BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

ISO 9001:2008

PHẠM XUÂN HINH

LUẬN VĂN THẠC SĨ
NGÀNH HỆ THỐNG THÔNG TIN

Hải Phòng - 2016

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

PHẠM XUÂN HINH

TRA CỨU ẢNH DỰA TRÊN NỘI DUNG SỬ DỤNG NHIỀU ĐẶC TRƯNG VÀ PHẢN HỒI LIÊN QUAN

LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ THÔNG TIN
CHUYÊN NGÀNH: HỆ THỐNG THÔNG TIN
MÃ SỐ: 60 48 01 04

NGƯỜI HƯỚNG DẪN KHOA HỌC:
PGS.TS. Ngô Quốc Tạo

MỤC LỤC

LỜI CẢM ƠN

Trong quá trình học tập và thực hiện luận văn, tôi đã được các Thầy cô trường Đại học Dân lập Hải Phòng, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tạo mọi điều kiện thuận lợi, đồng nghiệp và bạn bè đã thường xuyên động viên. Tôi xin bày tỏ sự cảm ơn chân thành với những sự hỗ trợ và giúp đỡ này.

Luận văn sẽ không thể hoàn thành nếu không có sự hướng dẫn tận tình của Thầy hướng dẫn khoa học PGS.TS Ngô Quốc Tạo - Trưởng phòng nhận dạng và Công nghệ tri thức- Viện Hàn lâm Khoa học và Công nghệ Việt Nam là người thầy mà tôi muốn bày tỏ lòng biết ơn sâu sắc nhất.

Xin chân thành cảm ơn Thầy giáo - Ths Ngô Trường Giang - Phó trưởng khoa CNTT trường Đại học Dân Lập Hải Phòng đã có nhiều ý kiến đóng góp, giúp đỡ quan trọng trong quá trình thực hiện luận văn.

Xin chân thành cảm ơn Ban giám hiệu, GS.TS.NGƯT Trần Hữu Nghị Hiệu trưởng nhà trường và tập thể Thầy Cô trong khoa Công Nghệ Thông Tin- Trường Đại Học Dân Lập Hải Phòng đã quan tâm tạo môi trường thuận lợi để học tập và nghiên cứu chuyên sâu về lĩnh vực Công nghệ thông tin.

Cuối cùng tôi cảm ơn tất cả những sự giúp đỡ của đồng nghiệp, bạn bè đã đóng góp ý kiến, động viên để tôi hoàn thành được luận văn này.

LỜI CAM ĐOAN

Tên tôi là: Phạm Xuân Hinh

Lớp: Cao học Công nghệ thông tin Khóa 1

Khóa học: *2014-2016*

Chuyên ngành: Hệ thống thông tin

Mã số chuyên ngành: *60 48 01 04*

Cơ sở đào tạo: Trường Đại học Dân Lập Hải Phòng

Người hướng dẫn khoa học: *PGS.TS Ngô Quốc Tạo*

Tôi xin cam đoan toàn bộ nội dung trình bày trong luận văn này là kết quả tìm hiểu và nghiên cứu của bản thân. Các số liệu, kết quả trình bày trong luận văn là hoàn toàn trung thực. Những tư liệu được sử dụng trong luận văn đều được tuân thủ theo luật sở hữu trí tuệ, có liệt kê rõ ràng các tài liệu tham khảo.

Tôi xin chịu hoàn toàn trách nhiệm với những nội dung viết trong luận văn này!

Hải Phòng, ngày 01 tháng 12 năm 2016

Tác giả luận văn

Phạm Xuân Hinh

DANH MỤC CHỮ VIẾT TẮT

Stt	Từ viết tắt	Diễn giải
1	CBIR	Content-Based Image Retrieval
2	RF	Relevance Feedback
3	ST	Semantic Template
4	RGB	Red-Green-Blue
5	SVM	Support Vector Machine
6	SVT	Semantic Visual Template
7	PCA	Principal Component Analysis
8	KL	Karhunen-Loeve
9	CSDL	Cơ sở dữ liệu
10	CCV	Color Coherence Vector
11	SIFT	Scale Invariant Feature Transform
12	PCA	Principal Component Analysis

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

MỞ ĐẦU

Những năm gần đây, với sự xuất hiện của Internet đã thay đổi hoàn toàn cách thức chúng ta tìm kiếm thông tin. Ví dụ khi cần tìm kiếm, đơn giản chỉ cần gõ một vài từ khóa vào máy tìm kiếm Google hay Bing, ngay lập lức có được một danh sách tương đối chính xác các trang web có liên quan đến thông tin cần tìm. Đối với hình ảnh, cũng đã có các hệ thống tương tự. Với hệ thống này, bằng cách lấy một ảnh đầu vào từ người dùng, hệ thống cố gắng tìm kiếm các ảnh giống nhất trong cơ sở dữ liệu rồi trả lại cho người sử dụng. Đây là hệ thống tra cứu ảnh theo nội dung hay đơn giản là tra cứu ảnh. Về cơ bản, hệ thống hoạt động theo cách thức sau: Đầu tiên ảnh đưa vào để tìm kiếm (hay gọi là ảnh truy vấn) và toàn bộ ảnh trong CSDL được hệ thống ánh xạ sang các vector (đặc trưng của ảnh). Hệ thống sẽ tính toán và đo khoảng cách giữa ảnh truy vấn với từng ảnh trong CSDL. Cuối cùng, các ảnh có khoảng cách gần nhất với ảnh truy vấn được hệ thống trả về. Tuy nhiên kết quả trả về vẫn còn xa so với sự mong đợi của người dùng. Ta thường gọi vấn đề này là vấn đề "*khoảng cách ngữ nghĩa*".

Để thu hẹp được khoảng cách ngữ nghĩa, nâng cao hiệu quả tra cứu, phương pháp phản hồi liên quan đã được gới thiệu trong CBIR[4]. Đã có nhiều nhà nghiên cứu bắt đầu xem phản hồi liên quan như là bài toán phân lớp hoặc bài toán học. Việc kết hợp nhiều đặc trưng để xây dựng truy vấn đã góp phần nâng cao hiệu quả của các phương pháp học máy, do vậy hiệu quả tra cứu đã được cải thiện. Tuy nhiên, để tận dụng đầy đủ lợi thế của các thông tin bổ sung, phát sinh từ tương tác người dùng, việc lựa chọn phương pháp kết hợp sử dụng nhiều đặc trưng hiệu quả là nhiệm vụ quan trọng và rất cần thiết. Đó cũng là lý do mà tôi chọn đề tài "*Tra cứu ảnh dựa trên nội dung sử dụng nhiều đặc trưng và phản hồi liên quan* ".

Nội dung luận văn gồm 3 chương:

Chương 1. KHÁI QUÁT VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

Chương này trình bày khái quát lý thuyết cơ bản về tra cứu ảnh dựa trên nội dung, tìm hiểu một số phương pháp trích chọn đặc trưng ảnh và tìm hiểu một số hệ thống tra cứu ảnh sẵn có.

Chương 2. KẾT HỢP NHIỀU ĐẶC TRƯNG TRONG TRA CỨU ẢNH SỬ DỤNG SVM VÀ PHẢN HỒI LIÊN QUAN

Chương này tìm hiểu một số kỹ thuật phản hồi liên quan trong tra cứu ảnh dựa trên nội dung, tìm hiểu các kỹ thuật kết hợp các đặc trưng hình ảnh trong trong CBIR.

Chương 3. THỰC NGHIỆM

Xây dựng chương trình thực nghiệm tra cứu ảnh theo nội dung kết hợp nhiều đặc trưng với phản hồi liên quan, đánh giá hiệu năng và một số kết quả đạt được.

Chương 1. KHÁI QUÁT VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

1.1 Giới thiệu tra cứu ảnh dựa trên nội dung

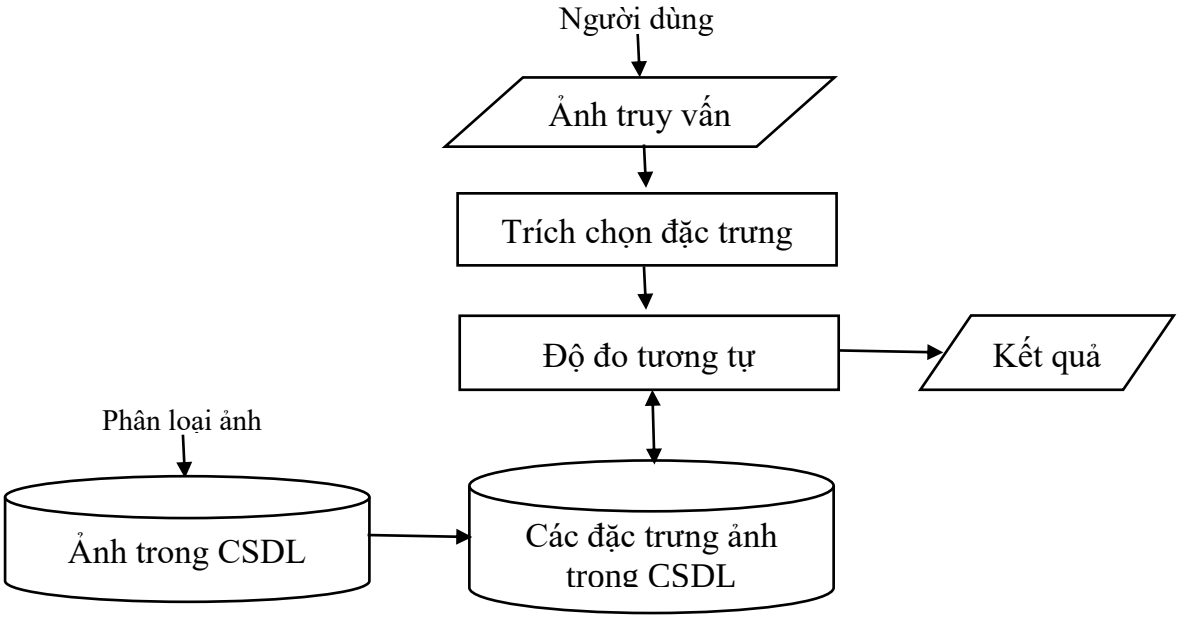
Thuật ngữ "*Tra cứu thông tin*" được đưa ra vào năm 1952 và đã giành được sự quan tâm đặc biệt của hội các nhà nghiên cứu từ năm 1961 [*Jones and Willet, 1977*]. Chúng ta có thể dễ dàng mô tả một hệ thống đó như là một hệ thống lưu trữ và tra cứu thông tin. Vì vậy nó gồm một tập hợp các thành phần tương tác lẫn nhau, mỗi thành phần được thiết kế cho một chức năng riêng, có mục đích riêng và tất cả các thành phần này có quan hệ với nhau để đạt được mục đích là tìm kiếm thông tin trong một phạm vi nào đó.

Trước đây, tra cứu thông tin hình ảnh là người ta nghĩ đến tra cứu thông tin theo kết cấu, nhưng định nghĩa trên vẫn được giữ khi ứng dụng vào việc tra cứu thông tin thị giác (*Visual Infomation Retrieval*). Mặc dù vậy vẫn có sự phân biệt giữa kiểu của thông tin và nét tự nhiên của tra cứu các đối tượng trực quan. Thông tin kết cấu là tuyến tính trong khi ảnh là hai chiều và video là ba chiều.

Có hai phương pháp để giải bài toán tra cứu thông tin thị giác dựa trên những thông tin trực quan đó là: Phương pháp dựa trên những thuộc tính và phương pháp dựa trên những đặc điểm. Phương pháp dựa trên thuộc tính là tra cứu dựa vào thông tin kết cấu truyền thống và những phương pháp quản lý cơ sở dữ liệu dựa trên lý trí cũng như là sự can thiệp của con người để trích chọn dữ liệu về đối tượng trực quan và sự chú thích kết cấu. Việc chú thích về đối tượng đều mất nhiều thời gian và tốn nhiều công sức. Hơn nữa lời chú thích phụ thuộc rất nhiều vào cảm nhận chủ quan của con người, mà sự cảm nhận chủ quan và sự giải thích mơ hồ chính là nguyên nhân của sự ghép đôi không cân xứng trong quá trình xử lý. Vấn đề tìm kiếm ảnh và video dựa trên lời chú thích đã thúc đẩy đến sự quan tâm, phát triển những giải pháp dựa trên

đặc điểm. Đó là thay sự giải thích thủ công bằng những từ khoá dựa trên văn bản, ảnh có thể được trích chọn ra bằng cách sử dụng một số đặc điểm thị giác như là màu sắc, kết cấu, hình dạng… và được đánh chỉ số dựa trên những đặc điểm thị giác này. Phương pháp này được gọi là tra cứu ảnh dựa trên nội dung CBIR [4]. Cách thức tìm kiếm ảnh của CBIR là việc trích chọn các đặc trưng được thực hiện một cách tự động và nội dung của ảnh luôn luôn nhất quán.

1.2 Các thành phần của hệ thống CBIR



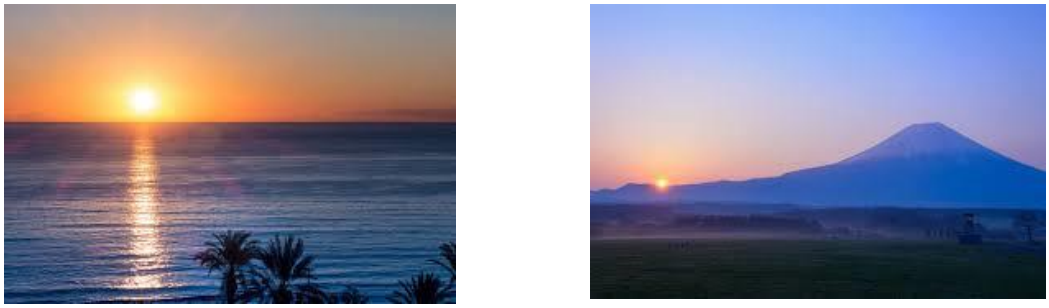
Hình 1.1. Kiến trúc tổng quan về hệ thống tra cứu ảnh

1.2.1 Trích chọn đặc trưng

Các đặc trưng của hình ảnh bao gồm các đặc trưng nguyên thủy và các đặc trưng ngữ nghĩa hoặc đặc trưng logic. Các đặc trưng cơ bản đó là: màu sắc (*color*), kết cấu (*texture*), hình dạng (*shape*), vị trí không gian (*spatial location*),… được định lượng trong tự nhiên, chúng có thể được trích xuất tự động hoặc bán tự động. Đặc trưng logic cung cấp mô tả trừu tượng của dữ liệu hình ảnh ở các cấp độ khác nhau. Thông thường, một hoặc nhiều đặc trưng có thể được sử dụng trong từng ứng dụng cụ thể trên thực tế.

1.2.2 Đo độ tương tự giữa các ảnh

Hệ thống CBIR dựa trên những đặc điểm nguyên thủy để so sánh độ tương tự giữa ảnh truy vấn và tất cả các ảnh trong CSDL. Mặc dù vậy sự tương tự hoặc sự khác nhau giữa các ảnh không chỉ xác định theo một cách. Số lượng của ảnh tương tự sẽ thay đổi khi yêu cầu truy vấn thay đổi. Chẳng hạn trong trường hợp hai hình ảnh, một là biển xanh mặt trời mọc và trường hợp khác là núi xanh với mặt trời mọc.



Hình 1.2. Hình ảnh minh họa độ tương tự giữa 2 hình ảnh

Khi mặt trời được xem xét thì độ tương tự giữa hai ảnh này là cao nhưng nếu đối tượng quan tâm là biển xanh thì độ tương tự giữa hai ảnh này là thấp. Như vậy rất khó khăn để tìm ra phương pháp đo độ tương tự giữa hai hình ảnh trên một cách chính xác đối với tất cả các kiểu yêu cầu của truy vấn. Hay nói cách khác mỗi một phương pháp tra cứu sẽ có giới hạn của chính nó. Ví dụ rất khó cho công nghệ tra cứu dựa trên màu sắc để tìm ra điểm khác nhau giữa một ảnh là bầu trời màu xanh với một ảnh là mặt biển xanh. Vì vậy khi đánh giá một phương pháp tra cứu ảnh dựa trên nội dung cần phải biết rằng hiệu quả của công nghệ đó phụ thuộc vào kiểu yêu cầu tra cứu mà người dùng sử dụng.

1.2.3 Đánh chỉ số

Đánh chỉ số là một công việc quan trọng trong tra cứu ảnh dựa trên nội dung, nó giúp tìm kiếm nhanh ảnh dựa trên đặc trưng trực quan, bởi vì các vector đặc trưng của ảnh có xu hướng, có số chiều cao và vì vậy nó không

thích hợp cho các cấu trúc đánh chỉ số truyền thống. Do đó trước khi lên kế hoạch đánh chỉ số ta phải tìm cách làm giảm số chiều của các vector đặc trưng.

Có nhiều phương pháp làm giảm số chiều của vector đặc trưng, một trong những công nghệ được sử dụng phổ biến là phân tích thành phần chính PCA. Nó là một công nghệ tối ưu trong việc ánh xạ tuyến tính dữ liệu đầu vào một không gian toạ độ, các trục được thẳng hàng để phản ánh các biến thể lớn nhất trong dữ liệu. Hệ thống QBIC sử dụng PCA để làm giảm số chiều của vector đặc trưng hình dạng từ nhiều chiều thành hai hoặc ba chiều. Ngoài phương pháp PCA ra, nhiều nhà nghiên cứu còn sử dụng biến đổi KL để làm giảm số chiều trong không gian đặc trưng. Ngoài hai phương pháp biến đổi PCA và KL, thì mạng nơ ron cũng là công cụ hữu ích cho việc giảm số chiều đặc trưng.

Khi đã giảm được số chiều thì dữ liệu đa chiều được đánh chỉ số. Có nhiều phương pháp đánh chỉ số bao gồm : *K-D-B tree, R-tree, linear quad-trees,...* các phương pháp này đều cho hiệu quả hợp lý với không gian có số chiều nhỏ.

1.2.4 Giao diện truy vấn (Query Interface)

Để biểu diễn ảnh tra cứu từ CSDL cho người dùng thì có rất nhiều cách. Và những cách thông thường nhất được sử dụng là: Duyệt qua mục; truy vấn bởi khái niệm; truy vấn bởi bản phác thảo và truy vấn bởi ví dụ,...

- Duyệt qua mục là phương pháp duyệt qua toàn bộ CSDL theo danh mục các ảnh. Mục đích của phương pháp này là ảnh trong CSDL được phân loại thành nhiều mục khác nhau theo ngữ nghĩa hoặc nội dung trực quan.
- Truy vấn bởi khái niệm là tra cứu ảnh theo mô tả khái niệm liên quan với từng ảnh trong CSDL [4] .

- Truy vấn bởi bản phác thảo và truy vấn bởi ví dụ là vẽ ra một bản phác thảo hoặc cung cấp một ảnh ví dụ từ những ảnh với độ tương tự đặc trưng trực quan sẽ được trích chọn từ CSDL.

Trong số các phương pháp trên thì phương pháp thì truy vấn bởi bản phác thảo hoặc bởi ví dụ là phương pháp quan trọng và khó khăn nhất. Phần lớn các nghiên cứu tra cứu ảnh dựa trên nội dung tập trung đi sâu vào phương pháp này.

1.3 Một số phương pháp trích chọn đặc trưng

Các đặc trưng cơ bản của hình ảnh bao gồm: màu sắc (*color*), kết cấu (*texture*), hình dạng (*shape*), vị trí không gian (*spatial location*),… được định lượng trong tự nhiên, chúng có thể được trích xuất tự động hoặc bán tự động. Dưới đây sẽ giới thiệu một số phương pháp trích chọn đặc trưng hình ảnh.

1.3.1 Trích chọn đặc trưng màu sắc

Hình ảnh bao gồm một mảng các điểm ảnh (*pixel*), và mỗi pixel thể hiện một màu sắc. Có nhiều không gian màu được sử dụng để tính toán các giá trị màu của pixel như: không gian chuẩn RGB, không gian trực giác HSV... Các đặc trưng được lưu giữ dưới dạng các vector biểu diễn cho các thông tin mô tả nội dung ảnh.

