên ảnh xám vì tính ổn định cao hơn, dù tồn tại nguyên tính toán hơn. Một ví dụ điển hình là phương pháp của Viola & Jones, sử dụng đặc trưng Haar và thuật toán AdaBoost để chọn lọc đặc trưng hiệu quả, giúp phát hiện khuôn mặt trong thời gian thực với ảnh nhỏ. Tuy nhiên, khi kích thước ảnh lớn, kết hợp thêm thông tin màu hoặc chuyển động có thể giúp tăng tốc quá trình phát hiện.

2.6 Tín hiệu màu cho bài toán nhận diện khuôn mặt

Vai trò của màu sắc trong nhận diện vật thể và đang là chủ đề đáng được tranh luận. Tuy nhiên, vẫn chưa có nhiều nghiên cứu về giá trị của màu sắc cho nhận dạng khuôn mặt. Đại đa số các nghiên cứu chỉ chú trọng vào cường độ ánh sáng của khuôn mặt, và do đó bỏ qua yếu tố màu sắc, vì nhiều lý do:

- Sự thiếu dẫn chứng về việc thị giác con người sử dụng màu sắc để nhận dạng khuôn mặt. Trong một bài nghiên cứu [10], các nhà nghiên cứu nhận thấy rằng các ứng viên có thể dễ dàng nhận diện các khuôn mặt kể cả khi sắc thái đã bị đảo ngược, và do đó đi đến kết luận rằng màu sắc có lẽ không mang lại lợi ích gì quá lớn ngoài thông tin về độ sáng. Một bài nghiên cứu khác [11] giải thích việc màu sắc ít được coi trọng trong các nghiên cứu này là vì hình dạng khuôn mặt đóng vai trò quá lớn trong nhận diện khuôn mặt, đến mức màu sắc là không cần thiết. Họ đi đến kết luận rằng tuy thành phần sáng trên khuôn mặt mang lại ý nghĩa lớn cho việc nhận dạng, màu sắc lại bị lược bỏ hoàn toàn trong quá trình nhận dạng khuôn mặt. Họ cho rằng màu sắc vẫn có đóng góp dưới điều kiện ánh sáng chất lượng thấp như ước lượng tốt hơn về hình dáng và kích thước của khuôn mặt.
- Sự khó khăn trong việc liên kết giữa nguồn sáng và việc cân bằng trắng của máy ảnh. Như đã trình bày trong những phần trước, nguồn sáng vẫn là một thách thức trong việc tự động hóa nhận dạng khuôn mặt, do đó ta không nên làm phức tạp hơn

vấn đề.

- Sự thiếu hụt ảnh màu trong bộ dữ liệu được sử dụng để kiểm chứng các thuật toán sử dụng màu sắc được đề xuất, đi kèm với sự miễn cưỡng trong việc đề xuất giải pháp mà không thể được sử dụng với kho dữ liệu ảnh trắng đen có sẵn.

Đã có vài nỗ lực trong việc sử dụng màu sắc trong nhận dạng khuôn mặt bao gồm:

- Một thí nghiệm được thực hiện bởi Torres và các cộng sự [12] tính các thành phần chính từ mỗi kênh màu từ 3 không gian màu khác nhau (RGB, YUV, và HSV). Sau đó, ta tính tổng có trọng số của khoảng cách Mahalanobis của từng kênh màu để phân lớp. Với bộ dữ liệu gồm 59 ảnh, kết quả thu được cho thấy tỉ lệ nhận dạng của không gian màu YUV và HSV là 88.14%, trong khi không gian màu RGB cho ra kết quả là 84.75%, giống với việc chỉ sử dụng thành phần độ sáng Y. Từ đó kết luận được đưa ra là màu sắc quan trọng trong nhận dạng khuôn mặt. Tuy nhiên, kết quả này không có sức thuyết phục, bởi độ nhỏ của cơ sở dữ liệu và phương pháp được sử dụng (mặt phẳng riêng) quá đơn giản.
- Một nghiên cứu khác [13] so sánh kết quả lấy được từ các cách chuyển đổi ảnh màu thành ảnh đen trắng như PCA, hồi quy tuyến tính, và giải thuật di truyền với kết quả từ việc chuyển đổi bằng cách lấy trung bình cộng của ba kênh màu RGB. Họ nhận thấy hiệu quả tăng lên từ 4% đến 14% với bộ dữ liệu 280 ảnh, một lần nữa chưa đủ lớn.
- Hai phương pháp khác sử dụng NMF cũng cho ra kết quả tốt hơn so với ảnh đen trắng, trên bộ dữ liệu nhỏ.

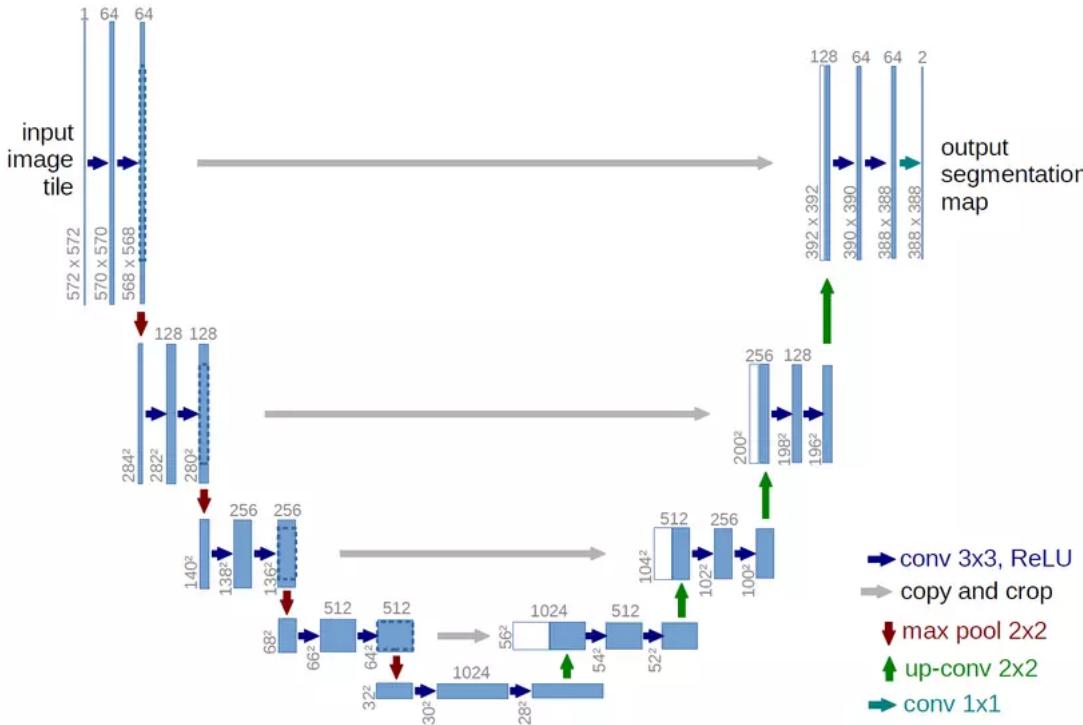
Cho đến gần đây, nhận dạng khuôn mặt bằng màu sắc lại được nhắc tới, với mục tiêu là tìm ra phương pháp sử dụng màu sắc để nâng cao hiệu quả nhận diện khuôn mặt, và cho ra các kết quả rất hứa hẹn:

