BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM

NGUYỄN ĐỨC DŨNG

TRÍCH CHỌN ĐẶC TRƯNG KẾT CẤU MÀU CỤC BỘ CHO BÀI TOÁN NHẬN DẠNG ẢNH MÀU MẶT NGƯỜI

LUẬN VĂN THẠC SĨ

Chuyên ngành: Công nghệ thông tin

Mã số ngành: 60480201

TP. HỒ CHÍ MINH, tháng 8 năm 2015

BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM

NGUYỄN ĐỨC DŨNG

TRÍCH CHỌN ĐẶC TRƯNG KẾT CẤU MÀU CỤC BỘ CHO BÀI TOÁN NHẬN DẠNG ẢNH MÀU MẶT NGƯỜI

LUẬN VĂN THẠC SĨ

Chuyên ngành: Công nghệ thông tin

Mã số ngành: 60480201

CÁN BỘ HƯỚNG DẪN KHOA HỌC: TS. ĐẶNG THANH DŨNG

TP. HỒ CHÍ MINH, tháng 8 năm 2015

CÔNG TRÌNH ĐƯỢC HOÀN THÀNH TẠI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM

Cán bộ hướng dẫn khoa học: TS. ĐẶNG THANH DŨNG

Luận văn Thạc sĩ được bảo vệ tại Trường Đại học Công nghệ TP. HCM ngày … tháng … năm …

Thành phần Hội đồng đánh giá Luận văn Thạc sĩ gồm:

TT	Họ và tên	Chức danh Hội đồng
1	PSG. TSKH Nguyễn Xuân Huy	Chủ tịch
2	PGS. TS Đỗ Phúc	Phản biện 1
3	TS. Võ Đình Bảy	Phản biện 2
4	TS. Trần Đức Khánh	Ủy viên
5	TS. Cao Tùng Anh	Ủy viên, Thư ký

Xác nhận của Chủ tịch Hội đồng đánh giá Luận sau khi Luận văn đã được sửa chữa (nếu có).

Chủ tịch Hội đồng đánh giá LV

TRƯỜNG ĐH CÔNG NGHỆ TP. HCM
PHÒNG QLKH – ĐTSĐH

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc

TP. HCM, ngày..... tháng..... năm 20.....

NHIỆM VỤ LUẬN VĂN THẠC SĨ

Họ tên học viên: Nguyễn Đức Dũng — Giới tính: Nam

Ngày, tháng, năm sinh: 07/05/1989 — Nơi sinh: TP.HCM

Chuyên ngành: Công nghệ thông tin — MSHV: 1341860004

I- Tên đề tài:

Trích chọn đặc trưng kết cấu màu cục bộ cho bài toán nhận diện ảnh màu mặt người

.....

.....

II- Nhiệm vụ và nội dung:

Nhiệm vụ tổng quát của đề tài là: Đề xuất được phương pháp trích chọn đặc trưng kết cấu màu cục bộ cho bài toán nhận dạng ảnh màu mặt người, trên cơ sở cải tiến phương pháp của Choi và các đồng sự, sao cho tăng độ chính xác nhận dạng (so với phương pháp của Choi và các đồng sự).

III- Ngày giao nhiệm vụ: *18/08/2014*

IV- Ngày hoàn thành nhiệm vụ: *15/6/2015*

V- Cán bộ hướng dẫn: TS. Đặng Thanh Dũng

.....

.....

CÁN BỘ HƯỚNG DẪN
(Họ tên và chữ ký)

KHOA QUẢN LÝ CHUYÊN NGÀNH
(Họ tên và chữ ký)

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu, kết quả nêu trong Luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện Luận văn này đã được cảm ơn và các thông tin trích dẫn trong Luận văn đã được chỉ rõ nguồn gốc.

Học viên thực hiện Luận văn

Nguyễn Đức Dũng

LỜI CÁM ƠN

Để hoàn thành luận văn này, tác giả xin tỏ lòng biết ơn sâu sắc nhất đến Thầy TS. Đặng Thanh Dũng, người đã tận tình hướng dẫn trong suốt quá trình viết luận văn.

Đồng thời, tác giả cũng xin chân thành cảm ơn gia đình, bạn bè và các anh chị em trong Công ty CP Giải pháp CNTT Tân Cảng đã giúp đỡ và tạo điều kiện thuận lợi để tác giả có thể hoàn thành luận văn thạc sỹ này.

Tác giả cũng xin chân thành cảm ơn quý Thầy, Cô trong khoa Công nghệ thông tin, trường Đại học Công Nghệ TP. Hồ Chí Minh đã tận tình truyền đạt kiến thức trong những năm tháng học tập. Với vốn kiến thức được tiếp thu trong quá trình học không chỉ là nền tảng cho quá trình nghiên cứu mà còn là hành trang quí báu để tác giả có thể bước vào đời một cách vững chắc và tự tin.

Cuối cùng, tác giả xin kính chúc quý Thầy, Cô dồi dào sức khỏe và thành công trong sự nghiệp, đạt được nhiều thành công tốt đẹp trong công việc và cuộc sống.

Học viên thực hiện Luận văn

Nguyễn Đức Dũng

TÓM TẮT

Trong bài cáo cáo *"Color Local Texture Features for Color Face Recognition"* của *Choi và các đồng sự* năm 2012, nhóm tác giả Choi đã đề xuất phương pháp kết hợp các đặc trưng kết cấu cục bộ LBP trích từ các kênh màu khác nhau của ảnh màu mặt người (được biểu diễn trên một không gian màu, chẳng hạn RQC_r) để làm đặc tả ảnh màu mặt người (color face descriptor). Đặc tả này được nhóm tác giả gọi là đặc trưng kết cấu màu cục bộ (color local texture feature – CLTF). Kết quả các thí nghiệm được tiến hành trong bài báo cáo trên chứng tỏ rằng CLTF giúp tăng độ chính xác nhận dạng.

Tuy nhiên, nhóm tác giả cũng chỉ ra trong phần kết luận một số hạn chế trong phương pháp xây dựng CLTF được nêu trong công trình này. Một trong những hạn chế đó là, khi xây dựng CLTF, các đặc trưng kết cấu cục bộ (local texture feature, LTF) thành phần được trích chọn từ cùng một không gian màu gồm 3 thành phần (ví dụ *RQC*r). Xử lý theo cách này sẽ không tận dụng được điểm mạnh của các không gian màu khác nhau. Mỗi không gian màu có các đặc điểm riêng có thể được tận dụng để làm tăng độ chính xác nhận dạng.

Trong nghiên cứu này, tác giả sẽ xây dựng đặc trưng kết cấu màu cục bộ (CLTF), trong đó các thành phần đặc trưng kết cấu cục bộ được chọn lọc từ các không gian màu khác nhau một cách phù hợp, sao cho chứa được nhiều thông tin phân loại (discriminant information) nhất. Tác giả sẽ xây dựng các không gian màu bằng cách kết hợp các thành phần màu từ các không gian màu cố định và sau đó thử nghiệm độ chính xác của các không gian màu kết hợp trên cùng một bộ dữ liệu hình ảnh. Ở đây tác giả sử dụng bộ dữ liệu Color Feret và so sánh các kết quả thực nghiệm giữa phương pháp của nhóm tác giả Choi và phương pháp cải tiến nhằm nâng cao hiệu suất nhận diện khuôn mặt màu một cách tốt nhất.

ABSTRACT

In the paper "*Color Local Texture Features for Color Face Recognition*" by *Choi et al* 2012, Choi et al have proposed a new color local texture feature was Color Local Binary Pattern (CLBP) by using LBP extract from image face's different color channels (represented on a color space, such as RQCr) to be a color face descriptor. The results of experiments conducted in paper prove that CLTF help improve the accuracy of face regconition.

However, Choi et al' authors also pointed out in the conclusion some limitations of this methods outlined in this work. One of them is, when constructing CLTF, local texture features – LTF selected components are extracted from the same color space consists of 3 components (eg RQCr). Handling it by this way will not take the strengths of the different color spaces. Each color space has its own characteristics that can be utilized to increase the accuracy of identification.

In this study, the author will construct specific Color local texture feature (CLTF), in which components featured local structures are selected from different color spaces appropriately, so that they contain more discriminant information as much as possible. Authors will build the color space by combining color components from the fixed color space and then test the accuracy of the color space combined on the same dataset of image. Here, author uses dataset ColorFeret and compare experimental results between the method of Choi và các đồng sự and innovative methods to enhance performance color face regconition.

MỤC LỤC

CHƯƠNG 1: TỔNG QUAN 1
1.1 Giới thiệu 1
1.1.1 Đặt vấn đề 1
1.1.2 Tính cấp thiết của đề tài 2
1.2 Mục tiêu của đề tài 3
1.3 Nội dung nghiên cứu 4
CHƯƠNG 2: CÁC KIẾN THỨC NỀN TẢNG VỀ NHẬN DIỆN ẢNH MẶT NGƯỜI 5
2.1 Các không gian màu 5
2.1.1 Không gian màu RGB 5
2.1.2 Không gian màu CMYK 7
2.1.3 Không gian màu HSV 7
2.1.4 Các không gian màu CIE 8
2.2 Bài toán nhận dạng ảnh mặt người 10
2.3.1. Các độ đo khoảng cách giữa các đối tượng 13
2.3.2. Độ đo khoảng cách giữa các dãy 14
2.3.3. Độ đo theo lý thuyết thông tin 18
CHƯƠNG 3: KỸ THUẬT TRÍCH CHỌN KẾT CẤU ĐẶC TRƯNG MÀU CỤC BỘ 22
3.1 Mẫu nhị phân cục bộ (Local Binary Pattern – LBP) 22
3.1.1 Thuật toán LBP 22
3.1.2 Thuật toán Opponent color LBP (OCLBP) 23
3.1.3 Mẫu nhị phân đồng dạng – Uniform Pattern 24
3.2 Các bước xử lý trong phương pháp trích trọn đặc trưng 25
3.2.1 Mô hình trích chọn đặc trưng của Choi và các đồng sự 25
3.2.2 Trích chọn đặc trưng LBP màu (CLBP) 26
3.2.5 Kết hợp 29
CHƯƠNG 4: MÔ HÌNH ĐỀ XUẤT VÀ THỰC NGHIỆM 32
4.1 Mô hình đề xuất 32
4.2. Các không gian màu lựa chọn. 34

4.2.1 Không gian màu *YIQ*.....34
4.2.2 Không gian màu *ZRG* 35
4.2.3 Không gian màu RC_rQ 35
4.2.4 Kết hợp các không gian màu.....35
4.3 Cơ sở dữ liệu ảnh màu Color Feret 36
4.3.1 Giới thiệu.....36
4.3.2 Quy tắc đặt tên hình ảnh 37
4.3.2 Ground Truth File 38
4.4 Cơ sở lý thuyết PCA 39
4.4.1 Giới thiệu.....39
4.1.2 Thuật toán PCA.....41
4.5 Khoảng cách Mahalanobis 43
4.5.1 Giới thiệu.....43
4.5.2 Định nghĩa và tính chất 44
4.6 Mô tả thí nghiệm 44
4.7 Kết quả thực nghiệm 47
4.8 Đánh giá thực nghiệm 47
CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....48
5.1 Kết luận 48
5.2 Thách thức và hướng phát triển 48
TÀI LIỆU THAM KHẢO.....50

DANH MỤC CÁC TỪ VIẾT TẮT

STT	Từ viết tắt	Từ gốc	Nghĩa tiếng việt
1	FR	Face Recognitions	Nhận diện khuôn mặt
2	LBP	Local Binary Pattern	Mẫu nhị phân cục bộ
3	CLTF	Color Local Texture Feature	Đặc trưng kết cấu màu cục bộ
4	CLBP	Color Local Binary Pattern	Mẫu nhị phân cục bộ màu
5	CLGW	Color Local Gabor Wavelet	Gabor Wavelet cục bộ màu
6	LTF	Local Texture Feature	Đặc trưng kết cấu cục bộ
7	HCI	Human–computer interaction	Tương tác người - máy
8	PIN	Personal identification number	Mã số định danh cá nhân
9	ATM	Automated Teller Machine	Máy rút tiền tự động
10	CIE	Commission Internationale de l'Eclairage	Ủy ban quốc tế về chiếu sáng
11	PCA	Principal component Analysis	Phương pháp phân tích thành phần chính
12	LDA	Linear Discriminant Analysis	phương pháp phân tích sự khác biệt tuyến tính
13	EBGM	Elastic Bunch Graph Matching	Phương pháp đối sánh bó đồ thị co giãn
14	OCLBP	Opponent Color Local Binary Pattern	Mẫu nhị phân màu đối lập cục bộ
15	NTSC	National Teltevision System Committee	Ủy ban quốc gia hệ thống truyền hình
16	NIST	National Institute of Standards and Technology	Viện tiêu chuẩn và công nghệ quốc gia Hoa Kỳ
17	NN	Nearest Neighbor	Phương pháp láng giềng gần nhất
18	PAL	Phase Alternative Line	Đảo pha theo từng dòng
19	CSDL		Cơ sở dữ liệu

DANH MỤC CÁC BẢNG

Bảng 2.1 Bảng tham chiếu các độ đo khoảng cách giữa các đối tượng.....13

Bảng 4.1 Bảng số lượng ảnh trong tập Color Feret 37

DANH MỤC CÁC BIỂU ĐỒ, ĐỒ THỊ, SƠ ĐỒ, HÌNH ẢNH

Hình 2.1 Không gian màu RGB 6
Hình 2.2 Không gian màu CYMK 7
Hình 2.3 Không gian màu HSV 8
Hình 2.4 Hệ thống quan sát CIE 9
Hình 2.5 Phạm vi quan sát trong mô hình của CIE 10
Hình 2. 6 Minh họa của khoảng cách Hausdorff giữa 2 tập hợp A và B: 19
Hình 3.1 Mô tả thuật toán LBP gốc 22
Hình 3.2 Số lượng điểm P cách đều trên các đường tròn với bán kính khác nhau ... 22
Hình 3.3 Các mẫu nhị phân đồng dạng và không dồng dạng. 24
Hình 3.4 Mô hình nhận diện khuôn mặt màu sử dụng đặc trưng kết cấu màu cục bộ. 25
Hình 3.6: Hình minh họa cách sử dụng thuật toán LBP Unichrome 28
Hình 3.7 Các histogram được trích xuất từ thuật toán Opponent LBP 29
Hình 4.1 Mô hình đề xuất cải tiến 33
Hình 4. 2 Ví dụ mô tả về số lượng chiều của dữ liệu. 40
Hình 4. 3 Minh họa sự biến thiên dữ liệu trong không gian đa chiều trước và sau khi sử dụng thuật toán PCA 41
Hình 5.1: Mô hình thực nghiệm 46
Hình 5.2: Đồ thị mô tả độ chính xác trên các không gian màu 47

CHƯƠNG 1: TỔNG QUAN

1.1 Giới thiệu

1.1.1 Đặt vấn đề

Nhận dạng ảnh mặt người (Face Recognition – FR) có rất nhiều ứng dụng thực tế, như Giám sát qua camera (Bowyer, 2004), Nhận dạng bằng đặc điểm sinh trắc (Jain et al., 2004), Đánh chỉ mục nội dung multimedia,…; vì vậy, trong những năm gần đây đã thu hút được nhiều quan tâm trong các nghiên cứu về thị giác máy tính (computer vision) và nhận dạng mẫu (pattern regconition).

Trích chọn đặc trưng (feature extraction) ảnh mặt người là một bước xử lý rất quan trọng trong toàn bộ quá trình xử lý tính toán cho FR. Gần đây, đặc trưng kết cấu cục bộ (*local texture feature*) được xem là một đặc tả mặt (*face descriptor*) tốt bởi nó ít thay đổi trong điều kiện có thay đổi về biểu hiện cảm xúc trên mặt, về độ nghiêng gương mặt được chụp, về các phần khuất trên gương mặt (do xõa tóc, đội nón, đeo kính)(Choi et al., 2010), v..v.. Đặc biệt, đặc trưng mẫu nhị phân cục bộ (*Local binary pattern texture feature – LBP texture feature*) đã được chứng minh là một đặc tả mặt rất tốt cho FR (Ahonen et al., 2006).

Trong (Choi et al., 2012), nhóm tác giả đã đề xuất phương pháp kết hợp các đặc trưng kết cấu cục bộ LBP trích từ các kênh màu khác nhau của ảnh màu mặt người (được biểu diễn trên một không gian màu nào đó, chẳng hạn RGB) để làm đặc tả mặt ảnh màu mặt người (*color face descriptor*). Đặc tả này được nhóm tác giả của (Choi et al., 2012) gọi là đặc trưng kết cấu màu vục bộ (*color local texture feature* – CLTF). Kết quả các thí nghiệm được tiến hành trong (Choi et al., 2012) chứng tỏ rằng CLTF giúp tăng độ chính xác nhận dạng.

Tuy nhiên, nhóm tác giả của Choi cũng chỉ ra (trong phần kết luận) một số hạn chế trong phương pháp xây dựng CLTF được nêu trong công trình này. Một trong những hạn chế đó là, khi xây dựng CLTF, các *đặc trưng kết cấu cục bộ* (local texture feature, *LTF*) thành phần được trích chọn từ cùng một không gian màu gồm 3 thành phần (ví dụ RQC_r). Xử lý theo cách này sẽ không tận dụng được điểm mạnh

của các không gian màu khác nhau. Theo (Stokman and Gevers, 2007), mỗi không gian màu có các đặc điểm riêng có thể được tận dụng để làm tăng độ chính xác nhận dạng.

Vì vậy, trong nghiên cứu này, tác giả sẽ tìm cách xây dựng CLTF, trong đó các LTF thành phần được chọn lọc từ các không gian màu khác nhau một cách phù hợp, sao cho chứa được nhiều thông tin phân loại (discriminant information) nhất. Điều này có thể đạt được bằng cách sử dụng giải pháp được mô tả trong (Choi et al., 2011) (multiclass boosting color feature selection).

