

## LỜI CAM ĐOAN

Tôi xin cam đoan đề tài nghiên cứu trong luận văn này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Đồng thời, kết quả này chưa từng xuất hiện trong bất cứ một nghiên cứu nào. Các số liệu, kết quả nêu trong luận văn là trung thực nếu sai tôi hoàn chịu trách nhiệm.

**Tác giả luận văn**

## LỜI CẢM ƠN

Đề tài "**Nghiên cứu giải thuật phân cụm phổ cho tra cứu ảnh dựa trên nội dung**" là nội dung tôi chọn để nghiên cứu và làm luận văn tốt nghiệp cao học sau hai năm theo học ở Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Trong quá trình làm và hoàn thiện luận văn tốt nghiệp tôi đã nhận được rất nhiều sự giúp đỡ.

Đầu tiên tôi xin gửi lời cảm ơn chân thành tới thầy giáo hướng dẫn Ngô Quốc Tạo. Thầy là người đã tận tình giúp đỡ, dẫn dắt, truyền đạt kiến thức cũng như kinh nghiệm cho tôi, không chỉ trong quá trình thực hiện luận văn tốt nghiệp mà còn trong suốt thời gian học tập tại Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Tiếp đó tôi cũng xin gửi lời cảm ơn đến các quý thầy cô Viện Công nghệ thông tin, những người đã giảng dạy tôi trong suốt thời gian theo học tại Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Tôi cũng xin gửi lời cảm ơn đơn vị chuyên môn, ban Lãnh đạo, phòng Đào tạo, các phòng chức năng của Học viện Khoa học và Công nghệ để luận văn được hoàn thành.

Sau cùng tôi muốn gửi lời cảm ơn đến gia đình, bạn bè, các đàn anh đàn chị, các bạn học cùng khóa ITT20B01 đã luôn động viên, giúp đỡ tôi trong quá trình làm luận văn.

Trong quá trình làm luận văn, tôi đã cố gắng hết sức để hoàn thành thật tốt đề tài này trong khả năng của mình. Tuy nhiên do điều kiện thời gian cũng như kiến thức còn hạn chế, nên luận văn chắc chắn vẫn còn nhiều thiếu sót. Tôi thực sự mong nhận được sự góp ý của thầy cô và các bạn.

**Một lần nữa, xin chân thành cảm ơn!**

*Hà Nội, ngày tháng năm 2022*

***Học viên thực hiện***

## MỤC LỤC

LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN .....	ii
MỤC LỤC.....	iii
DANH MỤC CÁC BẢNG.....	v
DANH MỤC HÌNH VẼ.....	vi
MỞ ĐẦU.....	viii
CHƯƠNG 1. TỔNG QUAN VỀ CÁC PHƯƠNG PHÁP TRA CỨU ẢNH ...	1
1.1.  MỘT SỐ KHÁI NIỆM CƠ BẢN.....	1
1.1.1.  Xử lý ảnh.....	1
1.1.2.  Hệ thống xử lý ảnh.....	2
1.1.3.  Một số thuật ngữ cơ bản trong xử lý ảnh.....	2
1.2.  CÁC PHƯƠNG PHÁP TRA CỨU ẢNH.....	4
1.2.1.  Tra cứu ảnh dựa trên văn bản.....	5
1.2.2.  Tra cứu ảnh dựa trên nội dung.....	6
1.2.3.  Trích chọn đặc trưng trong tra cứu ảnh .....	7
1.2.4.  Phản hồi liên quan trong tra cứu ảnh .....	13
1.2.5.  Một số độ đo khoảng cách .....	17
1.3.  KIẾN TRÚC CỦA HỆ THỐNG TRA CỨU ẢNH .....	19
1.3.1.  Các vấn đề nghiên cứu được quan tâm .....	20
1.3.2.  Một số hệ thống tra cứu ảnh dựa trên nội dung.....	22
1.4.  LĨNH VỰC ỨNG DỤNG CỦA TRA CỨU ẢNH .....	27
CHƯƠNG 2. CÁC PHƯƠNG PHÁP TRA CỨU ẢNH DỰA TRÊN NỘI DUNG VÀ PHƯƠNG PHÁP PHẢN HỒI LIÊN QUAN .....	28
2.1.  GIỚI THIỆU .....	28
2.1.1.  Trích chọn đặc trưng sử dụng các phương pháp gói (Wrapper methods) .....	28
2.1.2.  Trích chọn đặc trưng sử dụng các phương pháp lọc (Filter methods) .....	31
2.1.3.  Trích chọn đặc trưng sử dụng các phương pháp nhúng (Embedded methods) .....	36
2.1.4.  So sánh phương pháp Filter, Wrapper và Embedded .....	37
2.2.  PHƯƠNG PHÁP PHẢN HỒI LIÊN QUAN.....	38

2.2.1. Giới thiệu về phản hồi liên quan .....	38
2.2.2. CBIR với phản hồi liên quan .....	40
2.3. TRA CỨU ẢNH DỰA TRÊN PHÂN CỤM ĐỒ THỊ .....	45
2.3.1. Phân cụm đồ thị và các thuật toán phân cụm phổ.....	45
2.3.2. Mã giả thuật toán Spectral Clustering.....	50
2.3.3. Các thuật toán phân cụm phổ.....	54
2.4. PHƯƠNG PHÁP TRA CỨU ẢNH SỬ DỤNG PHÂN CỤM PHỔ..	55
2.4.1. Phát biểu bài toán.....	55
2.4.2. Phân tích và xây dựng mô hình.....	58
2.5. THUẬT TOÁN PHÂN CỤM TRONG PHƯƠNG PHÁP TRA CỨU ẢNH .....	59
2.5.1. Ý tưởng.....	59
2.5.2. Thuật toán phân cụm tập ảnh phản hồi từ người dùng .....	60
2.5.3. Tìm ảnh đại diện cho cụm.....	66
2.5.4. Khoảng cách từ một ảnh đến truy vấn đa điểm .....	66
2.5.5. Thuật toán tra cứu ảnh sử dụng phân cụm phổ trong phản hồi liên quan .....	67
CHƯƠNG 3. CHƯƠNG TRÌNH MÔ PHỎNG KỸ THUẬT PHÂN CỤM PHỔ VÀ ỨNG DỤNG ĐỂ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG.....	69
3.1. CHƯƠNG TRÌNH MÔ PHỎNG KỸ THUẬT PHÂN CỤM PHỔ ...	69
3.1.1. Các bước mô phỏng phân cụm phổ.....	69
3.1.2. Công cụ và dữ liệu .....	69
3.1.3. Các bước phân cụm với bộ dữ liệu blods.....	70
3.1.4. Kết quả phân cụm trên các bộ dữ liệu moons, circle và aniso.....	75
3.2. CHƯƠNG TRÌNH TRA CỨU ẢNH DỰA TRÊN NỘI DUNG SỬ DỤNG KỸ THUẬT PHÂN CỤM PHỔ.....	77
3.2.1. Mô hình của chương trình tra cứu ảnh dựa trên nội dung sử dụng kỹ thuật phân cụm phổ .....	77
3.2.2. Môi trường và kỹ thuật.....	78
3.2.3. Bộ dữ liệu ảnh .....	79
3.2.4. Thực nghiệm và thảo luận.....	84
KẾT LUẬN .....	91
TÀI LIỆU THAM KHẢO.....	92

## DANH MỤC CÁC BẢNG

Bảng 2. 1. Bảng so sánh phương pháp Filter, Wrapper và Embedded .....	37
Bảng 3. 1. Thông tin về các bộ dữ liệu moons, circle và aniso .....	75
Bảng 3. 2. Chi tiết về các lớp và số lượng mẫu trong bộ dữ liệu.....	81
Bảng 3. 3. Danh sách các lớp trong bộ dữ liệu Oxford.....	82
Bảng 3. 4. Kết quả truy vấn với 10 người hợp ảnh đầu vào của bộ Corel.....	87
Bảng 3. 5. Kết quả truy vấn với 06 ảnh đầu vào của bộ Oxford.....	89

## DANH MỤC HÌNH VẼ

Hình 1.1. Quá trình xử lý ảnh.....	1
Hình 1.2. Các giai đoạn chính trong hệ thống xử lý ảnh .....	2
Hình 1.3. Quan hệ trong vùng lân cận giữa các điểm ảnh .....	4
Hình 1.4. Không gian màu RGB .....	9
Hình 1.5. Không gian màu HSV .....	11
Hình 1.6. Thủ tục phản hồi liên quan.....	14
Hình 1.7. Phản hồi liên quan tìm kiếm trên các ảnh, người dùng xem các kết quả truy vấn ban đầu của truy vấn bike .....	15
Hình 1.8. Phản hồi liên quan tìm kiếm trên các ảnh, người dùng xem tập kết quả được hiệu chỉnh .....	15
Hình 1.9. Kiến trúc của hệ thống tra cứu ảnh dựa trên nội dung.....	20
Hình 1.10. Hệ thống QBIC .....	22
Hình 1.11. Hệ thống VisualSeek.....	25
Hình 1.12. Hệ thống WebSeek.....	26
Hình 2. 1. Sơ đồ khối phương pháp Wrapper .....	30
Hình 2. 2. Sơ đồ khối phương pháp Filter.....	31
Hình 2. 3. Sơ đồ khối phương pháp Embedded .....	36
Hình 2. 4. Kiến trúc của hệ thống CBIR sử dụng RF .....	41
Hình 2. 5. Phân loại các thuật toán phản hồi liên quan.....	42
Hình 2. 6(a) Dịch chuyển điểm truy vấn (b) Hình dạng lỗi (đa điểm)(c) Hình dạng lõm (đa điểm) .....	58
Hình 2. 7. Cấu trúc của phương pháp SCRF.....	59
Hình 2. 8. Đồ thị biểu diễn cơ sở dữ liệu ảnh .....	63
Hình 3. 1. Minh họa phân bố các điểm trong bộ dữ liệu .....	70
Hình 3. 2. Biểu đồ Heat Map của ma trận Laplacian trên bộ dữ liệu blods .....	71
Hình 3. 3. Biểu đồ giá trị riêng của đồ thị Laplacian trên tập blods .....	72
Hình 3. 4. Biểu đồ 10 giá trị riêng đầu tiên đại diện của đồ thị Laplacian trên bộ dữ liệu blods .....	72
Hình 3. 5. Độ lớn các phần tử trong mỗi vector riêng của đồ thị Laplacian trên bộ dữ liệu blods .....	73
Hình 3. 6. Phương pháp elbow xác định số cụm trên tập blods.....	74
Hình 3. 7. Kết quả phân thành 03 cụm trên tập blods.....	74
Hình 3. 8. Kết quả phân cụm trên tập moons với số cụm là 02 .....	75
Hình 3. 9. Kết quả phân cụm trên tập circle với số cụm là 02.....	76
Hình 3. 10. Kết quả phân cụm trên tập aniso với số cụm là 03 .....	76
Hình 3. 11. Mô hình hệ thống tra cứu ảnh dựa trên nội dung.....	77
Hình 3. 12. Quy trình tra cứu ảnh dựa trên phân cụm phổ .....	78
Hình 3. 13. Kiến trúc mạng VGG16 .....	79
Hình 3. 14. Các thư mục chứa dữ liệu từng lớp ảnh .....	80

Hình 3. 15. Các lớp tương ứng với mỗi giống loài trong bộ dữ liệu Oxford.....	83
Hình 3. 16. Trang chính hệ thống CBIR dựa trên kỹ thuật phân cụm phổ .....	84
Hình 3. 17. Thời gian cho trích xuất thuộc tính và phân cụm .....	85
Hình 3. 18. Thử nghiệm với ảnh truy vấn là động vật .....	86
Hình 3. 19. Thử nghiệm với ảnh truy vấn là lâu đài .....	87
Hình 3. 20. Thử nghiệm với ảnh truy vấn là bông hoa .....	87
Hình 3. 21. Truy vấn ảnh nhóm Chihuahua .....	88
Hình 3. 22. Truy vấn ảnh trong lớp american_pit_bull_terrier.....	89

## MỞ ĐẦU

### 1. Lý do chọn đề tài

Ngành khoa học máy tính ngày càng phát triển trong thời đại công nghệ 4.0. Chúng ta có thể nhận ra sự phát triển nhanh chóng của các lĩnh vực trong khoa học máy tính, từ trí tuệ nhân tạo, chế tạo robot, dữ liệu lớn, kỹ thuật đồ họa, công nghệ in ấn 3D, vân vân và tất nhiên trong các phân ngành đó, không thể nào thiếu được phân ngành xử lý ảnh.

Sự phổ biến lan rộng và nhanh chóng của Internet cũng như các thiết bị điện tử công nghệ cao đã khiến cho các hoạt động trao đổi, lưu trữ và tìm kiếm thông tin trở thành một hoạt động thường nhật không thể thiếu được trong đời sống hiện đại. Hình ảnh có thể được xem là một dạng thông tin thường gặp và thường được sử dụng, có tác động rất mạnh tới thị giác của con người. Song song với sự phát triển của khoa học công nghệ, thì nhu cầu sử dụng ảnh số của con người ngày càng tăng cao và chắc chắn sẽ không dừng lại. Chúng ta có thể nhận thấy sự hiện diện của ảnh số ở khắp mọi nơi, từ trong cuộc sống bình thường, ví dụ những bức ảnh được đăng tải trên các trang mạng, cho tới những hình ảnh mà mọi người chia sẻ với nhau qua mạng xã hội, hay là trong các lĩnh vực chuyên môn sâu, ví dụ như việc nhận dạng khuôn mặt, chữ viết, dấu vân tay trong công việc quản lý cơ sở dữ liệu dân cư hoặc an ninh điều tra. Ngay cả những ngành khoa học khác như viễn thám, vũ trụ, y sinh, thiết kế đồ họa, vân vân, cũng rất cần đến những thông tin ảnh số.

Ngành khoa học xử lý ảnh ra đời đã đáp ứng được nhu cầu sử dụng ảnh số của thế giới hiện đại. Xử lý ảnh bao gồm rất nhiều phân ngành nhỏ, như là nâng cao chất lượng hình ảnh, phân tích hình ảnh, phục hồi hình ảnh, nén và định dạng ảnh, đoán nhận ảnh và tất nhiên không thể không nhắc tới cả tra cứu ảnh.

Tra cứu ảnh hay tìm kiếm hình ảnh, là một chuyên ngành phổ biến và có tính ứng dụng cao của ngành khoa học xử lý ảnh. Chúng ta có thể đặt ra rất nhiều câu hỏi, ví dụ như với một lượng thông tin khổng lồ trên các nền tảng lưu trữ ví dụ như là không gian internet và trong đó có không ít là các dữ liệu ảnh số, làm thế nào để ta có thể tìm được đúng những hình ảnh theo như mong muốn của mình ? Và nếu như trong trường hợp chúng ta đã tìm thấy ảnh, thì làm cách nào có thể biết được liệu những dữ liệu ảnh đó có đúng với yêu cầu mà mình



mong muốn hay không ? Rồi giả sử như đã tìm được hình ảnh đúng như ý muốn, thì liệu có thể có một phương án nào đó giúp cho việc tra cứu ở những lần sau ngày càng nhanh chóng, hiệu quả hơn và ít tốn kém hơn lần trước không ?

Nhận ra đây là một vấn đề thiết thực và gần gũi với đời sống hằng ngày, cho nên tôi đã chọn làm về lĩnh vực tra cứu ảnh cho luận văn cao học của mình, với tên đề tài là "Nghiên cứu giải thuật phân cụm phổ cho tra cứu ảnh dựa trên nội dung".

## **2. Mục đích nghiên cứu**

Giống như những gì mà tên đề tài đã nêu ra, mục đích nghiên cứu của tôi chính là tìm ra được một phương án tra cứu ảnh hiệu quả, tối ưu và có thể ứng dụng được vào các lĩnh vực trong thực tế.

## **3. Nội dung nghiên cứu**

Nội dung đề tài nghiên cứu là tìm hiểu, giới thiệu tổng quát về các phương pháp tra cứu ảnh. Trong đó, trọng tâm chính là phương pháp tra cứu ảnh trong đó có áp dụng giải thuật phân cụm phổ (spectral clustering). Đồng thời dựa trên cơ sở các phương pháp tra cứu ảnh trên, tiến hành xây dựng chương trình thử nghiệm cho phép đọc vào một tập các mô tả hoặc các thẻ được dán nhãn theo yêu cầu của người dùng, rồi tìm kiếm những hình ảnh theo đúng chuẩn các mô tả đó trên bộ dữ liệu sẵn có.

## **4. Cơ sở khoa học và tính thực tiễn của đề tài**

Hoạt động tra cứu ảnh không chỉ đơn giản là một nhu cầu thiết yếu của con người mà còn nắm vai trò quan trọng trong rất nhiều các ngành khoa học khác. Ta có thể kể tới một vài ví dụ như sau. Trong ngành y tế, thông qua việc sử dụng và khám phá bộ dữ liệu hình ảnh 3D của cơ thể con người, thường được chụp qua các máy chụp cắt lớp hay máy chụp cộng hưởng, mà bác sĩ có thể tiến hành công việc chẩn đoán bệnh hay tìm ra các phương thức điều trị bệnh phù hợp với tình trạng của bệnh nhân. Hay như trong lĩnh vực quân sự, ứng dụng tra cứu ảnh có thể hỗ trợ việc nhận dạng máy bay hay các thiết bị quân sự khác từ màn hình radar, xác định mục tiêu từ ảnh vệ tinh hay cung cấp bản đồ dẫn đường cho tên lửa hành trình.

Có rất nhiều phương pháp tra cứu ảnh đã được đề ra, nghiên cứu và phát triển. Trong đó, phương pháp tra cứu ảnh dựa trên nội dung có thể xem là

phương pháp tra cứu ảnh phổ biến nhất. Thông qua luận văn của mình, tôi cũng muốn tìm hiểu và nghiên cứu phương pháp này, đồng thời tự phát triển một hệ thống tra cứu ảnh dựa trên những gì đã khám phá được.

### **5. Những đóng góp của luận văn**

Trong luận văn của mình, với chủ đề nghiên cứu là "Áp dụng giải thuật phổ cho tra cứu ảnh dựa trên nội dung", tôi đã trình bày những gì mình đã tìm hiểu và khám phá được. Bắt đầu từ những khái niệm và thuật ngữ cơ bản nhất của xử lý ảnh cho đến tra cứu ảnh, sau đây là những phương pháp tra cứu ảnh từ đơn giản cho đến phức tạp, rồi đến những kỹ thuật, những thuật toán được sử dụng để nâng cao độ chính xác của kết quả tra cứu.

Ngoài ra trong luận văn tôi còn đề cập đến phương pháp phản hồi liên quan. Đây không chỉ là một phương pháp cải tiến hiệu suất tra cứu dựa trên những phản hồi từ chính người dùng, mà còn là một quá trình tương tác, xây dựng câu hỏi để kết nối người dùng với công cụ tìm kiếm nhằm tạo ra kết quả có ý nghĩa hơn. Phản hồi liên quan không chỉ là một phương pháp được ứng dụng trong công việc xử lý và tra cứu ảnh, mà nó còn có thể được áp dụng cho các bộ môn nghiên cứu khác, ví dụ như trong lĩnh vực truy hồi thông tin hay khai phá dữ liệu. Phương pháp này giúp mô hình hóa nhận thức của con người theo cách tốt hơn.

Trong chương cuối của luận văn, tôi đã xây dựng được một chương trình mô phỏng, sử dụng kỹ thuật phân cụm phổ để ứng dụng tra cứu hình ảnh theo nội dung. Thông qua kết quả thực nghiệm, tôi đã đưa ra được kết luận rằng thuật toán phân cụm phổ mà mình sử dụng đã đem lại kết quả có độ chính xác cao, qua đó có thể ứng dụng vào các hệ thống tra cứu ảnh dựa trên nội dung trong tương lai.

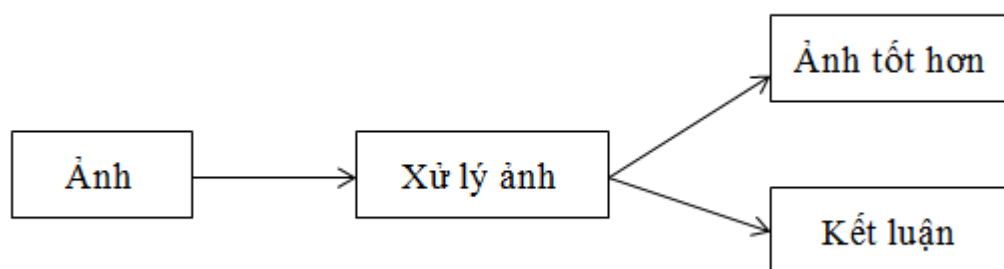
# CHƯƠNG 1. TỔNG QUAN VỀ CÁC PHƯƠNG PHÁP TRA CỨU ẢNH

## 1.1. MỘT SỐ KHÁI NIỆM CƠ BẢN

### 1.1.1. Xử lý ảnh

Con người cảm nhận thế giới xung quanh bởi năm giác quan. Trong đó, thị giác có thể xem là giác quan đem lại những cảm nhận "trực quan" nhất, bởi đây là giác quan phát hiện và diễn dịch thông tin ánh sáng được nhìn thấy bằng mắt. Trong thời đại công nghệ số, với sự phát triển của các ngành khoa học và đặc biệt là ngành công nghệ thông tin, hình ảnh đóng một vai trò rất quan trọng trong quá trình tương tác cũng như trao đổi thông tin của con người, bởi đây là những thông tin mà chúng ta có thể nhận được một cách trực quan thông qua thị giác của mình. Ngành khoa học xử lý ảnh vì vậy đã ra đời và ngày càng phát triển mạnh mẽ, song song với đó là nhu cầu sử dụng ảnh số ngày càng lớn của con người.

Xử lý ảnh là giải pháp được thực hiện trên các ảnh đầu vào, thông qua đó tạo ra các ảnh mới với một hay nhiều mục tiêu nào đó. Ảnh trả về có thể là ảnh mang chất lượng được nâng cao hoặc là được thay đổi kết cấu định dạng, cũng có thể là được trích rút một số thông tin hữu ích hay là theo nhiều kiểu khác.



Hình 1.1. Quá trình xử lý ảnh

Ảnh trong xử lý ảnh là ảnh kỹ thuật số (digital image). Ảnh số được tạo thành từ các điểm ảnh (pixel) và đây cũng có thể xem là một ảnh n chiều [1]. Mỗi pixel có một màu riêng biệt và các màu sẽ được xác định bởi các không gian màu.

### 1.1.2. Hệ thống xử lý ảnh

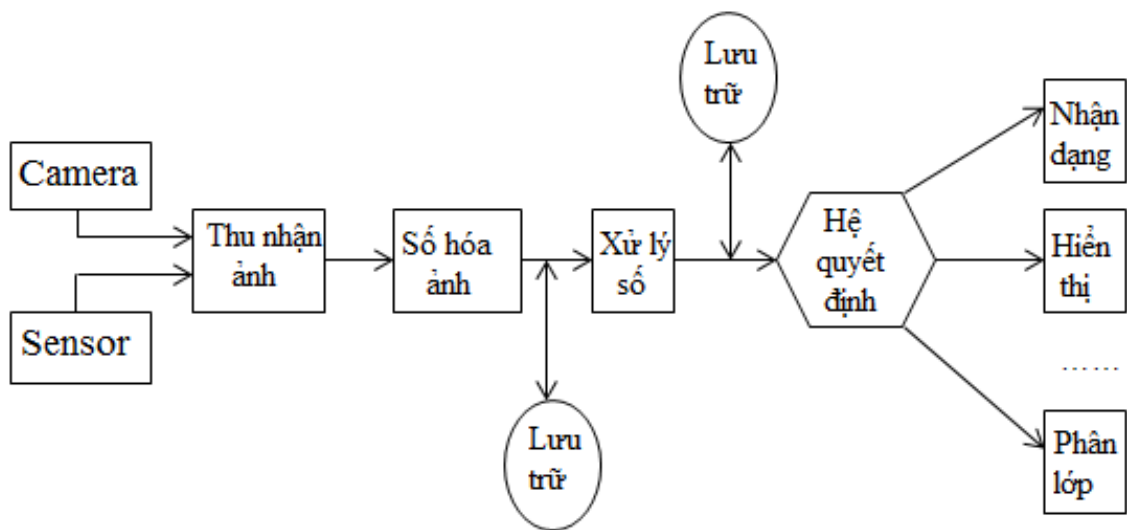
Một hệ thống xử lý ảnh thường xử lý qua các giai đoạn chính như sau [1]:

- Pha thu nhận ảnh: Thu nhận ảnh qua các thiết bị máy ảnh số, camera, máy quét ảnh hoặc các bộ cảm ứng.

- Pha số hóa ảnh: Thực hiện biến đổi hình ảnh tương tự trở thành hình ảnh rời rạc (rời rạc hóa) để có thể xử lý trên hệ thống máy vi tính thông qua quá trình lấy mẫu (Sampling) và quá trình lượng tử hóa (Quantization).

- Pha xử lý số: Là một quá trình bao gồm nhiều pha nhỏ hơn: tăng cường chất lượng ảnh (Enhancement), khôi phục hình ảnh (Restoration), dò biên (Edge Detection), phân đoạn ảnh (Segmentation), trích chọn các đặc trưng (Feature Extraction)...

- Hệ hỗ trợ ra quyết định tùy thuộc vào mục đích và ứng dụng mà quyết định chuyển tiếp sang các pha khác như là pha hiển thị, pha phân lớp, pha phân đoạn, pha nhận dạng, pha truyền thông...



Hình 2.2. Các giai đoạn chính trong hệ thống xử lý ảnh

### 1.1.3. Một số thuật ngữ cơ bản trong xử lý ảnh

- Điểm ảnh (Pixel): Phần tử điểm ảnh gọi tắt là phần tử ảnh được định nghĩa là tín hiệu (cường độ mức sáng) tại một vị trí tọa độ xác định trong không gian của đối tượng.

- Ảnh: Một ảnh được định nghĩa là một tập hợp của các điểm ảnh.

- Mức xám: Mức xám được định nghĩa là các giá trị (Value) có thể có của các điểm ảnh trong một ảnh.

- Lược đồ mức xám (Histogram): Lược đồ mức xám có thể gọi tắt là lược đồ xám được định nghĩa như là một lược đồ tần suất thống kê số lần xuất hiện các mức xám trong ảnh, hay nói cách khác một hàm trả về tần suất xuất hiện của mỗi mức xám trong ảnh chính là lược đồ xám của ảnh.

### **1.1.3.1. Thu nhận ảnh**

Gốc của ảnh (ảnh tự nhiên) là ảnh liên tục về không gian và độ sáng [2]. Để xử lý bằng máy tính (số), ảnh cần phải được số hoá, tức ta phải đưa được ảnh vào bên trong thiết bị thu nhận ảnh để tiến hành xử lý. Thu nhận ảnh là bước đầu tiên trong hệ thống xử lý ảnh.

Thiết bị thu nhận ảnh, là một thiết bị quang học có thể chuyển đổi hình ảnh đã thu nhận thành hình ảnh kỹ thuật số. Theo như sơ đồ ở Hình 1.2, chúng ta có thể thu thập hình ảnh bằng các thiết bị thu nhận như máy ảnh (camera), máy quét (scanner), máy fax... vân vân và thậm chí là bằng các cảm biến (sensor). Cảm biến thu nhận hình ảnh có ba dạng: Cảm biến đơn (Single Sensor), Cảm biến dòng (Line Sensor), Cảm biến mảng (Array Sensor).

### **1.1.3.2. Biểu diễn ảnh**

Có hai cách biểu diễn ảnh thường gặp nhất hiện nay, một là biểu diễn ảnh theo mô hình Vector và hai là biểu diễn ảnh theo mô hình Raster.

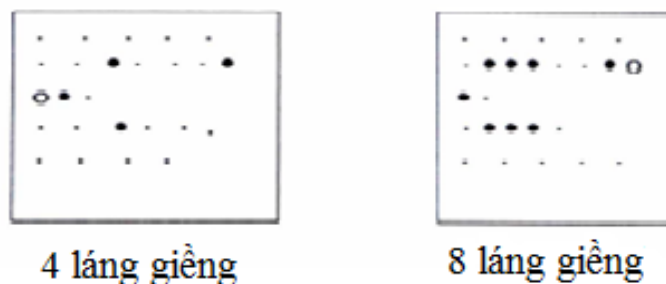
Ảnh Vector là ảnh được tạo ra từ các khái niệm hình học cơ bản như đường thẳng, đường cong,... Nó được lưu trữ dưới dạng chỉ dẫn thay vì các điểm ảnh. Ví dụ như khi ta vẽ một đường thẳng từ A đến B với độ dày T và màu C, hoặc là vẽ một đường cong tam X, bán kính R, độ dày T, màu C, vùng bên trong tô màu Z. Ảnh Vector phù hợp với ảnh vẽ, ảnh thiết kế đồ họa, ảnh hoạt hình hoặc ảnh hoạt hình 3D, ví dụ như khi sử dụng phần mềm Adobe Illustrator [3].

Ảnh Vector có kích thước file khá nhỏ, lưu thông tin thay vì điểm ảnh, không phụ thuộc vào độ phân giải, có dẫn tùy ý mà không mất thông tin. Các định dạng của ảnh Vector là .AI, .SVG, .WMF, .DWG, .EPS, .PDF [3].

Ảnh Raster, hay còn được gọi là ảnh Bitmap, là một định dạng ảnh kỹ thuật số sử dụng các pixel hình chữ nhật nhỏ, hoặc các phần tử hình ảnh, được sắp xếp theo hình lưới (grid) để hiển thị hình ảnh. Nói cách khác, ảnh được lưu trữ như một mảng 2 chiều các điểm ảnh, mỗi điểm ảnh có thông số về màu sắc. Ngày nay, hầu hết những hình ảnh chúng ta thấy trên Internet và các bức ảnh được chụp bởi máy ảnh kỹ thuật số đều là ảnh Raster. Bởi vì định dạng Raster có thể hỗ trợ nhiều màu sắc khác nhau và mô tả được tinh tế các tông màu, cho nên nó rất thích hợp cho ảnh chụp tự nhiên, ảnh chụp phong cách, ảnh chụp bóng mờ... [3]

Ảnh Raster có kích thước lớn hơn ảnh Vector, nó mất thông tin khi co giãn. Đối với ảnh chụp Raster, cần tối thiểu 8 bit cho mỗi thành phần màu cơ bản (R, G, B). Các máy ảnh chỉ có thể tin tại độ phân giải từ 300 đến 600 dpi. Các định dạng của ảnh Raster là .JPG, .BMP, .TIFF, .PNG, .PSD, EPS, .PDF [3].

Khi xử lý các ảnh Raster, chúng ta có thể sử dụng phương pháp "Toán tử láng giềng" hay còn gọi là "Toán tử hàng xóm". Trong Hình 1.3, cho một kích thước hình vuông (có thể là hình chữ nhật tùy theo bài toán), ta thấy toán tử có liên quan đến các điểm láng giềng, láng giềng nằm trong hình chữ nhật xung quanh điểm ảnh hiện tại [3].



Hình 3.3. Quan hệ trong vùng lân cận giữa các điểm ảnh

## 1.2. CÁC PHƯƠNG PHÁP TRA CỨU ẢNH

Sự phát triển của công nghệ thông tin và sự mở rộng của mạng lưới Internet đã đồng thời thúc đẩy sự phát triển của không ít các ngành nghề khác. Trong lĩnh vực xử lý ảnh, có rất nhiều phương pháp tra cứu ảnh đã được nghiên cứu, cải tiến và phát triển, phục vụ cho nhu cầu sử dụng ảnh số ngày càng lớn của con người.

### 1.2.1. Tra cứu ảnh dựa trên văn bản

Ảnh số (digital Image) hay còn gọi là ảnh kỹ thuật số, là một đơn vị biểu diễn hình ảnh, được ghi nhận bởi các thiết bị điện tử, đồng thời được lưu trữ trong bộ nhớ của các thiết bị này dưới dạng dữ liệu ảnh. Trong đời sống hiện đại, với sự phát triển nhanh chóng của Internet, nhu cầu sử dụng ảnh số ngày càng tăng cao. Chúng ta có thể dễ dàng nhận thấy sự hiện diện của ảnh số trên hầu hết các lĩnh vực của đời sống như y tế, giải trí, an ninh, thương mại, kiến trúc, xây dựng, giáo dục... vân vân. Nhờ sự tiện lợi của Internet, ngày càng có nhiều dữ liệu ảnh số được đưa lên không gian mạng để sử dụng, chia sẻ và lưu trữ cho các mục đích khác nhau. Không thể phủ nhận những bộ sưu tập dữ liệu ảnh số với số lượng lớn đã cung cấp cho chúng ta những nguồn thông tin, tài liệu phong phú và đa dạng. Tuy nhiên, điều này cũng đã đặt ra rất nhiều vấn đề khó khăn và thách thức. Ví dụ như trong một cơ sở dữ liệu ảnh rất lớn, làm thế nào để chúng ta có thể tìm được nhanh nhất và đúng nhất những dữ liệu ảnh phù hợp với mục đích sử dụng của mình? Hay là làm sao chúng ta có thể xác định được liệu những dữ liệu ảnh được tìm thấy đã đạt được chất lượng tốt như mong muốn hay chưa? Hoặc là đối với những dữ liệu ảnh được lưu trữ hoặc được tìm thấy nhưng chất lượng ảnh lại ở mức chưa tốt, làm thế nào để chúng ta có thể nâng cao chất lượng của những dữ liệu đó lên mức đạt yêu cầu?

Để giải quyết các vấn đề trên mà thường gặp nhất chính là việc làm sao có thể tra cứu được ảnh một cách chính xác, có rất nhiều phương pháp tra cứu ảnh đã được nghiên cứu và ứng dụng vào thực tế. Từ những phương pháp tra cứu cơ bản và là nền tảng cho rất nhiều phương pháp tra cứu ảnh chuyên sâu và mở rộng hơn sau này. Tiêu biểu là phương pháp *tra cứu ảnh dựa trên văn bản*.

Tra cứu ảnh dựa trên văn bản hay còn gọi là tra cứu ảnh dựa trên từ khóa - TBIR (Text-based image retrieval) được bắt đầu từ những năm 1960 và được nghiên cứu tích cực vào những năm 1970 [4]. Trong phương pháp này, hình ảnh được chú thích bằng các mô tả văn bản và việc tra cứu chúng sẽ dựa trên việc so khớp (match) các yêu cầu truy vấn nguyên bản/nguyên văn của người dùng với các chú thích của văn bản đó. Tuy nhiên, phương pháp này vẫn có những nhược

điểm rất lớn. Thứ nhất là trong một cơ sở dữ liệu lớn với những hình ảnh có quá nhiều chi tiết, thì việc đánh chú thích cho mỗi hình ảnh một là rất khó khăn. Thứ hai là việc đánh chú thích vẫn thường được thực hiện bằng cách thủ công cho nên sẽ tốn rất nhiều thời gian và hiệu suất truy vấn. Thứ ba là việc đánh chú thích có thể mang tính chủ quan do hoàn cảnh tác động, không nhất quán, không đầy đủ, dẫn đến kết quả trả về có thể không được đúng như yêu cầu.

Để khắc phục các nhược điểm của TBIR, thì CBIR hay còn gọi là phương pháp tra cứu ảnh dựa trên nội dung đã ra đời và sẽ được đề cập đến ở phần 1.2.2 tiếp sau.