- Một nghiên cứu đã chỉ ra rằng các không gian màu RGB và XYZ yếu hơn so với các không gian I1I2I3, YUV, và YIQ trên tập dữ liệu FRGC.
- Một nghiên cứu khác tạo ra các mô hình định danh bằng cách chiếu ảnh khuôn mặt tạo bởi các phép LBP đa quang phổ vào không gian LDA. Độ tương đồng sẽ được tính bằng cách hợp độ tương đồng cục bộ với các bộ định danh khu vực. Cách thức này đã được thử nghiệm trên tập dữ liệu XM2VTS và FRGC 2.0.
- Một thí nghiệm khác biến đổi không gian màu từ không gian RGB qua các phép biến đổi tuyến tính, sau đó nối các thành phần ảnh lại để tạo nên tập các vector, rồi

giảm chiều bằng PCA. Cuối cùng họ sử dụng mô hình EFM để nhận dạng. Kết quả cho được tốt hơn ảnh trắng đen và ảnh RGB. Họ cũng đã thử tạo ra không gian màu khác bằng cách sử dụng kênh màu R của không gian màu RGB và I và Q của không gian màu YIQ. Thí nghiệm trên tập dữ liệu FRGC 2.0 cho thấy không gian màu này tăng hiệu quả nhận dạng khuôn mặt một cách rõ rệt bởi tính bù trừ của các thành phần màu sắc.

2.7 Mô hình phân đoạn hình ảnh dựa theo ngữ nghĩa

Mô hình phân đoạn UNet [14] được đề xuất bởi Ronneberger cùng các cộng sự vào năm 2015. Mô hình này kế thừa kiến trúc của mạng tích chập với dạng chữ U tượng trưng cho Auto Encoder và Decoder, và chính mô hình này cũng từng đạt năm giữ vị trí mạnh mẽ về hiệu suất phân đoạn trong các bài toán phân đoạn hình ảnh y tế.



Hình 2.11 Kiến trúc mô hình UNet

Kiến trúc của mạng UNet này gồm hai phần chính: một bộ mã hóa đối xứng với bộ giải mã với kiến trúc hình chữ U độc đáo như hình 2.11. Bộ mã hóa bao gồm một chuỗi 3 lớp tích chập xen kẽ với một tầng max pool. Khi một ảnh đầu vào có độ phân giải 32×32 pixel được truyền xuống vào bộ mã hóa, bộ này sẽ trích xuất một vector đặc trưng có 1024 chiều dữ liệu. Sau đó, vector đặc trưng này được dẫn truyền qua bộ giải mã, bao gồm một bộ 3 lớp tích chập xen kẽ với một lớp up sampling, nhận lấy các đặc trưng được trích xuất và tạo sinh ra kết quả. Ý tưởng chính trong kiến trúc này là các skip connections qua giữa bộ mã hóa và bộ giải mã. Mạng học sâu này sao chép và cắt

các feature map tại mỗi lớp tích chập trong số 3 lớp tại bộ mã hóa và nối nó với bản đồ đặc điểm tại mỗi lớp tích chập lên tại bộ giải mã. Mục đích của các skip connections này là giúp bộ giải mã phục hồi cấu trúc không gian của hình ảnh đầu vào cho đầu ra.

Đối với mạng UNet, tuy là một trong những kiến trúc nổi bật nhất được sử dụng trong phân đoạn ảnh, đặc biệt là trong các ảnh y học, nó vẫn có một số ưu và nhược điểm như sau:

- **Ưu điểm:** Mô hình này có thể hoạt động tốt với một số lượng dữ liệu nhỏ, nhờ vào việc sử dụng cơ chế data augmentation mạnh mẽ kết hợp kiến trúc đối xứng. Thiết kế mô hình đơn giản, nhẹ nhưng hiệu quả, có thể được dùng để triển khai và huấn luyện nhanh chóng so với các kiến trúc phức tạp khác như TransUNet hay DeepLab.
- **Nhược điểm:** Mô hình khó có thể xử lý dữ liệu với ngữ cảnh rộng hoặc các đối tượng lớn. Các phép tích chập cục bộ là nguyên nhân chính khiến cho UNet không thể hiểu được thông tin tổng thể của toàn bộ ảnh, đặc biệt là áp dụng phân đoạn ảnh các vật thể lớn hoặc phân tán không theo quy luật nhất định. Mô hình dễ bị overfitting cũng như khó xử lý những dữ liệu đầu vào có độ phức tạp cao.

2.8 Mô hình phát hiện gương mặt kết hợp học sâu

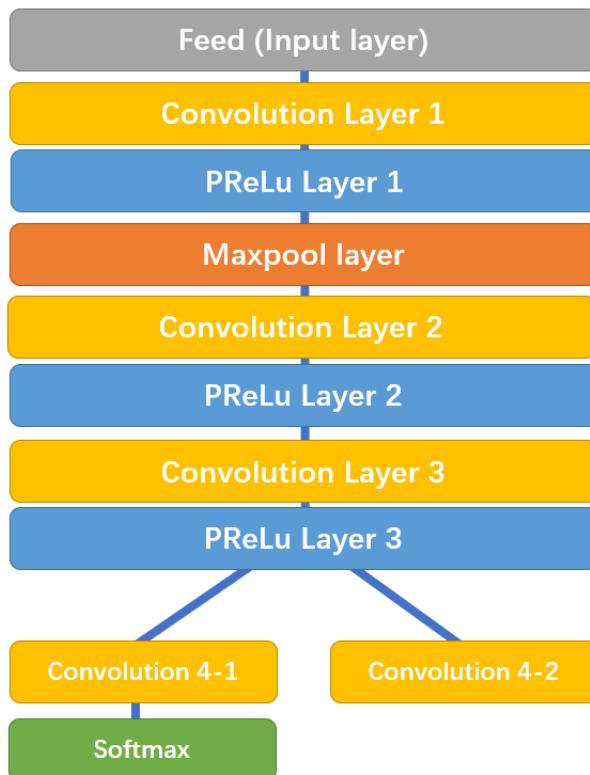
2.8.1 Mô hình MTCNN

Mô hình lần đầu tiên giới thiệu bởi Kaipeng Zhang và cộng sự trong bài báo "**Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks**" [15] vào năm 2016. Bài báo này đã trình bày một hệ thống phát hiện khuôn mặt và định vị landmarks qua ba giai đoạn mạng Convolutional (P-Net, R-Net, và O-Net).

2.8.1.1 Giai đoạn 1: Kiến trúc P-Net (Proposal Network)

Ở giai đoạn này, ta sẽ sử dụng mạng FCN (Fully Connected Network). Mạng khác mạng CNN ở chỗ mạng FCN không sử dụng lớp Dense layer. P-Net được sử dụng để có được các windows tiềm năng và bounding box regression vectors của chúng (tọa độ).

Ta sẽ sử dụng Bounding Box Regression là kỹ thuật dự đoán vị trí của bounding box khi chúng ta cần phát hiện đối tượng (ở đây là khuôn mặt). Sau khi có được tọa độ của bounding boxes một vài tinh chỉnh được thực hiện để loại bỏ một số bounding boxes overlap với nhau (xem trong code sẽ có). Đầu ra của bước này là tất cả bounding boxes sau khi đã thực hiện sàng lọc.



Hình 2.12 Kiến trúc P-Net

Đầu ra của giai đoạn này là một classification map và bounding box regression offsets.