1.1.2 Tính cấp thiết của đề tài

Nhận dạng ảnh mặt người (FR) có rất nhiều ứng dụng thực tế hiện nay, chẳng hạn như: Giám sát qua camera (Bowyer, 2004), Nhận dạng bằng đặc điểm sinh trắc (Jain et al., 2004), Đánh chỉ mục nội dung multimedia, v..v. Ngoài ra, các hệ thống tương tác người máy (HCI) thông minh cũng rất cần đến kết quả của bài toán nhận dạng mặt người để có thể hỗ trợ người dùng bảo quản được các thông tin nhạy cảm mà không cần phải lo lắng về việc đánh mất hoặc mất cắp thông tin xác thực (như số PIN để truy cập ATM, password để truy cập máy tính và các hệ thống trên mạng Internet).

Ưu điểm của việc sử dụng FR trong an ninh là nó cho phép xác thực người dùng mà không cần sự hợp tác của đối tượng. So với các phương pháp xác thực dựa trên sinh trắc học khác như: Nhận dạng dấu vân tay, tròng mắt, v.v… (mặc dù rất đáng tin cậy và chính xác, nhưng lại đòi hỏi sự hợp tác của người dùng), đây là một lợi thế lớn. Vì vậy, trong khoảng 10 năm trở lại đây, bài toán nhận dạng mặt người thu hút nhiều sự quan tâm của các nhà nghiên cứu trong các lĩnh vực nhận dạng, xử lý ảnh, thị giác máy tính. Việc tìm ra các phương pháp làm tăng độ chính xác nhận dạng mặt không những có giá trị khoa học, mà còn có giá trị thực tiễn rất lớn.

Một cách tiếp cận đối với việc tăng độ chính xác nhận dạng mặt là tối ưu hóa quá trình trích chọn vector đặc trưng ảnh mặt người, sao cho vector này chứa nhiều thông tin phân loại nhất (discriminant information).

Theo cách tiếp cận này, tác giả đặt mục tiêu nghiên cứu trong luận văn thạc sỹ là tìm cách xây dựng đặc trưng ảnh màu mặt người, bằng cách cải biên phương pháp do Choi và các đồng sự đưa ra vào năm 2012. Để đạt được mục tiêu này, luận văn sẽ phải giải quyết các câu hỏi sau đây:

- Tìm hiểu đặc điểm của các không gian màu quan trọng (chẳng hạn *RGB, ZRG, RQC*$_r$) trong bài toán nhận dạng ảnh màu mặt người.

- Chọn lọc các kênh màu (từ các không gian màu thích hợp) sẽ được sử dụng cho việc rút trích đặc trưng.

- Tìm cách kết hợp các đặc trưng từ các kênh màu nêu trên.

Các giả thuyết được dùng trong luận văn:

- Đặc trưng kết cấu màu cục bộ chứa nhiều thông tin nhận dạng hơn đặc trưng kết cấu toàn cục (Ahonen et al., 2006).

- Đặc trưng ảnh màu chứa thông tin kết cấu (texture information) từ các kênh màu khác nhau chứa nhiều thông tin nhận dạng hơn đặc trưng ảnh chứa thông tin kết cấu từ một kênh màu duy nhất (Choi et al., 2012).

- Các thông tin nhận dạng được rút trích từ các không gian màu khác nhau sẽ có ảnh hưởng khác nhau đến kết quả nhận dạng (Stokman and Gevers, 2007).

1.2 Mục tiêu của đề tài

- Mục tiêu tổng quát của đề tài là: Đề xuất được phương pháp trích chọn đặc trưng kết cấu màu cục bộ cho bài toán nhận dạng ảnh màu mặt người, trên cơ sở cải tiến phương pháp của Choi và các đồng sự, sao cho tăng độ chính xác nhận dạng (so với phương pháp của Choi và các đồng sự).

- Mục tiêu cụ thể của đề tài bao gồm:

1. Chọn các không gian màu phù hợp.

2. Chọn các kênh màu phù hợp trong từng không gian màu trên.

3. Trích chọn đặc trưng kết cấu màu cục bộ cho từng kênh màu nêu trong (2).

4. Tìm cách kết hợp các đặc trưng kết cấu cục bộ để cho ra được Đặc trưng kết cấu màu cục bộ cho ảnh màu mặt người.

5. Cài đặt thử nghiệm và tiến hành thí nghiệm, đánh giá.

1.3 Nội dung nghiên cứu

1. Cách biểu diễn màu và các không gian màu.

2. Khái quát về trích chọn đặc trưng ảnh và nhận dạng dựa trên đặc trưng ảnh.

3. Thuật toán LBP và Local Ternary Pattern (LTP) (Tan and Triggs, 2007).

4. Tìm hiểu phương pháp rút trích đặc trưng ảnh đề xuất bởi Choi và các đồng sự năm 2012.

5. Tìm hiểu phương pháp chọn lựa kênh màu được trình bày trong (Choi et al., 2011).

6. Đề xuất phương pháp rút trích đặc trưng kết cấu màu cục bộ.

7. Tìm hiểu một số Tập ảnh màu mặt người miễn phí dùng cho mục đích đánh giá các phương pháp nhận dạng ảnh màu mặt người.

8. Thực nghiệm và đánh giá phương pháp đề xuất.

CHƯƠNG 2: CÁC KIẾN THỨC NỀN TẢNG VỀ NHẬN DIỆN ẢNH MẶT NGƯỜI

Xác định khuôn mặt người (Face Detection) là một kỹ thuật máy tính để xác định các vị trí và các kích thước của các khuôn mặt người trong các ảnh bất kỳ (ảnh kỹ thuật số). Kỹ thuật này nhận biết các đặc trưng của khuôn mặt và bỏ qua những thứ khác, như: tòa nhà, cây cối, cơ thể.

2.1 Các không gian màu

Không gian màu là một mô hình toán học dùng để mô tả các màu sắc trong thực tế được biểu diễn dưới dạng số học. Trên thực tế có rất nhiều không gian màu khác nhau được mô hình để sử dụng vào những mục đích khác nhau.

2.1.1 Không gian màu RGB

RGB là không gian màu rất phổ biến được dùng trong đồ họa máy tính và nhiều thiết bị kĩ thuật số ("RGB color model," 2015). Ý tưởng chính của không gian màu này là sự kết hợp của 3 màu sắc cơ bản: màu đỏ (Red – R), xanh lục (Green – G) và xanh lơ (Blue – B) để mô tả tất cả các màu sắc khác.

Một trong những ứng dụng phổ biến nhất của mô hình màu RGB là việc hiển thị màu sắc trong các ống tia âm cực, màn hình tinh thể lỏng hay màn hình plasma, chẳng hạn như màn hình máy tính hay ti vi. Mỗi điểm ảnh trên màn hình có thể được thể hiện trong bộ nhớ máy tính như là các giá trị độc lập của màu đỏ, xanh lá cây và xanh lam. Các giá trị này được chuyển đổi thành các cường độ và gửi tới màn hình. Bằng việc sử dụng các tổ hợp thích hợp của các cường độ ánh sáng đỏ, xanh lá cây và xanh lam, màn hình có thể tái tạo lại phần lớn các màu trong khoảng đen và trắng. Các phần cứng hiển thị điển hình được sử dụng cho các màn hình máy tính trong năm 2003 sử dụng tổng cộng 24 bit thông tin cho mỗi điểm ảnh (trong tiếng Anh thông thường được biết đến như *bits per pixel* hay *bpp*). Nó tương ứng với mỗi 8 bit cho màu đỏ, xanh lá cây và xanh lam, tạo thành một tổ hợp 256 các giá trị có thể, hay 256 mức cường độ cho mỗi màu. Với hệ thống như thế, khoảng 16,7 triệu màu rời rạc có thể tái tạo.

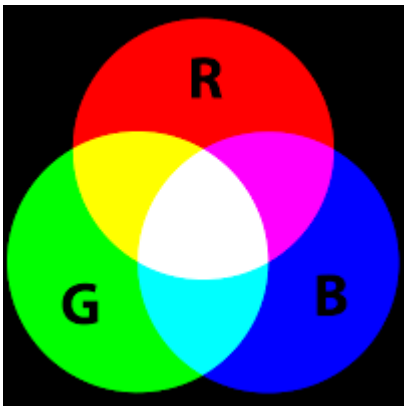
Khi biểu diễn dưới dạng số, các giá trị RGB trong mô hình 24 bpp thông thường được ghi bằng cặp ba số nguyên giữa 0 và 255, mỗi số đại diện cho cường độ của màu đỏ, xanh lá cây, xanh lam trong trật tự như thế.

Số lượng màu tối đa sẽ là:

$256 \times 256 \times 256$ hay 256^3 hay $2^{24} = 16.777.216$

Ví dụ:

- (0, 0, 0) là màu đen.
- (255, 255, 255) là màu trắng.
- (255, 0, 0) là màu đỏ.
- (0, 255, 0) là màu xanh lá cây.
- (0, 0, 255) là màu xanh lam.
- (255, 255, 0) là màu vàng.
- (0, 255, 255) là màu xanh ngọc.
- (255, 0, 255) là màu hồng cánh sen.

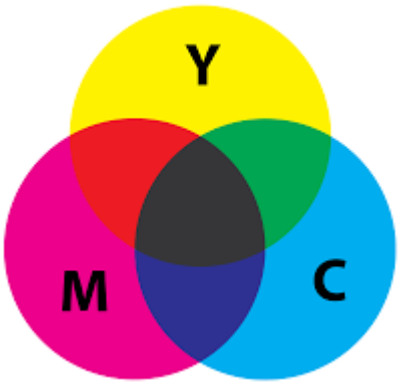


Hình 2.1 Không gian màu RGB

2.1.2 Không gian màu CMYK

CMYK là không gian màu được sử dụng phổ biến trong ngành công nghiệp in ấn. Ý tưởng cơ bản của hệ không gian này là dùng 4 màu sắc cơ bản để phục vụ cho việc pha trộn mực in. Trên thực tế, người ta dùng 3 màu là C: xanh lơ (Cyan), M: hồng sẫm (Magenta) và Y: vàng (Yellow) để biểu diễn các màu sắc khác nhau. Nếu lấy màu hồng sẫm cộng với vàng sẽ ra màu đỏ, màu sẫm kết hợp với xanh lơ sẽ cho xanh lam. Sự kết hợp của 3 màu trên sẽ cho ra màu đen, tuy nhiên màu đen ở đây không phải là đen tuyệt đối và thường có độ tương phản lớn, nên trong ngành in, để tiết kiệm mực in người ta thêm vào màu đen để in những chi tiết có màu đen thay vì phải kết hợp 3 màu sắc trên. Và như vậy ta có hệ màu CMYK. chữ K ở đây là để kí hiệu màu đen (Black) (“CMYK color model,” 2015).

Nguyên lý làm việc của hệ màu này như sau: Trên một nền giấy trắng, khi mỗi màu này được in lên sẽ loại bỏ dần đi thành phần màu trắng. Ba màu C, M, Y khác nhau in theo những tỉ lệ khác nhau sẽ loại bỏ đi thành phần đó một cách khác nhau và cuối cùng cho ta màu sắc cần in. Khi cần in màu đen, thay vì phải in cả 3 màu người ta dùng màu đen để in lên. Nguyên lý này khác với nguyên lý làm việc của hệ RGB ở chỗ hệ RGB là sự kết hợp của các thành phần màu, còn hệ CMYK là sự loại bỏ lẫn nhau của các thành phần màu.

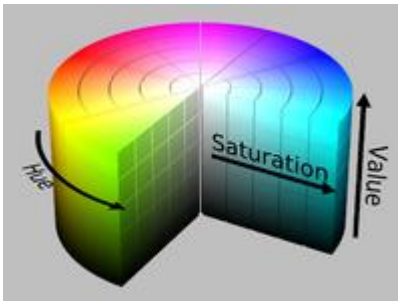


Hình 2.2 Không gian màu CYMK

2.1.3 Không gian màu HSV

HSV cũng gần tương tự như HSL là không gian màu được dùng nhiều trong việc chỉnh sửa, phân tích ảnh và là một phần của lĩnh vực thị giác máy tính. Hệ

không gian này dựa vào 3 thông số sau để mô tả màu sắc: H: màu sắc (Hue), S: độ bão hòa (Saturation), V: giá trị cường độ sáng (Value). Không gian màu này thường được biểu diễn dưới dạng hình trụ hoặc hình nón theo ("HSL and HSV," 2015).



Hình 2.3 Không gian màu HSV

Theo đó, đi theo vòng tròn từ 0 – 360 độ là trường biểu diễn màu sắc (Hue). Trường này bắt đầu từ màu đỏ chính (red primary) tới màu xanh lục chính (green primary) và nằm trong khoảng 0 – 120 độ, từ 120 – 240 độ là màu xanh lục chính (green primary) tới xanh lơ chính (blue primary). Từ 240 – 360 độ là từ màu đen tới lại màu đỏ.

Theo như cách biểu diễn không gian màu theo hình trụ như trên, giá trị độ sáng (V) được biểu diễn bằng cách đi từ dưới đáy hình trụ lên và nằm trong khoảng từ 0 – 1. Ở đáy hình trụ, V có giá trị là 0, là tối nhất và trên đỉnh hình trụ là độ sáng lớn nhất (V = 1).

Đi từ tâm hình trụ ra mặt trụ là giá trị bão hòa của màu sắc (S). S có giá trị từ 0 – 1. 0 ứng với tâm hình trụ là chỗ mà màu sắc là nhạt nhất. S = 1 ở ngoài mặt trụ là nơi mà giá trị màu sắc là đậm đặc nhất.

Như vậy với mỗi giá trị (H, S, V) sẽ cho ta một màu sắc mà ở đó mô tả đầy đủ thông tin về màu sắc, độ đậm đặc và độ sáng của màu đó.

2.1.4 Các không gian màu CIE

CIE là tên viết tắt của Commission Internationale de l'Eclairage, hiệp hội được thành lập vào năm 1913 với chức năng như một diễn đàn quốc tế cung cấp, trao đổi thông tin và các ý tưởng về mọi vấn đề liên quan đến chiếu sáng. Trong đó, nghiên cứu màu sắc là một trong những nhiệm vụ chính của tổ chức này ("International Commission on Illumination," 2015).

Mô hình màu CIE được phát triển để có thể độc lập với bất kì một phương tiện chiếu sáng nào và dựa trên nền tảng cảm thụ màu sắc của mắt người. Chìa khóa quan trọng của mô hình CIE là sự định nghĩa cụ thể các nguồn sáng chuẩn và sự quy định cụ thể đối với người quan sát chuẩn.

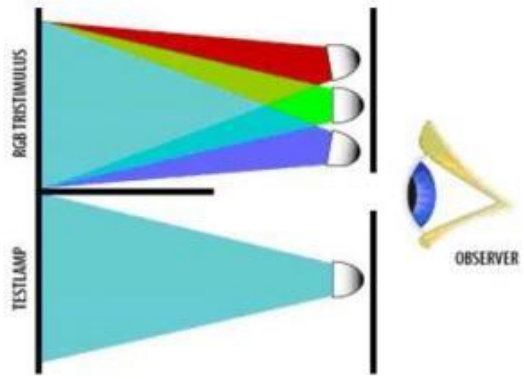
Các nguồn sáng chuẩn được định nghĩa năm 1913:

• Nguồn A: bóng đèn sợi tóc Tungsten với nhiệt độ màu 2854 K.

• Nguồn B: mô hình ánh sáng mặt trời vào buổi trưa với nhiệt độ 4800 K.

• Nguồn C: mô hình ánh sáng mặt trời ban ngày với nhiệt độ 6500 K.

Nguồn B và C thật ra xuất phát từ nguồn A thông qua sử dụng các tấm lọc thay đổi sự phân bố nguồn quang phổ. Bên cạnh đó, CIE cũng đã định nghĩa 1 loạt các nguồn sáng daylight gọi là daylight D. Trong đó, D65 với nhiệt độ màu ở 6500 K được sử dụng làm hệ tham chiếu phổ biến nhất.

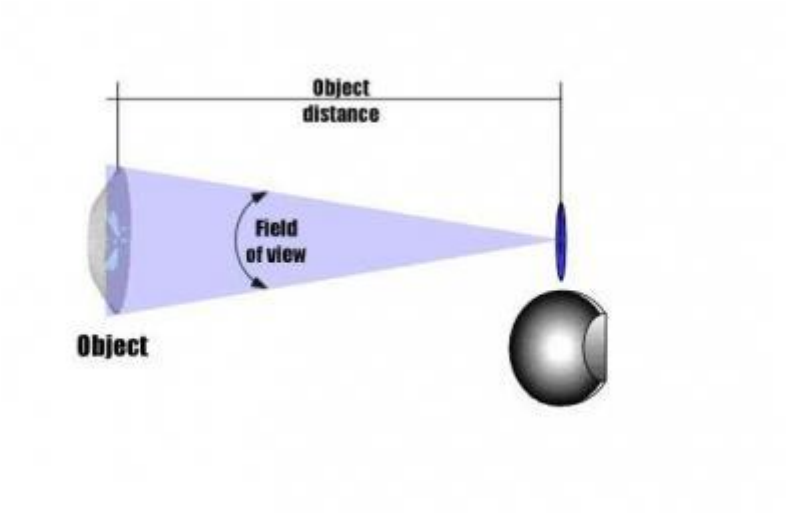


Hình 2.4 Hệ thống quan sát CIE

CIE thường sử dụng hai hệ thống người quan sát chuẩn: hệ quan sát 1913 và 1964. Trong cả hai trường hợp, người quan sát chuẩn là sự tổng hợp của những nhóm nhỏ từ 15 – 20 người và như thế, đại diện cho sự quan sát màu sắc của con người bình thường.

Người quan sát sẽ nhìn vào 1 màn hình bị chia đôi với độ phản xạ là 100% (trắng tinh khiết). Ở 1 nửa màn hình, 1 bóng đèn kiểm tra sẽ chiếu 1 màu sắc lên, ở nửa còn lại, 3 bóng đèn đỏ, xanh dương, xanh lá sẽ thay đổi hàm lượng ánh sáng để

có thể tạo ra màu giống với màu trong nửa màn hình thứ nhất. Người quan sát sẽ nhìn thông qua 1 lỗ hổng và quyết định khi nào thì màu sắc của 2 màn hình giống hệt nhau. Bộ 3 thông số (tritimulus) RGB cho mỗi 1 màu sắc sẽ có được theo cách thức này.