Lược đồ màu (*Histogram*) là đại lượng đặc trưng cho phân bố màu cục bộ của ảnh. Được định lượng:

$$H(I_D, C_i) = \frac{m(I_D, C_i)}{n(I_D)} \tag{1.1}$$

trong đó:

C_i : là màu của điểm ảnh

$n(I_D)$: tổng số điểm ảnh trong ảnh.

$m(I_D, C_i)$: Biểu diễn số điểm ảnh có giá trị màu C_i

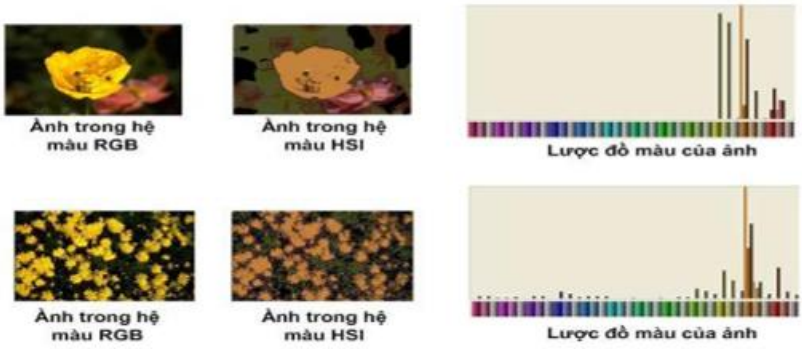
H : lược đồ màu của ảnh.

Độ đo tính tương tự về màu sắc giữa lược đồ màu của ảnh truy vấn $H(I_Q)$ và lược đồ màu của ảnh trong CSDL ảnh $H(I_D)$ được định nghĩa:

$$D_H(I_Q, I_D) = \frac{\sum_{j=1}^{M} \min(H(I_Q, j), H(I_D, j))}{\sum_{j=1}^{M} H(I_D, j)} \quad (1.2)$$

Công thức (1.2) cho ta thấy, tính tương tự về màu sắc được tính bằng phần giao của 2 lược đồ màu ảnh truy vấn $H(I_Q)$ và ảnh trong cơ sở dữ liệu ảnh $H(I_D)$. Kết quả sẽ là một lược đồ màu thể hiện độ giống nhau giữa 2 ảnh trên.

Tuy nhiên vì lược đồ màu chỉ thể hiện tính phân bố màu toàn cục của ảnh mà không xét đến tính phân bố cục bộ của điểm ảnh nên có thể có 2 ảnh trông rất khác nhau nhưng lại có cùng lược đồ màu.



Hình 1.3. Hình minh họa 2 ảnh có lược đồ giống nhau đến 70% nhưng khác nhau về ngữ nghĩa

Để khắc phục được tình trạng này, chúng ta dùng phân hoạch lưới ô vuông trên ảnh. Lược đồ màu của ảnh là không duy nhất.

1.3.1.1 Vector liên kết màu

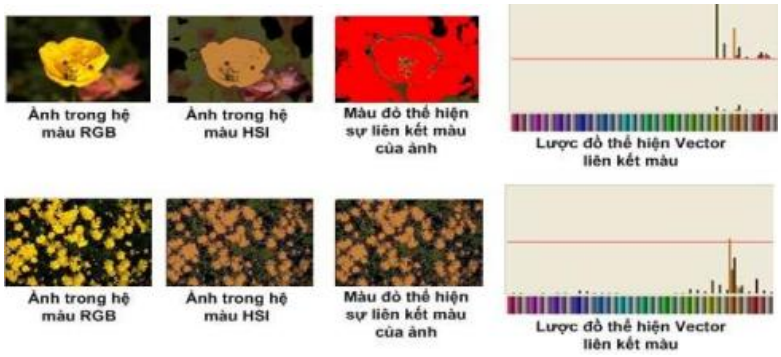
Vector liên kết màu (CCV) [5] là lược đồ tinh chế lược đồ màu, chia mỗi ô màu (bin) thành 2 nhóm điểm ảnh: Nhóm liên kết màu (coherence pixels) và nhóm không liên kết màu (non-coherence pixels).

Một pixel trong 1 ô màu (bin) được gọi là điểm liên kết màu (coherent) nếu nó thuộc vùng gồm các màu tương tự với kích thước lớn (thường bằng khoảng 1% kích thước ảnh). Với mỗi ô màu (bin) giả sử số điểm liên kết màu là α và số điểm không liên kết màu là β thì vector liên kết màu được xác định:

$V_c = [(\alpha_1, \beta_1), (\alpha_2, \beta_2), ..., (\alpha_n, \beta_n)]$, n là số ô màu (bin)

Trong tìm kiếm ảnh với việc sử dụng đặc trưng vectơ liên kết màu sẽ giúp ta tránh được tình trạng hai ảnh có cùng lược đồ màu nhưng khác nhau hoàn toàn về ngữ nghĩa.

Ngoài ra vector liên kết màu còn giúp giải quyết khuyết điểm về tính không duy nhất của lược đồ màu đối với ảnh. Hai ảnh có thể có chung lược đồ màu nhưng khác nhau hoàn toàn, đây là khuyết điểm của lược đồ màu. Nhưng với truy vấn theo đặc trưng vector liên kết màu thì nó sẽ giải quyết được khuyết điểm không duy nhất này



Hình 1.4 Hình minh họa vector liên kết mầu

1.3.1.2 Tương quan màu (*Correlogram*)

Như đã giới thiệu ở trên, lược đồ màu chỉ ghi nhận được sự phân bố màu trong ảnh mà không chứa các thông tin mối quan hệ về khoảng cách. Để khắc phục hạn chế đó, đặc trưng tương quan màu biểu diễn sự thay đổi mối quan hệ về không gian giữa các cặp màu theo khoảng cách.

Cũng giống như đặc trưng vectơ liên kết màu, đặc trưng tương quan màu thể hiện mối quan hệ chặt chẽ về sự phân bố màu trong ảnh. Chính vì vậy nếu truy tìm ảnh sử dụng đặc trưng này cũng tránh được tình trạng mà đặc trưng lược đồ màu vấp phải

So sánh với lược đồ màu và vector gắn kết màu, tương quan màu cho các kết quả tra cứu tốt hơn. Tuy nhiên, tương quan màu có độ phức tạp tính toán cao, do vector đặc trưng có số chiều cao.

1.3.1.3 Các màu trội

Các lược đồ màu thường rất thưa và thông thường chỉ cần số ít màu là đủ để miêu tả đặc trưng màu trong một ảnh màu, các màu trội [3, 10] được sử dụng để mô tả đặc trưng màu của một ảnh. Phân cụm màu được thực hiện để thu các màu trội đại diện và phần trăm tương ứng của nó. Mỗi màu đại diện và phần trăm tương ứng này tạo ra một cặp các thuộc tính mô tả các đặc trưng màu trong một vùng ảnh.

Ký hiệu mô tả đặc trưng lược đồ màu trội F được xác định bởi một tập các cặp thuộc tính:

$$F = \{\{c_i, p_i\}, i = 1,..., N\} \quad (1.3)$$

Ở đây N là tổng số các cụm màu trong ảnh, C_i là một vector màu ba chiều, p_i là phần trăm của nó, và $\sum_i p_i = 1$. Tuy nhiên, phương pháp này cũng cho kết quả tra cứu không cao khi cơ sở dữ liệu ảnh có kích thước lớn, do nó chỉ biểu thị phân bố xác suất của các màu trội trong ảnh.

1.3.1.4 Mô men màu

Mô men màu là các mô men thống kê của các phân bố xác suất của các màu. Các mô men màu được sử dụng trong nhiều hệ thống tra cứu ảnh như QBIC [11]. Các mô men màu bậc nhất (*trung bình*), bậc hai (*phương sai*) và bậc ba (*độ lệch*), đã được minh chứng là hiệu quả trong biểu diễn các phân bố màu của các ảnh.

Về mặt toán học, ba mô men đầu tiên được xác định bằng:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \tag{1.4}$$

$$\sigma_i = \left(\frac{1}{N}\sum (f_{ij} - \mu_i)^2\right)^{\frac{1}{2}} \tag{1.5}$$

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N} (f_{ij} - \mu_i)^3\right)^{\frac{1}{3}} \tag{1.6}$$

Ở đây f_{ij} là giá trị của thành phần màu thứ i của điểm ảnh j và N là số các điểm ảnh trong ảnh.

Do chỉ số (ba mô men cho một trong ba thành phần màu) được sử dụng để biểu diễn đặc trưng màu của mỗi ảnh, các mô men màu là một biểu diễn rất nén so với các đặc trưng màu khác. Do biểu diễn rất nén này, các mô men màu có thể làm giảm khả năng phân biệt các ảnh. Thông thường, các mô men màu có thể được sử dụng như sơ duyệt lần đầu để giảm không gian tra cứu trước khi các đặc trưng màu phức tạp khác được sử dụng.

1.3.1.5 Thông tin không gian

Các vùng hoặc đối tượng với các đặc trưng màu và kết cấu tương tự có thể được phân biệt tốt hơn bằng việc kết hợp các thông tin không gian. Chẳng hạn, các vùng bầu trời màu xanh và biển xanh có thể có các lược đồ màu tương tự, nhưng thông tin không gian của chúng trong các ảnh là khác nhau.

Do đó, thông tin không gian của các vùng (hoặc các đối tượng) hoặc quan hệ không gian giữa nhiều vùng (hoặc đối tượng) trong một ảnh rất quan trọng cho tra cứu các ảnh.

Thu nhận thông tin không gian của các đối tượng trong một ảnh là một quá trình quan trọng trong phân biệt các ảnh. Quá trình này bao gồm việc biểu diễn vị trí không gian tuyệt đối và vị trí không gian tương đối của các đối tượng. Bố cục màu kết hợp thông tin không gian với đặc trưng màu trong ảnh tạo ra một đặc trưng rất quan trọng trong quá trình tra cứu.

Trong [2] đã đề xuất kỹ thuật sử dụng lược đồ hình quạt. Tác giả đã đề xuất một cách tiếp cận dựa vào lược đồ màu có đưa thông tin không gian vào bản miêu tả ảnh. Ban đầu ảnh được lượng hóa thành *n* màu và sau đó ảnh được chia thành các khối hình quạt và tính toán lược đồ của mỗi màu. Các điểm ảnh tuy có cùng màu, song chúng được phân vào các dải khác nhau tùy thuộc vào điểm ảnh thuộc khối hình quạt nào.

1.3.2 Trích chọn đặc trưng kết cấu (texture)

Kết cấu (*texture*) hay còn gọi là vân, là một đối tượng dùng để phân hoạch ảnh ra thành những vùng được quan tâm và để phân lớp những vùng đó. Vân cung cấp thông tin sự sắp xếp về mặt không gian của màu sắc và cường độ của một ảnh. Vân được đặc trưng bởi sự phân bổ không gian của những mức cường độ trong một khu vực láng giềng với nhau. Vân của ảnh màu và vân đối với ảnh xám là như nhau. Vân gồm nhiều vân gốc hay vân phần tử gộp lại, đôi khi được gọi là texel. Xét về vấn đề phân tích vân, có hai đặc trưng chính yếu nhất:

Cấu trúc vân: là tập hợp những texel được sắp xếp theo một số quy luật nhất định hay có cấu trúc không gian lặp đi lặp lại.

Sự thống kê vân được định nghĩa như sau: là một độ đo về số lượng của sự sắp xếp những mức xám hay cường độ sáng trong vùng. Một vân bất kỳ có

thể coi như là một tập của những texel thô trong một quan hệ không gian đặc biệt nào đó. Một cấu trúc không gian của một vân bất kỳ sau đó có thể bao gồm một sự mô tả của texel và một đặc tả về không gian. Những texel đương nhiên phải được phân đoạn và quan hệ không gian phải được tính toán một cách thật hiệu quả. Texel là những vùng ảnh có thể trích rút từ một số hàm phân ngưỡng đơn giản. Đặc điểm quan hệ không gian của chúng có thể miêu tả như sau: Giả sử rằng chúng ta có tập những texel, với mỗi phần tử của tập hợp này ta có thể đặc trưng bởi một điểm ý nghĩa nhất, điểm này gọi là trọng tâm. Đặt S là tập của những điểm này. Với mỗi cặp điểm P và Q trong tập S, ta có thể xây dựng đường phân giác trực giao nối chúng lại với nhau. Đường phân giác trực giao này chia mặt phẳng thành hai nửa mặt phẳng, một trong chúng là tập của những điểm gần với P hơn và cái còn lại là tập những điểm gần với Q hơn. Đặt HQ(P) là nửa mặt phẳng gần P hơn. Ta có thể lặp lại quá trình này với mỗi điểm Q trong S. Đa giác Voronoi của P là vùng đa giác bao gồm tất cả những điểm gần P hơn những điểm khác của S và được định nghĩa: $V(P) = \bigcap_{Q \in S, Q \neq S} H^Q(P)$.

Các đặc trưng kết cấu có xu hướng ghi nhận các "hoa văn" dạng hạt, vân,…của những vùng cục bộ (local pattern) trong ảnh. Ví dụ, mặt sân cỏ, tường gạch, vân gỗ, vân đá,… là những dạng texture khác nhau. Tùy theo cơ sở dữ liệu ảnh như: ảnh không gian, ảnh y tế,… hệ thống truy vấn sử dụng các đặc trưng texture có tính chất, đặc thù riêng để đạt hiệu quả truy vấn cao nhất xét về độ chính xác, thời gian xử lý.

Các đặc trưng texture đã được nghiên cứu một thời gian dài trong các lĩnh vực như: xử lý ảnh, computer vision, đồ họa máy tính (computer graphic). Có rất nhiều giải pháp trích đặc trưng texture của ảnh đã được công bố và có thể phân loại thành hai dạng trích đặc trưng texture: trong miền không gian và trong miền biến đổi của ảnh

Ví dụ cấu trúc của vân của một số loại lá cây:

Hình 1.5. Cấu trúc vân của lá cây

1.3.2.1 Ma trận đồng hiện mức xám (Co-occurence Matrix)

Ma trận đồng hiện mức xám là ma trận lưu trữ số lần xuất hiện của những cặp điểm ảnh trên một vùng đang xét. Các cặp điểm này được tính theo những quy luật cho trước. Ví dụ với ảnh *f* như sau:

$$f=\begin{bmatrix}1 & 1 & 0 & 0\\ 1 & 1 & 0 & 0\\ 0 & 0 & 2 & 2\\ 0 & 0 & 2 & 2\end{bmatrix}$$ ta có ma trận đồng hiện mức xám $P(1,0)$, với $P(1,0)=\begin{bmatrix}4 & 0 & 2\\ 2 & 2 & 0\\ 0 & 0 & 2\end{bmatrix}$

(lưu ý là có rất nhiều ma trận đồng hiện mức xám khác nhau cho một ma trận ban đầu)

Ma trận đồng hiện mức xám trên tạo ra bởi những cặp điểm lệch nhau (*1,0*) nghĩa là 2 điểm kế nhau trên cùng hàng. Giá trị tại dòng 0, cột 0 của ma trận đồng hiện trên là 4 vì ảnh *f* có 4 cặp điểm 0 0 kế nhau trên cùng một hàng. Tương tự như vậy, giá trị ở dòng 1, cột 2 của ma trận là 0 vì không có cặp 1 2 nào xuất hiện nhau trên cùng một hàng.

Công thức tổng quát của ma trận đồng hiện mức xám là:

$$C_t[i,j]=|\{[r,c]\,|\,f(r,c)=i \text{ và } f(r+t_x,c+t_y)=j\}|$$

trong đó $t=(t_x,t_y)$

Ví dụ với ma trận *f* đã cho như trên thì khi *t*=(*1,0*) ta sẽ có ma trận đồng hiện như ví dụ trên, và khi *t*=(*1,1*), nghĩa là tìm những cặp điểm kế nhau trên cùng một đường chéo, ta có ma trận đồng hiện là:

$$P_{(1,1)} = \begin{bmatrix} 2 & 0 & 2 \\ 2 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Từ ma trận đồng hiện mức xám người ta định nghĩa ra các đặc trưng về vân như sau:

Energy (năng lượng):

$$\sum_i \sum_j P_t^2[i,j]$$

Entropy:

$$\sum_i \sum_j P_t[i,j] \log_2(P_t[i,j])$$

Maximan Probability:

$$\max_{(i,j)} P_t[i,j]$$

Contrast (thông thường k=2 và l=1):

$$\sum_i \sum_j [i-j]^k \; P_t^l[i,j]$$

Inverset difference moment:

$$\sum_i \sum_j \frac{P_t^l[i,j]}{|i-j|^k}, i \neq j$$

Correlation:

$$\sum_i \sum_j \frac{(i-u_i)(j-u_i)P_t[i,j]}{a_i a_j}$$

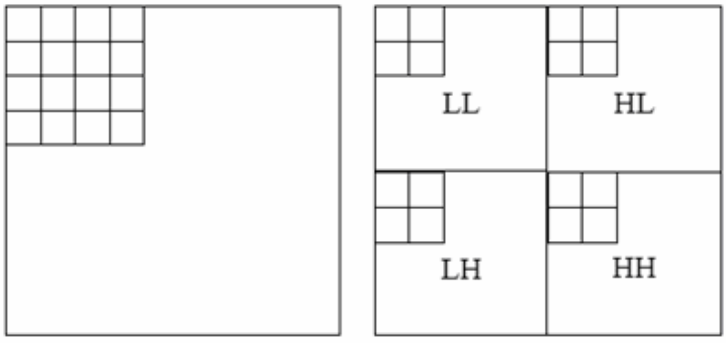
$$u_i = \sum_{i=1} (i \sum_j P_t[i,j], \; u_j = \sum_{j=1} (j \sum_i P_t[i,j]$$

$$a_i = \sum_{i=1} (i - u_i)^2 \sum_j P_t[i, j]$$

$$a_j = \sum_{j=1} (j - u_j)^2 \sum_i P_t[i, j]$$

1.3.2.2 Phép biến đổi Wavelet

Vân thu được từ phép biến đổi wavelet được hầu hết các nghiên cứu công nhận là đặc trưng tốt nhất cho việc phân đoạn ảnh. Từ một vùng kích thước n × n ta có thể thu được một vector có 3 thành phần đặc trưng cho texture với biến đối wavelet ở mức 1. Để có được 3 thành phần này, chúng ta áp dụng biến đổi wavelet Daubechies-4 hoặc bộ lọc Haar với thành phần L của ảnh. Sau khi áp dụng 1 mức biến đổi, chúng ta sẽ có 4 miền tần số (*frequency band*) thì khi đó một thành phần vector sẽ được tính bằng giá trị trung bình của vùng trên miền tần số tương ứng ấy. Ví dụ, ta xét trên vùng 4×4, thông qua biến đổi Daubechies-4, ta có 4 miền tần số là LL, HL, LH, HH như ở hình Hình 1.6, từ 4 miền đó, ta có được 3 thành phần tương ứng với giá trị ở các miền HL, LH và HH.



Hình 1.6. Decompostion để tạo ra các frequency bands bởi biến đổi Wavelet

Như vậy với một hình có kích thước 4×4 như trong ví dụ trên thì thành phần ứng với HL (giả sử HL bao gồm $C_{k,l}$,C_{k+1}, C_{l+1}, $C_{k,l,l+1}$) sẽ được tính:

$$f = (\frac{1}{4} \sum_0^1 \sum_0^1 c^2 + i,\ i + j)^{\frac{1}{2}} \qquad (1.8)$$

Tính toán tương tự cho các vùng LH, HH:

Thuật toán tính ra các đặc trưng vân theo biến đổi Wavelet:

- Tính biến đổi Wavelet trên toàn ảnh.
- Ứng với mỗi vùng cần tính, ta tính được 3 thành phần ứng với các miền HL, LH và HH
- Khi áp dụng biến đổi wavelet ở những mức sâu hơn, ta sẽ có tương ứng $3\times$ V thành phần ứng với V là chiều sâu của biến đổi Wavelet.

Lưu ý: Một cải tiến khác sẽ đem lại hiệu quả rất nhiều cho việc phân đoạn là áp dụng DWF (*Discrete Wavelet Frames*). Cách thức trên được khá nhiều nghiên cứu khác đã vận dụng và thành công.