### **1.2.2. Tra cứu ảnh dựa trên nội dung**

Tra cứu ảnh dựa trên nội dung CBIR (Content-based image retrieval) được giới thiệu vào năm 1990 [5]. Đây là phương pháp tự động truy vấn hình ảnh kỹ thuật số từ cơ sở dữ liệu lớn. CBIR sử dụng nội dung trực quan vốn có của hình ảnh để thực hiện tìm kiếm, tra cứu và truy vấn. Trái ngược với TBIR chủ yếu sử dụng các chú thích hình ảnh bằng phương pháp thủ công đã được đề cập phần 1.2.1 phía trên, CBIR xác định hình ảnh bằng các đặc điểm cú pháp được tự động rút gọn. Với sự phát triển vượt bậc của công nghệ, bao gồm sự mở rộng của Internet và các tiện ích của nó, sự phổ biến của các thiết bị thu nhận ảnh kỹ thuật số (ví dụ như máy ảnh kỹ thuật số, điện thoại thông minh máy tính), CBIR đã nhanh chóng được phát triển đồng thời được ứng dụng trong nhiều lĩnh vực khác nhau liên quan đến thị giác máy tính (ví dụ như thiết kế đồ họa) hay trí tuệ nhân tạo. Phương pháp này ở thời điểm hiện tại có thể hỗ trợ người dùng truy vấn hình ảnh hiệu quả mà dùng không cần phải mô tả hình ảnh một cách chủ quan, công kênh, giảm tỷ suất gặp lỗi khi tìm kiếm ảnh. Đặc biệt là có thể tra cứu ảnh dễ dàng hơn từ cơ sở dữ liệu quy mô lớn chỉ với một hình ảnh đầu vào. Qua đó đã cải thiện đáng kể khả năng sử dụng của hệ thống tra cứu ảnh.

CBIR sử dụng các đặc điểm trực quan (hay còn gọi là đặc trưng) của hình ảnh như màu sắc (color), hình dạng (shape), kết cấu (texture) hay các thông tin không gian (spatial information) để tìm kiếm hình ảnh truy vấn từ cơ sở dữ liệu lớn. Tuy nhiên, rất hiếm đặc trưng riêng lẻ có thể cho ra các kết quả chính xác



với mọi nhu cầu. Do vậy ta có thể kết hợp các đặc trưng lại để tìm được các kết quả tra cứu thích hợp cho ứng dụng tra cứu ảnh dựa trên nội dung.

### **1.2.3. Trích chọn đặc trưng trong tra cứu ảnh**

Trích chọn đặc trưng bao gồm trích chọn thông tin có nghĩa của ảnh. Điều này giúp làm giảm dung lượng lưu trữ, vậy nên hệ thống hoạt động sẽ nhanh và hiệu quả hơn, tiết kiệm thời gian truy vấn.

#### **1.2.3.1. Đặc trưng màu sắc**

Hệ thống CBIR được sử dụng để tự động lập chỉ số (index), tìm kiếm (search), tra cứu (retrieval), truy vấn (query) và duyệt (browse) cơ sở dữ liệu hình ảnh. Màu sắc là một trong những đặc trưng quan trọng nhất được sử dụng trong hệ thống CBIR.

Việc sử dụng màu sắc trong xử lý hình ảnh rất cần thiết, bởi vì màu sắc là một bộ miêu tả mạnh mẽ giúp đơn giản hóa việc nhận dạng và trích xuất đối tượng. Cảm nhận màu sắc phụ thuộc vào phổ của ánh sáng, trong đó, ánh sáng nhìn thấy có dải phổ rất hẹp, còn ánh sáng với tất cả các thành phần phổ nhìn thấy có năng lượng bằng nhau sẽ cho ra ánh sáng trắng [3]. Bộ não của con người có khả năng xử lý hình ảnh màu rất lớn, giúp cho chúng ta có thể phân biệt hàng ngàn sắc thái và độ đậm nhạt của màu sắc.

Xử lý ảnh màu được chia thành hai lĩnh vực chính [6]:

Thứ nhất là xử lý toàn màu (Full color): hình ảnh thu được bằng cảm biến đủ màu, chẳng hạn như máy ảnh TV màu hoặc máy quét màu. Việc này đã trở nên phổ biến trong khoảng một thập kỷ trở lại đây, do sự sẵn có, dễ tìm và giá thành rẻ của các cảm biến màu (color sensors) và phần cứng.

Thứ hai là xử lý màu giả (Pseudo color): Gán màu cho một loạt các cường độ đơn sắc cụ thể hoặc một phạm vi cường độ. Ví dụ, các đối tượng có các giá trị cường độ khác nhau có thể được gán các màu khác nhau, điều này sẽ cho phép con người dễ dàng xác định hay nhận dạng hình ảnh.

Trong tra cứu ảnh theo nội dung, có rất nhiều cách mô tả thông tin/dữ liệu màu khác nhau, bao gồm: màu chủ đạo (dominant colors), ký hiệu màu (color

signatures), khoảnh khắc màu (color moments), không gian màu (color spaces/color models) [7].

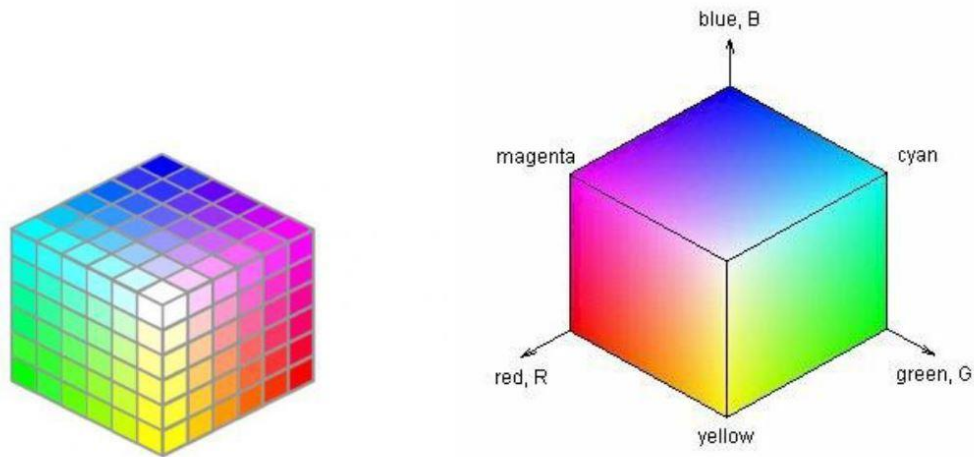
Hầu hết các không gian màu (color spaces/color models) có thể được ánh xạ thành hệ tọa độ 2-D, 3-D hoặc 4-D tương tự như hệ tọa độ Descartes. Do vậy, màu sắc có thể được xem là các tọa độ trong không gian 2-D, 3-D hoặc 4-D. Các thành phần màu trong không gian màu còn được gọi là kênh màu (color channels). Phần sau của luận văn sẽ đề cập đến một số không gian màu phổ biến.

➤ *a. Không gian màu*

- Không gian màu RGB (Red – Green – Blue)

Không gian màu RGB được sử dụng phổ biến trong đồ họa máy tính và các thiết bị kỹ thuật số (ví dụ camera màu, màn hình máy tính màu). Không gian dựa trên hệ tọa độ Descartes, là sự kết hợp của ba màu sắc cơ bản là màu đỏ (Red - R), xanh lục (Green - G) và xanh lam (Blue - B). Nó còn được xem như một khối lập phương ba chiều (như Hình 1.4), với ba màu tương ứng với ba góc của khối lập phương. Ngoài ra cả ba giá trị màu R, G và B đã được chuẩn hóa trong phạm vi [0, 1].

Số bit được sử dụng để biểu diễn mỗi pixel trong không gian màu RGB là được gọi là độ sâu pixel (pixel depth). Ví dụ, nếu mỗi hình ảnh màu đỏ, xanh lục và xanh lam là hình ảnh 8 bit, thì độ sâu pixel của hình ảnh RGB là 24 bit. Đây thường được gọi là hình ảnh đầy đủ màu sắc. Có thể có  $2^{24} = 16,777,216$  màu sắc độc đáo trong hệ thống RGB. [9]



Hình 4.4. Không gian màu RGB

- Không gian màu CIE:

Không gian màu CIE là một mô hình không gian màu được tạo ra bởi Ủy ban Quốc tế về Chiếu sáng có tên đầy đủ Commission Internationale de l'Éclairage. Ngoài ra nó còn được gọi là không gian màu CIE XYZ hoặc không gian màu CIE 1931 XYZ [10].

Không gian màu CIE là một hệ thống ánh xạ sử dụng tristimulus (sự kết hợp của 3 giá trị màu gần với giá trị đỏ (red)/ lục (green)/ lam (blue)), được vẽ trên không gian 3D. Khi các giá trị này được kết hợp với nhau, chúng có thể tái tạo bất kỳ màu nào mà mắt người có thể cảm nhận được. Đặc điểm kỹ thuật của CIE được cho là có thể thể hiện chính xác mọi màu sắc mà mắt người có thể cảm nhận được [10].

Trong không gian màu  $L^*a^*b^*$ , màu sắc được biểu thị bằng 3 giá trị số.  $L^*$  cho độ nhạt (L-lightness) và  $a^*$  cùng  $b^*$  cho các thành phần màu xanh lục-đỏ (green-red) và xanh lam-vàng (blue-yellow) tương ứng. Hệ thống được thiết kế để đồng nhất khả năng cảm nhận đối với thị giác màu của con người. Bản thân không gian màu CIE cũng là một không gian số thực 3 chiều, do đó bất kỳ sự biến đổi màu nào cũng có thể được biểu thị bằng tọa độ  $L^*a^*b^*$  [11].

- Không gian màu HSV:

Không gian màu HSV là một mô hình màu hình trụ, biến đổi các màu cơ bản RGB thành các chiều dễ hiểu hơn đối với con người. Ta có thể chuyển đổi

không gian RGB sang HSV bằng cách lấy các giá trị R, G, B chia cho 255 để thay đổi dải từ 0...255 thành 0...1 [3]. Không gian màu HSV xác định màu sắc theo ba thông số màu sắc. Trong đó:

+ H (Hue – Sắc độ): Đặc trưng cho màu sắc chủ đạo được người quan sát cảm nhận, là thuộc tính liên quan đến bước sóng chủ yếu trong hỗn hợp các bước sóng ánh sáng [3]. Sắc độ được đo bằng độ từ 0 đến 360. Ví dụ: Màu đỏ rơi vào khoảng từ 0 đến 60 độ. Màu vàng rơi vào khoảng từ 61 đến 120 độ. Màu xanh lá cây rơi vào khoảng từ 121 đến 180 độ. Màu lục lam (cyan) rơi vào khoảng từ 181 đến 240 độ. Màu xanh lam rơi vào khoảng từ 241 đến 300 độ. Màu đỏ tươi (magenta) rơi vào khoảng từ 301 đến 360 độ [12].

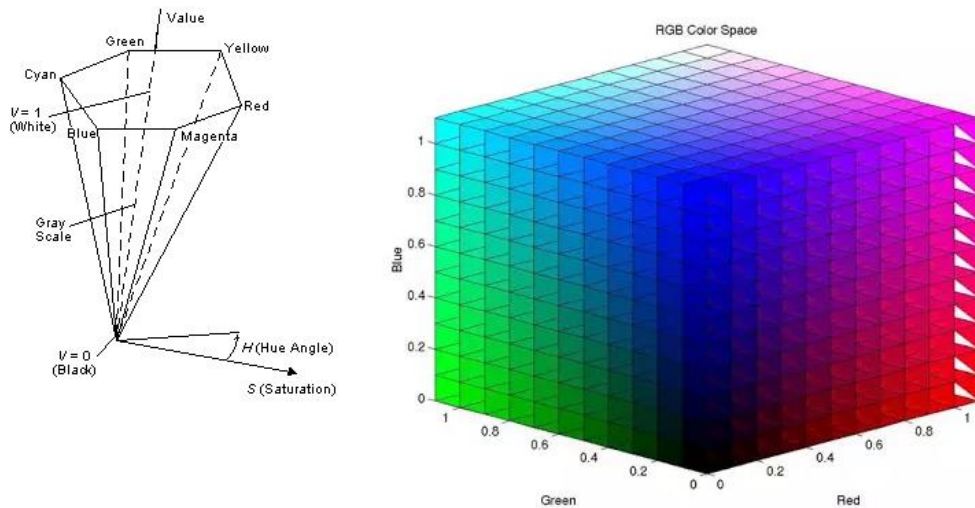
+ S (Saturation – Độ bão hòa): Thể hiện lượng màu trắng được hòa với sắc độ [3]. Độ bão hòa của một màu được phân tích trên thang điểm từ 0 đến 100 phần trăm [12].

+ V (Value – Độ sáng): Đặc trưng cho độ rơi cảm nhận [3]. Tất cả các giá trị trong khoảng từ 0 đến 1.

Hình 1.5 mô tả mô hình của không gian màu HSV dạng một hình chóp lục giác ngược. Mỗi mặt cắt ngang của nó là một hình lục giác. Mặt trên là một hình lục giác đều, thể hiện sự thay đổi sắc độ theo hướng H, từ 0 độ đến 360 độ là toàn bộ quang phổ của ánh sáng nhìn thấy. Sáu góc của hình lục giác đại diện cho vị trí của sáu màu đỏ (red), vàng (yellow), lục (green), lục lam (cyan), lam (blue) và đỏ tươi (magenta), mỗi màu cách nhau 60 độ [13].

Độ bão hòa S được biểu thị bằng hướng S từ tâm đến ranh giới lục giác, và giá trị thay đổi từ 0 đến 1. Càng gần ranh giới lục giác, độ bão hòa màu càng cao. Màu của ranh giới lục giác là bão hòa nhất, tức là  $S = 1$ . Độ bão hòa màu ở tâm của hình lục giác là 0, tức là  $S = 0$  [13].

Chiều cao của hình chóp lục giác (hay còn gọi là trục trung tâm) được ký hiệu là V, thể hiện sự chuyển màu đen sang trắng từ dưới lên trên. Đáy của V màu đen,  $V = 0$ ; đỉnh của V màu trắng,  $V = 1$  [13].



Hình 5.5. Không gian màu HSV

➤ *b. Biểu đồ màu*

Biểu đồ màu (color histogram) là một thực thể 3D mà trong đó mỗi pixel của hình ảnh (thay vì từng mẫu) được xác định bằng một tập các bin. Không gian màu (color space) được chia thành các khối thể tích bin, mỗi khối đại diện cho một dải màu [14].

Biểu đồ màu có cấu trúc linh hoạt. Nó có thể được xây dựng từ các hình ảnh trong các không gian màu khác nhau ví dụ RGB hay rg chromaticity (đây là không gian màu hai chiều được chuẩn hóa của RGB) hay bất kỳ không gian màu nào khác thuộc bất kỳ chiều nào. Biểu đồ của một hình ảnh được tạo ra đầu tiên bằng cách tùy biến các màu trong hình ảnh vào một số bin và đếm số pixel hình ảnh trong mỗi bin. Ví dụ: sắc độ đỏ (red) - xanh lam (blue). Trong lĩnh vực xử lý ảnh và nhiếp ảnh, biểu đồ màu là sự thể hiện sự phân bố của các màu trong một hình ảnh. Đối với hình ảnh kỹ thuật số, một biểu đồ màu thể hiện số lượng pixel có màu trong mỗi danh sách dải màu cố định, trải dài không gian màu của hình ảnh, tập hợp tất cả các màu có thể có [14].

Biểu đồ màu có thể được xây dựng cho bất kỳ loại không gian màu nào, mặc dù thuật ngữ này thường được sử dụng cho không gian ba chiều như RGB hoặc HSV. Đối với hình ảnh đơn sắc (monochromatic), thuật ngữ biểu đồ cường độ (intensity histogram) có thể được sử dụng thay thế. Đối với hình ảnh đa phổ

(multi-spectral images), trong đó mỗi pixel được biểu thị bằng một số phép đo tùy ý (ví dụ: trừ ba phép đo trong RGB), biểu đồ màu là N chiều, với N là số phép đo được thực hiện. Mỗi phép đo có dải bước sóng riêng của phổ ánh sáng, một số có thể nằm ngoài dải quang phổ nhìn thấy được [14].

### **1.2.3.2. Đặc trưng kết cấu**

Kết cấu là thuật ngữ chính được sử dụng để xác định các khái niệm hoặc đối tượng của một hình ảnh nhất định và đây là một tính năng quan trọng của ảnh, hỗ trợ tra cứu ảnh một cách hiệu quả hơn. Tính năng kết cấu được sử dụng để phân vùng hình ảnh thành các vùng quan tâm và phân loại các vùng đó. Đồng thời nó cũng cung cấp thông tin về sự sắp xếp không gian của màu sắc hoặc cường độ trong một hình ảnh. Ví dụ như trong lĩnh vực thị giác máy tính ta có thể phân tích kết cấu để nhận dạng vật thể, nhận dạng mẫu, nhận dạng khuôn mặt, phát hiện khuyết tật bề mặt, phân tích hình ảnh y tế hay thẩm mỹ...

Phân tích kết cấu đóng một vai trò quan trọng trong các trường hợp thị giác máy tính như nhận dạng vật thể, phát hiện khuyết tật bề mặt, nhận dạng mẫu, phân tích hình ảnh y tế... Kể từ đó, nhiều phương pháp tiếp cận đã được đề xuất để mô tả hình ảnh kết cấu một cách chính xác. Phương pháp phân tích kết cấu thường được phân thành bốn loại: phương pháp thống kê (statistical), phương pháp cấu trúc (structural), dựa trên mô hình (model-based) và dựa trên biến đổi (transform based).

### **1.2.3.3. Đặc trưng hình dạng**

Hình dạng trong thế giới thực là một đặc điểm quan trọng giúp cho con người có thể nhận dạng được sự vật và sự việc. Tương tự trong quy trình xử lý ảnh, hình dạng cũng là một đặc trưng quan trọng. Đặc trưng hình dạng của các vùng hay các đối tượng đã được sử dụng nhiều trong các hệ thống tra cứu CBIR. Khi so sánh với đặc trưng màu sắc và đặc trưng kết cấu, đặc trưng hình dạng thường được mô tả sau khi hình ảnh đã được phân đoạn thành các vùng đối tượng [15].

Mục đích của đặc trưng hình dạng là mã hóa các dạng hình học đơn giản (ví dụ như đường thẳng không định hướng). Các phương pháp trích xuất đặc

trung hình dạng có thể được chia thành hai nhóm, bao gồm phương pháp dựa trên đường biên và phương pháp dựa trên vùng. Phương pháp dựa trên đường biên tính toán các đặc điểm hình dạng bằng cách chỉ sử dụng ranh giới bên ngoài của hình dạng. Còn phương pháp dựa trên vùng thì tính toán các đặc điểm hình dạng bằng cách sử dụng toàn bộ hình dạng của vùng.

#### **1.2.4. Phản hồi liên quan trong tra cứu ảnh**

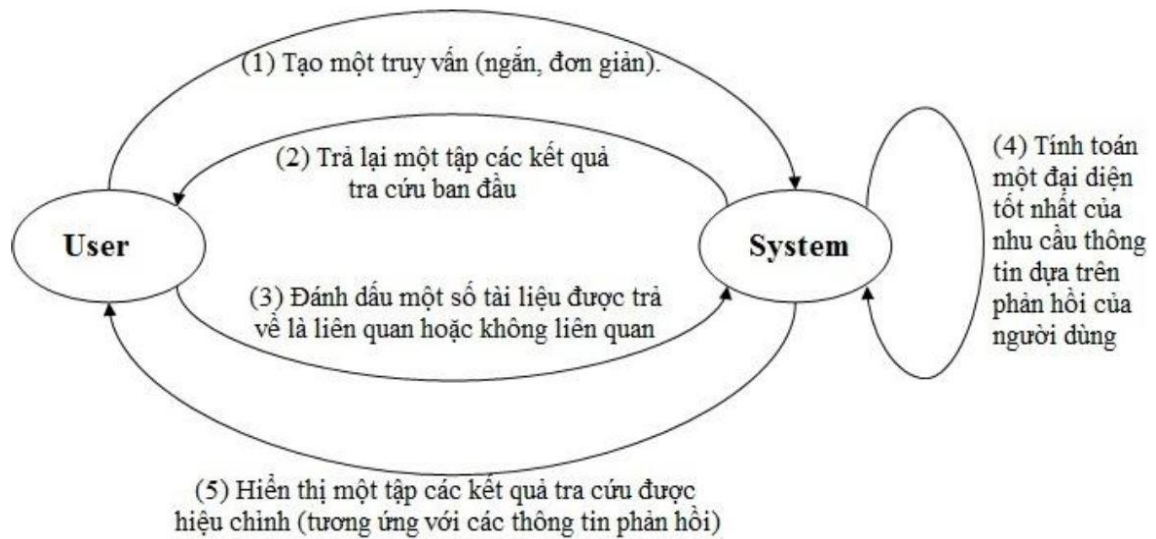
Phản hồi liên quan ( RF- Relevance feedback) vừa là đặc trưng của hệ thống truy hồi thông tin lẫn tra cứu ảnh. Phản hồi liên quan trong tra cứu ảnh dựa trên nội dung là một phương pháp mà phản hồi về hiệu năng/hiệu suất/hiệu quả (performance) đang được sử dụng để cải thiện chính nó [16].

Ý tưởng của phản hồi liên quan là để người dùng tham gia vào quá trình tra cứu để cải thiện tập kết quả cuối cùng, tức là để cho người dùng đưa ra phản hồi về mức độ liên quan của các tài liệu/dữ liệu trong một tập hợp kết quả ban đầu. Các bước cơ bản trong phương pháp phản hồi liên quan là [16]:

- Người dùng đưa ra một truy vấn (ngắn, đơn giản).
- Hệ thống trả về một bộ kết quả tra cứu ban đầu.
- Người dùng đánh dấu một số tài liệu/dữ liệu được trả lại là có liên quan hoặc không liên quan.
- Hệ thống tính toán/ước tính một phép biểu diễn tốt hơn về nhu cầu thông tin (information need) dựa trên phản hồi của người dùng.
- Hệ thống hiển thị một tập hợp các kết quả tra cứu đã được hiệu chỉnh tương ứng với những phản hồi được trả về.

Phản hồi liên quan có thể trải qua một hay nhiều vòng lặp. Quá trình này sử dụng ý tưởng rằng: "Có thể khó hình thành một truy vấn tốt khi chúng ta không biết rõ về toàn bộ các tập tài liệu/dữ liệu, nhưng rất dễ dàng để đánh giá các tài liệu/dữ liệu cụ thể. Do đó, việc tham gia vào việc sàng lọc truy vấn lặp đi lặp lại kiểu này là rất hợp lý." Trong kịch bản như vậy, phản hồi liên quan cũng có thể hiệu quả trong việc theo dõi nhu cầu thông tin đang phát triển của người dùng. Ví dụ như xem một số tài liệu/dữ liệu có thể giúp người dùng cải thiện vốn hiểu

biết của họ về thông tin mà đang tìm kiếm [16]. Hình 1.6 đã thể hiện rõ quá trình này.

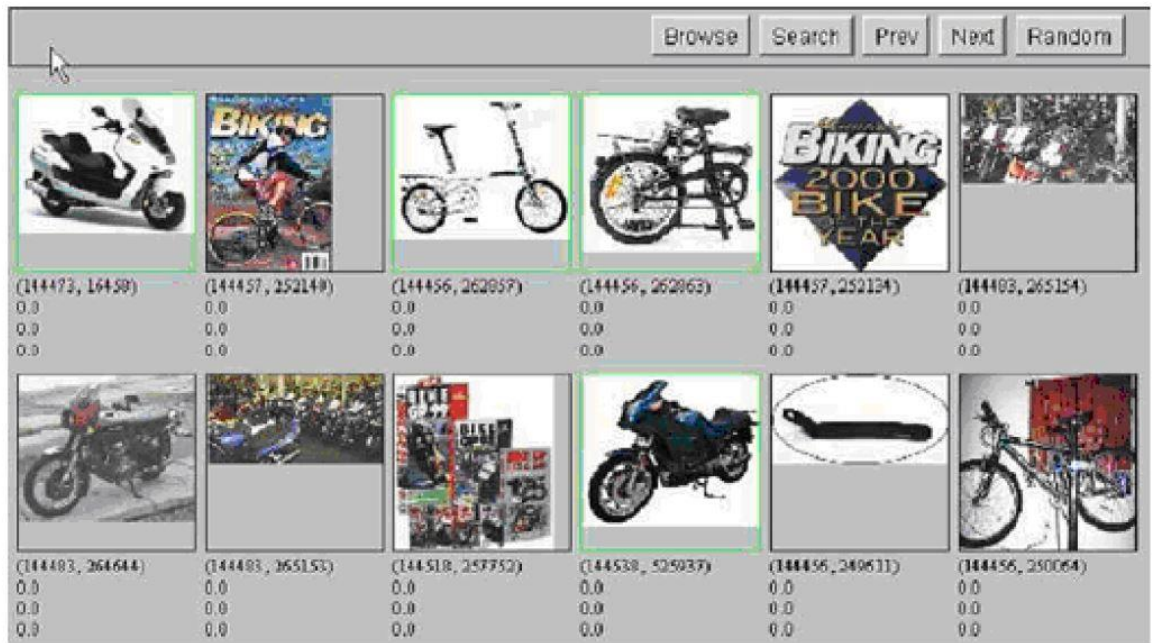


*Hình 6.6. Thủ tục phản hồi liên quan*

Tra cứu hình ảnh sử dụng phản hồi liên quan không chỉ dễ dàng xem được kết quả ngay tại nơi đã thực thi công việc, mà đây còn là miền nơi người dùng có thể gặp khó khăn trong việc hình thành những gì họ muốn bằng từ ngữ, nhưng lại có thể dễ dàng chỉ ra hình ảnh có liên quan hoặc không liên quan.

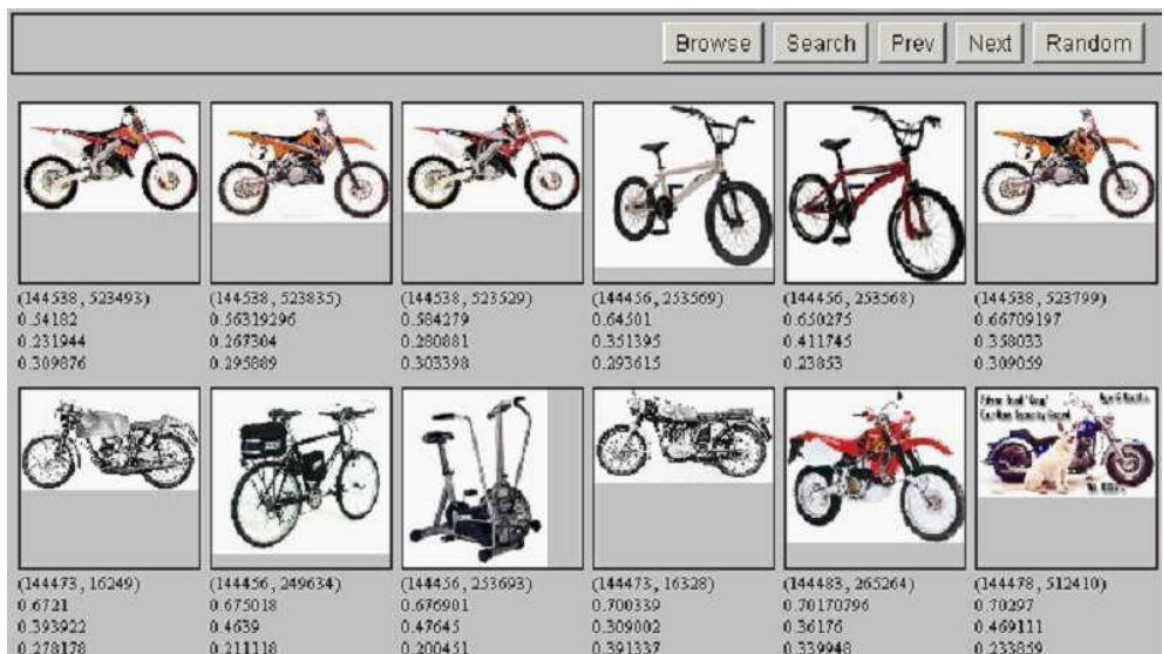
Ví dụ, giả sử như người dùng đã nhập vào một truy vấn khởi tạo là bike. Các kết quả ban đầu đã được trả về và dữ liệu trong trường hợp này là ảnh. Hình 1.7 cho thấy người dùng đã xem kết quả truy vấn ban đầu cho một truy vấn về xe bike, chọn kết quả đầu tiên, thứ ba và thứ tư ở hàng trên cùng và kết quả thứ tư ở hàng dưới cùng nếu có liên quan và gửi phản hồi này. Các hình ảnh liên quan này sẽ được sử dụng để sàng lọc truy vấn (refine the query), trong khi các kết quả hiển thị khác không ảnh hưởng đến việc tái diễn đạt (reformulation) lại truy vấn [16].





Hình 7.7. Phản hồi liên quan tìm kiếm trên các ảnh, người dùng xem các kết quả truy vấn ban đầu của truy vấn bike

Hình 1.8 chỉ ra các kết quả được phân hạng ở trên theo thứ tự mới được tính sau vòng phản hồi liên quan. Người dùng thấy tập hợp kết quả đã sửa đổi. Độ chính xác được cải thiện rất nhiều [16].



Hình 8.8. Phản hồi liên quan tìm kiếm trên các ảnh, người dùng xem tập kết quả được hiệu chỉnh

Nói tóm lại, phản hồi liên quan sẽ dựa vào khả năng của người dùng để truyền đạt những đánh giá của họ về mức độ liên quan giữa kết quả tìm kiếm với công cụ tìm kiếm. Sau đó, công cụ tìm kiếm sẽ sử dụng những phản hồi của người dùng về mức độ liên quan để xây dựng một truy vấn tối ưu, đồng thời sử dụng nó để tra cứu các kết quả đã được cải thiện rồi trả về cho người dùng, hoặc có thể xếp hạng lại kết quả. Mục đích của phản hồi liên quan chính là giảm tải cho người dùng khỏi các công đoạn tìm kiếm quá phức tạp, đẩy việc này cho các công cụ tìm kiếm và hỗ trợ các công cụ này cho ra được kết quả tốt hơn đúng như mong muốn của người dùng [17].

Cơ chế của phản hồi liên quan cho phép nó sửa đổi lặp đi lặp các tham số không gian (space parameter) và thông tin mô tả truy vấn, ví dụ như đặc trưng (features), ngữ nghĩa (semantics), phân lớp (classification), so khớp mẫu (matching models), số liệu (metrics) hay bất cứ loại hình siêu kiến thức (Meta knowledge) nào. Tất cả nhằm mục đích đưa người dùng tiếp cận được gần với những truy vấn tối ưu, đẩy độ chính xác của công đoạn tìm kiếm lên mức cao nhất có thể [17].

Có 4 cơ chế cho người dùng để có thể cung cấp mức độ về phản hồi liên quan [17]:

Số 1, phản hồi rõ ràng (Explicit feedback). Người dùng đưa ra phản hồi rõ ràng để đánh giá kết quả truy vấn. Hệ thống sử dụng nó để cải thiện công thức truy vấn.

Số 2, phản hồi ngầm (Implicit feedback). Hệ thống thu thập dữ liệu người dùng bằng hai cách, một là từ hành vi có thể quan sát được của người dùng, hai là xem xét kỹ lưỡng các hoạt động của người dùng trên chính hệ thống. Hoạt động của người dùng có thể là: sử dụng công cụ để tìm kiếm, tra cứu và lựa chọn hình ảnh; điều hướng trên web để xem các hình ảnh hoặc chọn các đường link; sử dụng các công cụ thao tác như là email, trình soạn thảo văn bản, xem các log file (một dạng như nhật ký cho phép chúng ta xem lại những sự kiện đã xảy ra trên hệ thống), xem cơ sở dữ liệu hoặc tài liệu...

Số 3, phản hồi xếp hạng (Ranking feedback). Người dùng xem xét một tập hợp con các kết quả tại một thời điểm và “sắp xếp” chúng theo thứ tự mà họ nghĩ rằng chúng sẽ xuất hiện. Theo một nghĩa nào đó, người dùng đang thực hiện nhiệm vụ của hệ thống tra cứu: để nó mô phỏng xếp hạng mà mình ưu thích hơn. Phản hồi xếp hạng cung cấp những phản hồi tuyệt vời cho hệ thống tra cứu, nhưng có xu hướng gây ra gánh nặng cho người dùng.

Số 4, phản hồi mù (Blind feedback), còn được gọi là phản hồi giả (pseudo feedback) hoặc phản hồi tự động (automatic feedback). Nó bao gồm việc tự động hóa phần thủ công của phản hồi liên quan, hệ thống sẽ mô phỏng tương tác của người dùng bằng cách giả định rằng K kết quả đầu tiên được hiển thị là phù hợp và phần còn lại thì không. Cơ chế phản hồi giả gặp phải việc tìm kiếm hội tụ đến mức tối ưu cục bộ.

### **1.2.5. Một số độ đo khoảng cách**

Kỹ thuật tra cứu hình ảnh dựa trên nội dung là giải pháp cho vấn đề tra cứu hình ảnh từ các cơ sở dữ liệu lớn. Ý nghĩa của tra cứu hình ảnh dựa trên nội dung (CBIR) là các phân tích tìm kiếm ảnh sẽ dựa trên nội dung thực tế của hình ảnh cụ thể đó chứ không phải là văn bản hay từ khóa được liên kết với nó. CBIR tra cứu hình ảnh bằng cách sử dụng các tính năng của đặc trưng, bao gồm: màu sắc (color), hình dạng (shape), kết cấu (texture) hay các thông tin không gian (spatial information) hoặc những thông tin có thể được tra cứu tự động từ chính hình ảnh. CBIR có thể sử dụng bất kỳ hoặc kết hợp tất cả các đặc trưng lại, sau đó tạo thành các ký số/ký hiệu (signature) mô tả nội dung của toàn bộ hình ảnh. Sau khi các đặc trưng được trích xuất, nó sẽ được lưu trữ trong cơ sở dữ liệu đặc trưng và các phép đo tương đồng sẽ được sử dụng để xác định khoảng cách giữa các ký số/ký hiệu của truy vấn và tất cả các hình ảnh trong tập dữ liệu. Việc này sẽ tra cứu các hình ảnh có khoảng cách nhỏ hơn đến hình ảnh truy vấn, và tất cả những hình ảnh được tra cứu này đều là những hình ảnh quan trọng đối với hình ảnh truy vấn [18].

Như vậy ta có thể thấy rằng, các phép đo tương đồng hay còn gọi là độ đo tương tự hoặc độ đo khoảng cách, đóng một vai trò quan trọng trong quá trình tra cứu hình ảnh dựa trên nội dung. Bởi nó tính toán sự tương đồng trực quan

giữa hình ảnh truy vấn và hình ảnh trong cơ sở dữ liệu được xếp hạng theo sự tương đồng của chúng với hình ảnh truy vấn. Các độ đo khác nhau có tác dụng khác nhau trong một hệ thống tra cứu hình ảnh. Phần này sẽ trình bày sơ lược một số độ đo khoảng cách thường được sử dụng khi tra cứu ảnh.