2.8.1.2 Giai đoạn 2: Kiến trúc R-Net (Refine Network)

Tất cả các bounding boxes từ P-Net được đưa vào R-Net. Mạng R-Net có kiến trúc CNN sâu hơn P-Net nhằm loại bỏ các false positives, chọn ra những vùng có xác suất cao. R-Net giảm số lượng bounding boxes xuống, tinh chỉnh lại tọa độ, và áp dụng Non-max suppression.

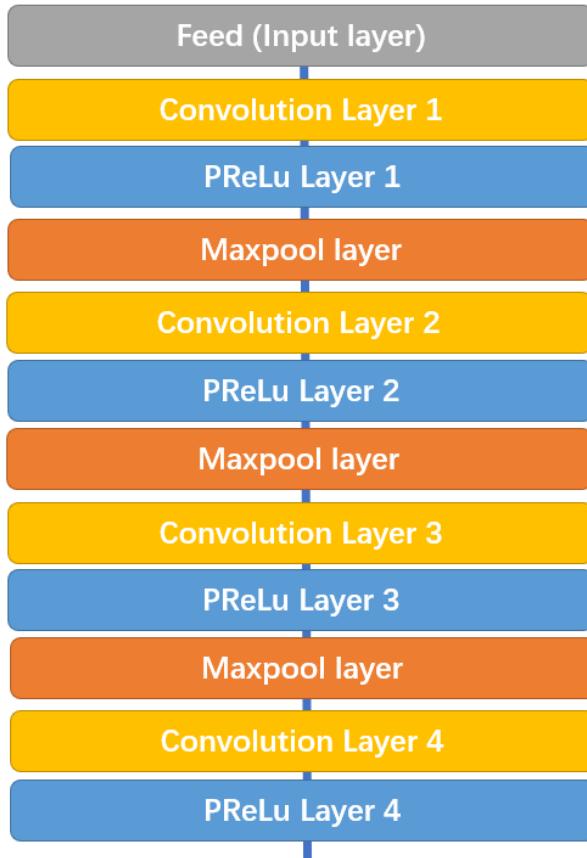
Đầu ra của giai đoạn này là xác suất của candidate window chứa khuôn mặt và cập nhật, tinh chỉnh lại bounding box offsets.



Hình 2.13 Kiến trúc R-Net

2.8.1.3 Giai đoạn 3: Kiến trúc O-Net (Output Network)

Đầu ra của R-Net được sử dụng làm đầu vào của O-Net. O-Net là mạng CNN có kiến trúc sâu và phức tạp nhất trong chuỗi cascade của MTCNN. Ngoài việc thực hiện phân loại và bounding box regression như các mạng trước đó, O-Net còn dự đoán facial landmarks, cung cấp 5 điểm đặc trưng của khuôn mặt (mắt trái, mắt phải, mũi, miệng trái, miệng phải).



Hình 2.14 Kiến trúc O-Net

2.8.1.4 Ưu điểm và nhược điểm

Mạng trên có những ưu điểm và nhược điểm sau:

- **Ưu điểm**

- Mô hình này có thể tích hợp phát hiện khuôn mặt, tính toán landmarks trong cùng một pipeline, hỗ trợ cản chỉnh gương mặt.
- Mô hình này có thể chạy real-time trên CPU với các ảnh kích thước nhỏ.

- **Nhược điểm**

- Mặc dù kiến trúc cascade giúp giảm false positive, nhưng việc thiết kế, training và điều chỉnh các ngưỡng giữa các stage khá phức tạp.
- Khi gặp các khuôn mặt có chất lượng thấp, nghiêng mạnh hoặc có che khuất, hiệu suất có thể giảm sút so với các mô hình mới hơn.
- Các giai đoạn khác nhau yêu cầu ảnh có kích thước đầu vào cố định.

2.8.2 Mô hình RetinaFace

Kiến trúc này lần đầu tiên được đề xuất trong bài báo "**RetinaFace: Single-stage Dense Face Localisation in the Wild**" [16] vào năm 2019. RetinaFace là một mô hình anchor-based single-stage detector, mở rộng từ kiến trúc RetinaNet, kết hợp với feature pyramid network (FPN) và có khả năng dự đoán landmarks chính xác.

2.8.2.1 Kiến trúc Backbone

Backbone của RetinaFace dùng để trích xuất các đặc trưng từ ảnh đầu vào. Các kiến trúc thông dụng là ResNet-50 hay các biến thể nhẹ hơn như MobileNet.

Backbone được huấn luyện trước (pretrained trên ImageNet) và giữ lại các tầng convolution ban đầu, tạo ra một bộ đặc trưng đa cấp.

2.8.2.2 Kiến trúc mạng Feature Pyramid Network (FPN)

FPN được sử dụng để lấy được các đặc trưng từ nhiều tầng khác nhau của Backbone, từ đó xử lý các đối tượng (trong trường hợp này là khuôn mặt) với kích thước khác nhau.

FPN sẽ hợp nhất các đặc trưng cấp cao (đầy đủ ngữ cảnh nhưng chi tiết thấp) với các đặc trưng cấp thấp (chi tiết cao nhưng ngữ cảnh hạn chế), qua đó cung cấp một bộ đặc trưng phong phú cho các bước dự đoán phía sau.

2.8.2.3 Context Module và Extra Layers

Các context module (hoặc extra convolution layers) được thêm vào từ FPN nhằm nâng cao khả năng nhận biết các vùng khuôn mặt, đặc biệt là khi khuôn mặt ở các vị trí phức tạp hoặc trong điều kiện ánh sáng biến đổi.

2.8.2.4 Prediction Heads (Anchor-based)

RetinaFace có ba nhánh dự đoán chính từ mỗi mức của FPN, theo dạng anchor-based:

- **Class Head:** nhằm dự đoán xác suất chứa khuôn mặt cho mỗi anchor. Sử dụng Focal Loss (thường được dùng trong RetinaNet) để giải quyết vấn đề mất cân bằng giữa các anchors chứa đối tượng và không chứa đối tượng.
- **BBox Regression Head:** dự đoán các hiệu chỉnh (offsets) cho bounding box của từng anchor, giúp tinh chỉnh vị trí, kích thước dự đoán cho chính xác hơn.
- **Landmark Head:** giúp dự đoán các tọa độ của facial landmarks. Điều này giúp căn chỉnh khuôn mặt cho các bước nhận diện tiếp theo.

Mỗi prediction head được áp dụng trên nhiều cấp độ của FPN để có thể phát hiện khuôn mặt với kích thước đa dạng.

2.8.2.5 *Ưu điểm và nhược điểm*

Mạng kiến trúc này cũng có những ưu và khuyết điểm sau:

- **Ưu điểm**

- Kiến trúc RetinaFace đạt độ chính xác tốt trên các benchmark.
- Nhờ kiến trúc độc đáo FPN, mô hình có khả năng phát hiện khuôn mặt ở nhiều kích thước khác nhau.
- Độ chính xác dự đoán landmark rất chính xác.

- **Nhược điểm**

- Mô hình này yêu cầu độ tính toán khá cao, GPU mạnh để đạt được tốc độ real-time.
- Cài đặt phức tạp, đòi hỏi bộ nhớ lớn.

CHƯƠNG 03. PHƯƠNG PHÁP NGHIÊN CỨU

Sau khi chúng ta đã nghiên cứu về phần cơ sở lý thuyết, chúng ta biết được màu sắc là một đặc trưng hữu dụng trong xử lý ảnh khuôn mặt, thể hiện rõ nhất trong việc phân vùng da và phát hiện khuôn mặt, mặc dù độ hữu ích của màu sắc trong nhận dạng khuôn mặt vẫn chưa có câu trả lời tối ưu nhất định.