1.3.2.3 Các đặc trưng Tamura

Các đặc trưng Tamura, bao gồm thô, độ tương phản, hướng, giống nhất, tính chất đều và nhám, được thiết kế phù hợp với các nghiên cứu tâm lý về nhận thức của người đối với kết cấu. Trong đó, thô, độ tương phản, hướng được sử dụng trong một số hệ thống tra cứu ảnh nổi tiếng như QBIC và Photobook.

1.3.2.3.1 Thô (Coarseness)

Thô là một độ đo tính chất hột của kết cấu. Để tính toán thô, các trung bình động $A_k(x,y)$ được tính đầu tiên sử dụng cỡ $2k\times 2k\,(k=0,1,...,5)$ tại mỗi pixel (x,y) ta có:

$$A_k(x,y)=\sum_{i=x-2^{k-1}}^{x+2^{k-1}-1}\sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i,j)/2^{2k} \qquad (1.9)$$

Trong đó, $g(i,j)$ là cường độ pixel tại (i, j)

Sự khác nhau giữa các cặp trung bình động không theo hướng ngang và đứng cho mỗi pixel được tính toán đó là:

$$E_{k,h}(x,y) = \left| A_k\left(x + 2^{k-1}, y\right) - A_k\left(x - 2^{k-1}, y\right) \right| \quad (1.10)$$

$$E_{k,v}(x,y) = \left| A_k\left(x, y + 2^{k-1}\right) - A_k\left(x, y - 2^{k-1}\right) \right| \quad (1.11)$$

Giá trị của *k* cực đại hoá *E* theo một trong hai hướng được sử dụng để đặt cỡ tốt nhất cho mỗi pixel đó là: $S_{best}(x,y) = 2^k$. Thô được tính bằng *Sbest* trên toàn bộ ảnh đó là:

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} S_{best}(i,j) \quad (1.12)$$

Cải tiến của đặc trưng thô có thể thu được bởi sử dụng một lược đồ để mô tả phân bố của *Sbest* . Đã làm tăng đáng kể hiệu năng tra cứu và làm cho đặc trưng có khả năng xử lý với một ảnh hoặc vùng có đa đặc tính kết cấu. Do vậy, nó là hữu ích hơn đối với các ứng dụng tra cứu ảnh.

1.3.2.3.2 Độ tương phản

Công thức cho tương phản là:

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}} \quad (1.13)$$

Trong đó: $\alpha_4 = \frac{\mu_4}{\sigma^4}$

μ_4: là mô men thứ tư về trung bình

σ^4: phương sai

1.3.2.3.3 Hướng

Độ lớn và góc của vector được định nghĩa như sau:

$$|\Delta G| = |\Delta_h| + |\Delta_v|$$

$$\theta = \tan^{-1}\left(\frac{\Delta_v}{\Delta_h}\right) + \frac{\pi}{2}$$

Trong đó Δ_h và Δ_v là các khác biệt ngang và dọc của chập. Sau đó, bằng lượng hoá θ và đếm số các pixel với độ lớn tương ứng |ΔG| lớn hơn một ngưỡng, một lược đồ của θ, biểu thị bằng *HD*, có thể được xây dựng. Lược đồ này sẽ cho biết các đỉnh bền vững cho các ảnh hướng cao và sẽ là tương đối phẳng với các ảnh không có hướng bền vững. Toàn bộ lược đồ được tóm lược để thu toàn bộ độ đo hướng dựa trên tính nhọn của các đỉnh:

$$F_{dir} = \sum_{p}^{n_p} \sum_{\phi \grave{o} \varpi_p}^{n} \left(\phi - \phi_p\right)^2 H_D(\phi) \qquad (1.14)$$

Trong đó *p* : là tổng các phạm vi trên *np* đỉnh

Mỗi đỉnh *p*, *wp* là tập các bin màu được phân bố trên nó.

ϕ_p : là bin màu nhận giá trị đỉnh.

1.3.2.4 Các đặc trưng lọc Gabor

Lọc Gabor được sử dụng rộng rãi để trích rút các đặc trưng ảnh, đặc biệt là các đặc trưng kết cấu. Nó tối ưu về mặt cực tiểu hoá sự không chắc chắn chung trong miền không gian và miền tần số, và thường được sử dụng như một hướng và tỷ lệ biên điều hướng và phát hiện đường. Có nhiều cách tiếp cận đã được đề xuất để mô tả các kết cấu của các ảnh dựa trên các lọc Gabor. Ý tưởng cơ bản của sử dụng các lọc Gabor để trích rút các đặc trưng kết cấu.

Hàm Gabor hai chiều *g*(*x*, *y*) được định nghĩa:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2pjwx\right] \qquad (1.15)$$

Trong đó

- σx : là độ lệch chuẩn của các bao Gaussian dọc theo hướng *x*
- σy : là độ lệch chuẩn của các bao Gaussian dọc theo hướng *y*

- Sau đó một tập các lọc Gabor có thể thu được bởi sự co giãn và quay thích hợp của $g(x, y)$:

$$g_{mn}(x, y) = a^{-m} g(x', y')$$

$$x' = a^{-m}(-x\cos\theta + y\sin\theta)$$

$$y' = a^{-m}(-x\cos\theta + y\sin\theta)$$

Trong đó: $a > 1, \theta = \frac{nr}{K}, n = 0,1,...,K-1$ và $m = 0,1,...,S-1$

K và S: là số các hướng và các tỷ lệ

a^{-m} : là nhân tố tỷ lệ nhằm để đảm bảo rằng năng lượng là độc lập của m. Một ảnh $I(x, y)$ đã cho, biến đổi Gabor của nó được định nghĩa bằng:

$$W_{mn} = \int I(x, y) g^*_{mn}(x - x_1, y - y_1) dx_1 dy_1 \qquad (1.16)$$

Trong đó *: chỉ ra số liên hợp phức.

μ_{mn} : là trung bình.

σ_{mn} : là độ lệch chuẩn của độ lớn $W_{mn}(x, y)$

$$f = \mu_{00}, \sigma_{00}, ..., \mu_{mn}, \sigma_{mn}, \Lambda, \mu_{S-1K-1}, \sigma_{S-1K-1}$$

có thể được sử dụng để biểu diễn đặc trưng kết cấu của một vùng kết cấu thuần nhất.

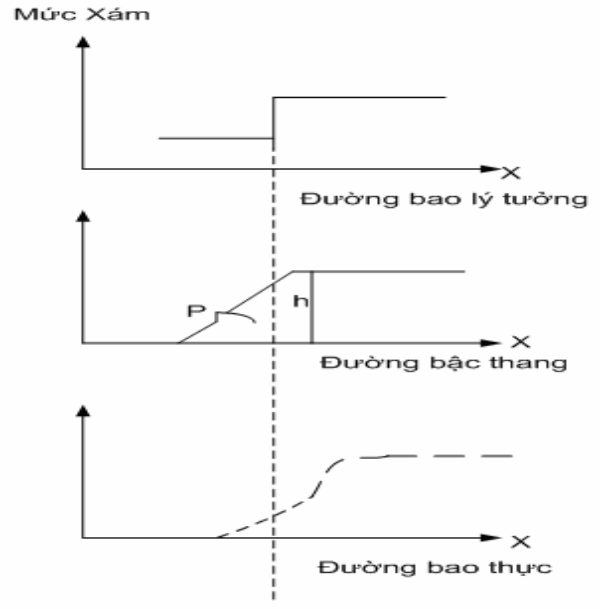
1.3.3 Trích chọn đặc trưng hình dạng (shape)

Phân đoạn ảnh là quá trình phân nhóm các pixel trong ảnh dựa trên các tiêu chuẩn tương đồng về màu, về texture, hoặc dựa trên các đường biên kết nối,…Khi đó, shape (dạng) là thuộc tính chính của các vùng ảnh phân đoạn, và đặc trưng shape có thể dùng để biểu diễn cho vùng phân đoạn. Đặc trưng shape cũng đóng vai trò quan trọng trong nhiều hệ thống truy vấn ảnh.

Màu sắc và kết cấu là những thuộc tính có khái niệm toàn cục của một bức ảnh. Trong khi đó, hình dạng không phải là một thuộc tính của ảnh. Do

đó, hình dạng thường được mô tả sau khi các ảnh được phân đoạn thành các vùng hoặc các đối tượng. Hay hình dạng chỉ là biên của đối tượng nào đó trong ảnh. Một biểu diễn đặc trưng hình dạng tốt cho một đối tượng phải bất biến với dịch chuyển, quay và tỷ lệ. Các bài toán trích trọn đặc trưng dựa trên hình dạng thường được bắt đầu với việc tìm và phát hiện biên của đối tượng, qua đó định hình cấu trúc và các thông tin bất biến của đối tượng ảnh.

Biên cạnh là đối tượng phân cách giữa 2 vùng ảnh thuần nhất có độ sáng khác nhau (Biên là nơi có biến thiên về độ sáng). Tập hợp các điểm biên tạo thành biên hay đường bao của ảnh (*boundary*). Ví dụ, trong một ảnh nhị phân, một điểm có thể gọi là biên nếu đó là điểm đen và có ít nhất một điểm trắng lân cận. Trong bài toán truy tìm ảnh, biên được sử dụng cho việc tìm kiếm những ảnh có cùng hình dáng với nhau. Để hình dung tầm qua trọng của biên ta xét đến ví dụ sau: khi người hoạ sĩ vẽ một cái bàn gỗ, chỉ cần vài nét phát thảo về hình dáng như mặt bàn chân bàn mà không cần thêm các chi tiết khác, người xem đã có thể nhận ra đó là cái bàn. Nếu ứng dụng của ta là phân lớp nhận diện đối tượng, thì coi như nhiệm vụ đã hoàn thành. Tuy nhiên nếu đòi hỏi thêm các chi tiết khác như vân gỗ hay màu sắc, … thì với chừng ấy thông tin là chưa đủ. Nhìn chung về mặt toán học, người ta có thể coi điểm biên của ảnh là điểm có sự biến đổi đột ngột về độ xám như chỉ ra trong hình sau:



Hình 1.7. Đường bao của ảnh

Như vậy phát hiện biên một cách lý tưởng là xác định được tất cả các đường bao trong các đối tượng. Định nghĩa toán học ở trên là cơ sở cho các kỹ thuật phát hiện biên.

1.3.3.1 Lược đồ hệ số góc (Edge Direction Histogram)

Lược đồ gồm 73 phần tử trong đó:72 phần tử đầu chứa số điểm ảnh có hệ số gốc từ 0 - 355 độ, các hệ số góc này cách nhau 5 độ. Phần tử cuối chứa số phần tử không nằm trên biên cạnh. Cần chuẩn hóa các đặc trưng này để thích hợp với kích thước khác nhau của ảnh:

$$H(I_D,i)=\frac{m(I_D,i)}{n_E(I_D)},\ i\in[0,1,\ldots,71] \tag{1.17}$$

$$H(72)=\frac{H(72)}{n(I_D)} \tag{1.18}$$

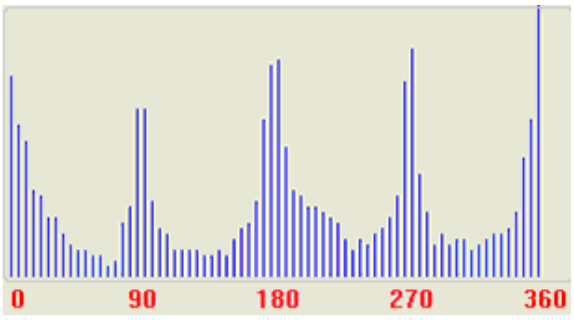
$m(I_D,i)$: là số điểm ảnh thuộc biên cạnh có hệ số gốc là $\alpha i=i*5$

$n_E(I_D)$: là tổng số các điểm ảnh thuộc biên cạnh

$n(I_D)$: là tổng số điểm ảnh của ảnh I_D

Ví dụ minh hoạ về lược đồ hệ số góc của ảnh:

Hình 1.8. Đường biên của ảnh



Hình 1.9. Lược đồ hệ số góc của ảnh

1.3.3.2 Vector liên kết hệ số góc

Là lược đồ tinh chế lược đồ hệ số góc, chia mỗi ô chứa (*bin*) thành 2 nhóm điểm ảnh: Nhóm điểm liên kết hệ số góc (*coherent pixels*) và nhóm điểm không liên kết hệ số góc (*non-coherence pixels*).

Một pixel trong một ô chứa (*bin*) được gọi là điểm liên kết hệ số góc (*coherent*) nếu nó thuộc vùng gồm các điểm thuộc cạnh có hệ số góc tương tự với kích thước lớn (thường vào khoảng 0.1% kích thước ảnh).

Với mỗi ô chứa (*bin*) giả sử số điểm liên kết hệ số góc là α và số điểm không liên kết hệ số góc là β thì vector liên kết hệ số góc được xác định:

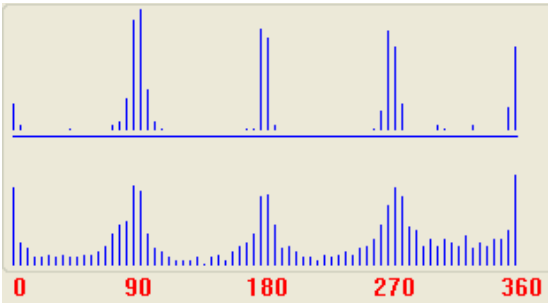
$V_E = [(\alpha_1, \beta_1), (\alpha_2, \beta_2), ..., (\alpha_n, \beta_n)]$, n là số ô màu (bin)

Độ đo tính tương tự giữa 2 ảnh dựa trên đặc trưng vector liên kết hệ số góc:

$$D_E\left(I_Q, I_D\right) = \sum_{j=1}^{n}\left(\left|\alpha_{Q_j} - \alpha_{D_j}\right| + \left|\beta_{Q_j} - \beta_{D_j}\right|\right) \quad (1.19)$$

Ví dụ minh họa ảnh và lược đồ vector liên kết hệ số góc:

Hình 1.10. Ảnh minh họa sự liên kết giữa các biên cạnh



Hình 1.11. Lược đồ vector liên kết hệ số góc của ảnh

1.3.4 Trích chọn đặc trưng cục bộ bất biến

SIFT là viết tắt của cụm từ *Scale-Invariant Feature Transform* là một trong những thuật toán nổi tiếng nhất hiện nay dùng để phát hiện và mô tả các đặc trưng của ảnh số. Thuật toán này được công bố bởi David Lowe vào năm 1999.

Hình 1.12. Hình ảnh sau khi SIFT

Hai hình trên có thể được nhận ra là của cùng một cảnh bởi SIFT. Giống như nhiều thuật toán về xử lý ảnh, SIFT là thuật toán khá phức tạp, phải trải qua nhiều bước xử lý và sử dụng nhiều kiến thức về toán học. Sau đây sẽ là các bước chính trong thuật toán:

- Dò tìm cực trị trong không gian đo (Scale space Extrema Detection)
- Lọc và trích xuất các điểm đặc biệt (Keypoint Localization)
- Gán hướng cho các điểm đặc trưng (Oriented Assignment)
- Bộ mô tả điểm đặc trưng (Keypoint Descriptor)

1.4 Khoảng cách ngữ nghĩa trong CBIR

Trong lĩnh vực tra cứu ảnh hiện nay có hai hệ thống đang được phát triển là: hệ thống tra cứu dựa trên nội dung và hệ thống dựa trên từ khoá. Điểm khác biệt duy nhất giữa hai hệ thống này chính là sự tương tác của người dùng. Con người thì luôn có xu hướng sử dụng các khái niệm đặc trưng mức cao như là: từ khoá, mô tả văn bản, giải thích hình ảnh và đo độ tương tự. Trong khi đó, các đặc trưng ảnh được tự động trích chọn bằng kỹ thuật thị giác máy tính thì chủ yếu là các đặc trưng mức thấp (màu sắc, kết cấu, hình dạng, vị trí không gian, v.v...). Nói chung là không có mối liên quan trực tiếp giữa đặc trưng mức thấp và đặc trưng mức cao.

Mặc dù các nhà nghiên cứu đã phát triển rất nhiều các thuật toán phức tạp để mô tả các đặc trưng hình ảnh như: màu sắc, kết cấu, hình dạng nhưng cũng không thể mô tả đầy đủ ngữ nghĩa và có nhiều hạn chế khi giải quyết trong một cơ sở dữ liệu có số lượng ảnh lớn. Các thí nghiệm mở rộng trên hệ thống CBIR cho thấy nội dung đặc trưng mức thấp thường không thể mô tả các khái niệm ngữ nghĩa mức cao trong suy nghĩ người dùng. Do đó, hiệu suất của CBIR vẫn chưa đáp ứng được nhu cầu của người dùng. Tác giả Eakins vào năm 1999 đã đề xuất ra ba mức độ của các truy vấn trong CBIR.

Mức 1: Tra cứu bởi các đặc trưng cơ bản như: màu sắc, kết cấu, hình dạng hoặc bố trí không gian của các phần tử ảnh.

Mức 2: Tra cứu bởi các đối tượng được xác định bằng đặc trưng nguyên thuỷ, với một mức độ suy luận logic. Ví dụ: "tìm một bức ảnh có chứa hình ảnh máy vi tính"

Mức 3: Tra cứu bởi các thuộc tính trừu tượng, bao hàm số lượng mục đích các đối tượng trong ảnh, hoặc nội dung của ảnh được miêu tả. Điều này có nghĩa là tra cứu tên các sự kiện, ý nghĩa của ảnh, hoặc các dấu hiệu nổi bật,... Ví dụ như: "tìm một bức ảnh có đám đông vui vẻ".

Có thể thấy mức 2 và mức 3 được gọi là tra cứu ảnh ngữ nghĩa. Khoảng cách giữa mức 1 và mức 2 là khoảng cách ngữ nghĩa. Sự khác biệt giữa giới hạn mô tả đặc trưng ảnh mức thấp và sự phong phú ngữ nghĩa người dùng, được gọi là "*Khoảng cách ngữ nghĩa*".

Các phương pháp thu hẹp khoảng cách ngữ nghĩa:

Làm thế nào để chúng ta có thể liên kết các đặc trưng mức thấp của ảnh với các ngữ nghĩa mức cao? Câu hỏi này đã thúc đẩy các nhà nghiên cứu cố gắng phát triển các công nghệ để giải quyết vấn đề này. Các công nghệ mới trong việc làm giảm khoảng cách ngữ nghĩa hiện nay có thể được phân ra theo các tiêu chí khác nhau. Bằng cách áp dụng vào các lĩnh vực khác nhau, các công nghệ tra cứu ảnh có thể được chia ra là: tra cứu ảnh nghệ thuật, tra cứu ảnh phong cảnh, tra cứu ảnh web, v.v.. Dưới đây là một số kỹ thuật thường được sử dụng để suy ra ngữ nghĩa mức cao:

- Sử dụng bản thể đối tượng để định nghĩa khái niệm mức cao.
- Sử dụng phương pháp học có giám sát hoặc không có giám sát để gắn đặc trưng mức thấp với các khái niệm truy vấn.
- Giới thiệu phản hồi liên quan (RF) vào vòng lặp tra cứu ảnh cho việc học liên tục ý định của người dùng.

- Sinh mẫu ngữ nghĩa (ST) để hỗ trợ tra cứu ảnh mức cao.
- Sử dụng cả hai cách là thông tin văn bản từ trên web và nội dung trực quan của ảnh để tra cứu ảnh web.