Hình 2.5 Phạm vi quan sát trong mô hình của CIE

Một khi các giá trị tristimulus RGB đã thu được, chúng sẽ được sử dụng trong các mô hình mẫu của màu sắc. Do sự ràng buộc của các gam màu, mô hình màu sắc *RGB* không thể sao chép toàn bộ ánh sáng quang phổ mà không đưa vào các giá trị thông số *RGB* âm. CIE đã chuyển đổi các giá trị của bộ 3 thông số *RGB* vào trong 1 hệ khác chỉ sử dụng hoàn toàn các giá trị dương, được gọi là *XYZ*. Đây cũng chính là mô hình mẫu đầu tiên của CIE, còn gọi là không gian màu CIE *XYZ* (Yang et al., 2010).

2.2 Bài toán nhận dạng ảnh mặt người

Trong nhiều năm qua, có rất nhiều công trình nghiên cứu về bài toán nhận dạng mặt người. Các nghiên cứu đi từ bài toán đơn giản, từ việc nhận dạng một mặt người trong ảnh đen trắng cho đến mở rộng cho ảnh màu và có nhiều mặt người trong ảnh. Đến nay các bài toán xác định mặt người đã mở rộng với nhiều miền nghiên cứu như nhận dạng khuôn mặt, định vị khuôn mặt, theo dõi mặt người hay nhận dạng cảm xúc mặt người…(Bowyer, 2004)

Phát hiện mặt người trong ảnh là phần đầu tiên của một hệ thống nhận dạng mặt người. Các hệ thống nhận dạng khuôn mặt được bắt đầu xây dựng từ những năm 1970, tuy nhiên do còn hạn chế về các luật xác định mặt người nên chỉ được áp dụng trong một số ứng dụng như nhận dạng thẻ căn cước. Nó chỉ được phát triển mạnh mẽ từ những năm 1990 khi có những tiến bộ trong công nghệ video và ngày nay thì các ứng dụng của xác định mặt người đã trở nên phổ biến trong cuộc sống.

Bài toán xác định khuôn mặt người (Face Regconition) là một kỹ thuật máy tính để xác định các vị trí và các kích thước của khuôn mặt trong ảnh bất kỳ (ảnh kỹ thuật số). Kỹ thuật này nhận biết các đặc trưng khuôn mặt và bỏ qua những thứ khác như: toà nhà, cây cối cơ thể.

Một số ứng dụng của bài toán xác định khuôn mặt là: hệ thống tương tác giữa người và máy (điều kiển máy tính qua các cử động của khuôn mặt), hệ thống nhận dạng người (giúp cho các cơ quan an ninh quản lý con người), hệ thống quan sát theo dõi, hệ thống quản lý việc ra vào cho các cơ quan và công ty, hệ thống kiểm tra người lái xe có ngủ gật hay không, hệ thống phân tích cảm xúc trên khuôn mặt, và hệ thống nhận dạng khuôn mặt cho các máy ảnh kỹ thuật số…

Thuật toán nhận diện khuôn mặt hiện chia làm hai loại là hình học (geometric) và trắc quang (photometric). Hình học nhận diện khuôn mặt dựa trên các đặc trưng trên khuôn mặt như mắt, mũi, miệng, gò má; trong khi trắc quang là phương pháp biến hình ảnh thành các giá trị và so sánh với giá trị mẫu để nhận diện. Các nhà nghiên cứu ngày nay đã phát triển những kỹ thuật nhận diện khuôn mặt riêng, nhưng phổ biến nhất hiện có ba loại chính là phân tích thành phần chính (PCA), phân tích phân lớp tuyến tính (LDA) và phương pháp đồ thị đàn hồi (EBGM).

Cách nhận diện khuôn mặt sử dụng phương pháp PCA phụ thuộc rất nhiều vào cơ sở dữ liệu ban đầu chứa các ảnh mẫu và góc quay camera cũng như ánh sáng. Sử dụng các thuật toán đại số để tìm giá trị mặt riêng và vector riêng rồi so sánh với giá trị mẫu, ta thu được khuôn mặt cần nhận diện. Đặc điểm của phương pháp này là giảm thiểu được dữ liệu cần sử dụng làm mẫu. Trong khi đó, phương

pháp LDA lại phân loại các lớp chưa biết thành các lớp đã biết, mà ở đó các khuôn mặt tạo thành một lớp và sự khác biệt giữa các khuôn mặt trong một lớp là rất nhỏ. Cả PCA và LDA đều chọn cách thống kê lấy mẫu, chọn lọc để nhận diện khuôn mặt.

Phương pháp còn lại EBGM chia mặt thành mạng lưới gồm các nút với mỗi khuôn mặt có khoảng 80 điểm nút. Vị trí của các nút giúp xác định khoảng cách giữa hai mắt, độ dài của sống mũi, độ sâu của hốc mắt, hình dạng của gò má… Điểm khó của phương pháp này là cần tính toán chính xác khoảng cách giữa các điểm nút, và do đó đôi khi nó phải dùng kết hợp với các phương pháp như PCA hay LDA.

Với những hạn chế khi sử dụng công nghệ nhận diện khuôn mặt truyền thống, phương pháp nhận diện 3D đã trở thành hướng đi mới trong việc ứng dụng công nghệ nhận diện khuôn mặt. Phương pháp này lưu lại hình ảnh 3D của khuôn mặt với các điểm đặc trưng như độ cong của cằm, mũi, hốc mắt… Ưu điểm của nó là có thể nhận diện khuôn mặt ở nhiều góc độ khác nhau, không bị ảnh hưởng bởi ánh sáng.

Cũng như những phương pháp truyền thống, phương pháp nhận diện khuôn mặt 3D cũng dựa trên các thuật toán. Nó tính toán các đường cong, những điểm đặc trưng trên khuôn mặt để tạo thành những dòng lệnh duy nhất và so sánh với cơ sở dữ liệu. Chúng ta có thể dễ dàng bắt gặp quá trình so sánh này trong các bộ phim hành động của Mỹ, khi hình ảnh của một người được camera ghi lại và ngay lập tức nó được so sánh liên tục với hàng triệu khuôn mặt trong cơ sở dữ liệu của cảnh sát.

Ngày nay, các công ty của Mỹ đã cải tiến phương pháp nhận diện 3D bằng việc bổ sung thêm nhận diện mẫu da, được gọi là phương pháp phân tích vân bề mặt. Phương pháp này cũng sử dụng các thuật toán chia nhỏ vùng da thành các không gian có thể đo đếm được, giúp xác định danh tính của cả những cặp sinh đôi.

Nhận diện khuôn mặt 3D vẫn chưa hoàn hảo, nó vẫn bị hạn chế đáng kể bởi các yếu tố bên ngoài như khuôn mặt bị tóc che phủ, đeo kính, hình ảnh quá mờ. Các công ty của Mỹ hiện vẫn đang liên tục tìm cách cải tiến để tăng độ chính xác cho

công nghệ nhận diện khuôn mặt mà không gây khó chịu cho người bị nhận diện.

2.3 Một số độ đo khoảng cách

2.3.1. Các độ đo khoảng cách giữa các đối tượng

Được mô tả trong (Pękalska and Duin, 2005), để phân tích sự khác biệt giữa các đối tượng được mô tả bởi các vector trong một không gian đặc trưng, một số độ đo khác nhau có thể được xem xét. Nếu các vectơ trung bình được sử dụng để làm đại diện cho toàn bộ các đối tượng, chúng có thể được sử dụng để tính toán khoảng cách giữa các nhóm theo các công thức từ bảng 2.1.

Bảng 2.1 Bảng tham chiếu các độ đo khoảng cách giữa các đối tượng

Ref.	D	Độ đo tương tự d(x,y)	M	E
Euclidean	D_E, D_2	$\sqrt{(x-y)^T(x-y)}$	Yes	Yes
Weight Euclidean	$D_{\omega E}$	$\sqrt{(x-y)^T diag(w_i^2)(x-y)}$	Yes	Yes
City Block	D_1	$\sum_{i=1}^{m} \lvert x_i - y_i \rvert$	Yes	No
Max norm	$D_{\max}$	$\max_i \lvert x_i - y_i \rvert$	Yes	No
ℓ_p or Minkowski	D_p	$\left(\sum_{i=1}^{m} \lvert x_i - y_i \rvert^p\right)^{\frac{1}{p}}, p \geq 1, p \neq 2$	Yes	No
Mahalanobis	D_M	$\sqrt{(x-y)^T C^{-1}(x-y)}$; C is psd	Yes	Yes
Median distance	D_{med}	$D_{\lfloor \frac{n}{2} \rfloor - rank}$	No	No
Correlation-base	D_{corr}	$\frac{1}{2}\left(1 - \frac{x^T y}{\lVert x \rVert^2 + \lVert y \rVert^2}\right)$	No	No
Correlation-base	D_{corr2}	$\frac{1}{2}\left(1 - \frac{x^T y}{\lVert x \rVert^2 + \lVert y \rVert^2 - 2x^T y}\right)$	No	No
Cosine	D_{cos}	$\frac{1}{2}\left(1 - \frac{x^T y}{\lVert x \rVert \lVert y \rVert}\right)$	No	No

Divergene	D_{div}	$\sqrt{\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{(x_i+y_i)^2}}$	No	No
Bray and Curtis	D_{BC}	$\frac{\sum_{i=1}^{n}\lvert x_i-y_i\rvert}{\sum_{i=1}^{n}x_i+y_i}$	No	No
Soergel	D_S	$\frac{\sum_{i=1}^{n}\lvert x_i-y_i\rvert}{\sum_{i=1}^{n}\max\{x_i,y_i\}}$	No	No
Ware and Hedges	D_{WH}	$\sum_{i=1}^{n}\left(1-\frac{\min\{x_I,y_i\}}{\max\{x_i,y_i\}}\right)$	No	No

Một khả năng khác là đặc trưng cho một đối tượng bằng một hàm phân bố xác suất nhiều biến. Sau đó, sự khác biệt giữa hai quần thể được đo bằng sự khác nhau giữa hai hàm phân bố xác suất F_1 và F_2. Độ đo Kolmogorov thường được sử dụng (Gibbs et al., n.d.). Cho hai hàm phân phối F_1 và F_2, độ đo Kolmogorov được định nghĩa như sau:

$$D_K(F_1, F_2) = \sup|F_1(x) - F_2(x)|$$

Như một phần mở rộng, việc đánh giá sự khác nhau giữa các đối tượng cũng có thể dựa vào mô tả từng phân phối như là một điểm trong một không gian Riemann với các tọa độ xác định bởi các thông số đối tượng. Ví dụ: một đối tượng đặc trưng bởi một hàm mật độ bình thường được xác định bởi các tọa độ (μ, Σ) trong $m + m(m + 1)/2$ không gian chiều. Đối tượng được mô tả bởi các thông số tương tự sẽ được ánh xạ thành các điểm lân cận trong không gian này. Với điều kiện một độ đo metric phù hợp có thể được xác định, sự khác nhau giữa các nhóm là chiều dài trắc địa (kết nối con đường ngắn nhất hai điểm) giữa các điểm biểu diễn cho đối tượng.

2.3.2. Độ đo khoảng cách giữa các dãy

Gọi A là một bảng chữ cái, ví dụ như một bộ sưu tập hữu hạn các ký hiệu, còn được gọi là chữ cái, từ đó trình tự hoặc chuỗi được tạo thành. Cho $s = s_1, s_2, \dots s_n$ là một chuỗi các ký tự từ A. Một từ rỗng được ký hiệu là ε và nó có

một chiều dài vô hạn. Chuỗi này được sử dụng trong các mô hình nhận biết và khu vực máy học để mã hóa các đối tượng của cấu trúc tương đối đồng nhất. Ở đây sẽ giới thiệu ngắn gọn các độ đo khoảng cách phổ biến nhất.

- **Khoảng cách Euclid**

- *Khoảng cách Euclidean (Euclidean Distance)*: Khoảng cách giữa 2 điểm là chiều dài của đường thẳng nối chúng. Trong mặt phẳng, khoảng cách giữa 2 điểm (x_1, y_1) và (x_2, y_2) được cho bởi định lý Pythagorean như sau:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Trong không gian ba chiều Euclidean, khoảng cách giữa hai điểm (x_1, y_1, z_1) và (x_2, y_2, z_2) là:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

Một cách tổng quát, khoảng cách giữa 2 điểm x, y trong không gian Euclidean R với n chiều được tính như sau:

$$d = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$

- **Khoảng cách Hamming**

Đây là một trong những độ đo đơn giản nhất: cho hai chuỗi chiều dài bằng nhau, nó đếm vị trí ký hiệu mà chúng khác nhau. Để không mất tính tổng quát, cho $\boldsymbol{s = s_1 s_2 \ldots s_n}$ và $= \boldsymbol{t_1 t_2 \ldots t_n}$, là chuỗi nhị phân. Khoảng cách Hamming sau đó được xác định như sau: $d_{Ham}(s, t) = \sum_{k=1}^{n} I(s_k \neq t_k)$.

Nó không phải là một độ đo linh hoạt vì nó giả định các chuỗi có chiều dài cố định. Tuy nhiên, trong nhiều bài toán, các chuỗi có chiều dài thay đổi, và hơn thế nữa, có thể giữa các vị trí biểu tượng sự tương ứng không ổn định. Sự thay đổi nhỏ của các vị trí của một trong hai chuỗi gần như giống hệt nhau và có thể dẫn đến các giá trị phóng đại trong khoảng cách Hamming.

- **Khoảng cách Hamming mờ**

Một khoảng cách Hamming mờ đã được đề xuất để làm cho khoảng cách Hamming nhạy cảm với khu vực lân cận (Bookstein et al., 2002). Đây là kiểu của khoảng cách chỉnh sửa cho trình tự của chiều dài bằng nhau. Sửa khoảng cách dựa trên chuyển một chuỗi thành chuỗi khác bằng cách sử dụng phép toán sửa đổi. Các hoạt động của phép toán chỉnh sửa: chèn, xóa và thay đổi, với giá trị $\boldsymbol{c_{ins}}$, $\boldsymbol{c_{del}}$ và $\boldsymbol{c_{sub}}$ được phân công tương ứng. Các phép toán dịch chuyển cho phép chuyển đổi một 1 bit trong một chuỗi đến gần 1 bit trong chuỗi khác với chi phí nhỏ hơn bởi có cả hai phép xóa và chèn. Các phép toán được sử dụng để chuyển đổi một chuỗi ký tự thành chuỗi khác và sự khác nhau của kết quả $\boldsymbol{d_{fHam}}$ được tính bằng cách cộng các chi phí của các phép toán, như vậy nó có tổng chi phí tối thiểu. Khoảng cách Hamming mờ là độ đo metric nếu $\boldsymbol{c_{del} = c_{ins}}$ và cho kích thước tuyệt đối của sự chuyển dịch $h \geq 0$, $\boldsymbol{c_{sub}}(h) \geq 0$ và $\boldsymbol{c_{sub}}(h) = 0$ khi và chỉ khi $h = 0$, $\boldsymbol{c_{sub}}(h)$ tăng đơn điệu và lõm trên các số nguyên (Bookstein et al., 2002).

- **Khoảng cách Levenshtein**

Khoảng cách chỉnh sửa phổ biến nhất là khoảng cách Levenshtein, thể hiện một sự tương đồng nội vùng giữa các chuỗi có độ dài tùy ý. Nó được dựa trên ba phép toán sửa: chèn, xóa và thay thế. Các giá trị $\boldsymbol{c_{ins}}$, $\boldsymbol{c_{del}}$ và $\boldsymbol{c_{sub}}$ tương ứng với mỗi phép toán trong số đó, dẫn đến một phiên bản trọng lượng của khoảng cách này. Trong khoảng cách chỉnh sửa, $\boldsymbol{c_{sub} > c_{del}}$.

+ $\boldsymbol{c_{ins}}$, có nghĩa là xóa của a và chèn của b là sự thay thế của a cho b. Nếu tất cả các giá trị như vậy là một biến đơn không lớn hơn tổng của hai giá trị khác, sau đó d_L là độ đo metric. Tương tự như $\boldsymbol{d_{fHam}}$, trọng lượng khoảng cách Levenshtein d_L được xác định bằng tổng chi phí tối thiểu liên quan đến hoạt động chuyển đổi một chuỗi s thành t. Giả định rằng một chuyển đổi như vậy đòi hỏi phải thay thế $\boldsymbol{n_{sub}}$, chèn $\boldsymbol{n_{ins}}$ và phép xóa $\boldsymbol{n_{del}}$, d_L được thể hiện như sau:

$$d_L(s,t) = \min_{\mathrm{n_{sub}},\mathrm{n_{ins}},n_{del}} (n_{sub}c_{sub} + n_{ins}c_{ins} + n_{del}c_{del})$$

Khoảng cách soạn thảo truyền thống với tất cả các giá trị tương đương thường được xét đến. Tuy nhiên vấn đề chính là d_L phụ thuộc vào độ dài của chuỗi so sánh và có thể bị ảnh hưởng nhiều bằng cách so sánh hai chuỗi, trong đó một là ngắn và chuỗi khác là rất dài.

- **Các khoảng cách liên quan khác**

Hai dãy có thể được so sánh dựa trên tiền tố chung dài nhất, hậu tố hay chỉ là một dãy. Giả sử chúng ta được cho hai chuỗi s và t của chiều dài n và m ≤ n, tương ứng. Sau đó, khoảng cách giữa chúng có thể được định nghĩa là $d(s,t) = m + n - 2|common(s,t)|$. Vấn đề của việc tìm kiếm dãy chung dài nhất là bổ sung cho việc xác định khoảng cách soạn thảo. Điều đó cũng có thể được giải quyết bằng việc sử dụng các quy hoạch động.