Vấn đề ở đây đầu tiên là việc lựa chọn không gian màu và mô hình màu da, bởi chưa có một giải pháp nào tối ưu cho vấn đề này, mà việc lựa chọn còn bị phụ thuộc vào yêu cầu của phần mềm và môi trường xung quanh. Một khi ta chọn được mô hình màu da, màu sắc có vai trò quan trọng trong phát hiện khuôn mặt, thể hiện trong quá trình tiền xử lý và lựa chọn những vùng trên ảnh có màu da. Nhờ đó, những bước lọc tiếp theo có thể được thực hiện nhằm tìm ra những khuôn mặt trong những vùng có màu da đó. Nhờ đó, các thuật toán phát hiện khuôn mặt sử dụng màu sắc sẽ hoạt động nhanh hơn rất nhiều so với các thuật toán sử dụng ảnh đen trắng, đặc biệt là với các ảnh có kích thước lớn.

Trong lĩnh vực nhận dạng khuôn mặt, việc thông tin về màu sắc có đem lại lợi ích cho việc nhận dạng hay không còn mang đến nhiều tranh cãi. Các kết quả thu được cho thấy rằng màu sắc chưa cho thấy hết toàn bộ tiềm năng và cần được nghiên cứu sâu hơn. Do đó, vào thời điểm hiện tại, có lẽ các hệ thống nhận dạng khuôn mặt chưa thể sử dụng màu sắc.

Đó cũng chính là lý do và động lực giúp nhóm chúng em thử nghiệm đề tài này. Trong bài báo cáo này, nhóm chúng em sẽ thử qua cả hai phương án nghiên cứu ban đầu lẫn cải tiến để xem xét hiệu quả sau đó, nhóm chúng em sẽ đi đến kết luận chọn phương án nghiên cứu tối ưu nhất. Phần này sẽ giúp giải đáp câu hỏi "Which state-of-the-art method will you deliver?".

3.1 Phương án nghiên cứu ban đầu

Phương án chuyển đổi không gian màu, sử dụng các không gian màu HSV và YCbCr để phát hiện màu da, và tạo một mask cho vùng da. Sau đó ta map tọa độ của mask này áp dụng lên ảnh gốc, các pixel nào nằm ngoài mask sẽ được filter về màu đen (tức giá trị pixel bằng 0), chỉ giữ lại các pixel ở ảnh gốc mà khớp với vùng mask.

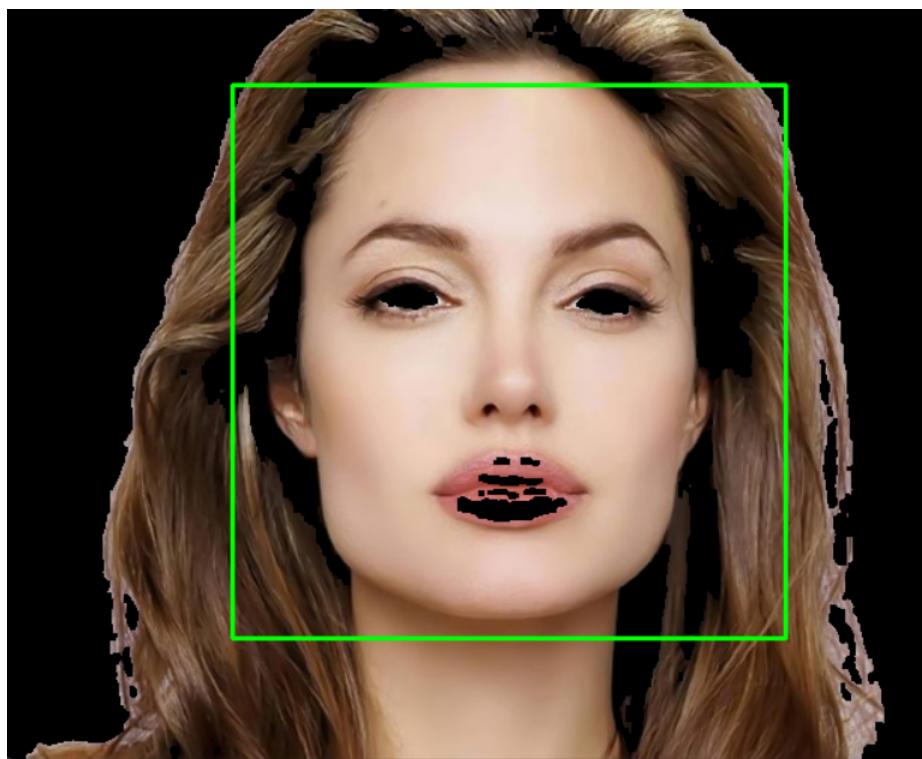
Sau đó, ta sẽ sử dụng một mô hình phát hiện khuôn mặt hiện đại để có thể phát hiện được khuôn mặt dựa trên ảnh mà ta đã áp dụng mask vào.

Việc sử dụng các phương pháp để phát hiện vùng da trước khi đưa vào các mô hình hiện đại để phát hiện khuôn mặt sẽ giúp máy sẽ loại bỏ được các vùng pixel không là

ứng cử viên có thể là khuôn mặt (bởi các vùng này không phải là các vùng da). Các mô hình phát hiện khuôn mặt của ta sẽ chỉ tập trung vào những vùng mà chúng ta đã giới hạn giúp nó, việc này sẽ giúp cho việc xác định và phát hiện khuôn mặt trở nên chính xác hơn và ít tốn chi phí hơn vì nó không tập trung quá nhiều vào những phần chi tiết không đáng có.

Tuy nhiên thì phương pháp này tồn tại cả những ưu điểm cũng như khuyết điểm, sau đây sẽ là các ưu điểm và khuyết điểm của phương pháp này mà chúng tôi nhận thấy được sau quá trình thử nghiệm:

- **Ưu điểm :** Ta có thể thấy được ngay rằng phương pháp này rất ít tốn chi phí về cả bộ nhớ lẫn thời gian để chúng ta có thể xác định và giới hạn lại phạm vi tìm kiếm. Bởi việc chỉ dùng các ngưỡng của các không gian màu như HSV và YCbCr để có thể lọc các pixel ít có khả năng là da người thì nó chỉ mất chi phí tính toán của chúng là 0(1) cho từng pixel với việc chỉ cần so sánh giá trị pixel với các ngưỡng đã đề sẵn.
- **Nhược điểm:** Nhưng chính việc đơn giản là chỉ sử dụng các ngưỡng để lọc các pixel không phải là vùng da thì nó sẽ dễ cho ra các sai sót và không tối ưu được việc giới hạn phạm vi tìm kiếm. Đây là một vấn đề mà chúng tôi gặp phải khi thử nghiệm phương pháp này.