1.5 Một số hệ thống CBIR

Ứng dụng của tra cứu ảnh dựa trên nội dung có rất nhiều trong đời sống xã hội, phục vụ cho nhiều mục đích khác nhau, nhằm xác nhận, tra cứu thông tin. Nhờ đó mà giảm bớt công việc của con người, nâng cao hiệu suất làm việc, ví dụ như: Album ảnh số của người dùng, ảnh y khoa, bảo tàng ảnh, tìm kiếm nhãn hiệu, logo, mô tả nội dung video, truy tìm ảnh tội phạm, hệ thống tự nhận biết điều khiển luồng giao thông… Một vài hệ thống lớn đại diện cho các lĩnh vực bao gồm :

1.5.1 Hệ thống QBIC của hãng IBM

Là một hệ thống tra cứu ảnh thương mại đầu tiên và nổi tiếng nhất trong số các hệ thống tra cứu ảnh dựa trên nội dung. Nó cho phép người sử dụng tra cứu ảnh dựa vào màu sắc, hình dạng và kết cấu. QBIC cung cấp một số phương pháp: Simple, Multi-feature, và Multi-pass. Trong phương pháp truy vấn Simple chỉ sử dụng một đặc trưng. Truy vấn Multi-feature bao gồm nhiều hơn một đặc trưng và mỗi đặc trưng đều có trọng số như nhau trong suốt quá trình tìm kiếm. Truy vấn Multi-pass sử dụng đầu ra của các truy vấn trước làm cơ sở cho bước tiếp theo. Người sử dụng có thể vẽ ra và chỉ định màu, kết cấu mẫu của hình ảnh yêu cầu. Trong hệ thống QBIC màu tương tự được tính toán bằng thước đo bình phương sử dụng biểu đồ màu *k* phần tử (k-element) và màu trung bình được sử dụng như là bộ lọc để cải tiến hiệu quả của truy vấn. Bản demo của QBIC tại địa chỉ *wwwqbic.almaden.ibm.com*

1.5.2 Hệ thống Photobook

Hệ thống này được phát triển ở Massachusetts Institute of Technology cho phép người sử dụng tra cứu ảnh dựa trên màu sắc, kết cấu và hình dạng. Hệ thống này cung cấp một tập các thuật toán đối sánh gồm: Euclidean, Mahalanobis, Vector space angle, Histogram, Fourier peak và Wavelet tree distance như là những đơn vị đo khoảng cách. Trong hầu hết các phiên bản, đã có thể định nghĩa những thuật toán đối sánh của họ. Hệ thống như là một công cụ bán tự động và có thể sinh ra một mẫu truy vấn dựa vào những ảnh mẫu được cung cấp bởi người sử dụng. Điều này cho phép người sử dụng trực tiếp đưa những yêu cầu truy vấn của họ với những lĩnh vực khác nhau, và mỗi lĩnh vực họ có thể thu được những mẫu truy vấn tối ưu.

1.5.3 Hệ thống VisualSEEK và WebSEEK

Cả hai hệ thống này đều được phát triển tại Trường Đại học Colombia. VisualSEEK là hệ thống cơ sở dữ liệu ảnh; nó cho phép người sử dụng tra cứu ảnh dựa trên màu sắc, không gian miền và đặc điểm kết cấu. Tập màu và chuyển đổi wavelet dựa trên kết cấu được sử dụng để thực hiện những đặc điểm này. Thêm vào đó VisualSEEK còn cho phép người sử dụng tạo truy vấn bằng việc chỉ định vùng màu và những không gian vị trí của chúng. WebSEEK là một catalog ảnh và là công cụ tìm kiếm cho web. Hệ thống này cung cấp mẫu cho danh sách ảnh và video trên trang web sử dụng kết hợp xử lý dựa trên text và phân tích dựa trên nội dung.

1.5.4 Hệ thống RetrievalWare

Hệ thống này được phát triển bởi tập đoàn công nghệ Excalibur cho phép người sử dụng tra cứu ảnh bởi nội dung màu, hình dạng, kết cấu, độ sáng, kết cấu màu và hệ số co. Người sử dụng có thể điều chỉnh tỷ trọng của những đặc điểm này trong suốt quá trình tìm kiếm.

1.5.5 Hệ thống Imatch

Hệ thống này cho phép người sử dụng tra cứu ảnh bởi nội dung màu, hình dạng và kết cấu. Nó cung cấp một số phương pháp để tra cứu ảnh tương tự: Màu tương tự, màu và hình dạng (*Quick*), màu và hình dạng (*Fuzzy*) và sự phân bố màu. Màu tương tự truy vấn những ảnh tương tự với ảnh mẫu dựa trên sự phân bố màu toàn cục.

- Màu và hình dạng (*Quick*) tìm hình ảnh tương tự bởi việc kết hợp cả hình dạng, kết cấu và màu.
- Màu và hình dạng (*Fuzzy*) thực hiện thêm những bước xác định đối tượng trong ảnh mẫu.
- Phân bố màu cho phép người sử dụng vẽ ra sự phân bố màu hoặc xác định tỷ lệ phần trăm của một màu trong hình ảnh mong muốn.
- Imatch cũng cung cấp những đặc điểm khác nội dung để xác định ảnh: ảnh nhị phân, ảnh co kích thước, lưu trữ trong những định dạng khác và những ảnh có tên tương tự.

Ngoài ra, còn một số hệ thống khác như: Virage system, Stanford SIMPLICity system, NEC PicHunter system, v.v…

Kết luận chương 1

Chương này tập trung tìm hiểu khái quát về tra cứu ảnh dựa trên nội dung, trong đó các nội dung đã tìm hiểu bao gồm: Các phương pháp tra cứu ảnh truyền thống; một số phương pháp trích chọn đặc trưng ảnh; khoảng cách ngữ nghĩa và phương pháp làm giảm khoảng cách ngữ nghĩa; tìm hiểu một số hệ thống CBIR lớn theo các lĩnh vực đã ứng dụng rộng rãi.

Với các kết quả tìm hiểu ở trên chúng ta có thể dễ dàng nhận ra những hạn chế của các hệ thống CBIR nêu trên, nó chỉ phù hợp với từng lĩnh vực cụ thể, các kết quả trả về còn xa so với sự mong đợi của người dùng.

Để khắc phục những hạn chế trên, việc kết hợp nhiều đặc trưng ảnh để xây dựng truy vấn cùng với sự phản hồi liên quan từ người dùng làm nâng cao hiệu quả của các phương pháp máy học là nhiệm vụ, hướng nghiên cứu tiếp theo trong chương 2.

Chương 2. KẾT HỢP NHIỀU ĐẶC TRƯNG TRONG TRA CỨU ẢNH SỬ DỤNG SVM VÀ PHẢN HỒI LIÊN QUAN

2.1 Phản hồi liên quan trong CBIR

2.1.1 Giới thiệu về phản hồi liên quan

Phương pháp tra cứu ảnh dựa trên nội dung ra đời đã mở ra một hướng đi triển vọng trong tra cứu ảnh, tuy nhiên các kết quả tra cứu mới chỉ dựa trên điểm tương đồng của các đặc trưng trực quan thuần túy, mỗi loại đặc trưng trực quan có xu hướng chỉ nắm bắt một khía cạnh của thuộc tính hình ảnh và nó thường khó khăn cho người sử dụng để xác định rõ những khía cạnh khác nhau được kết hợp cũng như khoảng cách ngữ nghĩa. Để khắc phục được nhược điểm này, kỹ thuật dựa trên phản hồi liên quan (RF) được giới thiệu vào năm 2007 bởi Liu cùng các cộng sự. Đây là kỹ thuật học trực tuyến có giám sát mà được sử dụng rộng rãi trong hệ thống CBIR để khắc phục các nhược điểm trên. RF sẽ thay đổi nhiều lần thông tin mô tả truy vấn (*đặc trưng, mô hình đối sánh, metrics,...*) như là hồi đáp phản hồi của người dùng trên kết quả tra cứu, thiết lập liên kết giữa các khái niệm mức cao và đặc trưng mức thấp.

Ý tưởng chính của phương pháp này là khi đưa vào một truy vấn, đầu tiên hệ thống sẽ trả về một danh sách các hình ảnh được xếp theo một độ tương tự xác định trước. Sau đó, người dùng đánh dấu những hình ảnh có liên quan đến truy vấn (*mẫu dương*) hoặc không có liên quan (*mẫu âm*). Hệ thống sẽ chọn lọc kết quả tra cứu dựa trên những phản hồi và trình bày một danh sách mới của hình ảnh cho người dùng. Do đó, vấn đề quan trọng trong phản hồi liên quan là làm thế nào để kết hợp các *mẫu dương* và *mẫu âm* để tinh chỉnh các truy vấn, điều chỉnh các biện pháp cho phù hợp.

Để cải thiện hơn nữa, hệ thống CBIR dựa trên RF lần đầu tiên cập nhật trọng số đặc trưng [12] tương ứng một cách tự động để nắm bắt mục đích của người dùng trong truy vấn và nhận thức chủ quan sau mỗi vòng lặp truy vấn. Kết quả đã cải thiện đáng kể hiệu năng tra cứu ảnh so với các hệ thống không dựa trên RF khác. Người dùng đóng một vai trò quan trọng trong hệ thống *CBIR* dựa trên *RF,* những phản hồi chính xác từ người dùng sẽ làm tăng hiệu năng của hệ thống. Các nhà nghiên cứu đang tập trung áp dụng các kỹ thuật phản hồi liên quan để cải thiện hiệu năng tra cứu.

2.1.2 Các kỹ thuật phản hồi liên quan

Trong các hệ thống CBIR với phản hồi liên quan, người dùng đóng một vai trò quan trọng. Các thông tin phản hồi chính xác từ người dùng sẽ góp phần làm tăng đáng kể hiệu năng của hệ thống tra cứu. Chọn lọc truy vấn sử dụng thông tin phản hồi liên quan đã đạt được nhiều sự chú ý trong nghiên cứu và phát triển của các hệ thống CBIR. Các nghiên cứu đã tập trung vào điều chỉnh truy vấn trong mỗi phiên tra cứu. Điều này thường được gọi là học trong nội bộ truy vấn hoặc học ngắn hạn. Ngược lại, liên truy vấn, còn được gọi là học dài hạn là chiến lược cố gắng để phân tích mối quan hệ giữa các phiên tra cứu hiện tại và quá khứ. Các kỹ thuật học máy trên những phản hồi của người dùng cũng được các nhà nghiên cứu tập trung áp dụng để cải thiện hiệu năng tra cứu. Kỹ thuật cập nhật truy vấn và kỹ thuật học thống kê là những kỹ thuật được sử dụng phổ biến trong các hệ thống CBIR với phản hồi liên quan .

2.1.2.1 Kỹ thuật cập nhật truy vấn

Kỹ thuật cập nhật truy vấn cải thiện việc biểu diễn chính truy vấn bằng cách sử dụng thông tin được gán nhãn chủ quan của người dùng. Các ví dụ của kỹ thuật cập nhật truy vấn bao gồm cập nhật trọng số truy vấn, di chuyển truy vấn, và mở rộng truy vấn.

Cập nhật trọng số truy vấn làm thay đổi trọng số tương đối của các đặc trưng khác nhau trong biểu diễn truy vấn. Kỹ thuật cập nhật vector trọng số cho phép hệ thống học sự giải thích của người dùng về hàm khoảng cách. Ý tưởng trung tâm đằng sau phương pháp cập nhật trọng số rất là đơn giản và trực quan. Mỗi ảnh được đại diện bởi một vector đặc trưng *N* chiều. Nó có thể được xem như là một điểm trong không gian *N* chiều. Các chiều đặc trưng quan trọng để giúp tra cứu các ảnh liên quan sẽ được nâng cấp tầm quan trọng trong khi các chiều khác cản trở tiến trình này sẽ bị giảm tầm quan trọng. Vào năm 2004, Kushki và các cộng sự đã sử dụng kỹ thuật cập nhật trọng số để học ánh xạ tối ưu giữa đặc trưng trực quan mức thấp và khái niệm ngữ nghĩa mức cao của ảnh. Kỹ thuật này hoạt động bằng cách tinh chỉnh các trọng số (*hoặc sự quan trọng*) của từng thành phần đặc trưng hoặc bằng cách thay đổi đo độ tương tự một cách tương ứng. Cũng trong năm 2004, Muneesawang và cộng sự đã áp dụng kỹ thuật di chuyển truy vấn để cho phép người dùng thay đổi trực tiếp đặc trưng của ảnh truy vấn bằng cách chỉ định các thuộc tính của các ảnh liên quan hoặc không liên quan được đánh dấu bởi người dùng. Có nghĩa là, các đặc trưng của nội dung ảnh truy vấn được thay đổi theo hướng biểu diễn ngữ nghĩa chính xác hơn được cung cấp bởi người dùng trong suốt quá trình tra cứu. Vào năm 2005, Widyantoro và các cộng sự đã áp dụng kỹ thuật mở rộng truy vấn để thêm vào một tập các ảnh liên quan mà không được gán nhãn bởi người dùng để bù đắp cho sự thiếu hụt những ảnh đã được gán nhãn bởi người dùng giúp hệ thống nắm bắt ý nghĩa của ảnh truy vấn một cách chính xác hơn.

2.1.2.2 Những kỹ thuật học thống kê

Kỹ thuật học thống kê đã cải thiện giới hạn phân loại giữa những ảnh liên quan và không liên quan hoặc dự đoán những ảnh liên quan mà chưa

được gán nhãn trong suốt quá trình huấn luyện. Các ví dụ của kỹ thuật học thống kê bao gồm học quy nạp và học chuyển đổi.

Học quy nạp được định nghĩa như là một quá trình tiếp thu tri thức bằng cách vẽ ra các suy luận quy nạp từ giáo viên hoặc môi trường cung cấp sự kiện. Đây là một quá trình liên quan đến hoạt động khái quát, biến đổi, hiệu chỉnh, tinh chỉnh biểu diễn tri thức. Phương pháp học quy nạp được áp dụng trong hệ thống CBIR nhằm tạo ra các bộ phân lớp khác nhau để phân tách thành ảnh có liên quan (mẫu dương) và không có liên quan (mẫu âm), và khái quát tốt hơn những ảnh chưa gán nhãn. Ở đây, những ảnh có liên quan và không có liên quan là nhãn ảnh tra cứu dương và âm một cách tương ứng bởi người dùng trong suốt phiên tra cứu. Các kỹ thuật học quy nạp điển hình bao gồm mạng neural, học cây quyết định, học Bayesian, Boosting, Support vector machine (SVM), học SVM mờ (FSVM). Vào năm 2000, MacArthur và cộng sự đã sử dụng cây quyết định trong ứng dụng CBIR. Các ảnh liên quan và không liên quan được đánh dấu bởi người dùng được sử dụng để phân chia không gian đặc trưng cho đến khi tất cả các ví dụ trong một phân vùng là cùng lớp. Năm 2003, Su và các cộng sự đã cung cấp phản hồi liên quan và không liên quan từ người dùng vào bộ phân loại Bayesian. Những ảnh liên quan được sử dụng để ước lượng một phân bố Gaussian. Phân bố này dùng để biểu diễn những ảnh mà người dùng mong muốn trong khi những ảnh không liên quan thì lại được sử dụng để duyệt lại việc xếp hạng những ứng cử đã được tra cứu. Năm 2001, Tong và cộng sự đã đề xuất một hệ thống CBIR với sự trợ giúp của SVM để học đường bao quyết định sử dụng mẫu liên quan và không liên quan đã thu thập được từ vòng lặp tra cứu trước đó. Đường bao quyết định này sau đó được sử dụng để phân tách ảnh trong cơ sở dữ liệu thành hai phân vùng liên quan và không liên quan. Năm 2006, Wu và các cộng sự đã áp dụng FSVM để học đường bao quyết định để phân tách ảnh

huấn luyện dương và âm dựa trên các trọng số mờ tương ứng. Đường bao quyết định sau đó được dùng để phân chia cơ sở dữ liệu ảnh thành ảnh liên quan và không liên quan. Những ảnh liên quan với khoảng cách lớn nhất tới đường bao quyết định được coi như là những ảnh tương tự nhất với ảnh truy vấn. Năm 2004, Tieu và cộng sự đã đề xuất một hệ thống CBIR mà sử dụng kỹ thuật học “boosting" để sinh ra một số lượng lớn các đặc trưng chọn lọc cao cho việc nắm bắt nhiều dạng của khái niệm trực quan ảnh. Một loạt các phương pháp học yếu dựa trên một số lượng nhỏ các đặc trưng đã được huấn luyện trong suốt thời gian truy vấn. Bằng việc kết hợp các phân loại yếu, hệ thống cuối cùng thu được một bộ phân loại mạnh có độ tương quan tốt hơn với phân lớp lý tưởng.

2.1.2.3 Phương pháp học ngắn hạn

Trong học ngắn hạn, chỉ những phản hồi của phiên tìm kiếm hiện tại được sử dụng cho thuật toán học và các đặc trưng ảnh là nguồn dữ liệu chính. Thách thức chính trong phương pháp này là tìm sự kết hợp tốt nhất các đặc trưng biểu diễn truy vấn của người dùng. Ví dụ một bộ các đặc trưng tối ưu sẽ bao gồm những đặc trưng mà có thể bắt lấy sự tương tự giữa các mẫu dương hoặc những đặc trưng mà có thể phân biệt các mẫu dương và mẫu âm. Do đó nhiều thuật toán học máy cổ điển được sử dụng trong học ngắn hạn như là SVMs, mô hình học Bayes, boosting và đánh trọng số đặc trưng, phân tích sự khác biệt v.v.. Tuy nhiên, cách tiếp cận học ngắn hạn là nhiệm vụ rất khó bởi vì trước hết kích thước của dữ liệu huấn luyện là nhỏ hơn nhiều so với độ dài không gian đặc trưng, thứ hai là có quá nhiều sự mất cân bằng giữa phản hồi của những người dùng khác nhau. Và cuối cùng quá trình học là trực tuyến sẽ đòi hỏi nhiều thời gian thực hơn.

2.1.2.4 Phương pháp học dài hạn

Phương pháp học dài hạn có thể đạt được độ chính xác tra cứu tốt hơn so với các kỹ thuật RF truyền thống. Có thể sử dụng học tập dài hạn để vượt qua những khó khăn như không có khả năng nắm những ngữ nghĩa hiếm hoi và mất cân bằng giữa các ví dụ phản hồi, và thiếu cơ chế bộ nhớ v.v.. Trên thực tế, khái niệm học dài hạn trong CBIR được thông qua từ công việc của lọc cộng tác. Phương pháp học dài hạn sử dụng các thông tin phản hồi thu thập được từ trước. Nó là một quá trình tích lũy cho việc thu thập thông tin phản hồi nhanh chóng và được lưu trữ trong các hình thức của ma trận. Một ma trận lưu trữ các nhãn được cung cấp bởi người dùng cho mỗi hình ảnh trong mỗi lần lặp. Thông thường kích thước của ma trận lịch sử tìm kiếm là lớn, mô hình thống kê và các phương pháp như phân tích thành phần chính và phân tích ngữ nghĩa tiềm ẩn rất phổ biến trong các phương pháp học tập dài hạn. Tuy nhiên, có những vấn đề trong phương pháp học tập dài hạn.

Những hạn chế của phương pháp học dài hạn :

- Thứ nhất, đây là phương pháp thể hiện sự không phù hợp với những ứng dụng mà hình ảnh thường xuyên được thêm vào hoặc gỡ bỏ. Một cách tiếp cận tốt hơn là sử dụng mô hình vector đặc trưng và phân tích mối quan hệ liên truy vấn.

- Thứ hai, là sự thưa thớt của thông tin phản hồi được ghi lại. Chất lượng học dài hạn phụ thuộc rất nhiều vào số lượng người dùng đăng nhập mà hệ thống lưu trữ. Do thiếu các tương tác và cơ sở dữ liệu lớn, nó không phải là dễ dàng để thu thập thông tin đăng nhập một cách đầy đủ.

