

MỤC LỤC

MỤC LỤC	1
LỜI CẢM ƠN.....	3
LỜI MỞ ĐẦU	4
CHƯƠNG 1: TỔNG QUAN VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG	6
1.1. Những thành phần của một hệ thống tra cứu ảnh.....	6
1.1.1 Công nghệ tự động trích chọn metadata	6
1.1.2 Giao diện để lấy chương trình truy vấn của người sử dụng.....	6
1.1.3 Phương pháp để so sánh độ tương tự giữa các ảnh.....	6
1.1.4 Công nghệ tạo chỉ số và lưu trữ dữ liệu hiệu quả	7
1.2. Đặc điểm tra cứu ảnh.....	8
1.3. Những ứng dụng cơ bản của tra cứu ảnh.....	10
1.4. Tra cứu ảnh dựa trên nội dung	11
1.4.1 Những phương pháp quản lý dữ liệu ảnh truyền thống	11
1.4.2 Các chức năng của hệ thống tra cứu ảnh dựa trên nội dung	12
1.4.3 Trích chọn những đặc điểm.....	15
1.4.4 Những khoảng cách tương ứng.....	18
1.4.5 Các phương pháp tra cứu ảnh dựa trên nội dung	22
1.5. Những hệ thống tra cứu ảnh dựa trên nội dung.....	26
1.5.1 Hệ thống QBIC(Query By Image Content)	26
1.5.2 Hệ thống PhotoBook	27
1.5.3 Hệ thống VisualSEEK và WebSEEK	27
1.5.4 Hệ thống RetrievalWare.....	27
1.5.5 Hệ thống Imatch.....	28
CHƯƠNG 2: TRA CỨU ẢNH DỰA TRÊN NỘI DUNG	29
2.1. Không gian màu	29
2.1.1 Không gian màu RGB.....	29
2.1.2 Không gian màu HSx.....	31
2.1.3 Không gian màu YUV và YIQ	32

2.1.4	Không gian ma trận CIEXYZ và LUV	32
2.2.	Biểu đồ màu.....	32
2.3.	Lượng tử hóa màu	33
2.4.	Thước đo khoảng cách biểu đồ màu.....	34
2.4.1	Thước đo khoảng cách Minkowski.....	35
2.4.2	Thước đo khoảng cách Quadratic	36
2.4.3	Thước đo khoảng cách Non-histogram.....	37
2.5.	Tra cứu ảnh dựa trên biểu đồ màu.....	38
2.5.1	Phương pháp truyền thống dựa trên màu sắc.....	38
2.5.2	Phương pháp Harbin	40
2.5.3	Sự nâng cấp phương pháp Harbin.....	45
2.6.	Cải tiến hiệu quả tra cứu.....	49
CHƯƠNG 3: CHƯƠNG TRÌNH THỬ NGHIỆM		51
3.1.	Bài toán.....	51
3.2.	Lựa chọn công cụ	51
3.3.	Một số kết quả chương trình.....	52
3.3.1	Giao diện chương trình	52
3.3.2	Kết quả	53
KẾT LUẬN		54
TÀI LIỆU THAM KHẢO		55

LỜI CẢM ƠN

Để có thể hoàn thành được đề án tốt nghiệp này, em đã được học hỏi những kiến thức quý báu từ các thầy, cô giáo của Trường Đại Học Dân Lập Hải Phòng trong suốt bốn năm đại học. Em vô cùng biết ơn sự dạy dỗ, chỉ bảo tận tình của các thầy, các cô trong thời gian học tập này.

Em xin bày tỏ lòng biết ơn tới thầy Ngô Trường Giang - Khoa công nghệ thông tin – Trường Đại Học Dân Lập Hải Phòng đã tận tình chỉ bảo và định hướng cho em nghiên cứu đề tài này. Thầy đã cho em những lời khuyên quan trọng trong suốt quá trình hoàn thành đề án. Cuối cùng, em xin cảm ơn gia đình và bạn bè luôn tạo điều kiện thuận lợi, động viên và giúp đỡ em trong suốt thời gian học tập, cũng như quá trình nghiên cứu, hoàn thành đề án này.

Do hạn chế về thời gian thực tập, tài liệu và trình độ bản thân, bài đề án của em không thể tránh khỏi những thiếu sót, rất mong các thầy cô góp ý và sửa chữa để bài đề án tốt nghiệp của em được hoàn thiện hơn. Em xin chân thành cảm ơn!

Hải Phòng ... tháng ... năm 2010

Sinh viên

Phạm Duy Thành

LỜI MỞ ĐẦU

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 tra cứu thông tin như là một hệ thống lưu trữ và tra cứu thông tin. Như là một hệ thống, 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 có nghĩa là 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 (VIR-Visual Information Retrieval). Mặc dù vậy vẫn có sự phân biệt giữa kiểu của thông tin và nét tự nhiên của tra cứu của văn bản và các đối tượng trực quan. Thông tin kết cấu là tuyến tính trong khi ảnh là hai chiều và video là ba chiều. Một cách chính xác hơn là văn bản được cung cấp với một điểm bắt đầu và kết thúc vốn có và với một chuỗi phân tích cú pháp tự nhiên. Chiến lược phân tích cú pháp tự nhiên như vậy không thích hợp với ảnh và video.

Có hai phương pháp chung 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 dựa vào tra cứu 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 metadata về đối tượng trực quan và sự chú thích kết cấu. Thật không may là việc phân tích kết cấu 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 truy cập ảnh và video dựa trên text đã thúc đẩy quan tâm đến sự 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 chủ yếu dựa trên kết quả của đồ hoạ máy tính.

Tra cứu ảnh được ứng dụng trong rất nhiều lĩnh vực, những lĩnh vực thành công bao gồm: ngăn chặn tội phạm, quân sự, quản lý tài sản trí tuệ, thiết kế kiến trúc máy móc, thiết kế thời trang và nội thất, báo chí quảng cáo, chuẩn đoán y học ... Nhận biết được sự quan trọng của nhận dạng ảnh nên khóa luận này em muốn “Tìm hiểu tra cứu ảnh dựa trên biểu đồ màu”. Trong bài đồ án này, bản luận của tôi sẽ tập trung vào một số đặc điểm cụ thể đặc biệt là những đặc điểm dựa trên màu sắc và kết cấu ứng dụng cho tra cứu ảnh nói chung hoặc cho tra cứu ảnh dựa trên nội dung. Mặc dù vậy không có không có đặc điểm riêng lẻ nào tốt nhất có thể cho ra những kết quả chính xác trong bất kỳ một thiết lập chung nào. Một kết hợp thông thường của các đặc điểm là cần thiết để cung cấp những kết quả tra cứu thích đáng đối với ứng dụng tra cứu ảnh dựa trên nội dung. Nội dung khóa luận bao gồm, Phần mở đầu, Phần kết luận và 3 chương nội dung, cụ thể:

Chương I : Tổng quan về tra cứu ảnh

Chương này nói về các kiến thức cơ bản về tra cứu ảnh

Chương II : Tra cứu ảnh dựa trên biểu đồ màu

Chương này đi vào chi tiết tra cứu ảnh dựa trên biểu đồ màu

Chương III: Chương trình thử nghiệm

CHƯƠNG 1: TỔNG QUAN VỀ TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

1.1. Những thành phần của một hệ thống tra cứu ảnh

1.1.1 Công nghệ tự động trích chọn metadata

Mỗi đặc điểm nguyên thủy của ảnh có định dạng đặc trưng của nó như biểu đồ màu được sử dụng rộng rãi để biểu thị đặc điểm màu sắc. Một ví dụ khác đặc điểm hình dạng có thể biểu thị bằng một tập các đoạn biên liền nhau. Với metadata thích hợp hệ thống tra cứu ảnh dựa trên nội dung có thể tra cứu ảnh bởi màu sắc, hình dạng, kết cấu và bởi sự kết hợp các đặc tính trên.

1.1.2 Giao diện để lấy chương trình truy vấn của người sử dụng

Trong bất kỳ một hệ thống tra cứu nào thì quá trình tra cứu đều bắt đầu từ một yêu cầu tra cứu. Vì vậy, nó là vấn đề cốt yếu để lấy yêu cầu truy vấn của người sử dụng một cách chính xác và dễ dàng. Với hệ thống tra cứu ảnh dựa trên nội dung thì quá trình tra cứu thường được thực hiện thông qua một hình ảnh mẫu được cung cấp bởi người sử dụng gọi là truy vấn bởi mẫu. Mặc dù vậy người sử dụng không thể luôn luôn đưa ra một ảnh mẫu cho hệ thống tra cứu. Hệ thống tra cứu ảnh dựa trên nội dung hiện nay giải quyết vấn đề này bằng cách đưa ra một giao diện để chỉ định hoặc chọn một số đặc điểm cơ bản cho việc cung cấp ảnh mẫu. Chẳng hạn như khi sử dụng hệ thống QBIC của IBM người sử dụng có thể chỉ định truy vấn đặc điểm màu sắc bằng cách chọn ra số lượng thành phần RED, BLUE, GREEN liên quan hoặc là có thể lựa chọn màu sắc ảnh mong muốn từ bảng màu, đồng thời người sử dụng có thể chọn kết cấu mong muốn cho đặc điểm kết cấu và vẽ ra một phác họa cho truy vấn đặc điểm hình dạng.

1.1.3 Phương pháp để so sánh độ tương tự giữa các ảnh

Hệ thống Tra cứu ảnh dựa trên nội dung yêu cầu những phương pháp dựa trên những đặc điểm nguyên thủy để so sánh độ tương tự giữa ảnh mẫu và

tất cả những hình ảnh trong tập ảnh. 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 bức tranh, một là biển xanh mặt trời mọc và trường hợp khác là núi xanh với mặt trời mọc. 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 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 công nghệ 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.

1.1.4 Công nghệ tạo chỉ số và lưu trữ dữ liệu hiệu quả

Đối với một tập dữ liệu ảnh lớn thì không gian lưu trữ cho metadata là rất cần thiết. Một hệ thống tra cứu ảnh dựa trên nội dung phải có những công nghệ hiệu quả để quản lý metadata đồng thời phải có chuẩn để mô tả nó. Chuẩn MP7 đang là chuẩn quan trọng nhất để mô tả metadata cho cả dữ liệu ảnh và dữ liệu video. Khi một truy vấn được xử lý trên một cơ sở dữ liệu lớn, việc so sánh độ tương tự giữa ảnh truy vấn và tất cả các hình ảnh từng cặp là không thể thực hiện được bởi người dùng chỉ cần những ảnh có độ tương tự cao so với ảnh mẫu. Những chỉ số cấu trúc có thể giúp tránh được việc tìm kiếm tuần tự và cải thiện truy vấn một cách hiệu quả nên được sử dụng trong hệ thống tra cứu ảnh dựa trên nội dung. Hơn nữa với những cơ sở dữ liệu ảnh thường xuyên thay đổi thì chỉ số cấu trúc động là rất cần thiết. Khi nội dung của ảnh được thể hiện bởi các vector low dimension và khoảng cách giữa các ảnh được định nghĩa(chẳng hạn như khoảng không gian được tính toán bằng

khoảng cách Euclidean) cây R và các thành phần của nó có thể được sử dụng để đánh chỉ số cho ảnh. Khi khoảng cách không được định nghĩa như không gian vector hoặc khi không gian vector là High dimension hoặc khi mà những gì chúng ta có chỉ là một hàm khoảng cách tức là khoảng không metric thì những phương pháp để đánh chỉ số ảnh dựa trên hàm khoảng cách trong không gian metric là thích hợp.

1.2. Đặc điểm tra cứu ảnh

Kiểu truy vấn nào thích hợp để người sử dụng đưa vào cơ sở dữ liệu ảnh? Để trả lời câu hỏi này một cách sâu sắc đòi hỏi phải có sự hiểu biết chi tiết về nhu cầu của người sử dụng: Tại sao những người dùng lại tìm kiếm ảnh, họ sử dụng chúng để làm gì, và họ đánh giá lợi ích của hình ảnh mà họ tìm được như thế nào. Cảm giác chung gợi ra rằng ảnh tĩnh được yêu cầu bởi một loạt các lý do gồm:

Minh họa của những bài báo, truyền đạt thông tin hoặc cảm xúc khó mô tả bằng từ

Hiện thị dữ liệu chi tiết cho phân tích

Ghi lại dữ liệu thiết kế cho việc sử dụng sau này.

Truy cập tới một ảnh yêu cầu từ một kho dữ liệu ảnh có thể liên quan đến việc tìm kiếm ảnh mô tả kiểu đặc biệt của đối tượng hoặc đơn giản bao gồm kết cấu hoặc màu đặc biệt. Vì vậy ảnh có rất nhiều thuộc tính có thể sử dụng cho việc tra cứu bao gồm:

- Sự kết hợp đặc biệt của đặc tính màu sắc, kết cấu, hình dạng (ví dụ những ngôi sao mà xanh)

- Sự sắp xếp của các kiểu riêng biệt của đối tượng(ví dụ những chiếc ghế xung quanh cái bàn)

- Sự mô tả kiểu sự kiện (Trận bóng đá)

- Tên cá nhân, vị trí, sự kiện(ví dụ Nữ hoàng đón nhận vương miện)
- Những cảm xúc chủ quan kết hợp với hình ảnh(ví dụ niềm hạnh phúc)
- Metadata giống như ai đã tạo ra ảnh, ở đâu, khi nào?