Khi đo khoảng cách giữa hai điểm ảnh, khoảng cách giữa hai điểm ảnh  $p$  tọa độ  $(x, y)$ ,  $q$  tọa độ  $(s, t)$  là hàm khoảng cách (Distance) hoặc Metric nếu [2]:

1.  $D(p, q) \geq 0$  (Với  $D(p, q) = 0$  và chỉ nếu  $p = q$ )
2.  $D(p, q) = D(q, p)$
3.  $D(p, z) \leq D(p, q) + D(q, z)$ ;  $z$  là một điểm ảnh khác

#### **1.2.5.1. Độ đo khoảng cách Euclide**

Khoảng cách Euclide (Euclidean Distance) là khoảng cách ngắn nhất giữa hai điểm. Các giá trị khoảng cách giữa các pixel so với pixel đang xét được tổ chức như một hình tròn. Khoảng cách Euclide giữa hai điểm ảnh  $p(x, y)$  và  $q(s, t)$  được định nghĩa như sau:

$$D_e(p, q) = [(x - s)^2 + (y - t)^2]^{1/2} \quad (1.1)$$

#### **1.2.5.2. Độ đo khoảng cách khối đồ thị**

Khoảng cách khối đồ thị (City-Block Distance) còn được gọi là khoảng cách Manhattan, đây là tổng giá trị tuyệt đối của hiệu giữa hai tọa độ của hai điểm. Các giá trị khoảng cách giữa pixel so với pixel đang xét được tổ chức như hình thoi. Gọi  $D_4(p, q)$  là khoảng cách giữa hai điểm  $p, q$  ta có công thức tính như sau:

$$D_4(p, q) = |x - s| + |y - t| \quad (1.2)$$

Trong trường hợp khoảng cách khối đồ thị, khoảng cách giữa 2 điểm được xác định là tổng khoảng cách ta đi dọc theo hướng  $x$  cộng với khoảng cách dọc theo hướng  $y$ . Tất cả các điểm có khoảng cách khối đồ thị từ điểm  $p$  nhỏ hơn hoặc bằng một số giá trị  $r$ , sẽ từ một trung tâm hình thoi có tâm tại điểm  $p$ . Vậy

$p$  là điểm ở tâm, thì tất cả các điểm có khoảng cách khối đồ thị, chúng là 4 lân cận của điểm  $p$ . Tương tự, tất cả các điểm có khoảng cách khối đồ thị bằng 2, chúng chỉ đơn giản là các điểm ở khoảng cách 2.

### 1.2.5.3. Độ đo khoảng cách bàn cờ

Khoảng cách bàn cờ (Chess-Board Distance) là khoảng cách lớn nhất giữa hai trị tuyệt đối của hiệu hai tọa độ. Các giá trị khoảng cách giữa pixel so với pixel đang xét được tổ chức như hình vuông. Gọi  $D_8(p, q)$  là khoảng giữa điểm ảnh  $p, q$  ta có công thức tính như sau:

$$D_8(p, q) = \max(|x - s|, |y - t|) \quad (1.3)$$

### 1.2.5.4. Độ đo khoảng cách Minkowski

Khoảng cách Minkowski là phép đo khoảng cách / độ tương tự giữa hai điểm trong không gian vectơ chuẩn (không gian thực N chiều). Nó còn là tổng quát của khoảng cách Euclide và khoảng cách City-Block. Với  $p$  là một tham số khác 0,  $p = 1$  là khoảng cách City-Block còn  $p = 2$  là khoảng cách Euclide, khoảng cách được xác định như sau:

$$D = ((|x - s|)^p + (|y - t|)^p)^{\frac{1}{p}} \quad (1.4)$$

## 1.3. KIẾN TRÚC CỦA HỆ THỐNG TRA CỨU ẢNH

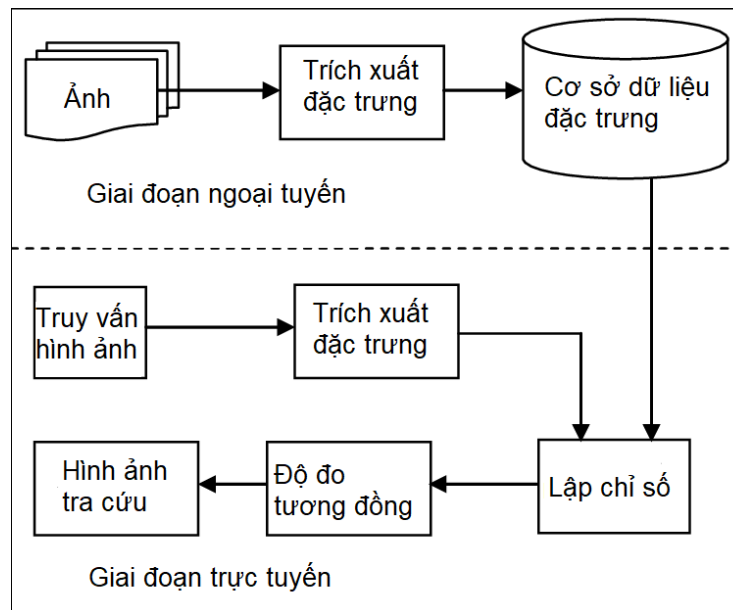
Hình 1.9 đã minh họa một kiến trúc điển hình của một hệ thống tra cứu ảnh dựa trên nội dung. Hệ thống bao gồm hai giai đoạn: giai đoạn ngoại tuyến và giai đoạn trực tuyến [4].

Trong giai đoạn ngoại tuyến, hệ thống tự động trích xuất đặc trưng của từng hình ảnh trong cơ sở dữ liệu và lưu trữ chúng trong cơ sở dữ liệu đặc trưng (các đặc trưng của hình ảnh được trích xuất và biểu diễn bằng các vector đặc trưng).

Trong giai đoạn trực tuyến, người dùng nhập một truy vấn hình ảnh vào hệ thống. Các đặc trưng của hình ảnh truy vấn được trích xuất và biểu diễn. Sự

tương đồng được đo giữa vector đặc trưng của hình ảnh truy vấn và vector đặc trưng của ảnh trong cơ sở dữ liệu. Sau đó, quá trình tra cứu được thực hiện bằng cách áp dụng một lược đồ lập chỉ số để tìm kiếm cơ sở dữ liệu hình ảnh một cách hiệu quả.

Cuối cùng, hệ thống trả về những hình ảnh giống nhất với hình ảnh truy vấn.



Hình 9.9. Kiến trúc của hệ thống tra cứu ảnh dựa trên nội dung

### 1.3.1. Các vấn đề nghiên cứu được quan tâm

Các vấn đề nghiên cứu và phát triển trong CBIR rất đa dạng và rộng mở, trong số đó có không ít chủ đề có liên quan và được ứng dụng trong lĩnh vực khai phá dữ liệu (data mining) hay truy hồi thông tin (information retrieval). Một số vấn đề được quan tâm nhất hiện nay là [19]:

1. Hiểu được nhu cầu của người dùng hình ảnh và hành vi tìm kiếm thông tin.
2. Xác định được cách nhận dạng phù hợp để mô tả nội dung hình ảnh.
3. Trích chọn các đặc trưng từ hình ảnh thô.
4. Cung cấp dung lượng lưu trữ nhỏ gọn cho cơ sở dữ liệu hình ảnh lớn.
5. So khớp truy vấn và hình ảnh được lưu trữ bằng cách ánh xạ/phản chiếu các phán đoán giống nhau của con người.

6. Truy cập hiệu quả tệp hình ảnh được lưu trữ theo nội dung.

7. Cung cấp các giao diện mà con người có thể sử dụng được cho hệ thống CBIR.

8. Lập chỉ số hiệu quả và tìm kiếm nhanh hình ảnh dựa trên các đặc trưng trực quan.

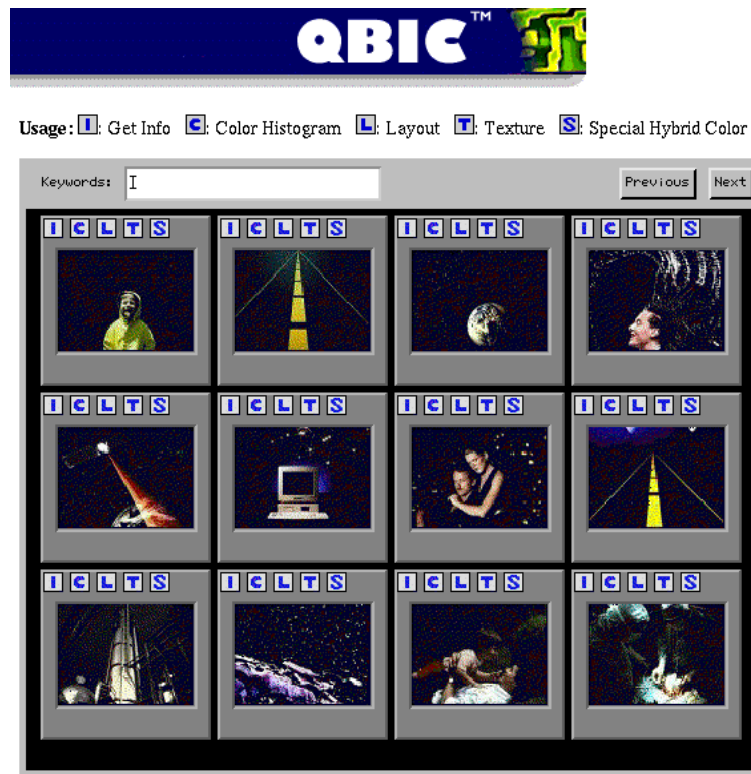
Ở mục số 7, đối với tra cứu hình ảnh dựa trên nội dung, tương tác của người dùng với hệ thống tra cứu là rất quan trọng bởi ta chỉ có thể có được sự hình thành và sửa đổi linh hoạt các truy vấn bằng cách cho người dùng tham gia vào quá trình tra cứu. Giao diện người dùng trong hệ thống tra cứu hình ảnh thường bao gồm phần xây dựng truy vấn và phần trình bày kết quả [20].

Các hình thức truy vấn thường được sử dụng là: duyệt theo danh mục (category browsing), truy vấn theo khái niệm (query by concept), truy vấn bằng phác thảo (query by sketch) và truy vấn bằng ví dụ (query by example). *Duyệt theo danh mục* là duyệt qua cơ sở dữ liệu theo danh mục của ảnh. *Truy vấn bằng khái niệm* là tra cứu hình ảnh theo mô tả khái niệm được liên kết với mỗi hình ảnh trong cơ sở dữ liệu. *Truy vấn bằng phác thảo* và *truy vấn bằng ví dụ* là vẽ một bản phác thảo hoặc cung cấp một hình ảnh ví dụ mà từ đó các hình ảnh có các đặc điểm trực quan tương tự sẽ được trích xuất từ cơ sở dữ liệu. *Truy vấn bằng bản phác thảo* cho phép người dùng vẽ bản phác thảo hình ảnh bằng công cụ chỉnh sửa đồ họa được cung cấp bởi hệ thống truy xuất hoặc một số phần mềm khác. *Truy vấn bằng ví dụ* cho phép người dùng hình thành truy vấn bằng cách cung cấp hình ảnh ví dụ [20].

Ở mục số 8, chỉ số trong CBIR cũng đóng một vai trò vô cùng quan trọng. Hình ảnh được xây dựng tương ứng với giá trị chỉ số trong mỗi pixel dựa trên sự kết hợp của các giá trị độ sáng trong kênh nhất định. Lập chỉ số (Indexing) là một kỹ thuật trong đó chỉ số cơ sở dữ liệu của mỗi hình ảnh được duy trì ở định dạng bảng có thể được sử dụng để truy cập và tìm bất kỳ hình ảnh nào [20] Các kỹ thuật lập chỉ số hiệu quả với một số cấu trúc dữ liệu như R\*-tree, Quad-tree, K-d-tree, Grid... đều là những vấn đề rất được quan tâm hiện nay.

## 1.3.2. Một số hệ thống tra cứu ảnh dựa trên nội dung

### 1.3.2.1. Hệ thống QBIC



Hình 10.10. Hệ thống QBIC

QBIC hay Query By Image Content (Truy vấn bằng nội dung hình ảnh) do hãng IBM và trung tâm nghiên cứu ALmaden hợp tác phát triển. Đây là ứng dụng của thị giác máy tính vào bài toán tra cứu ảnh, tức là bài toán tìm kiếm số hình ảnh trong cơ sở dữ liệu lớn và tìm kiếm sẽ dựa trên phân tích nội dung thực tế của ảnh. Hệ thống QBIC cho phép người dùng sử dụng những công cụ đồ họa để mô tả và hiệu chỉnh truy vấn trên nhiều đặc trưng như là màu sắc, hình dạng, kết cấu [21].

Có ba công đoạn trong quy trình của một ứng dụng QBIC: tập hợp cơ sở dữ liệu (database population), tính toán đặc trưng (feature calculation) và truy vấn hình ảnh (image query) [22].

Bước đầu tiên trong việc *tập hợp cơ sở dữ liệu* là chỉ cần tải các hình ảnh vào hệ thống. Điều này liên quan đến việc thêm hình ảnh vào cơ sở dữ liệu. Ví dụ, ta có thể chuẩn bị một hình ảnh thu nhỏ (thumbnail) 100x100 được giảm bớt



và thêm bất kỳ thông tin nào từ văn bản có sẵn vào cơ sở dữ liệu. Việc nhận dạng đối tượng/phần tử là một phần không bắt buộc của bước này. Hệ thống QIBC cho phép người dùng xác định thủ công hoặc bán tự động các đối tượng (object), khu vực (area) hoặc các cạnh (edge) quan trọng trong một cảnh của hình ảnh bằng cách sử dụng chuột. Bên trong mỗi đối tượng hoặc khu vực được xác định sẽ trở thành một binary mask. Binary mask sẽ xác định vùng quan tâm (ROI - region of interest) của một hình ảnh. Bước thứ hai là phác thảo tương tác (Interactive Outlining). Trong thao tác này, người dùng cung cấp một số thông tin ban đầu mà hệ thống sử dụng, cùng với các phương pháp phân tích hình ảnh, để tính toán phác thảo đối tượng. [22]

Quy trình *tính toán đặc trưng* do hệ thống thực hiện. Các đặc trưng bao gồm màu sắc, kết cấu, hình dạng. QBIC tính toán biểu đồ màu (color histogram) qua các đặc trưng màu và đặc trưng kết cấu của nó dựa trên không gian đặc trưng ba chiều: độ thô (coarseness), tương phản (contrast) và định hướng (directionality). Đối với đặc trưng hình dạng, QBIC sử dụng kết hợp tính tuần hoàn, độ lệch tâm, hướng trục chính và một tập hợp các đại số bất biến. [23]

Khi tập hợp các đặc trưng cho các đối tượng và hình ảnh đã được tính toán, các truy vấn hình ảnh có thể được chạy. QBIC hỗ trợ truy vấn toàn cảnh (full scene query). Các hàm tương tự mà nó sử dụng chủ yếu là các chỉ số về khoảng cách, ví dụ như City-Block, Euclidean. QBIC cũng hỗ trợ truy vấn bằng phác thảo (query by sketch), thực hiện bằng cách ghép các cạnh do người dùng vẽ để tự động trích xuất các cạnh từ hình ảnh trong cơ sở dữ liệu [23].

### **1.3.2.2. Hệ thống Photobook**

Hệ thống Photobook là một bộ công cụ tương tác dành cho duyệt, truy vấn tìm kiếm hình ảnh cùng video clip được phát triển tại phòng thí nghiệm truyền thông MIT. Giống như QIBC, Photobook hoạt động bằng cách so sánh các đặc trưng được trích xuất từ hình ảnh, bao gồm đặc trưng hình dạng, kết cấu và màu sắc được trích xuất tương ứng. Các tiêu chí so khớp bao gồm Euclidean, Mahalanobis, phân kỳ (divergence), biểu đồ (histogram), các đỉnh Fourier hoặc bất kỳ sự kết hợp tuyến tính nào của chúng [23][24].

Photobook có ba kiểu mô tả hình ảnh, giống như ba khối con (sub-books), người dùng có thể truy vấn hệ thống dựa trên các đặc trưng từ ba khối con này. Mô tả ngoại hình được áp dụng cho cơ sở dữ liệu khuôn mặt và khung hình chính. Mô tả kết cấu được áp dụng cho kết cấu Brodatz - tập dữ liệu kết cấu được sử dụng rất phổ biến, đặc biệt là trong lĩnh vực xử lý tín hiệu và thị giác máy tính và khung hình chính của cơ sở dữ liệu. Mô tả hình dạng được áp dụng cho công cụ thủ công và cơ sở dữ liệu fish (FishBase) - cơ sở dữ liệu toàn cầu và trực tuyến lớn nhất về các loại cá (đặc biệt là bộ cá vây tay) [23][24].

Đặc trưng chính của Photobook là nó bao gồm cả FourEyes (Minka 1996). FourEyes là phiên bản phát triển của Photobook, nó bao gồm phản hồi liên quan (relevance feedback) [23][24].

### **1.3.2.3. Hệ thống VisualSEEK và WebSEEK**

Hệ thống VisualSeek là một phần của hệ thống CBVIR (Content-based visual information retrieval – Tra cứu thông tin trực quan dựa trên nội dung) được phát triển tại đại học Columbia. Nó hỗ trợ các truy vấn dựa trên các đặc trưng trực quan và các mối quan hệ không gian (spatial relationship). Phiên bản hiện tại của VisualSeek chủ yếu dành cho việc tìm kiếm hình ảnh thông qua World Wide Web và nó có kho lưu trữ lên tới 12000 hình ảnh [23].

VisualSeek bao gồm ba phần. Phần một là *client application* hay *ứng dụng khách*, đây là một bộ ứng dụng Java, thực thi trong trình duyệt WWW, thu thập truy vấn từ người dùng và tạo chuỗi truy vấn (query string). Phần hai là *network and communication application* hay *ứng dụng máy chủ*, nó xử lý tất cả thông tin liên lạc thông qua WWW. Phần ba là *server application* hay *ứng dụng máy chủ*, nó nhận chuỗi truy vấn và trả về kết quả cho người dùng. Chương trình máy chủ tạo code Hypertext Markup Language (HTML) hiển thị kết quả truy vấn cho người dùng [24].

# VisualSEEK

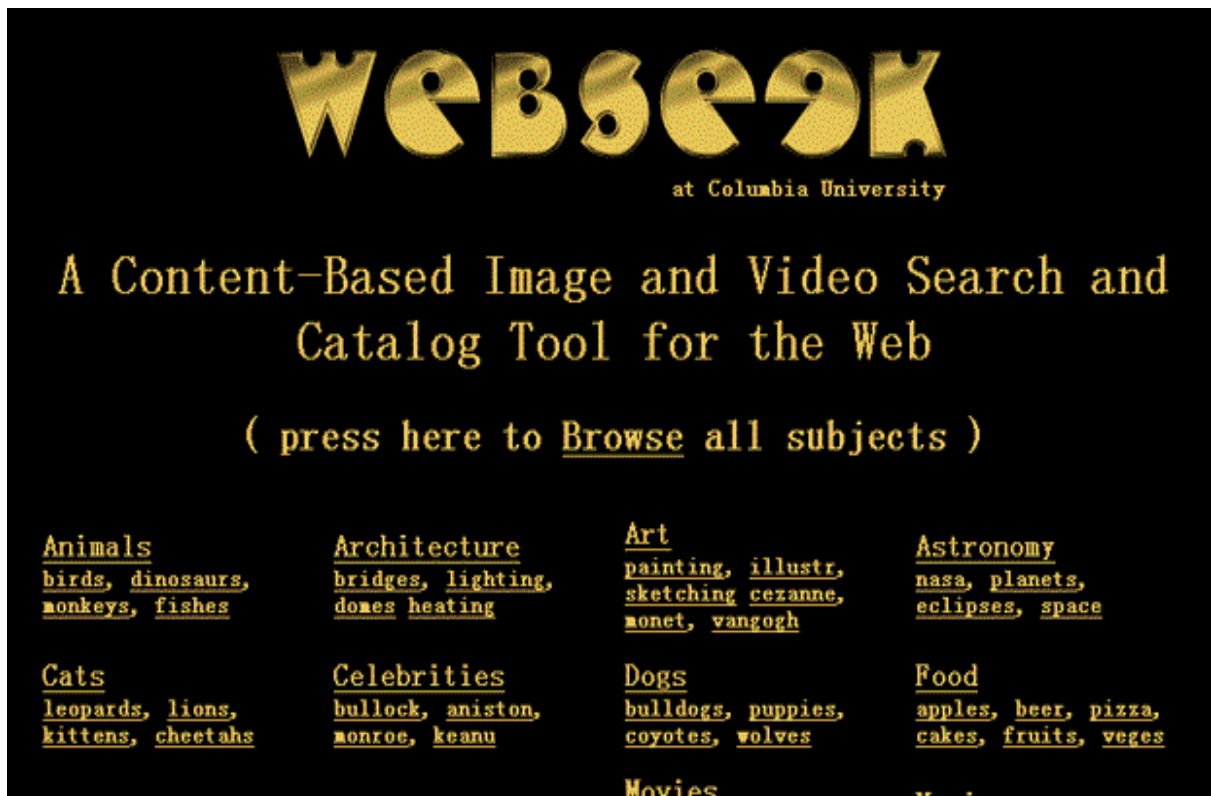
*Colour histogram query*



Hình 11.11. Hệ thống VisualSeek

WebSEEK là một hệ thống truy vấn ảnh và video dựa trên nội dung đồng thời là một công cụ tìm kiếm cho World Wide Web được phát triển tại đại học Columbia. Nó thu thập hình ảnh và video bằng cách sử dụng một số tác nhân Web tự động, tự động phân tích, lập chỉ số và gắn hình ảnh và video cho các lớp chủ đề. Hệ thống này sử dụng đồng thời cả thông tin văn bản lẫn hình ảnh để cung cấp cho việc lập danh chỉ số lẫn tìm kiếm hình ảnh và video [25].

Một hệ thống Webseek sẽ có những chức năng sau: Tìm kiếm sử dụng kỹ thuật dựa trên nội dung hình ảnh; Sửa đổi truy vấn bằng cách sử dụng phản hồi liên quan dựa trên nội dung, Tự động thu thập thông tin trực quan; Trình bày nhỏ gọn hình ảnh và video để hiển thị kết quả truy vấn; Tìm kiếm chủ đề hình ảnh và video cùng điều hướng; Tìm kiếm dựa trên văn bản. Kết quả tìm kiếm bao gồm cả các thao tác như tương giao (intersection), loại trừ (subtraction) và ghép nối (concatenation). Hiện tại, hệ thống Weekseek đã lập danh mục hơn 650.000 hình ảnh và 10.000 video từ Web [25].



Hình 12.12. Hệ thống WebSeek

#### 1.3.2.4. Hệ thống Virage

Virage hay Virage Engine là một công cụ tìm kiếm hình ảnh dựa trên nội dung thương mại được phát triển tại Virage, Inc và được thực hiện tại đại học California tại San Diego. Đây là một công cụ mở, có kiến trúc di động và có thể mở rộng để kết hợp với bất kỳ miền cụ thể nào trên lược đồ (schema) [24].

Virage áp dụng ba cấp độ trừu tượng hóa thông tin cho hình ảnh: *Hình ảnh thô* hay *raw image* (cấp độ đại diện hình ảnh); *Hình ảnh đã xử lý* hay *processed image* (cấp độ đối tượng hình ảnh); *Đặc điểm người dùng đã quan tâm* (cấp độ đối tượng miền). Các đặc trưng trực quan được tính toán của hình ảnh có tên là "Nguyên thủy (primitives)". Một số đặc trưng nguyên thủy phổ biến của Virage là global color, local color, cấu trúc và kết cấu [24].

Virage được chuyển giao dưới dạng thống kê hoặc thư viện có thể liên kết động cho nhiều nền tảng khác nhau như SUN, SGI, Windows và MAC. Thư viện này chứa nhiều dạng cơ sở dữ liệu và ứng dụng framework khác nhau. Những cơ sở dữ liệu như Oracle, Sybase hay các công cụ tổng hợp và xử lý hình ảnh như Photoshop, CorelDraw đều có thể sử dụng cho Virage để tra cứu và quản lý hình ảnh [24].

#### 1.4. LĨNH VỰC ỨNG DỤNG CỦA TRA CỨU ẢNH

Trong xã hội hiện đại, khi nhu cầu sử dụng ảnh số ngày càng trở nên phổ biến thì các hệ thống tra cứu ảnh cũng được nghiên cứu, phát triển và cải tiến mỗi ngày. Chúng ta có thể dễ dàng tra cứu hình ảnh thông qua các app ứng dụng trên di động ví dụ như Google Lens, hoặc sử dụng các trang tìm kiếm như là trang <http://images.google.com/>. Các hệ thống này sử dụng đa dạng các truy vấn đầu vào như chú thích ảnh, hoặc kết hợp của nhiều đặc trưng như màu sắc, kết cấu, vị trí không gian...

Tra cứu ảnh được sử dụng hiệu quả trong lĩnh vực an ninh điều tra. Một số cơ sở dữ liệu phục vụ cho công việc này: cung cấp công nghệ xác minh danh tính - (<http://www.viisage.com/>), dấu vân tay - (<http://www.east-shore.com/>), dấu chân (<http://www.fosterfreeman.co.uk/sicar.html>)

Ngoài ra, tra cứu ảnh cũng có thể được ứng dụng trong lĩnh vực y tế sức khỏe - (<http://www.brisbio.ac.uk/>), đơn cử là trong các hoạt động khám chữa bệnh như siêu âm, chụp cắt lớp, chụp cộng hưởng, chụp nhũ ảnh, chụp X-quang. Thông qua các hình ảnh được chụp chiếu và phân tích, bác sĩ có thể dựa vào đó để tiến hành phương pháp điều trị bệnh phù hợp cho bệnh nhân. Không chỉ vậy, tra cứu ảnh còn có thể được sử dụng trong lĩnh vực địa - vật lý, ví dụ như nghiên cứu địa chất. Hoặc rộng hơn nữa là ngành khoa học vũ trụ, ta có thể chụp được ảnh ngoài vũ trụ thông qua kính viễn vọng. Không chỉ thế, tra cứu ảnh còn được sử dụng trong lĩnh vực mỹ thuật đồ họa, phim ảnh, trò chơi điện tử, vân vân.

#### **Kết luận Chương 1:**

Trong kỷ nguyên công nghệ 4.0, các yếu tố như Internet vạn vật (Internet of things), Trí tuệ nhân tạo (Artificial intelligence), và Dữ liệu lớn (Big data) ngày càng trở nên quen thuộc. Internet xuất hiện tạo nên một sự đổi mới trong công cuộc tìm kiếm, lưu trữ và chia sẻ thông tin. Ảnh số là một dạng thông tin được sử dụng rất phổ biến. Sự tiện dụng của các cơ sở dữ liệu ảnh số yêu cầu những phương pháp tra cứu phù hợp. Chương 1 của luận văn đã trình bày một số khái niệm, thuật ngữ, các phương pháp tra cứu ảnh cơ bản, bao gồm: tra cứu ảnh dựa trên văn bản, nội dung, kết hợp với một số độ đo khoảng cách, kỹ thuật phản hồi liên quan. Đây là chương nền tảng để mở rộng nghiên cứu sang các chương tiếp.

## **CHƯƠNG 2. CÁC PHƯƠNG PHÁP TRA CỨU ẢNH DƯỠNG TRÊN NỘI DUNG VÀ PHƯƠNG PHÁP PHẢN HỒI LIÊN QUAN**

### **2.1. GIỚI THIỆU**

Trích chọn đặc trưng (Feature Selection) là một kỹ thuật được sử dụng rất rộng rãi trong các ngành khoa học, đặc biệt là ngành khoa học liên quan đến dữ liệu. Cơ chế của trích chọn đặc trưng, chính là làm giảm đi các yếu tố đầu vào để xử lý và phân tích, hoặc tìm kiếm các yếu tố đầu vào có ý nghĩa nhất. Mục đích của việc này là để lựa chọn các đặc trưng phù hợp nhất, cải thiện hiệu suất mô hình dự đoán và làm chi phí tính toán của mô hình.

Ví dụ, trong lĩnh vực trí tuệ nhân tạo (AI - Artificial intelligence), khi xây dựng mô hình học máy (machine learning) trong đời thực, rất hiếm có trường hợp mà tất cả các biến trong tập dữ liệu đều hữu ích để xây dựng mô hình. Việc có thêm các biến dư thừa sẽ làm giảm khả năng tổng quát hóa của mô hình và cũng có thể làm giảm độ chính xác tổng thể của bộ phân loại (classifier). Hơn nữa, việc thêm ngày càng nhiều biến vào một mô hình sẽ làm tăng độ phức tạp tổng thể của chính mô hình đó.

Vậy nên việc trích chọn đặc trưng ở đây là rất cần thiết. Trích chọn đặc trưng không chỉ làm cho quá trình học máy chính xác hơn, mà còn làm tăng khả năng dự đoán của các thuật toán bằng cách chọn các biến quan trọng nhất, đồng thời loại bỏ những biến dư thừa và không liên quan.

#### **2.1.1. Trích chọn đặc trưng sử dụng các phương pháp gói (Wrapper methods)**

Phương pháp gói phụ thuộc vào hiệu suất của các bộ phân loại (classifier) để thu được một tập hợp con đặc trưng. Phương pháp gói chọn tập hợp con đặc trưng bằng cách sử dụng thuật toán cảm ứng (induction algorithm) như một hộp đen (tức là không cần kiến thức về thuật toán, chỉ cần giao diện là được). Độ chính xác của bộ phân loại cảm ứng được ước tính bằng cách sử dụng các kỹ thuật ước tính độ chính xác. Vấn đề chính của phương pháp gói là tìm kiếm không gian trạng thái và các công cụ tìm kiếm khác nhau cho các phương pháp khác nhau [26].

Số lượng kỹ thuật tìm kiếm khác nhau có thể được sử dụng để tìm tập hợp con tốt nhất của các đặc trưng tối đa hóa hiệu suất phân loại. Ví dụ: Phương pháp nhánh và ràng buộc (Branch and Bound); Thuật toán di truyền (GA - Genetic Algorithm); Tối ưu bầy đàn (PSO - Particle Swarm Optimization). Mục tiêu của tìm kiếm là tìm trạng thái có đánh giá tối đa, sử dụng phương pháp thử và sai (trial and error) để hướng dẫn nó. *Thuật toán lựa chọn tuần tự (Sequential selection)* và *thuật toán tìm kiếm tiến hóa (Evolutionary search)* là hai loại chính của phương pháp Wrapper [26].

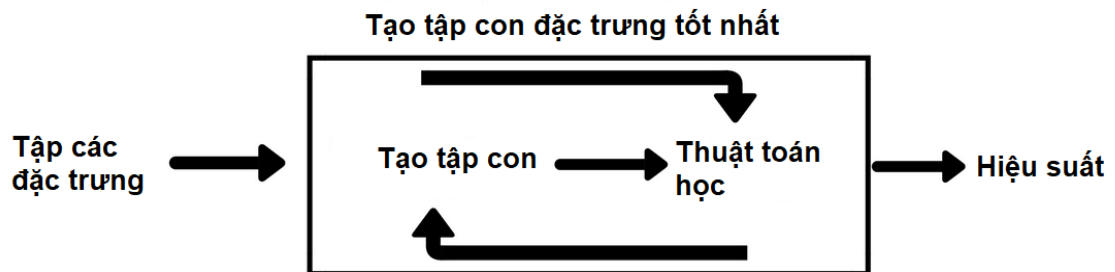
*Thuật toán lựa chọn tuần tự* tìm các đặc trưng tối thiểu (hoặc tối đa) bằng cách lặp lại quy trình. Thuật toán chọn tính năng tuần tự (SFS - Sequential Feature Selection) bắt đầu với một tập hợp trống và thêm một đặc trưng cho bước đầu tiên để tăng hiệu suất của hàm chức năng. Từ bước thứ hai trở đi, các đặc trưng còn lại được thêm riêng lẻ vào tập hợp con hiện tại và hiệu suất của tập hợp con mới được tính toán. Bằng quy trình này, ta có thể tìm thấy tập hợp con đặc trưng tốt nhất mang lại độ chính xác tối đa của bộ phân loại [26].

Mục tiêu của *thuật toán tìm kiếm tiến hóa* là tạo ra một giải pháp trong một khung thời gian hợp lý đủ tốt để giải quyết vấn đề trong tầm tay. Các thuật toán tìm kiếm tiến hóa đánh giá các tập hợp con khác nhau để tối ưu hóa hiệu suất của hàm mục tiêu. Các tập hợp con đặc trưng khác nhau được tạo ra bằng cách tìm kiếm xung quanh trong không gian tìm kiếm hoặc bằng cách tạo ra các giải pháp cho vấn đề tối ưu hóa. Các thuật toán tiến hóa dựa trên những ý tưởng của quá trình tiến hóa sinh học, chẳng hạn như tái tạo (reproduction), đột biến (mutation) và tái tổ hợp (recombination), để tìm kiếm giải pháp của một bài toán tối ưu hóa. Vòng lặp chính của các thuật toán tiến hóa bao gồm các bước sau [26]:

- Bước 1, khởi tạo và ước tính tập hợp ban đầu.
- Bước 2, thực hiện lựa chọn cạnh tranh (competitive selection).
- Bước 3, áp dụng các toán tử tiến hóa khác nhau để tạo ra các giải pháp mới.
- Bước 4, ước tính các giải pháp cho tập hợp.

- Bước 5, lặp lại từ bước thứ hai, cho đến khi một số tiêu chí hội tụ được đáp ứng.

Mô hình hoạt động của phương pháp gói được thể hiện trong Hình 2.1:



Hình 2. 1. Sơ đồ khối phương pháp Wrapper

Một số ví dụ phổ biến của phương pháp gói:

- Lựa chọn chuyển tiếp (Forward Selection): Lựa chọn chuyển tiếp là một phương pháp lặp đi lặp lại. Nó được bắt đầu với việc không có đặc trưng nào trong mô hình, tức là tập hợp các đặc trưng trống. Sau mỗi lần thực hiện vòng lặp, các đặc trưng tốt nhất sẽ được lựa chọn ra và thêm vào trong tập trống. Nhược điểm chính kỹ thuật này là không thể loại bỏ các đặc trưng không quan trọng sau khi bổ sung các đặc trưng khác.

- Loại bỏ theo chiều ngược (Backward Elimination): Trong loại bỏ theo chiều ngược, quy trình bắt đầu với tất cả các đặc trưng của dữ liệu. Sau mỗi vòng lặp thì những đặc trưng ít quan trọng nhất sẽ được loại bỏ để cải thiện hiệu suất của mô hình. Loại bỏ theo chiều ngược hoạt động tốt nhất với một số lượng lớn các đặc trưng trong tập dữ liệu.

- Loại bỏ đặc trưng đệ quy (Recursive Feature elimination): Đây là một thuật toán tối ưu hóa tham lam (greedy optimization) với mục đích tìm ra tập hợp con đặc trưng hoạt động tốt nhất. Trong các vòng lặp, nó liên tục tạo ra các mô hình, xác định và giữ lại các đặc trưng tốt nhất hoặc tệ nhất. Sau đây, nó xây dựng mô hình tiếp theo với các tập đặc trưng được chọn cho đến khi tất cả các đặc trưng đã được sử dụng hết. Cuối cùng, nó xếp hạng các đặc trưng dựa trên thứ tự loại bỏ của chúng. Đối với trường hợp xấu nhất, giả sử như có một tập dữ



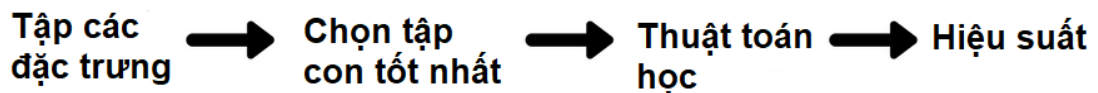
liệu với số lượng đặc trưng là  $K$ , thì phương pháp loại bỏ đệ quy này có thể tìm kiếm lên tới  $2^K$  tổ hợp thuộc tính.

### 2.1.2. Trích chọn đặc trưng sử dụng các phương pháp lọc (Filter methods)

Phương pháp lọc (Filter method) lọc thường được sử dụng như một bước tiền xử lý. Phương pháp lọc đánh giá mức độ liên quan của các đặc trưng bằng cách sử dụng quy trình xếp hạng hay phương pháp xếp hạng biến số (variable ranking) làm tiêu chuẩn để chọn biến số theo thứ tự. Các đặc trưng được xếp hạng dựa trên điểm của chúng trong các bài test thống kê khác nhau về mối tương quan của chúng với lớp (class). Các đặc trưng đạt điểm dưới một ngưỡng nhất định sẽ bị xóa, trong khi các đặc trưng đạt điểm trên ngưỡng đó sẽ được chọn. Tập hợp con các đặc trưng được chọn có thể trở thành một vào đầu vào cho thuật toán phân loại đã chọn.