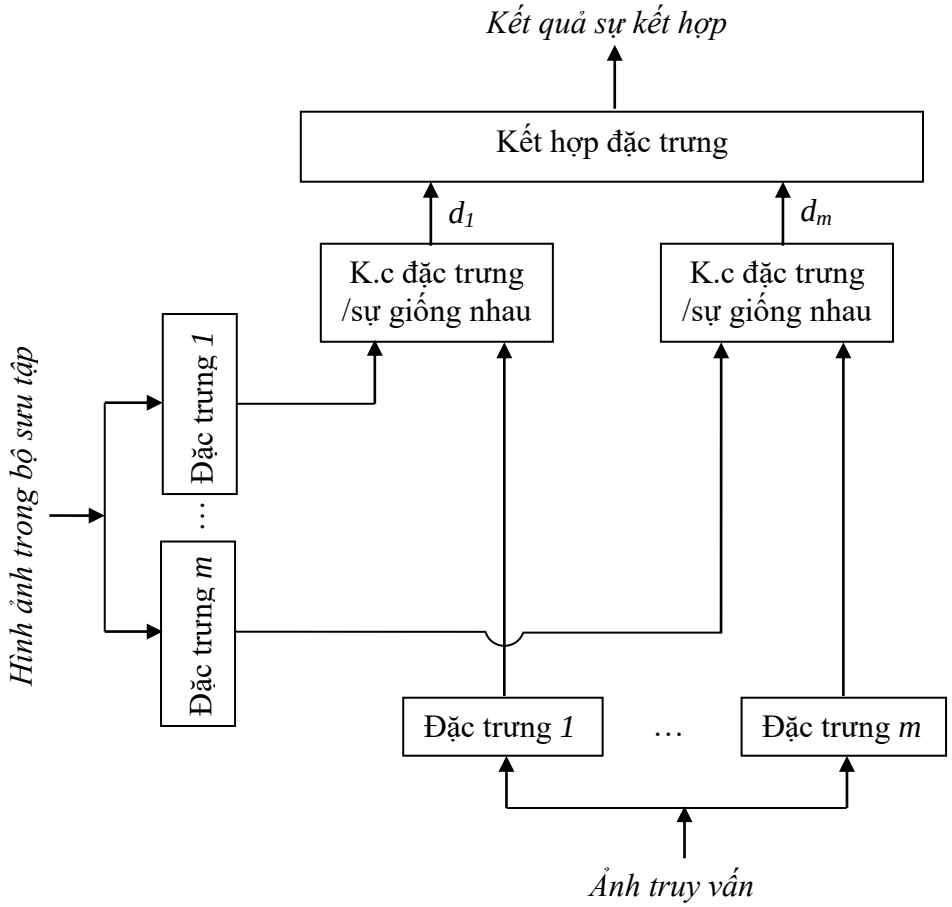
Cuối cùng, vấn đề khác là hầu hết các giải pháp học dài hạn chỉ giới thiệu các kiến thức ngữ nghĩa được ghi nhớ cho người sử dụng nhưng thiếu khả năng học tập để dự đoán ngữ nghĩa ẩn trong các mẫu ngữ nghĩa thu được.

2.2 Kết hợp nhiều đặc trưng trong CBIR

Trong những hệ thống tra cứu ảnh dựa trên nội dung sử dụng nhiều đặc trưng ảnh trong một bộ sưu tập sẽ được sắp xếp theo độ tương tự đối với ảnh truy vấn, trong đó câu truy vấn được mô tả bằng những đặc trưng liên quan đến thị giác, chẳng hạn như màu sắc, kết cấu và hình dạng,... Mỗi đặc trưng liên quan đến thị giác thường mô tả một khía cạnh của nội dung, và sự kết hợp của nhiều đặc trưng [13] cho ta một cách mô tả đầy đủ nội dung ảnh. Có hai phương pháp kết hợp các đặc trưng đó là kết hợp trước và kết hợp sau.

- Phương pháp kết hợp trước: Kết hợp nhiều đặc trưng để hình thành một vector đặc trưng chung và sử dụng một độ đo thống nhất để đo độ tương tự giữa các ảnh. Ưu điểm của phương pháp này là tiện lợi trong tính toán và phân tích toán học. Tuy nhiên, phương pháp này không áp dụng cho các đặc trưng có độ đo khác nhau.
- Phương pháp kết hợp sau [1]: Là mỗi một không gian đặc trưng được sử dụng một độ đo khác nhau, sau đó các độ đo này được kết hợp lại thành một độ đo chung để đo độ tương tự của ảnh như minh họa trong *Hình 2.1*. Ưu điểm của phương pháp này là sử dụng nhiều độ đo khác nhau để đo độ tương tự giữa các ảnh, ngược lại phương pháp này độ tính toán phức tạp hơn, phải tính toán nhiều lần.

Trong [13] Hình 2.1 di, (i=1,2,...) là ký hiệu của khoảng cách đặc trưng i^{th} giữa ảnh truy vấn với ảnh trong CSDL. Những hình ảnh trong CSDL này sẽ được sắp xếp lại theo độ đo tương tự của chúng. Vì vậy, việc tra cứu ảnh sẽ phụ thuộc phần lớn vào việc sắp xếp theo hệ thống sự kết hợp các đặc trưng.



Hình 2.1. Mô hình sự kết hợp các đặc trưng trong hệ thống CBIR

Để đo độ tương tự giữa các đặc trưng, thông thường người ta sử dụng độ đo có trọng số. Dưới đây sẽ trình bày cụ thể phương pháp này.

2.2.1 Độ đo có trọng số

Sự hình thành công thức của độ đo có trọng số [9] là liên quan đến sự tính toán của độ đo tương tự từ véc tơ truy vấn, được trình bày như sau:

Một hình ảnh x có thể được mô tả như là $x = x(F)$, trong đó F là một tập của các đặc trưng mức thấp liên quan tới hình ảnh, như màu sắc, kết cấu, hình dạng,… Với mỗi đặc trưng f_i có thể được mô hình bởi một số các biểu diễn, ví dụ biểu đồ mầu sắc (*Color Histogram)* và khả năng mở rộng màu (*Scalable Color)* là biểu diễn của các đặc trưng màu sắc. Mỗi biểu diễn f_{ij} chính là một vector với nhiều thành phần:

$$f_{ij} = \left[f_{ij1}, ..., f_{ijn}, ..., f_{ijk}, ..., f_{ijN} \right] \quad (2.1)$$

trong đó N là độ dài vector. Đối với mỗi f_i , f_{ij} và f_{ijk} có thể được gán tương ứng với một trọng số w_i , w_{ij} và w_{ijk}. Nếu ta coi M là độ đo tương tự, thì có thể đánh giá sự tương tự giữa hai ảnh trong điều kiện của f_{ij} và trọng số w_{ijk} của nó là :

$$S(f_{ij}) = M(f_{ij,} W_{i,k}) \text{ với } k = 1...N \quad (2.2)$$

Ví dụ, để so sánh hai ảnh được biểu diễn bằng một vector có thể sử dụng Minkowski.

$$d_p^{f_{ij}}(x, y) = \left(\sum_{k=1}^{N} | x(f_{ijk}) - y(f_{ijk}) |^p \right)^{1/p} \quad (2.3)$$

Với p> 1 công thức (2.2) có thể được viết lại như sau:

$$S(f_{ij}) = \left(\sum_{k=1}^{N} w_{ijk} | x(f_{ijk}) - y(f_{ijk}) |^p \right)^{1/p} \quad (2.4)$$

Hơn nữa nó có thể được định nghĩa ở mức độ khác nhau giữa sự mô tả của nó và các thành phần vector được liên kết giữa chúng trong sự phù hợp với những nét đặc trưng của những tập được mô tả. Ví dụ như sự mô tả một hình ảnh trong "*Color Histogram Layout"* được chia thành G các hình ảnh nhỏ (*G/4* phần chia nhỏ theo chiều ngang và *G/4* phần chia nhỏ theo chiều dọc) và mỗi hình ảnh nhỏ được tính toán trong các biểu diễn biểu đồ mầu. Do đó, công thức (2.1) trở thành:

$$f_{ij} = \left[g_{ij1}, ..., g_{ijk}, ..., g_{ijG} \right] \quad (2.5)$$

với

$$g_{ij1} = \left[g_{ij1}, ..., g_{ijh} \right] ;$$
$$g_{ijG} = \left[g_{ijk}, ..., g_{ijN} \right] \quad (2.6)$$

Tương tự như trên, có thể gán một trọng số w_g cho tập của các thành phần, với $g = 1...G$ và công thức (2.2) có thể được mở rộng. Ví dụ trường hợp

mà các trọng số được gán cho những tập con *G* của các biểu diễn đại diện cho giao diện của biểu đồ mầu. Trong trường hợp này công thức (2.3) trở thành:

$$S(f_{\mathrm{ij}}) = \sum_{g=1}^{G} \mathrm{w}_g . d_p^g(x,y) \tag{2.7}$$

Trong đó $d_p^g(x,y)$ là khoảng cách giữa *x* và y trong không gian bé *g*. Cuối cùng trọng số cũng có thể được ước lượng cho độ đo tương tự ở mức của đặc trưng f_i như sau:

$$S(f_{\mathrm{i}}) = \sum_{j} \mathrm{w}_{\mathrm{ij}} S(f_{\mathrm{ij}}) \tag{2.8}$$

và hơn nữa ở mức độ cao nhất

$$S = \sum_{i} \mathrm{w}_{\mathrm{i}} S(f_{\mathrm{i}}) \tag{2.9}$$

Trọng số này không chỉ được sử dụng để đo độ tương tự ảnh, mà còn có thể được sử dụng để đo thứ tự xếp hạng (*score*) của ảnh. Đặc biệt, các trọng số có thể được ước lượng lẫn nhau từ nhóm những điểm số liên quan được gán cho những hình ảnh, cho mỗi đặc trưng hoặc cho nhóm các điểm số bắt đầu từ những tập đặc trưng con để kết hợp các điểm số có nguồn gốc từ các tập con đặc trưng (2.6). Trong những trường hợp trên công thức (2.4) và (2.7) có thể được viết lại như sau:

$$rel\left(x(f_{\mathrm{ij}})\right) = \sum_{k=1}^{N} \mathrm{w}_{\mathrm{ij}k} . rel\left(x(f_{\mathrm{ijk}})\right) \tag{2.10}$$

$$rel\left(x(f_{\mathrm{ij}})\right) = \sum_{g=1}^{G} \mathrm{w}_g . rel\left(x(g_{\mathrm{ijg}})\right) \tag{2.11}$$

2.2.2 Ước lượng độ liên quan của các đặc trưng

Ước lượng độ liên quan của các đặc trưng (Estimation of feature relevance) [9] tập trung vào vấn đề của nhiệm vụ các trọng số để hợp thành

(*hoặc thiết lập tập các thành phần*) của véc tơ thành phần và các điểm số (*scores*) liên quan.

Mục đích của trọng số đặc trưng trong một độ đo tương tự là việc chỉ ra những đặc trưng quan trọng hơn để cho phép thu hồi một số lượng hình ảnh liên quan lớn hơn. Thông thường, các trọng số được gán cho các đặc trưng khác nhau thì được tính bằng cách so sánh các giá trị các đặc trưng của vector truy vấn với các đặc trưng tương ứng của hình ảnh có liên quan. Dưới đây sẽ trình bày một số kỹ thuật phổ biến được sử dụng để xác định trọng số:

2.2.2.1 Nghịch đảo của độ lệch chuẩn

Kỹ thuật đơn giản nhất là dựa trên việc tính toán độ lệch chuẩn cho mỗi đặc trưng [9], bằng cách tham gia vào tính toán chỉ những hình ảnh có liên quan. Tôi đưa ra một ma trận $R \times F$ đặc trưng I, trong đó R là số lượng hình ảnh có liên quan và F là số những đặc trưng. Mỗi cột của I là một vector gồm các giá trị của các thành phần vector đặc trưng tương tự cho tất cả các hình ảnh R có liên quan. Nếu các giá trị của vector cột thứ j-th là tương tự như nhau, điều đó có nghĩa là những hình ảnh có liên quan có những giá trị tương tự với đặc trưng j, và những đặc trưng này có liên quan mật thiết tới truy vấn. Độ tương tự giữa các thành phần đặc trưng càng lớn thì độ liên quan của các đặc trưng đó với ảnh truy vấn càng lớn, và nó được gán trọng số lớn hơn. Do đó nghịch đảo của độ lệch chuẩn của mỗi giá trị của vector j là một trọng số tốt cho các đặc trưng j.

$$w_j = \frac{1}{\sigma_j} \tag{2.12}$$

trong đó j là đặc trưng thứ j-th và σ_j là độ lệch chuẩn của nó.

2.2.2.2 Học xác suất

Học xác suất (Probabilistic learning) [9], một kỹ thuật khác để xác định trọng số đặc trưng là dựa trên các lỗi do dự đoán xác suất mà các giá trị $x(f_i)$ của thành phần đặc trưng *i*-th của hình ảnh *x* là bằng với giá trị $z(f_i)$ của thành phần đặc trưng *i*-th của truy vấn z [7]. Giả sử y∈{0,1} là lớp nhãn, và x_j là ảnh tra cứu thứ *j*-th, với *j = 1. . .K*. Độ đo liên quan của thành phần đặc trưng *i*-th của truy vấn z là [7]:

$$r_i(z) = \frac{\sum_{j=1}^{K} y_i 1\left(|x_j(f_i) - z(f_i)| \le \Omega\right)}{\sum_{j=1}^{K} 1\left(|x_j(f_i) - z(f_i)| \le \Omega\right)} \quad (2.13)$$

trong đó y_j có thể giả định giá trị 1 (*có liên quan*) hoặc 0 (*không liên quan*) tùy thuộc vào việc gán nhãn bởi người dùng, 1 (.) là một hàm trả về 1 nếu đối số của nó là đúng hoặc 0 nếu ngược lại. Ω được chọn sao cho:

$$\sum_{j=1}^{K} 1\left(|x_j(f_j) - z(f_j)| \le \Omega\right) = C \quad (2.14)$$

trong đó $C \le K$ là một hằng số. Trọng số cho thành phần đặc trưng *i*-th sau đó được cho bởi công thức:

$$w_i(z) = \frac{e^{(T.r_i(z))}}{\sum_{l=1}^{F} e^{(T.r_i(z))}} \quad (2.15)$$

trong đó *F* là số lượng các đặc trưng và *T* là một tham số có thể được chọn để tăng cường ảnh hưởng của r_i đối với w_i. Trong thực tế, nếu *T=0*, $w_i = 1/F$ cho tất cả các đặc trưng, mặt khác nếu *T* lớn, ảnh hưởng của r_i sẽ lớn hơn.

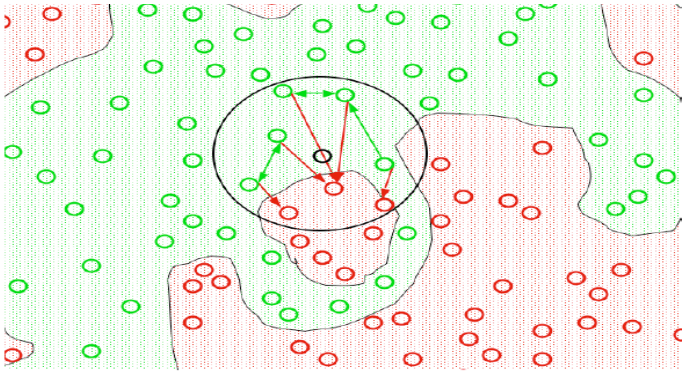
Các trọng số đã nêu trên phản ánh sự liên quan một số đặc trưng có trên truy vấn. Nếu trọng số nhỏ, đặc trưng không hữu ích để dự đoán truy vấn, ngược lại trọng số lớn thì sự đóng góp để dự đoán truy vấn là rất lớn. Điều

đáng chú ý là trọng số ước tính trong nghiên cứu xác suất liên quan chặt chẽ với các vị trí của các truy vấn, không phụ thuộc vào vị trí tương đối của các hình ảnh có liên quan và không liên quan trong không gian đặc trưng.

2.2.2.3 Cập nhật trọng số đặc trưng dựa trên láng giềng gần nhất

Cập nhật trọng số đặc trưng dựa trên láng giềng gần nhất (*Neighborhood - Based feature weighting*)[9] mục đích là để thay đổi độ đo khoảng cách thông qua những trọng số phù hợp sao cho các ảnh có liên quan sẽ gần lại nhau hơn so với những ảnh không liên quan [8]. Kỹ thuật này dựa trên cơ sở sau khi tính toán liên quan láng giền gần nhất [6]. Vì vậy, đầu tiên sự có liên quan của không gian đặc trưng khác nhau được ước tính trong điều kiện của khả năng sự biểu diễn những hình ảnh có liên quan như là láng giềng gần nhất, sau đó sự liên quan của một hình ảnh được ước tính theo các hình ảnh có liên quan và không liên quan trong vùng láng giềng gần nhất của nó. Độ liên quan của một hình ảnh có thể được tính như sau :.

$$rel_{NN}(x) = P(relevant \mid x) = \frac{P^{r}_{NN}(x)}{P^{r}_{NN}(x) + P^{nr}_{NN}(x)} = \frac{\| x - NN^{nr}(x) \|}{\| x - NN^{r}(x) \| + \| x - NN^{nr}(x) \|} \quad (2.16)$$



Hình 2.2 Xem xét vị trí các trọng số mà hình ảnh có liên quan và không liên quan giả định nhau

P_{NN} được định nghĩa như sau:

$$P_{NN}(x) = \frac{1/t}{V(\| x - NN(x) \|)} \quad (2.17)$$

trong đó *t* là số lượng hình ảnh và *x* là hình ảnh được xem xét. $NN(x)$ biểu thị láng giềng gần nhất, $NN^r(x)$ láng giềng gần nhất có liên quan, và $N^{nr}(x)$ láng giềng gần nhất không liên quan của *x* tương ứng, *V* là tập của các điểm rất nhỏ nằm ở trung tâm hình ảnh x, bao gồm cả *NN*(*x*). Tập *V (r)* của (hypersphere) trong một không gian d chiều được thể biểu diễn là $V(r) = V_d.r^d$ trong đó r là bán kính của hypersphere và V_d là một hằng số bằng cho bởi công thức $\frac{\pi^{d/2}}{r\left(\frac{d}{2}+1\right)}$ và Γ(.) là hàm Gamma. Theo công thức (2.16) thì độ liên quan của các đặc trưng được đánh giá bằng công thức sau:

$$rel_{NN}(f_x) = \frac{P^r_{NN}(f_x)}{P^r_{NN}(f_x) + P^{nr}_{NN}(f_x)} \quad (2.18)$$

trong đó $P^r_{NN}(f_x)$ và $P^{nr}_{NN}(f_x)$ nó được đánh giá theo công thức dưới đây:

$$P^r_{NN}(f_x) = \frac{1/t}{V^r_{NN}(f_x)} \quad ; \quad P^{nr}_{NN}(f_x) = \frac{1/t}{V^{nr}_{NN}(f_x)} \quad (2.19)$$

Với t là chỉ số của các ảnh, $V^r_{NN}(f_x)$ được đánh giá như là số lượng trung bình xung quanh các hình ảnh có liên quan mà chỉ chứa hình ảnh có liên quan gần nhất với nó, và $V^{nr}_{NN}(f_x)$ là số ượng trung bình xung quanh các hình ảnh có liên quan mà chỉ chứa hình ảnh không liên quan gần nhất với nó

$$V^r_{NN}(f_x) = \frac{1}{|R|}\sum_{i \in R} d^{f_x}_{\min}(x_i, R) \quad (2.20)$$

$$V^{nr}_{NN}(f_x) = \frac{1}{|R|}\sum_{i \in R} d^{f_x}_{\min}(x_i, N) \quad (2.21)$$

trong đó *R* và *N* là tập theo thứ tự của những hình ảnh liên quan và không liên quan, *x* là một hình ảnh đại diện, $d_{\min}^{f_x}(.)$ là một hàm trả về khoảng cách tối thiểu giữa một hình ảnh và một tập của các hình ảnh, và |.| là một hàm trả về lực lượng trong tập. Khoảng cách này được đo theo thành phần f_x của không gian đặc trưng và nó được tính toán bằng công thức sau:

$$d_{\min}^{f_x}(x_i, M) = \min\left[d_p^{f_x}(x_i, x_k) \right] \forall x_k \in M \tag{2.22}$$

Với *M* là một ảnh đại diện của tập các hình ảnh. Tóm lại, các trọng số được liên kết lại tới mỗi đặc trưng f_x có thể được tính toán như dưới đây theo những công thức ở trên:

$$\mathrm{w}f_x = rel_{NN}(f_x) = \frac{\sum_{i \in R} d_{\min}^{f_x}(x_i, R)}{\sum_{i \in R} d_{\min}^{f_x}(x_i, R) + \sum_{i \in R} d_{\min}^{f_x}(x_i, N)} \tag{2.23}$$

Trong sự phù hợp với các đặc trưng trong tập, mỗi tập được chia làm ba hoặc bốn tập nhỏ (mục 2.2.1). Theo cách đó thì công thức (2.23) có thể được viết lại bằng công thức sau:

$$\mathrm{w}_g = rel_{NN}(g) = \frac{\sum_{i \in R} d_{\min}^{g}(X_i, R)}{\sum_{i \in R} d_{\min}^{g}(x_i, R) + \sum_{i \in R} d_{\min}^{g}(x_i, N)} \tag{2.24}$$

trong đó $d_{\min}^{g}(\square)$ là một hàm trả về khoảng cách tối thiểu giữa hai hình ảnh có cùng độ đo trong tập đặc trưng g-th của một không gian đặc trưng cố định.