Mỗi kiểu truy vấn được liệt kê bên dưới miêu tả mức trừu tượng cao hơn mức trước đó. Và mỗi mức rất khó để trả lời mà không tham khảo thêm tri thức bên ngoài. Điều này dẫn đến kiểu truy vấn được phân làm ba mức tăng dần theo độ phức tạp.

Mức 1: Gồm tra cứu bởi những đặc điểm nguyên thủy như màu sắc, kết cấu, hình dạng hoặc những vị trí đặc biệt của những phần tử ảnh. Ví dụ “Tìm một bức tranh với một đối tượng dài, màu xám ở trên đỉnh góc trái”, “Tìm ảnh chứa ngôi sao màu vàng được xếp thành một dãy” hoặc “Tìm bức tranh giống như thế này”... Mức tra cứu này sử dụng các đặc điểm từ chính những ảnh đó mà không cần tham khảo bất kỳ tri thức bên ngoài nào. Nó thường được ứng dụng trong lĩnh vực chuyên gia như việc đăng kí thương hiệu, nhận dạng các bộ sưu tập thiết kế.

Mức 2: Gồm những tra cứu bằng những đặc điểm biến đổi liên quan đến một số kết luận logic về sự đồng nhất của các đối tượng được mô tả trong ảnh. Nó có thể được chia thành:

Khôi phục các đối tượng theo kiểu nhất định(ví dụ tìm ảnh của chiếc xe buýt 2 tầng

Tra cứu những đối tượng đặc biệt hoặc người (ví dụ tìm bức ảnh của tháp Eiffel)

Để trả lời truy vấn ở mức này cần phải tham khảo một số tri thức bên ngoài, đặc biệt là truy vấn ở mức 2b. Trong ví dụ đầu tiên ở trên hiểu biết trước tiên cần thiết để xác định đối tượng là một chiếc xe buýt hơn là một

chiếc xe tải. Trong ví dụ thứ 2 cần một tri thức về một cấu trúc có tên là “tháp Eiffel”. Truy vấn mức này thường gặp hơn so với mức 1.

Mức 3: Gồm tra cứu bởi những thuộc tính trừu tượng liên quan đến một số lượng đáng kể suy luận ở mức cao về ý nghĩa và mục đích của đối tượng. Mức này có thể được chia làm:

- Tra cứu tên gọi của những sự kiện hoặc kiểu của hành động (ví dụ Tìm bức tranh về điệu nhảy dân gian Scottish)
- Tra cứu ảnh với những cảm xúc (“Tìm bức tranh mô tả sự đau khổ”)

Những thành công trong trả lời truy vấn ở mức này đòi hỏi một vài sự tinh tế của công cụ dò tìm. Để tạo ra sự kết nối giữa nội dung ảnh và những khái niệm trừu tượng thì cần phải có những lập luận phức hợp và những ý kiến chủ quan để minh họa. Nhưng truy vấn ở mức độ này ít phổ biến hơn mức độ 2 và thường gặp ở báo chí và những thư viện nghệ thuật.

Chúng ta nhận thấy rằng sự phân lớp của các kiểu truy vấn này có thể có lợi cho việc minh họa điểm mạnh cũng như những hạn chế của các công nghệ tra cứu ảnh khác nhau. Khoảng cách đáng kể hiện nay nằm giữa mức 1 và mức 2. Một số tác giả đề cập tới mức 2 và mức 3 như là tra cứu ảnh dựa trên ngữ nghĩa, và vì vậy khoảng cách giữa mức 1 và mức 2 là khoảng cách ngữ nghĩa.

1.3. Những ứng dụng cơ bản của tra cứu ảnh

Tra cứu ảnh được ứng dụng trong rất nhiều lĩnh vực, những lĩnh vực thành công bao gồm:

- Ngăn chặn tội phạm
- Quân sự
- Quản lý tài sản trí tuệ
- Thiết kế kiến trúc máy móc

- Thiết kế thời trang và nội thất
- Báo chí quảng cáo
- Chuẩn đoán y học
- Hệ thống thông tin địa lý
- Di sản văn hóa
- Giáo dục và đào tạo
- Giải trí
- Tìm kiếm trang web

1.4. Tra cứu ảnh dựa trên nội dung

1.4.1 Những phương pháp quản lý dữ liệu ảnh truyền thống

Sự cần thiết của việc lưu trữ và tra cứu ảnh một cách có hiệu quả đã được những nhà quản lý tập hợp ảnh lớn như thư viện ảnh, bộ sưu tập thiết kế...quan tâm từ nhiều năm nay. Trong khi việc xác định một ảnh mong muốn từ một tập ảnh nhỏ hoàn toàn có thể thực hiện được một cách đơn giản bằng cách duyệt qua thì với một tập ảnh lớn gồm hàng ngàn các đề mục thì cần phải có một công nghệ hiệu quả hơn. Công nghệ thường được sử dụng là gán mô tả dữ liệu bằng hình thức từ khóa, tiêu đề hoặc là mã phân lớp đối với mỗi ảnh khi nó được đưa vào tập hợp ảnh lần đầu tiên và sau đó dùng những ký hiệu mô tả này như là khóa để tìm kiếm.

Nhiều thư viện ảnh dùng từ khóa như là hình thức tra cứu chính của họ. Sơ đồ chỉ số thường được phát triển trong một nhóm phản ánh nét tự nhiên của tập ảnh. Một ví dụ điển hình là hệ thống được phát triển bởi Getty Image [Bjarnestam,1998]. Từ điển chuyên đề của họ trên 10.000 từ khóa được phân thành chín nhóm nghĩa gồm: Địa lý, con người, hoạt động và khái niệm...Lĩnh vực hay sử dụng sơ đồ chỉ số nhất là nghệ thuật và từ điển chuyên đề về nghệ thuật và kiến trúc (AAT), nó có nguồn gốc từ viện Rensselaer

Polytechnic vào đầu những năm 80, và ngày nay nó được sử dụng trong các thư viện nghệ thuật trên khắp thế giới. AAT gồm 120.000 thuật ngữ cho việc mô tả đối tượng, kết cấu vật liệu hình ảnh, kiến trúc và các di sản văn hóa khác. Các thuật ngữ được sắp xếp thành hệ thống phân cấp khái niệm như thuộc tính vật lý, kiểu, giai đoạn, chất liệu...

Một số sơ đồ chỉ số dùng mã phân lớp nhiều hơn từ khóa để mô tả nội dung ảnh bởi vì chúng có thể đưa ra ngôn ngữ độc lập hơn và chỉ ra khái niệm hệ thống phân cấp rõ ràng hơn ví dụ như: CONCLASS của trường Đại học Leiden [Gordon, 1990]

Công nghệ đánh chỉ số ảnh hiện thời có nhiều điểm mạnh đặc biệt là chỉ số từ khóa, nó có thể được sử dụng để mô tả hầu hết các khía cạnh của nội dung ảnh. Nó có thể mở rộng một cách dễ dàng phù hợp với những khái niệm mới và có thể sử dụng để mô tả nội dung ảnh ở những mức độ biến đổi phức tạp. Có rất nhiều phần mềm tra cứu văn bản có sẵn để tự động hóa quá trình tìm kiếm nhưng quá trình đánh chỉ số hướng dẫn (manual indexing) hoặc là từ khóa hoặc là mã phân lớp đều gặp phải hai hạn chế:

- Thứ nhất: Nó vốn là công việc rất tỉ mỉ, thời gian đánh chỉ số được đưa ra cho ảnh tĩnh là khoảng 7-40 phút/1 ảnh[Eakins and Graham, 1999].
- Thứ hai: Nó không xuất hiện một cách đáng tin cậy như là nghĩa của vấn đề tra cứu chẳng hạn những người khác nhau lấy những từ khác nhau để gán cho những ảnh bức ảnh giống nhau.

1.4.2 Các chức năng của hệ thống tra cứu ảnh dựa trên nội dung

Một hệ thống tra cứu ảnh dựa trên nội dung tiêu biểu không chỉ liên quan tới các nguồn thông tin trong những dạng khác nhau (ví dụ như văn bản, ảnh, video) mà còn liên quan đến nhu cầu của người sử dụng. Về cơ bản nó phân tích cả nội dung của nguồn thông tin cũng như truy vấn của người sử

dụng và sau đó đối sánh chúng để tìm ra những tiêu chí có liên quan này. Những chức năng chính của một hệ thống bao gồm:

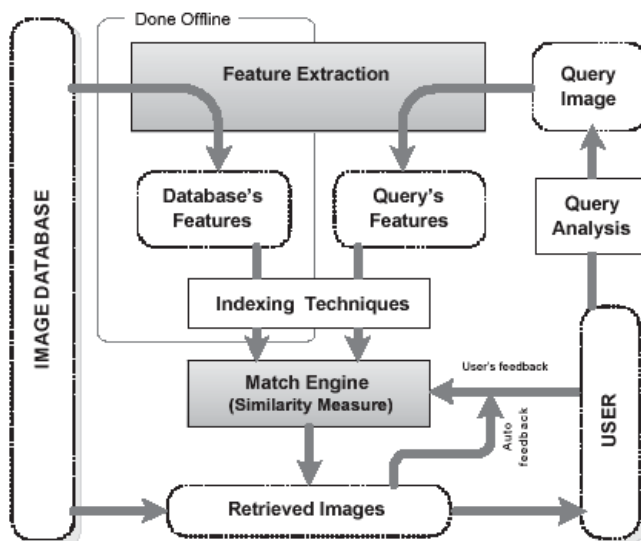
Phân tích nội dung của nguồn thông tin và biểu diễn nội dung của các nguồn thông tin được phân tích phù hợp với sự đối sánh truy vấn của người sử dụng (không gian của thông tin nguồn được chuyển đổi thành không gian đặc điểm với mục đích đối sánh nhanh trong bước tiếp theo). Bước này thường là mất nhiều thời gian cho việc xử lý tuần tự các thông tin nguồn (ảnh) trong cơ sở dữ liệu. Nó chỉ phải làm một lần và có thể làm độc lập.

Phân tích các truy vấn của người dùng và biểu diễn chúng thành các dạng phù hợp với việc đối sánh với cơ sở dữ liệu nguồn. Nhiệm vụ của bước này giống với bước trước nhưng chỉ được áp dụng với những ảnh truy vấn.

Xác định chiến lược để đối sánh tìm kiếm truy vấn với thông tin được lưu trữ trong cơ sở dữ liệu. Bước này có thể thực hiện trực tuyến và thực hiện rất nhanh. Công nghệ đánh chỉ số hiện tại có thể được sử dụng để nhận dạng không gian đặc điểm để tăng tốc độ xử lý đối sánh.

Tạo ra sự điều chỉnh cần thiết trong hệ thống (thường là bằng cách đổi chiều các tham số trong công nghệ đối sánh) dựa trên phản hồi từ người sử dụng hoặc những hình ảnh được tra cứu.

Rõ ràng là từ sự trình bày ở trên ta thấy một mặt hệ thống tra cứu ảnh dựa trên nội dung có các nguồn thông tin trực quan trong các dạng khác nhau, mặt khác lại có cả các yêu cầu của người sử dụng. Chúng được liên kết với nhau qua một loạt các công việc như được minh họa trong hình 1.1.



Hình 1.1. Các chức năng chính của hệ thống tra cứu ảnh dựa trên nội dung

Người sử dụng yêu cầu:

Có rất nhiều cách có thể đưa truy vấn trực quan. Một phương pháp truy vấn tốt là phương pháp tự nhiên với người sử dụng tức là cung cấp đầy đủ thông tin từ người sử dụng để trích chọn những kết quả có ý nghĩa. Những phương pháp dưới đây thường được sử dụng trong kỹ thuật tra cứu ảnh dựa trên nội dung:

- **Truy vấn bởi ví dụ** (QBE-Query By Examble): Trong kiểu truy vấn này người sử dụng chỉ định một ảnh truy vấn gốc dựa trên cơ sở dữ liệu ảnh được tìm kiếm và so sánh. Ảnh truy vấn có thể là một ảnh chuẩn, một ảnh quét với độ phân giải thấp, hoặc người sử dụng vẽ bằng cách sử dụng công cụ vẽ đồ họa. Ưu điểm của kiểu hệ thống này là rất tự nhiên đối với người sử dụng để tra cứu ảnh trong cơ sở dữ liệu ảnh.

- **Truy vấn bởi đặc điểm** (QBF- Query By Feature): Trong hệ thống kiểu này người dùng chỉ định câu hỏi bởi những đặc điểm chỉ định rõ ràng đó là những đặc điểm được quan tâm trong tìm kiếm. Ví dụ người dùng có thể truy vấn cơ sở dữ liệu ảnh bởi việc đưa ra một câu lệnh “Đưa ra tất cả những ảnh có góc bên trên trái chứa 25% điểm màu vàng”. Truy vấn này được người

dùng chỉ định bởi việc sử dụng công cụ giao diện đồ họa đặc biệt. Những người sử dụng chuyên nghiệp thì có thể tìm kiếm kiểu truy vấn tự nhiên này nhưng những người không chuyên thì rất khó. QBIC là một ví dụ về hệ thống tra cứu ảnh dựa trên nội dung mà người sử dụng truy vấn kiểu này.