- **Khoảng cách thông tin và xấp xỉ thông tin**

Giả sử một tập hợp các xâu nhị phân. Độ tính phức tạp Kolmogorov $K(s)$ của một chuỗi nhị phân s là chiều dài (theo bit) của các chương trình máy tính nhanh nhất của một tham chiếu cố định hệ thống máy tính sản xuất s như là một kết quả. Sự thay đổi của một hệ thống máy tính thay đổi giá trị này bằng một chất phụ gia cố định liên tục xem trong. Một cách giải thích có thể có của $K(s)$ là chiều dài của phiên bản nén cuối cùng của s mà s có thể được phục hồi bằng một chương trình giải nén. Để đo sự khác biệt giữa hai xâu, s và t, khoảng cách thông tin chuẩn đã được đề xuất (Pękalska and Duin, 2005):

$$NID(s,t) = \frac{K(s,t) - \min\{K(s), K(t)\}}{\max\{K(s), K(t)\}}$$

Lưu ý rằng $K(s,t)$ là độ dài của chương trình ngắn nhất mà bản in s và t mô tả làm thế nào để phân biệt chúng. Từ khoảng cách NID là không thể tính được, một xấp xỉ được đề nghị sử dụng chương trình nén dữ liệu để ước tính K. Điều này dẫn đến khoảng cách nén chuẩn được định nghĩa là (Cilibrasi and Vitanyi, 2005).

$$NCD(s,t) = C(s,t) - \frac{\min\{C(s), C(t)\}}{\max\{C(s), C(t)\}}$$

trong đó C là chương trình nén lựa chọn và $C(s)$ là chiều dài của chuỗi nén. Bất kỳ chuỗi (sau khi mã hóa lại thích hợp để các chuỗi nhị phân) có thể được so sánh bằng khoảng cách này, chẳng hạn như các tập tin nhị phân như phần âm nhạc ở định dạng MIDI (Musical Instrument Digital Interface: Giao Diện Số Hoá Nhạc Cụ).

2.3.3. Độ đo theo lý thuyết thông tin

Theo ý nghĩa lý thuyết thông tin, một định nghĩa phổ biến của sự tương đồng, áp dụng đối với các lĩnh vực trong đó có một mô hình xác suất, được đề xuất bởi (Lin, 1998). Nó được dựa trên quan sát chung rằng sự tương đồng giữa hai đối tượng là kết nối phổ biến và chúng khác nhau và hai đối tượng đồng nhất dẫn đến sự giống nhau tối đa. Điều này dẫn đến các giả định sau (Lin, 1998)

(1) Sự tương đồng giữa A và B được đo bằng $I\big(com(A,B)\big)$, trong đó I là số lượng thông tin, thường là logarit tiêu cực xác suất của các sự kiện nó đề cập đến.

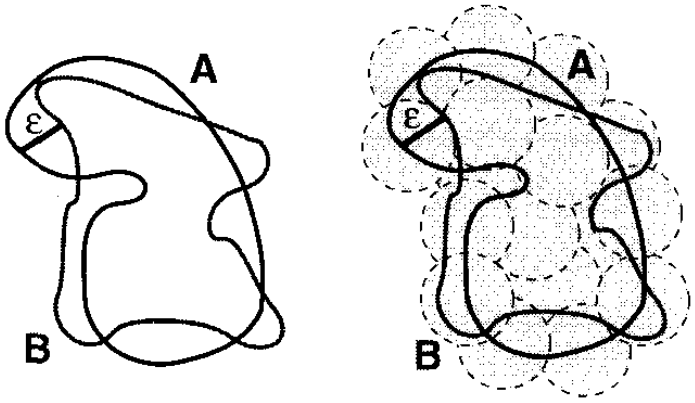
(2) Sự khác biệt giữa A và B được đo bằng $I(desc(A,B)) - I\big(com(A,B)\big) \geq 0$, trong đó $desc(A,B)$ là một đề xuất mô tả A và B.

(3) Sự tương đồng là một hàm $f: R^0 + x\ R_+ \rightarrow [0,1]$ của tương đồng và sự khác biệt cho là $sim(A,B) = f\left(I\big(com(A,B)\big), I\big(desc(A,B)\big)\right)$, chẳng hạn $f(x,x) = 1$ và $f(0,y) = 0$.

(4) Sự tương tự chung của hai đối tượng là trung bình có trọng số được tính tương tự từ những quan điểm khác nhau.

Sự giống nhau bắt nguồn từ những giả định được tính như tỷ lệ giữa số lượng thông tin cần thiết tới tính phổ biến của hai đối tượng và số lượng thông tin cần thiết để mô tả chúng. Nó được gọi là $sim(A,B) = \log P\big(com\,(A,B)/$ $\log P\big(desc(A,B)\big)\big)$ (Lin, 1998) trình bày định nghĩa chung này được áp dụng cho một số lĩnh vực, kết quả là một sự tương đồng giữa các chuỗi, từ hoặc khái niệm trong phân loại.

Khái quát và khoảng cách phổ biến độ đo metric đã được đề xuất trong và tiếp tục khám phá trong (Cilibrasi and Vitanyi, 2005). Như các tác giả khẳng định, độ đo metric của chúng là tổng thể, nó có thể được áp dụng trong nhiều lĩnh vực như: âm nhạc, văn bản, hệ gen, thực thi chương trình hoặc mô tả ngôn ngữ tự nhiên và nó không tập trung vào đặc trưng đặc biệt hoặc tương đồng giữa các trường, nhưng phải mất tất cả chúng cùng một lúc vào tài khoản. Ý tưởng cơ bản là để thể hiện sự gần gũi của hai đối tượng nếu chúng có ý nghĩa 'nén' cho các thông tin khác. Điều này được chính thức hóa bởi quan điểm của Kolmogorov phức tạp. Trong thực tế, nó được xấp xỉ bằng khoảng cách nén chuẩn. Cùng một nguyên tắc là tiếp tục sử dụng để xác định một độ đo khoảng cách, Googlebased so sánh hai thuật ngữ tìm kiếm x và y như lập chỉ mục của (Cilibrasi and Vitanyi, 2007).



Hình 2. 6 Minh họa của khoảng cách Hausdorff giữa 2 tập hợp A và B:

$$\mathrm{dH(A, B)} = \varepsilon$$

- **Khoảng cách Hausdorff**

Cho(X, p) là một không gian độ đo metric và $C(X) \subseteq X$ là một không gian rỗng, tập con đóng kín và bị chặn của X. Cho $N_\varepsilon(A) = \bigcup_{x \in A} B\varepsilon(x)$ là lớp bảo vệ của $A \in X$ bằng cách mở ε-balls. Từ $B_\varepsilon(x) = y \in X: p(x, y) < \varepsilon$ là lân cận của A. Khoảng cách Hausdorff giữa A và B được định nghĩa là ε-lân cận nhỏ nhất của A bao gồm B và ngược lại. Mặt khác, hướng khoảng cách Hausdorff giữa A và B, có thể được thể hiện như tối đa thực hiện trên bộ sưu tập của khoảng cách tối thiểu

giữa các phần tử của A và tập B. Sau đó, khoảng cách Hausdorff $d_H(AB)$ là lớn nhất trong hai khoảng cách chỉ dẫn. Chính thức, chúng ta có:

Trong không gian (nửa) metric (X, p), khoảng cách Hausdorff với các cơ sở p được xác định cho tất cả $A, B \in C(X)$ theo những cách sau đây:

$$d_H(A,B) = \inf_{\varepsilon>0}\{A \subset N_\varepsilon(B)\ \&\ B \subset N_\varepsilon(A)\}$$

$d_H(A,B) = \max\{d_H^{\triangleright}(A,B), d_H^{\triangleright}(B,A)\}$ trong đó $d_H^{\triangleright}$ là một khoảng cách Hausdorff định hướng $d_H^{\triangleright}(A,B) = \sup_{\mathrm{a}\in\mathrm{A}} \inf_{b\in B} p(a,b)$

Nếu miền $d_H^{\triangleright}$ bị hạn chế, thì cận trên trở thành tối đa và vô cùng trở thành tối thiểu, cụ thể là: $d_H^{\triangleright}(A,B) = \max_{a\in A} \min_{\mathrm{b}\in\mathrm{B}} p(a,b)$

- **Các độ đo trên tập mờ**

Một khoảng cách Hausdorff-cùng tên cũng có thể được xác định cho tập mờ. Xét hai tập hợp không mờ A_f và B_f trên một tập hợp hỗ trợ S trong một thước đo không gian. Cho $x^* = \max\{A_f(t): A_f \in S\}$ là thành viên lớn nhất của x. Cho $A_{max} = \{\, t: A_f\,(t) = x^*\}$ là tập không mờ và cho A_a là một tập hợp không trống, không mờ của S, như vậy $A_{max} = B_{max}$, và cho hai tập mờ A_f và B_f , $A_a = B_a$ khi và chỉ khi $A_{max} = B_{max}$. Xác định các nhóm tập hợp không mờ $A_\mu, \mu \in [0,1]$ bằng cách:

$$A_\mu = \begin{cases} \{t: A_f(t) \in [\mu, x^*]\}, if\ \mu \le x^* \\ A_a, \qquad\qquad if\ \mu > x^* \end{cases}$$

Lưu ý rằng $A_\mu = A_{max}$ nếu $\mu = x^*\ for\ x^* \neq 1$.

Giả định rằng tập mờ có thể chỉ có giá trị từ một tập hợp rời rạc của các giá trị thành viên $\mu 1, \mu 2, \ldots, \mu c$. Cho $d_H(A_{\mu i}, B_{\mu i})$ là khoảng cách Hausdorff sắc nét giữa các bộ $A_{\mu i}$ và $B_{\mu i}$. Sau đó, khoảng cách Hausdorff-cùng tên mờ giữa A_f và B_f được định nghĩa là:

$$d_{H_f}(A_f, B_f) = \frac{\sum_{i=1}^{c} \mu_i d_H(A_{\mu i}, B_{\mu i})}{\sum_{i=1}^{c} \mu_i}$$

đó là độ đo metric (Chaudhuri and Rosenfeld, 1999). d_{H_f} có thể được xem như một trọng số thành viên trung bình của khoảng cách Hausdorff giữa mức độ

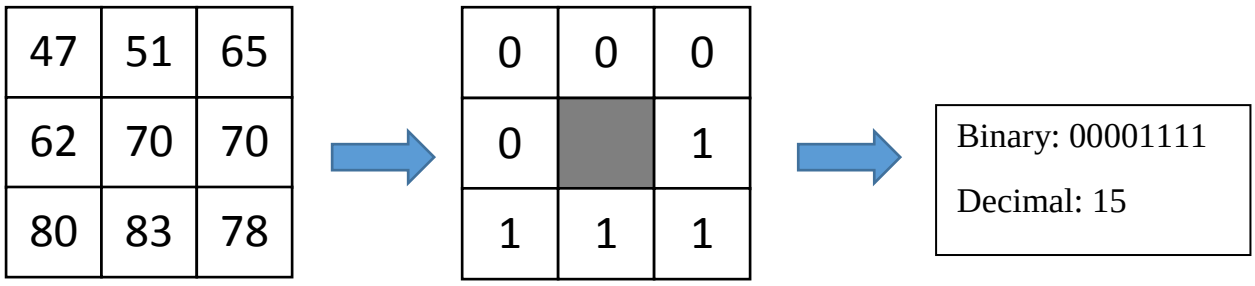
biến đổi tập mờ đặt ra xem xét. Lưu ý khoảng cách được thay đổi Hausdorff-cùng tên mờ có thể được định nghĩa bằng cách sử dụng các d_{MH} thay vì d_H trong công thức trên.

CHƯƠNG 3: KỸ THUẬT TRÍCH CHỌN KẾT CẤU ĐẶC TRƯNG MÀU CỤC BỘ

3.1 Mẫu nhị phân cục bộ (Local Binary Pattern – LBP)

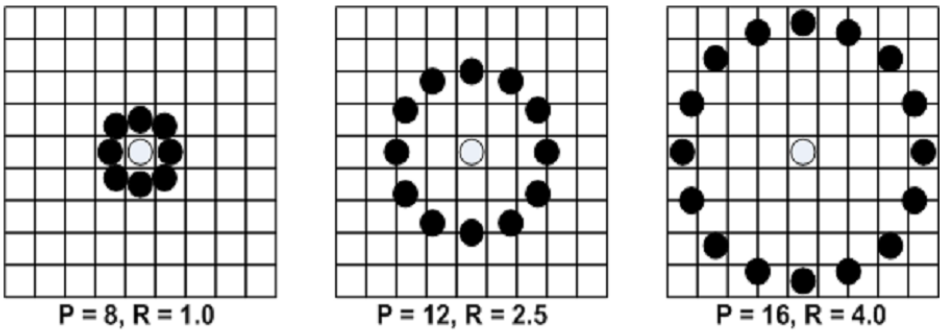
3.1.1 Thuật toán LBP

Thuật toán LBP được giới thiệu bởi Ojala et al. (1996) đánh nhãn cho các pixel hình ảnh bằng cách xác định ngưỡng neighborhood 3 x 3 của mỗi pixel làm giá trị trung tâm và xem xét kết quả như là một số nhị phân. Histogram trong số 2^8 = 256 nhãn khác nhau đó có thể được sử dụng như một mô tả kết cấu.



Hình 3.1 Mô tả thuật toán LBP gốc

Thuật toán LBP gốc của Ojala chỉ có 8 điểm lận cận 3x3 (có 2^8=256 label). Sau này, thuật toán đã được mở rộng để sử dụng các neighborhood với độ lớn khác nhau (Ojala et al. 2002)(Ojala et al., 2002). Mở rộng ra kích thước bất kì, khi đó kí hiệu (P, R) nghĩa là P điểm cách đều nhau nằm trên đường tròn bán kính R.



Hình 3.2 Số lượng điểm P cách đều trên các đường tròn với bán kính khác nhau

$$LBP_{P,R}^{i}(z_c) = \sum_{n=0}^{P-1} \delta\left(r_n^{(i)} - r_c^{(i)}\right)2^n$$

Trong đó:

(i) là hình ảnh grayscale cần tìm giá trị LBP.

(r_n) là giá trị grayscale của các pixel lân cận.

(r_c) là giá trị grayscale của các pixel trung tâm. (δ) là hàm nhị phân được xác định như sau: $\delta(x) = 1$ nếu giá trị $x \geq 0$.

Nếu tọa độ pixel trung tâm là (x_c, x_y) thì tọa độ của P pixel lân cận trên đường tròn tâm (x_c, x_y) bán kính R (theo đường tròn lượng giác) là:

$$x_p = x_c + Rcos(2\pi p/P),$$

$$y_p = y_c + Rsin(2\pi p/P).$$

3.1.2 Thuật toán Opponent color LBP (OCLBP)

Opponent color LBP được phát triển như là một toán tử kết cấu màu chung dùng để so sánh mức xám và đặc trưng kết cấu màu sắc (Jain and Healey, 1998). Ngoài ra, thuật ngữ "opponent color" cũng được sử dụng để chỉ ra rằng: Từng cặp các kênh màu trong một không gian màu được gọi là opponent color. Ví dụ: Không gian màu RGB có các cặp opponent color: RG, RB và GB.

Trong Opponent color LBP, các toán tử LBP được áp dụng trên mỗi kênh màu sắc một cách riêng biệt. Ngoài ra, mỗi cặp kênh màu opponent color còn được sử dụng để thu thập mẫu màu bằng cách sử dụng giá trị pixel trung tâm của kênh màu này và các giá tri neighborhood từ kênh màu còn lại.

Tương tự như công thức tính LBP, công thức tính Opponent color LBP được tính như sau:

$$LBP_{P,R}^{(i,j)}(z_c) = \sum_{n=0}^{P-1} \delta\left(r_n^{(j)} - r_c^{(i)}\right)2^n$$

Trong đó:

$(i), (j)$ là 2 hình ảnh trong không gian màu đối lập.

(r_n) là giá trị grayscale của pixel lân cận.

(r_c) là giá trị grayscale của pixel trung tâm.

(δ) là hàm nhị phân được xác định như sau: $\delta(x) = 1$ nếu giá trị $x \geq 0$.

3.1.3 Mẫu nhị phân đồng dạng – Uniform Pattern

Mẫu đồng dạng (uniform pattern): dựa trên quan sát một số mẫu xuất hiện tương đối phổ biến hơn các mẫu khác, Một mẫu nhị phân được gọi là đồng dạng khi xét chuỗi bit xoay vòng thì có *nhiều nhất là 2 lần thay đổi* (transitions) từ giá trị bit 0 sang 1 hoặc từ giá trị bit 1 sang 0.

Ví dụ:

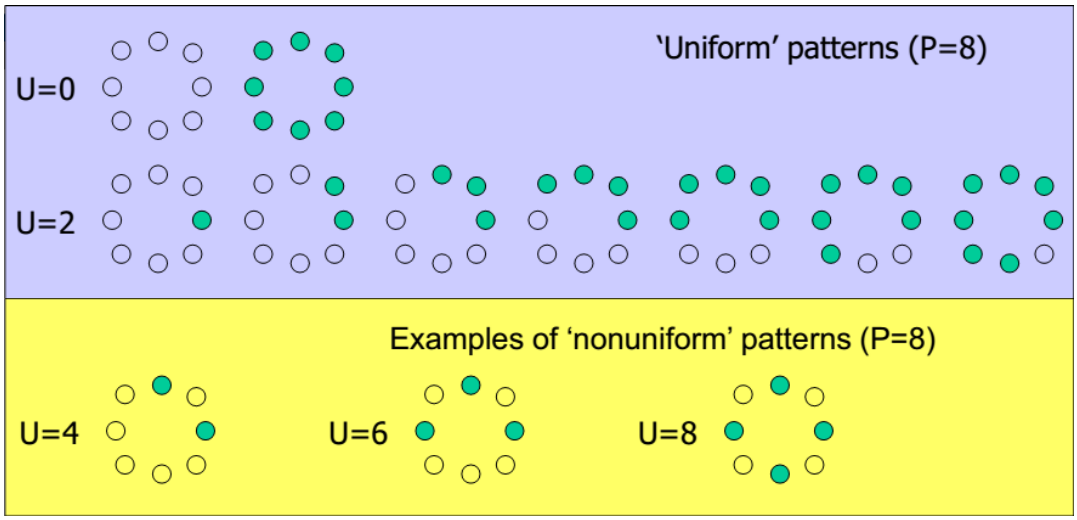
Mẫu 00000000 (0 lần dịch chuyển) -->uniform.