Theo cách này hơn mỗi một hình ảnh có liên quan thì cách xa hình ảnh không liên quan gần nhất với nó và gần với hình ảnh có liên quan gần nhất, phần lớn trọng số được gán cho đặc trưng hoặc (tập đặc trưng) sử dụng để đánh giá khoảng cách. Trong thực tế hai hình ảnh có liên quan với nhau thì khoảng cách là nhỏ (và lớn hơn khoảng cách từ những hình ảnh không liên quan) xung quanh các thành phần cố định hoặc (tập của các thành phần) của

không gian đặc trưng, lớn hơn thì xác suất tìm ra những hình ảnh liên quan khác xung quanh các thành phần cố định hàng xóm của chúng hoặc (tập của các thành phần). Vì vậy một trọng số lớn hơn sẽ được gán cho các đặc trưng đó hoặc (tập các đặc trưng). Nó có thể dễ dàng nhận ra rằng khi khoảng cách giữa những hình ảnh có liên quan là lớn và khi khoảng cách giữa những hình ảnh có liên quan và không liên quan là nhỏ, $w_{f_x} \approx 1$. Mặt khác, khi những hình ảnh có liên quan là cụm lại trong một miền của không gian đặc trưng và những hình ảnh không liên quan nằm ngoài miền có liên quan, thì $w_{f_x} \approx 0$.

2.3 Kết hợp nhiều đặc trưng dựa trên SVM và phản hồi liên quan

2.3.1 Kỹ thuật máy học (SVM)

SVM đã được giới thiệu đầu tiên bởi Vapnik vào cuối những năm 90 và đến nay vẫn còn được quan tâm bởi cộng đồng nghiên cứu học máy. Với nền tảng lý thuyết mạnh mẽ và chặt chẽ, nó đang được sử dụng cho nhiều ứng dụng và là một phương pháp học mẫu nhỏ phổ biến có hiệu năng tốt cho bài toán phân loại mẫu. Giả sử có một tập n mẫu được gán nhãn $L = \{(x_1, y_1), \ldots, (x_l, y_l)\}$, với $x \in \mathbb{R}^d$ biểu diễn một ảnh trong không gian *d* chiều và $y \in \{1, -1\}$ là các nhãn. Ý tưởng chính của SVM là tìm siêu phẳng

$$f(x) = (w.x) + b \qquad (2.25)$$

để chia tách các điểm có y_i=1 và các điểm có y_i=-1 sao cho siêu phẳng phân cách có lề cực đại trong khi tỷ lệ lỗi phân lớp là nhỏ nhất. Đây là bài toán quy hoạch toàn phương và nó có thể được giải từ việc tìm w và b sao cho cực tiểu hóa hàm

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \qquad s.t.\ y_i(w.x_i + b) \geq 1 - \xi_i,\ \ \xi_i \geq 0\ ,\ \ i = 1 \ldots n \qquad (2.26)$$

Nếu viết điều kiện phân loại dưới dạng đối ngẫu thì bài toán đối ngẫu của SVM chính là bài toán tối ưu tìm các tham số α_i ($i=1..n$) để cực đại hóa hàm

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i\ x_j) \tag{2.27}$$

sao cho $$\sum_{i=1}^{n} y_i \alpha_i = 0,\ 0 \le \alpha_i \le C,\ \ i = 1 \ldots n \tag{2.28}$$

Ở đây $K(x_i.x_j)$ là hàm kernel. Khi đó hàm phân lớp SVM sẽ là :

$$f(x) = sign\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i\ x_j) + b \right) \tag{2.29}$$

và đường bao quyết định sẽ là: $\sum_{i=1}^{l} \alpha_i y_i K(x_i\ x_j) + b = 0$

Trong tra cứu ảnh với phản hồi liên quan dựa trên SVM, đường bao quyết định đã được sử dụng để đo sự liên quan giữa ảnh truy vấn và các mẫu đưa vào. Nói chung, các mẫu có giá trị tuyệt đối của hàm phân lớp càng lớn thì khả năng tin cậy dự đoán sẽ cao.

2.3.2 Cập nhật trọng số đặc trưng dựa trên phản hồi liên quan

Mục đích của việc cập nhật trọng số đặc trưng [12] là để nhấn mạnh các thông số phân biệt nhất. Trong thực tế, ý tưởng là thực hiện một sự lựa chọn đặc trưng luôn thay đổi theo sự phản hồi liên quan. Thuật toán cập nhật trọng số đặc trưng sử dụng trong công việc này là tương tự như một trong những đề xuất trong (*Mattia and Francesco, 2010)* và (*Wu and Zhang, 2002)* và được dựa trên một tập hợp các đặc tính thống kê.

Đối với một hình ảnh truy vấn được đưa ra với sự tối ưu hóa các điểm truy vấn q^k và sự phản hồi có liên quan được lặp đi lặp lại sau lần lặp thứ k

một tập các hình ảnh $I^k = \left\{ i_i^k, ..., i_N^k \right\}$ được thu hồi. Mỗi hình ảnh $i_n^k \in I^k$ được đại diện bởi m các đặc trưng $f\frac{n}{k} = f_{n,1}^k, ..., f_{n,M}^k$

Hãy để tập các hình ảnh có liên quan và không liên quan sau k^{th} phản hồi được lặp đi lặp bởi R^k và U^k theo thứ tự, trong đó $I^k = R^k \cup U^k$. Đối với tất cả các hình ảnh trong R^k, chúng ta sắp xếp thành phần s^{th} của các đặc trưng m vào tập $F_{m,s}^{k,U}$. Phạm vi ảnh hưởng lớn của R^k trên trục của thành phần s^{th} của đặc trưng m được định nghĩa là:

$$\Phi_{m,s}^k = \left[\phi_{m,s}^{k,1}, \phi_{m,s}^{k,2} \right] \tag{2.30}$$

với $\phi_{m,s}^{k,1} = Min(F_{m,s}^{k,R})$; $\phi_{m,s}^{k,2} = Max(F_{m,s}^{k,R})$ nơi Min và Max là các hàm trả về giá trị tối thiểu và tối đa của một tập, theo thứ tự $\Phi_{m,s}^k$ là phạm vi trên trục của thành phần s^{th} của của đặc trưng m kéo dài bởi những hình ảnh có liên quan sau khi k^{th} được lặp lại.

Tập hỗn độn các thành phần s^{th} của các đặc trưng m sau lần lặp thứ k^{th} được xác định bằng công thức:

$$\Psi_{m,s}^{k,u} = \left\{ \forall f_{m,s}^k \mid f_{m,s}^k \in \Phi_{m,s}^k \text{ và } f_{m,s}^k \in F_{m,s}^{k,U} \right\} \tag{2.31}$$

Các thiết lập nhầm lẫn là tập hợp con của $F_{m,s}^{k,U}$ rơi vào trong phạm vi chi phối $\Phi_{m,s}^k$ sau sự lặp đi lặp lại k^{th}. Tỷ lệ biểu thức của thành phần s^{th} với đặc trưng của m được định nghĩa là:

$$\delta_{m,s}^k = 1 - \frac{\sum_{l=1}^{k} \mid \Psi_{m,s}^{l,U}}{\sum_{l=1}^{k} \mid F_{m,s}^{l,U}} \tag{2.32}$$

Tỷ lệ biểu thức cho thấy tỷ lệ của hình ảnh không liên quan nằm ngoài phạm vi ảnh hưởng trên tất cả hình ảnh không thích hợp, và nó cho thấy khả năng của các thành phần s của đặc trưng m trong việc tách hình ảnh không liên quan từ những hình ảnh có liên quan. Biểu thị độ lệch chuẩn của $F_{m,s}^{k,R}$ là $\delta_{m,s}^{k,R}$ thì trọng số $w_{m,s}$ được cập nhật bằng công thức:

$$\mathrm{w}_{m,s}^{k} = \frac{\delta_{m,s}^{k}}{\sigma_{m,s}^{k,R}} \tag{2.33}$$

Phản hồi có liên quan có thể hội tụ chỉ với vài hoặc thậm chí chỉ là một hình ảnh có liên quan. Điều này là do các trọng số đặc trưng đang bị mắc kẹt trong một trạng thái tối ưu, có thể được phát hiện bởi các điều kiện sau:

- $| R^k | = 1$, với mọi $k \geq 1$
- $R^k = R^{k-1}$ và $|R^k| < \gamma$ với mọi $k > 1$, trong đó γ là một ngưỡng xác định trước.

Khi một trong hai điều kiện trên được thỏa mãn, các trọng số trặc trưng bị kẹt trong một tình trạng ở một điểm cực thuận (*Mattia and Francesco, 2010; Wu and Zhang, 2002*). Để đẩy trọng số đặc trưng ra khỏi tình trạng tối ưu nhất, sử dụng một độ đo "*disturbing factor"* từ sự phân tán của từ các hình ảnh có liên quan và những hình ảnh không liên quan. Trong thực tế, hình ảnh không liên quan có xu hướng đa mô hình (multimodel), nhưng chúng ta đơn giản hóa chúng bằng cách đưa chúng về một lớp từ khi chúng có xu hướng khôi phục quá trình cập nhật trọng số đặc trưng khi nó đang bị mắc kẹt.

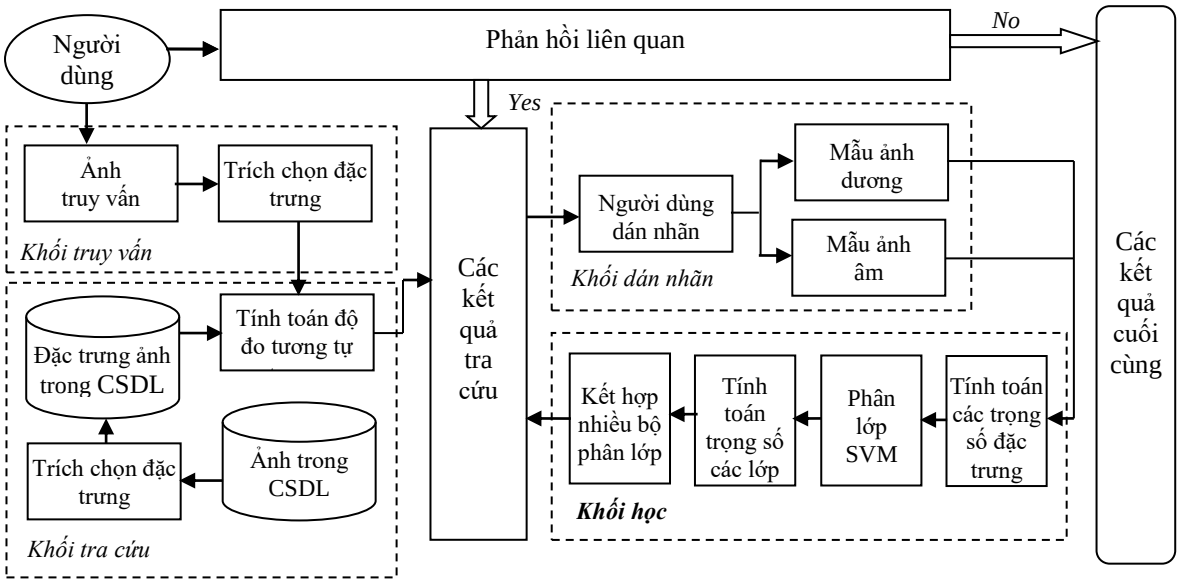
Biểu thị các giá trị trung bình của $F_{m,s}^{k,R}$ và $F_{m,s}^{k,U}$ bởi $\mu_{m,s}^{k,R}$ và $\mu_{m,s}^{k,U}$ tương ứng, và cho phép độ lệch chuẩn là $\sigma_{m,s}^{k,U}$. "*Disturbing factor"* được cho bởi công thức:

$$\lambda_{m,s}^{k} = \frac{\left(\mu_{m,s}^{k,R} - \mu_{m,s}^{k,U}\right)^2}{\left(\sigma_{m,s}^{k,R}\right)^2 + \left(\sigma_{m,s}^{k,U}\right)^2} \quad (2.34)$$

Công thức trên là tiêu chuẩn "Fisher" (Wu và Zhang, 2002 ; Fukunage, 1990), đã được sử dụng rộng rãi trong việc đo lường phân tán giữa hai lớp. Trọng số $\omega_{m,s}^{k}$ sau đó được cập nhật bởi công thức:

$$\omega_{m,s}^{k} = \lambda_{m,s}^{k} \times \omega_{m,s}^{k} \quad (2.35)$$

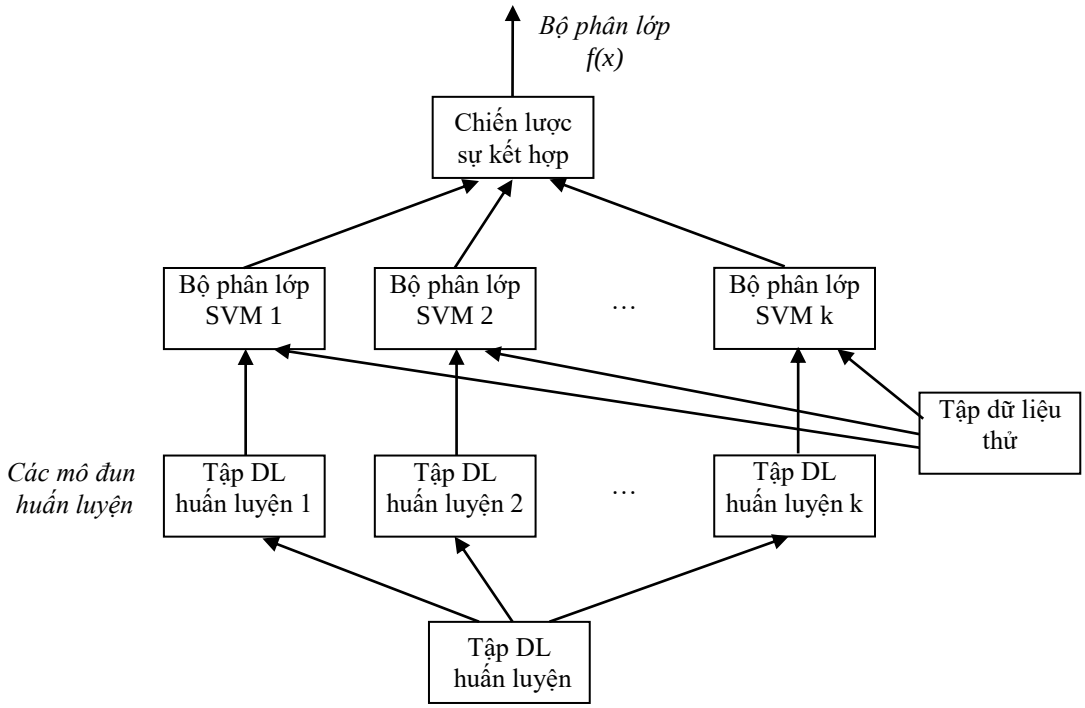
Sau mỗi lần sự phản hồi được lặp lại, các vector đặc trưng của ảnh truy vấn được thiết lập lại là tập các giá trị trung bình của các vector đặc trưng có liên quan.



Hình 2.3 Sơ đồ hệ thống tra cứu ảnh sử dụng phản hồi liên quan [12]

2.3.3 Kết hợp nhiều bộ phân lớp SVM dựa trên RF

Trong những năm gần đây, nhiều nhà nghiên cứu đã chú ý nhiều đến sự kết hợp nhiều bộ phân lớp SVM (Kim et al., 2003). Hình 2.4 dưới đây chỉ ra một cấu trúc tổng thể về sự kết hợp này [12].



Hình 2.4. Một cấu trúc tổng thể của sự kết hợp nhiều bộ phân lớp SVM

Trong suốt giai đoạn huấn luyện, mỗi SVM riêng lẻ được huấn luyện độc lập bằng một tập hợp dữ liệu huấn luyện được tái tạo riêng của nó. Tất cả các thành phần SVM được hợp thành sẽ được tập hợp lại bằng chiến lược kết hợp khác nhau. Trong suốt giai đoạn thử nghiệm, mỗi mẫu thử nghiệm được áp dụng đến toàn bộ các SVM cùng một lúc và đạt được một quyết định chung dựa vào chiến lược kết hợp này.

Nhiều phương pháp xây dựng một sự kết hợp của các phân loại này được phát triển. Một điều quan trọng nhất trong việc này là mỗi một SVM riêng lẻ trở lên khác biệt với SVM khác càng nhiều càng tốt. Các nhu cầu này có thể được đáp ứng bằng cách sử dụng những kiểu cách huấn luyện khác nhau cho những SVM khác nhau. Một số những phương pháp đang được lựa chọn để huấn luyện ví dụ như là: “bagging”, “boosting”, “randomization”, “stacking” và “dagging etc”. Trong số những phương pháp này chúng ta đặt trọng tâm vào những phương pháp điển hình “Adaboosting” và “Majority

voting". "Adaboosting" là thuật toán tăng nhanh dựa trên nguyên lý cơ bản về giảm thiểu lỗi. Nó có thể thu được những kết quả phân loại tốt nhất bằng cách lặp đi lặp lại nhiều lần việc điều chỉnh những trọng số của các phân loại riêng lẻ, nhưng nó thường tốn nhiều thời gian. Một phương pháp khác là dựa trên kỹ thuật của "Majority voting" thu được những điểm số (scores) của các phân lớp riêng lẻ, và khi đó chọn lấy một phương pháp có điểm số cao nhất. Sự bất lợi của phương pháp bỏ phiếu "Majority voting" là các điểm số (scores) của các phân lớp riêng lẻ có thể rất giống nhau khi người dùng tập trung trên nhiều phân lớp tại cùng một thời điểm. Các đặc điểm chung của hai phương pháp trên là khả năng phân loại các phân lớp riêng lẻ của những mẫu dương và những mẫu âm là như nhau. Do đó để cải thiện hiệu suất của việc phân loại trên cần điều chỉnh những tham số của những mẫu dương và những mẫu âm.

Giả sử có n không gian đặc trưng có thể nhìn bằng mắt được, chúng ta xây dựng một cấu trúc phân loại SVM cho mỗi không gian đặc trưng thị giác trong mỗi lần phản hồi, việc phân loại được mô tả như sau:

$$P_i^t = C\big(S(t), X(i)\big), \;\; i = 1,2,...,n; \;\; t = 1,2,...,m \tag{2.36}$$

Trong đó t là thời gian của các thông tin phản hồi, $S(t)$ thể hiện quá trình huấn luyện trong tập của thông tin phản hồi của thời gian i^{th}, $X(i)$ là không gian đặc trưng trong i^{th}. Mỗi một hình ảnh trong CSDL ảnh sử dụng X như là những vector riêng thích hợp của nó và $P_i^t(x)$ biểu diễn khả năng có thể xảy ra của hình ảnh có liên quan trong không gian đặc trưng i^{th} này. Khả năng của hình ảnh có thể xảy ra này có liên quan sau ý kiến phản hồi thì được mô tả là:

$$P^t(x) = \omega_i P_i^t(x) \tag{2.37}$$

Trong đó ω_i là trọng số của các lớp tương ứng. Những mô tả dưới đây cho cách tính toán trọng số.