- ***Những truy vấn dựa trên thuộc tính*** (Attribute-based queries):

Những truy vấn dựa trên thuộc tính sử dụng những chú giải kết cấu được trích chọn đầu tiên bởi sự nỗ lực của con người như khoá tra cứu. Mô tả kiểu này đòi hỏi phải có mức trừu tượng cao, cái rất khó đạt được mức độ tự động hoá hoàn toàn bởi vì ảnh gồm rất nhiều thông tin và rất khó có thể tổng kết bằng một ít từ khoá. Trong khi phương pháp này nhìn chung là nhanh hơn và dễ thực thi hơn thì nó vốn có sự chủ quan và mơ hồ ở mức cao như đã giới thiệu phần trước.

Phương pháp truy vấn nào là tự nhiên nhất? Với người sử dụng nói chung thì chắc chắn là truy vấn dựa trên những thuộc tính. Người sử dụng điển hình chắc chắn thích hỏi hệ thống tra cứu ảnh dựa trên nội dung bởi câu hỏi tự nhiên “Đưa ra cho tôi tất cả những ảnh từ hai năm trước”, hoặc là “tìm tất cả các ảnh trên Internet mà có bàn phím của máy tính”. Việc ánh xạ câu hỏi bằng ngôn ngữ tự nhiên này thành truy vấn trên cơ sở dữ liệu ảnh là vô cùng khó đối với việc sử dụng những phương pháp được tự động. Khả năng những máy tính thực hiện nhận dạng đối tượng tự động trên những ảnh vẫn đang là vấn đề nghiên cứu mở. Hầu hết những nghiên cứu cũng như các hệ thống mang tính thương mại đều tập trung xây dựng những hệ thống thực hiện tốt với những phương pháp QBE.

1.4.3 Trích chọn những đặc điểm

Trích chọn đặc điểm là cơ sở của tra cứu ảnh dựa trên nội dung. Theo một nghĩa rộng, những đặc điểm có thể gồm cả những đặc điểm dựa trên text (Từ khoá, những chú giải) và những đặc điểm trực quan (màu sắc, kết cấu, hình dạng). Trong phạm vi đặc điểm trực quan, những đặc điểm này lại được

phân thành những đặc điểm mức thấp và những đặc điểm mức cao. Những đặc điểm mức thấp bao gồm: màu sắc, kết cấu, hình dạng trong khi đặc điểm mức cao được ứng dụng dựa trên những đặc điểm này ví dụ mặt người, vân tay. Bởi nhận thức chủ quan, nên không tồn tại cách biểu diễn tốt nhất cho mỗi đặc điểm và vì vậy với mỗi đặc điểm có nhiều cách để biểu diễn mô tả những đặc điểm từ những ngữ cảnh khác nhau.

1.4.3.1 Màu sắc

Màu là đặc điểm trực quan đầu tiên và dễ nhất cho việc đánh chỉ số và tra cứu của ảnh và nó cũng là đặc điểm hay được sử dụng nhất trong lĩnh vực này.

Một ảnh màu điển hình được lấy từ camera số hoặc download từ Internet thường có ba kênh màu (ảnh xám chỉ có một kênh), những giá trị của dữ liệu ba chiều này từ ảnh màu có thể cho ta biết vị trí của những điểm ảnh này trong không gian màu. Những điểm ảnh có giá trị $(1, 1, 1)$ cho những màu khác nhau trong những không gian màu khác nhau. Như vậy mô tả đầy đủ của một ảnh màu điển hình gồm thông tin không gian hai chiều với điểm ảnh trong vùng không gian này và dữ liệu màu ba chiều với điểm ảnh màu trong không gian mà chúng ta đang đề cập. Ở đây giả thiết không gian màu là có định, bỏ qua thông tin không gian, thông tin màu trong ảnh có thể coi như là tín hiệu ba chiều đơn giản.

Nếu chúng ta coi thông tin màu của ảnh là tín hiệu một, hai, hoặc ba chiều đơn giản thì việc phân tích các tín hiệu sử dụng ước lượng mật độ xác suất là một cách dễ nhất để mô tả thông tin màu của ảnh. Biểu đồ màu là một công cụ đơn giản nhất, những cách khác mô tả thông tin màu trong tra cứu ảnh dựa trên nội dung gồm những đại diện màu, những moment màu.

1.4.3.2 Kết cấu

Kết cấu được sử dụng rộng rãi và rất trực quan nhưng không có định nghĩa chính xác bởi tính biến thiên rộng của nó. Có rất nhiều cách để mô tả

kết cấu: Những phương pháp thống kê thường sử dụng tần số không gian, ma trận biên cố, tần số biên... Từ những đặc điểm đơn giản này như là năng lượng, entropy, độ tương phản, độ thô, tính đồng nhất, tính tương quan, đẳng hướng, pha, độ ráp, đã được nhận ra. Những phương pháp mô tả kết cấu này tính toán các thuộc tính kết cấu khác nhau và hoàn toàn phù hợp nếu cơ của kết cấu gốc có thể được so sánh với cơ của điểm ảnh..

1.4.3.3 Hình dạng

Định nghĩa hình dạng của đối tượng thường là rất khó. Hình dạng thường được biểu diễn bằng lời nói hoặc hình vẽ, và mọi người thường sử dụng thuật ngữ như là tròn, méo. Xử lý hình dạng dựa trên máy tính đòi hỏi rất phức tạp, trong khi rất nhiều phương pháp mô tả hình dạng thực tế đang tồn tại nhưng không có một phương pháp chung nào cho mô tả hình dạng. Có hai kiểu đặc điểm hình dạng chính thường được sử dụng: những đặc điểm dựa trên biên và những đặc điểm dựa trên vùng. Đặc điểm dựa trên biên chỉ sử dụng đường bao ngoài của hình dạng trong khi đó đặc điểm vùng sử dụng toàn bộ vùng của hình dạng. Ví dụ những đặc điểm biên bao gồm mã xích, mô tả fourier, những đường viền hình học đơn giản như uốn cong, chiều dài biên,..., đặc điểm vùng như số chu trình, độ lệch tâm...

1.4.3.4 Những đặc điểm mức cao

Phần lớn những nghiên cứu tra cứu ảnh dựa trên nội dung đều tập trung vào những phương pháp ở mức thấp. Mặc dù vậy, một vài nghiên cứu đã cố gắng làm giảm khoảng cách giữa mức thấp và mức cao, chúng có hướng tập trung vào một trong hai vấn đề sau. Thứ nhất là nhận dạng cảnh, nó thường rất quan trọng để xác định tất cả các kiểu cảnh miêu tả ảnh, nó thường được sử dụng để tìm kiếm và có thể giúp xác định đối tượng một cách rõ ràng. Một trong những hệ thống kiểu này là IRIS (Hermes-1995), chúng sử dụng màu, kết cấu, vùng và thông tin không gian lấy ra từ phân tích hợp nhất của cảnh, tạo ra kí hiệu text để có thể đưa vào bất kỳ hệ thống tra cứu dựa trên text. Những nghiên cứu khác đã đưa ra những kỹ thuật đơn giản cho phân tích

cảnh, sử dụng những thành phần tần số thấp của ảnh để huấn luyện mạng neural, hoặc những thông tin màu lân cận được trích chọn từ những ảnh độ phân giải thấp để tạo ra những mẫu do người dùng định nghĩa.

Hướng thứ hai tập trung nghiên cứu nhận dạng đối tượng. Những công nghệ đang được phát triển cho nhận dạng và phân lớp đối tượng với cơ sở dữ liệu trực quan. Kỹ thuật tốt nhất được biết đến trong lĩnh vực này là kỹ thuật cho nhận dạng người trong ảnh. Tất cả những công nghệ này đều dựa trên ý tưởng phát triển mẫu cho mỗi lớp của những đối tượng được nhận dạng, xác định những vùng ảnh chứa đựng những mẫu của những đối tượng và xây dựng lên những mẫu chốt để xóa nhận hoặc loại bỏ sự có mặt của đối tượng.

1.4.4 Những khoảng cách tương ứng

Khi những đặc điểm của ảnh trong cơ sở dữ liệu được trích chọn và truy vấn của người dùng được thực hiện thì kết quả tìm kiếm được đưa ra bởi việc đo độ tương tự giữa những đặc điểm được trích chọn trong cơ sở dữ liệu và truy vấn của người sử dụng được phân tích. Những thước đo lý tưởng có một số những thuộc tính cơ bản sau:

- **Độ tương tự trực quan:** Đặc điểm khoảng cách giữa hai ảnh là lớn chỉ khi những ảnh không tương tự và ngược lại khoảng cách giữa hai ảnh là nhỏ nếu chúng tương tự. Những ảnh thường được mô tả trong không gian đặc điểm và sự tương tự giữa các ảnh thường được đo bởi những thước đo khoảng cách trong không gian đặc điểm. Số thuộc tính của không gian này cho cảm nhận của con người và hiểu những thuộc tính của những đặc điểm vectơ mô tả ảnh là rất quan trọng trong việc cải thiện thuộc tính độ tương tự trực quan của những thước đo độ tương tự được đề xuất.

- **Hiệu quả:** Sự đo đạc cần phải được tính toán nhanh để nhanh chóng đưa ra kết quả. Những ứng dụng tra cứu ảnh dựa trên nội dung tiêu biểu đòi hỏi phản hồi nhanh. Trong khoảng thời gian ngắn công nghệ tìm

kiểm thường phải tính toán hàng ngàn khoảng cách phụ thuộc vào cỡ của cơ sở dữ liệu ảnh, bởi vậy độ phức tạp tính toán là rất quan trọng.

- **Khả năng biến đổi:** Quá trình hệ thống thực hiện không nên bị giảm hiệu quả quá nhiều đối với cơ sở dữ liệu lớn bởi vì một hệ thống có thể tìm kiếm trong cơ sở dữ liệu chứa hàng triệu ảnh. Một sự thi hành đơn giản của một hệ thống tra cứu ảnh dựa trên nội dung là tính toán tất cả khoảng cách giữa ảnh truy vấn và ảnh trong cơ sở dữ liệu ảnh, sau đó những khoảng cách này được sắp xếp để tìm ra những ảnh tương tự nhất với ảnh truy vấn. Độ phức tạp của công nghệ tìm kiếm này tương ứng với cỡ của cơ sở dữ liệu ảnh (hoặc là $O(N)$ với N là số ảnh). Công nghệ đánh chỉ số đa chiều có thể được sử dụng để làm giảm độ phức tạp xuống $O(\log(N))$. Tuy nhiên, theo báo cáo rằng việc thực hiện của những công nghệ đánh chỉ số hiện thời đã giảm bớt được việc quét liên tục khi số chiều cần để đánh chỉ số là lớn hơn 20. Bởi vậy cần phải xem xét nhân tố này khi làm việc với cơ sở dữ liệu lớn.

- **Hệ thước đo:** Vấn đề khoảng cách tương tự là có lên là hệ mét hay không vẫn chưa được quyết định chính thức khi sự nhìn nhận của con người là rất phức tạp và chưa được hiểu một cách đầy đủ. Chúng ta thích khoảng cách tương tự là một hệ đo khi chúng ta xem xét những thuộc tính sau như là những yêu cầu rất tự nhiên:

- **Sự bất biến của tương tự với chính nó:** Khoảng cách giữa một ảnh với chính nó là hằng số độc lập với ảnh.,,

$$d(A,A)=d(B,B)$$

- **Sự tối thiểu:** Một ảnh giống với nó hơn là với những ảnh khác

$$d(A,A)<d(A,B)$$

- **Sự đối xứng:** Là vô lý nếu chúng ta nói rằng ảnh A giống với ảnh B nhưng ảnh B không giống với ảnh A

$$d(A,B)=d(B,A)$$

- **Sự bắc cầu**: Là vô lý nếu nói rằng ảnh A rất giống với ảnh B, ảnh B không giống với ảnh C nhưng ảnh C rất giống với ảnh A. Tuy nhiên, thuộc tính bắc cầu này có thể không đúng cho một dãy các ảnh. Thậm chí nếu ảnh I_i là giống với ảnh I_{i+1} với tất cả $i=1..N$ thì điều này không có nghĩa rằng ảnh I_i tương tự với ảnh I_N , ví dụ trong băng video mỗi khung tương tự với khung kế nó nhưng khung đầu tiên và khung cuối cùng có thể là rất khác nhau.

- **Sự mạnh mẽ**: Hệ thống cần có khả năng để thay đổi những điều kiện ảnh trong cơ sở dữ liệu ảnh, ví dụ nếu ảnh trong cơ sở dữ liệu ảnh được lấy dưới ánh sáng đèn điện (hơi đỏ) thì hệ thống phải có thể tìm được những đối tượng này ngay cả khi đối tượng truy vấn được lấy dưới ánh sáng ban ngày (hơi xanh).