Các phương thức lọc được cho là nhanh, có thể mở rộng, tính toán đơn giản và độc lập với bộ phân loại. Ngoài ra thì kỹ thuật xếp hạng thống kê (statistical ranking) cũng được sử dụng.

Quy trình của một mô hình lọc được thể hiện trong Hình 2.2:



Hình 2. 2. Sơ đồ khối phương pháp Filter

Không giống như phương pháp gói (wrapper) ở phần 2.1.1 hay phương pháp nhúng (embedded) sẽ được giới thiệu ở phần 2.1.3, phương pháp lọc (filter) độc lập và tách biệt với thuật toán phân loại. Sự tách biệt này có nghĩa là các phương pháp lọc không bị ảnh hưởng bởi sự thiên vị hay sự chệch lệch của bộ phân loại (classifier's bias), mà điều này làm giảm việc trang bị quá mức (overfitting). Tuy nhiên, sự độc lập này cũng có nghĩa là sự tương tác với bộ phân loại không được xem xét trong quá trình lựa chọn đặc trưng. Do vậy tập đặc trưng được chọn sẽ tổng quát hơn và không được tinh chỉnh cho phù hợp với bất kỳ bộ phân loại cụ thể nào. Việc thiếu điều chỉnh này làm cho các phương pháp lọc có xu hướng tạo ra các mô hình làm giảm hiệu suất dự đoán so với các mô hình được tạo ra bởi phương pháp lọc hay phương pháp nhúng [27].

Ưu điểm chính phương pháp lọc so với các phương pháp lựa chọn đặc trưng khác là chúng thường ít đòi hỏi tính toán hơn và do đó có thể dễ dàng được chia tỷ lệ thành kiểu dữ liệu có chiều rất cao hay *high dimensional data*. Đây là kiểu dữ liệu có số lượng đặc trưng vượt qua số lượng quan sát (observation) [27].

Các phương pháp lọc được chia thành hai loại: *phương pháp lọc đơn biến (univariate filter)* và *phương pháp lọc đa biến (multivariate filter)*. Các phương pháp đơn biến đánh giá các đặc trưng một cách độc lập, do đó bỏ qua các phụ thuộc của đối tượng và dẫn đến các tập con tính năng kém. Không giống như các phương pháp đơn biến bỏ qua các phụ thuộc đặc trưng và tương tác với thuật toán phân loại, các phương pháp đa biến xem xét hai yếu tố này ở một mức độ nhất định [28].

Phần dưới đây sẽ giới thiệu một số phương pháp lọc phổ biến, trong đó, 2.1.2.1 và 2.1.2.2 nằm trong *phương pháp lọc đơn biến* và 2.1.2.3 và 2.1.2.4 nằm trong *phương pháp lọc đa biến* [29].

### **2.1.2.1 Information Gain (IG)**

*Information Gain (IG)* hay *độ lợi thông tin*, đây là một phương pháp dựa trên lý thuyết thông tin của entropy, được sử dụng rộng rãi trên *dữ liệu chiều cao* hay *high dimensional data*, ví dụ như phân loại văn bản. Entropy là đơn vị đo lường tính ngẫu nhiên hay sự rối loạn của một hệ thống. Phương pháp IG tính toán lượng thông tin theo từng bit liên quan đến dự đoán lớp (class prediction), trong trường hợp khi thông tin duy nhất có sẵn là sự hiện diện của một biến và mục tiêu phân bố (hoặc lớp) tương ứng. Ngoài ra, nó đo lường mức giảm entropy dự kiến để quyết định mức độ quan trọng của một đặc trưng nhất định. Một hàm entropy tăng khi phân bố lớp trở nên thưa thớt hơn và nó có thể được áp dụng đệ quy để tìm các tập con entropy. Phương trình sau cung cấp một hàm entropy thỏa mãn hai yêu cầu:

$$H(D) = - \sum_{i=1}^c \frac{n_i}{n} \log \left( \frac{n_i}{n} \right) \quad (2.1)$$

Trong đó  $D$  là tập dữ liệu,  $n$  là số thực thể/đối tượng/phiên bản (instance) có trong  $D$ ,  $n_i$  là đại diện cho các thành viên trong lớp  $i$ ,  $C$  là số lớp. Phương trình tiếp sau đây biểu diễn entropy của các tập con:

$$H(D|X) = \sum_j \left( \frac{|x_j|}{n} \right) H(D|x = x_j) \quad (2.2)$$

Trong đó  $H(D|x = x_j)$  đại diện cho entropy tương quan với tập con của các thực thể/đối tượng/phiên bản, với giả định là giá trị  $x_j$  cho đặc trưng  $x$ . Ví dụ: Khi  $x$  cung cấp một mô tả tốt về lớp (class), giá trị được liên kết với đặc trưng đó sẽ giả định một giá trị thấp của entropy trong phân phối lớp của nó. Cuối cùng, IG được định nghĩa là sự giảm entropy như sau:

$$IG(X) = H(D) - H(D|X) \quad (2.3)$$

Giá trị cao của  $IG$  chỉ ra rằng  $X$  là một đặc trưng quan trọng đối với hiện tượng được xem xét.

### 2.1.2.2. Chi-squared Statistics

Chi-squared Statistics hay còn gọi là thống kê chi bình phương, là phương pháp đánh giá các biến riêng lẻ bằng cách đo lường thống kê chi bình phương (chi-squared) của chúng. Phép kiểm định cung cấp điểm số tuân theo phân phối chi bình phương với mục tiêu để xếp hạng tập hợp các đặc trưng đầu vào. Cách tiếp cận này mặc dù được sử dụng rộng rãi nhưng nó không tính đến sự tương tác của các đặc trưng.

Nếu chúng ta giả sử rằng biến lớp (class variable) là nhị phân thì giá trị chi bình phương để tính điểm thuộc về biến  $v$  cho lớp  $k$  được đánh giá như sau:

$$X^2(D, k, v) = \sum_{i=1}^N \left[ \frac{(n_{i+} - u_{i+})^2}{u_{i+}} + \frac{(n_{i-} - u_{i-})^2}{u_{i-}} \right] \quad (2.4)$$

Trong đó  $D$  là tập dữ liệu được xem xét,  $N$  là số biến đầu vào,  $n_{i+}$  là số mẫu có lớp dương cho biến  $i$  và cuối cùng  $u_{i+}$  đại diện cho giá trị mong đợi (expected value) nếu có bất kỳ mối quan hệ nào giữa  $v$  và  $k$ .

Trong thống kê, kiểm định chi bình phương được sử dụng để xác minh xem hai sự kiện có độc lập hay không. Trong lựa chọn đặc trưng, thống kê chi bình phương thực hiện một kiểm định giả thuyết về phân phối của lớp (class), bởi vì nó liên quan đến độ đo của biến đang được xem xét. Giả thuyết null thể hiện sự thiếu của mối tương quan.

### 2.1.2.3. Correlation

*Correlation* hay *cách tiếp cận tương quan*, được sử dụng trong lựa chọn đặc trưng, bao gồm tính toán hệ số tương quan giữa đặc trưng và mục tiêu - target (hoặc lớp - class trong trường hợp có vấn đề về phân loại). Một đặc trưng được chọn nếu nó có tương quan cao với lớp nhưng không tương quan với các đặc trưng còn lại. Có hai cách tiếp cận khác nhau để đánh giá mối tương quan giữa hai biến số: *mối tương quan tuyến tính cổ điển* và *mối tương quan dựa trên lý thuyết thông tin*. Đối với hệ số tương quan tuyến tính, nó được tính theo phương trình sau:

$$c = \frac{\sum(x_i - \mu_{xi})(y_i - \mu_{yi})}{\sqrt{\sum(x_i - \mu_{xi})^2} \sqrt{\sum(y_i - \mu_{yi})^2}} \quad (2.5)$$

Trong đó  $x, y$  là hai biến được xem xét, trong khi  $\mu_x$  và  $\mu_y$  là giá trị trung bình của chúng. Hệ số tương quan tuyến tính  $c$  nằm trong khoảng  $[-1,1]$ . Nếu hai biến có tương quan tuyến tính thì  $|c| = 1$ , còn nếu chúng độc lập thì  $c$  giả sử là giá trị null. Cách tiếp cận này có hai ưu điểm chính: nó loại bỏ các đặc trưng có hệ số tương quan rất thấp và giảm bớt sự dư thừa. Mặt khác, phương pháp tương quan tuyến tính không phác thảo đầy đủ các tương quan phi tuyến tính, thường xảy ra khi xử lý với các bộ dữ liệu trong thực tế.

### 2.1.2.4. Fast Correlation Based Filter (FCBF)

*Fast Correlation Based Filter (FCBF)* hay còn gọi là *bộ lọc nhanh dựa trên tương quan*, là một phương pháp lựa chọn đặc trưng đa biến. Nó bắt đầu với một tập hợp đầu đủ các đặc trưng, sử dụng độ bất định đối xứng (symmetrical uncertainty) để tính toán độ phụ thuộc của các đặc trưng và tìm ra tập hợp con

tốt nhất bằng cách sử dụng kỹ thuật *lựa chọn ngược* (*backward selection*) với *chiến lược tìm kiếm tuần tự* (*sequential search strategy*).

Phương pháp này có tiêu chí dừng (stopping criterion) bên trong để ngăn nó loại bỏ các đặc trưng, tức là làm cho chính nó dừng lại khi không còn đặc trưng nào để loại bỏ. FCBF là phương pháp lựa chọn tập con đặc trưng dựa trên tương quan, và nói chung nó chạy nhanh hơn đáng kể so với các phương pháp lựa chọn tập con khác.

*Độ bất định đối xứng* (*Symmetrical Uncertainty - SU*) là một phép đo lý thuyết thông tin chuẩn hóa sử dụng các giá trị entropy có điều kiện để tính toán các phụ thuộc của các đối tượng. Nếu  $X$  là một biến ngẫu nhiên và  $P(x)$  là xác suất của  $x$  thì entropy  $X$  của là:

$$H(X) = - \sum_i P_{(x_i)} \log_2(P_{(x_i)}) \quad (2.5)$$

Entropy có điều kiện (conditional entropy) hoặc độ bất định (onditional uncertainty) có điều kiện của  $X$  với một biến ngẫu nhiên khác  $Y$ , là entropy có điều kiện trung bình của  $X$  so với  $Y$ :

$$H(X|Y) = - \sum_j P(y_j) \sum_i P(x_i|y_j) \log_2(P(x_i y_j)) \quad (2.6)$$

$$SU(X|Y) = 2 \left[ \frac{H(X) - H(X|Y)}{H(X) + H(Y)} \right] \quad (2.7)$$

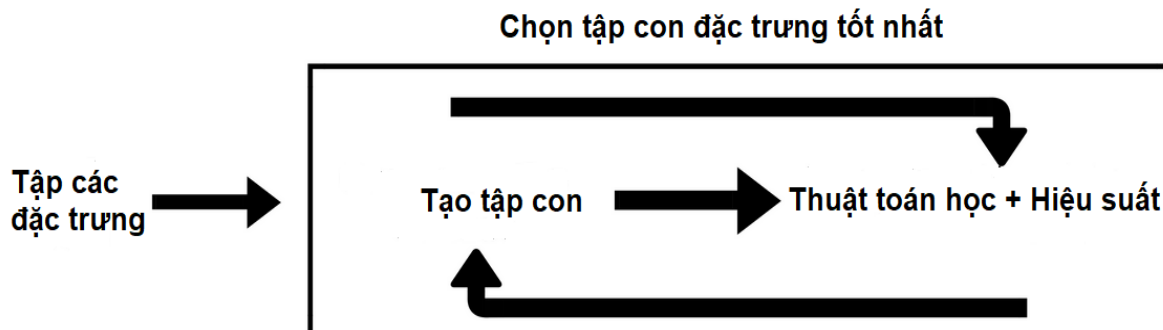
Giá trị  $SU$  bằng 1 cho biết rằng khi sử dụng một đặc trưng thì giá trị của đặc trưng khác có thể được dự đoán hoàn toàn và giá trị 0 cho biết hai đặc trưng hoàn toàn độc lập. Các giá trị  $SU$  là đối xứng cho cả hai đặc trưng. Để tính toán các giá trị  $SU$ , các đặc trưng được định danh (nominal) nhưng các đặc trưng liên tục cũng có thể được sử dụng nếu các giá trị của chúng được phân loại đúng cách.

### 2.1.3. Trích chọn đặc trưng sử dụng các phương pháp nhúng (Embedded methods)

Phương pháp nhúng (Embedded) bao gồm tiện ích của cả phương pháp gói (Wrapper) và phương pháp lọc (Filter). Phương pháp nhúng khác với phương pháp gói và trình bao bọc ở chỗ nó tương tác với thuật toán học (learning algorithm) để lựa chọn đặc trưng nhưng thời gian tính toán nhỏ hơn hai phương pháp kia.

Phương pháp nhúng thực hiện việc lựa chọn biến trong học máy (learning machine). Các biến được chọn trong quá trình đào tạo (training phase), do đó giảm chi phí tính toán và nâng cao hiệu quả trong giai đoạn lựa chọn biến. Thêm một sự khác biệt giữa phương pháp nhúng với phương pháp gói và phương pháp lọc là phương pháp nhúng yêu cầu sự cập nhật lặp đi lặp lại và sự phát triển của các mô hình tham số dựa trên hiệu suất của mô hình được xem xét. Và phương pháp gói chỉ xem xét hiệu suất mô hình của tập hợp các biến đã chọn [29].

Mô hình hoạt động của phương pháp nhúng được thể hiện trong Hình 2.3:



Hình 2. 3. Sơ đồ khối phương pháp Embedded

Phương pháp nhúng này có thể được phân loại thành ba phương pháp con, đó là phương pháp cắt tỉa (pruning), cơ chế tích hợp (built-in mechanism) và các mô hình chính quy hóa (regularization) [26].

Trong phương pháp dựa cắt tỉa, ban đầu tất cả các đặc trưng được đưa vào quy trình đào tạo (training) để xây dựng mô hình phân loại và các đặc trưng có giá trị hệ số tương quan ít hơn sẽ bị loại bỏ để quy bằng cách sử dụng máy vector hỗ trợ (SVM - support vector machine).

Trong phương pháp cơ chế tích hợp, một phần của quá trình huấn luyện (training phase) của thuật toán học tập (learning) có giám sát C4.5 và ID3 được sử dụng để chọn các đặc trưng.

Trong phương pháp chính quy hóa, các lỗi điều chỉnh (fitting) được giảm thiểu bằng cách sử dụng các hàm mục tiêu và loại bỏ các đặc trưng có hệ số hồi quy gần bằng 0.

#### 2.1.4. So sánh phương pháp Filter, Wrapper và Embedded

Phương pháp lựa chọn đặc trưng được sử dụng để đạt được một tập con các đặc trưng tối ưu. Bảng 2.1 sẽ đưa ra sự so sánh cũng như sự tổng kết về các phương pháp này [30].

Bảng 2. 1. Bảng so sánh phương pháp Filter, Wrapper và Embedded

Phương pháp	Ưu điểm	Nhược điểm
<b>Wrapper</b>	<ul style="list-style-type: none"> <li>- Đánh giá độ tương đồng giữa các đặc trưng và các nhãn lớp (class label).</li> <li>- Phân tích mức phụ thuộc giữa các đặc trưng.</li> <li>- Mức độ xác thực cao hơn Filter.</li> </ul>	<ul style="list-style-type: none"> <li>- Thời gian tính toán cao cho một tập dữ liệu có nhiều đặc trưng.</li> <li>- Tập con đặc trưng được chọn bị phân tích lặp lại nhiều lần.</li> <li>- Một vài các đặc trưng bị loại bỏ ở các vòng lặp đầu thì lại không được đánh giá.</li> <li>- Gây ra overfitting. Đây là hiện tượng quá khớp, xảy ra khi một mô hình quá khớp đối với <i>dữ liệu đào tạo – training data</i> nhưng lại có hiệu suất kém với <i>dữ liệu thử nghiệm – test data</i>.</li> </ul>
<b>Filter</b>	<ul style="list-style-type: none"> <li>- Tính toán nhanh hơn Wrapper và Embedded,</li> <li>- Phù hợp với <i>low dimensional data (dữ liệu chiều thấp)</i>.</li> </ul>	<ul style="list-style-type: none"> <li>- Bỏ qua độ tương đồng hay sự tương tác giữa các bộ phân loại (classifier).</li> <li>- Bỏ qua sự phụ thuộc đặc trưng do các đặc trưng được xem xét một cách độc lập.</li> <li>- Trong một số trường hợp, không rõ cách xác định điểm ngưỡng cho thứ hạng để chọn được các đặc trưng đúng yêu cầu cũng như loại trừ được nhiều.</li> </ul>

<p><b>Embedded</b></p>	<ul style="list-style-type: none"> <li>- Xem xét độ tương quan của các đặc trưng giống phương pháp gói.</li> <li>- Tính toán nhanh hơn phương pháp Wrapper, điểm này giống phương pháp Filter.</li> <li>- Có tính chính xác hơn các phương pháp Filter.</li> <li>- Tìm được tập hợp con đặc trưng cho các thuật toán đào tạo (training algorithm).</li> <li>- Ít bị over-fitting.</li> </ul>	<ul style="list-style-type: none"> <li>- Không phù hợp với <i>high dimensional data</i> (dữ liệu chiều cao).</li> <li>- Độ bao quát thấp.</li> <li>- Chi phí cao hơn Filter.</li> </ul>
------------------------	--	---

## 2.2. PHƯƠNG PHÁP PHẢN HỒI LIÊN QUAN

### 2.2.1. Giới thiệu về phản hồi liên quan

Phương pháp tra cứu ảnh dựa trên nội dung (CBIR) ra đời đã khắc phục được những điểm chưa hoàn thiện của các phương pháp tra cứu ảnh trước đó, mở ra một hướng đi mới cho lĩnh vực tra cứu ảnh nói riêng và ngành khoa học xử lý ảnh nói chung. Trong hệ thống CBIR, các thuật toán xử lý hình ảnh (thường là tự động) được sử dụng để trích xuất các vectơ đặc trưng đại diện cho các đặc trưng của hình ảnh như màu sắc, kết cấu và hình dạng.

Không thể phủ nhận những tiện ích mà CBIR mang lại. Trong một hệ thống CBIR, chúng ta có thể rất lập chỉ mục và tra cứu hình ảnh với số lượng lớn. Kết quả tìm kiếm hiệu quả và ít tốn kém. Tuy nhiên, song song với những ưu điểm, CBIR còn có những yếu điểm vẫn còn tồn đọng. Một trong số đó chính là vấn đề các đặc trưng cấp thấp không có khả năng mô tả và giải thích ngữ nghĩa của bối cảnh hình ảnh, vậy nên những hình ảnh có các đặc trưng cấp thấp này có thể khác với hình ảnh truy vấn ở các vị trí của ngữ nghĩa mà người dùng cảm nhận được. Hoặc là giữa các đặc trưng cấp thấp và các đặc trưng cấp cao sẽ có khoảng cách ngữ nghĩa quá khác biệt. Tất cả dẫn đến việc sẽ khó hiểu nhu cầu thông tin của người dùng từ hình ảnh truy vấn, do đó không thể phục vụ được cho người dùng phổ thông.

Vậy nên để khắc phục những nhược điểm đó, phương án "Phản hồi liên quan cả người dùng" đã được sử dụng kết hợp, nhằm hỗ trợ và phát triển quá



trình tra cứu với mục đích tạo ra hiệu ứng tra cứu đáng kể hơn về mặt tri giác và ngữ nghĩa.

Phản hồi liên quan (Relevance feedback - RF) là một kỹ thuật dùng để đánh dấu các kết quả đã tra cứu, xem xét chúng có liên quan hoặc không liên quan đến người dùng. Hay nói cách khác, cả người dùng và hệ thống tra cứu lúc này, nhờ có phản hồi liên quan mà đã được đưa vào một quá trình "cộng sinh". Cơ chế vận hành của hệ thống RF như sau. Sau khi xây dựng công thức của truy vấn ban đầu, đối với các lần lặp lại tiếp theo của việc sàng lọc truy vấn, hệ thống sẽ trình bày một tập hợp các kết quả và người dùng sẽ đánh giá kết quả đó để điều chỉnh tập hợp các hình ảnh được tra cứu theo mong muốn của mình.

Một kịch bản thường gặp của hệ thống phản hồi liên quan trong CBIR sẽ bao gồm bốn bước như sau [31]:

➤ Bước 1: Người dùng cung cấp một ảnh mẫu truy vấn hoặc từ khóa mô tả đối với hệ thống.

➤ Bước 2: Hệ thống thực hiện tìm kiếm ban đầu đối với ảnh mẫu truy vấn do người dùng cung cấp, dựa trên các công thức đã xây dựng trước đó (ví dụ như áp dụng các độ đo tương tự), sau đó trả về các kết quả đã tra cứu được, thường là một số lượng nhỏ hình ảnh.

➤ Bước 3: Người dùng đánh dấu hoặc cho biết các hình ảnh đã được truy vấn, hình ảnh có liên quan hay là không liên quan đến yêu cầu của mình và liên quan ở mức độ nào.

➤ Bước 4: Dựa trên các đánh giá và phản hồi của người dùng, hệ thống tự động định dạng lại truy vấn ban đầu để tra cứu một danh sách các ảnh mới cho người dùng. Thuật toán lặp lại Bước 3 và sẽ tiếp tục lặp lại cho đến khi người dùng hài lòng.

Phương pháp phản hồi liên quan giúp giảm bớt vấn đề về khoảng cách ngữ nghĩa, vì nó cho phép hệ thống CBIR tìm hiểu nhận biết của người dùng về hình ảnh mong muốn. Các kỹ thuật RF thường xử lý các mẫu đào tạo (training sample) nhỏ mà thông thường là ít hơn khoảng hai mươi mẫu ở mỗi vòng tương

tác, tính không đối xứng trong mẫu đào tạo (ví dụ như các phản hồi tiêu cực được cung cấp cho hệ thống) và yêu cầu thời gian thực. Vậy nên các thuật toán RF phải đủ nhanh để hỗ trợ tương tác người dùng trong thời gian thực.

Có bốn cơ chế để người dùng phản hồi về mức độ liên quan, bao gồm: Phản hồi rõ ràng (Explicit feedback); Phản hồi ngầm (Implicit feedback); Phản hồi xếp hạng (Ranking feedback), Phản hồi mù (Blind feedback). Cả bốn cơ chế này đều đã được trình bày trong mục 1.2.4 của luận văn.

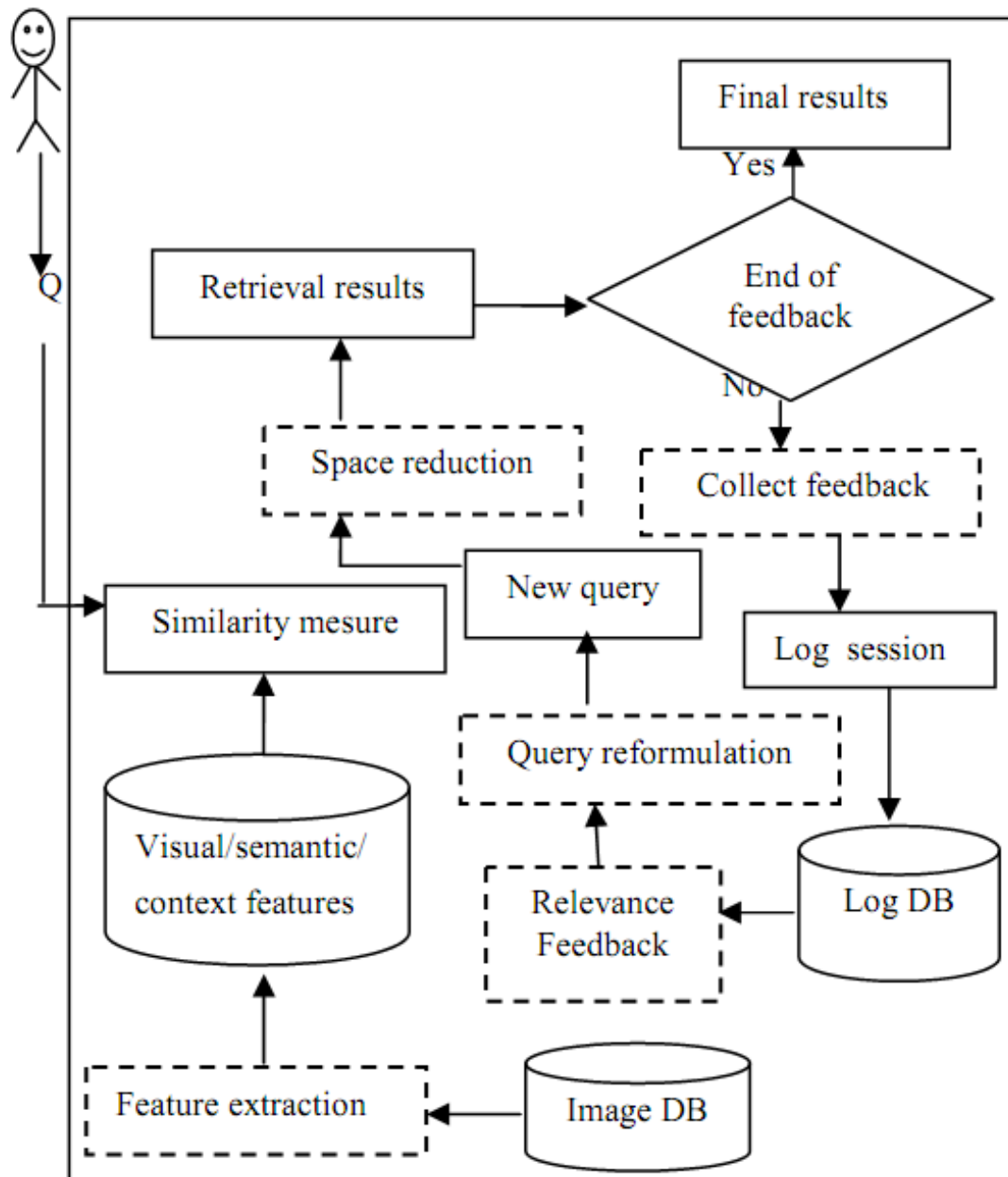
### **2.2.2. CBIR với phản hồi liên quan**

Để cải thiện hiệu suất tra cứu trong hệ thống CBIR, cơ chế phản hồi liên quan đã được đề xuất như một bước tiến lớn trong công nghệ tương tác người dùng để truy cứu hình ảnh.

Ngày nay, nó là một lĩnh vực mở rộng và phát triển hơn, bao gồm nhiều chủ đề phụ khác nhau, hỗ trợ giải quyết một số mối quan tâm thực tế trong khi vẫn ghi nhớ người dùng trong vòng lặp.

#### **2.2.2.1. Kịch bản thuật toán RF**

Các bước quan trọng nhất của một hệ thống CBIR sử dụng phản hồi liên quan được mô tả trong Hình 2.4 [32].



Hình 2. 4. Kiến trúc của hệ thống CBIR sử dụng RF

Như mô tả trong Hình 2.4, người dùng thiết lập một truy vấn ban đầu (Q) cho hệ thống tra cứu, hệ thống này sẽ tạo ra một tập hợp các câu trả lời. Sau đó, người dùng sẽ kiểm tra các câu trả lời và đưa ra đánh giá về chất lượng hoặc mức độ liên quan của các câu trả lời. Hệ thống sử dụng phản hồi của người dùng để tạo một truy vấn mới nhằm hạn chế không gian tra cứu.

#### 2.2.2.2. Các mô hình phản hồi liên quan trong CBIR

Các phương pháp tiếp cận phản hồi liên quan cho CBIR đã được phân loại thành hai mô hình: học ngắn hạn – short term learning (hay còn gọi là học truy

vấn nội bộ - intraquery learning) và học dài hạn (hay còn gọi là học truy vấn liên quan - interquery learning). Hình 2.5 minh họa sự phân loại của các thuật toán phản hồi liên quan được phát triển theo con đường từ các kỹ thuật dựa trên cơ sở kinh nghiệm đến các thuật toán học tối ưu.



Hình 2. 5. Phân loại các thuật toán phản hồi liên quan.

Các thuật toán học máy được sử dụng rộng rãi trong mô hình học ngắn hạn hoặc học dài hạn [32].

**\* Mô hình học ngắn hạn**

Là mô hình đầu tiên. Nó tập trung vào việc điều chỉnh truy vấn (sàng lọc) trong một phiên tra cứu duy nhất. Bao gồm:

1. Trọng số Heuristic: Cơ chế tính toán trọng số của đặc trưng, giả định xây dựng công thức heuristic với điều chỉnh tham số truy vấn theo kinh nghiệm, chủ yếu dọc theo đường trọng số trục độc lập trong không gian đặc trưng. Ý tưởng chính đằng sau *phương pháp tính toán trọng số Heuristic của đặc trưng* là sửa đổi điểm truy vấn trong mỗi lần lặp tìm kiếm thành biểu diễn phù hợp hơn để đưa yêu cầu của người dùng đến truy vấn tối ưu gần nhất. Hơn nữa, nó cung cấp một bảng xếp hạng các hình ảnh được tra cứu theo mức độ tương đồng của chúng với truy vấn được tìm kiếm. *Phương pháp tiếp cận theo trọng số đặc trưng Heuristic* có thể đạt được bằng cách sử dụng các thuật toán sau: Di chuyển điểm truy vấn (Query point movement), Định hình lại hàm khoảng cách, Lựa chọn hàm khoảng cách, Mở rộng truy vấn (Query Expansion), Phân rã truy vấn (Query Decomposition), Đánh giá lại truy vấn (Query Reweighting), Thêm hoặc xóa truy vấn đặc trưng đơn lẻ, Chuẩn hóa khoảng cách.

2. Ước tính mật độ: Tính toán sự phân bố của các hình ảnh có nhãn tích cực và (có thể) tiêu cực bằng cách mô hình hóa hàm mật độ xác suất được cập nhật sau mỗi vòng lặp và các hàm phân loại để xác định lớp (class) “tốt nhất” để gán một hình ảnh mới.

3. Tìm kiếm phân loại: Phương pháp tìm kiếm phân loại dựa trên hình ảnh mẫu được cung cấp bởi người dùng được sử dụng để đào tạo (train) một bộ phân loại (classifier), sau đó được sử dụng để phân loại cơ sở dữ liệu một cách tối ưu thành những hình ảnh có liên quan đến truy vấn và những hình ảnh không liên quan. Các phương pháp tiếp cận khác nhau đã được áp dụng trong các kỹ thuật học có giám sát khác nhau bao gồm: Cây quyết định, Mạng thần kinh, Phương pháp Bayes, Phân tích phân biệt (discriminate analysis), <ô hình tổng hợp / EM, Tăng cường (boosting), Máy vector hỗ trợ (SVM), Kết hợp người học, So khớp đồ thị (Graph matching), Đặc trưng ảo (Visual Feature). Ngoài ra, vấn đề phân loại dựa trên phản hồi liên quan có thể được chuyển đổi thành một bài toán tối ưu hóa. Tối ưu hóa coi phản hồi liên quan giống như một vấn đề tối thiểu hóa (minimization) và tối đa hóa (maximization) có giải pháp là các tham số của mô hình giúp ta có thể tìm ra truy vấn lý tưởng, tính toán các đặc trưng và chuyển đổi không gian đặc trưng thành một không gian mới phù hợp hơn với người dùng.

4. Kernels (Hạt nhân): Phương pháp này có thể được áp dụng trong các kỹ thuật học máy khác nhau để thu được kết quả tốt hơn. Nó được sử dụng để chuyển đổi không gian đặc trưng sang không gian chiều cao hơn hoặc thấp hơn, mà lý tưởng nhất chính là các hình ảnh được quan tâm có thể được phân tách tuyến tính khỏi tất cả các hình ảnh khác. Các kernel (hạt nhân) thường gặp nhất là nhân đa thức, nhân tuyến tính, nhân sigmoid hàm Gaussian, hoặc các nhân kết hợp.

5. Tìm kiếm so sánh: Tìm kiếm so sánh dựa trên việc tính đến các phán đoán tương đối so với các phán đoán liên quan để tìm ra ảnh mục tiêu hoặc phần tử gần nhất với ảnh đích (target image) trong một tập hợp các phần tử trong cơ sở dữ liệu.

6. Trục quan tương tác: Cung cấp khả năng xem toàn cục các hình ảnh được tra cứu trên sơ đồ hai chiều theo cách phản ánh mối quan hệ giữa các hình ảnh trong tập hợp trả về, và sau đó người dùng có thể chọn vùng màn hình để tinh chỉnh truy vấn của mình.

7. Cắt đồ thị: Bao gồm việc tách cơ sở dữ liệu thành một nhóm có liên quan và một nhóm không liên quan, vì vậy các hình ảnh phản hồi được sử dụng làm hạt giống (seed) cho quá trình cắt đồ thị, với các hình ảnh có liên quan tạo thành nhóm nguồn (source group) và các hình ảnh không liên quan là nhóm chìm (sink group). và một biểu đồ có trọng số mô hình hóa cấu trúc liên kết của cơ sở dữ liệu được xây dựng.

8. Học đa tạp (Manifold learning): Hiệu suất tìm kiếm (Search performance) giảm mạnh khi kích thước của không gian đặc trưng tăng lên. Rất nhiều kỹ thuật đã được đề xuất trong phương pháp tiếp cận học tập đa tạp để giảm kích thước không gian: các kỹ thuật tuyến tính bao gồm Phân tích thành phần chính (PCA - Principal Component Analysis), Phân tích thành phần độc lập (ICA - Independent Component Analysis), Phép chiếu ngẫu nhiên (Random projection), Phép chiếu bảo toàn cục bộ (LPP - Locality-preserving projection), Phân biệt tuyến tính Phân tích (LDA - Linear Discriminated Analysis), và Đa tạp biên tối đa (MMD - Maximum Marginal Diversity)... Liên quan đến các kỹ thuật phi tuyến, các phương pháp nổi tiếng là: Chia tỷ lệ đa chiều (MDS - Multidimensional Scaling), Phân tích thành phần đường cong (CCA - Curvilinear Component Analysis) và FastMap.

9. Học tập tích cực: Kỹ thuật này có thể nâng cao đáng kể tốc độ học tập của khái niệm mà người dùng có trong đầu. Bằng cách chủ động chọn một tập hợp hình ảnh thích hợp, thường được gọi là những hình ảnh nhiều thông tin nhất, người dùng sẽ chỉ cần dán nhãn (label) cho một số lượng hình ảnh tương đối nhỏ để người học hiểu khái niệm truy vấn.

**\* Mô hình học dài hạn**

Mô hình này phân tích mối quan hệ giữa các phiên tra cứu hiện tại và trước đây bằng cách sử dụng lịch sử tương tác hoặc nghiên cứu để lập mô hình mối quan tâm của người dùng, bao gồm:

1. RF lai (Hybrid RF-HRF): Phương pháp này tìm kiếm một sơ đồ kết hợp của hai phương pháp khác nhau miễn là hiệu suất tổng thể được cải thiện.

2. Phương pháp tiếp cận RF bổ sung (Complementary RF-CRF): Phương pháp bao gồm việc kết hợp hai cách tiếp cận khác nhau bổ sung cho nhau với hy vọng loại bỏ điểm yếu của cách tiếp cận riêng lẻ.

3. Học gia tăng (Incremental learning-IL): Phương pháp này sử dụng phương pháp học tăng dần. Điều đó có nghĩa là, các thuật toán gia tăng có thể học dần dần, ví dụ phản hồi thứ  $n$  từ thứ  $(n - 1)$  mà không cần thực hiện lại toàn bộ thuật toán có hiệu quả trong học tập và hiệu quả trong tính toán.