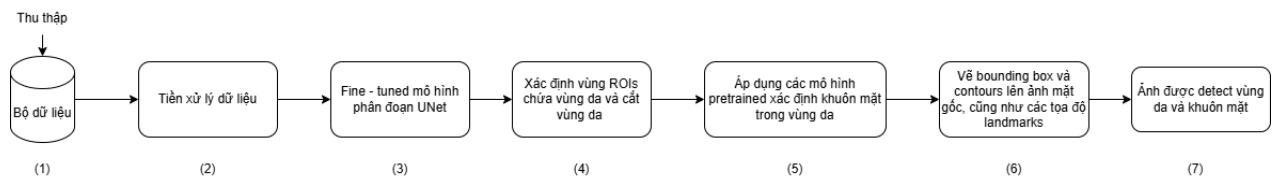
Hình 3.1 Ví dụ thử nghiệm với phương pháp chuyển không gian màu

- Ở đây ta có thể thấy là các background của ảnh, mắt, môi đều đã được thuật toán lọc

vùng da của ta filter đi bớt và chỉ còn lại màu đen thôi, nhưng mà do ngưỡng màu của da và tóc nó cùng nằm trong một ngưỡng nên là ở đây tóc của chúng ta cũng không được lọc đi bởi thuật toán chuyển đổi không gian màu này. Chính vì vậy mà ta không thể tối ưu được việc giới hạn phạm vi phát hiện khuôn mặt. Do vậy mà chúng tôi đã nghiên cứu một phương pháp cải tiến mới đó là dùng Segmentation để có thể phân đoạn vùng da người một cách tối ưu hơn mà không dính những vật thể khác vào, để có thể tối ưu được vùng tìm kiếm khuôn mặt.

3.2 Phương án nghiên cứu cải tiến

Trong phương án cải tiến này, nhóm chúng em sẽ trình bày các giai đoạn thực hiện nghiên cứu như hình 3.2.



Hình 3.2 Pipeline xử lý nghiên cứu Face Detection based on Skin Segmentation

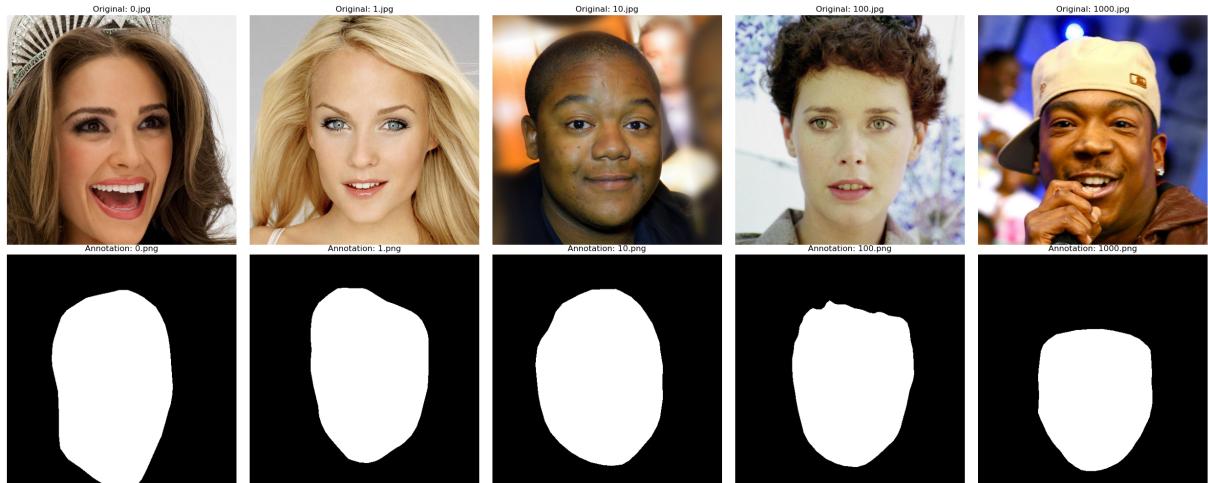
3.2.1 Thu thập dữ liệu

Về phần thu thập dữ liệu, nhóm chúng em sẽ đi khảo sát các dữ liệu được gán nhãn sẵn vùng da mặt người, trong đó nhóm đã tìm hiểu được những bộ dataset sau đây có thể khả thi cho quá trình thực hiện đồ án.

3.2.1.1 CelebAMask-HQ Dataset

CelebAMask-HQ [17] được giới thiệu lần đầu trong bài báo **MaskGAN: Towards Diverse and Interactive Facial Image Manipulation** [18] bởi các tác giả Cheng-Han Lee and Ziwei Liu và Lingyun Wu and Ping Luo vào năm 2020, là một bộ dữ liệu về mặt người theo diện rộng bao gồm 30000 ảnh mặt người với độ phân giải cao được chọn từ bộ dữ liệu CelebA theo CelebA-HQ. Mỗi ảnh có một bộ mask phân đoạn các đặc trưng khuôn mặt dựa theo CelebA. Những bộ masks của bộ dữ liệu CelebAMask-HQ đều được gán nhãn thủ công với kích thước ảnh mask là 512 x 512 và 19 lớp nhãn bao gồm tất cả các thành phần sinh học và phụ kiện có trên khuôn mặt từng người như da mặt, mũi, mắt trái, mắt phải, lông mày, tai, miệng, môi, tóc, nón, mắt kính, bông tai, dây chuyền, cổ và quần áo.

Bộ dữ liệu này được sử dụng rộng rãi trong nhiều nghiên cứu về phân đoạn khuôn mặt (face segmentation) và phân tích khuôn mặt (face parsing), được dùng để huấn luyện và đánh giá các mô hình nhận diện gương mặt và GANs cho việc tạo sinh và chỉnh sửa gương mặt.



Hình 3.3 Ảnh minh họa bộ dữ liệu CelebAMask-HQ

Tuy nhiên, do đề tài bài toán về phát hiện khuôn mặt dựa trên vùng da được phân đoạn nên nhóm chúng em chỉ sử dụng các ảnh gốc và ảnh annotations về vùng da mặt được gán nhãn là skins từ bộ dữ liệu dataset như được biểu diễn trong hình 3.4. Trong hình trên, hàng đầu tiên hiển thị toàn bộ ảnh gốc RGB có trong bộ dữ liệu CelebAMask-HQ, đồng thời từng annotation tương ứng về vùng da mặt được hiển thị ở dòng dưới.

Bộ dữ liệu này cũng tồn tại những ưu điểm và khuyết điểm nhất định như sau:

- **Ưu điểm**

- Độ phân giải và chất lượng hình ảnh của bộ dữ liệu này rất tốt, cho phép các mô hình segmentation học được rất nhiều chi tiết đặc trưng.
- Dữ liệu chứa sự đa dạng về ánh sáng, góc chụp, biểu cảm và các yếu tố khác, hỗ trợ các mô hình học được các đặc trưng tổng quát và có khả năng ứng dụng trong nhiều tình huống thực tế.
- Bộ dữ liệu được công bố rộng rãi và sử dụng làm benchmark trong nhiều nghiên cứu, tạo điều kiện so sánh hiệu năng của các mô hình khác nhau.
- **Nhược điểm:** Bộ dữ liệu này chỉ tập trung nhận diện vùng da mặt của một cá nhân nhất định chứ chưa có thể nhận diện nhiều vùng da mặt khác nhau có trong khung ảnh.

3.2.1.2 Pratheepan Dataset

Bộ dữ liệu Pratheepan được xây dựng với mục tiêu nghiên cứu việc phát hiện da người từ các hình ảnh gương mặt, được công bố lần đầu tiên trong bài báo **A Fusion Approach for Efficient Human Skin Detection** [19] vào năm 2012.

Các ảnh trong bộ dữ liệu này được tải ngẫu nhiên từ Google cho nghiên cứu phát hiện da người, bao gồm các ảnh chụp bằng nhiều loại máy ảnh khác nhau với các kỹ thuật cải thiện màu sắc và dưới các điều kiện ánh sáng khác nhau.