Dựa vào đặc trưng i có thể nhìn được bằng mắt đối với mỗi phân lớp, khi việc phân lớp càng chính xác hơn trong quá trình huấn luyện, nghĩa là lỗi huấn luyện càng nhỏ thì mức độ của các đặc trưng mà người dùng cần lại càng cao. Đối với những hình ảnh liên quan (hay không liên quan) trong tập huấn luyện thì giá trị của những hình ảnh có thể xảy ra là những hình ảnh có liên quan (hoặc không liên quan) phản ánh mức độ chính xác của việc phân lớp. Trọng số của lớp phân loại được biểu diễn là:

$$\omega_i = \sum_{\substack{x \in S(t) \\ y=1}} \alpha_i P_i^t(x) + \sum_{\substack{S \in x(t) \\ y=-1}} \beta_i (1 - P_i^t(x)) \qquad (2.38)$$

$$\alpha_i + \beta_i = 1$$

Khi x thuộc về hình ảnh có liên quan, y=1; ngược lại y=-1. α_i và β_i thể hiện tầm quan trọng của hình ảnh liên quan và không liên quan đối với bộ phân lớp i^{th}. α_i càng lớn thì việc phân lớp lại càng quan trọng hơn đối với những hình ảnh có liên quan, sự phân lớp của những hình ảnh có liên quan càng chính xác thì trọng số của các bộ phân lớp tương ứng càng lớn. Trường hợp đặc biệt, nếu $\alpha_i = 1$ nó hiển thị bộ phân lớp này chỉ để xem xét những hình ảnh có liên quan mà thôi.

Kết luận chương 2

Nhằm khắc phục những hạn chế của các hệ thống CBIR như đã nêu ở chương 1, việc kết hợp nhiều đặc trưng ảnh để xây dựng truy vấn sử dụng kỹ thuật phản hồi liên quan nâng cao hiệu quả của các truy vấn là vấn đề trong chương 2 này tập trung nghiên cứu. Kết quả đã trình bày được một số kỹ thuật phản hồi liên quan và các phương pháp kết hợp nhiều đặc trưng trong tra cứu ảnh dựa vào nội dung có sử dụng SVM và phản hồi liên quan.

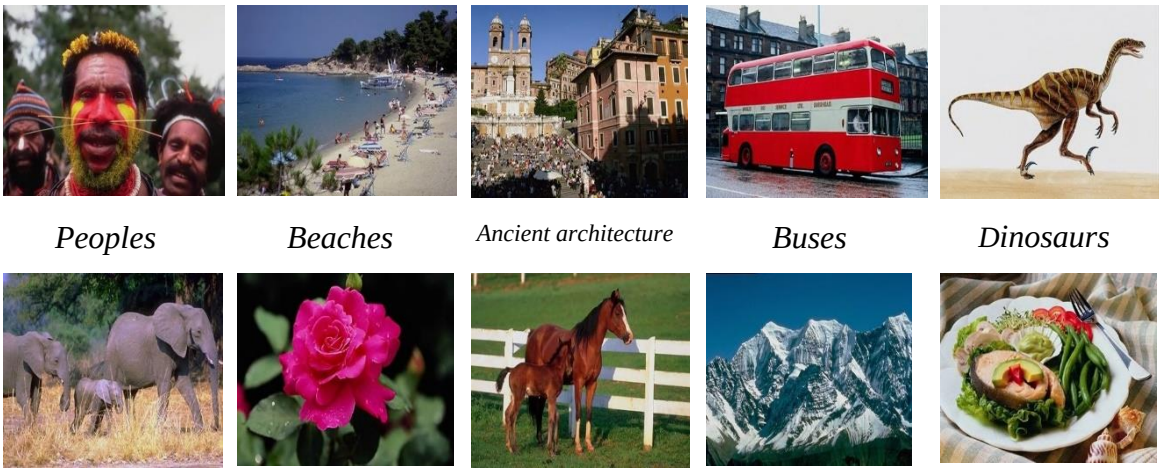
Để có thể đánh giá được hiệu quả của việc sử dụng từng đặc trưng ảnh như: đặc trưng mầu sắc, kết cấu, hình dạng và việc kết hợp cả 3 đặc trưng này để truy vấn ảnh sử dụng SVM và phản hồi liên quan trong CBIR, trong chương 3 sẽ viết chương trình thực nghiệm để so sánh, đánh giá các kết quả.

Chương 3. THỰC NGHIỆM

3.1 Môi trường thực nghiệm

3.1.1 Cơ sở dữ liệu

Các thực nghiệm trong chương trình được thực hiện với bộ công cụ Matlab 7.11.0 (R2010b). Tập dữ liệu sử dụng trong thử nghiệm là tập dữ liệu Wang bao gồm 1000 ảnh tự nhiên được tổ chức thành 10 thể loại khác nhau (10 lớp): *Peoples*, *Beaches*, *Ancient architecture*, *Buses*, *Dinosaurs*, *Elephants*, *Flowers*, *Horses*, *Mountains*, *Foods*. Mỗi lớp có 100 ảnh, tất cả các ảnh là ảnh mầu, định dạng *.jpg*, kích thước mỗi ảnh là 384 pixel × 256 pixel. Hình 3.1 minh họa một số ảnh mẫu trong tập dữ liệu này.



Hình 3.1. Các ảnh minh họa cho 10 thể loại trong tập ảnh Wang

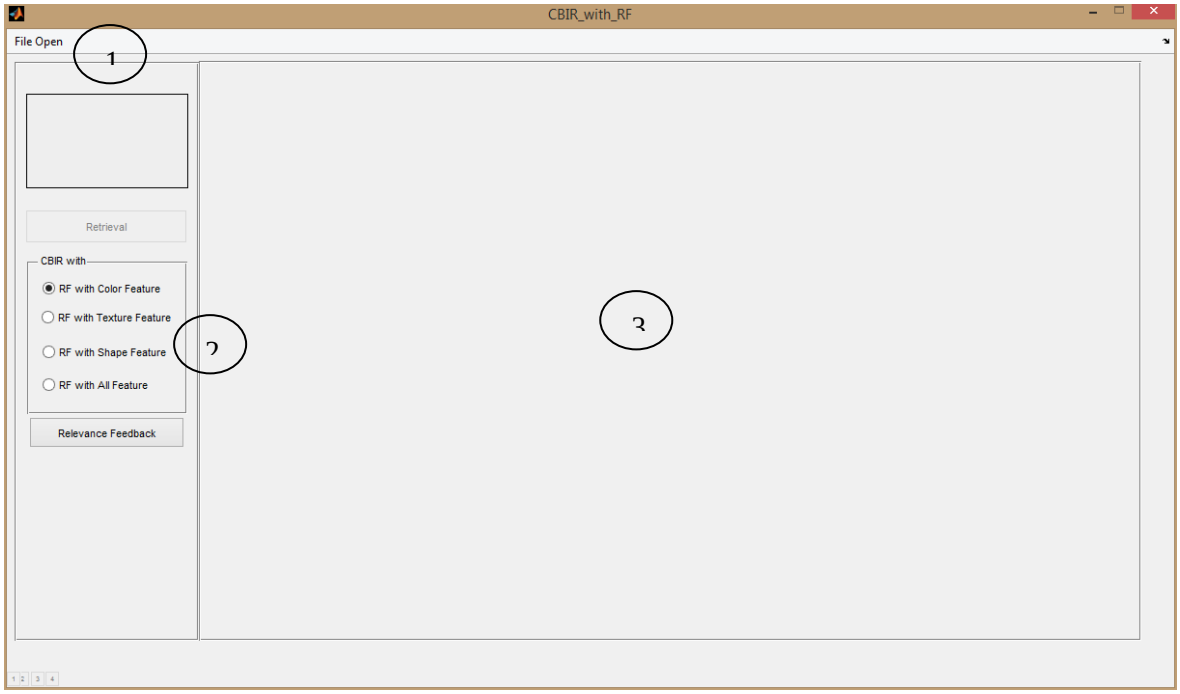
3.1.2 Trích chọn đặc trưng

Để trích chọn các đặc trưng của ảnh chương trình thực nghiệm trích chọn 3 kiểu đặc trưng: Mầu sắc, kết cấu, hình dạng.

Các đặc trưng được trích chọn Offline và được lưu vào tệp Wang1K.bin.

3.2 Mô tả chương trình thực nghiệm

3.2.1 Giao diện chương trình



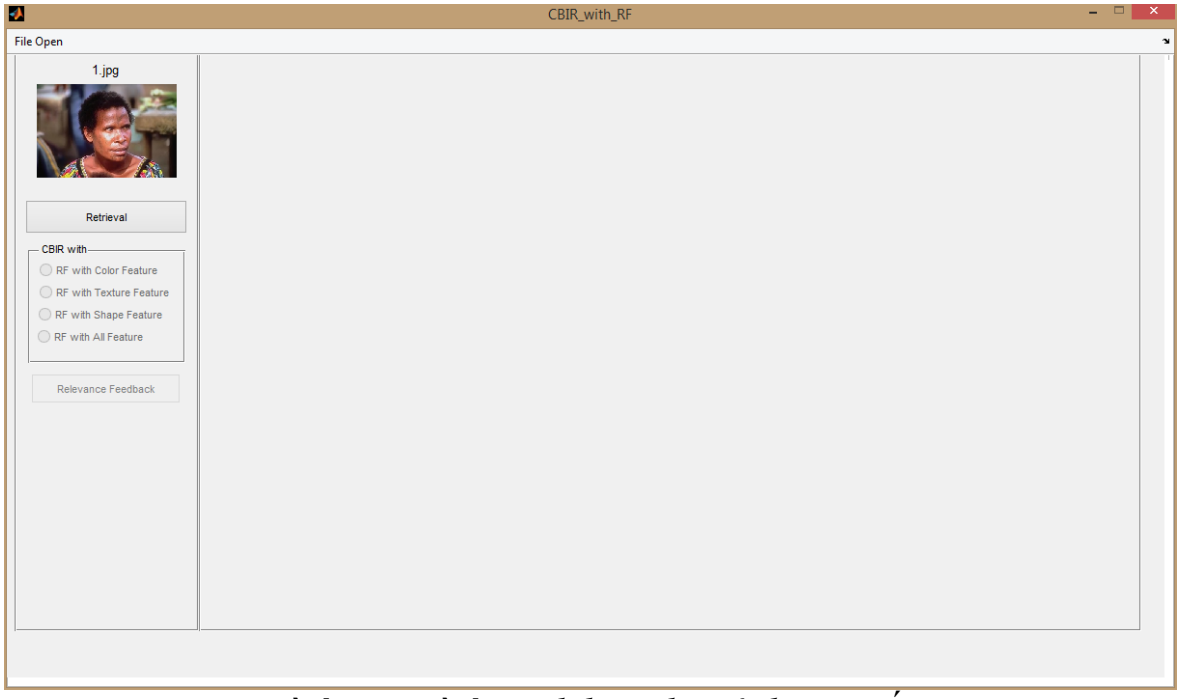
Hình 3.2. Hình ảnh giao diện chương trình thực nghiệm

- Vùng 1: Chọn, hiển thị ảnh truy vấn
- Vùng 2: Lựa chọn phương pháp truy vấn
- Vùng 3: Hiển thị hình ảnh để người dùng gán nhãn theo ngưỡng xác định đồng thời hiển thị kết quả trả về.

3.2.2 Các bước thực hiện truy vấn

Bước 1. Mở ảnh truy vấn

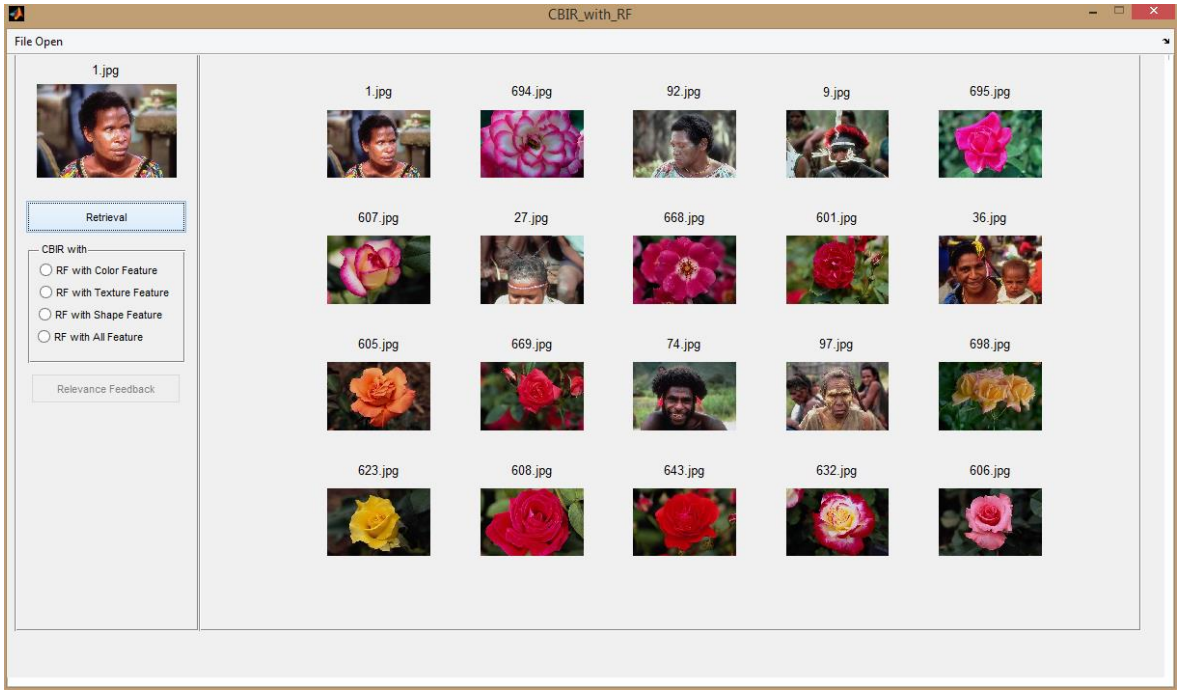
- Chọn ảnh truy vấn (lấy ảnh trong CSDL làm ảnh truy vấn) bằng cách chọn File -> Open trên Menu chức năng
- Chương trình tự động trích chọn đặc trưng ảnh truy vấn
- Hiển thị tên, ảnh truy vấn lên khung ảnh truy vấn.



Hình 3.3. Hình minh họa chọn ảnh truy vấn

Bước 2. Tra cứu ảnh

- Với ảnh truy vấn đã chọn, từ giao diện chương chọn nút Retrival (thực hiện truy vấn), chương trình tính toán khoảng cách giữa ảnh truy vấn với tất cả các ảnh trong CSDL bằng hàm ***distance***, sắp xếp theo thứ tự tăng dần của các khoảng cách đồng thời cập nhật chỉ số của các ảnh trong CSDL theo thứ tự khoảng cách đã sắp xếp, hiển thị 20 hình ảnh đầu tiên có khoảng cách nhỏ nhất lên vùng 3.

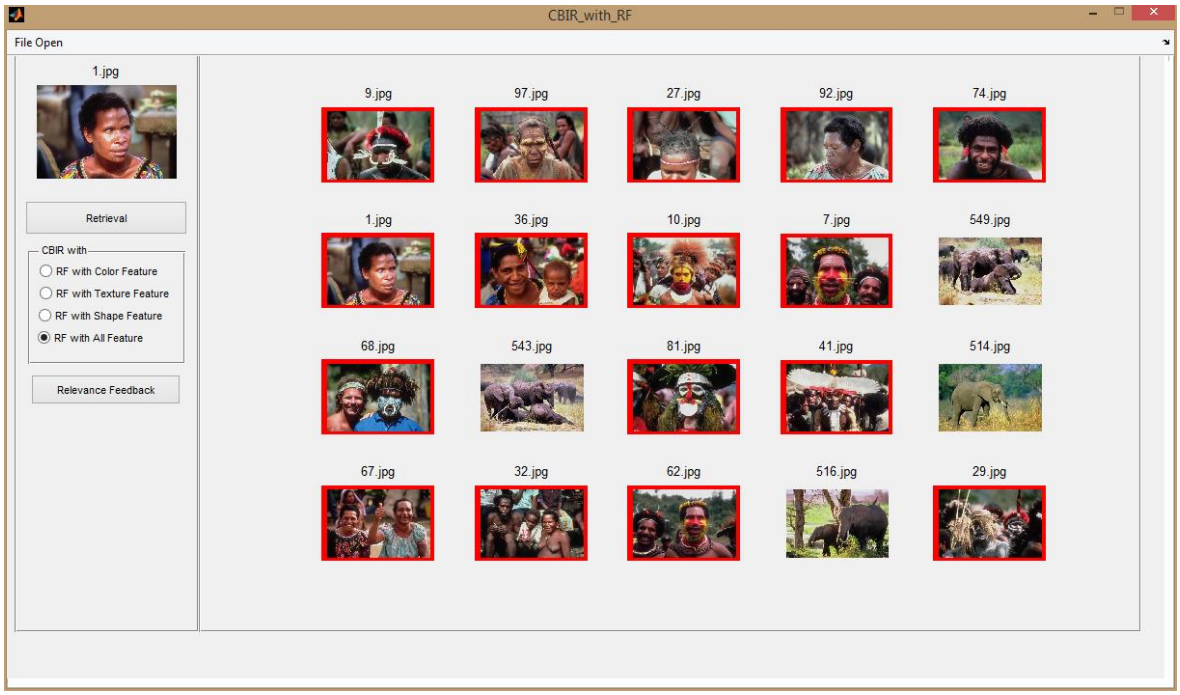


Hình 3.4. Hình minh họa sau khi chọn nút Retrival

Bước 3. Phản hồi liên quan

- Ảnh hiển thị trên giao diện ở vùng 3, nếu người dùng chưa hài lòng với kết quả truy vấn thì tiếp tục thực hiện việc kích chọn (gọi hàm selected) những ảnh có liên quan đến ảnh truy vấn (mẫu ảnh dương), số ảnh còn lại không được chọn tự động được hệ thống gán nhãn không liên quan (mẫu ảnh âm).
- Người dùng chọn một trong 4 phương pháp tra cứu ảnh ở khung CBIR_with sau đó chọn nút *Relevance Feedback* để thực hiện truy vấn ảnh. Hệ thống tự động tính toán tổng số (mẫu ảnh dương) người dùng gán nhãn sau các vòng lặp và tổng số (mẫu ảnh âm) hệ thống tự gán nhãn (các mẫu ảnh âm, ảnh dương được cộng dồn sau mỗi lần lặp). Các ảnh đã gán nhãn sau đó sử dụng hàm *svmtrain* và hàm *svmpredict* để huấn luyện mô hình phân lớp mới tìm ra giá trị quyết định của hàm phân lớp, sắp xếp theo chiều giảm dần của giá trị quyết định và hiểu thị hình ảnh kết quả.

- Quá trình này được lặp đi lặp lại cho đến khi người dùng hài lòng với kết quả tra cứu thì dừng.