4. Khai thác mô hình điều hướng (Navigation Pattern Mining-NPM): Nghiên cứu gần đây đã bắt đầu tập trung vào Khai thác mô hình điều hướng. Nó tập trung vào việc tìm kiếm các mẫu thú vị từ cơ sở dữ liệu, chẳng hạn như quy tắc kết hợp, tương quan, chuỗi, tập, bộ phân loại, cụm bằng cách sử dụng các mẫu thường xuyên, biểu đồ, điều hướng hoặc tuần tự để hội tụ không gian tìm kiếm theo ý định của người dùng.

Kết lại, mô hình học dài hạn có thể đạt được độ tra cứu chính xác tốt hơn so với học ngắn hạn. Chúng ta có thể sử dụng mô hình học dài hạn để khắc phục tình trạng hệ thống không nắm bắt được ngữ nghĩa, các ví dụ phản hồi không cân bằng và thiếu cơ chế ghi nhớ.

## 2.3. TRA CỨU ẢNH DỰA TRÊN PHÂN CỤM ĐỒ THỊ

### 2.3.1. Phân cụm đồ thị và các thuật toán phân cụm phổ

#### 2.3.1.1. Giới thiệu

Phân cụm (Clustering) là một phương pháp học tập không giám sát còn được gọi là phân khúc hoặc phân loại. Đây là một kỹ thuật dùng để phân tích hoặc thống kê dữ liệu và được ứng dụng vào nhiều chuyên ngành nghiên cứu như truy hồi thông tin, khai phá dữ liệu, xử lý ảnh, học máy, học sâu, v.v. Lý thuyết hoạt động của phân cụm chính là phân/nhóm các quần thể hoặc các

đối tượng thành các cụm, sao cho các điểm trong một cụm tương tự/tương đồng nhau và ít giống với các điểm trong các cụm khác. Một thuật toán phân cụm tốt đảm bảo độ tương tự trong cụm cao và độ tương tự giữa các cụm thấp.

Trong khoa học dữ liệu, chúng ta có thể sử dụng phân tích phân cụm để có được một số thông tin chi tiết có giá trị từ dữ liệu của mình bằng cách xem các điểm dữ liệu rơi vào nhóm nào khi áp dụng thuật toán phân nhóm.

Đặc điểm của phân cụm:

- Cụm dữ liệu là tập hợp các đối tượng hay cá thể giống nhau có chung đặc điểm.

- Số cụm dữ liệu được phát hiện và xác định khi thực hiện hoạt động phân cụm vậy nên số lượng của chúng là không thể biết trước.

- Có rất nhiều kỹ thuật và thuật toán phân cụm khác nhau, và mỗi một cái lại mang lại những hiệu quả khác nhau.

### ***2.3.1.2. Thuật toán phân cụm phổ***

Phân cụm phổ (Spectral Clustering) là một dạng thuật toán phân cụm. Thuật toán này hoạt động tốt hơn một số thuật toán phân cụm truyền thống trong nhiều trường hợp. Nó coi mỗi điểm dữ liệu như một nút đồ thị và do đó biến bài toán phân cụm thành bài toán phân vùng đồ thị.

Thuật toán phân cụm phổ được chia thành ba bước cơ bản như sau:

- Bước 1: Pre-processing (Tiền xử lý). Xây dựng biểu diễn ma trận của đồ thị.

Bước 2: Decomposition (Khai triển). Tính toán các giá trị riêng và vector đặc trưng của ma trận. Ánh xạ từng điểm thành biểu diễn chiều thấp (lower-dimensional) hơn dựa trên một hoặc nhiều vector đặc trưng.

- Bước 3: Grouping (Phân nhóm). Gán điểm cho hai hoặc nhiều cụm, dựa trên cách biểu diễn mới.

Kỹ thuật chính để phân cụm phổ là ma trận đồ thị Laplacian. Ma trận Laplacian là một tương tự rời rạc (discrete analog) của toán tử Laplacian trong



phép tính đa biến. Ma trận Laplacian cho phép liên kết tự nhiên giữa các biểu diễn rời rạc, ví như đồ thị, phép biểu diễn liên tục (continuous representation), hay là không gian vector và đa tạp (manifold). Ứng dụng quan trọng nhất của ma trận Laplacian là phân cụm phổ, tương đương với một giải pháp có thể tính toán được cho bài toán phân vùng đồ thị.

Giả sử ta gọi ma trận Laplacian là ma trận của đồ thị  $G$ . Trong đó  $G = (V, E)$  là đồ thị vô hướng, không có trọng số cũng không có vòng lặp đồ thị  $(i, i)$  hoặc nhiều cạnh từ nút này sang nút khác.  $V$  là tập đỉnh,  $n = |V|$ .  $E$  là tập cạnh, là một ma trận đối xứng  $n \times n$  với một hàng và cột cho mỗi nút.  $G$  được xác định bởi:

$$L = D - A \quad (2.8)$$

Trong đó  $D = \text{diag}(d_1, \dots, d_n)$  là ma trận bậc, tức ma trận đường chéo được hình thành từ các độ đỉnh và  $A$  là ma trận kề. Do đó, các phần tử đường chéo  $l_{ij}$  của  $L$  với tung độ của đỉnh  $v_i$  và các phần tử nằm ngoài đường chéo  $l_{ij}$  là  $-1$  nếu đỉnh  $v_i$  tiếp giáp với  $v_j$  và  $0$  nếu ngược lại.

Ma trận Laplacian không dấu (signless Laplacian matrix) được định nghĩa như sau, trong đó  $D$  là ma trận độ (degree matrix),  $A$  là ma trận kề (adjacency matrix):

$$L_S = D + A \quad (2.9)$$

Ma trận kề chuẩn hóa (normalized adjacency) Laplacian hay biểu đồ chuẩn hóa được định nghĩa như sau:

$$L^\wedge = D^{-\frac{1}{2}}AD^{-1/2} \quad (2.10)$$

Trong đó  $L^\wedge$  là ma trận kề của của  $L$  và  $D = \text{diag}(d)$  với  $d(i)$  là bậc nút của  $i$ .

$L$  và  $L^\wedge$  là các đối tượng chính trong lý thuyết đồ thị phổ. Các bất biến đồ thị khác nhau có thể được ước tính theo giá trị riêng của các ma trận này. Tương tự, ta sẽ thấy rằng phân cụm phổ cho phép người ta thay thế các bài toán tối ưu

hóa khác nhau bằng đại số tuyến tính tiêu chuẩn. Cho vector  $f = (f_1, \dots, f_n) \in R^n$ , nhận dạng khóa sau có thể dễ dàng xác minh:

$$fLf^t = 2 \sum_{[i] \sim [j]} w_{ij} (f_i - f_j)^2 \quad (2.11)$$

Lưu ý rằng phương trình này cho thấy rằng  $L$  là nửa xác định dương, và vì  $D^{\frac{1}{2}}$  là xác định dương nên điều này cũng đúng với  $L^\wedge$ . Bây giờ ta sẽ thấy rằng các thuộc tính phân cụm của biểu đồ Laplacian theo trực tiếp từ thuộc tính này. Cho một tập con  $S \subset V$ , xác định véc tơ cột  $f_S = (f_{S_1}, \dots, f_{S_n})' \in R^n$  như sau:

$$f_{S_i} = \begin{cases} 1, [i] \in S \\ -1, [i] \in \bar{S} \end{cases} \quad (2.12)$$

Sau đó tiếp tục:

$$f_S^t L f_S = \sum_{[i] \sim [j]} w_{ij} (f_{S_i} - f_{S_j})^2 = 4 \text{vol}(\delta S) \quad (2.13)$$

và đối với tất cả  $S \subset V$ , ta có:

$$f_S^t D f_S = \sum_{[i] \in V} w_{ij} = \text{vol}(V) \quad (2.14)$$

Bây giờ hãy xem xét trường hợp cắt cân bằng có trọng số. Ta biểu thị bằng  $1$  véc tơ cột của tất cả các véc tơ đó và có:

$$f_S^t D 1 = \sum_{[i] \in S} \text{deg}[i] - \sum_{[i] \in \bar{S}} \text{deg}[i] = \text{vol}(S) - \text{vol}(\bar{S}) \quad (2.15)$$

Do đó,  $f_S^t D 1 = 0$  nếu và chỉ khi phần cắt tương ứng với  $S$  là cân bằng thể tích, đó là  $\text{vol}(S) = \text{vol}(\bar{S})$ . Do đó, chúng ta có thể định dạng lại vấn đề tính toán phần cắt cân bằng có trọng số như:

$$b_w(G) = \min_{f \in \{-1,1\}^n, f^t D 1 = 0} f^t L f \quad (2.16)$$

Hơn nữa, vì  $f^t D f$  có giá trị không đổi  $1/4 \text{ vol}(V)$  với mọi  $f \in \{-1,1\}^n$ , chúng ta có thể viết lại điều này dưới dạng:

$$b_w(G) = \frac{1}{4} \min_{f \in \{-1,1\}^n, f^t D 1 = 0} \frac{f^t L f}{f^t D f} \quad (2.17)$$

Được phát biểu ở dạng này, bài toán tối ưu hóa rời rạc thừa nhận một sự đơn giản bằng cách để  $f$  nhận các giá trị thực thay vì  $\{-1,1\}$ . Nhận thấy rằng  $L1 = 0$ , một đối số đại số tuyến tính tiêu chuẩn cho thấy rằng:

$$\lambda_2 = \min_{f \in \mathbb{R}^n, f^t D 1 = 0} \frac{f^t L f}{f^t D f} \quad (2.18)$$

Trong đó  $\lambda_2$  là giá trị riêng nhỏ nhất thứ hai của bài toán hiệu chỉnh tổng quát  $Lf = \lambda Df$ . Rõ ràng là giá trị riêng nhỏ nhất  $\lambda_1$  của  $L$  là 0 và giá trị riêng tương ứng là 1. Hơn nữa, dễ dàng chỉ ra rằng đối với một đồ thị liên thông, giá trị riêng thứ hai thỏa mãn  $\lambda_2 > 0$ . Do đó, vector  $f$  mà giá trị nhỏ nhất trong Phương trình ở trên đạt được là vector riêng tương ứng với  $\lambda_2$ . Những suy luận trên dẫn trực tiếp đến thuật toán phân vùng sau đây:

1. Tính các ma trận  $L$  và  $D$ .

2. Tìm giá trị riêng  $e$  tương ứng với giá trị riêng nhỏ nhất thứ hai của bài toán giá trị riêng tổng quát sau:

$$L f = \lambda D f$$

3. Lấy phân hoạch:  $S = \{[i]: e_i > 0\}, \bar{S} = \{[i]: e_i \leq 0\}$ .

Đây là thuật toán cơ bản để chuẩn hóa phân cụm phổ.

Do vậy, chúng ta có thể đồng nhất ba ký hiệu: đối tượng  $i$ , vector  $x_i$  và đỉnh  $v_i$  là tương đương với nhau theo nghĩa biểu diễn cho cùng một thực thể dữ liệu.

### 2.3.2. Mã giả thuật toán Spectral Clustering

**Input:** Đồ thị G

**Do:**

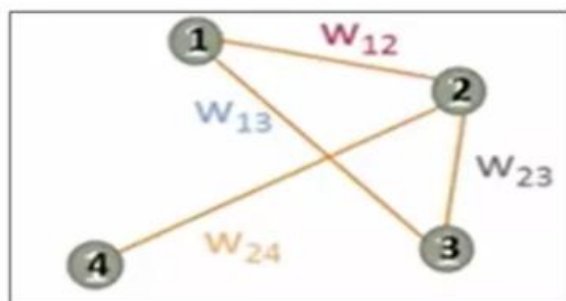
1. Tính ma trận kề W.
2. Tính ma trận Laplacian,  $L=D-W$  với D là ma trận bậc.
3. Tính k vector đầu tiên của ma trận L.
4. Sử dụng K-means để phân cụm dựa trên các vector riêng trên.

**Output:** Kết quả phân cụm của K-means.

#### 2.3.2.1. Tính ma trận kề

Cho đồ thị  $G = (V, E)$  với V là tập gồm các đỉnh và E là tập gồm các cạnh. Mỗi cạnh thuộc E sẽ gồm 2 đỉnh trong tập V - một cặp đỉnh  $(v_i, v_j)$ , với trọng số cạnh là  $w_{ij}$ .

$$A = \begin{cases} w_{ij} : \text{trọng số của cặp đỉnh } (v_i, v_j) \\ 0 : \text{nếu không có cạnh giữa hai đỉnh } (v_i, v_j) \end{cases}$$



	1	2	3	4
1	0	$w_{12}$	$w_{13}$	0
2	$w_{12}$	0	$w_{23}$	$w_{24}$
3	$w_{13}$	$w_{23}$	0	0
4	0	$w_{24}$	0	0

#### 2.3.2.2. Tính ma trận Laplacian

Ma trận laplacian:  $L = D - A$  với D là ma trận bậc. Ma trận bậc D được tính từ ma trận kề A, có số chiều giống với ma trận A. Mỗi phần tử trên đường chéo chính của ma trận bậc D là tổng của các phần tử trên một hàng của ma trận A tương ứng. Các phần tử khác ngoài đường chéo chính đều bằng 0.

$$d_i = \sum_{\{j|(i,j) \in E\}} w_{ij} \quad (2.19)$$

$$\begin{aligned}
 d_1 &= w_{12} + w_{13} \\
 d_2 &= w_{12} + w_{23} + w_{24} \\
 d_3 &= w_{13} + w_{23} \\
 d_4 &= w_{24}
 \end{aligned}$$

$$D =$$

	1	2	3	4
1	$d_1$	0	0	0
2	0	$d_2$	0	0
3	0	0	$d_2$	0
4	0	0	0	$d_4$

Ma trận laplacian:

$$L = D - A = \begin{cases} d_i & \text{nếu } i = j \\ -w_{ij} & \text{nếu } (i, j) \text{ là một cạnh} \\ 0 & \text{nếu không có cạnh nối giữa } (i, j) \end{cases} \quad (2.20)$$

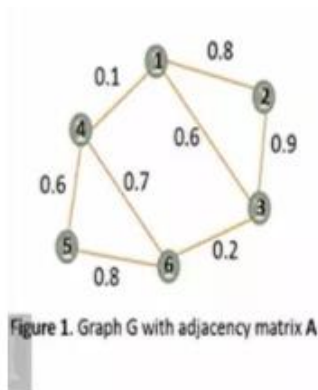
$$L =$$

	1	2	3	4
1	$d_1$	$-w_{12}$	$-w_{13}$	0
2	$-w_{12}$	$d_2$	$-w_{23}$	$-w_{24}$
3	$-w_{13}$	$-w_{23}$	$d_2$	0
4	0	$-w_{24}$	0	$d_4$

### 2.3.2.3. Tính k vector riêng đầu tiên của ma trận Laplacian

Cho một ma trận A, ta có  $\lambda$  là một giá trị riêng và  $v$  là vector riêng của A nếu:  $Av = \lambda v$

Cho một đồ thị G có n nút, ma trận kề của nó sẽ có n giá trị riêng  $\{\mu_1, \mu_2, \mu_3, \dots, \mu_n\}$  với  $\mu_1 \geq \mu_2 \geq \mu_3 \geq \dots \geq \mu_n$  và n vector riêng  $\{x_1, x_1, x_2, x_3, \dots, x_n\}$ :



$$A =$$

	1	2	3	4	5	6
1	0	0.8	0.6	0.1	0	0
2	0.8	0	0.9	0	0	0
3	0.6	0.9	0	0	0	0.2
4	0.1	0	0	0	0.6	0.7
5	0	0	0	0.6	0	0.8
6	0	0	0.2	0.7	0.8	0

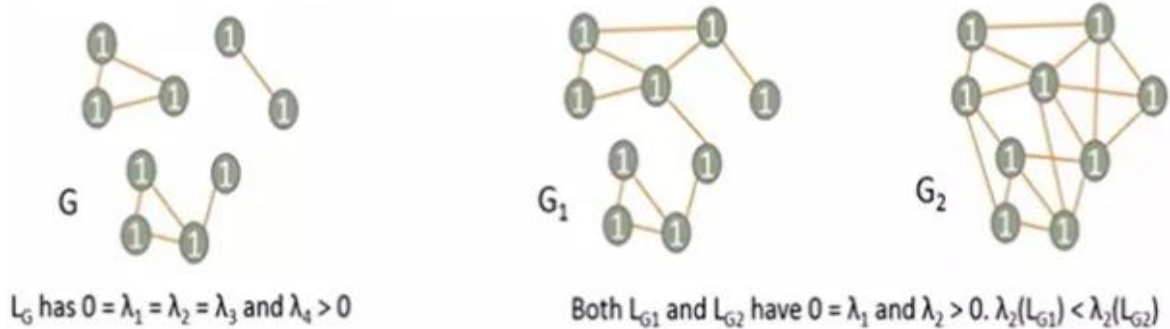
$\mu_1 = -0.993$	$x_1 = [-0.17, 0.65, -0.56, -0.11, -0.25, 0.39]$
$\mu_2 = -0.832$	$x_2 = [-0.46, 0.44, 0.00, 0.30, 0.37, -0.61]$
$\mu_3 = -0.642$	$x_3 = [-0.37, -0.03, 0.36, 0.53, -0.66, 0.13]$
$\mu_4 = -0.482$	$x_4 = [-0.57, 0.06, 0.48, -0.56, 0.19, 0.31]$
$\mu_5 = 1.355$	$x_5 = [0.25, 0.30, 0.24, -0.48, -0.52, -0.52]$
$\mu_6 = 1.584$	$x_6 = [-0.48, -0.53, -0.52, -0.26, -0.25, -0.30]$

Eigenvalues and eigenvectors of A

Ở đây ta sẽ đi tính véc tơ riêng và giá trị riêng cho ma trận Laplacian, sau đó lấy ra k vector đầu tiên. Có thể nói việc tính ma trận Laplacian và tính k giá trị riêng và vector riêng của ma trận này là trái tim của thuật toán Spectral Clustering. Các giá trị riêng cho biết các thuộc tính toàn cục không rõ ràng của đồ thị từ cấu trúc cạnh. Xét đồ thị Laplacian của G,  $L_G$  có tập giá trị riêng  $\{\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n\}$  và tập vector  $\{x_1, x_2, x_3, \dots, x_n\}$ :

- Nếu 0 là giá trị riêng của L ( $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$ ) với k vector riêng thì đồ thị G có k kết nối thành phần

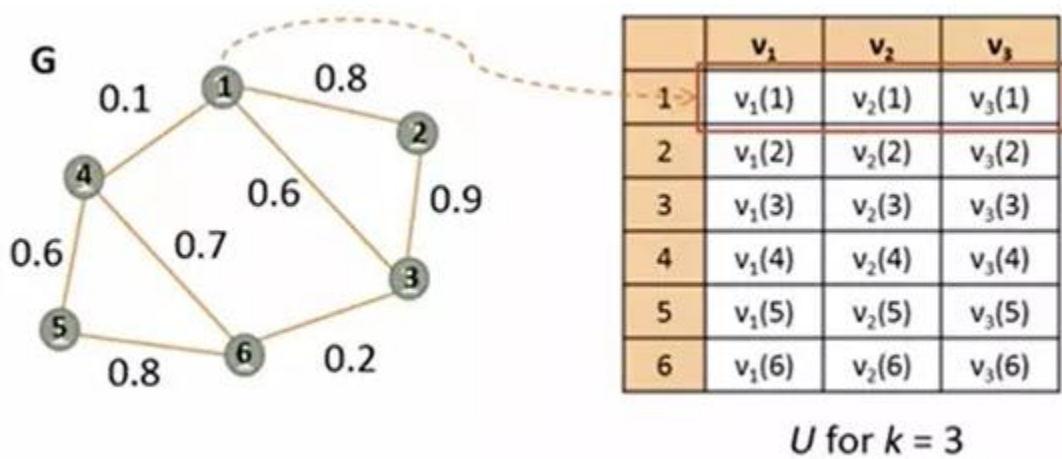
- Nếu đồ thị đã được kết nối,  $\lambda_2 > 0$  và  $\lambda_2$  là kết nối đại số của G,  $\lambda_2$  càng lớn thì G càng có nhiều kết nối.



Dễ thấy rằng  $L_G$  có  $\lambda_1 = \lambda_2 = \lambda_3 = 0$  và  $\lambda_4 > 0$  nên G có 3 đường kết nối.  $L_{G_1}$  và  $L_{G_2}$  có  $\lambda_1 = 0$  nên 2 đồ thị  $G_1$  và  $G_2$  có một đường kết nối. Lại có  $\lambda_2(L_{G_1}) < \lambda_2(L_{G_2})$  nên  $G_2$  có nhiều đường kết nối hơn.

#### 2.3.2.4. Sử dụng K-means để phân cụm

Sắp xếp thành ma trận U với k véc tơ riêng đầu tiên của ma trận L và các đỉnh của đồ thị G (ký hiệu v trong hình là véc tơ riêng, không phải đỉnh). Sau đó áp dụng phân cụm đối với ma trận U sử dụng thuật toán K-mean để phân các đỉnh của G (hàng của U).



Thuật toán K-mean:

**Input:**  $K$  (số lượng các cụm),

Training set  $\{x^{((1))}, x^{((2))}, \dots, x^{((m))}\} x^{((i))} \in R^n$

**Output:** Các center  $\mu_K$  và label cho từng điểm dữ liệu  $x^{((i))}$ .

1. Chọn  $K$  điểm bất kỳ làm các center ban đầu.
2. Phân mỗi điểm dữ liệu vào cụm có center gần nó nhất.
3. Nếu việc gán dữ liệu vào từng cụm ở bước 2 không thay đổi so với vòng lặp trước nó thì ta dừng thuật toán.
4. Cập nhật center cho từng cụm bằng cách lấy trung bình cộng của tất các các điểm dữ liệu đã được gán vào cluster đó sau bước 2.
5. Quay lại bước 2.

### 2.3.2.5. Ưu và nhược điểm của thuật toán

Spectral clustering khắc phục hai vấn đề chính trong phân cụm: một là hình dạng của cụm và hai là xác định trọng tâm của cụm. K-means giả định các cụm có hình cầu, với bán kính bằng khoảng cách giữa tâm và điểm dữ liệu xa nhất, và không hoạt động tốt khi các cụm có hình dạng khác nhau, ví như cụm hình elip. Do vậy để xác định tâm của cụm thì K-means sẽ sử dụng nhiều lần lặp. Các cụm trong Spectral clustering không có hình dạng hoặc mẫu cố định. Các điểm cách xa nhau nhưng được kết nối sẽ thuộc về cùng một cụm, trong khi các điểm gần nhau hơn nhưng không được kết nối có thể thuộc về các cụm khác nhau.

Điều này chứng tỏ rằng thuật toán hiệu quả với dữ liệu có hình dạng và kích cỡ khác nhau.

Về mặt tính toán, nhanh với các tập dữ liệu thưa thớt (khoảng vài ngàn data point) và đắt với các tập dữ liệu lớn. Nguyên nhân do các giá trị riêng và vector riêng cần được tính toán, và sau đó ta mới thực hiện phân cụm trên các vector này, cho nên đã làm tăng độ phức tạp về thời gian lên.

### 2.3.3. Các thuật toán phân cụm phổ

#### 2.3.3.1. Thuật toán phân cụm phổ phi chuẩn

**Input:** + Ma trận liên kề  $W \in \mathbb{R}^{n \times n}$  của đồ thị  $G = (V, E)$

+ Ma trận bậc  $D$  của đồ thị  $G$

+ Số cụm  $K$

**Output:** Các cụm  $C_1, \dots, C_K$

1. Tính ma trận Laplacian  $L = D - W$
2. Tính  $K$  vector giá trị đặc trưng  $u_1, \dots, u_K$  của  $L$  ứng với  $K$  giá trị đặc trưng nhỏ nhất
3. Đặt  $U \in \mathbb{R}^{n \times K}$  là ma trận có các cột là các vector giá trị đặc trưng  $u_1, \dots, u_K$
4. for  $i = 1$  to  $n$  do
5.  $y_i \in \mathbb{R}^K$  là vector ứng với hàng thứ  $i$  của  $U$
6. Gọi thuật toán  $K$ -means đối với các tâm cụm  $y_1, \dots, y_K$  để phân  $V$  thành  $K$  cụm.

#### 2.3.3.2. Thuật toán phân cụm phổ chuẩn hóa

**Input:** + Ma trận liên kề  $W \in \mathbb{R}^{n \times n}$  của đồ thị  $G = (V, E)$

+ Ma trận bậc  $D$  của đồ thị  $G$

+ Số cụm  $K$

**Output:** Các cụm  $C_1, \dots, C_K$

1. Tính ma trận Laplacian chuẩn hóa  $L_{\text{sym}} = D^{-1/2} L D^{-1/2}$



2. Tính K vector giá trị đặc trưng đầu  $u_1, \dots, u_K$  của  $L_{\text{sym}}$
3. Đặt  $U \in \mathbb{R}^{n \times K}$  là ma trận có các cột là các vector giá trị đặc trưng  $u_1, \dots, u_K$
4. Tạo ra ma trận  $T \in \mathbb{R}^{n \times K}$  từ  $U$  bằng cách chuẩn hóa các hàng theo chuẩn 1:

$$t_{ij} = \frac{u_{ij}}{\sqrt{\sum_{k=1}^n u_{ik}^2}}$$

5. for  $i = 1$  to  $n$  do
6.  $y_i \in \mathbb{R}_K$  là vector ứng với hàng thứ  $i$  của  $T$
7. Gọi thuật toán K-means đối với các tâm cụm  $y_1, \dots, y_K$  để phân  $V$  thành  $K$  cụm.

## 2.4. PHƯƠNG PHÁP TRA CỨU ẢNH SỬ DỤNG PHÂN CỤM PHỔ

### 2.4.1. Phát biểu bài toán

Tra ra cứu ảnh dựa trên nội dung là một đề tài được đầu tư nghiên cứu và phát triển trong những năm gần đây trong ngành khoa học xử lý ảnh. Rất nhiều hệ thống CBIR đã được ra đời, có thể kể tới QIBC, PhotoBook, VisualSeek, WeekSeek, RetrievalWare, Virage, ImageRover, PicHunter, Amore, Blobworld, vân vân [33].

Nguyên lý hoạt động của một hệ thống CBIR: Các đặc trưng cấp thấp của ảnh (bao gồm màu sắc, kết cấu, hình dạng) sẽ được tự động trích xuất theo mục tiêu được đánh chỉ số và mô tả ảnh. Trong trường hợp tra cứu ảnh trong cơ sở dữ liệu ảnh, một ảnh truy vấn mẫu sẽ được đưa vào hệ thống, sau đó nó sẽ được xử lý để sinh ra một vectơ phù hợp. Sau đây, dựa trên độ đo tương tự, một tập các ảnh được xếp loại theo thứ tự giảm dần so với ảnh truy vấn mẫu, sẽ được trả về.

Độ đo tương tự sẽ định lượng mức độ phù hợp và tương đồng giữa hai hay nhiều hình ảnh. Trong CBIR, dựa trên các kiểu đặc trưng của ảnh tra cứu mà ta sẽ sử dụng các độ đo thích hợp. Hiệu suất tra cứu của hệ thống CBIR cũng phụ thuộc vào phép đo độ tương tự của vectơ đặc trưng ảnh cơ sở dữ liệu và vectơ đặc trưng ảnh truy vấn. Độ đo tương tự và khoảng cách ngữ nghĩa có mối quan

hệ tương hỗ lẫn nhau. Khoảng cách ngữ nghĩa thể hiện sự khác biệt trong việc mô tả đối tượng bằng các cách biểu diễn ngôn ngữ khác nhau, đồng thời phản ánh sự khác biệt giữa khả năng mô tả hạn chế của các đặc trưng cấp thấp với các đặc trưng cấp cao. Việc giảm khoảng cách ngữ nghĩa giữa các đặc trưng cấp thấp và cấp cao là một thách thức lớn của lĩnh vực tra cứu ảnh dựa trên nội dung [33].

Có rất nhiều phương pháp và kỹ thuật đã được áp dụng để rút ngắn khoảng cách ngữ nghĩa, trong đó có thể kể tới như là: sử dụng mô hình học máy (machine learning) để liên kết các đặc trưng cấp thấp với đặc trưng cấp cao, áp dụng mô hình tra cứu dựa trên cây hoặc đồ thị, xây dựng đối tượng bản thể học (ontology), kết hợp phản hồi liên quan (relevance feedback - RF), v.v.

Trong một hệ thống CBIR có kết hợp phản hồi liên quan, người dùng có thể gửi vào một hình ảnh mẫu hoặc một yêu cầu truy vấn đơn giản tới hệ thống tra cứu, với mục đích tìm kiếm những hình ảnh đúng như mong muốn. Bước một, hệ thống tra cứu sẽ biểu thị ảnh đầu vào hoặc yêu cầu truy vấn bằng một vector đặc trưng. Bước hai, khoảng cách (chính là các điểm tương đồng) giữa các vector đặc trưng của ví dụ truy vấn và các vector đặc trưng của hình ảnh trong cơ sở dữ liệu sẽ được tính toán và xếp hạng. Bước ba, tra cứu được thực hiện bằng cách áp dụng lược đồ lập chỉ mục. Bước bốn, hệ thống xếp hạng các kết quả tìm kiếm và sau đó trả về các kết quả gần giống với ví dụ truy vấn nhất. Bước 5, người dùng sẽ nhận được những kết quả tìm kiếm này và đánh giá xem liệu nó có liên quan tới hình ảnh mình muốn tìm hay có phù hợp với mong muốn của mình hay không. Vòng lặp sẽ được lặp lại ở bước một cho đến khi nào người dùng tìm được đúng hình ảnh theo yêu cầu của mình mới thôi.

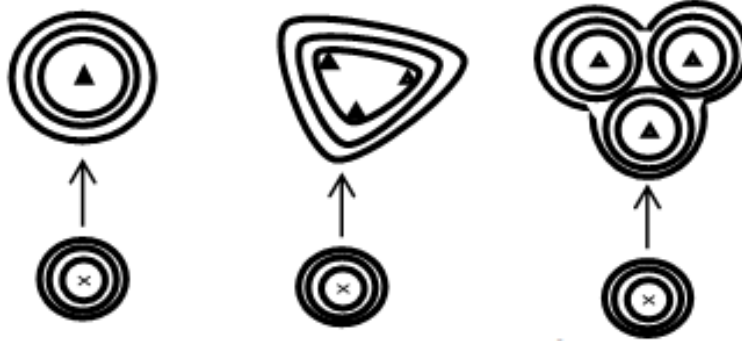
Mặc dù đã cải thiện được tính năng tra cứu ảnh, tuy nhiên kỹ thuật phản hồi liên quan vẫn có những hạn chế nhất định. Nguyên lý hoạt động của hệ thống CBIR truyền thống kết hợp RF, chính là những ảnh được trả về, sẽ có liên quan hoặc có điểm tương đồng với ảnh mẫu truy vấn trong không gian đặc trưng. Tuy nhiên, sự tương đồng giữa các hình ảnh trong tâm trí người dùng, sẽ có sự sai biệt với khoảng cách giữa chúng trong không gian đặc trưng. Điều này có nghĩa là, những hình ảnh liên quan đến ảnh mẫu, có thể nằm rải rác ở trên toàn bộ

không gian đặc trưng chứ không tập trung ở một miền nhất định, phân thành nhiều cụm chứ không tập hợp ở duy nhất một cụm. Để khắc phục hạn chế này, chúng ta sẽ áp dụng phương pháp phân cụm phổ trong CBIR, kết hợp cùng RF [33].

Như đã trình bày ở trên, chúng ta có thể thấy rằng, kỹ thuật phản hồi liên quan truyền thống không hoạt động tốt khi di chuyển tâm truy vấn. Cơ chế hoạt động của phản hồi liên quan, chính là tính toán một hay nhiều điểm truy vấn mới trong không gian đặc trưng, đồng thời thay đổi hàm khoảng cách. Áp dụng phương pháp phân cụm phổ, ta có thể tính được các điểm truy vấn mới, những truy vấn mới này có sử dụng các phản hồi/đánh giá của người dùng về việc hình ảnh có liên quan hay không [33].

Hình 2.6(a) đã biểu diễn một truy vấn mới bằng một điểm đơn, đồng thời thay đổi các trọng số của các thành phần đặc trưng, qua đó tìm ra một điểm truy vấn tối ưu và một hàm khoảng cách tối ưu. Ở trường hợp này, một điểm đơn được tính toán bằng cách sử dụng trung bình trọng số của tất cả các ảnh liên quan trong không gian đặc trưng. Các đường viền biểu diễn các đường có độ tương tự tương đương [33].

Hình 2.6(b) thì biểu diễn một truy vấn mới bằng nhiều điểm (đa điểm), qua đó xác định hình của đường viền. Ở trường hợp này, một phương pháp phân cụm đã được sử dụng để tính toán các điểm truy vấn mới, trong đó các điểm truy vấn mới này có sử dụng các phản hồi/đánh giá của người dùng về việc hình ảnh có liên quan hay không. Trên giả định là: Các hình ảnh liên quan được ánh xạ sang các điểm gần nhau theo độ đo tương tự. Một đường viền rộng được xây dựng để phủ tất cả các điểm truy vấn và hệ thống tìm các ảnh tương tự với các truy vấn này [33].



Hình 2. 6(a) Dịch chuyển điểm truy vấn (b) Hình dạng lõi (đa điểm)(c) Hình dạng lỗm (đa điểm)

Mục đích của hệ thống CBIR là để cho người dùng có thể tra cứu hình ảnh từ các cơ sở dữ liệu ảnh lớn. Trong hệ thống, các hình ảnh sẽ được biểu diễn bằng các đặc trưng từ mức thấp đến mức cao (giữa chúng là khoảng cách ngữ nghĩa), và các hàm khoảng cách sẽ được sử dụng để điều khiển các truy vấn. Vậy nên không thể phủ nhận sự tương tác, hỗ trợ lẫn nhau của khoảng cách ngữ nghĩa và độ đo tương tự. Hạn chế của hệ thống CBIR là đôi khi các ảnh cần tìm lại được xếp hạng theo độ đo tương tự đối với các ảnh truy vấn, vậy nên những hình ảnh này sẽ không được trả về nếu như chúng nằm rải rác trong không gian đặc trưng, và trong khi kết quả tìm kiếm chỉ là các ảnh trong khoảng lân cận với ảnh truy vấn mà thôi.

Để giải quyết những hạn chế này, luận văn sẽ đề xuất và nghiên cứu một phương pháp cải thiện sự tương tác giữa người dùng và hệ thống tra cứu ảnh, thông qua việc thu thập, tận dụng ý kiến của người dùng và các độ đo tương tự. Người dùng sẽ không phải dùng quá nhiều hình ảnh để biểu đạt yêu cầu tra cứu của mình, và thời gian tra cứu cũng sẽ không tăng trong trường hợp có nhiều ý kiến hoặc ảnh phản hồi.

#### 2.4.2. Phân tích và xây dựng mô hình

Ta gọi phương pháp sử dụng là Spectral Clustering Relevance Feedback, viết tắt là SCRF. Hình 2.7 là sơ đồ thể hiện quá trình hoạt động của SCRF [34].

Bước 1, tập ảnh truy vấn được đưa vào, hệ thống tiến hành trích xuất đặc trưng của ảnh. Các ảnh này sẽ được lưu trữ thành tập các vectơ đặc trưng.

Bước 2, áp dụng độ đo tương tự với các véctơ đặc trưng có được ở bước một, tiến hành so sánh, phân hạng độ tương đồng của các ảnh truy vấn với các ảnh có trong cơ sở dữ liệu.