Bộ dữ liệu được tổ chức thành 4 thư mục chính:

- **FacePhoto:** chứa ảnh khuôn mặt của một cá nhân duy nhất, gồm 32 ảnh mẫu.
- **FamilyPhoto:** chứa ảnh nhiều đối tượng (gia đình), cấu trúc nền ảnh phức tạp hơn, bao gồm 46 ảnh mẫu.
- **GroundT_FacePhoto:** chứa ảnh groundtruth tương ứng cho thư mục FacePhoto.
- **GroundT_FamilyPhoto:** chứa ảnh groundtruth tương ứng cho thư mục FamilyPhoto.



Hình 3.4 Ảnh minh họa bộ dữ liệu Pratheepan

Bộ dữ liệu này cũng có những ưu điểm và khuyết điểm như sau:

- **Ưu điểm:** Bộ dữ liệu này không chỉ cung cấp vùng da cho từng cá nhân nhưng cung cấp vùng da cho nhiều người trong cùng một khung ảnh.
- **Nhược điểm:** Bộ dữ liệu này có số lượng ảnh hạn chế, với chỉ khoảng 78 ảnh, dẫn đến bộ dữ liệu này có thể không đủ lớn để huấn luyện các mô hình deep learning phức tạp, đặc biệt khi so sánh với các bộ dữ liệu quy mô lớn khác. Đồng thời, các annotation này được thực hiện trên vùng da toàn bộ các bộ phận cơ thể chứ không chỉ cho da mặt, nên sẽ có rất nhiều nhiễu và ảnh mask sẽ không được xử lý sạch sẽ.

3.2.1.3 Các bộ dữ liệu khác

Ngoài ra, nhóm cũng xem xét các bộ dữ liệu khác trên Kaggle như bộ Skin Tone Classification hoặc Monk Skin Tone Examples (MST-E) Dataset.

MST-E là một tập dữ liệu gồm các ví dụ của 19 người trải dài trên thang điểm MST 10 điểm. Nó chứa 1515 hình ảnh và 31 video. Mỗi người được chụp ảnh ở nhiều tư thế và điều kiện ánh sáng khác nhau và có/không có phụ kiện như mặt nạ và kính. Sau đó, Tiến sĩ Monk đã chú thích hình ảnh của những người này, cung cấp cho chúng tôi tông màu da MST thực tế. Tuy nhiên các bộ dữ liệu này chỉ tập trung về màu da trên gương mặt chứ chưa tập trung sâu sắc về vùng da người.

3.2.1.4 Kết luận

Nhóm sẽ tiếp tục khảo sát nhiều bộ dữ liệu nữa và kết hợp chúng lại để đạt được hiệu quả cao nhất đáp ứng yêu cầu bài toán trên.

3.2.2 Tiền xử lý dữ liệu

Đầu tiên, ta sẽ đọc ảnh gốc dưới định dạng RGB và ảnh mask dưới dạng grayscale với threshold. Nếu kích thước ảnh gốc không bằng với kích thước ảnh mask, thì ta sẽ resize ảnh mask theo kích thước ảnh gốc.

Sau đó, ta sẽ sử dụng albumentations để resize ảnh và mask theo kích thước tiêu chuẩn 256 x 256 cũng như áp dụng augmentation (horizontal flip) để tăng độ chính xác trong quá trình huấn luyện.

Sau đó ta sẽ trộn nhiều bộ dataset lại với nhau với kết hợp nhiều cặp (ảnh, mask) trên các folder tương ứng và nạp vào DataLoader để sử dụng trong quá trình huấn luyện và validation.

Trước khi huấn luyện, ta sẽ chia bộ dữ liệu gốc thành 3 tập khác nhau như tập huấn luyện với tỷ lệ 70%, tập validation với tỷ lệ 15% và tập testing với tỷ lệ 15%.

3.2.3 Fine-tuned mô hình phân đoạn UNet

Việc sử dụng mô hình UNet nhằm để phân đoạn vùng da trên mặt người. Mục tiêu phân đoạn nhằm để xác định vùng da mặt người một cách chính xác nhất, loại bỏ các khống gian nền nhiều hay các ánh sáng ngoại lai chiếu vào khuôn mặt làm giảm hiệu phát hiện khuôn mặt. Sau khi xác định được vùng da thì ta chỉ cần áp dụng các mô hình phát hiện gương mặt huấn luyện sẵn để xác định khuôn mặt nhằm làm tăng hiệu suất phát hiện khuôn mặt nhất có thể.

3.2.3.1 Kiến trúc mô hình

Trong kiến trúc mạng UNet này, nhóm em sẽ giữ nguyên kiến trúc UNet, cụ thể như sau:

- **Encoder Layers:** gồm 4 block, mỗi block gồm 2 lớp convolution kernel 3x3, được theo sau bởi các Batch Normalization, ReLU và một lớp Dropout 2d nhằm giảm overfitting. Sau mỗi block trừ block cuối, ta sẽ sử dụng một Max Pooling 2x2 để giảm kích thước không gian.
- **Bottleneck Layer:** là tầng nằm giữa encoder và decoder, có số lượng kênh gấp đôi so với encoder cuối cùng, nhằm giúp học các đặc trưng trừu tượng nhất.
- **Decoder Layers:** cũng gồm 4 block. Mỗi block bắt đầu bằng một lớp ConvTranspose2d (up-convolution) để tăng kích thước không gian gấp đôi, sau đó kết hợp (concatenate) kết quả với đầu ra của block encoder tương ứng (skip connection) nhằm tái sử dụng thông tin không gian chi tiết. Sau đó, áp dụng 2 lớp convolution (cùng cấu trúc như các block encoder: Conv2d, BatchNorm, ReLU, và Dropout2d sau lớp đầu tiên).
- **Output Layer:** Một lớp convolution 1x1 được sử dụng để chuyển đổi số kênh từ số features của block cuối cùng của decoder thành số kênh mong muốn của output (trong trường hợp bài toán nhị phân segmentation, out_channels=1). Sau đó, áp dụng hàm sigmoid để chuyển các giá trị output về dạng xác suất cho mỗi pixel.

3.2.3.2 Huấn luyện mô hình

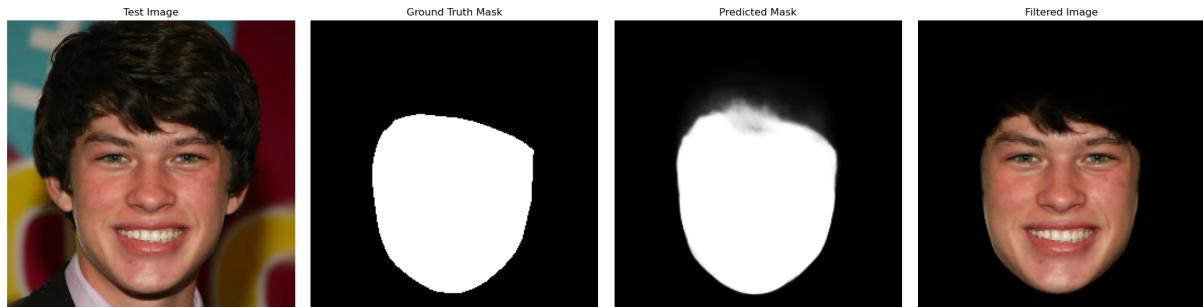
Trong quá trình huấn luyện cũng như đánh giá mô hình, chúng ta sẽ sử dụng các hàm mất mát như binary cross entropy để đo lường sự khác biệt pixel-wise cũng như Dice Loss để đo độ trùng khớp về mặt hình học của mask so với ảnh gốc. Đồng thời, ta cũng áp dụng các scheduler như ReduceLROnPlateau để giúp tự động giảm learning rate hay early stopping nhằm tránh overfitting.