Mẫu 01110000 (2 lần dịch chuyển) -->uniform.

Mẫu 11001111 (2 lần dịch chuyển) --> uniform.

Mẫu 11001001 (4 lần dịch chuyển) --> không phải.

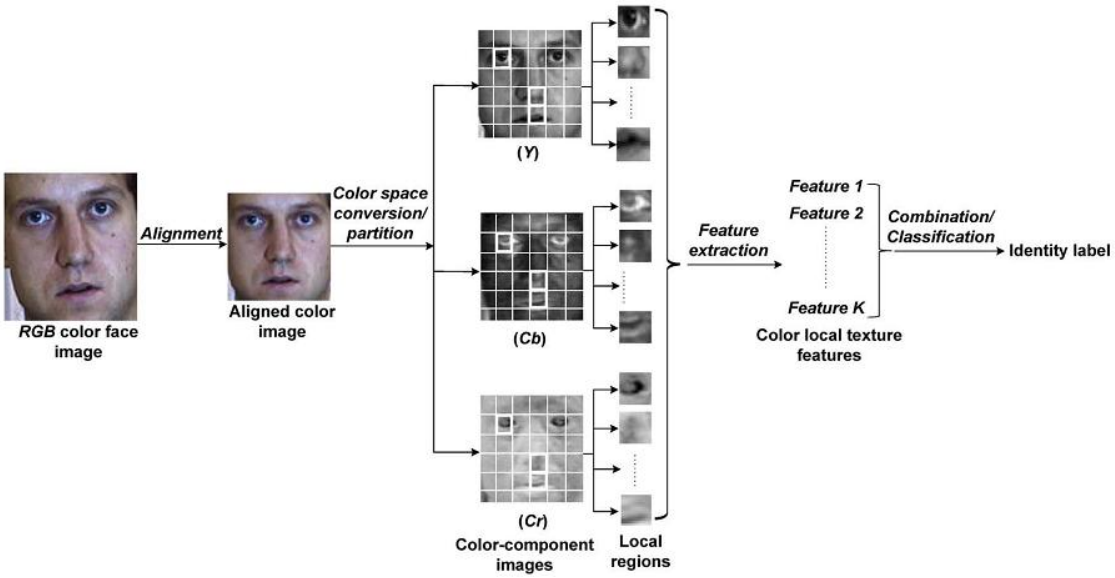
Mẫu 01010010 (6 lần dịch chuyển) --> không phải.



Hình 3.3 Các mẫu nhị phân đồng dạng và không dồng dạng.

3.2 Các bước xử lý trong phương pháp trích trọn đặc trưng

3.2.1 Mô hình trích chọn đặc trưng của Choi và các đồng sự



Hình 3.4 Mô hình nhận diện khuôn mặt màu sử dụng đặc trưng kết cấu màu cục bộ.

Phương pháp nhận diện khuôn mặt được Choi và các đồng sự đề xuất sử dụng đặc trưng kết cấu màu cục bộ bao gồm ba bước chính: chuyển đổi không gian màu - phân vùng, trích xuất đặc trưng, kết hợp - phân loại.

Một hình ảnh gương mặt trong không gian màu RGB đầu tiên sẽ được biên dịch, xoay và thay đổi tỷ lệ một mẫu cố định (Phillips et al., 2000), cho hình ảnh khuôn mặt phù hợp tương ứng. Sau đó, hình ảnh màu RGB được chuyển đổi một cách phù hợp thành một hình ảnh đại diện trong một không gian màu. Lưu ý rằng không chỉ các không gian màu tuyến tính hay phi tuyến tính (ví dụ YC_bC_r, hoặc $L^*a^*b^*$) mà nhưng những không gian màu mới được đưa ra với mục đích nhận diện khuôn mặt (ví dụ, không gian màu chuẩn hóa (normalized color space) được đề xuất trong (Yang et al., 2010) cũng có thể được sử dụng cho việc chuyển đổi không gian màu sắc . Mỗi thành phần hình ảnh màu của mô hình hiện tại được phân vùng thành 64 các vùng cục bộ (mỗi vùng có kích thước 15x15 pixel) như được đề xuất bởi (Zou et al., 2007).

Trong bước tiếp theo, việc trích chọn kết cấu đặc trưng được thực hiện một cách riêng biệt và độc lập trên từng khu vực cục bộ. Kể từ lúc đặc trưng kết cấu được trích chọn từ các vùng trên khuôn mặt và thu được từ kênh màu khác nhau, chúng được gọi là "Đặc trưng kết cấu màu cục bộ". Lưu ý rằng chìa khóa nhận diện khuôn mặt chính là sử dụng thông tin màu sắc để trích chọn cái gọi là đặc trưng kết cấu đối lập (opponent texture features) (Jain and Healey, 1998) giữa mỗi cặp của hai hình ảnh quang phổ. Điều này cho phép thu thập các đặc trưng kết cấu bổ sung nhiều hơn để cải thiện hiệu suất nhận diện khuôn mặt, so với trích chọn đặc trưng kết cấu mức xám khai thác tính năng kết cấu, chỉ có độ sáng của hình ảnh được đưa vào.

Từ N đặc trưng kết cấu màu cục bộ (mỗi đặc trưng thu được từ các khu vực cục bộ liên quan và các kênh phổ) có sẵn, tác giả phải kết hợp chúng để phân loại sau cùng. Để làm được điều này, kỹ thuật kết hợp đa phương thức (Jain et al., 2005) được sử dụng để tổng hợp các đặc trưng kết cấu màu sắc cục bộ để cải thiện hiệu suất nhận dạng khuôn mặt.

3.2.2 Trích chọn đặc trưng LBP màu (CLBP)

Ở đây, tác giả trình bày các phương pháp trích chọn các đặc trưng kết cấu màu cục bộ đã đề xuất từ một ảnh màu. Phương pháp trích chọn nổi bật được sử dụng đại diện cho đặc trưng kết cấu là LBP, các đặc trưng kết cấu mức xám được mở rộng ra đặc trưng kết cấu đa phổ sử dụng thông tin màu sắc. Cụ thể ở đây đưa ra một hình ảnh màu sắc, toán tử kết cấu được áp dụng trên mỗi kênh màu sắc riêng biệt. Ngoài ra, tác giả còn sử dụng toán tử kết cấu của màu sắc đối lập (opponent color). Thuật ngữ "màu đối lập" ở đây được hiểu theo định nghĩa được đề xuất bởi (Jain and Healey, 1998), tức là tất cả các cặp kênh màu khác nhau trong một không gian màu là "màu sắc đối lập."

Trước khi mô tả phương pháp trích chọn đặc trưng kết cấu màu cục bộ, tác giả giới thiệu các ký hiệu thường được sử dụng trong suốt các phần phụ sau. Giả định có K hình ảnh quang phổ khác nhau được tạo ra từ một hình ảnh khuôn mặt màu RGB thông qua chuyển đổi màu sắc đã được xác định trước. $\boldsymbol{S}_i$ là hình ảnh

quang phổ thứ i (ví dụ, thành phần không gian màu chrominance"Q" từ không gian màu YIQ), với $i = 1, \dots, K$. Hơn nữa, giả sử $\mathbf{S}_i$ được chia thành M các khu vực khuôn mặt (như mô tả trong Hình 3.4), tác giả biểu thị vục cục bộ thứ m của nó bằng $\mathbf{S}_i^{(m)}$.

Cho K không gian hình ảnh màu khác nhau $\mathbf{S}_i (i = 1, \dots, K)$, đặc trưng LBP Unichrome được tính toán riêng biệt và độc lập từ mỗi $\mathbf{S}_i$. Biểu thị z_c là vị trí pixel trung tâm của $\mathbf{S}_i$ và $z_n (n = 0, \dots, P-1)$ là các khoảng pixel cách đều nhau (hoặc các điểm mẫu) trên một vòng tròn bán kính R ($R > 0$) hình thành một khu vực neighborhood vòng cung xung quanh pixel trung tâm z_c. Toán tử LBP unichrome LBP cho vị trí pixel trung tâm z_c của $\mathbf{S}_i$ được định nghĩa như sau:

$\mathrm{LBP}_{P,R}^{i}(z_c) = \sum_{n=0}^{P-1} \delta\left(r_n^{(i)} - r_c^{(i)}\right) 2^n$, (1)

và $\delta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x \leq 0 \end{cases}$, $r_n^{(i)} (n = 0, \dots, P-1)$ biểu thị giá trị pixel của $\mathbf{S}_i$ tại z_n, và $r_c^{(i)}$biểu thị giá trị pixel tại z_c của các neighborhood xung quanh. Lưu ý rằng mỗi giá trị pixel của $L_i(x, y)$ được lấp đầy $LBP_{P,R}^{(i)}(z_c)$ tại vị trí pixel nhất định của nó $z_c = (x, y)$. Toán tử LBP nói trên thực hiện trên mỗi hình ảnh thành phần màu từ không gian màu được minh họa trong hình 3.5.

Để ánh xạ và mã hóa các đặc tính cục bộ (local properties) của một thành phần hình ảnh màu $\mathbf{S}_i$, tác giả đã tính toán LBP histogram cho từng vùng cục bộ $\mathbf{S}_i^{(m)} (m = 1, \dots, M)$. Histogram LBP cho cho vùng cục bộ thứ m, $\mathbf{S}_i^{(m)}$ được tính toán như sau:

$$h_i^{(m)}(k) = \sum_{z_c=(x,y)\in \mathbf{S}_i^{(m)}} T\left(\mathbf{L}_i(z_c) = k\right) \text{ for } 1 \leq k \leq P(P-1) + 3 \quad (2)$$

$$\text{với } T(A) = \begin{cases} 1, & \text{nếu } A \text{ đúng} \\ 0, & \text{nếu } A \text{ sai} \end{cases} \quad (3)$$

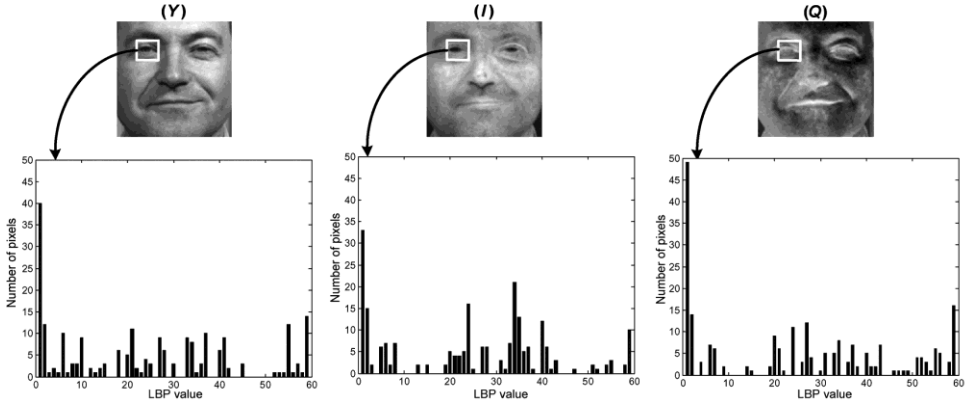
và k biểu thị giá trị LBP thứ k trong phạm vi $[0, P(P-1)+2]$ do đó, $h_i^{(m)}(k)$ biểu thị số lượng pixel với giá trị LBP k trong vùng cục bộ $\mathbf{S}_i^{(m)}$.

Sử dụng (2), đặc tả vùng LBP (regional LBP descriptor) cho $\mathbf{S}_i^{(m)}$ có thể được định nghĩa như sau:

$$\mathbf{h}_i^{(m)} = \left[h_i^{(m)}(1), h_i^{(m)}(2), \dots, h_i^{(m)}(P(P-1)+3)\right]^T. \quad (4)$$

Lưu ý rằng, trong (4), $\mathbf{h}_i^{(m)}$ cung cấp thông tin histogram LBP cho $\mathbf{S}_i^{(m)}$. Để giữ các thông tin về mối quan hệ không gian của vùng cục bộ của khuôn mặt, tất cả các giá trị $\mathbf{h}_i^{(m)}(m = 1, \dots, M)$ được nối vào một cột vector gọi đặc trưng LBP unichrome (unichrome LBP feature) cho $\mathbf{S}_i$, tức là:

$$\mathbf{x}_i^{U_LBP} = \left[(\mathbf{h}_i^{(1)})^T (\mathbf{h}_i^{(2)})^T \dots (\mathbf{h}_i^{(M)})^T\right]^T \quad (5)$$



Hình 3.5: Hình minh họa cách sử dụng thuật toán LBP Unichrome

Để tìm được đặc trưng LBP đối lập (opponent LBP feature) phải sử dụng mỗi cặp kênh màu khác nhau. Khi đó, pixel trung tâm kết hợp với neighborhood xung quanh lấy từ các kênh màu khác. Dựa trên (1), đưa ra một cặp $\mathbf{S}_i$ và $\mathbf{S}_j$, toán tử LBP đối lập được định nghĩa như sau:

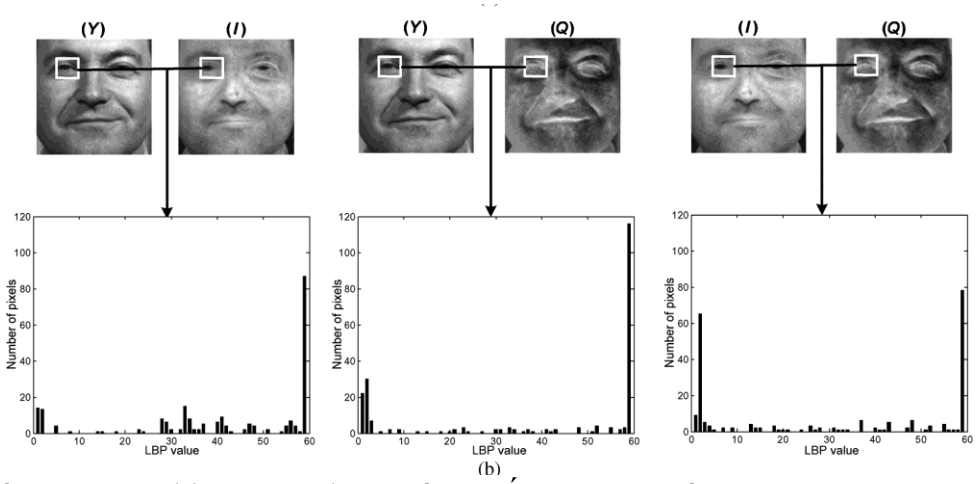
$$LBP_{P,R}^{(i,j)}(z_c) = \sum_{n=0}^{P-1} \delta\,(r_n^{(j)} - r_c^{(i)})2^n, \quad (6)$$

với $r_c^{(i)}$ biểu thị giá trị pixel của $\mathbf{S}_i$ và kết hợp z_c của nó (pixel trung tâm của các neighborhood xung quanh) và $r_n^{(i)}(n = 0, \dots, P-1)$ biểu thị giá trị pixel của $\mathbf{S}_j$ và hình thành các neighborhood xung quanh pixel trung tâm z_c của $\mathbf{S}_i$. Đặc trưng Opponent LBP được biểu thị bằng $\mathbf{x}_{i,j}^{O_LBP}$ cho các cặp $\mathbf{S}_i$ và $\mathbf{S}_j$ và có thể dễ dàng tính toán từ $LBP_{P,R}^{(i,j)}$.

Cuối cùng, đặc trưng LBP màu được đề xuất (CLBP) cho $\mathbf{S}_i$ có thể được định nghĩa như sau:

$$\mathbf{x}_i^{\mathrm{CLBP}} = \left[(\mathbf{x}_i^{\mathrm{U_LBP}})^T (\mathbf{x}_{i,j}^{\mathrm{O_LBP}})^T \dots (\mathbf{x}_{i,K}^{\mathrm{O_LBP}})^T\right]^T \text{ với } i \neq j \quad (7).$$

Lưu ý rằng, trong (7), $\mathbf{x}_i^{\mathrm{CLBP}}$ chứa một đặc trưng LBP unichrome (được trích chọn từ $\mathbf{S}_i$) và $K-1$ đặc trưng LBP đối lập được tính toán giữa $\mathbf{S}_i$ (với i cụ thể) và $(K-1)$ các $\mathbf{S}_j$ khác nhau ($j=1,\dots,K$ và $i \neq j$).



Hình 3.6 Các histogram được trích xuất từ thuật toán Opponent LBP

Hình 3.6 và 3.7 cho thấy sáu histogram LBP mẫu khác nhau được xác định. Ba histogram đầu tiên có nguồn gốc từ việc thực hiện thuật toán LBP unichrome trên ba khu vực cục bộ tương ứng, mỗi khu vực tương ứng với mỗi kênh trong không gian màu YIQ, trong khi ba histogram còn lại được tạo thành từ việc thực hiện thuật toán LBP đối lập trên cặp hai kênh màu khác nhau trên vùng cục bộ tương ứng. Dễ dàng thấy rằng các histogram này có giá trị LBP khác nhau nên chúng bổ sung cho nhau cho mục đích phân loại. Thật vậy, $\mathbf{x}_i^{\mathrm{CLBP}}$ đề xuất chứa nhiều thông tin phân biệt hơn một thuật toán LBP mức xám có thể cung cấp.