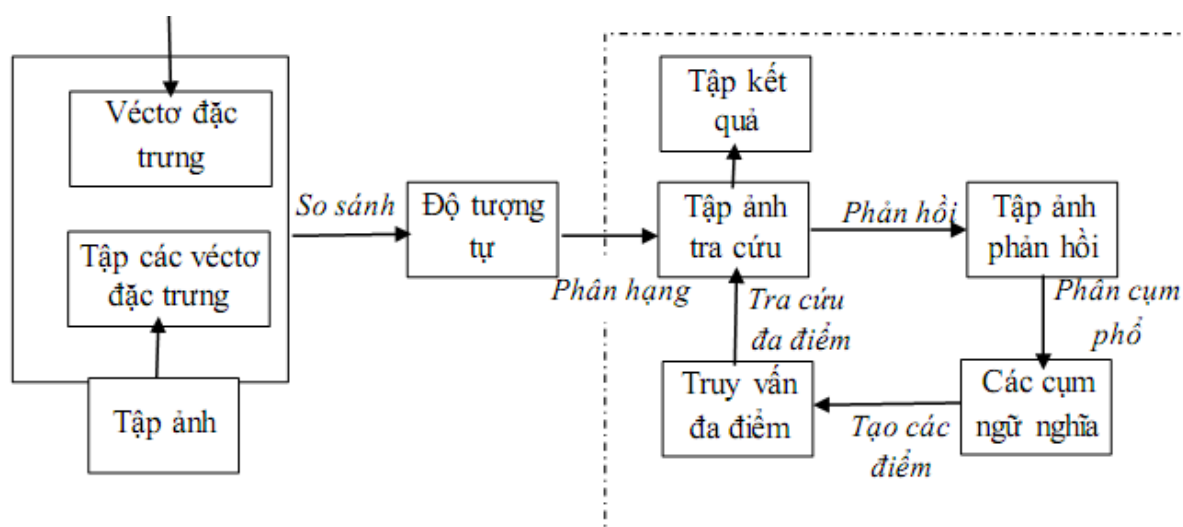
Bước 3, xây dựng được tập ảnh lân cận với tập ảnh truy vấn, tạo thành tập kết quả và chuyển nó cho người dùng.

Bước 4, người dùng chọn ra những ảnh phù hợp với yêu cầu của mình, tập ảnh phản hồi dựa trên ý kiến của người dùng được hình thành từ đây.

Bước 5, ở tập ảnh phản hồi, thuật toán phân cụm phổ sẽ được áp dụng. Các cụm ngữ nghĩa từ đó được hình thành.

Bước 6, với mỗi cụm ngữ nghĩa được tạo dựng, SCRF sẽ tiến hành tìm đại diện cho mỗi cụm để hình thành truy vấn đa điểm.

Bước 7, quá trình tra cứu được lặp lại ở bước 3, chỉ dừng lại khi người dùng ngừng phản hồi và đưa ra tập kết quả cuối cùng.



Hình 2. 7. Cấu trúc của phương pháp SCRF

## 2.5. THUẬT TOÁN PHÂN CỤM TRONG PHƯƠNG PHÁP TRA CỨU ẢNH

### 2.5.1. Ý tưởng

Ý tưởng: SCRF tiến hành phân cụm tập ảnh phản hồi của người dùng, nhờ đó mà có được các cụm ngữ nghĩa. Thông qua các cụm ngữ nghĩa, SCRF tìm đại diện cho mỗi cụm. Mỗi đại diện được sử dụng để thực hiện truy vấn đa điểm ở lần lặp tra cứu tiếp theo [34].

Tính cải thiện: SCRF sẽ tìm các ảnh tương đồng với bất cứ điểm nào hay đại diện nào của truy vấn đa điểm. Qua đó sẽ lấy được tập đa dạng các ảnh nằm phân tán trong toàn bộ không gian đặc trưng, chứ không chỉ tập trung tìm kiếm vào truy vấn trung tâm dựa trên các mẫu tích cực mà người dùng chọn nữa [34].

### 2.5.2. Thuật toán phân cụm tập ảnh phản hồi từ người dùng

Ở bước 4 của phương pháp SCRF, người dùng sẽ chọn ra tập ảnh phù hợp với yêu cầu của mình, qua đó hình thành nên tập ảnh phản hồi. Để khai thác thông tin tương đồng giữa các ảnh trong tập ảnh phản hồi, ta sử dụng thuật toán Clustering Relevant Images Set Using Eigenvectors, viết tắt là CRISE, hay còn gọi là  $k$  véctơ riêng, để hình thành các cụm ngữ nghĩa. Những ảnh được chọn làm đại diện cho mỗi cụm ngữ nghĩa phải là ảnh tương đồng nhất với tất cả các ảnh trong cụm. Sau đó là đến bước 5, 6, 7 của mục 2.4.2.

Nguyên nhân sử dụng CRISE, là do thuật toán này hoạt động tốt hơn so với các thuật toán phân cụm truyền thống. Ưu điểm của CRISE chính là: nó phân cụm các hình ảnh được kết nối với nhau, và những ảnh này không nhất thiết phải bị phân nhóm vào trong các vùng có hình dạng lồi (đa điểm) - xem lại Hình 2.6(b). Vậy nên ta có thể dễ dàng tìm kiếm được những hình ảnh có liên quan nằm phân tán trong tất cả các phần của không gian đặc trưng, đồng thời cải thiện được tính chính xác cũng như thời gian tra cứu [34].

Dưới một biểu diễn đồ thị, phân cụm có thể được phát biểu tự nhiên như một bài toán phân hoạch đồ thị. Ở đây, chúng ta sử dụng phương pháp sử dụng  $k$  véctơ riêng và tính trực tiếp phân hoạch  $k$ -way. So với phương pháp sử dụng một véctơ riêng tại một thời điểm và gọi đệ quy, phương pháp sử dụng  $k$  véctơ riêng được chỉ ra là tốt hơn về mặt thực hành. Nói chung, một phương pháp phân hoạch đồ thị cố gắng tổ chức các nút thành các nhóm sao cho độ tương tự trong phạm vi nhóm là cao, và/hoặc độ tương tự giữa các nhóm là thấp. Một đồ thị đã cho  $G = (V, E)$  với ma trận affinity  $A$ , một cách đơn giản để định lượng giá cho các nút phân hoạch thành hai tập rời nhau  $C_1$  và  $C_2$  ( $C_1 \cap C_2 = \emptyset$  và  $C_1 \cup C_2 = V$ ) là tổng có trọng số của các cạnh mà kết nối hai tập.

Đầu tiên, từ  $n$  điểm dữ liệu ảnh, phương pháp xây dựng ma trận affinity  $A$

$$a_{ij} = e^{-\frac{\|s_i - s_j\|^2}{2\sigma^2}} \quad (i \neq j, a_{ii} = 0) \quad (2.19)$$

Ở đây tham số tỉ lệ  $\sigma^2$  điều khiển mức độ áp lực  $a_{ij}$  giảm nhanh thế nào với khoảng cách giữa  $s_i$  và  $s_j$ . Một giá trị  $a_{ij}$  giữa hai ảnh là “cao” nếu hai ảnh là rất tương tự.

Xây dựng ma trận đường chéo  $D$  trong đó phần tử  $(i,i)$  là tổng hàng thứ  $i$  của ma trận  $A$ .  $D$  là một ma trận chéo với  $D_{ii} = \sum_{j=1, \dots, n} a_{ij}$

Tính ma trận Laplace chuẩn hóa:  $L = D^{-1/2} A D^{-1/2}$

Tìm  $k$  vectơ riêng  $x_1, x_2, \dots, x_k$  lớn nhất của ma trận  $L$ , trong đó  $x_1 = (x_{11}, x_{12}, x_{13}, \dots, x_{1n})$ ,  $x_2 = (x_{21}, x_{22}, x_{23}, \dots, x_{2n})$ ,  $\dots$ ,  $x_k = (x_{k1}, x_{k2}, x_{k3}, \dots, x_{kn})$  và xây dựng ma trận  $X = [x_1^T, x_2^T, \dots, x_k^T] \in \mathbb{R}^{n \times k}$ , cụ thể:

	$x_1^T$	$x_2^T$	$x_3^T$	...	$x_k^T$
$x_1$	$x_{11}$	$x_{21}$	$x_{31}$	...	$x_{k1}$
$x_2$	$x_{12}$	$x_{22}$	$x_{32}$	...	$x_{k2}$
$x_3$	$x_{13}$	$x_{23}$	$x_{33}$	...	$x_{k3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_k$	$x_{1n}$	$x_{2n}$	$x_{3n}$		$x_{kn}$

Xây dựng ma trận  $Y$  từ  $X$  bằng việc chuẩn hóa mỗi dòng của  $X$  là chiều dài đơn vị của ma trận  $Y$ .

$$Y_{ij} = \frac{x_{ij}}{(\sum_j x_{ij}^2)^{1/2}} \quad (2.20)$$

$y_1$	$y_{11}$	$y_{12}$	$y_{13}$	...	$y_{1k}$
$y_2$	$y_{21}$	$y_{22}$	$y_{23}$	...	$y_{2k}$
$y_3$	$y_{31}$	$y_{32}$	$y_{33}$	...	$y_{3k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_k$	$y_{n1}$	$y_{n2}$			$y_{nk}$

Mỗi dòng của ma trận Y được xem như là một điểm trong không gian véctor k chiều. Đến đây, sẽ có n điểm trong không gian  $R^k$ , phân cụm  $(y_i)_{i=1 \dots n}$  trong không gian  $R^k$  thành k cụm  $C_1, C_2, \dots, C_k$  thông qua K-Means. Cuối cùng, gán điểm si tới cụm j nếu và chỉ nếu hàng thứ i của ma trận Y tương ứng với cụm j.

Dưới đây mô tả thuật toán phân cụm sử dụng k véctor riêng CRISE (Clustering Images Set Using Eigenvectors) thực hiện việc phân cụm tập các ảnh liên quan mà người dùng chọn thành k cụm.

**Thuật toán CRISE**

**Input:** - Tập các ảnh  $S = \{s_1, s_2, \dots, s_n\}; s_i \in R^*$   
 - Số cụm k;

**Output:** k cụm:  $C_1, C_2, \dots, C_k$ ;

**Bước 1: Xây dựng ma trận affinity**  
 For i ← 1 to n do  
     For j ← 1 to n do  
         If  $i \neq j$   $a_{ij} \leftarrow \exp(-\frac{d_{ij}}{2})$   
         Else  $a_{ij} \leftarrow 0$

**Bước 2: Xây dựng ma trận đường chéo và ma trận Laplace L**  
 For i ← 1 to n do  
      $d_{ii} \leftarrow \sum_{j=1, \dots, n} a_{ij}$   
      $L \leftarrow D^{-1/2} A D^{-1/2}$

**Bước 3 : Tìm k véctor riêng lớn nhất  $x_1, x_2, \dots, x_k$  của ma trận Laplace**  
 For i ← 1 to k do

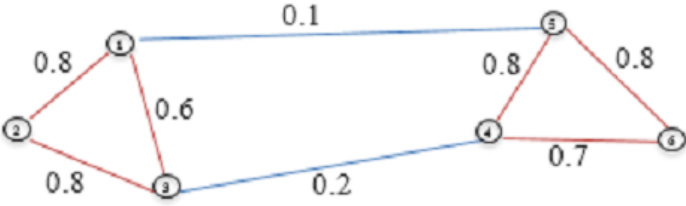


```

 $x_i \leftarrow \text{Largest\_eigen\_vector}(L)$ 
 $X \leftarrow [x_1^T, x_2^T, \dots, x_k^T]$ 
Bước 4: Xây dựng ma trận Y từ X
For i  $\leftarrow$  1 to n do
    For j  $\leftarrow$  1 to n do
         $y_{ij} \leftarrow x_{ij} / \sqrt{\sum_{k=1}^k x_{ik}^2}^{1/2}$ 
     $Y \leftarrow [y_1, y_2, \dots, y_k]$ 
Bước 5: Phân thành k cụm thông qua K-Means
 $P \leftarrow \emptyset$ 
For i  $\leftarrow$  1 to n do
     $p_i \leftarrow y_i$ 
     $P \leftarrow P \cup p_i$ 
    K-Mean(P)
Bước 6: Gán các  $s_i$  vào các cụm
For i  $\leftarrow$  1 to n do
    If  $( \quad )$ 
         $p_i \in C_j, j=1, \dots, k$ 
         $C_j \leftarrow C_j \cup s_i$ 
Return  $C_1, C_2, \dots, C_k$ 

```

**Ví dụ:** Giả sử tập ảnh trong cơ sở dữ liệu được biểu diễn bởi đồ thị như Hình 2.5, nhiệm vụ là phân tập ảnh thành hai cụm (k=2).



Hình 2. 8. Đồ thị biểu diễn cơ sở dữ liệu ảnh

Tính ma trận Affinity A, ta được:

0	0.8	0.6	0	0.1		0
0.8	0	0.8	0	0		0
0.6	0.8	0	0.2	0		0
0	0	0.2	0	0.8		0.7

0.1	0	0	0.8	0		0.8
0	0	0	0.7	0.8		0

Tính ma trận đối xứng D, ta có:

**D =**

<b>1.5</b>	0	0	0	0	0
0	<b>1.6</b>	0	0	0	0
0	0	<b>1.6</b>	0	0	0
0	0	0	<b>1.7</b>	0	0
0	0	0	0	<b>1.7</b>	0
0	0	0	0	0	<b>1.5</b>

Sau khi có D, ta có D-1/2:

**D-1/2 =**

<b>1.224</b>	0	0	0	0	0
0	<b>1.264</b>	0	0	0	0
0	0	<b>1.264</b>	0	0	0
0	0	0	<b>1.303</b>	0	0
0	0	0	0	<b>1.303</b>	0
0	0	0	0	0	<b>1.22</b>

Từ D-1/2 và A, tính ma trận Laplace chuẩn hóa L, ta có:

**L =**

<b>0</b>	1.237	0.928	0	0.159	0
1.237	<b>0</b>	1.278	0	0	0
0.928	1.278	<b>0</b>	0.329	0	0
0	0	0.329	<b>0</b>	2.489	1.116
0.159	0	0	1.358	<b>1.292</b>	1.275
0.151	0	0	2.408	2.505	<b>1.214</b>

Tìm giá trị riêng của ma trận Laplace L, ta có:

$\lambda_1$	<b>4.453078</b>	$\lambda_4$	-1.289313
$\lambda_2$	<b>2.294362</b>	$\lambda_5$	-0.735609+0.1143889i
$\lambda_3$	-1.480126	$\lambda_6$	-0.735609- 0.1143889i

Chọn 2 véc tơ riêng lớn nhất tương ứng với hai giá trị riêng lớn nhất là  $\lambda_1 = 4.453078$  và  $\lambda_2 = 2.294362$ , ta có:

$$+ x_1 = (0.03431904, 0.02337397, 0.04820157,$$

0.46421601, 0.49628174, 0.73086278)

+  $x_2 = (0.55156401, 0.60938701, 0.55977393,$

$-0.01995990, -0.05256799, -0.08897172)$

Sau khi có hai véc tơ riêng  $x_1, x_2$ , xây dựng ma trận  $X = [x_1^T, x_2^T] \in \mathbb{R}^{n \times k}$

$X =$

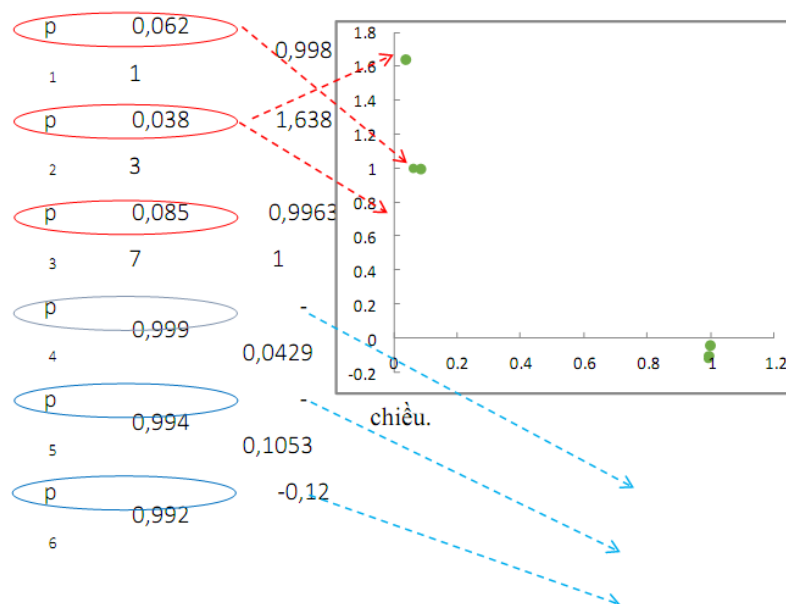
0.03431904	0.55156401
0.02337397	0.60938701
0.04820157	0.55977393
0.46421601	-0.01995990
0.49628174	-0.05256799
0.73086278	-0.08897172

Tiếp theo, chuẩn hóa ma trận  $X$  để có ma trận  $Y$

$Y =$

0,062101223	0,998069856
0,038328344	1,638582578
0,08579152	0,996313111
0,999076908	-0,042957319
0,994436856	-0,105334415
0,992671624	-0,12084307

Mỗi dòng của  $Y$  sẽ tương ứng với một điểm  $p_i$  trong không gian véc tơ 2:

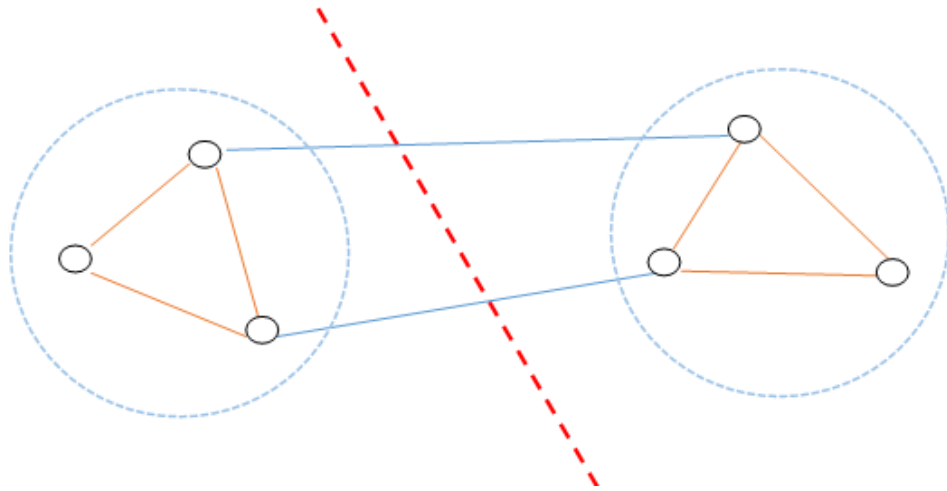


Phân cụm K-Means chúng ta được:

- Cụm A gồm:  $p_1, p_2, p_3$ ;

- Cụm B gồm: p4, p5, p6

Gán si vào các cụm:



### 2.5.3. Tìm ảnh đại diện cho cụm

Để thực hiện việc tra cứu ảnh hiệu quả, phải thu được một ảnh đại diện thích hợp cho mỗi cụm. Ở đây, một ảnh được chọn là đại diện cho một cụm phải là ảnh mà tương đồng nhất với tất cả các ảnh trong cụm. Phát biểu này được minh họa bằng toán học như sau: Với một biểu diễn đồ thị của các ảnh được cho  $G=(V,E)$  với ma trận affinity  $A$ , cho tập các cụm ảnh là  $\{C_1, C_2, \dots, C_k\}$  (tập các cụm này cũng này cũng là một phân hoạch của  $V$ , tức là  $(C_i \cap C_j = \emptyset$  và  $\bigcup_{i=1}^k C_i = V)$  và của là:

$$\arg \max_{j \in C_i} \sum_{j \in C_i} a_{jt} \quad (2.15)$$

Như vậy, với một cụm, ảnh đại diện là ảnh mà có tổng độ tương tự trong phạm vi cụm là cực đại.

### 2.5.4. Khoảng cách từ một ảnh đến truy vấn đa điểm

Khác với các phương pháp tra cứu ảnh khác, phương pháp này sẽ hình thành lên truy vấn đa điểm  $M_Q = (Q_1, Q_2, \dots, Q_k)$  từ các đại diện của mỗi cụm. Khi đó, khoảng cách từ một ảnh đến truy vấn đa điểm  $M_Q = (Q_1, Q_2, \dots, Q_k)$  là cực tiểu của các khoảng cách có trọng số từ một ảnh đến mỗi  $Q_j$  trong truy vấn đa điểm và được tính theo công thức:

$$D(DI_i, MQ) = \min_{j=1\dots k} \text{dist}(DI_i, Q_j) \quad (2.16)$$

Trong công thức (2.14),  $\text{dist}(DI_i, Q_j)$  với  $i=1\dots, N, j=1\dots, k$  là khoảng cách từ một ảnh  $DI_i$  đến một điểm truy vấn  $Q_j$  trong truy vấn đa điểm MQ.

### 2.5.5. Thuật toán tra cứu ảnh sử dụng phân cụm phổ trong phản hồi liên quan

Dưới đây mô tả "Thuật toán tra cứu ảnh sử dụng phân cụm phổ trong phản hồi liên quan", viết tắt là SCRF. Khi người dùng thực hiện truy vấn, phương pháp sẽ sử dụng thuật toán MQMRBR để tra cứu trên tập các ảnh cơ sở dữ liệu DI và cho kết quả là tập các ảnh S. Người dùng thực hiện việc chọn tập các ảnh liên quan E trong tập S thông qua hàm *User\_Choose\_RelevanceImage()*, phương pháp sẽ phân cụm tập E này thành k cụm thông qua thuật toán *CRIES* và tìm đại diện cho k cụm đó thông qua hàm *Compute\_Representative()* và gán cho tập đại diện. Khoảng cách giữa ảnh cơ sở dữ liệu  $DI_i$  và truy vấn đa điểm MQ được tính theo công thức (2.14). Quá trình này tiếp tục cho đến khi người dùng dừng việc chọn các ảnh liên quan.

<b>Thuật toán tra cứu ảnh hiệu quả sử dụng phân cụm phổ trong phản hồi liên quan</b>		
<b>Input</b>		
Tập N ảnh cơ sở dữ liệu		DI
Ảnh truy vấn		Q
<b>Output</b>		
Tập ảnh kết quả		S'

```

MQMRBR (DI, Q, S) // thực hiện trên tập ảnh DI với truy vấn Q để cho ra tập kết
quả S
  Repeat
    E ← User_Choose_Relevancelmage (S, n) // người dùng chọn các ảnh liên
quan từ tập S
    C ← CRIES (E, k) // phân tập ảnh liên quan E thành k cụm
    RI ← Compute_Representative (C, M)
    For I ← 1 to N do
      For j ← 1 to k do
        Tính  $dis_i$  theo công thức sau:  $dis_i = \min_{j=1 \dots k} dis_{ij}$ 
      Sort (DI) // Sắp xếp các ảnh trong tập ảnh cơ sở dữ liệu DI theo thứ tự tăng
dần của khoảng cách so với truy vấn đã điểm MQ.
    Return S' // danh sách ảnh có khoảng cách nhỏ nhất với MQ
  Untill (User dừng phản hồi)

```

## Kết luận chương 2

Mục đích nghiên cứu của luận văn là sử dụng kỹ thuật phân cụm phổ kết hợp cùng phản hồi liên quan để cải thiện tính năng tra cứu ảnh dựa trên nội dung. Nghiên cứu của học viên được thực hiện dựa trên sự kế thừa và phát triển của một nghiên cứu trước đó là "Nghiên cứu phương pháp tra cứu ảnh dựa trên phương pháp phân cụm đồ thị" của tác giả Nguyễn Thị Hường, thực hiện ở trường đại học Công Nghệ Thông Tin và Truyền Thông Thái Nguyên.

Phương pháp SCRF và ứng dụng cùng ưu điểm của thuật toán CRISE được đề cập trong luận văn cũng là một sự tiếp nối và phát triển từ nghiên cứu của tác giả Nguyễn Thị Hường. Đây là những tiền đề quan trọng để học viên áp dụng và xây dựng một chương trình tra cứu ảnh mô phỏng. Chương trình mô phỏng áp dụng kỹ thuật phân cụm phổ kết hợp phản hồi liên quan sẽ được trình bày trong chương 3 luận văn.

## **CHƯƠNG 3. CHƯƠNG TRÌNH MÔ PHỎNG KỸ THUẬT PHÂN CỤM PHỔ VÀ ỨNG DỤNG ĐỂ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG**

### **3.1. CHƯƠNG TRÌNH MÔ PHỎNG KỸ THUẬT PHÂN CỤM PHỔ**

#### **3.1.1. Các bước mô phỏng phân cụm phổ**

Thuật toán phân cụm phổ - Spectral Clustering sử dụng thông tin là các giá trị riêng (phổ) của các ma trận được xây dựng từ đồ thị hoặc tập dữ liệu, tìm ra các điểm có độ tương đồng với nhau và tạo thành một cụm [42]. Trong bài toán tra cứu ảnh, đồ thị hoặc tập dữ liệu này được tạo nên từ thông tin được trích xuất từ ảnh.

Khái quát các bước trong kỹ thuật phân cụm phổ đã trình bày ở phần lý thuyết Chương 2 như sau:

- Dữ liệu vào: Đồ thị  $G$ , gồm đỉnh và các cạnh.
- Quá trình thực hiện:
  1. Tính ma trận kề  $W$ .
  2. Tính ma trận Laplacian  $L$ .
  3. Tính toán  $k$  vector riêng của ma trận  $L$ .
  4. Sử dụng thuật toán K-means để phân cụm dựa trên các vector riêng.
- Dữ liệu ra: Nhãn của cụm đối với từng điểm dữ liệu.

#### **3.1.2. Công cụ và dữ liệu**

Để mô phỏng thuật toán phân cụm phổ, học viên sử dụng các công cụ và dữ liệu như sau:

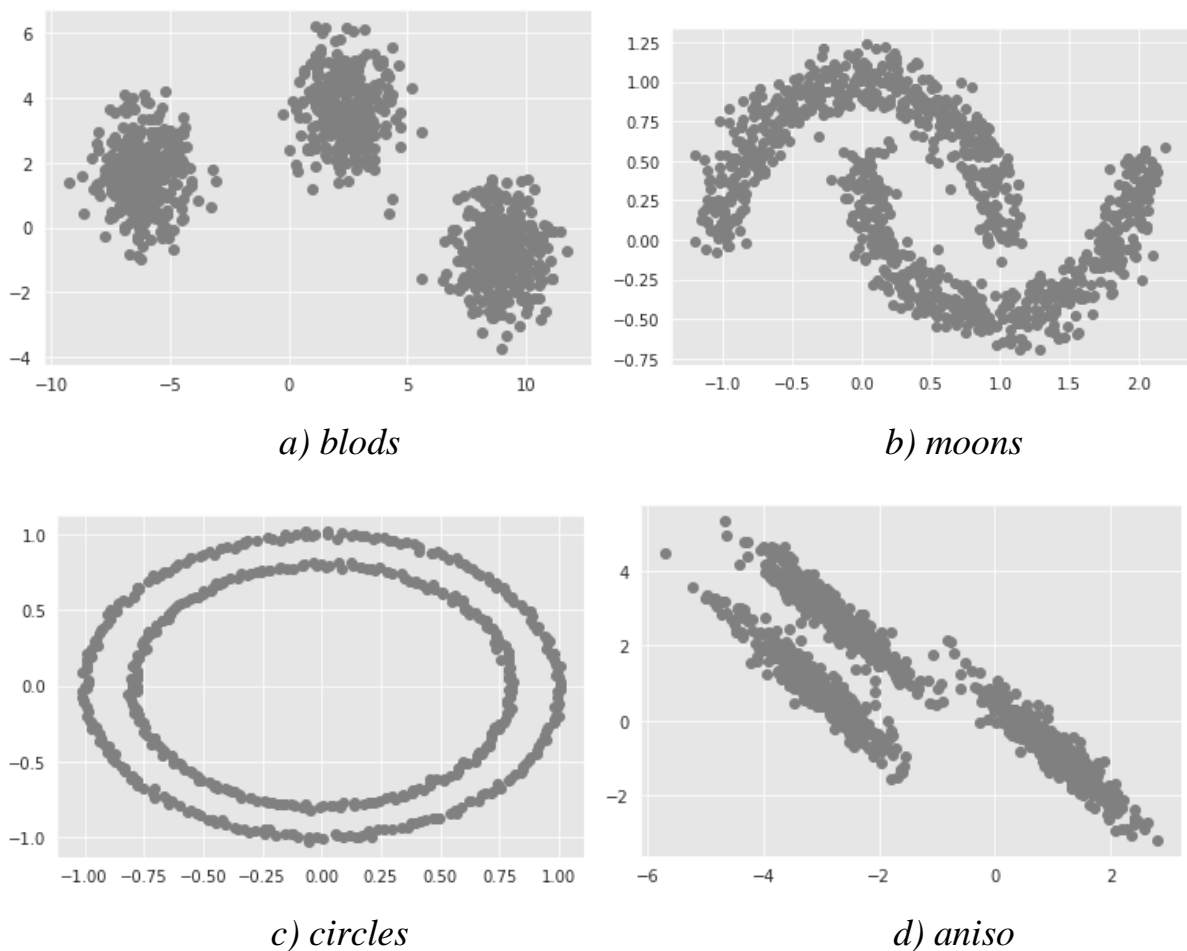
- Ngôn ngữ lập trình: Python: Là ngôn ngữ lập trình bậc cao, có tính linh hoạt, nhiều thư viện hỗ trợ, phù hợp cho các giải thuật khai phá dữ liệu, học máy và học sâu.

- Công cụ: Google CoLab Pro [35]: Là một dịch vụ điện toán đám mây của Google Research. Google CoLab vừa là một IDE vừa là môi trường Linux, cho phép các người dùng nào viết mã nguồn trong trình và chạy nó từ trình duyệt.

Công cụ này hỗ trợ ngôn ngữ lập trình Python được hỗ trợ CPU/GPU có tốc độ tính toán cao hơn nhiều so với các máy tính cá nhân.

- Thư viện Scikit-learn (Sklearn) [36]: Là thư viện học máy miễn phí cho các thuật toán học được viết trên ngôn ngữ Python. Chúng cung cấp một tập các công cụ xử lý các bài toán gồm: Phân lớp, phân cụm, hồi quy, giảm chiều dữ liệu. Thư viện được cấp phép bản quyền chuẩn FreeBSD và chạy được trên nhiều nền tảng bản phân phối của Linux.

- Dữ liệu: Thuật toán được chạy trên các dữ liệu thuộc Toy Dataset, bao gồm 04 bộ dữ liệu: blobs, moons, circles và aniso [37].



Hình 3. 1. Minh họa phân bố các điểm trong bộ dữ liệu

### 3.1.3. Các bước phân cụm với bộ dữ liệu blobs

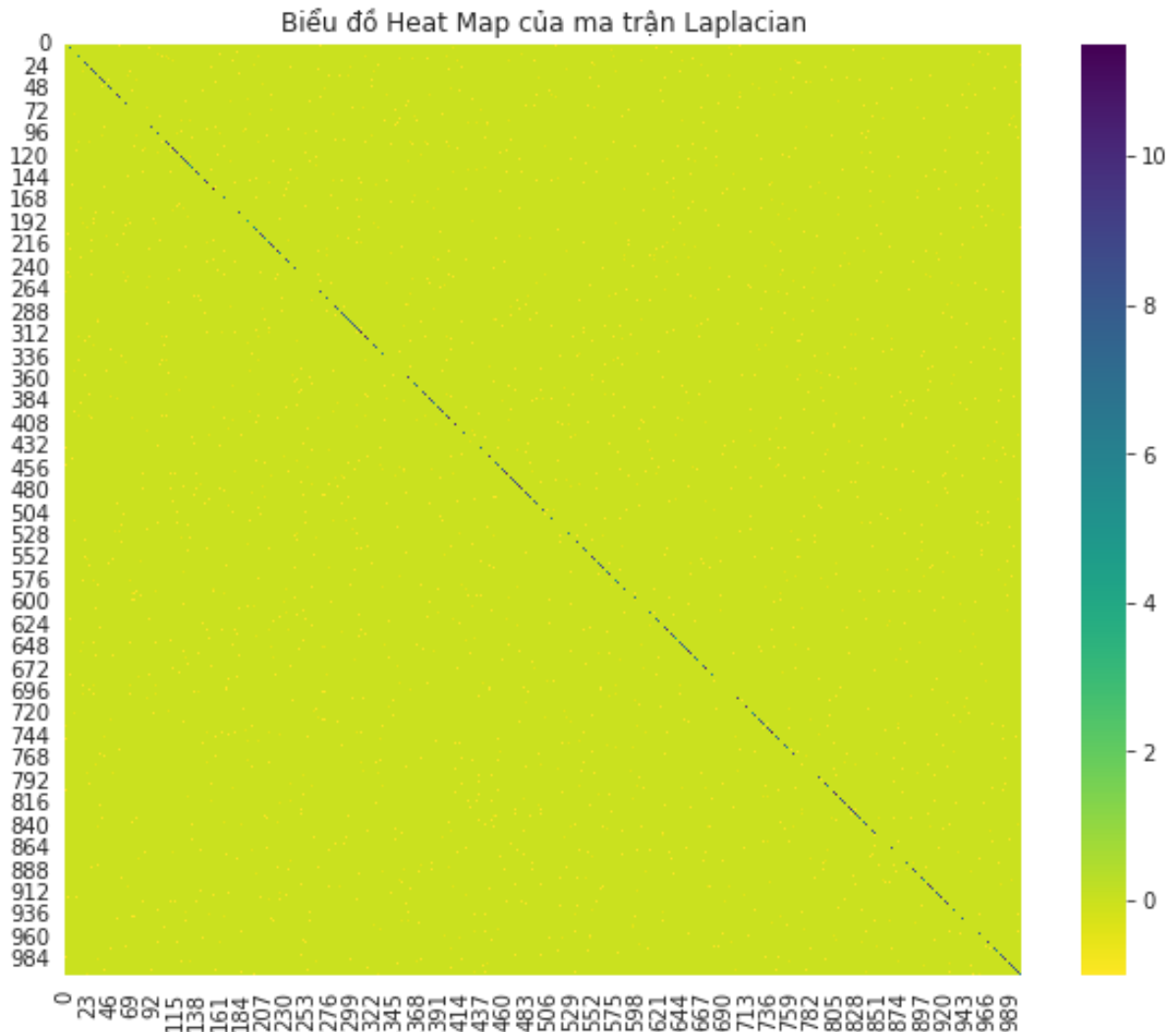
Lấy bộ dữ liệu blobs làm đại diện, thực hiện các bước trong kỹ thuật phân cụm phổ như sau:

- Tính ma trận Laplacian:



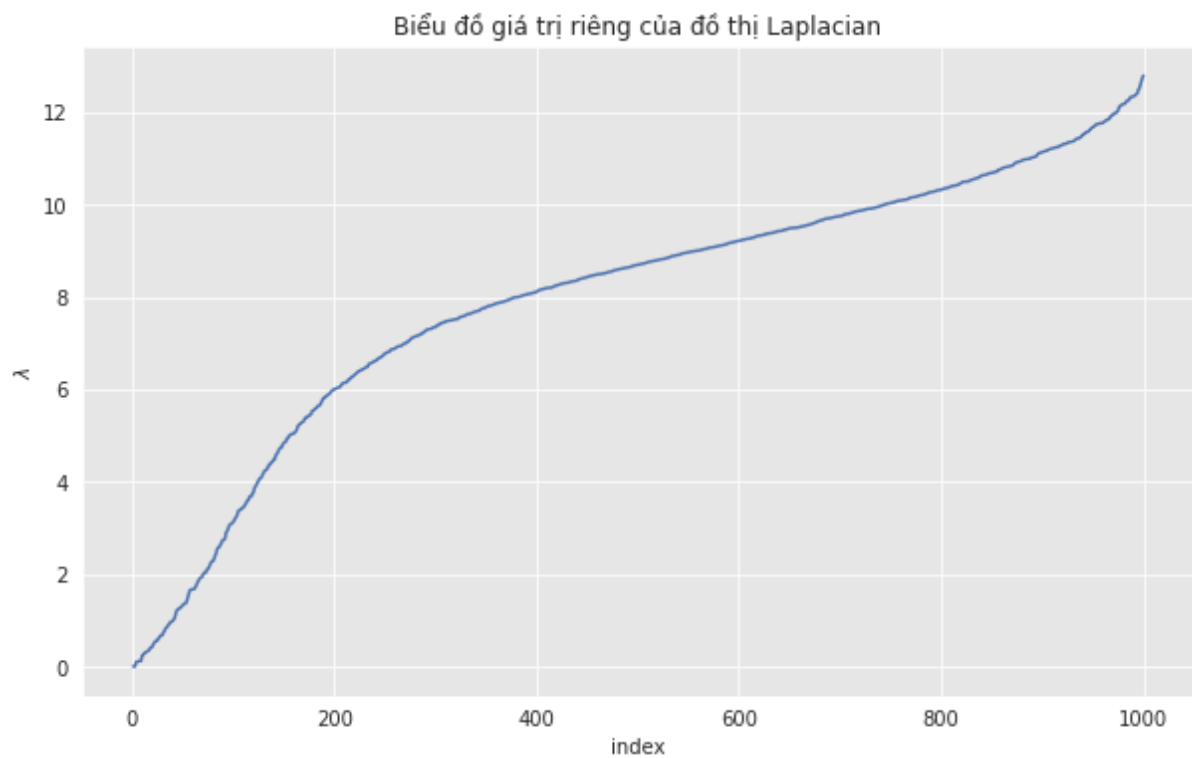
- + Tính ma trận kề  $W$ ;
- + Tính ma trận Laplacian từ ma trận kề  $W$ .