Có rất nhiều thước đo khoảng cách tương tự đã được đưa ra nhưng chúng đều không có đầy đủ các thuộc tính trên. dưới đây là một vài thước đo chung nhất thường được sử dụng:

Histogram intersection Distanc (Swain and Ballard 1991):

Đây là một trong những thước đo khoảng cách đầu tiên trong tra cứu ảnh dựa trên màu sắc. Khoảng cách được định nghĩa dựa trên cỡ phần chung của hai biểu đồ màu. Cho hai biểu đồ màu h_1, h_2 , khoảng cách giữa chúng có thể được định nghĩa như sau:

$$\text{dis}_{\text{HI}} = 1 - \sum_{i=1}^N \min(h_{1i}, h_{2i})$$

Việc đo khoảng cách này rất nhanh bởi nó dựa trên công thức đơn giản. Tuy nhiên thông tin màu không được sử dụng khi nhận được khoảng cách bởi vậy có thể dẫn tới những kết quả không tốt.

L1 Distanc (Stricker and Orengo, 1996):

Khoảng cách dạng Minkowski L_p giữa hai biểu đồ màu được định nghĩa như sau:

$$\text{dis}_{M_p} = \left(\sum_i |h_{1i} - h_{2i}|^p \right)^{1/p}$$

Quadratic form Distanc (Hafner, 1995):

Khoảng cách giữa hai biểu đồ màu N chiều h_1 và h_2 được định nghĩa như sau:

$$\mathbf{dist}_{QF} = (h_1 - h_2)A(h_1 - h_2)$$

Với $A=[a_{ij}]$ là ma trận với trọng số biểu thị sự giống nhau giữa bin i và bin j , a_{ij} được tính như sau:

$$a_{ij} = 1 - (d_{ij} / d_{\max})^k$$

Ở đây d_{ij} là khoảng cách giữa màu i và màu j (thường d_{ij} là khoảng cách Euclidean giữa hai màu trong một vài không gian màu đồng dạng) và $d_{\max} = \max_{ij}(d_{ij})$. k là hằng số điều khiển trọng số giữa những màu lân cận.

Earth Mover Distance (Rubner, 1998):

Thước đo này dựa trên chi phí tối thiểu để chuyển một phân bố thành phân bố khác. Nếu chi phí của việc di chuyển một đơn vị đặc điểm đơn trong không gian đặc điểm là khoảng cách chung thì khoảng cách giữa hai phân bố sẽ là tổng cực tiểu của giá trị để di chuyển những đặc điểm riêng. Khoảng cách EMD có thể được định nghĩa như sau:

$$\text{dist}_{EMD} = \frac{\sum_{ij} g_{ij} d_{ij}}{\sum_{ij} g_{ij}}$$

Ở đây g_{ij} biểu thị khoảng cách tương tự giữa bin i và bin j và $g_{ij} \geq 0$ là sự tối ưu hoá giữa hai phân bố như là tổng giá trị được cực tiểu hoá,

$$\sum_i g_{ij} \leq h_{1i}$$

$$\sum_j g_{ij} \leq h_{2i}$$

$$\sum_{ij} g_{ij} = \min(h_{1i}, h_{2i})$$

Kolmogorov-Smirnov Distance (German,1990) :

Được định nghĩa như là sự khác nhau lớn nhất giữa những phân bố lũy tiến

$$dist_{KS} = \max_i |h_{1i}^c - h_{2i}^c|$$

Ở đây h^c là biểu đồ lũy tiến của biểu đồ h .

A Statistics of the Cramer/Von Mises:

$$dist_C = \sum_i (h_{1i}^c - h_{2i}^c)^2$$

Kullback-Leibler:

$$dist_{KL} = \sum_i h_{1i} \log \frac{h_{1i}}{h_{2i}}$$

Jeffrey- divergence:

$$dist_{KL} = \sum_i \left(h_{1i} \log \frac{h_{1i}}{h_{2i}} + h_{2i} \log \frac{h_{2i}}{h_{1i}} \right)$$

1.4.5 Các phương pháp tra cứu ảnh dựa trên nội dung

1.4.5.1 Tra cứu ảnh dựa trên màu sắc

Tra cứu ảnh dựa trên nền tảng màu sắc tương tự hầu hết là biến đổi dựa trên ý tưởng giống nhau. Mỗi ảnh khi đưa vào tập hợp ảnh đều được phân tích, tính toán một biểu đồ màu đó là tỷ lệ của những điểm ảnh của mỗi màu trong ảnh. Sau đó biểu đồ màu của mỗi ảnh sẽ được lưu trữ trong cơ sở dữ

liệu. Khi tìm kiếm người sử dụng có thể xác định tỷ lệ của mỗi màu mong muốn (ví dụ 75% Blue, 25% Red) hoặc đưa ra một ảnh mẫu với biểu đồ màu đã được tính toán. Đồng thời khi đó quá trình đối sánh tra cứu những biểu đồ màu của những hình ảnh này so sánh với biểu đồ màu của truy vấn gần nhất. Kỹ thuật đối sánh được sử dụng phổ biến nhất là biểu đồ màu giao nhau được phát triển đầu tiên bởi Swain and Ballard's[1991]. Những kỹ thuật cải tiến từ kỹ thuật này ngày nay được sử dụng rộng rãi trong các hệ thống tra cứu ảnh hiện thời.

Phương pháp cải tiến dựa trên công nghệ độc đáo của Swain and Ballard's gồm cách sử dụng biểu đồ màu tích lũy [Stricker and Orengo, 1995], kết hợp biểu đồ màu giao nhau với một số thành phần đối sánh không gian [Stricker and Dimai, 1996] và sử dụng vùng truy vấn dựa trên màu sắc [Carrson et al, 1997]. Kết quả của các hệ thống này đã tạo những ấn tượng khá sâu sắc.

1.4.5.2 Tra cứu ảnh dựa trên kết cấu

Khả năng tra cứu ảnh dựa trên kết cấu tương tự đường như không hiệu quả nhưng khả năng đối sánh dựa trên đặc điểm này thường có lợi cho việc phân biệt các vùng ảnh với màu tương tự (ví dụ như bầu trời và biển hoặc lá cây và cỏ). Một loạt các kỹ thuật được sử dụng cho việc đo kết cấu tương tự; công nghệ tốt nhất được thiết lập dựa trên sự so sánh những giá trị đã được biết đến như là số liệu thống kê thứ hai được tính toán từ truy vấn và những ảnh được lưu trữ. Từ đó có thể tính toán được khoảng cách của kết cấu ảnh như mức độ tương phản, độ thô, phương hướng và tính cân đối [Tamura et al, 1978] hoặc chu kỳ, phương hướng và tính ngẫu nhiên [Liu and Picard, 1996]. Các phương pháp phân tích kết cấu cho tra cứu bao gồm sử dụng những bộ lọc Gabor [Manjunath and Ma, 1996] và những Fractal [Kaplan et al, 1998]. Các truy vấn kết cấu có thể được trình bày tương tự như truy vấn màu sắc bằng việc lựa chọn những mẫu kết cấu như mong muốn từ bảng màu hoặc bằng việc cung cấp ảnh truy vấn mẫu. Hệ thống sau đó sẽ tra cứu những ảnh

với giá trị độ đo kết cấu giống nhau nhất với truy vấn. Gần đây có một sự mở rộng của công nghệ là cuốn từ điển kết cấu được phát triển bởi Ma and Manjunath, nó tra cứu những vùng kết cấu rõ ràng trong ảnh dựa trên nền tảng của sự tương tự để nhận lấy từ mã mô tả các lớp quan trọng của kết cấu trong tập ảnh một cách tự động.

1.4.5.3 Tra cứu ảnh dựa trên hình dạng

Khả năng tra cứu bởi hình dạng có lẽ là nhu cầu hiển nhiên nhất ở mức độ nguyên thủy. Không như kết cấu, hình dạng là một khái niệm hoàn toàn rõ ràng, và bằng chứng là những vật thể tự nhiên đầu tiên được nhận thấy bởi hình dạng của chúng [Biederman, 1987]. Số lượng những đặc điểm tiêu biểu của hình dạng đối tượng được tính toán cho mỗi đối tượng xác định trong mỗi ảnh được lưu trữ. Sau đó truy vấn được trả lời bởi việc tính toán tập những đặc điểm cho ảnh truy vấn, và việc tra cứu đặc điểm của những hình ảnh được lưu trữ này phải phù hợp với đặc điểm của truy vấn. Hai kiểu chính của đặc điểm hình dạng thường được sử dụng là đặc điểm tổng thể như tỷ lệ bên ngoài, hình tròn [Niblack et al, 1993] và những đặc điểm cục bộ như tập các đoạn biên liên tiếp [Mehrotra and Gary, 1995]. Các phương pháp khác đề cập tới sự đối sánh hình dạng bao gồm sự biến dạng co giãn của các khuôn dạng ([Pentland et al, 1996], [delBimbo et al, 1996]), sự so sánh của những biểu đồ định hướng của những biên được trích chọn từ ảnh [jain and Vailaya, 1996], khung biểu diễn hình dạng của đối tượng có thể được so sánh bằng việc sử dụng những kỹ thuật đồ sánh đồ thị [Kimia et al, 1977], Tirthap et al, 1998]. Những truy vấn đối với hệ thống tra cứu hình dạng thường được biểu diễn bằng cách xác định một hình ảnh mẫu để thực hiện như là hình thức truy vấn hoặc như là một bản phác thảo được vẽ ra bởi người sử dụng [Hirata and Kato, 1992], [Chan and Kung, 1997].

Việc đối sánh hình dạng của các đối tượng 3 chiều là một công việc khó khăn hơn. Trong khi chưa có giải pháp chung cho vấn đề này thì một số cách hữu ích đã được tạo thành cho việc xác định độ đo của đối tượng từ

nhiều khía cạnh khác nhau. Một phương pháp đã được sử dụng để xây dựng tập mô hình 3 chiều thích hợp từ ảnh 2 chiều có sẵn và đối sánh chúng với các mẫu khác trong cơ sở dữ liệu [Chen and Stokman, 1996]. Một cách khác nhằm tạo ra một loạt các ảnh 2 chiều khác của mỗi cơ sở dữ liệu đối tượng và mỗi ảnh này được đối sánh với ảnh truy vấn [Dickinson et al, 1998]. Những vấn đề nghiên cứu có liên quan đến lĩnh vực này gồm định nghĩa những độ đo tương tự hình dạng 3 chiều [Shum et al, 1996] và cung cấp phương tiện cho người sử dụng tạo ra những truy vấn hình dạng 3 chiều [Horikoshi and Kasahara, 1990].

1.4.5.4 Tra cứu ảnh bởi các đặc điểm khác

Một trong những phương tiện truy cập dữ liệu có hình ảnh cổ điển nhất là tra cứu bởi vị trí của nó trong ảnh. Truy cập dữ liệu bởi không gian vị trí là một khía cạnh chủ yếu của hệ thống thông tin địa lý, và các phương pháp hiệu quả để thực hiện công việc này đã được áp dụng trong nhiều năm gần đây (ví dụ Chock et al [1984], Roussopoulos et al [1988]). Những công nghệ tương tự cũng đã được áp dụng cho những tập ảnh, cho phép người sử dụng tìm kiếm những ảnh chứa các đối tượng có mối quan hệ không gian xác định với các đối tượng khác (Chang et al [1998], Chang and Jungert [1991]). Các thuật toán được cải tiến cho việc tra cứu thuộc lĩnh vực không gian vẫn đang được đề xuất. Việc đánh chỉ số không gian riêng nó thì ít hiệu quả, mặc dù nó chứng tỏ được hiệu quả của nó trong việc kết hợp với các dạng khác như màu sắc và hình dạng.

Một vài kiểu khác của đặc điểm ảnh được đưa ra như là nền tảng cho việc tra cứu ảnh dựa trên nội dung. Hầu hết những kiểu này đều dựa vào sự biến đổi phức tạp của cường độ của điểm ảnh. Đa số các công nghệ đều hướng về việc trích chọn ra những đặc điểm phản ánh một số khía cạnh của hình ảnh tương tự mà đối tượng con người có thể cảm nhận được, ngay cả khi người đó cảm thấy rất khó để mô tả. Kỹ thuật thành công nhất của loại này là sử dụng cách biến đổi wavelet. Kết quả tra cứu đầy hứa hẹn đã được báo cáo

bằng việc đối sánh những đặc điểm wavelet được tính toán từ truy vấn và những ảnh được lưu trữ. Một phương pháp khác cũng cho kết quả rất tốt là tra cứu bởi hình thức. Hai phiên bản của phương pháp này đã được phát triển, một cho đối sánh toàn bộ và một cho đối sánh những phần được lựa chọn của ảnh. Kỹ thuật đối sánh từng phần ảnh liên quan tới việc lọc ảnh với đạo hàm Gaussian nhiều mức, rồi sau đó tính toán lượng chênh lệch bất biến. Kỹ thuật đối sánh toàn bộ ảnh sử dụng sự phân bố độ cong và giai đoạn cục bộ.