3.2.3.3 Đánh giá mô hình

Sau khi huấn luyện, mô hình được đánh giá trên tập test để kiểm tra khả năng tổng quát hóa. Các metrics nhóm em dùng như sau:

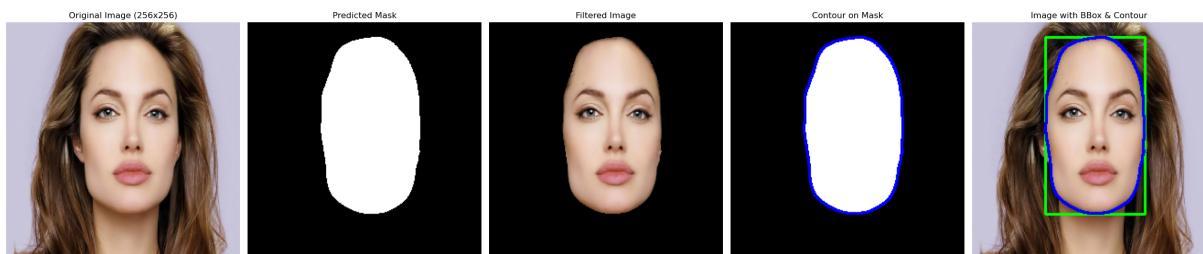
- **IoU (Intersection over Union):** xác định tỷ lệ giữa vùng giao nhau và vùng hợp nhất của dự đoán và nhãn thực tế.
- **Precision và Recall:** nhằm đánh giá độ chính xác và khả năng phát hiện đúng các vùng quan tâm.
- Các độ đo metrics khác cũng được sử dụng.

Hình 3.5 là ảnh kết quả dự đoán của mô hình sau khi được huấn luyện.



Hình 3.5 Kết quả dự đoán của mô hình

3.2.4 Xác định và hiển thị vùng ROIs da

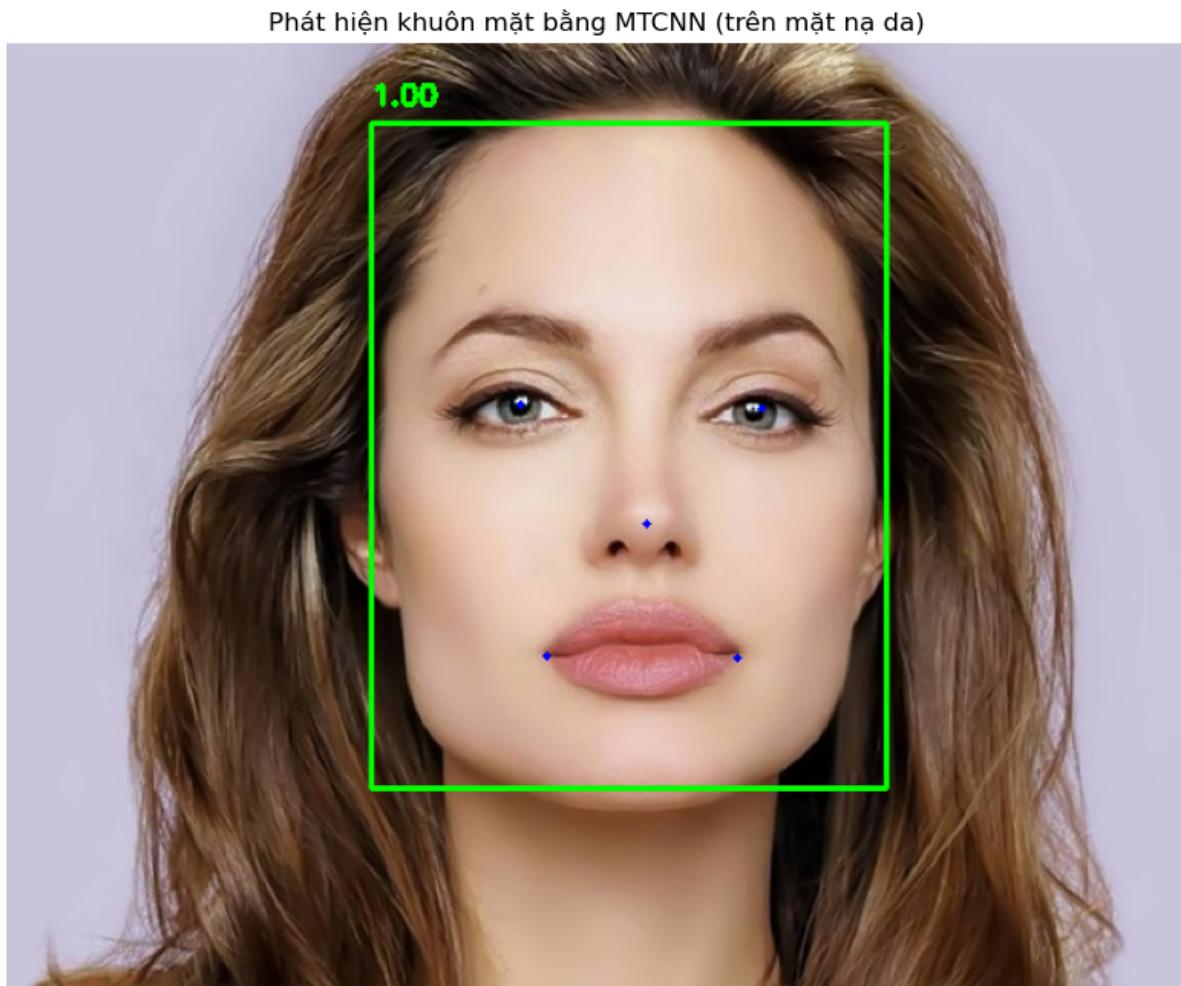


Hình 3.6 Ánh sau khi được xử lý hiển thị vùng ROIs và contours

Sau khi ta đã phân đoạn được vùng da mặt, ta sẽ xác định vùng tọa độ ROIs vùng da trên ảnh mask và áp dụng lên ảnh gốc thông qua vùng liên thông, sau đó ta sẽ hiển thị lên ảnh gốc. Mục đích của bước này nhằm xác định hết tất cả các vùng da mặt trên ảnh nhằm loại bỏ đi các ảnh nền thừa và nhiễu bởi các chi tiết giống khuôn mặt.

Hình 3.6 là ảnh kết quả của bước này. Trong đó, ảnh ban đầu là ảnh gốc, ảnh tiếp theo là ảnh mask được dự đoán bởi mô hình, ảnh thứ ba là ảnh sau khi cắt bỏ vùng da mặt, ảnh thứ 4 là ảnh sau khi vẽ contours, và ảnh cuối cùng là ảnh kết quả của quá trình trên.

3.2.5 Áp dụng các mô hình pretrained phát hiện gương mặt



Hình 3.7 Kết quả cuối cùng của pipeline

Sau đó, cuối cùng ta sẽ dùng các mô hình phát hiện gương mặt trên vùng ROIs đã được xác định bước trên. Ảnh kết quả cuối cùng sẽ hiển thị confidence của mô hình sau khi detect gương mặt, các điểm landmarks lên ảnh gốc.