Hình 3.5. Hình minh họa sau khi người dùng gán nhãn phản hồi liên quan

3.3 Đánh giá hiệu năng

Để đánh giá hiệu năng của hệ thống tra cứu, người ta có thể dựa trên các tiêu chí khác nhau. Trong khuôn khổ luận văn tác giả tập trung đánh giá độ chính xác trung bình (*Average Precicion*) của các phương pháp tra cứu và thời gian tính toán. Các thông số đo đạc này được lấy từ chương trình thực nghiệm trên 2 CSDL (*1000 ảnh*) và Oliva (*với 2688 ảnh*) để so sánh. Cụ thể như sau:

- Chương trình thực nghiệm được thiết kế chạy tự động trên mỗi CSDL riêng biệt sau đó ghi kết quả ra tệp để thực hiện so sánh, đánh giá sau này.
- Đối với mỗi CSDL ảnh, chương trình thực nghiệm trên cửa sổ chọn ảnh lần lượng là 5, 10, 15, 20 ảnh và tương ứng với số lượng ảnh trả

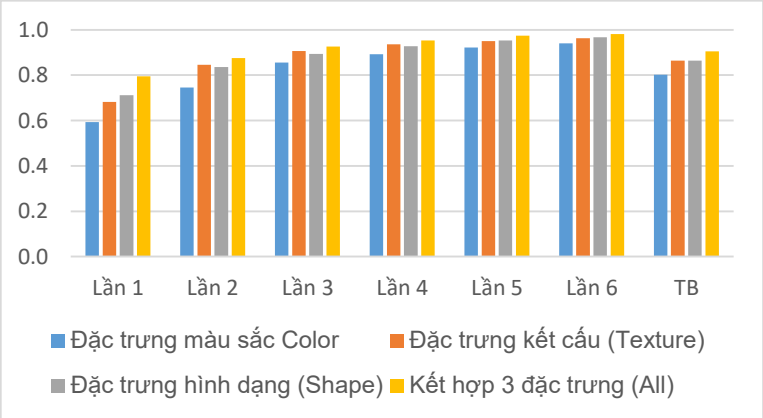
về khác nhau trong CSDL là 20, 40, 60 80, 100 ảnh qua 6 lần phản hồi để tính độ chính xác trung bình và thời gian thực hiện truy vấn

- Trong phần thực nghiệm này, độ đo Average Precision được định nghĩa bởi NISTTREC video sẽ được sử dụng để đánh giá hiệu năng của các phương pháp tra cứu: theo mầu sắc, kết cấu, hình dạng và kết hợp 3 đặc trưng trên
- Dưới đây là một số bảng kết quả và biểu đồ mô phỏng thực nghiệm chi tiết:

3.3.1 Thực nghiệm trên CSDL Wang

Bảng 1. So sánh độ chính xác trung bình của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn (20 ảnh) với số ảnh trả về 20, 40, 60, 80, 100 trên CSDL ***Wang*** qua 6 lần phản hồi

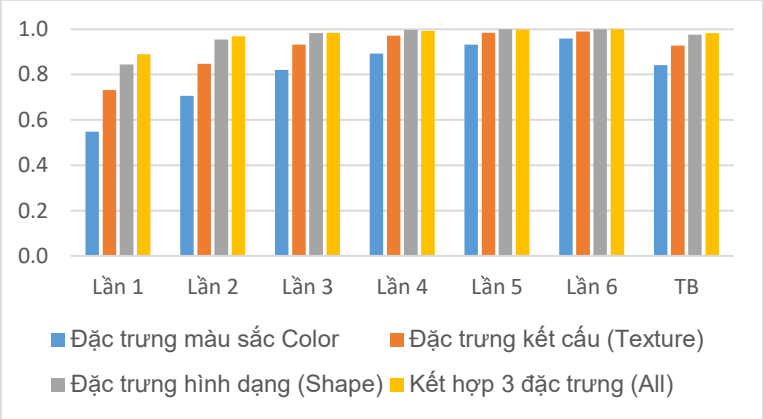
Phương pháp	*Kết quả trung bình sau các lần phản hồi*						***TB***
	1	2	3	4	5	6	
Đặc trưng màu sắc Color	59,29%	74,55%	85,62%	89,22%	92,13%	94,03%	80,16%
Đặc trưng kết cấu(Texture)	68,22%	84,58%	90,61%	93,56%	95,06%	96,25%	86,40%
Đặc trưng hình dạng(Shape)	71,10%	83,62%	89,35%	92,77%	95,35%	96,70%	86,44%
Kết hợp 3 đặc trưng (All)	79,56%	87,48%	92,56%	95,23%	97,46%	98,13%	90,46%



*Hình 3.6.. Kết quả truy vấn của các phương pháp thực nghiệm trên cỡ cửa sổ chọn (20 ảnh) với số ảnh trả về 20, 40, 60, 80, 100 trên CSDL **Wang** qua 6 lần phản hồi*

Bảng 2. So sánh độ chính xác trung bình của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn (20 ảnh) với số ảnh trả về 20, 40, 60, 80, 100 trên CSDL ***Oliva*** qua 6 lần phản hồi

Phương pháp	*Kết quả trung bình sau các lần phản hồi*						***TB***
	1	2	3	4	5	6	
Đặc trưng màu sắc Color	54,87%	70,58%	82,03%	89,21%	93,21%	95,88%	84,09%
Đặc trưng kết cấu (Texture)	73,16%	84,74%	93,17%	97,08%	98,35%	98,95%	92,70%
Đặc trưng hình dạng(Shape)	84,45%	95,42%	98,29%	99,64%	99,96%	100,00%	97,53%
Kết hợp 3 đặc trưng (All)	88,88%	96,89%	98,35%	99,31%	99,84%	100,00%	98,29%



Hình 3.7. Kết quả truy vấn của các phương pháp thực nghiệm trên cỡ cửa sổ chọn (20 ảnh) với số ảnh trả về 20, 40, 60, 80, 100 trên CSDL ***Oliva*** *qua 6 lần phản hồi*

Bảng 3. So sánh ***độ chính xác trung bình*** của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn (20 ảnh) với CSDL ***Wang*** và ***Oliva*** qua 6 lần phản hồi.

Phương pháp	***CSDL Wang*** *(Độ chính xác TB %)*	***CSDL Oliva*** *(Độ chính xác TB %)*
Đặc trưng màu sắc Color	80,2%	84,1%
Đặc trưng kết cấu (Texture)	86,4%	92,7%
Đặc trưng hình dạng (Shape)	86,4%	97,5%
Kết hợp 3 đặc trưng (All)	**90,5%**	**98,3%**

Bảng 4. So sánh thời ***gian tính toán trung bình*** của các phương pháp, thực nghiệm trên cửa sổ chọn (20 ảnh) với CSDL ***Wang*** và ***Oliva*** qua 6 lần phản hồi.

Phương pháp	***CSDL Wang*** *(thời gian)*	***CSDL Oliva*** *(thời gian)*
Đặc trưng màu sắc Color	0,021	0,059
Đặc trưng kết cấu (Texture)	0,036	0,082
Đặc trưng hình dạng (Shape)	0,125	0,275
Kết hợp 3 đặc trưng (All)	**0,198**	**0,445**

3.3.2 Thực nghiệm trên 2 CSDL Wang và Olivavới

Bảng 5. . So sánh độ chính xác trung bình của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn ảnh [5, 10, 15, 20] với số ảnh trả về [20, 40, 60, 80, 100] trên CSDL ***Wang và Oliva*** *qua 6 lần phản hồi*

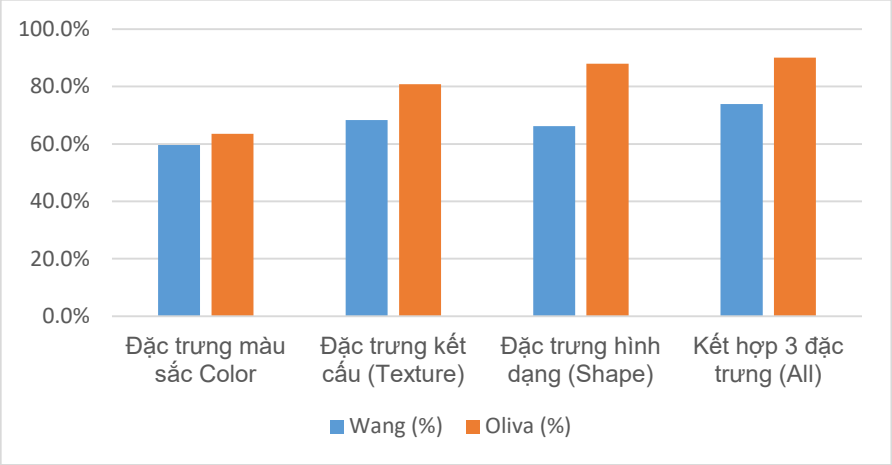
Cửa sổ chọn ảnh gán nhãn	Đặc trưng màu sắc *(Color)*		Đặc trưng kết cấu *(Texture)*		Đặc trưng hình dạng *(Shape)*		Kết hợp 3 đặc trưng *(All)*	
	Wang	*Oliva*	*Wang*	*Oliva*	*Wang*	*Oliva*	*Wang*	*Oliva*
5	59,6%	63,5%	68,3%	80,8%	66,2%	87,9%	73,9%	90,0%
10	69,1%	72,1%	76,4%	86,3%	76,5%	93,1%	82,7%	94,9%
15	76,1%	78,8%	82,6%	90,8%	82,6%	95,9%	87,2%	97,3%
20	80,2%	84,1%	86,4%	92,7%	86,4%	97,5%	90,5%	98,3%

*Bảng 6. So sánh **thời gian tính toán trung bình** của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn ảnh [5, 10, 15, 20] với số ảnh trả về [20, 40, 60, 80, 100] trên CSDL **Wang và Oliva** qua 6 lần phản hồi*

Cửa sổ chọn ảnh gán nhãn	Đặc trưng màu sắc *(Color)*		Đặc trưng kết cấu *(Texture)*		Đặc trưng hình dạng *(Shape)*		Kết hợp 3 đặc trưng *(All)*	
	Wang	*Oliva*	*Wang*	*Oliva*	*Wang*	*Oliva*	*Wang*	*Oliva*
5	0,008	0,019	0,013	0,032	0,042	0,107	0,063	0,164
10	0,012	0,032	0,019	0,050	0,069	0,168	0,103	0,265
15	0,017	0,047	0,028	0,071	0,102	0,226	0,150	0,372
20	0,021	0,059	0,036	0,082	0,125	0,275	0,198	0,445

*Bảng 7. Tổng hợp độ chính xác trung bình của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn ảnh [5, 10, 15, 20] với số ảnh trả về [20, 40, 60, 80, 100] trên CSDL **Wang và Oliva** qua 6 lần phản hồi*

Phương pháp	***Wang (%)***	***Oliva (%)***
Đặc trưng màu sắc Color	59,6%	63,5%
Đặc trưng kết cấu (Texture)	68,3%	80,8%
Đặc trưng hình dạng (Shape)	66,2%	87,9%
Kết hợp 3 đặc trưng (All)	**73,9%**	**90,0%**

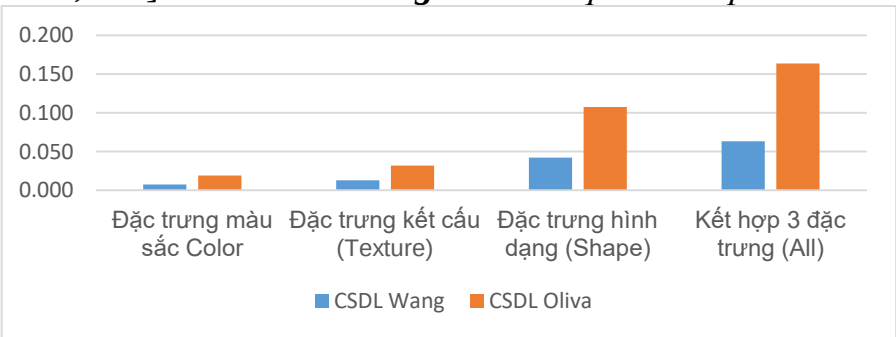


*Hình 3.8. Biểu đồ thể hiện độ chính xác trung bình của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn ảnh [5, 10, 15, 20] với số ảnh trả về [20, 40, 60, 80, 100] trên CSDL **Wang và Oliva** qua 6 lần phản hồi*

*Bảng 8. Thời gian tính toán trung bình của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn ảnh [5, 10, 15, 20] với số ảnh trả về [20, 40, 60, 80, 100] trên CSDL **Wang và Oliva** qua 6 lần phản hồi*

Phương pháp	***CSDL Wang***	***CSDL Oliva***
Đặc trưng màu sắc Color	0,008	0,019
Đặc trưng kết cấu (Texture)	0,013	0,032
Đặc trưng hình dạng (Shape)	0,042	0,107
Kết hợp 3 đặc trưng (All)	0,063	0,164

*Hình 3.9. Biểu đồ thể hiện thời gian trung bình của các phương pháp, thực nghiệm trên cỡ cửa sổ chọn ảnh [5, 10, 15, 20] với số ảnh trả về [20, 40, 60, 80, 100] trên CSDL **Wang và Oliva** qua 6 lần phản hồi*



Kết luận chương 3

Nhận xét, đánh giá các kết quả thực nghiệm:

Với mục đích so sánh, đánh giá độ chính xác, thời gian thực hiện truy vấn, trong chương 3 này luận văn đã xây dựng một chương trình thực nghiệm trên 2 tập dữ liệu *Wang (1000 ảnh)* và *Oliva (2688 ảnh)*. Các kết quả bước đầu có thể tóm tắt như sau:

- Kết quả thực nghiệm đã cho thấy mỗi phương pháp tra cứu ảnh dựa trên nội dung sử dụng SVM với phản hồi liên quan với các đặc trưng mầu sắc, kết cấu, hình dạng về cơ bản cho hiệu quả và độ chính xác không cao theo nhu cầu của người dùng. Nó chỉ hiệu quả khi người dùng có nhu cầu truy vấn những hình ảnh liên quan đến các đặc trưng đó, tuy nhiên thời gian thực hiện truy vấn lại rất nhanh.

- Phương pháp tra cứu ảnh kết hợp nhiều đặc trưng sử dụng SVM với phản hồi liên quan đạt hiệu quả cao, độ chính xác nhất so với các phương pháp trên, tuy nhiên lại mất nhiều thời gian hơn.

KẾT LUẬN

Tra cứu ảnh dựa trên nội dung đang là lĩnh vực được nhiều người quan tâm nghiên cứu và phát triển mạnh mẽ trong nước và nước ngoài. Nó cần phải nghiên cứu, phát triển mạnh hơn nữa mới có thể đáp ứng được nhu cầu ngày càng cao của người dùng trong thực tế.

Trong khuôn khổ của luận văn này tác giả tập trung tìm hiểu, nghiên cứu một số nội dung cơ bản của CBIR.

Các kết quả chính đạt được:

- Đã nắm được một số phương pháp trích chọn đặc trưng hình ảnh, một số phương pháp phản hồi liên quan trong tra cứu ảnh dựa vào nội dung
- Trình bày được phương pháp tìm kiếm hình ảnh theo đặc trưng mầu sắc, kết cấu, hình dạng và phương pháp kết hợp các đặc trưng trên áp dụng trong tra cứu ảnh theo nội dung sử dụng SVM và phản hồi liên quan.
- Đã viết được chương trình thực nghiệm, thực hiện tìm kiếm ảnh theo đặc trưng mầu sắc, kết cấu, hình dạng và kết hợp 3 đặc trưng trên sử dụng SVM và phản hồi liên quan, sử dụng bộ công cụ và thư viện Matlab. Chương trình được chạy thực nghiệm trên 2 CSDL Wang và Oliva và đã so sánh, đánh giá được độ chính xác và thời gian thực hiện tìm kiếm ảnh của các phương pháp trên.

Một số vấn đề cần tiếp tục giải quyết

Vấn đề nghiên cứu bước đầu đã đạt được một số kết quả khả quan trên tập dữ liệu ảnh đã thử nghiệm, nhưng đối với các truy vấn cấp cao thì chưa áp dụng vào. Hơn nữa, vấn đề thời gian trong truy vấn ảnh cũng cần được quan tâm khi thư viện ảnh của hệ thống được mở rộng

Hướng nghiên cứu tiếp theo

Trong thời gian tới, ngoài việc tiếp tục giải quyết các vấn đề còn tồn tại, tôi định hướng một số nghiên cứu tiếp theo:

- Truy vấn ảnh dựa theo theo vùng, đối. Ảnh gồm tập hợp các vùng hay còn gọi là vần. Đây là đặc trưng cấp cao của ảnh. Với đặc trưng vùng sẽ giúp cho chúng ta có thể giải quyết được một vấn đề lớn đang cản trở bước phát triển việc truy tìm ảnh dựa vào nội dung là dữ liệu nhập chưa được mô phỏng gần gũi hơn với suy nghĩ của con người và ảnh tìm được có thể mang nội dung ngữ nghĩa rất khác so với ảnh truy vấn.

- Với đặc trưng vùng, con người có thể tiến thêm một bước trong việc truy tìm ảnh dựa vào nội dung là tìm kiếm dựa vào ngữ nghĩa. Với việc áp dụng mô hình học vào bài toán vùng. Khi đó, mỗi vùng sẽ mang một ngữ nghĩa, từ đó làm cho dữ liệu đầu vào được mô phỏng gần gũi với con người hơn.

- Quá trình phân đoạn vùng của ảnh đòi hỏi phải tốn nhiều thời gian, do đó khi tìm kiếm trên một cơ sở dữ liệu ảnh lớn thì vấn đề thời gian là một trong những vấn đề gây khó khăn cho bài toán, do đó cần phải có biện pháp tổ chức cơ sở dữ liệu hiệu quả giúp cho việc tìm kiếm được nhanh hơn, hiệu quả hơn

- Truy vấn theo ngữ nghĩa dẫn đến một khả năng truy tìm mới là dựa vào câu truy vấn dạng ngôn ngữ, liên kết giữa các vùng đại diện theo một lý luận để truy tìm ảnh trong từng nhóm ảnh đã được phân loại bằng tay, trích ra các vùng đặc thù cho nhóm ảnh để khi truy vấn ta có thể diễn dịch từ ngôn ngữ sang vần ảnh. Ví dụ như tìm cảnh bãi biển lúc bình minh, thì liên kết nhóm từ bãi biển với vùng đại diện cho bãi biển và liên kết nhóm từ bình minh với vùng đại diện cho bình minh. Đây có thể xem là hướng truy tìm ảnh

theo phương pháp mới. Với phương pháp này ta có thể truy tìm ảnh chỉ bằng những câu chữ mà không cần hình ảnh có sẵn. Việc tìm kiếm này đem đến sự tiện dụng cho người dùng;

- Ngoài ra còn có một vấn đề lớn đó chính là truy vấn ảnh động (phim ảnh). Đây cũng là một lĩnh vực rất được quan tâm trong nước và trên thế giới. Điều khác biệt giữa ảnh tĩnh và ảnh động đó chính là dung lượng. Chính phần dung lượng đã ảnh hưởng đến thời gian truy vấn ảnh cũng như tiêu tốn chi phí cho phần cứng để quản lý và lưu trữ ảnh tĩnh.

TÀI LIỆU THAM KHẢO

1. C. G. M. Snoek, M. Worring, and A.W.M. Smeulders, *Early versus late fusion in semantic video analysis*. November, 2005, In ACM International Conference on Multimedia: Singapore, pages. 399-402.
2. D. N. F.Awang Iskandar, James A.Thom, and S.M.M. Tahaghoghi, *Content-based Image Retrieval Using Image Regions as Query Examples*. 2008: CRPIT Volume 75- Database technologies
3. Deng, Y., et al., *An efficient color representation for image retrieval*. 2001, IEEE Trans. on Image Processing, 10, pages. 140-147.
4. Dr Fuhui Long, Dr Hongjiang Zhang, and P.D.D. Feng, *Fundamentals of content-based image retrieval*. 2012: International journal of computer science and information technologies, 3, pages. 3260 - 3263.
5. G. Pass and R. Zabith, *Histogram refinement for content-based image retrieval*. 1996, IEEE Workshop on Applications of Computer Vision, pages. 96-102.
6. Giacinto, G., *A nearest-neighbor approach to relevance feedback in content based image retrieval*. 2007, In CIVR ’07: Proceedings of the 6th ACM international conference on Image and video retrieval: New York, NY, USA, pages. 456–463.
7. Jing Peng, Bir Bhanu, and S. Qing, *Probabilistic feature relevance learning for content-based image retrieval*. July/August 1999, Computer Vision and Image Understanding, pages. 150–164.

8. Luca Piras and G. Giacinto, *Neighborhood-based feature weighting for relevance feedback in content-based retrieval*. 2009, IEEE Computer Society: In WIAMIS, pages. 238–241.

9. LucaPiras, *Interactive search techniques for content-based retrieval from archives of images*, in *Electronic and Computer Engineering*. 2011: Electrical and Electronic Engineering University of Cagliari, pages. 63-68.

10. Ma, W.-Y.a.M., B. S, *Netra: A toolbox for navigating large image databases*. 1997, In Proc. of IEEE Int. Conf. on Image Processing, 1, pages. 568-571.

11. Quynh, N.H.a.T., N. Q., Giang, N. T, *A efficient method for content based image retrieval using histogram graph*. 2008, In Proc. of IEEE on Control, Automation, Robotics and Vision, pages. 874-878.

12. Xiang-Yang Wang, B.-B.Z., Hong-Ying Yang, *Active SVM-based relevance feedback using multiple classifiers ensemble and features reweighting*. 2012: School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China, pages.

13. Zhang, J., *Robust content-based image retrieval of multiexample queries.*, in *Doctor of Philosophy thesis*. 2011, School of Computer Science and Software Engineering: University of Wollongong, pages.