3.2.5 Kết hợp

Sau khi đã có các đặc trưng kết cấu, việc cần thiết là kết hợp các đặc trưng kết cấu màu cục bộ để đạt được chất lượng nhận dạng khuôn mặt tốt nhất. Tác giả biểu thị một đặc trưng kết cấu màu cục bộ (đã tìm được từ ảnh màu $\mathbf{S}_i$) là $\mathbf{x}_i$. Do đó, $\mathbf{x}_i$ có thể đại diện cho $\mathbf{x}_i^{\mathrm{CLBP}}$. Thêm vào đó, cho $\mathbf{I}^P$ là ảnh mặt màu RGB chưa biết (được nhận diện hoặc kiểm tra), nó được biểu thị như là một hình ảnh cần được xác minh. Cho $\mathbf{G}$ là một bộ ảnh bao gồm các mẫu thử đưa vào các ảnh mặt màu RGB (được nhận diện), mỗi hình ảnh được biểu hiện bởi $\mathbf{I}^G$ và được nhận biết như các mẫu riêng lẻ (ví dụ $\mathbf{I}^G \in \mathbf{G}$). Hơn nữa, giả sử không mất đi đặc điểm chung, các

đặc trưng kết cấu màu cục bộ cá thể của $\mathbf{I}^P$ và $\mathbf{I}^G$ lần lượt được biểu hiện bằng $\mathbf{x}_i^{\mathrm{P}}$ và $\mathbf{x}_i^{\mathrm{G}}$, trong đó $i = 1, \ldots, K$.

Các đặc trưng kết cấu màu cục bộ K được đưa ra để nhận diện $\mathbf{I}^P$, với mục tiêu đạt được chất lượng nhận diện khuôn mặt tốt nhất bằng việc kết hợp các đặc trưng dựa trên lý thuyết hợp nhất thông tin (Kittler et al., 1998) để phân loại mẫu. Các kỹ thuật hợp nhất đa dấu hiệu (ví dụ các kết quả phân loại đa dạng hoặc các tính năng đa dạng) có thể được phân loại chung thành 2 loại, ví dụ hợp nhất tại mức độ "đặc trưng" hoặc tại mức độ "quyết định". Trong phương pháp đề xuất, nhóm tác giả sử dụng chiến lược hợp nhất mức độ các nét đặc trưng vì 2 lý do: (1) Theo báo cáo trong (Jain et al., 2005), "hợp nhất các nét đặc trưng" có thể đạt được kết quả phân loại tốt hơn so với phương pháp thực hiện bằng "mức độ quyết định". (2) Theo thí nghiệm sử dụng phổ biến như là bộ dữ liệu FRCG 2.0, điều này chỉ ra phương pháp hợp nhất tại mức độ "đặc trưng" cho ra kết quả FR tốt hơn thực hiện bằng phương pháp "quyết định".

Theo quan sát được đề cập ở trên, kỹ thuật hợp nhất thông tin được thực hiện ở mức độ đặc trưng. Sử dụng phương pháp này, các đặc trưng kết cấu màu sắc cục bộ của $\mathbf{I}^P$ (hoặc $\mathbf{I}^G$) có thể liên kết thành vector đặc trưng toàn cục. Tuy nhiên, cần phải lưu ý rằng việc áp dụng trực tiếp phân loại láng giềng lân cận gần nhất (Neighbor nearest – NN) để liên kết các vector tính tăng có thể làm giảm chất lượng nhận diện khuôn mặt, do số chiều không gian lớn và các thông tin không cần thiết. Để giải quyết giới hạn trên, kỹ thuật trích xuất làm giảm số chiều đặc trưng (low-dimensional feature extraction techniques) được sử dụng. Biểu thị đặc trưng khuôn mặt thứ i được trích xuất bởi φ_i (PCA) để trích xuất đặc trưng sử dụng số lượng chiều thấp cho đặc trưng kết cấu màu cục bộ $\mathbf{x}_i$. Lưu ý rằng φ_i có thể hình thành với một tập huấn luyện các đặc trưng kết cấu màu cục bộ $\mathbf{x}_i$, tất cả chúng được tính từ thành phần màu thứ i và hình ảnh huấn luyện $\mathbf{S}_i$. Sau đó, các đặc trưng của $\mathbf{x}_i^{\mathrm{P}}$ và $\mathbf{x}_i^{\mathrm{G}}$ được tính như sau:

$$\mathbf{f}_i^P = \varphi_i\,(\mathbf{x}_i^P) \qquad\qquad \mathbf{f}_i^G = \varphi_i(\mathbf{x}_i^G)$$

trong đó $\mathbf{f}_i^P, \mathbf{f}_i^G \in \Re^{Di}$ và $\Re^{Di}$ biểu thị D_i chiều không gian thực và $i = 1, \dots, K$. Khi đó đặc trưng sử dụng kỹ thuật làm giảm số chiều ở mức độ đặc trưng bằng cách kết hợp các đặc trưng theo thứ tự cột.

$$f^P = [(f_1^P)^T \dots (f_K^P)^T]^T \qquad f^G = [(f_1^G)^T \dots (f_K^G)^T]^T$$

với $\mathbf{f}_i^P, \mathbf{f}_i^G \in \Re^D$ và $D = \sum_{i=1}^K D_i$

Để thực hiện nhiệm vụ nhận diện khuôn mặt FR (xác định hoặc xác minh), một phương pháp phân loại láng giềng gần nhất NN (Nearest Neighbor) được áp dụng để xác định danh tính của bằng cách xác định độ phù hợp gần nhất hoặc tương ứng (Choi et al., 2009).

CHƯƠNG 4: MÔ HÌNH ĐỀ XUẤT VÀ THỰC NGHIỆM

Hầu hết các phương pháp FR màu hiện có (bao gồm cả công việc trước đây và báo cáo của nhóm tác giả của Choi) đều bị hạn chế do việc sử dụng một cấu hình cố định thành phần màu sắc bao gồm chỉ "hai" hoặc "ba" thành phần màu (như YC_bC_r và YIQ). Đặc biệt, việc sử dụng màu sắc lựa chọn thành phần chủ yếu hiện đang được thực hiện thông qua sự kết hợp giữa khả năng trực giác và so sánh thực nghiệm mà không có bất kỳ chiến lược lựa chọn hệ thống. Như vậy, các phương pháp hiện tại có thể có một giới hạn để đạt được kết quả tốt nhất cho cho nhiệm vụ FR. Điều này là do các thành phần màu sắc cụ thể có hiệu quả đối với một công việc nhận dạng này nhưng không thể làm việc tốt theo điều kiện nhận dạng khác (ví dụ, biến thể chiếu sáng) khác với những người xem xét trong quá trình xác định các thành phần màu sắc cụ thể. Do đó, vấn đề quan trọng trong bài toán nhận diện khuôn mặt màu là: làm thế nào người ta có thể chọn các thành phần màu sắc từ các mô hình màu khác nhau để đạt được hiệu quả tốt nhất cho một bài toán nhận diện khuôn mặt cụ thể.

4.1 Mô hình đề xuất

Trong phần kết luận, nhóm tác giả của Choi đã chỉ ra một số hạn chế trong phương pháp xây dựng CLTF được nêu trong công trình này. Một trong những hạn chế đó là, khi xây dựng CLTF, các đặc trưng kết cấu cục bộ thành phần được trích chọn từ cùng một không gian màu gồm 3 thành phần (ví dụ RQC_r). Xử lý theo cách này sẽ không tận dụng được điểm mạnh của các không gian màu khác nhau. Theo (Stokman and Gevers, 2007), mỗi không gian màu có các đặc điểm riêng có thể được tận dụng để làm tăng độ chính xác nhận dạng.

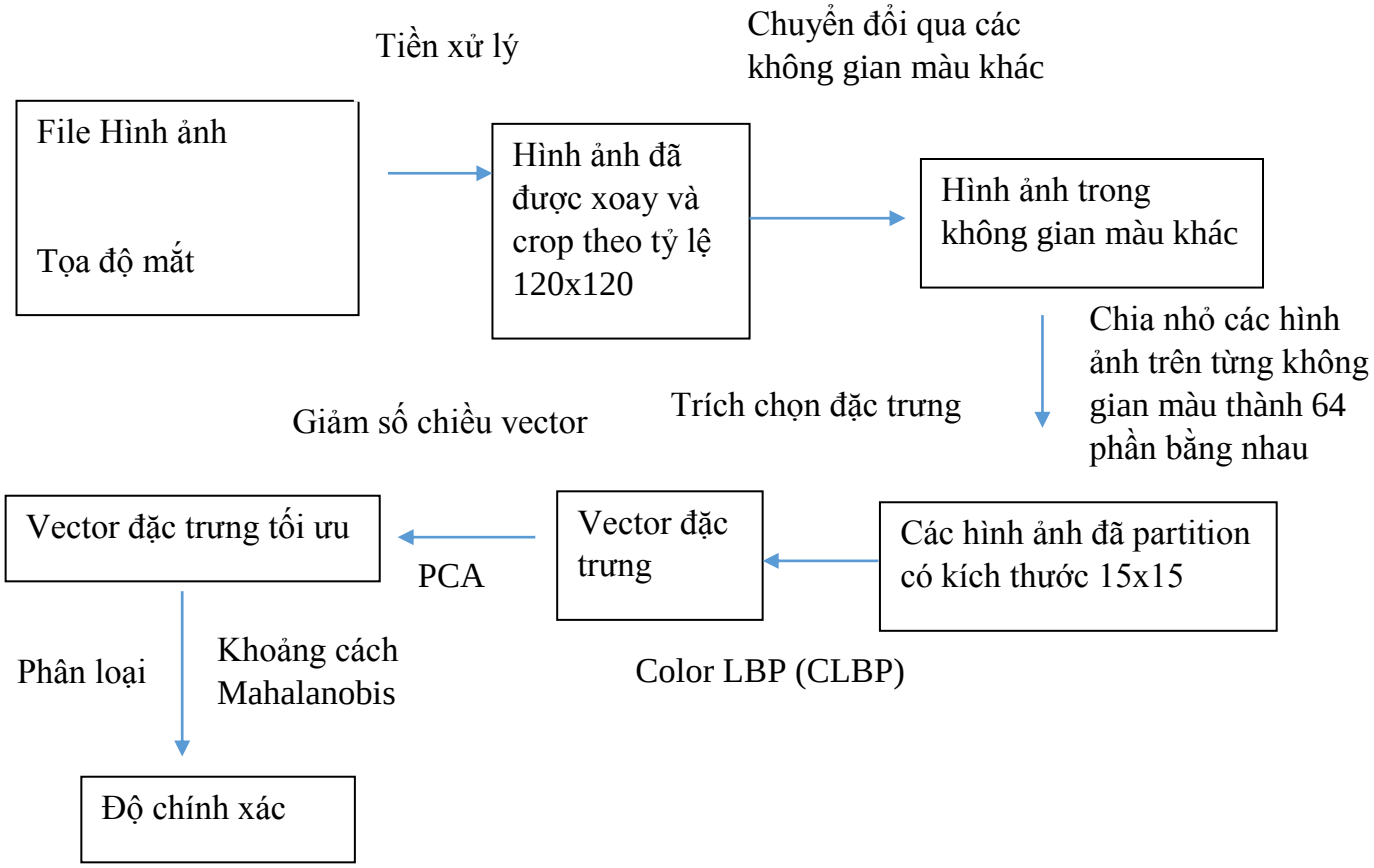
Bằng cách sử dụng giải pháp được mô tả trong (Choi et al., 2011) (multiclass boosting color feature selection). Mô hình mới đề xuất sẽ trộn lẫn các kênh màu từ các list màu chính để tìm ra màu sắc tốt hơn cho bài toán nhận diện khuôn mặt.

Tuy nhiên, do quy mô bài toán trong nghiên cứu của Choi và các đồng sự là khá lớn, nhóm tác giả Choi sử dụng đồng thời hai phương pháp CLBP và CLGW để trích xuất đặc trưng, cũng như tiến hành thực nghiệm trên khá nhiều bộ dữ liệu ví

dụ như CMU-PIE, Color FERET, XM2VTSDB, SCface, FRGC 2.0 với nhiều độ phân giải khác nhau, cũng như sử dụng nhiều kỹ thuật làm giảm số chiều cho vector đặc trưng như PCA, FLDA (Fisher's linear discriminant analysis), EFM (Enhanced Fisher linear discriminant Model), ERE (eigenfeature regularization and extraction), KDDA. Do đó, tác giả chỉ có thể sử dụng phương pháp trích chọn đặc trưng kết cấu màu cục bộ bằng phương pháp sử dụng mẫu nhị phân màu cục bộ (CLBP) để phân tích, so sánh, đồng thời sử dụng bộ dữ liệu Color Feret và kỹ thuật làm giảm số chiều vector bằng PCA.

Tương tự mô hình đề xuất thuật toán của Choi và các đồng sự. Mô hình cải tiến này gồm 3 bước chính:

- Chuyển đổi các không gian màu và chia nhỏ hình ảnh.
- Trích xuất đặc trưng kết cấu màu cục bộ.
- Kết hợp, phân loại và kiểm tra độ chính xác.



Hình 4.1 Mô hình đề xuất cải tiến

4.2. Các không gian màu lựa chọn.

4.2.1 Không gian màu *YIQ*

YIQ là không gian màu được sử dụng bởi các hệ thống truyền hình màu NTSC, chủ yếu ở khu vực Bắc, Trung Mỹ, và Nhật Bản. *I* là viết tắt của in-phase, trong khi *Q* là viết tắt của phép cầu phương – quadrature, đề cập đến các thành phần được sử dụng trong mô hình độ lớn cầu phương (quadrature amplitude modulation). Một số dạng NTSC hiện nay sử dụng không gian màu *YUV*, một số sử dụng các hệ thống khác như PAL .

Thành phần *Y* đại diện cho các thông tin luma, và là thành phần chỉ được sử dụng bởi máy thu hình đen trắng. *I* và *Q* đại diện cho thông tin màu. Trong *YUV*, các thành phần *U* và *V* có thể được coi như là *X* và Y tọa độ trong không gian màu. *I* và *Q* có thể được coi như là một cặp thứ hai của trục trên cùng một đồ thị, xoay 33°; do đó chỉ số *IQ* và *UV* đại diện cho các hệ tọa độ khác nhau trên cùng một mặt phẳng.

Hệ thống *YIQ* được thiết kế để tận dụng lợi thế của màu sắc đặc trưng phản xạ của con người. Mắt là nhạy cảm hơn với những thay đổi trong dải màu cam-xanh (orange – blue) (*I*) nhiều hơn trong dải màu tím-xanh (purple-green)(*Q*) - do đó *Q* đòi hỏi ít băng thông hơn *I*. Băng tần NTSC giới hạn I đến 1,3 MHz và Q đến 0,4 MHz . I và Q là tần số xen kẽ vào tín hiệu Y 4 MHz, và giữ các băng thông của tín hiệu tổng thể xuống đến 4,2 MHz. Trong các hệ thống *YUV*, *U* và *V* đều chứa thông tin trong phạm vi dải màu cam – xanh, cả hai thành phần phải được đưa ra cùng một lượng băng thông như *I* để đạt được độ trung thực màu sắc tương tự.

Trong xử lý hình ảnh, không gian màu *YIQ* đôi khi được sử dụng trong việc chuyển đổi màu sắc hình ảnh. Ví dụ, áp dụng một cân bằng histogram trực tiếp đến các kênh màu trong một hình ảnh RGB sẽ làm thay đổi sự cân bằng màu sắc của hình ảnh. Thay vào đó, việc cân bằng histogram được áp dụng cho các kênh Y của *YIQ* hoặc *YUV* của hình ảnh, mà chỉ chuẩn hóa (normalize) mức độ sáng của hình ảnh.

Công thức biến đổi

$R, G, B, Y \in [0,1], I \in [-0.5957,0.5957], Q \in [-0.5226,0.5226]$

Từ RGB thành *YIQ*

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.528 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Từ *YIQ* thành RGB

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0.956 & 0.621 \\ 1 & -0.272 & -0.647 \\ 1 & -1.108 & 1.705 \end{bmatrix} \begin{bmatrix} Y \\ I \\ Q \end{bmatrix}$$

4.2.2 Không gian màu *ZRG*

Là không gian màu lai được ghép từ các thành phần màu của các không gian màu *XYZ* và *RGB*, Z là thành phần màu từ không gian màu *XYZ*, R, G là thành phần màu từ không gian màu *RGB*.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.000 & 1.116 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

4.2.3 Không gian màu RC_rQ

Là không gian màu lai được ghép từ các thành phần màu của các không gian màu *RGB*, YC_bC_r và *YIQ*, R là thành phần màu từ không gian màu *RGB*, C_r là thành phần màu lấy từ không gian màu YC_bC_r và *Q* lấy từ không gian màu *YIQ*.

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

4.2.4 Kết hợp các không gian màu

Các không gian màu chính: *YIQ*, *ZRG*, QC_bC_r.

List = ['Y','I','Q','Z','R','G','b','r'];

Các không gian màu mới được kết hợp bằng công thức tổ hợp không lặp chập k có thứ tự của n phần tử. Ở đây là tổ hợp không lặp chập 3 của 8 phần tử

$$C_n^k = \frac{n!}{k!\,(n-k)!} = \frac{8!}{3!\,(8-3)!} = 56$$

Danh sách các không gian màu kết hợp:

YIQ YIZ YIR YIG YIb YIr YQZ YQR YQG YQb YQr
YZR YZG YZb YZr YRG YRb YRr YGb YGr Ybr IQZ
IQR IQG IQb IQr IZR IZG IZb IZr IRG IRb IRr
IGb IGr Ibr QZR QZG QZb QZr QRG QRb QRr QGb
QGr Qbr ZRG ZRb ZRr ZGb ZGr Zbr RGb RGr Rbr
Gbr

4.3 Cơ sở dữ liệu ảnh màu Color Feret

4.3.1 Giới thiệu

Color FERET là bộ dữ liệu hình ảnh được thực hiện bởi National Institute of Standards and Technology (NIST). Bộ cơ sở dữ liệu này chứa tổng cộng 11.338 ảnh màu mặt người được thu thập từ việc chụp 994 người ở các góc độ khác nhau từ năm 1993 đến năm 1996. Bộ dữ liệu này là phiên bản ảnh màu của bộ cơ sở dữ liệu FERET gốc được thực hiện vào năm 2001 với 14.051 bức ảnh grayscale.