Thuận lợi của tất cả những kỹ thuật này là chúng có thể mô tả một hình ảnh ở những mức khác nhau của chi tiết (có lợi trong những cảnh tự nhiên nơi mà các đối tượng quan tâm có thể xuất hiện dưới nhiều kiểu) và tránh việc cần thiết phân đoạn ảnh thành những vùng quan tâm trước khi những mô tả hình dạng có thể được tính toán. Bất chấp những tiến bộ gần đây trong kỹ thuật phân đoạn ảnh vẫn tồn tại vấn đề phức tạp.

1.5. Những hệ thống tra cứu ảnh dựa trên nội dung

1.5.1 Hệ thống QBIC(Query By Image Content)

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 điểm. Truy vấn Multi-feature bao gồm nhiều hơn một đặc điểm và mọi đặc điểm 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 www.qbic.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, hình dạng và kết cấu. 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 thốn 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.

CHƯƠNG 2: TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

Mắt của con người rất nhạy cảm với màu sắc, và đặc điểm màu là một trong những thành phần quan trọng nhất giúp con người có khả năng nhận biết hình ảnh. Vì vậy đặc điểm màu sắc là đặc điểm cơ bản của nội dung ảnh. Đặc điểm màu đôi khi có thể cung cấp những thông tin rất hữu hiệu cho việc phân loại ảnh và chúng cũng rất hữu ích cho việc tra cứu ảnh. Cũng vì thế mà tra cứu ảnh dựa trên màu sắc được sử dụng rộng rãi trong các hệ thống tra cứu ảnh dựa trên nội dung. Biểu đồ màu thường được sử dụng để thể hiện những đặc điểm màu của những ảnh. Mặc dù vậy trước khi sử dụng biểu đồ màu chúng ta cần phải lựa chọn và xác định kiểu không gian màu và lựa chọn thước đo khoảng cách.

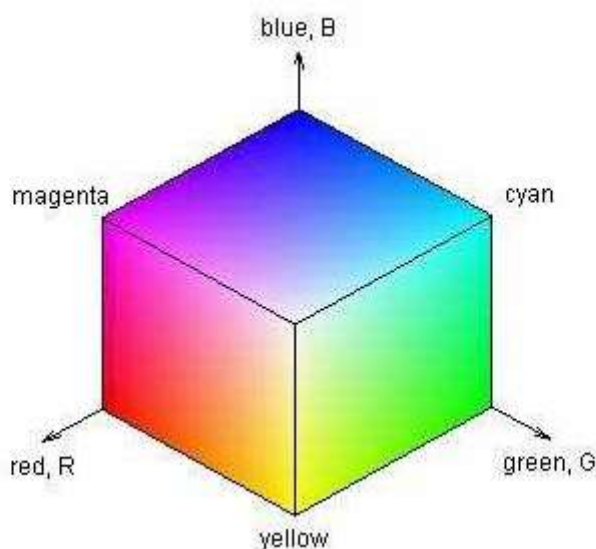
2.1. Không gian màu

Không gian màu là sự biểu diễn tập các màu, một số không gian màu được sử dụng rộng rãi trong đồ họa máy tính. Màu sắc thường được xác định trong không gian màu 3 chiều. Những mô hình không gian màu có thể được phân biệt thành hướng phần cứng và hướng người sử dụng. Mô hình không gian màu hướng phần cứng gồm RGB, CMY, YIQ dựa trên học thuyết 3 màu. Mô hình không gian màu hướng người sử dụng bao gồm HLS, HCV, HSV, HSB, MTM, CIE-LAB và CIE_LUV dựa trên 3 tỷ lệ phần trăm của màu đó là sắc màu, độ bão hòa và cường độ sáng. Mỗi mô hình không gian màu đồng thời có thể được phân biệt đồng dạng hoặc không đồng dạng dựa vào sự khác nhau trong không gian màu được nhận biết bởi con người. (Trong thực tế không có không gian màu đồng dạng đúng). Không gian màu đồng dạng xấp xỉ gồm MTM, CIE-LAB, CIE-LUB.

2.1.1 Không gian màu RGB

Không gian màu RGB được định nghĩa như là một hình lập phương đơn vị với 3 trục tương ứng là Red, Green và Blue, vì vậy, một màu trong

không gian màu RGB đại diện cho một véc tơ với ba tọa độ. Khi tất cả ba giá trị đều bằng 0 thì cho màu Black, khi tất cả ba giá trị đều bằng 1 thì cho màu White.

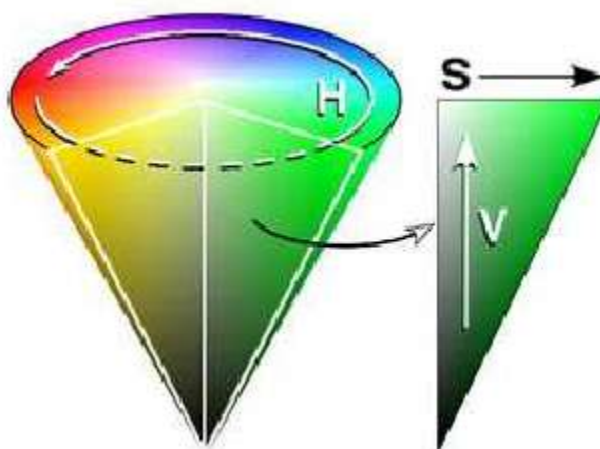


Hình 2.1. Không gian màu RGB

Không gian màu RGB là sự lựa chọn phổ biến nhất cho ảnh kỹ thuật số, bởi vì màn hình máy tính sử dụng những phospho red, green, blue để tạo ra màu mong muốn. Đồng thời người lập trình cũng dễ dàng hiểu và lập trình vì vậy không gian màu RGB được sử dụng rộng rãi trong nhiều năm nay. Những định dạng ảnh thường được sử dụng như GIF, JPEG và BMP luôn luôn lưu trữ và hiển thị màu trong không gian màu RGB. Vì vậy, tra cứu ảnh dựa trên không gian màu RGB sẽ không cần sự chuyển đổi không gian màu, do đó nó rất thuận tiện. Mặc dù vậy, do mô hình không gian màu RGB khó cảm nhận chẳng hạn như người sử dụng khó có được sự cảm nhận về màu $R = 100, G = 80, B = 50$ và khó tìm ra được sự khác nhau giữa hai màu $R = 100, G = 50, B = 50$ và $R = 100, G = 150, B = 150$ nên trong tra cứu ảnh dựa trên nội dung mô hình không gian màu RGB có thể được chuyển đổi thành mô hình không gian màu khác để cải thiện sự cảm nhận.

2.1.2 Không gian màu HSx

Không gian màu HSI, HSV, HSB, HLS thường được gọi là HSx có mối liên quan gần gũi với sự nhận thức về màu sắc của con người hơn là không gian màu RGB. Những trục từ không gian màu HSx mô tả những đặc tính của màu như sắc độ, độ bão hoà và độ sáng. Sự khác nhau giữa những không gian màu HSx khác nhau là sự biến đổi của chúng từ không gian màu RGB, chúng thường được mô tả bằng những hình dạng khác nhau (như hình nón, hình trụ). Trong hình 2.2 không gian màu HSV được mô tả như hình nón.



Hình 2.2. Mô tả không gian màu HSV

Sắc độ là thành phần của không gian màu HSx. Sắc độ là góc giữa những đường tham chiếu và điểm gốc màu trong không gian màu RGB như hình 2.1. Vùng giá trị này từ 0^0 đến 360^0 . Theo uỷ ban quốc tế về màu sắc CIE (Commission International de l'Éclairage) thì sắc độ là thuộc tính của cảm giác có liên quan đến thị giác, qua đó một vùng xuất hiện tương tự với một màu được cảm nhận như red, green, blue hoặc là sự kết hợp của hai trong số những màu được cảm nhận. Cũng theo CIE độ bão hoà là màu được đánh giá theo tỷ lệ độ sáng của nó. Trong hình nón độ bão hoà là khoảng cách từ tâm đến cạnh hình nón. Chiều cao của đường cắt chính là Value đây chính là độ sáng hoặc độ chói của màu. Khi độ bão hoà $S = 0$ thì H không xác định, giá trị nằm trên trục V biểu diễn ảnh xám. Không gian màu HSV dễ dàng

lượng tử hoá. Mức lượng tử hoá thông dụng trong không gian màu này là 162 bin với H nhận 18 mức, S và V nhận 3 mức.

2.1.3 Không gian màu YUV và YIQ

Không gian màu YUV và YIQ được phát triển cho truyền hình quảng bá. Không gian màu YIQ cũng giống như YUV với mặt phẳng I-Q là mặt phẳng U-V quay 33° . Y mô tả độ chói của điểm ảnh và chỉ sử dụng kênh đen trắng, U, V, I, Q là những thành phần màu. Kênh Y được định nghĩa bởi trọng số của giá trị R(0.299), G(0.587), B(0.144). Sơ đồ lượng tử hoá cho không gian màu YUV và YIQ thường được sử dụng là $125(5^3)$ hoặc $216(6^3)$ mức.

2.1.4 Không gian màu CIEXYZ và LUV

Không gian màu được phát triển đầu tiên bởi CIE là không gian màu XYZ. Thành phần Y là thành phần độ chói được định nghĩa bởi tổng trọng số của R(0.212671), G(0.715160), B(0.072169), X và Z là các thành phần màu. Không gian màu CIE LUV là sự biến đổi của không gian màu XYZ. Kênh L là độ chói của màu, kênh U và V là những thành phần màu. Vì vậy khi U và V được đặt bằng 0 thì kênh L biểu diễn ảnh xám. Trong lượng tử hoá không gian màu LUV mỗi trục được lượng tử hoá với mức xác định. Sơ đồ lượng tử hoá thường được sử dụng cho hai không gian màu này là 64, 125, 216 mức.

2.2. Biểu đồ màu

Biểu đồ màu là một tập hợp của các bin mà mỗi bin chỉ ra số pixel của một màu riêng biệt trong ảnh. Biểu đồ màu H cho ảnh được định nghĩa như là một vector:

$$H = \{H[0], H[1], H[2], \dots, H[i], \dots, H[N]\}$$

I: đại diện cho một màu trong biểu đồ màu và tương ứng với một hình vuông nhỏ trong không gian màu RGB

$H[i]$: là số điểm có màu i trong ảnh

N : là số bin trong biểu đồ màu tức là số màu trong không gian màu được chấp nhận.

Trong biểu đồ màu của ảnh, giá trị của mỗi bin sẽ là tổng số điểm ảnh có cùng màu tương ứng. Để so sánh những ảnh có kích cỡ khác nhau biểu đồ màu nên được chuẩn hóa và biểu đồ màu chuẩn hóa được định nghĩa như sau:

$$H' = \{H'[0], H'[1], H'[2], \dots, H'[i], \dots, H'[N]\}$$

Với $H'[i] = \frac{H[i]}{P}$ và P là tổng số các điểm trong ảnh.

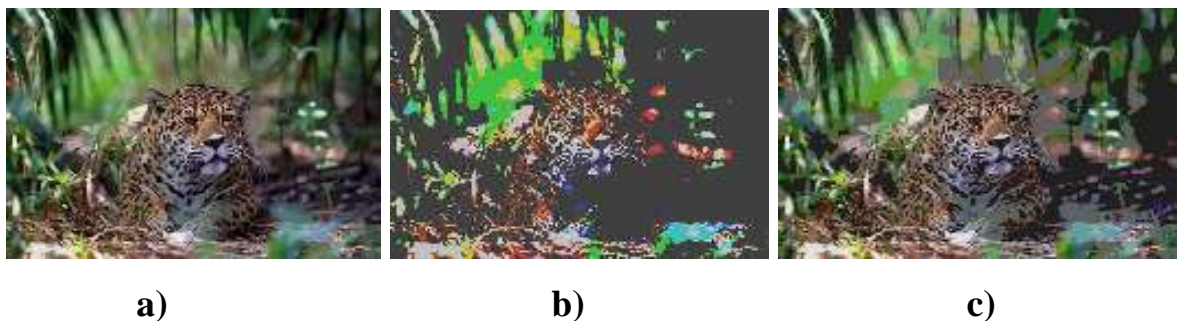
Một lượng tử hóa không gian màu lý tưởng sao cho những màu khác biệt không nên được định vị trong cùng một hình khối nhỏ, và những màu tương tự nên gán cho cùng một hình khối nhỏ. Sử dụng ít màu sẽ làm giảm khả năng những màu tương tự được gán cho những bin khác nhau nhưng nó làm tăng khả năng những màu phân biệt được gán cho những bin giống nhau và vì vậy nội dung thông tin của ảnh sẽ bị giảm đáng kể. Mặt khác biểu đồ màu với số lượng lớn các bin sẽ chứa nhiều thông tin về nội dung ảnh hơn, vì vậy, nó làm giảm khả năng các màu riêng biệt sẽ được gán cho các bin khác nhau, tăng không gian lưu trữ cơ sở dữ liệu, tăng thời gian tính toán khoảng cách giữa các biểu đồ. Chính vì thế cần phải có sự cân nhắc trong việc xác định bao nhiêu bin nên được sử dụng trong biểu đồ màu. Một con số điển hình được tìm thấy trong các tài liệu liên quan là 64.