Việc áp dụng các mô hình pretrained hiện nay vẫn được nhóm em nghiên cứu và đưa ra hướng giải quyết, trong đó có hai mô hình phát hiện gương mặt được nhóm em quan tâm là RetinaFace và MTCNN, cả hai kiến trúc này đều đã được trình bày trong các phần trước đó.

3.3 Kết luận

Sau khi bàn bạc và thảo luận, nhóm chúng em quyết định dùng phương án cải tiến với mong phương án này có thể đạt được hiệu quả cao trong quá trình nghiên cứu.

CHƯƠNG 04. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Trong chương này, nhóm chúng em sẽ trình bày việc trình bày chương trình chạy kiểm thử như nào trong quá trình vấn đáp, cũng như phần này sẽ giải đáp câu hỏi "How will you conduct the demonstration?"

Nhóm chúng em sẽ xây dựng một ứng dụng đơn giản mang tên **Face Detection based on Skin Segmentation**.

Đầu tiên chương trình sẽ cho phép người dùng bấm nút quay video bản thân và chương trình sẽ hiện ra cửa sổ camera để quay hình ảnh nhận được.

Sau đó, chương trình sẽ qua các giai đoạn xử lý và trả ra kết quả cho người dùng tương ứng với đoạn video được segmentation gương mặt và được detect gương mặt với bounding box và contours.

Đồng thời, nhóm chúng em cũng sẽ hiện các bản đánh giá kết quả quá trình mô hình cũng như các số liệu thống kê chi tiết trong quá trình huấn luyện cũng như kiểm thử.



Hình 4.1 Ứng dụng phát hiện gương mặt

KẾT LUẬN VÀ ĐỀ NGHỊ

Kết luận chung

Sau một thời gian tìm hiểu và nghiên cứu đề tài này, nhóm em đạt được một số kết quả sau đây:

- Tìm hiểu được các hướng tiếp cận dùng màu sắc để phát hiện và nhận dạng gương mặt.
- Tìm hiểu về các phương pháp chuyển đổi kênh màu trong bài toán phát hiện gương mặt.
- Áp dụng một số kỹ thuật xử lý ảnh để xây dựng một chương trình thử nghiệm phát hiện khuôn mặt dựa vào màu da.
- Tìm hiểu các mô hình phân đoạn hình ảnh theo ngữ nghĩa hiện đại ngày nay và các mô hình phát hiện gương mặt kết hợp học sâu.

Tuy nhiên, vẫn còn tồn tại một số bất cập sau đây:

- Chương trình vẫn chưa xử lý được phân đoạn nhiều khuôn mặt trong cùng một khung ảnh.
- Dữ liệu vẫn còn thiếu và chưa thể gán nhãn label cho từng ảnh cụ thể.
- Chương trình chạy chậm do phải dùng nhiều model xử lý.
- Vẫn còn phải sử dụng nhiều model hiện đại nên dung lượng để lưu trữ dữ liệu cũng như GPUs cho quá trình huấn luyện còn hạn chế.

Hướng phát triển

Nhóm chúng em sẽ cố gắng thực hiện nhiều thực nghiệm trên nhiều bộ dataset khác nhau để chọn ra bộ dataset tối ưu nhất.

Thực nghiệm trên nhiều mô hình tốt hơn để đưa ra sự lựa chọn xác đáng nhất.

Nhóm cũng đã nghĩ đến hướng sau khi phát hiện gương mặt thì nhóm có thể làm thêm tác vụ để phát hiện màu da dựa trên khuôn mặt đã phát hiện. Bởi lẽ, sau khi xác định gương mặt thì việc phát hiện màu da cũng rất quan trọng trong các ứng dụng phân tích gương mặt.

Kiến nghị và đề xuất

Nhóm em cũng chưa chắc chắn hướng đi của nhóm sẽ là chính xác tuyệt đối cho bài toán phát hiện gương mặt dựa trên vùng da. Mong các thầy có thể đưa ra những lời khuyên bổ ích cho nhóm em.

TÀI LIỆU THAM KHẢO

- [1] B. Funt, K. Barnard, and L. Martin, “Is machine colour constancy good enough?” in *Computer Vision — ECCV’98*, H. Burkhardt and B. Neumann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 445–459.
- [2] A. K. Jain and S. Z. Li, *Handbook of Face Recognition*. Berlin, Heidelberg: Springer-Verlag, 2005.
- [3] J. B. Martinkauppi, A. Hadid, and M. Pietikäinen, *Skin Color in Face Analysis*. London: Springer London, 2011, pp. 223–249. [Online]. Available: https://doi.org/10.1007/978-0-85729-932-1_9
- [4] R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain, “Face detection in color images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, 2002.
- [5] C. Garcia and G. Tziritas, “Face detection using quantized skin color regions merging and wavelet packet analysis,” *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 264–277, 1999.
- [6] K. Sobottka and I. Pitas, “Face localization and facial feature extraction based on shape and color information,” in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 3, 1996, pp. 483–486 vol.3.
- [7] H. Wu, Q. Chen, and M. Yachida, “Face detection from color images using a fuzzy pattern matching method,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 557–563, 1999.
- [8] A. Hadid, M. Pietikainen, and B. Martinkauppi, “Color-based face detection using skin locus model and hierarchical filtering,” in *2002 International Conference on Pattern Recognition*, vol. 4, 2002, pp. 196–200 vol.4.
- [9] A. Hadid and M. Pietikainen, “A hybrid approach to face detection under unconstrained environments,” in *18th International Conference on Pattern Recognition (ICPR’06)*, vol. 1, 2006, pp. 227–230.
- [10] R. Kemp, G. Pike, P. White, and A. Musselman, “Perception and recognition of normal and negative faces: The role of shape from shading and pigmentation cues,” *Perception*, vol. 25, no. 1, pp. 37–52, 1996, pMID: 8861169. [Online]. Available: <https://doi.org/10.1068/p250037>

- [11] A. W. Yip and P. Sinha, “Contribution of color to face recognition,” *Perception*, vol. 31, no. 8, pp. 995–1003, 2002, pMID: 12269592. [Online]. Available: <https://doi.org/10.1068/p3376>
- [12] L. Torres, J. Reutter, and L. Lorente, “The importance of the color information in face recognition,” in *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, vol. 3, 1999, pp. 627–631 vol.3.
- [13] C. F. Jones and A. L. Abbott, “Optimization of color conversion for face recognition,” *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 4, p. 948790, 2004. [Online]. Available: <https://doi.org/10.1155/S1110865704401073>
- [14] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015. [Online]. Available: <https://arxiv.org/abs/1505.04597>
- [15] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” *IEEE Signal Processing Letters*, vol. 23, no. 10, p. 1499–1503, Oct. 2016. [Online]. Available: <http://dx.doi.org/10.1109/LSP.2016.2603342>
- [16] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, “Retinaface: Single-stage dense face localisation in the wild,” 2019. [Online]. Available: <https://arxiv.org/abs/1905.00641>
- [17] C. M. Laboratory. (2019) Celebamask-hq: A benchmark for high-quality face segmentation. Accessed: April 12, 2025. [Online]. Available: https://mmlab.ie.cuhk.edu.hk/projects/CelebA/CelebAMask_HQ.html
- [18] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, “Maskgan: Towards diverse and interactive facial image manipulation,” 2020. [Online]. Available: <https://arxiv.org/abs/1907.11922>
- [19] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, “A fusion approach for efficient human skin detection,” *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 138–147, 2012.