Các bức ảnh trong bộ Color FERET có kích thước 512 x 768 pixel, và ở định dạng PPM. Có 13 hướng chụp khác nhau trong bộ cơ sở dữ liệu này.

fa: regular frontal image: Hướng chụp phía trước mặt thông thường

fb: alternative frontal image: Hướng chụp phía trước mặt, được chụp ngay sau hình fa tương ứng

pl: profile left: Hướng chụp về bên trái

hl: half left: Hướng chụp chếch về bên trái, đầu quay 1 góc 67.5 độ

ql: quarter left: Hướng chụp chếch về bên trái, đầu quay 1 góc 22.5 độ

pr: profile right: Hướng chụp về bên phải

hr: half right - Hướng chụp chếch về bên phải, đầu quay 1 góc 67.5 độ

qr: quarter right - Hướng chụp chếch về bên phải, đầu quay 1 góc 22.5 độ

ra: random image – Hình ảnh ngẫu nhiên, đầu quay về bên trái 45 độ

rb: random image - Hình ảnh ngẫu nhiên, đầu quay về bên trái 15 độ

rc: random image - Hình ảnh ngẫu nhiên, đầu quay về bên phải 15 độ

rd: random image - Hình ảnh ngẫu nhiên, đầu quay về bên phải 45 độ

re: random image - Hình ảnh ngẫu nhiên, đầu quay về bên phải 75 độ

Bảng 4.1 Bảng số lượng ảnh trong tập Color Feret

Hướng ảnh	Góc chụp	Số lượng ảnh	Số lượng chủ thể
fa	0	1364	994
fb	0	1358	993
hl	+67.5	1267	917
hr	-67.5	1320	953
pl	+90	1312	960
pr	-90	1363	994
ql	+22.5	761	501
qr	-22.5	761	501
ra	+45	321	261
rb	+15	321	261
rc	-15	610	423
rd	-45	290	236
re	-75	290	236

4.3.2 Quy tắc đặt tên hình ảnh

data/images/00012/00012_930831_fb_a.ppm.bz2

| | | | |

| | | | _____ optional flag

| | | _____ FERET pose name

| | _____ the image capture date

| _____ the subject ID

Đối với một số bức ảnh, tên file sẽ chứa ký tự cờ (flag).

- a Chủ thể bức ảnh có đeo mắt kiếng
- b Chủ thể thay đổi kiểu tóc và không đeo mắt kiếng

- c Chủ thể thay đổi kiểu tóc và đeo mắt kiếng

4.3.2 Ground Truth File

Là một văn bản thô (plain text) hoặc file XML chứa thông tin cần thiết cho mỗi chủ thể bức ảnh về giới tính, chủng tộc, năm sinh, vv.

Ví dụ: Thông tin về chủ thể ‘00012’ là đường dẫn

data/ground_truths/name_value/00012/00012.txt

data/ground_truths/xml/00012/00012.xml

Đối với khoảng 90% bức ảnh Color FERET có chứa thông tin tọa độ của mắt, mũi, miệng của chủ thể.

Ví dụ: Thông tin về chủ thể ‘00012_930831_fa’ là:

```
recording=cfrR00118
subject=cfrS00012
disc=1
relative=data/images/00012/00012_930831_fa.ppm.bz2
format=ppm
compression=bzip2
capture_date=08/31/1993
capture_time=00:00:00
pose=fa
expression=fa
yaw=0
pitch=0
roll=0
glasses=No
beard=No
mustache=No
left_eye_coordinates=348 350
right_eye_coordinates=242 344
nose_coordinates=296 390
mouth_coordinates=288 466
stage=cfrT00001
collection=cfrC00001
environment=cfrE00001
sensor=cfrN00002
illuminant=cfrI00001 cfrI00002 cfrI00003
weather=inside
```

4.4 Cơ sở lý thuyết PCA

4.4.1 Giới thiệu

Phương pháp phân tích thành phần chính (Principal Component Analysis – PCA) là một trong những phương pháp phân tích dữ liệu nhiều biến đơn giản nhất. Trong các bài toán thống kê, thông thường cần phải "nghiên cứu" dữ liệu trước khi xây dựng các mô hình suy diễn dựa trên dữ liệu đó. Tuy nhiên đôi khi dữ liệu có số chiều lớn, không thể suy diễn trong không gian có số chiều lớn. Do đó, cần phải tìm cách đưa dữ liệu về không gian có số chiều nhỏ hơn.

PCA là một trong những phương pháp như thế, nhưng hơn thế, PCA còn thể hiện sự đa năng với nhiều đặc tính tốt:

Làm giảm số chiều của dữ liệu. Thay vì giữ lại các trục tọa độ của không gian cũ, PCA xây dựng một không gian mới ít chiều hơn, nhưng lại có khả năng biểu diễn dữ liệu tốt tương đương không gian cũ, nghĩa là đảm bảo *độ biến thiên* (variability) của dữ liệu trên mỗi chiều mới.

Các trục tọa độ trong không gian mới là tổ hợp tuyến tính của không gian cũ, do đó về mặt ngữ nghĩa, PCA xây dựng feature mới dựa trên các feature đã quan sát được. Điểm hay là những feature này vẫn biểu diễn tốt dữ liệu ban đầu.

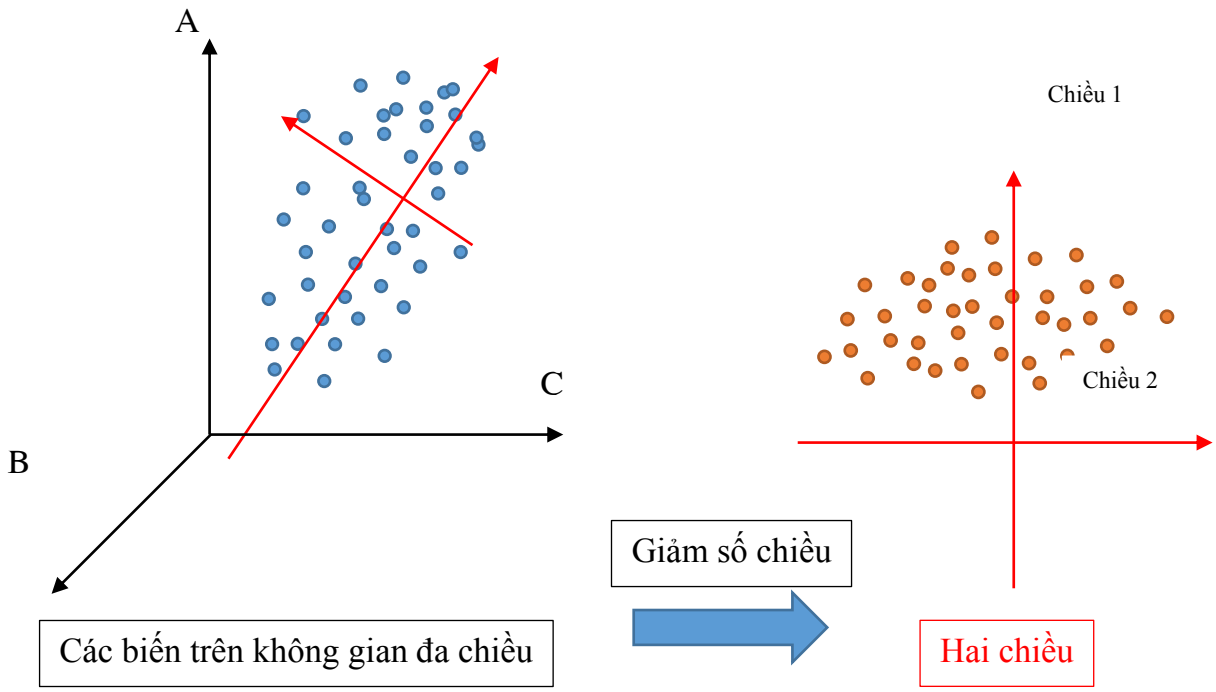
Trong không gian mới, các liên kết tiềm ẩn của dữ liệu có thể được khám phá, mà nếu đặt trong không gian cũ thì khó phát hiện hơn, hoặc những liên kết như thế không thể hiện rõ.

Nói một cách ngắn gọn hơn, mục tiêu của PCA là tìm một không gian mới (với số chiều nhỏ hơn không gian cũ). Các trục tọa độ của không gian mới được xây dựng sao cho trên mỗi trục, độ biến thiên của dữ liệu trên đó là lớn nhất có thể (maximize the variability).

Hình 4.2 Ví dụ mô tả về số lượng chiều của dữ liệu.

Một ví dụ kinh điển là về hình ảnh con lạc đà. Cùng là một con lạc đà, nhưng nếu nhìn từ bên hông thì ta có được đầy đủ thông tin nhất, trong khi nếu nhìn từ phía trước thì thật khó để xác định được đó là con lạc đà.

Một ví dụ thuyết phục hơn được minh họa trong hình sau:



Hình 4.3 Minh họa sự biến thiên dữ liệu trong không gian đa chiều trước và sau khi sử dụng thuật toán PCA

Giả sử tập dữ liệu ban đầu (tập điểm màu xanh) được quan sát trong không gian 3 chiều (trục màu đen) như hình bên trái. Rõ ràng 3 trục này không biểu diễn được tốt nhất mức độ biến thiên của dữ liệu. PCA do đó sẽ tìm hệ trục tọa độ mới (là hệ trục màu đỏ trong hình bên trái). Sau khi tìm được không gian mới, dữ liệu sẽ được chuyển sang không gian này để được biểu diễn như trong hình bên phải. Rõ ràng hình bên phải chỉ cần 2 trục tọa độ nhưng biểu diễn tốt hơn độ biến thiên của dữ liệu so với hệ trục 3 chiều ban đầu.

Một điểm nữa của PCA là các trục tọa độ trong không gian mới luôn đảm bảo trực giao đôi một với nhau, mặc dù trong không gian ban đầu, các trục có thể không trực giao.

4.1.2 Thuật toán PCA

Cho ma trận $= \{x_{ij}\} \in \mathcal{R}^{n \times p}$. Các bước của PCA lần lượt như sau:

- **Tiền xử lý**

Dữ liệu ban đầu có thể có giá trị thay đổi bất thường. Ví dụ trên feature 1 (cột 1 của $\mathbf{X}$) giá trị thay đổi trong khoảng (0, 1), trên feature 2 lại biến thiên trong đoạn (-100, 100). Rõ ràng cần phải có một bước tiền xử lí để chuẩn hóa giá trị trên các cột của ma trận X. Có 2 cách tiền xử lí thường được dùng cho PCA là *Centered PCA* và *Normed PCA*.

Centered PCA mang tất cả các feature (các cột của X) về cùng một gốc tọa độ:

$$\hat{X} = \{\hat{x}_{ij}\}$$

$$\hat{x}_{ij} = \frac{x_{ij} - g_j}{\sqrt{n}}$$

$$g_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$$

Trong đó n là số dòng của X, g_j là mean của cột thứ j của X, được tính như trên.

Normed PCA mang tất cả các feature về cùng một gốc tọa độ, đồng thời chuẩn hóa về cùng một quãng standard-deviation bằng 1:

$$\hat{X} = \{\hat{x}_{ij}\}$$

$$\hat{x}_{ij} = \frac{x_{ij} - g_j}{\sqrt{n}\sigma_j}$$

Trong đó σ_j là *độ lệch chuẩn* (*standard deviation*) của cột thứ j trong X.

Thông thường Normed PCA hay được dùng. Sau bước tiền xử lí, ma trận $\hat{X}$ sẽ là đầu vào cho bước tiếp theo.

- **Xây dựng không gian mới**

Tính ma trận hiệp phương sai (covariance) của các feature trong $\hat{X}$:

$$V = \hat{X}^T \hat{X}$$

Do là tích của ma trận $\hat{X}$ với chuyển vị của nó nên $V \in \mathcal{R}^{p \times p}$ là ma trận positive semidefinite kích thước $p \times p$. Hơn nữa V có p trị riêng $\lambda_i \geq 0, i = 1 \dots p$ .Tiếp theo, PCA tìm trị riêng và vector riêng tương ứng của V, sắp xếp theo thứ tự giảm dần của trị riêng. Giả sử p trị riêng của V là:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$$

và p vector riêng tương ứng là

$$\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_p$$

Khi đó các trục của không gian mới chính là các vector riêng $\mathbf{u}_i$ở trên, đương nhiên các vector riêng hoàn toàn độc lập tuyến tính (nghĩa là trực giao đôi một).

Chuyển dữ liệu từ không gian ban đầu vào không gian mới

Thông thường không gian mới không được xây dựng bằng tất cả *p* vector riêng trong (4), mà thông thường chỉ từ *k* vector riêng đầu tiên, với k $<$ p. Tại sao là các vector đầu tiên, và chọn k bao nhiêu thì tốt, ta sẽ bàn trong phần cuối.

Như vậy gọi

$$\mathbf{U} = [u_1|u_2| \ldots |u_k] \in \mathcal{R}^{p \times k}$$

Khi đó tọa độ các điểm trong hệ tọa độ mới là:

$$\mathbf{F} = \widehat{\mathbf{X}}\mathbf{U}$$

- **Một số hạn chế của PCA:**

- Chỉ làm việc với dữ liệu numeric,
- Nhạy cảm với các điểm outlier/extreme,
- Không phù hợp với các mô hình phi tuyến, do PCA hoàn toàn dựa trên các biến đổi tuyến tính.

4.5 Khoảng cách Mahalanobis

4.5.1 Giới thiệu

Khoảng cách Mahalanobis là một thước đo khoảng cách giữa một điểm P và bộ phân phối chuẩn D, được giới thiệu bởi P. C. Mahalanobis vào năm 1936.(“Mahalanobis distance,” 2015). Nó là một sự tổng quát đa chiều của việc có bao nhiêu độ lệch chuẩn (standard deviations) ra khỏi P từ bình quân (mean) D. Khoảng cách này là 0 nếu P là mean của D, và khoảng cách này càng ngày càng lớn khi P càng cách xa khỏi mean.

Dọc theo mỗi trục thành phần chính (PCA), nó đo lường số lượng độ lệch chuẩn từ P đến mean của D. Nếu mỗi trục được thay đổi tỷ lệ để có đơn vị có

phương sai thì khoảng cách Mahalanobis tương đương với khoảng cách Euclide tiêu chuẩn trong các không gian đã chuyển đổi.

4.5.2 Định nghĩa và tính chất

Khoảng cách Mahalanobis của một observation $x = (x_1, x_2, x_3, \dots, x_N)^T$ từ một tập hợp các observation với với mean $\mu = (\mu_1, \mu_2, \mu_3, \dots, \mu_N)^T$và ma trận hiệp phương sai S được định nghĩa như sau:

$$D_M(x) = \sqrt{\{(x-\mu)^T S^{-1} (x-\mu)\}}.$$

Khoảng cách Mahalanobis có thể được định nghĩa như là một độ đo tương tự giữa hai vectơ ngẫu nhiên $\vec{x}$ và $\vec{y}$ của cùng một phân phối với các ma trận hiệp biến S:

$$d(\vec{x}, \vec{y}) = \sqrt{\left((\vec{x}-\vec{y})^T S^{-1} (\vec{x}-\vec{y})\right)}$$

Nếu ma trận hiệp phương sai là ma trận nhận dạng (identity matrix), khoảng cách Mahalanobis làm giảm khoảng cách Euclide. Nếu ma trận hiệp phương sai là đường chéo (diagonal), thì độ đo khoảng cách thu được được gọi là khoảng cách Euclide chuẩn hóa (normalized Euclidean distance):

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^{N} \frac{(x_i - y_i)^2}{s_i^2}}$$

Với s_i là độ lệch chuẩn của x_i và y_i trên tập mẫu.

4.6 Mô tả thí nghiệm

Sử dụng chương trình Matlab và bộ dữ liệu Color Feret với 5429 dữ liệu hình chủ yếu ở dup fa, fb và rc của 1028 chủ thể. Các hình ảnh này trước tiên sẽ được tinh chỉnh và cắt gương mặt từ hình ảnh gốc dựa vào vị trí tọa độ mắt. Các vị trí tọa độ này được cung cấp kèm theo bộ dữ liệu và sau đó sẽ được điều chỉnh về kích thước 120x120 pixel.

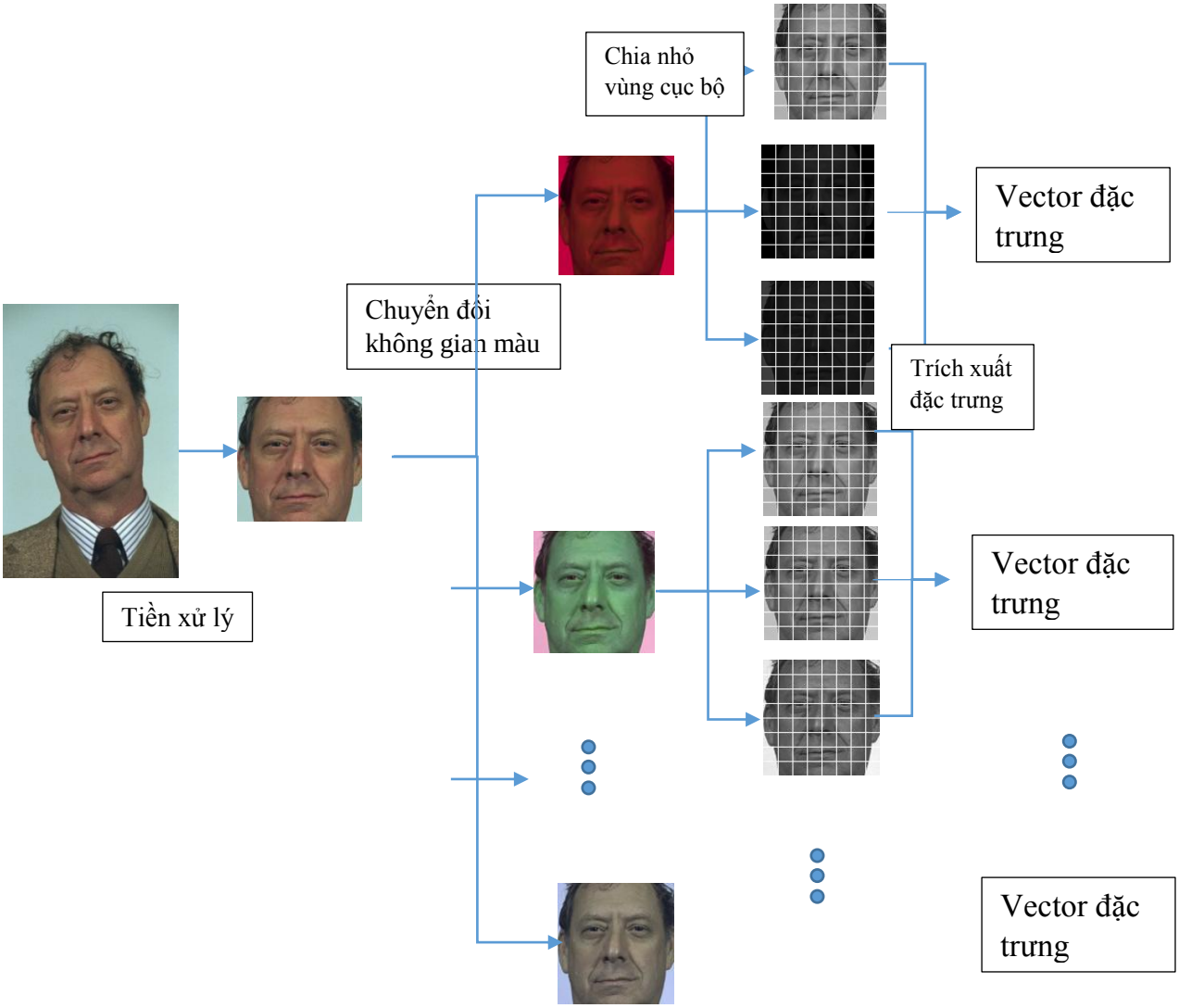
Sau bước tiền xử lý, mỗi hình ảnh mặt người sẽ được chuyển đổi về các không gian màu khác nhau. Trong báo cáo này, các hình ảnh mặt người sẽ được

chuyển đổi lần lượt thành 56 không gian màu và được tách thành các không gian màu đơn sắc (gray scale). Để chuyển đổi các thành phần không gian màu. Các thành phần không gian màu được lựa chọn: *YIQ*, *ZRG*, YC_bC_r. Lưu ý không gian màu *ZRG* (Z được lấy từ không gian màu "*XYZ*", R và G được lấy từ "*RGB*").