2.3. Lượng tử hóa màu

Lượng tử hóa màu là quá trình làm giảm số màu sắc được sử dụng để mô tả ảnh. Việc lượng tử hóa màu trong không gian màu RGB được thực hiện bằng cách chia khối hình lập phương lớn thành những khối nhỏ và mỗi khối nhỏ có thể đại diện cho một màu đơn. Ví dụ chia hình lớn thành $64(4^3)$ hình nhỏ bằng cách chia các trục Red, Green, Blue mỗi trục thành 4 phần nhỏ và

tất cả các màu sắc được xác định trong một hình khối nhào sẽ đại diện cho một màu đơn.

Với hệ thống máy tính hiện thời thì không gian RGB thường thể hiện bởi hệ thống màu thực 24 bit. Trong hệ thống màu 24 bit thì mỗi màu được xác định bằng 3 số nguyên: {Red, Green và Blue} và 3 số nguyên này nằm trong khoảng từ 0 - 2^7 như vậy nó cho ta khoảng 16.777.216 màu (2^{24}). Bởi vì quá trình lượng tử hóa không gian màu RGB tương tự như quá trình làm giảm số màu nên có thể xác định số màu trong không gian màu một cách đơn giản là giảm số màu từ 24 bit màu xuống còn n^3 màu như sau:



Hình 2.3. Ảnh được lượng tử hoá

- a) ảnh màu gốc 256^3 màu trong không gian màu RGB.**
- b) ảnh lượng tử hóa 8 bit trong không gian màu RGB.**
- c) ảnh lượng tử hóa 64 bit trong không gian màu RGB.**

Khi giảm một màu {R, G, B} 24 bit màu thành màu mới {R', G', B'} với n^3 màu ta đặt:

$$R' = \frac{n * R}{2^8} \quad G' = \frac{n * G}{2^8} \quad B' = \frac{n * B}{2^8}$$

Vì vậy sau khi giảm số màu sẽ có $n*n*n=n^3$ màu.

2.4. Thước đo khoảng cách biểu đồ màu

Có rất nhiều thước đo đã được đưa ra để tính toán khoảng cách giữa các biểu đồ màu. [Smi97] đã liệt kê thước đo khoảng cách làm 3 loại có tên là:

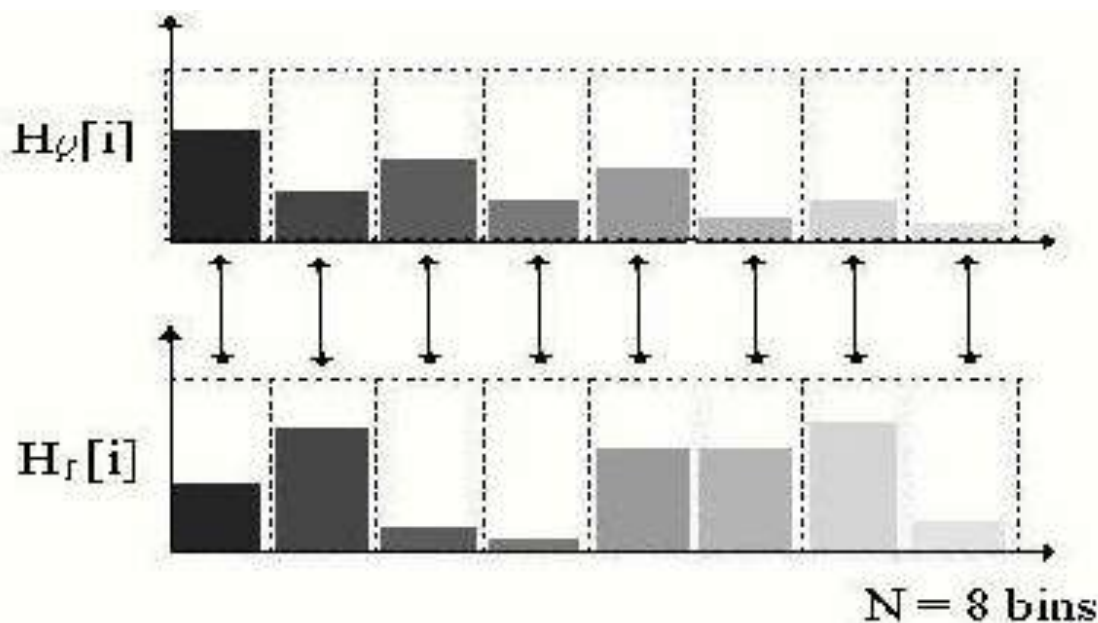
khoảng cách Minkowski, khoảng cách Quadratic và khoảng cách Non-histogram

2.4.1 Thước đo khoảng cách Minkowski

Trong [Sim97] độ đo khoảng cách dạng Minkowski chỉ so sánh những bin giống nhau giữa các biểu đồ màu (như hình 2.3) và nó được định nghĩa như sau:

$$d(Q,I) = \sum_{i=1}^N |H_Q[i] - H_I[i]|^r$$

Với Q và I là 2 ảnh, N là số bin trong biểu đồ màu (đối với mỗi ảnh số lượng màu được giảm xuống N màu trong không gian màu RGB, bởi vậy mỗi biểu đồ màu có N bin). $H_Q[i]$ là giá trị của bin i trong biểu đồ màu H_Q nó đại diện cho ảnh Q và $H_I[i]$ là giá trị của bin i trong biểu đồ màu H_I nó đại diện cho ảnh I.



Hình 2.4. Độ đo khoảng cách Minkowski

Khi $r=1$ thì khoảng cách Minkowski trở thành L_1 . Khi $r=2$ thì khoảng cách đó trở thành khoảng cách Euclidean. Trong thực tế khoảng cách Euclidean có thể được xem như khoảng cách không gian trong không gian đa chiều.

Trong đồ án này căn bậc hai của bình phương khoảng cách Euclidean

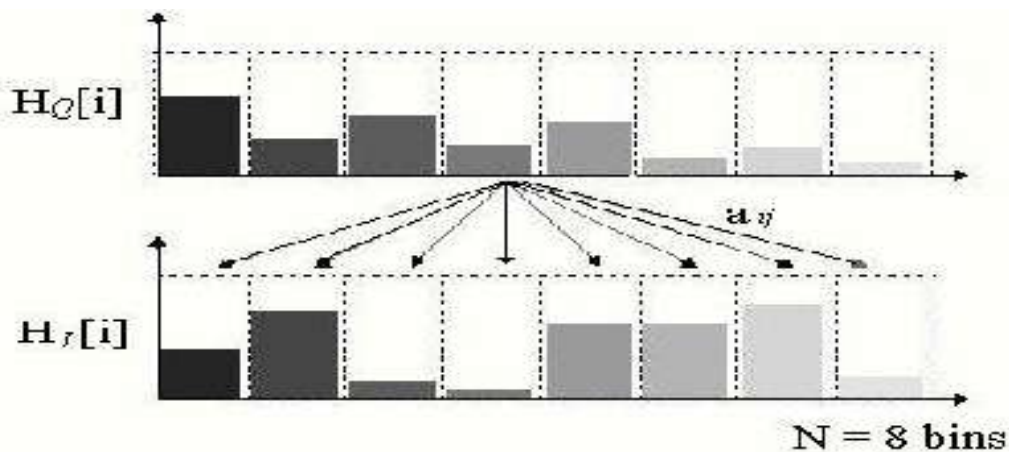
$$d(Q,I)=\sqrt{\sum_{i=1}^N(H_Q[i]-H_I[i])^2}$$

Sẽ được sử dụng để tính toán khoảng cách giữa hai biểu đồ màu

2.4.2 Thước đo khoảng cách Quadratic

Dự án QBIC [NBE +93] sử dụng thước đo khoảng cách màu Quadratic so sánh không chỉ các bin giống nhau mà còn so sánh nhiều bin khác nhau giữa các biểu đồ màu (như hình 2.4) và nó được định nghĩa như sau:

$$d(Q,I)=(H_Q-H_I)^t A(H_Q-H_I)$$



Hình 2.5.Độ đo khoảng cách Quadretic

Với Q và I là 2 ảnh, \$H_Q\$ là biểu đồ màu của ảnh IQ và \$H_I\$ nó là biểu đồ màu của ảnh I, \$A = [a_{i,j}]\$ lfa một ma trận cỡ \$N*N\$ với \$N\$ là số bin trong các biểu đồ

màu, và \$a_{i,j}\$ biểu thị sự tương tự giữa màu i và màu j. Độ đo khoảng cách này khắc phục được sự thiếu sót của thước đo khoảng cách dạng Minkowski đó là các bin trong biểu đồ màu hoàn toàn không liên quan đến nhau.

2.4.3 Thước đo khoảng cách Non-histogram

Stricker và Orengo [SO95] đã đưa ra phương pháp Color Moments nhằm vượt qua hiệu quả của lượng tử hóa biểu đồ màu. Trong phương pháp này những đặc điểm phân bố màu của ảnh được biểu diễn bởi những đặc điểm chủ yếu của chúng (gọi là các moment) đó là độ trung bình, sự thay đổi và tính đối xứng. Moment đầu tiên là màu trung bình của ảnh, thứ hai là độ lệch chuẩn của mỗi kênh màu và thứ ba là mối liên hệ thứ ba (third root) của mỗi kênh màu, và chúng được định nghĩa như sau:

$$E_i = \frac{1}{F} \sum_{j=1}^F P_{ij} ;$$

$$\delta_i = \left(\frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^2 \right)^{\frac{1}{2}} ;$$

$$S_i = \left(\frac{1}{F} \sum_{j=1}^F (P_{ij} - E_i)^3 \right)^{\frac{1}{3}} ;$$

P_{ij} : là giá trị của kênh màu thứ j và điểm ảnh thứ i .

E_i : là màu trung bình của kênh màu thứ i .

δ_i : là độ lệch chuẩn của kênh màu thứ i .

S_i : là giá trị của mối liên hệ thứ ba của kênh màu thứ i .

F là tổng số điểm ảnh.

Nếu Q và I là hai ảnh và đặc điểm màu của chúng được biểu diễn bởi r kênh màu thì sự tương tự giữa hai ảnh này được định nghĩa như:

$$d(Q,I) = \sum_{i=1}^r (W_{i1} |E_i^Q - E_i^I| + W_{i2} |\delta_i^Q - \delta_i^I| + W_{i3} |S_i^Q - S_i^I|)$$

Với W_{i1} , W_{i2} , W_{i3} là các trọng số được xác định bởi người sử dụng.

Stricker và Orengo chỉ ra rằng phương pháp của họ cho những kết quả tốt hơn và chạy nhanh hơn so với các phương pháp dựa trên biểu đồ màu (từ đó mỗi ảnh sẽ được biểu diễn bằng 9 số thực).

2.5. Tra cứu ảnh dựa trên biểu đồ màu

2.5.1 Phương pháp truyền thống dựa trên màu sắc

Có hai kỹ thuật truyền thống được sử dụng trong tra cứu ảnh dựa trên màu sắc đó là biểu đồ màu tổng thể biểu diễn những ảnh với những biểu đồ đơn và biểu đồ màu cục bộ đó là phân chia ảnh thành những khối cố định và mỗi khối có biểu đồ màu riêng cho khối đó. Biểu đồ màu toàn bộ không lấy được nội dung của những ảnh thích ứng, ngược lại biểu đồ màu cục bộ chứa nhiều thông tin hơn và nó còn cho phép so sánh khoảng cách màu giữa các vùng giữa các ảnh. Những kỹ thuật này thích hợp với các kiểu truy vấn khác nhau.