Hình 3.2 thể hiện biểu đồ heat map của ma trận Laplacian trên bộ dữ liệu blods. Biểu đồ này giúp biểu diễn trực quan độ lớn của các phần tử trong ma trận. Nhận xét rằng, ngoài các phần tử trên đường chéo chính, các phần tử còn lại phần đều nhỏ hơn hoặc bằng 0 (thể hiện bằng màu vàng nhạt).



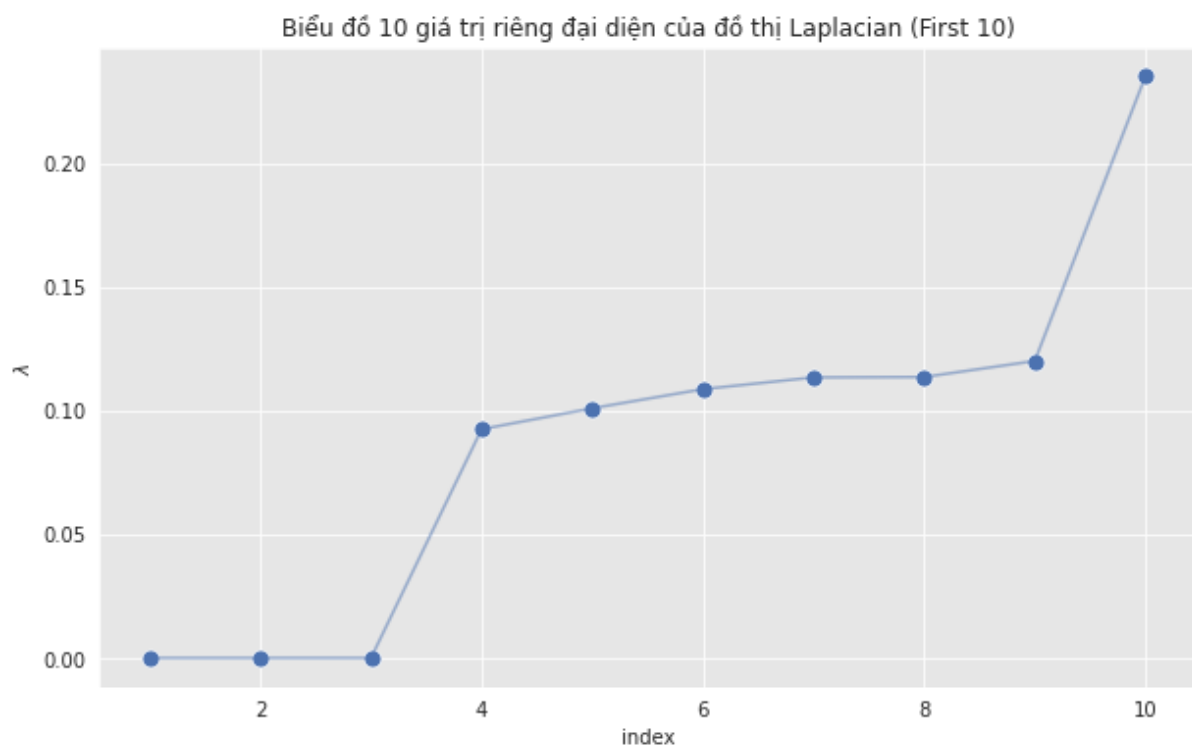
Hình 3. 2. Biểu đồ Heat Map của ma trận Laplacian trên bộ dữ liệu blods

Hình 3.3 thể hiện giá trị riêng của đồ thị Laplacian trên bộ dữ liệu blods, theo giá trị tăng dần. Lấy 10 giá trị đầu tiên từ biểu đồ giá trị riêng ở trên làm đại diện, ta thấy rõ được các giá trị riêng gần bằng 0 của đồ thị Laplacian, được thể hiện ở Hình 3.4.



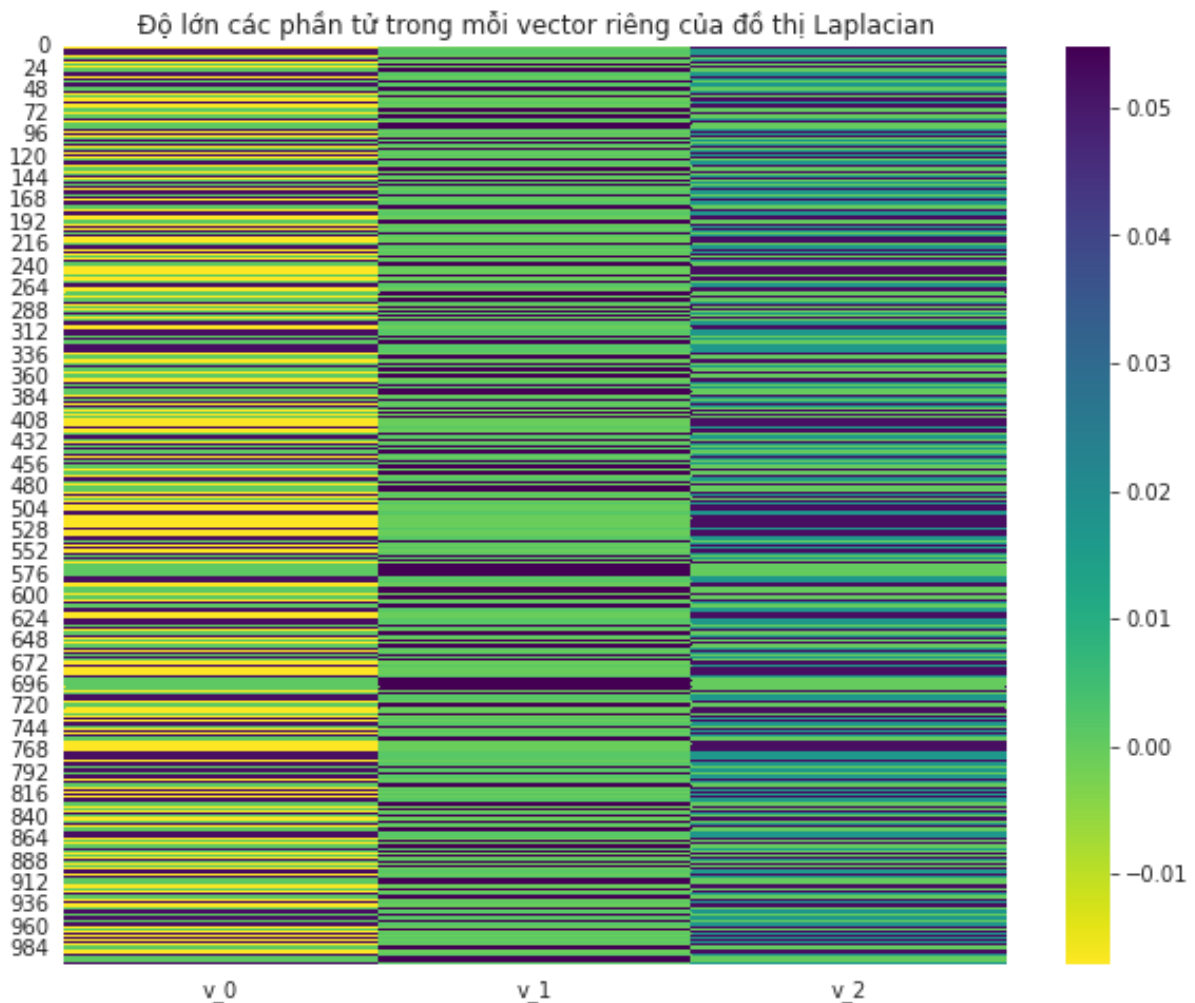
Hình 3. 3. Biểu đồ giá trị riêng của đồ thị Laplacian trên tập blods

Có thể thấy rằng, các điểm có index là 0, 1 và 2 có giá trị riêng  $\lambda$  bằng 0.



Hình 3. 4. Biểu đồ 10 giá trị riêng đầu tiên đại diện của đồ thị Laplacian trên bộ dữ liệu blods

Tính toán các vector riêng từ giá trị riêng, ta có biểu đồ thể hiện độ lớn các phần tử trong mỗi vector riêng của đồ thị Laplacian trên tập dữ liệu blods, được thể hiện tại Hình 3.5.

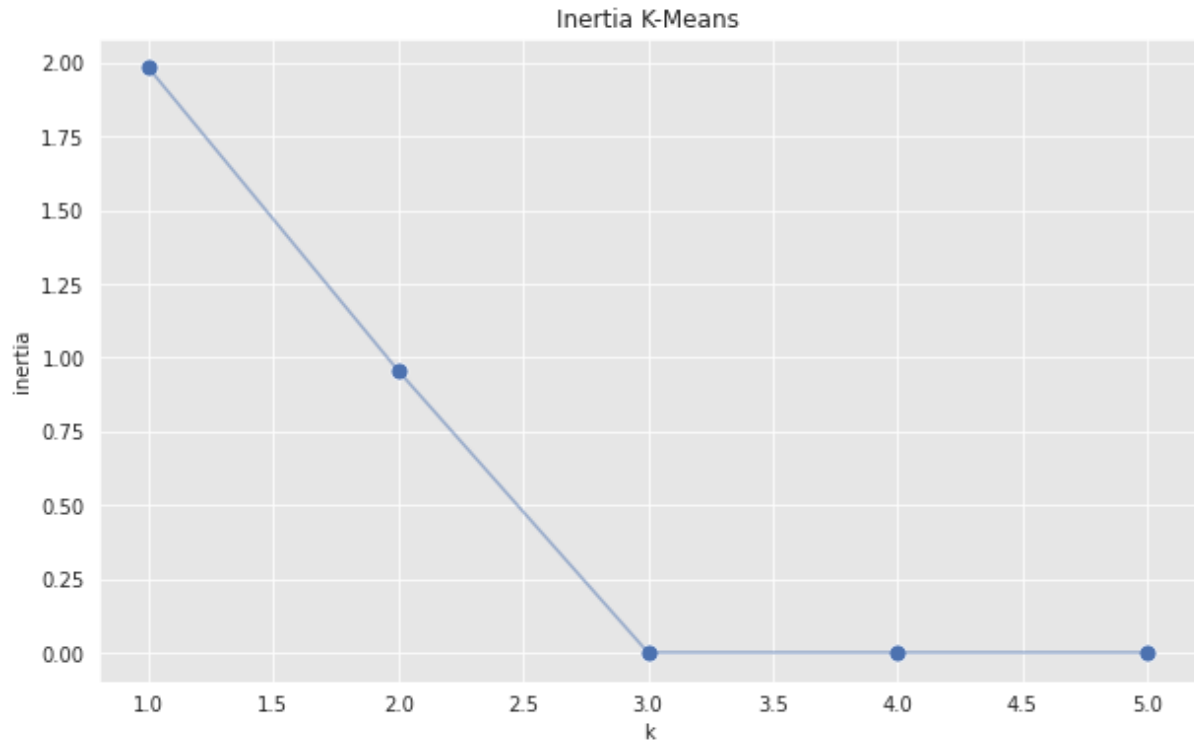


Hình 3. 5. Độ lớn các phần tử trong mỗi vector riêng của đồ thị Laplacian trên bộ dữ liệu blods

Biểu đồ trên cho thấy độ lớn của các phần tử, được nhóm vào 03 cụm là  $v_0$ ,  $v_1$  và  $v_2$ . Độ lớn của các giá trị được thể hiện bằng màu sắc tương ứng.

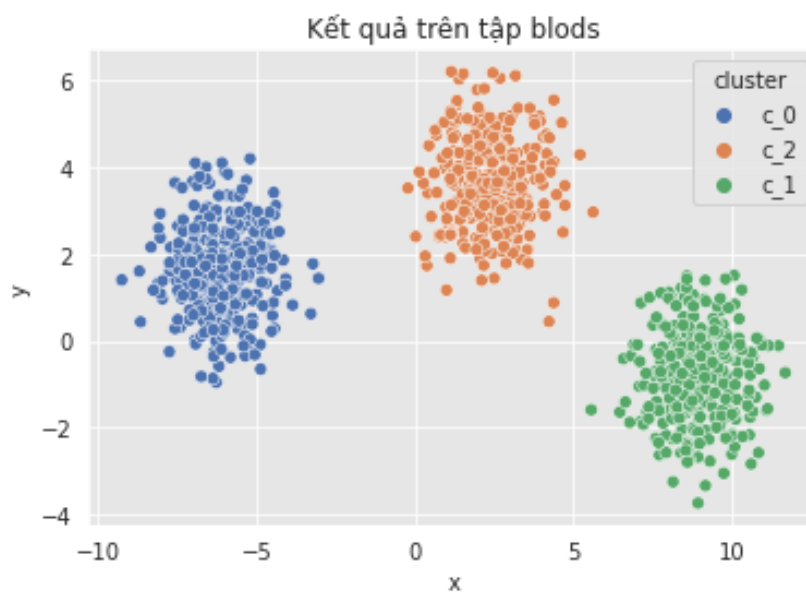
- Phân cụm với K-means: Ta sử dụng các vector riêng đầu tiên của ma trận Laplacian để phân cụm, áp dụng thuật toán với K-means. Với  $k$  là số cụm, thuật toán tiến hành phân các điểm trên đồ thị thành  $k$  cụm, sao cho các điểm trong cùng một cụm có nhiều sự tương đồng nhất.

Có thể sử dụng phương pháp elbow để xác định số cụm tương ứng. **Error! Reference source not found.** thể hiện kết quả xác định số cụm trên tập blods, với các số cụm đề nghị là 3, 4 hoặc 5 khi xem xét các số cụm trong khoảng (0, 6).



Hình 3. 6. Phương pháp elbow xác định số cụm trên tập blods

Kết quả phân cụm với K-means cho kết quả các điểm được chia thành 03 cụm (tương ứng với ba màu khác nhau), được thể hiện tại Hình 3.7.



Hình 3. 7. Kết quả phân thành 03 cụm trên tập blods

Nhận xét rằng, thuật toán phân cụm phổ với K-means hoạt động hiệu quả trên bộ dữ liệu blods, các điểm được nhóm thành ba cụm với các đặc trưng riêng rất rõ ràng.

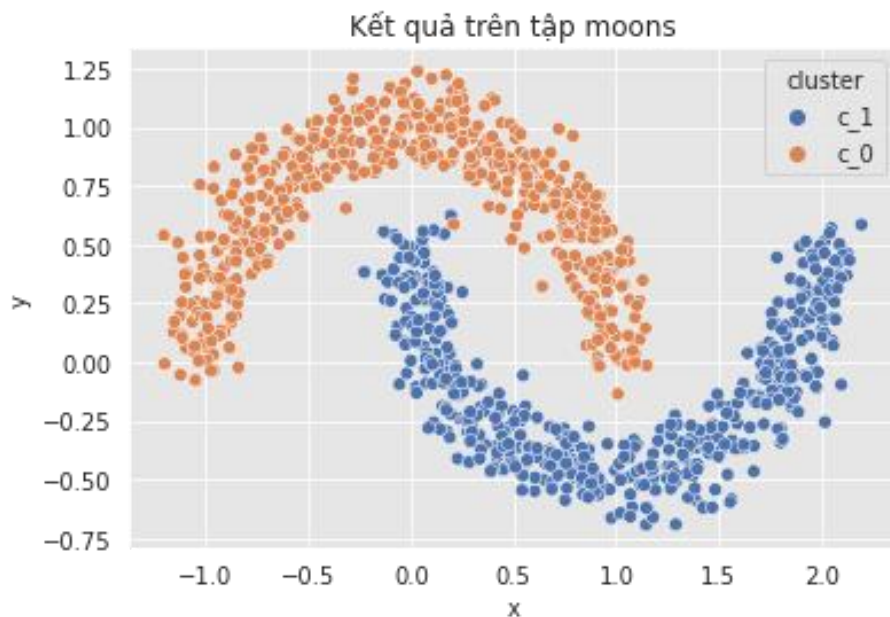
### 3.1.4. Kết quả phân cụm trên các bộ dữ liệu moons, circle và aniso

Tương tự các bước phân cụm đối với bộ dữ liệu blods, tiến hành phân cụm trên các bộ moons, circle và aniso. Thông tin về số lượng điểm và số cụm của các bộ dữ liệu được cho tại Bảng 3.1:

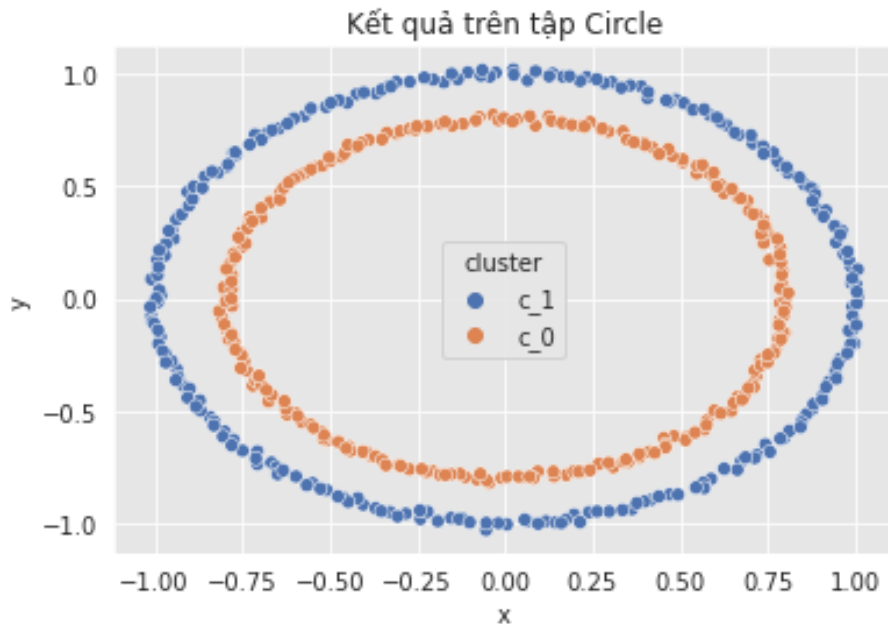
Bảng 3. 1. Thông tin về các bộ dữ liệu moons, circle và aniso

	moons	circle	aniso
Số điểm dữ liệu			
Số cụm	2	2	3

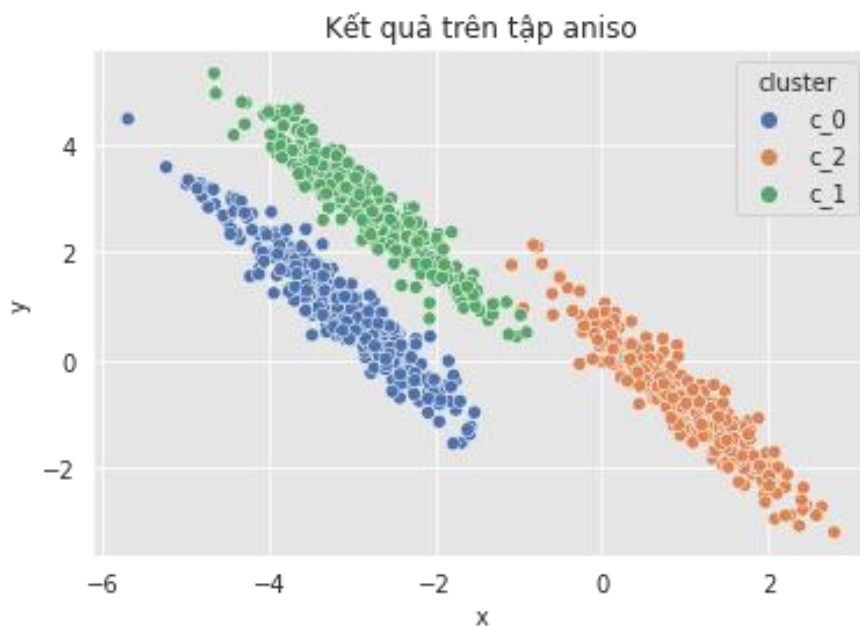
Kết quả phân cụm trên các bộ dữ liệu moons, circle và aniso lần lượt được thể hiện tại Hình 3.8, Hình 3.9 và Hình 3.10 như sau:



Hình 3. 8. Kết quả phân cụm trên tập moons với số cụm là 02



Hình 3. 9. Kết quả phân cụm trên tập circle với số cụm là 02



Hình 3. 10. Kết quả phân cụm trên tập aniso với số cụm là 03

Từ kết quả phân cụm trên 04 tập dữ liệu blods, moons, circle và aniso, ta thấy rằng thuật toán phân cụm phổ với K-means hoạt động hiệu quả. Các điểm trong cùng một cụm khi biểu diễn trên tọa độ thể hiện được sự tương đồng nhiều hơn so với các điểm không thuộc trong cụm.

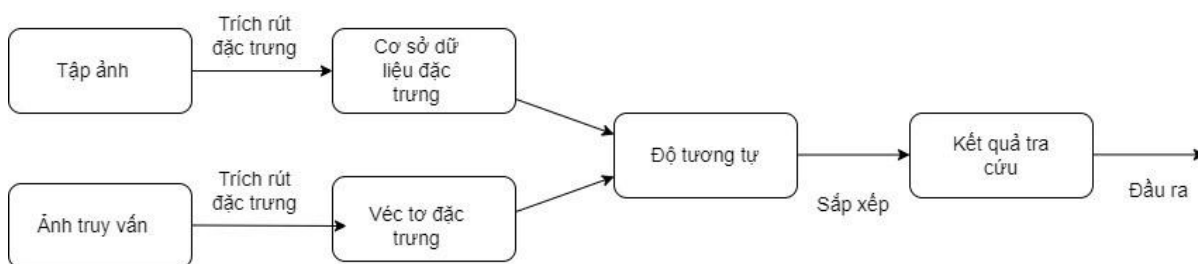
Thuật toán được ứng dụng cho bài toán tra cứu ảnh sử dụng kỹ thuật phân cụm phổ, được trình bày ở phần tiếp theo của chương này.

## 3.2. CHƯƠNG TRÌNH TRA CỨU ẢNH DỰA TRÊN NỘI DUNG SỬ DỤNG KỸ THUẬT PHÂN CỤM PHỔ

### 3.2.1. Mô hình của chương trình tra cứu ảnh dựa trên nội dung sử dụng kỹ thuật phân cụm phổ

#### 3.2.1.1. Hệ thống tra cứu ảnh dựa trên nội dung

Hệ thống tra cứu ảnh dựa trên nội dung - Content-Based Image Retrieval (CBIR) có chức năng tra cứu tìm các ảnh tương đồng với một ảnh được truy vấn. Thuật toán tiến hành phân tích nội dung của ảnh thay vì dựa vào các siêu dữ liệu như từ khóa, thẻ và các mô tả của ảnh. Ưu điểm của phương pháp tra cứu ảnh này là việc trích chọn đặc trưng được thực hiện một cách tự động và giữ được sự nhất quán trong nội dung của ảnh [42].

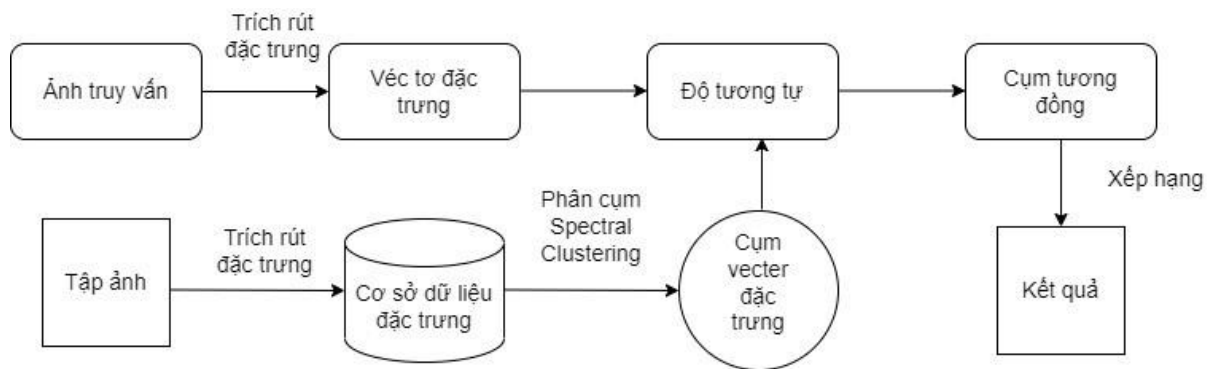


Hình 3. 11. Mô hình hệ thống tra cứu ảnh dựa trên nội dung

Hình 3.11 thể hiện mô hình hệ thống tra cứu ảnh dựa trên nội dung. Trong đó, các ảnh trong cả tập ảnh và ảnh truy vấn được trích xuất đặc trưng, thể hiện bằng các vector đặc trưng được lưu trữ trong cơ sở dữ liệu. Sự tương đồng giữa hai ảnh được đánh giá thông qua độ tương tự, là phép tính “khoảng cách” giữa hai vector đặc trưng của hai ảnh. Thông qua kết quả đánh giá này, các ảnh có độ tương đồng cao nhất sẽ được lựa chọn làm đầu ra cho quá trình tra cứu.

#### 3.2.1.2. Quá trình tra cứu ảnh dựa trên phân cụm phổ

Trong chương trình tra cứu ảnh dựa trên kỹ thuật phân cụm phổ, các bước chính của chương trình được thể hiện ở Hình 3.12:



Hình 3. 12. Quy trình tra cứu ảnh dựa trên phân cụm phổ

Xem xét tập ảnh là dưới dạng một đồ thị, với mỗi ảnh là một đỉnh, độ tương đồng giữa các cặp ảnh là trọng số của cạnh nối hai đỉnh đó. Thuật toán phân cụm phổ sẽ chia các bức ảnh trong tập ảnh thành các cụm khác nhau với nội dung tương đồng với nhau, với mỗi cụm được đặc trưng bởi một tâm cụm.

- Tập ảnh được rút trích đặc trưng, tạo nên cơ sở dữ liệu đặc trưng. Áp dụng kỹ thuật phân cụm phổ để tạo nên cụm vector đặc trưng.

- Ảnh truy vấn sau khi rút trích đặc trưng, sẽ được đánh giá độ tương đồng với cụm vector đặc trưng để tìm ra các cụm tương đồng.

Việc xếp hạng các cụm tương đồng này cho phép hệ thống đưa ra các ảnh tra cứu được có độ tương đồng cao nhất so với ảnh truy vấn.

Trong mô hình này, khi tra cứu ảnh, thay vì so sánh độ tương đồng giữa véc tơ đặc trưng của ảnh truy vấn với từng ảnh trong tập ảnh, thì ta chỉ cần so sánh nó với véc tơ đặc trưng của tâm các cụm, bởi tâm này có thể đại diện cho một cụm [42].

### 3.2.2. Môi trường và kỹ thuật

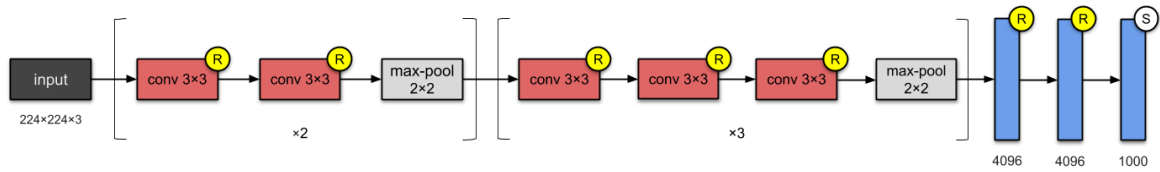
Môi trường phát triển chương trình như sau:

- Ngôn ngữ lập trình Python.
- Flask Framework [38]: Phát triển dựa trên Python để xây dựng chương trình hoạt động trên nền tảng Web.

Các kỹ thuật sử dụng:



- Trích xuất đặc trưng: Sử dụng mạng VGG16 ImageNet [39] để trích xuất đặc trưng cho hình ảnh, các đặc trưng tạo thành vector đặc trưng và lưu dưới dạng DataFrame.



Hình 3. 13. Kiến trúc mạng VGG16

Kiến trúc của mạng VGG16 được thể hiện ở Hình 3.13. Kiến trúc này bao gồm 13 layers tích chập hai chiều và 03 lớp kết nối đầy đủ. Trong VGG-16 ta cũng có khối tích chập. Đây là những kiến trúc gồm một tập hợp các lớp của CNN được lặp lại.

- Kỹ thuật phân cụm phổ, với thuật toán phân cụm K-means.

- Độ tương đồng giữa hai ảnh được đánh giá bằng công thức độ đo Euclid, tính bởi công thức sau:

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

Trong đó:

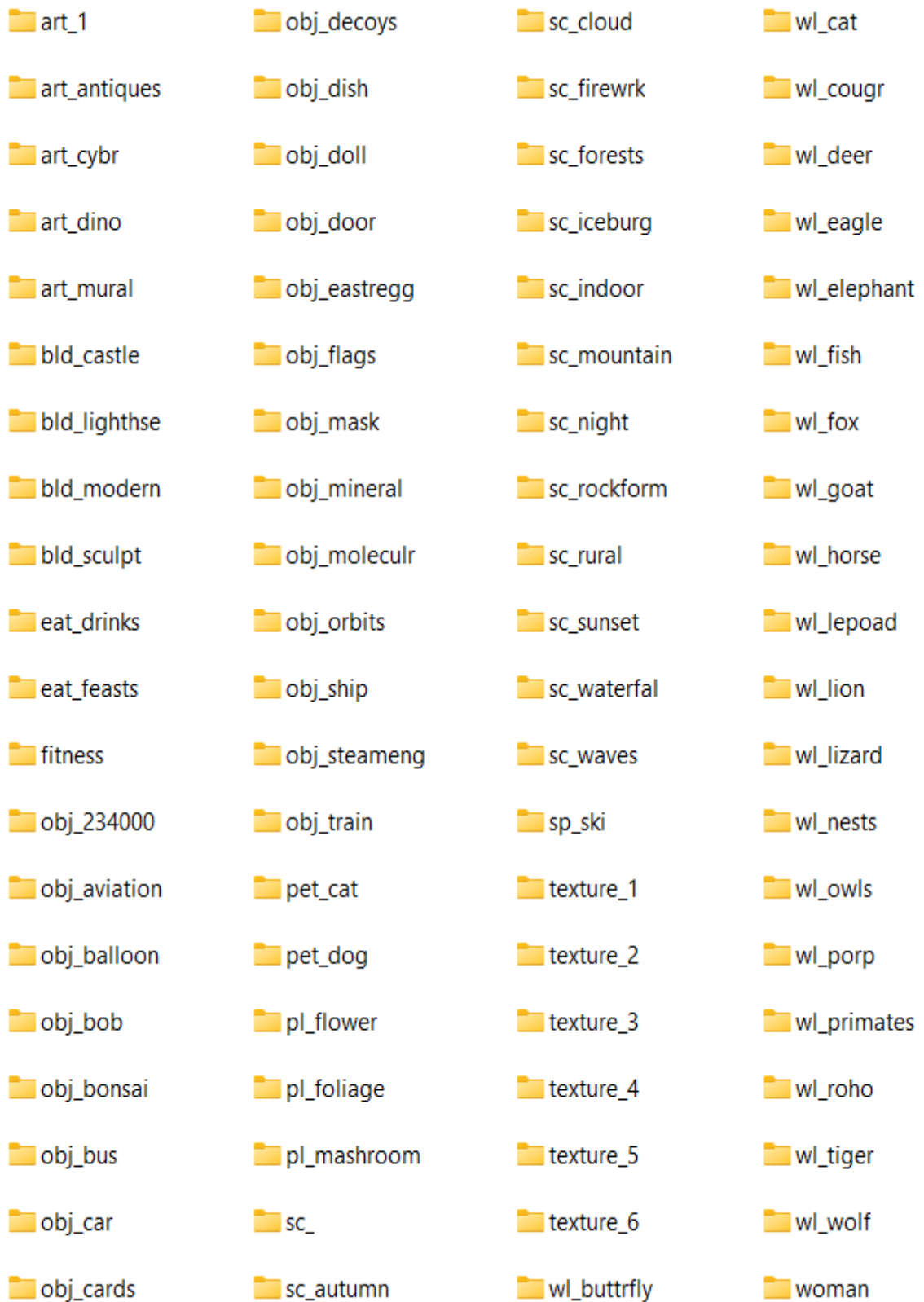
-  $D(x, y)$  là độ tương đồng giữa hai ảnh  $x$  và  $y$ ;

-  $(x_1, x_2, \dots, x_n)$  và  $(y_1, y_2, \dots, y_n)$  lần lượt là vector đặc trưng của hai ảnh  $x$  và  $y$ ;

### 3.2.3. Bộ dữ liệu ảnh

#### 3.2.3.1. Bộ dữ liệu Corel

Bộ dữ liệu ảnh thử nghiệm là Corel Database [40]. Bao gồm 10.800 ảnh có kích thước khác nhau, được chia thành 80 thư mục tương ứng với 80 lớp, thể hiện 80 loại đối tượng khác nhau trong bức ảnh. Bộ dữ liệu có tổng dung lượng là 39.9 MB.



*Hình 3. 14. Các thư mục chứa dữ liệu từng lớp ảnh*

Bảng 3.2 thể hiện chi tiết về các lớp (loại ảnh) và số lượng mẫu của từng lớp trong bộ dữ liệu ảnh được sử dụng.