Tại đây, mỗi hình ảnh màu đơn sắc 120 x 120 pixel này sẽ được chia nhỏ thành 64 vùng cục bộ bằng nhau (mỗi vùng có kích thước 15 x 15 pixel) và tính toán đặc trưng kết cấu trên từng vùng cục bộ bằng phương pháp CLBP đã nêu ở mô hình đề xuất, sử dụng toán tử LBP(8,1) (P=8, R=1). Từ đây, các vector đặc trưng trở thành vector đặc trưng kết cấu màu cục bộ.

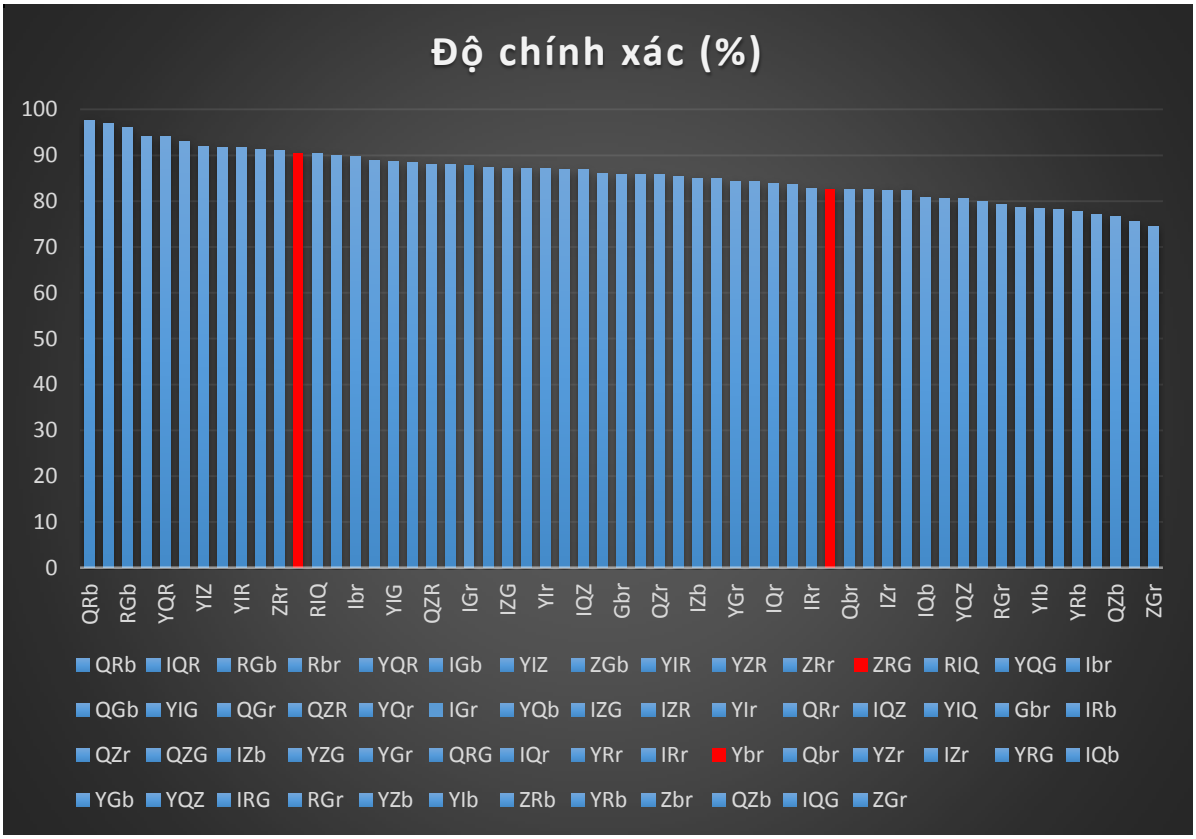
Để giảm số chiều đặc trưng màu cục bộ và làm tăng tính phân biệt, phương pháp PCA được áp dụng trên các vector đặc trưng. Do không gian làm việc (workspace) của chương trình Matlab có hạn, nên tất cả dữ liệu vector đặc trưng của tập dữ liệu hình ảnh trên 56 không gian màu sau khi được xử lý bằng kỹ thuật giảm số chiều sẽ được lưu riêng biệt thành file *.mat bên ngoài để tránh tình trạng tràn dữ liệu dẫn đến sai lệch kết quả khi tính toán và sau đó sử dụng độ đo khoảng cách Mahalanobis để xác định hình ảnh cần tìm gần nhất.

Kết quả thu được từ các kênh màu trộn lẫn được thống kê và so sánh với thuật toán gốc của Mr.Choi để xác định màu tối ưu cho bài toán nhận dạng ảnh màu mặt người.



Hình 5.1: Mô hình thực nghiệm

4.7 Kết quả thực nghiệm



Hình 5.2: Đồ thị mô tả độ chính xác trên các không gian màu

4.8 Đánh giá thực nghiệm

Qua hơn 7 ngày thực hiện chương trình, kết quả cho ra được độ chính xác trên từng không gian màu. Dựa vào kết quả thực nghiệm, có thể thấy một số không gian màu cho kết quả nhận diện tốt hơn so với các không gian màu cố định mà nhóm nghiên cứu của Choi và các đồng sự đã sử dụng trong (Choi et al., 2012) (*ZRG* và YC_bC_r). Một số không gian màu cho kết quả nhận diện tốt như *RQCb, RIQ, RGCb* cho kết quả khá tốt (hơn 96%).

CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

5.1 Kết luận

Thuật toán nhận dạng khuôn mặt bằng phương pháp trích chọn đặc trưng kết cấu màu cục bộ của (Choi et al., 2012) đã cho kết quả nhận dạng khuôn mặt rất tốt trong các điều kiện cụ thể (ví dụ như thiếu sáng hay các hình ảnh có độ phân giải thấp). Tuy nhiên thuật toán cũng có mặt hạn chế khi sử dụng một cấu hình có thành phần màu sắc cố định như YC_bC_r và *ZRG* (Choi et al., 2012).

Nhằm cải tiến các hạn chế trên, qua quá trình thực nghiệm, báo cáo của tác giả đã đóng góp một phần nhỏ vào việc nâng cao hiệu suất nhận dạng khuôn mặt bằng việc kết hợp các không gian màu và tìm kiếm một vài không gian màu cho kết quả nhận diện khuôn mặt khá tốt với độ chính xác cao, cụ thể như *RIQ*, RQC_b. Có thể thấy các không gian màu khác nhau sẽ cho độ chính xác nhận diện khuôn mặt khác nhau.

Tuy nhiên, báo cáo này vẫn còn một số mặt hạn chế như: (1) Phạm vi bài toán chỉ khai thác được một phần trong trích chọn đặc trưng của Choi và các đồng sự bằng phương pháp sử dụng Mẫu đặc trưng màu cục bộ (Color LBP) nên hiệu suất bài toán nhận dạng khuôn mặt chưa thật sự chính xác. (2) Hạn chế về mặt thời gian không cho phép tác giả thực hiện thuật toán trên nhiều bộ cơ sở dữ liệu nên kết quả còn chưa chính xác và còn mang tính chất tương đối. Trong tương lai, để cải tiến hiệu suất nhận dạng khuôn mặt (độ chính xác tăng ~ 98%), tác giả sẽ kết hợp với việc trích xuất kết cấu màu cục bộ, chẳng hạn như kết hợp với phương pháp Color Gabor Wavelet nhằm tăng tính phân biệt cho các đặc trưng kết cấu cục bộ hoặc sử dụng trọng số cho các kênh màu, từ đó xác định được không gian màu sắc tối ưu một cách tự động cho các trường hợp cụ thể khi nhận diện khuôn mặt (ví dụ: sử dụng không gian màu *RQCr* cho ảnh khuôn mặt ở độ sáng thấp, hoặc sử dụng không gian màu *XYZ* nếu ảnh khuôn mặt ở độ phân giải thấp,…)

5.2 Thách thức và hướng phát triển

Hiện nay vẫn còn nhiều các thách thức đối với đề tài:

+ Vấn đề hướng (pose variations): Các kết quả với các ảnh có hướng thay đổi (>45 độ, không phải chính diện) còn khá khiêm tốn và thiếu sự chính xác. Nhận dạng hình ảnh 3D có lẽ là một hướng giải quyết tốt về vấn đề này.

+ Vấn đề ảnh có độ phân giải thấp (low resolution): Ảnh thu được từ các camera giám sát (surveillance camera) thường có kích thước và chất lượng rất thấp, và kết quả nhận diện khuôn mặt với độ phân giải thấp còn thiếu chính xác.

+ Làm việc với dữ liệu video (video based face recognition): Với sự phát triển của các phương tiện multimedia, thông tin mặt người trong các dữ liệu video là vô cùng nhiều, tuy nhiên hầu hết các phương pháp nhận dạng vẫn làm việc với ảnh tĩnh trích xuất từ dữ liệu video, chưa có phương pháp tốt tận dụng hết ưu thế của dữ liệu video, một phương pháp nhận dạng thời gian thực có thể là một hướng đi mà đề tài này theo đuổi.

+ Vấn đề với các hệ thống cực lớn (very large scale systems): cơ sở dữ liệu (CSDL) ảnh mặt người được thực hiện bởi đề tài này còn khá nhỏ (vài chục nghìn ảnh mặt), tuy nhiên trên thực tế các CSDL có thể rất lớn, ví dụ CSDL ảnh mặt của cảnh sát của một nước có thể chứa từ hàng triệu tới hơn 1 tỉ ảnh, thời gian xử lý ảnh trên CSDL như vậy sẽ mất rất nhiều thời gian và công sức.

+ Vấn đề về tuổi tác **(**aging condition**)**: việc nhận dạng ảnh mặt thay đổi theo thời gian thực sự vẫn còn là một vấn đề lớn ngay cả đối với khả năng nhận dạng của con người.

+ Vấn đề về sự thiếu sáng (illumination)**:** là một trong những thách thức lớn của nhận dạng mặt, chưa có phương pháp tốt cho các ảnh chụp ở điều kiện ngoài trời.

TÀI LIỆU THAM KHẢO

Ahonen, T., Hadid, A., Pietikainen, M., (2006). *Face Description with Local Binary Patterns: Application to Face Recognition*. IEEE Trans. Pattern Anal. Mach. Intell. 28, 2037–2041. doi:10.1109/TPAMI.2006.244

Bowyer, K.W., (2004). *Face recognition technology: security versus privacy*. IEEE Technol. Soc. Mag. 23, 9–19. doi:10.1109/MTAS.2004.1273467

Chaudhuri, B.B., Rosenfeld, A., (1999). *A modified Hausdorff distance between fuzzy sets*. Inf. Sci. 118, 159–171. doi:10.1016/S0020-0255(99)00037-7

Choi, J.Y., De Neve, W., Ro, Y.M., Plataniotis, K.N., (2010). *Automatic Face Annotation in Personal Photo Collections Using Context-Based Unsupervised Clustering and Face Information Fusion*. IEEE Trans. Circuits Syst. Video Technol. 20, 1292–1309. doi:10.1109/TCSVT.2010.2058470

Choi, J.Y., Ro, Y.M., Plataniotis, K.N., (2012). *Color Local Texture Features for Color Face Recognition*. Trans Img Proc 21, 1366–1380. doi:10.1109/TIP.2011.2168413

Choi, J.Y., Ro, Y.M., Plataniotis, K.N., (2011). *Boosting Color Feature Selection for Color Face Recognition*. IEEE Trans. Image Process. 20, 1425–1434. doi:10.1109/TIP.2010.2093906

Choi, J.Y., Ro, Y.M., Plataniotis, K.N., (2009). *Color Face Recognition for Degraded Face Images*. Trans Sys Man Cyber Part B 39, 1217–1230. doi:10.1109/TSMCB.2009.2014245

Wikipedia Free Encycl (2015) CMYK color model, [online] from: http://en.wikipedia.org/w/index.php?title=CMYK_color_model&oldid=662358911

Wikipedia Free Encycl (2015) .HSL and HSV, [online] from: http://en.wikipedia.org/w/index.php?title=HSL_and_HSV&oldid=657413113

Wikipedia Free Encycl (2015) International Commission on Illumination, [online] from: http://en.wikipedia.org/w/index.php?title=International_Commission_on_Illumination&oldid=661018747

Wikipedia Free Encycl. (2015). Mahalanobis distance, [online] from: http://en.wikipedia.org/w/index.php?title=Mahalanobis_distance&oldid=661196628

Ojala, T., Pietikainen, M., Maenpaa, T., (2002). *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Trans. Pattern Anal. Mach. Intell. 24, 971–987. doi:10.1109/TPAMI.2002.1017623

Pękalska, E., Duin, R.P.W., (2005). *The Dissimilarity Representation for Pattern Recognition*: Foundations and Applications. World Scientific.

Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J., (2000). *The FERET Evaluation Methodology for Face-Recognition Algorithms*. IEEE Trans Pattern Anal Mach Intell 22, 1090–1104. doi:10.1109/34.879790

Wikipedia Free Encycl (2015) RGB color model, [online] from: http://en.wikipedia.org/w/index.php?title=RGB_color_model&oldid=660520690

PHỤ LỤC

Bảng thống kê độ chính xác trên tập dữ liệu của các không gian màu

Không gian màu	Độ chính xác (%)	Số hình ảnh nhận diện đúng	Số hình ảnh nhận diện sai	Số lượng ảnh tập Probe	Số lượng ảnh tập Gallery
QRb	97.69	2483	59	2542	5433
IQR	97.06	2467	75	2542	5433
RGb	96.07	2442	100	2542	5433
Rbr	94.2	2395	147	2542	5433
YQR	94.09	2392	150	2542	5433
IGb	93.1	2367	175	2542	5433
YIZ	91.92	2337	205	2542	5433
ZGb	91.81	2334	208	2542	5433
YIR	91.61	2329	213	2542	5433
YZR	91.21	2319	223	2542	5433
ZRr	90.99	2313	229	2542	5433
RIQ	90.4	2298	244	2542	5433
YQG	90.05	2289	253	2542	5433
Ibr	89.86	2284	258	2542	5433
QGb	88.91	2260	282	2542	5433
YIG	88.6	2252	290	2542	5433
QGr	88.36	2246	296	2542	5433
QZR	88.07	2239	303	2542	5433
YQr	88.01	2237	305	2542	5433
IGr	87.85	2233	309	2542	5433
ZRG	87.46	2223	319	2542	5433
YQb	87.37	2221	321	2542	5433
IZG	87.23	2217	325	2542	5433
IZR	87.19	2216	326	2542	5433
YIr	87.06	2213	329	2542	5433
QRr	86.85	2208	334	2542	5433
IQZ	86.84	2207	335	2542	5433
YIQ	85.95	2185	357	2542	5433
Gbr	85.91	2184	358	2542	5433
IRb	85.88	2183	359	2542	5433
QZr	85.72	2179	363	2542	5433
QZG	85.4	2171	371	2542	5433

IZb	84.97	2160	382	2542	5433
YZG	84.92	2159	383	2542	5433
YGr	84.32	2143	399	2542	5433
QRG	84.2	2140	402	2542	5433
IQr	83.81	2130	412	2542	5433
YRr	83.59	2125	417	2542	5433
IRr	82.83	2106	436	2542	5433
Ybr	82.68	2102	440	2542	5433
Qbr	82.51	2097	445	2542	5433
YZr	82.49	2097	445	2542	5433
IZr	82.44	2096	446	2542	5433
YRG	82.43	2095	447	2542	5433
IQb	80.84	2055	487	2542	5433
YGb	80.58	2048	494	2542	5433
YQZ	80.51	2047	495	2542	5433
IRG	79.89	2031	511	2542	5433
RGr	79.21	2014	528	2542	5433
YZb	78.66	2000	542	2542	5433
YIb	78.39	1993	549	2542	5433
ZRb	78.16	1987	555	2542	5433
YRb	77.78	1977	565	2542	5433
Zbr	77.19	1962	580	2542	5433
QZb	76.74	1951	591	2542	5433
IQG	75.66	1923	619	2542	5433
ZGr	74.56	1895	647	2542	5433