2.5.1.1 Biểu đồ màu toàn bộ(Global Color Histogram)

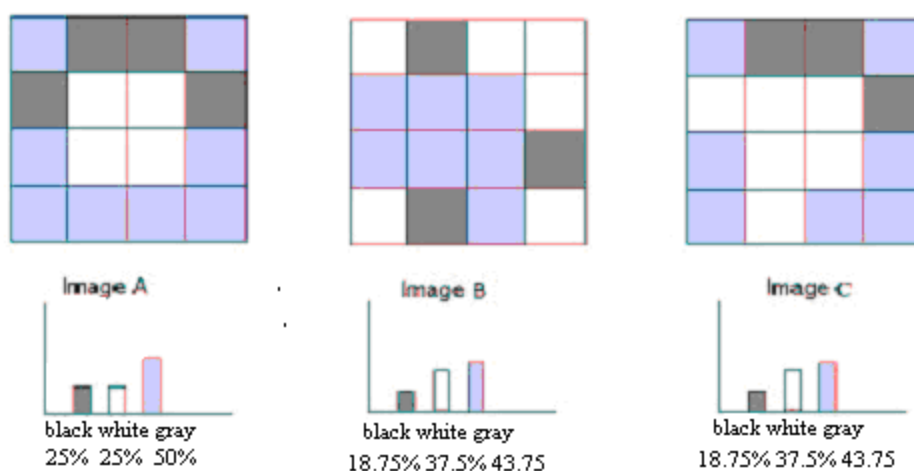
Biểu đồ màu loại này mô tả phân bố màu sử dụng tập các bin. Việc sử dụng biểu đồ màu toàn bộ (gọi tắt là GCH trong đề án này) thì một ảnh sẽ được mã hóa với biểu đồ màu của nó và khoảng cách giữa hai ảnh sẽ được xác định bởi khoảng cách giữa những biểu đồ màu của chúng. Với kỹ thuật này chúng ta có thể sử dụng các thước đo khác nhau để tính toán khoảng cách giữa hai biểu đồ màu. Ví dụ dưới đây sẽ mô tả hoạt động của kỹ thuật này:

Trong biểu đồ màu mẫu có 3 bin: Black, white and grey. Ta kí hiệu biểu đồ màu của ảnh A: {25%, 25%, 50%}; biểu đồ màu của ảnh B: {18.75%, 37.5%, 43.75} và ảnh C có biểu đồ màu như ảnh B. Nếu sử dụng thước đo khoảng cách Euclidean để tính toán khoảng cách biểu đồ thì khoảng cách giữa hai ảnh A và B cho biểu đồ màu toàn bộ là:

$$d_{GCH}(A,B) = \sqrt{(0.25 - 0.1785)^2 + (0.25 - 0.375)^2 + (0.5 - 0.4375)^2}$$

$$d_{GCH}(A,C) = d_{GCH}(A,B), d_{GCH}(B,C) = 0.$$

GCH là một phương pháp truyền thống cho việc tra cứu ảnh dựa trên màu sắc. Mặc dù vậy, nó không chứa các thông tin liên quan đến sự phân bố màu của các vùng. Vì vậy khoảng cách giữa các ảnh đôi khi không thể chỉ ra được sự khác nhau thực sự giữa các ảnh. Ví dụ khoảng cách giữa ảnh A và C khác so với khoảng cách giữa ảnh A và B nhưng bằng việc xây dựng GCH thì lại thu được khoảng cách tương tự. Ngoài ra còn có trường hợp hai ảnh khác nhau có GCH giống nhau như ví dụ trên ảnh B và C. và đây chính là hạn chế của biểu đồ màu toàn bộ.



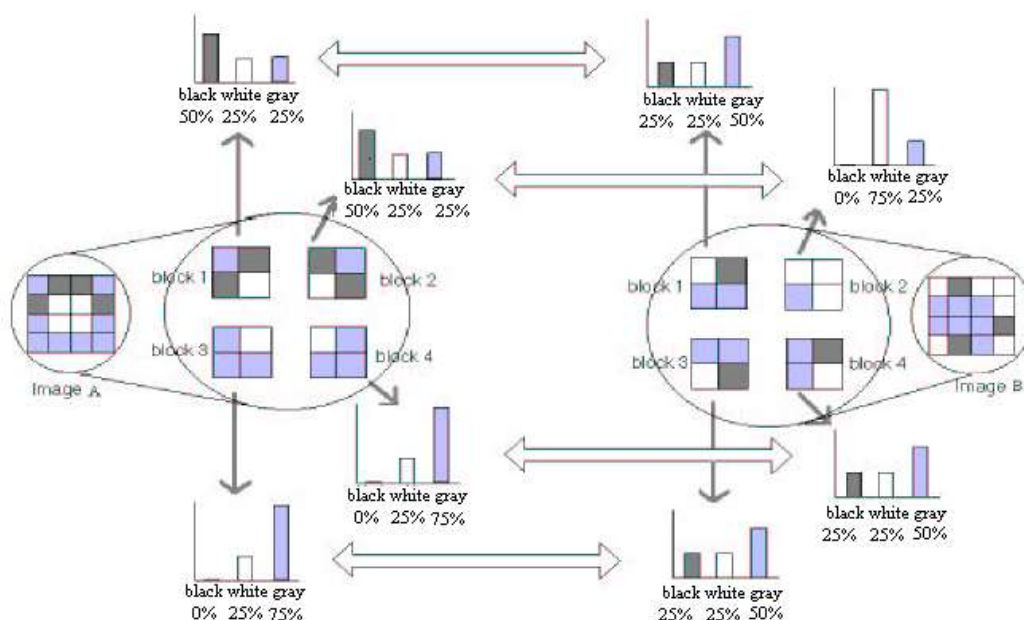
Hình 2.6. Ba ảnh và biểu đồ màu của chúng

2.5.1.2 Biểu đồ màu cục bộ

Phương pháp này được đề cập (gọi tắt là LCH) bao gồm thông tin liên quan đến sự phân bố màu của các vùng. Trước tiên là nó phân đoạn ảnh thành nhiều khối và sau đó biểu diễn biểu đồ màu cho mỗi khối, một ảnh sẽ được biểu diễn bởi những biểu đồ màu này. Khi so sánh hai hình ảnh, khoảng cách được tính toán bằng cách sử dụng những biểu đồ của chúng giữa một vùng trong một ảnh và một vùng tương ứng trong ảnh khác. Khoảng cách giữa hai ảnh được xác định bằng tổng tất cả các khoảng cách này. Nếu sử dụng căn bậc hai của khoảng cách Euclidean để tính toán khoảng cách biểu đồ thì khoảng cách giữa hai ảnh Q và I cho biểu đồ màu cục bộ là:

$$D(Q,I) = \sum_{k=1}^M \sqrt{\sum_{i=1}^N (H_Q^k[i] - H_I^k[i])^2}$$

ở đây M là số vùng được phân đoạn trong ảnh, N là số bin trong biểu đồ màu và H[i] là giá trị của bin i trong biểu đồ màu đại diện cho vùng k của ảnh. Những ví dụ dưới đây sử dụng những hình ảnh giống nhau như hình 2.5 để chỉ ra hoạt động của LCH và minh họa việc phân đoạn ảnh thành 4 khối có kích cỡ bằng nhau như thế nào.



Hình 2.7. Sử dụng LCH để tính toán khoảng cách giữa ảnh A và B

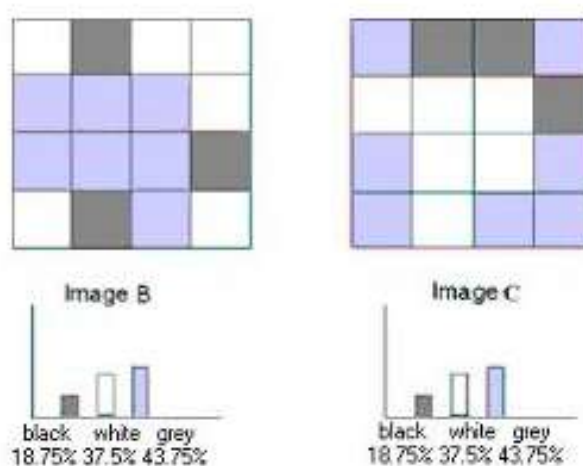
$$d_{LHC}(A,B) = 1.768, \quad d_{GHC}(A,B) = 0.153$$

$$\begin{aligned}
 d_{LCH}(A, B) &= \sqrt{(0.50 - 0.25)^2 + (0.25 - 0.25)^2 + (0.25 - 0.5)^2} + \\
 &\quad \sqrt{(0.50 - 0)^2 + (0.25 - 0.75)^2 + (0.25 - 0.25)^2} + \\
 &\quad \sqrt{(0 - 0.25)^2 + (0.25 - 0.25)^2 + (0.75 - 0.5)^2} + \\
 &\quad \sqrt{(0 - 0.25)^2 + (0.25 - 0.25)^2 + (0.75 - 0.5)^2} = 1.768
 \end{aligned}$$

2.5.2 Phương pháp Harbin

Hai công nghệ truyền thống tra cứu ảnh dựa trên nội dung là biểu đồ màu toàn cục (GCH - Global Color Histogram) và biểu đồ màu cục bộ (LCH -

Local Color Histogram) là điểm khởi đầu cho phương pháp này. GCH chỉ đưa ra biểu đồ màu cho ảnh mà không có thông tin vùng do vậy hiệu quả tra cứu thường là bị giới hạn, ví dụ ở hình 2.8 ảnh B và C có cùng biểu đồ màu do vậy khoảng cách giữa ảnh B và ảnh C theo $GCH = 0$. Mặc dù vậy hai ảnh này nhìn là khác nhau.



Hình 2.8. Hai ảnh và biểu đồ màu toàn cục của chúng

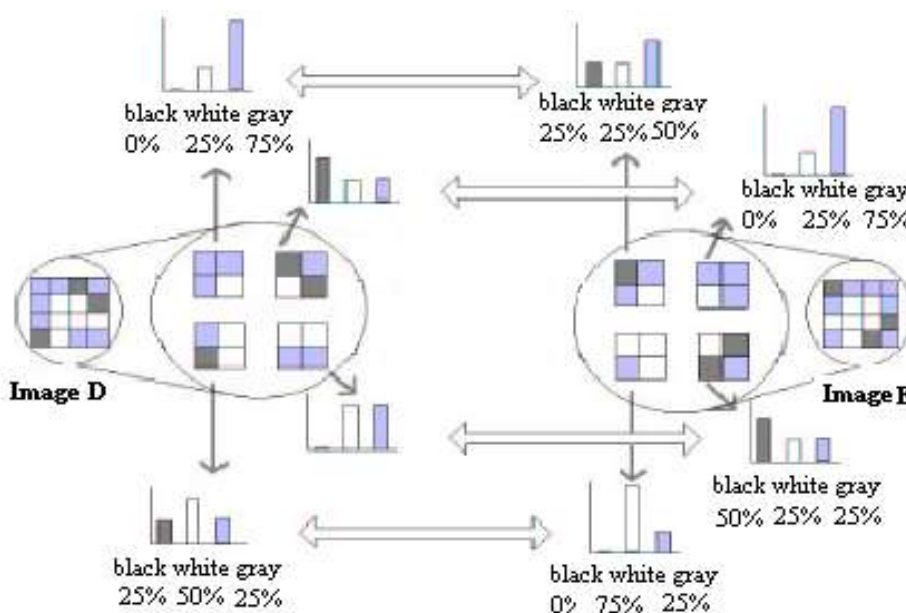
Phương pháp LCH đề cập ba bước:

- 1) Phân chia ảnh thành nhiều khối và xây dựng biểu đồ màu cho từng khối.
- 2) So sánh các khối trong vị trí tương ứng của hai ảnh (khoảng cách giữa hai ảnh là khoảng cách giữa biểu đồ màu của chúng)
- 3) Tổng hợp khoảng cách của tất cả các khối

Với phương pháp này khoảng cách giữa hai ảnh B và C được tính toán hợp lý hơn. Mặc dù vậy trong một số trường hợp như quay hoặc dịch chuyển ảnh thì tất cả các khối trong ảnh mới sẽ bị thay đổi vị trí và vì vậy sẽ không thích hợp khi sử dụng phương pháp LCH để so sánh độ tương tự những khối ở những vị trí giống nhau. Từ hình 2.9 ta thấy ảnh D gần giống với ảnh E, chỉ có hai khối khác nhau nhưng phương pháp LCH không thể so sánh những khối trong những vị trí thích hợp. Ví dụ này không những chỉ ra sự hạn chế của LCH mà nó còn đưa ra câu hỏi “có thể so sánh những khối trong những vị trí khác nhau một cách thích hợp không?”. Phương pháp Harbin đưa ra

nhằm trả lời câu hỏi này bằng việc xây dựng đồ thị vô hướng có trọng số và tìm kiếm giá trị đối sánh cực đại hoặc cực tiểu trên đồ thị này.

Những công nghệ tìm kiếm những đối sánh trong đồ thị vô hướng đã được sử dụng trong nhiều ứng dụng thực tế như những bài toán gán. Bài toán gán đòi hỏi phép gán tốt nhất của một nhóm công nhân với một nhóm công việc sao cho mỗi công nhân được gán với một công việc và mỗi công việc kết thúc bằng một công nhân cho đến khi không còn công nhân được gán với một công việc bất kỳ nào. Mỗi công nhân có thể kết thúc một công việc với một “giá trị” (giá trị ở đây có thể là năng suất hoặc chi phí hoặc giá trị sản phẩm). Với cách gán này ta có thể tính tổng giá trị cực đại hoặc cực tiểu của các giá trị này. Trong phần này sẽ chỉ rõ làm thế nào để xây dựng bài toán tính khoảng cách giữa hai ảnh trên cơ sở tìm kiếm giá trị đối sánh cực đại hoặc cực tiểu trong đồ thị vô hướng và phương pháp này được gọi là phương pháp Harbin.