Bảng 3. 2. Chi tiết về các lớp và số lượng mẫu trong bộ dữ liệu

Tên file	Số lượng ảnh	Nội dung bức ảnh	Tên file	Số lượng ảnh	Nội dung bức ảnh
art_1	100	Tranh	sc_cloud	107	Mây
art_antiques	100	Đồ cổ	sc_firewrk	185	Pháo hoa
art_cybr	100	Vũ trụ, viễn tưởng	sc_forests	100	Rừng
art_dino	100	Khủng long	sc_iceburg	100	Tảng băng trôi
art_mural	100	Tranh treo tường	sc_indoor	200	Trong nhà
bld_castle	160	Lâu đài	sc_mountain	398	Núi
bld_lighthse	100	Hải đăng	sc_night	100	Đêm
bld_modern	100	Công trình hiện đại	sc_rockform	100	Đá phong
bld_sculpt	100	Tượng	sc_rural	100	Nông thôn
eat_drinks	100	Đồ uống	sc_sunset	91	Hoàng hôn
eat_feasts	100	Lễ tiệc	sc_waterfal	115	Thác nước
fitness	100	Thể dục	sc_waves	100	Sóng
obj_234000	100	Bí ngô haloween	sp_ski	100	Trượt tuyết
obj_aviation	100	Máy bay cá nhân	texture_1	100	Kết cấu 1
obj_balloon	99	Khinh khí cầu	texture_2	100	Kết cấu 2
obj_bob	100	Xe mini	texture_3	100	Kết cấu 3
obj_bonsai	100	Cây cảnh	texture_4	100	Kết cấu 4
obj_bus	100	Xe buýt	texture_5	100	Kết cấu 5
obj_car	136	Xe ô tô con	texture_6	100	Kết cấu 6
obj_cards	100	Thẻ bài	wl_butterfly	200	Con bướm bướm
obj_decoys	100	Tượng	wl_cat	100	Mèo hoang dã
obj_dish	100	Ấm, bát sứ trung hoa	wl_cougr	100	Báo
obj_doll	100	Búp bê	wl_deer	200	Nai
obj_door	100	Cửa chính	wl_eagle	100	Đại bàng
obj_eastregg	100	Trứng giáng sinh	wl_elephant	100	Con voi
obj_flags	100	Cờ	wl_fish	200	Cá
obj_mask	1000	Mặt nạ	wl_fox	100	Cáo
obj_mineral	100	Khoáng sản	wl_goat	100	Dê
obj_molecular	100	Cấu trúc phân tử	wl_horse	100	Ngựa

obj_orbits	100	Quỹ đạo	wl_leopard	100	Báo
obj_ship	100	Thuyền	wl_lion	100	Sư tử
obj_steameng	100	Xe công nông	wl_lizard	100	Con thằn lằn
obj_train	100	Tàu hỏa	wl_nests	100	Chim yến
pet_cat	100	Mèo	wl_owls	100	Con chim cú mèo
pet_dog	150	Chó	wl_porpoise	100	Cá heo
pl_flower	400	Hoa	wl_primates	100	Loại linh trưởng
pl_foliage	100	Lá	wl_rohde	100	Con hà mã
pl_mushroom	100	Nấm	wl_tiger	100	Con hổ
Sc_	100	Hang động	wl_wolf	100	Chó sói
sc_autumn	103	Mùa thu	woman	550	Phụ nữ

Bộ dữ liệu Corel có đặc điểm là số lớp nhiều, với 80 lớp tương ứng với 80 chủ đề ảnh. Có sự khác nhau khá rõ ràng giữa chủ đề nội dung ảnh của mỗi lớp.

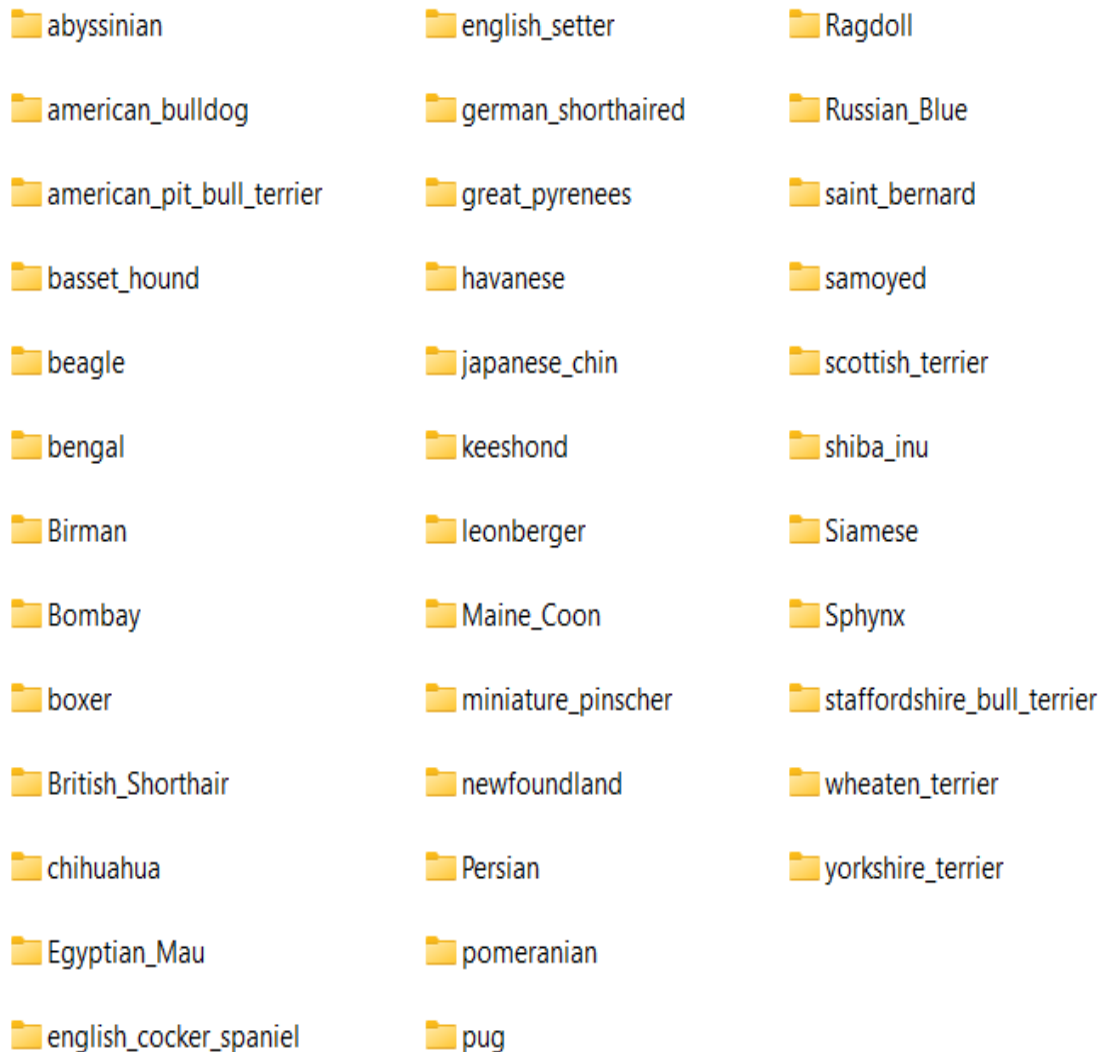
### 3.2.3.2. Bộ dữ liệu Oxford-IIIT Pet

Bộ dữ liệu Oxford IIIT-Pet [41] là bộ dữ liệu về các loại thú nuôi trong nhà, gồm 37 lớp tương ứng với 37 giống loài (Hình 3.15), chứa nhiều giống của cùng một loài. Mỗi lớp bao gồm 200 ảnh với tổng dung lượng là 755 MB.

*Bảng 3. 3. Danh sách các lớp trong bộ dữ liệu Oxford*

STT	Tên lớp	STT	Tên lớp
1.	Abyssinian	2.	Leonberger
3.	American_bulldog	4.	Maine_Coon
5.	American_pit_bill_terrier	6.	Miniature_pinscher
7.	Basset_hound	8.	newfoundland
9.	Beagle	10.	Persian
11.	Bengal	12.	Pomeranian
13.	Birman	14.	Pug
15.	Bombay	16.	Ragdoll
17.	Boxer	18.	Russian_Blue
19.	British_Shorthair	20.	Saint_bernard
21.	Chihuahua	22.	Samoyed
23.	Egyptian_Mau	24.	Scottish_terrier

25.	English_coket_spaniel	26.	Shiba_inu
27.	English_setter	28.	Siamese
29.	German_shorthaired	30.	Sphynx
31.	Great_pyrenees	32.	Staffordshine_bull_terrier
33.	Havanese	34.	Wheaten_terrier
35.	Japanese_chin	36.	Yorkshine_terrier
37.	keshound		



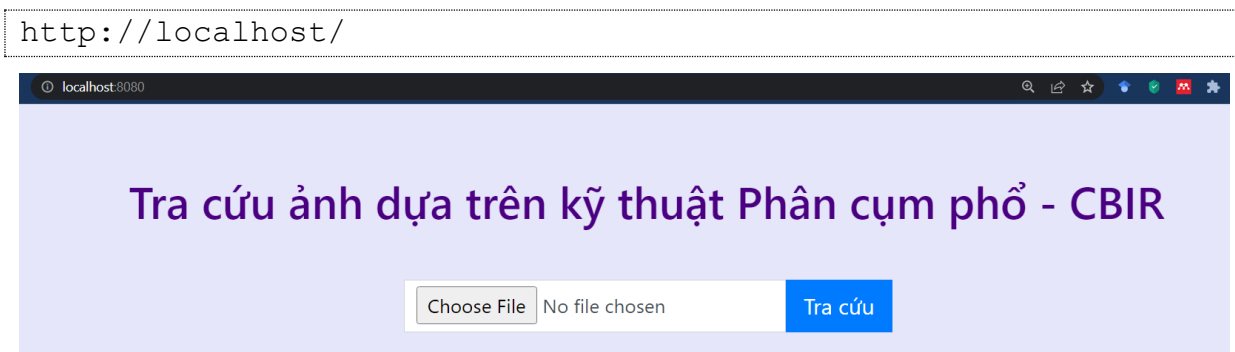
*Hình 3. 15. Các lớp tương ứng với mỗi giống loài trong bộ dữ liệu Oxford*

Nhìn chung, bộ dữ liệu Oxford chứa các ảnh có độ tương đồng khá cao giữa hai lớp. Ví dụ, cùng một loài có thể được chia thành nhiều lớp khác nhau tương ứng với mỗi loài như pug, chihuahua, shiba\_inu hay belgal.

### 3.2.4. Thực nghiệm và thảo luận

Chương trình tra cứu ảnh dựa trên nội dung được xây dựng và hoạt động trên nền tảng Web, triển khai trên localhost như sau:

Truy cập chương trình ra cứu truy cập qua localhost, với giao diện được thể hiện tại Hình 3.16:



Hình 3. 16. Trang chính hệ thống CBIR dựa trên kỹ thuật phân cụm phổ

Sau khi hoàn thành trích xuất đặc trưng cho bộ dữ liệu ảnh, chương trình hoạt động như sau:

- Lựa chọn ảnh cần tra cứu (gọi là ảnh truy vấn), có thể là một ảnh trong bộ dữ liệu trước đó hoặc một ảnh bất kỳ;
- Hệ thống CBIR tra cứu trong cơ sở dữ liệu và đưa ra các ảnh có độ tương đồng cao nhất với ảnh truy vấn;
- Hệ thống tính toán độ chính xác *Precision*, được tính bằng công thức sau:

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

Trong đó:

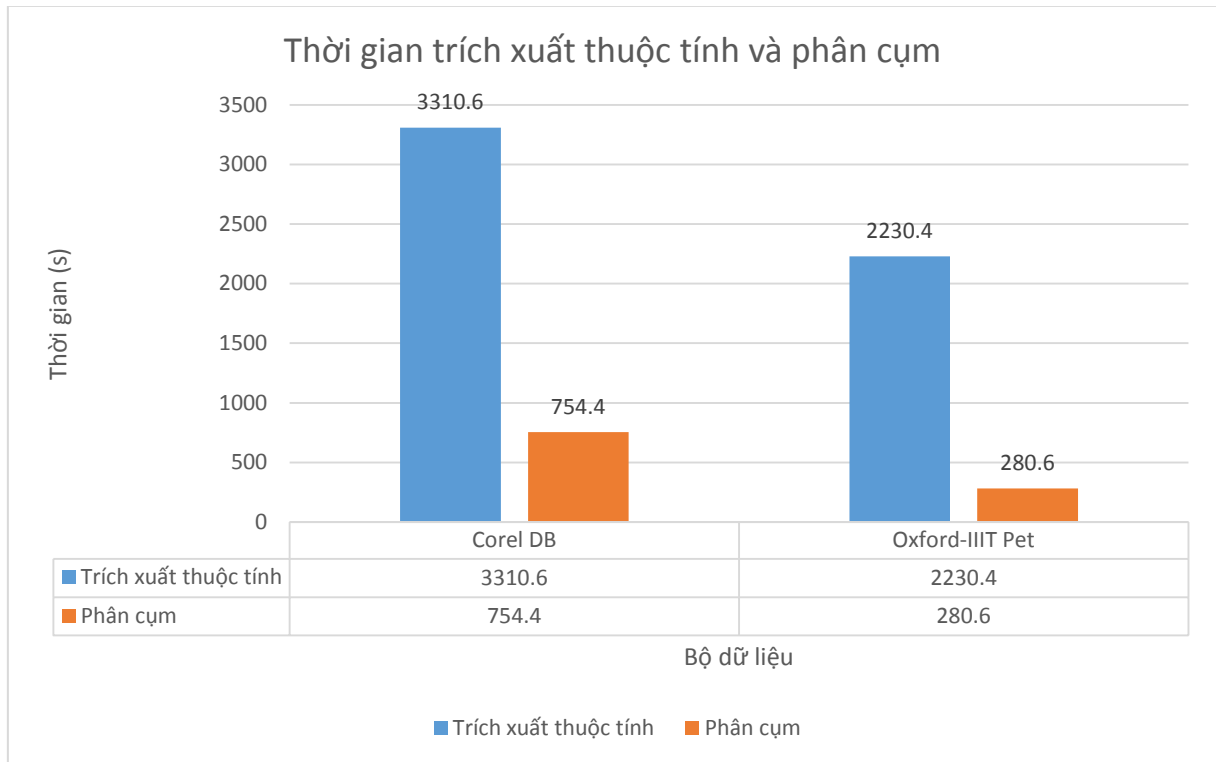
- *TP*: Là số ảnh đã được chọn mà đúng với ảnh truy vấn;
- *FP*: Là số ảnh đã được chọn mà sai với ảnh truy vấn;

#### 3.2.4.2. Trích xuất thuộc tính và phân cụm

Thuật toán thực hiện quá trình trích xuất thuộc tính và phân cụm các dữ liệu ảnh trước khi cho phép truy vấn. Quá trình trích xuất thuộc tính diễn ra tương tự

trên hai bộ dữ liệu. Quá trình phân cụm sử dụng thuật toán k-means, tương ứng là 80 cụm với bộ dữ liệu Corel và 37 cụm với bộ dữ liệu Oxford.

Thời gian dành cho trích xuất thuộc tính và phân cụm được thể hiện tại Hình 3.7. Các ảnh được trích xuất trên môi trường Windows 11, cấu hình chip Intel Core i5 8250U, RAM 16GB.



*Hình 3. 17. Thời gian cho trích xuất thuộc tính và phân cụm trên hai bộ dữ liệu Corel và Oxford*

Nhận xét rằng, việc trích xuất dữ liệu chiếm nhiều thời gian trong toàn bộ quá trình. Tuy nhiên, việc trích xuất được đánh giá là nhanh mặc dù chỉ sử dụng chipset dành cho thiết bị laptop di động. Thời gian trích xuất thuộc tính của bộ dữ liệu Oxford là ít hơn so với Corel, lần lượt là 2230.4s và 3310.6s, điều này có thể được giải thích bởi số lượng ảnh là ít hơn.

Đối với thời gian phân cụm, bộ dữ liệu Corel có thời gian phân cụm là 754.4s, chiếm 22.8% so với thời gian trích xuất, còn bộ dữ liệu Oxford là 280.6s, chiếm 12.6% so với thời gian trích xuất. Điều này một phần giải thích bởi số cụm trên bộ dữ liệu Corel là 80, nhiều hơn rất nhiều so với số cụm trên bộ Oxford là 37.


### 3.2.4.3. Truy vấn trên bộ dữ liệu Corel

Lần lượt thử nghiệm với các ảnh truy vấn trên bộ dữ liệu Corel, ta có kết quả minh họa ở các như sau:


**Tra cứu ảnh dựa trên kỹ thuật Phân cụm phổ - CBIR**

Choose File No file chosen **Tra cứu**

**ẢNH TRUY VẤN:**



**KẾT QUẢ:**  
Precision: 1.0



True True True True True True

*Hình 3. 18. Thử nghiệm với ảnh truy vấn là động vật*



ẢNH TRUY VẤN:



KẾT QUẢ:

Precision: 0.97



True



True



True

Hình 3. 19. Thử nghiệm với ảnh truy vấn là lâu đài

ẢNH TRUY VẤN:



KẾT QUẢ:

Precision: 1.0



True



True



True

Hình 3. 20. Thử nghiệm với ảnh truy vấn là bông hoa

Thử nghiệm với một số ảnh truy vấn khác, thu được kết quả ở Bảng 3.4:

Bảng 3. 4. Kết quả truy vấn với 10 người hợp ảnh đầu vào của bộ Corel

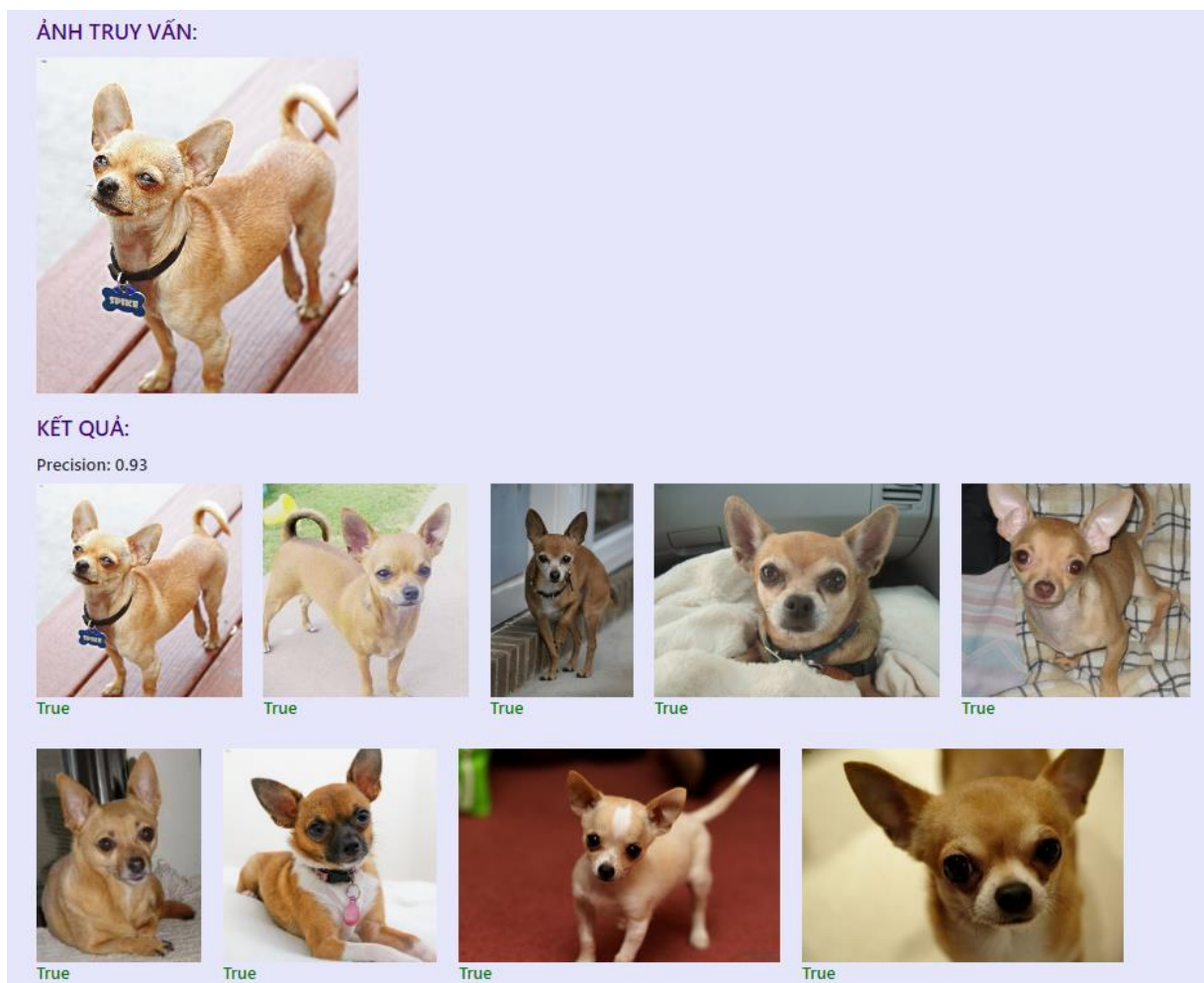
STT	Ảnh truy vấn	Độ chính xác	STT	Ảnh truy vấn	Độ chính xác
1	Lâu đài	0.97	6	Bàn ghế	1.0
2	Ly rượu	1.0	7	Con cá	0.98

3	Thức ăn	1.0	8		Con chim	0.85
4	Con chó	1.0	9		Người phụ nữ	1.0
5	Bông hoa	1.0	10		Con bướm	0.98

Kết quả thử nghiệm cho thấy, với hầu hết các trường hợp ảnh truy vấn đầu vào, hệ thống CBIR tìm được các ảnh trong cơ sở dữ liệu với độ chính xác là 1.00, bao gồm: Ly rượu, thức ăn, con chó, bông hoa, bàn ghế và người phụ nữ. Ảnh truy vấn có độ chính xác thấp nhất là Con chim khi đạt 0.84.

#### 3.2.4.4. Đánh giá trên bộ dữ liệu Oxford-IIIT Pet

Truy vấn với một số ảnh đại diện cho mỗi lớp, ta có thể quả được thể hiện như sau:



Hình 3. 21. Truy vấn ảnh nhóm Chihuahua

ẢNH TRUY VẤN:



KẾT QUẢ:

Precision: 0.75



True



True



True



True



True



True



True



True



True



True

Hình 3. 22. Truy vấn ảnh trong lớp *american\_pit\_bull\_terrier*

Bảng 3.5 tóm tắt kết quả truy vấn với 06 ảnh đầu vào của bộ dữ liệu Oxford, tương ứng với 6 lớp.

Bảng 3. 5. Kết quả truy vấn với 06 ảnh đầu vào của bộ Oxford

STT	Ảnh truy vấn	Độ chính xác	STT	Ảnh truy vấn	Độ chính xác
1	american_pit_bull_terrier	0.75	4	shiba_inu	1.0
2	chihuahua	0.93	5	havanese	1.0
3	persian	0.93	6	basset_hound	0.84

Kết quả thử nghiệm một số trường hợp trên bộ Oxford cho thấy, hệ thống vẫn đạt được khả năng phân loại chính xác cao với một số lớp như shiba\_inu

hay havanese, với 1.0. Tuy nhiên, cũng có một số nhóm có độ chính xác thấp hơn như american\_pit\_bull\_terrier với độ chính xác là 0.75 hay basset\_hound là 0.84. Nhìn chung, các lớp khác nhau trong bộ dữ liệu Oxford có sự tương đồng khá đáng kể, điều này là một phần ảnh hưởng tới khả năng tra cứu của hệ thống.

Kết luận, hệ thống tra cứu ảnh dựa trên phân cụm phổ đã trình bày ở trên cho phép truy vấn các ảnh với độ chính xác cao và tốc độ nhanh. Đồng thời, việc triển khai trên nền Web giúp cho hệ thống có thể dễ dàng phân phối và truy cập bởi số đông người dùng mà ít yêu cầu thêm các cài đặt phức tạp.

### **Kết luận Chương 3:**

Kết quả của Chương 3 gồm hai phần, trong đó, phần thứ nhất là chương trình mô phỏng thuật toán phân cụm phổ, phần thứ hai là chương trình ứng dụng kỹ thuật phân cụm phổ để tra cứu ảnh dựa trên nội dung.

Ở phần thứ nhất, thuật toán phân cụm phổ được minh họa sử dụng ngôn ngữ Python, chạy trên môi trường Colab, sử dụng thuật toán K-means để phân cụm. Bộ dữ liệu được thử nghiệm gồm blods, moons, circle và aniso. Kết quả cho thấy thuật toán hoạt động hiệu quả và các cụm được phân loại rõ ràng, các phần tử thuộc cùng một cụm có các tính chất tương đồng.

Ở phần thứ hai, chương trình ứng dụng kỹ thuật phân cụm phổ để phân loại ảnh dựa trên nội dung, hoạt động trên nền Web. Chương trình cho phép lựa chọn ảnh truy vấn và đưa ra các kết quả là những ảnh có độ tương đồng cao nhất với ảnh truy vấn. Thuật toán được thử nghiệm trên hai bộ dữ liệu là Corel Database và Oxford-IIIT Pet Dataset. Kết quả thử nghiệm trên bộ dữ liệu Corel cho thấy, thuật toán phân cụm phổ có độ chính xác cao, đạt từ 0.84 đến 1.00 trong các truy vấn thử nghiệm. Trên bộ dữ liệu Oxford, độ chính xác đạt được từ 0.75 đến 0.93. Nhìn chung, độ chính xác trên phụ thuộc vào các yếu tố như: Ảnh truy vấn, số lượng cụm và mức độ tương đồng giữa các cụm trong một bộ dữ liệu. Các kết quả đánh giá thể hiện sự khả quan và hoàn toàn có thể được ứng dụng vào các hệ thống tra cứu ảnh dựa trên nội dung trong tương lai.

## KẾT LUẬN

Phương pháp tra cứu ảnh dựa trên nội dung là một phương pháp tra cứu ảnh rất phổ biến. Một trong những ưu điểm của phương pháp tra cứu này, chính là khả năng tra cứu tự động thay vì cách tiếp cận dựa trên từ khóa truyền thống như phương pháp tra cứu ảnh dựa trên văn bản. Đã có rất nhiều đề tài nghiên cứu xung quanh việc tra cứu ảnh dựa trên nội dung, với rất nhiều kỹ thuật tiên tiến cũng như thuật toán được áp dụng và luận văn của cá nhân học viên cũng chỉ là một trong số đó. Thông qua luận văn của mình, học viên đã thực hiện được các việc như sau:

Thứ nhất, tìm hiểu, nghiên cứu và trình bày được những phương pháp tra cứu ảnh phổ biến. Bao gồm các khái niệm, định nghĩa căn bản trong xử lý ảnh nói chung và tra cứu ảnh nói riêng, kiến trúc thành phần của một hệ thống tra cứu ảnh, các chức năng cơ bản của hệ thống tra cứu đồng thời trình bày một số hệ thống tra cứu ảnh dựa trên nội dung đã có.

Thứ hai, tìm hiểu, nghiên cứu và trình bày được một số kỹ thuật hiệu quả để áp dụng vào hệ thống tra cứu ảnh dựa trên nội dung, ví dụ như là các độ đo khoảng cách, cách cách trích chọn đặc trưng như là gói, lọc, nhúng, phương pháp phản hồi liên quan và các kỹ thuật áp dụng.

Thứ ba, nghiên cứu, phân tích bài toán tra cứu ảnh dựa trên phân cụm đồ thị và các thuật toán phân cụm phổ (ví như như K-means). Xây dựng được chương trình mô phỏng kỹ thuật cụm phổ và áp dụng nó để tra cứu ảnh dựa trên nội dung trên các tập dữ liệu sẵn có. Thông qua thực nghiệm thấy được thuật toán đã đem lại độ chính xác cao, do đó có thể áp dụng vào những hệ thống tra cứu ảnh dựa trên nội dung trong tương lai.

Ngoài ra học viên có tham gia vào nhóm nghiên cứu của thầy hướng dẫn và có bài báo được báo cáo tại Hội Thảo Quốc Gia lần thứ XXV - VNICT 2022, với tên là "Nâng cao hiệu quả tra cứu ảnh sử dụng phân hoạch đồ thị."

Do khả năng còn hạn chế cũng như bản thân chưa có nhiều kinh nghiệm, luận văn chắc chắn sẽ còn những điều còn thiếu sót cần cải thiện. Trong tương lai, học viên hy vọng có thể phát triển được các nghiên cứu của mình, như là thử nghiệm chương trình trên các tập dữ liệu lớn và phức tạp hơn, hay như thử sử dụng một số độ đo khoảng cách khác.

## TÀI LIỆU THAM KHẢO

### I. Tài liệu tiếng Việt

- [1] PGS. TS. Đỗ Năng Toàn, *Giáo trình Xử lý ảnh*.
- [2] PGS. TS. Nguyễn Quang Hoan, *Giáo trình Xử lý ảnh*.
- [3] TS. Đặng Thành Trung, *Giáo trình Xử lý ảnh*.
- [33] Đ. T. Thuý Quỳnh, N. H. Quỳnh, P. V. Cảnh, and N. Q. Tạo, “Một Phương Pháp Tra Cứu Ảnh Hiệu Quả Sử Dụng Phân Cụm Phổ Trong Phản Hồi Liên Quan”, tháng 8 năm 2017, DOI:10.15625/vap.2017.00071, *Hội thảo: NGHIÊN CỨU CƠ BẢN VÀ ỨNG DỤNG CÔNG NGHỆ THÔNG TIN*.
- [34] Nguyễn Thị Hường, 2020, *Luận văn "Nghiên cứu phương pháp tra cứu ảnh dựa trên phương pháp phân cụm đồ thị"*, đại học Thái Nguyên.

### II. Tài liệu tiếng Anh

- [4] Mohammed Elmogy, Hazem M El – Bakry, January 2015, Text-based, Content-based, and Semantic-based Image Retrievals: A Survey, *International Journal of Computer and Information Technology (ISSN: 2279 – 0764)*, (Volume 04 – Issue 01).
- [5] Dr T. Santha - Principal Dr GRD College of Science, M. Abhayadev - PhD Research Scholar Dr GRD College of Science, 27<sup>th</sup> March 2015, *Proceedings of the UGC Sponsored National Conference on Advanced Networking and Applications*.
- [6] Digital Image Processing, 2nd ed. www.imageprocessingbook.com © 2002 R. C. Gonzalez & R. E. Woods Chapter 6 Color Image Processing Chapter 6 Color Image.
- [7] Linh Viet Tran, 2003, Efficient Image Retrieval with Statistical Color Descriptors, *Linköping Studies in Science and Technology Dissertation No. 810*.
- [9] Asst. Lec. Wasseem Nahy Ibrahim, Image Processing Lecture.

- [11] Taiki Aoyama, Akira Fukumoto, Kenjiro Shigita, Kenjiro Shigita, Bile pigment in small-bowel water content may reflect bowel habits: a retrospective analysis of a capsule endoscopy imaging series.
- [14] Kranti Burman<sup>1</sup>, Rahul Gedam<sup>2</sup>, 2013, Histogram Based Color Image Authentication By Digital Image Watermark Technique, *International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622*, (Vol. 3).
- [15] Fuhui Long, Hongjiang Zhang, David Dagan Feng Feng (The University of Sydney), *Fundamentals of Content-Based Image Retrieval*.
- [16] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, *An Introduction to Information Retrieval*, Online edition (c) 2009 Cambridge UP.
- [17] Khadidja BELATTAR, Sihem MOSTEFAI, 2013, *5<sup>th</sup> Conference: Computer Science and Information Technology (CSIT)*, CBIR using Relevance Feedback: Comparative Analysis and Major Challenges.
- [18] Nehal M. Varma, Prof. Anamika Choudhary, Evaluation Of Distance Measures In Content Based Image Retrieval, *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*.
- [19] John Eakins, Margaret Graham (University of Northumbria at Newcastle), *Content-based Image Retrieval*.
- [20] Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng, 2003, *FUNDAMENTALS OF CONTENT-BASED IMAGE RETRIEVAL*.
- [21] TS. Lê Thị Kim Nga (Chủ biên), PGS. TS. Đỗ Năng Toàn (Đồng chủ biên), ThS. Phạm Trần Thiện, ThS. Phùng Văn Minh, ThS. Nguyễn Thành Đạt, *Xử lý ảnh số và ứng dụng*, nhà xuất bản Khoa Học Kỹ Thuật.
- [22] Jian Kang Wu, Mohan S. Kankanhalli, Joo-Hwee Lim, Dezhong Hong, *Perspectives on Content-Based Multimedia Systems*.

- [23] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, G. Taubin, The QBIC Project: Querying Images by Content, Using Color, Texture, and Shape.
- [24] *Distributed Multimedia Databases: Techniques and Applications* edited by Timothy K. Shih.
- [25] John R. Smith, Shih Fu Chang, An image and video search engine for the world-wide web.
- [26] Rabia Musheer Aziz, C.K. Verma, Namita Srivastava, 2017, Dimension reduction methods for microarray data: a review.
- [27] Nicholas Pudjihartono, Tayaza Fadason, Andreas W. Kempa-Liehr, Justin M. O'Sullivan, 2022, A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction.
- [28] Yap Bee Wah, Nurain Ibrahim, Hamzah Abdul Hamid, Shuzlina Abdul-Rahman, 2018, Simon Fong, Feature Selection Methods: Case of Filter and Wrapper Approaches for Maximising Classification Accuracy.
- [29] *Multivariate Analysis in Management, Engineering and the Sciences* edited by Leandro Valim de Freitas and Ana Paula Barbosa Rodrigues de Freitas.
- [30] Hà Thị Minh Phương (Trường ĐH Công nghệ Thông tin và Truyền thông Việt-Hàn), Phan Thị Quỳnh Hương (Đại học Đà Nẵng), 2020, Nghiên cứu các kỹ thuật lựa chọn đặc trưng trong tập dữ liệu.
- [31] Ricardo da Silva Torres, Alexandre Xavier Falcão, 2006, Content-Based Image Retrieval: Theory and Applications.
- [32] Khadidja BELATTAR, Sihem MOSTEFAI, CBIR using Relevance Feedback: Comparative Analysis and Major Challenges, *2013 5th International Conference on Computer Science and Information Technology (CSIT)*.



[35] P. R. Arantes, M. D. Polêto, C. Pedebos, and R. Ligabue-Braun, “Making it Rain: Cloud-Based Molecular Simulations for Everyone,” *J. Chem. Inf. Model.*, vol. 61, no. 10, pp. 4852–4856, 2021, doi: 10.1021/acs.jcim.1c00998.

[36] E. Bisong, “Introduction to Scikit-learn,” *Build. Mach. Learn. Deep Learn. Model. Google Cloud Platf.*, pp. 215–229, 2019, doi: 10.1007/978-1-4842-4470-8\_18.

[37] M. Kumagai, K. Komatsu, M. Sato, and H. Kobayashi, “Ising-Based Combinatorial Clustering Using the Kernel Method,” *Proc. - 2021 IEEE 14th Int. Symp. Embed. Multicore/Many-Core Syst. MCSoc 2021*, pp. 197–203, 2021, doi: 10.1109/MCSoc51149.2021.00037.

[38] M. Grinberg, “Flask Web Development: Developing Web Applications with Python,” p. 258, 2014.

[39] H. Qassim, A. Verma, and D. Feinzimer, “Compressed residual-VGG16 CNN model for big data places image recognition,” *2018 IEEE 8th Annu. Comput. Commun. Work. Conf. CCWC 2018*, vol. 2018-Janua, pp. 169–175, 2018, doi: 10.1109/CCWC.2018.8301729.

[40] D. Tao, “The corel database for content based image retrieval,” 2012, [Online]. Available: <https://sites.google.com/site/dctresearch/Home/content-%5Cnbased-image-retrieval>.

[41] C. V. Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, “Cats and dogs,” in *2012 IEEE conference on computer vision and pattern recognition*, 2012, pp. 3498–3505.

### **III. Trang web**

[8] <https://docs.microsoft.com/en-us/windows/win32/wcs/color-spaces>

[10] <https://www.techopedia.com/definition/477/cie-color-model#:~:text=The%20CIE%20color%20model%20is%20a%20mapping%20system%20that%20uses,a%20human%20eye%20can%20perceive.>

[12] <https://www.lifewire.com/what-is-hsv-in-design-1078068>

[13] <https://www.dynamsoft.com/blog/insights/image-processing/image-processing-101-color-models/>

[42] <https://viblo.asia/newest>