Hình 2.9. Ví dụ chỉ ra LCH bị lỗi

Vấn đề của việc tìm giá trị cực đại và cực tiểu có thể chuyển đổi thành thay trọng số của mỗi cạnh w_{ij} bằng $w_{\max} - w_{ij}$. Với w_{ij} là trọng số của cạnh

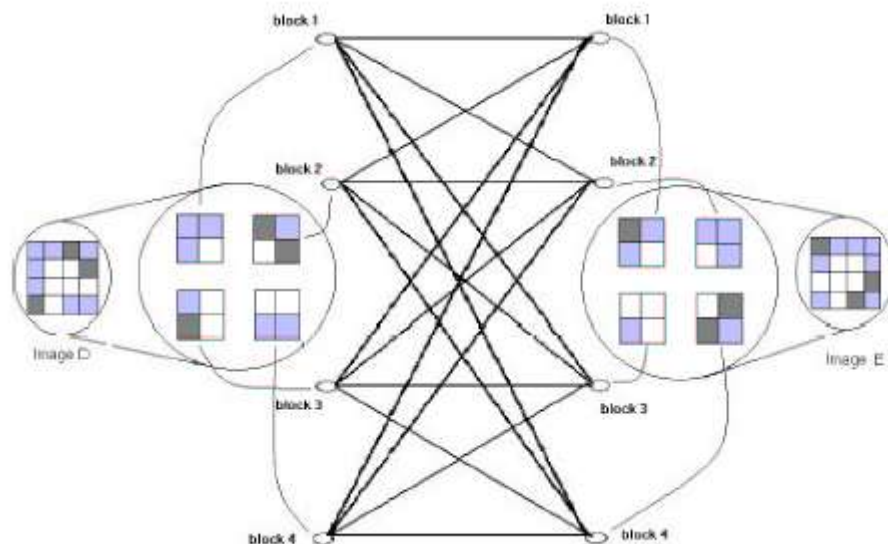
nối giữa đỉnh i và đỉnh j , w_{\max} là trọng số cực đại cho tất cả các cạnh. Phương pháp này gồm ba bước chính để so sánh khoảng cách giữa hai ảnh:

Chia ảnh thành các khối và tính biểu đồ màu cho từng khối

Xây dựng đồ thị vô hướng có trọng số

Tìm kiếm giá trị đối sánh cực đại hoặc cực tiểu, khoảng cách giữa hai ảnh sẽ là giá trị đối sánh nhỏ nhất.

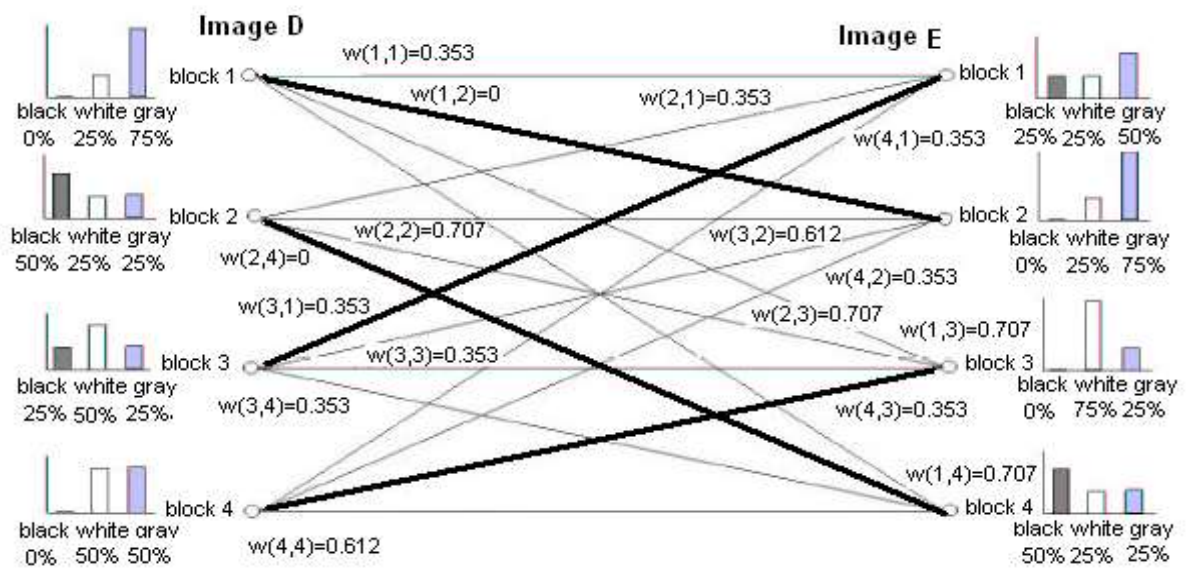
Trong hình 2.10 chỉ rõ cách xây dựng đồ thị vô hướng, bước đầu giống như bước đầu của phương pháp LCH. Trong việc xây dựng đồ thị vô hướng $G(X,Y,E)$, mỗi khối tương ứng với một đỉnh. Trong ví dụ này, mỗi ảnh được chia làm bốn khối và như vậy đồ thị vô hướng sẽ có tám đỉnh, mỗi khối sẽ trong một ảnh sẽ được nối với các khối của ảnh khác.



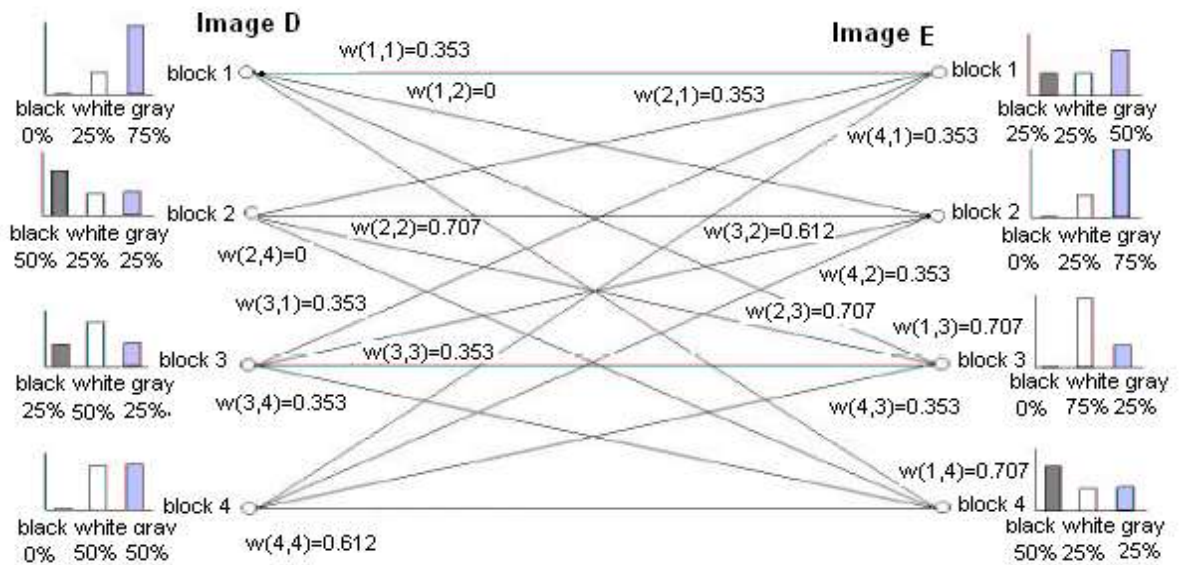
Hình 2.10. Xây dựng đồ thị vô hướng

Nếu trọng số của mỗi cạnh chỉ rõ khoảng cách giữa hai khối được nối bởi cạnh thì phương pháp Harbin sẽ tìm kiếm giá trị đối sánh nhỏ nhất và xử lý giá trị này như là khoảng cách giữa hai ảnh. Trong thực tế sự tương tự giữa các khối có thể được tính toán bằng công thức $S_{ij} = d_{\max} - d_{ij}$ với d_{\max} là khoảng cách lớn nhất giữa các khối, khối i và j thuộc hai ảnh tương ứng, d_{ij} là khoảng cách giữa khối i và khối j , S_{ij} là độ tương tự giữa khối i và j .

Hình 2.11 chỉ ra đồ thị vô hướng với trọng số là khoảng cách giữa các đỉnh. ở đây khoảng cách giữa các biểu đồ được tính toán sử dụng hàm khoảng cách Euclidean, giá trị này nằm trong khoảng 0 và 1. Sau khi tìm kiếm được giá trị đối sánh nhỏ nhất (đường nét đậm trong hình 2.12) tổng hợp lại sẽ được giá trị đối sánh. Giá trị đối sánh này sẽ là tổng các giá trị đối sánh của các cạnh tìm được: $Cost = w(1,2) + w(2,4) + w(3,1) + w(4,3) = 0.707$. Phương pháp này coi giá trị này là khoảng cách giữa ảnh D và ảnh E.



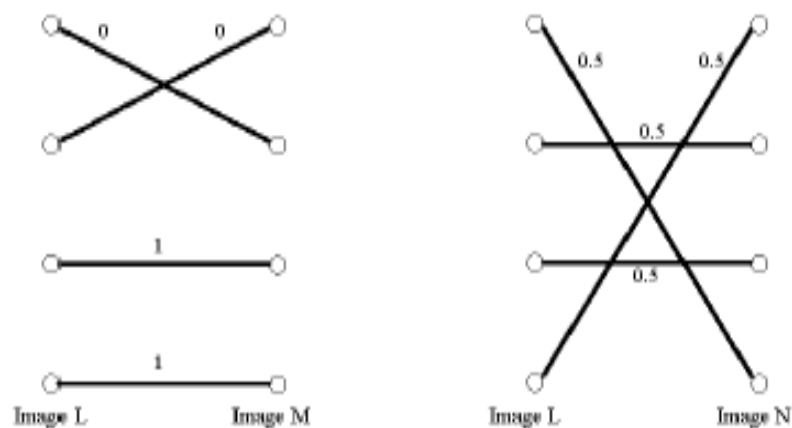
Hình 2.11. Đồ thị vô hướng biểu diễn mối quan hệ giữa các khối



Hình 2.12. Giá trị đối sánh nhỏ nhất trong đồ thị vô hướng

2.5.3 Sự nâng cấp phương pháp Harbin

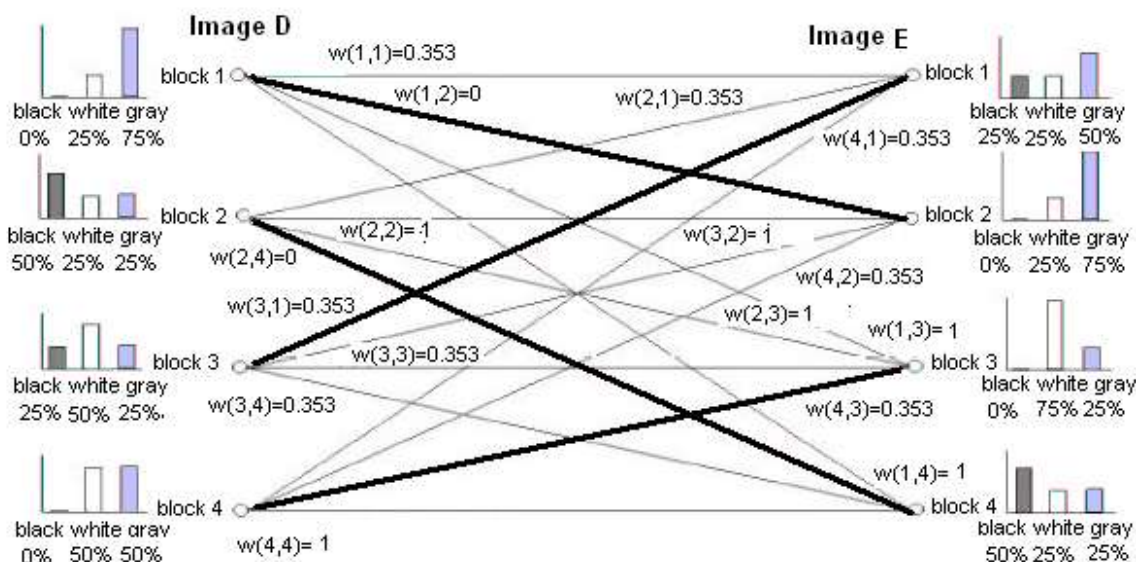
Nếu hai ảnh giống nhau thì chúng phải có một vài vùng giống nhau. Nếu một vùng trong một ảnh tương tự với một vùng trong ảnh khác thì ta nói rằng hai vùng được đối sánh. Những vùng đối sánh giữa hai ảnh sẽ xác định khoảng cách hoặc sự tương tự giữa chúng, khoảng cách hoặc sự tương tự sẽ được tính bằng tổng các khoảng cách hoặc sự tương tự giữa tất cả các vùng được đối sánh. Ý tưởng cơ bản của phương pháp Harbin là cố gắng đối sánh giữa các vùng của hai ảnh, bởi vậy chúng ta có thể cực tiểu hoá khoảng cách hoặc cực đại hoá độ tương tự giữa hai ảnh.



Hình 2.13. Sử dụng giá trị đối sánh cực tiểu để tính toán khoảng cách giữa hai ảnh L&M và L&N

Khi phương pháp Harbin sử dụng giá trị đối sánh cực tiểu để tính toán khoảng cách giữa các ảnh thì những đối sánh có thể bao gồm các cạnh với khoảng cách lớn hoặc những độ tương tự nhỏ. Mặc dù vậy, khi khoảng cách giữa hai khối là khá lớn (lớn hơn 0.5) thì những khối tương ứng sẽ không giống nhau chút nào và những cạnh với khoảng cách lớn sẽ chỉ làm tăng nhiều cho khoảng cách cuối cùng giữa hai ảnh. Điều này có nghĩa là không cần phân biệt những cái ít giống nhau và vì vậy không cần thiết phải phân biệt giữa các giá trị của khoảng cách lớn hoặc độ tương tự nhỏ. Ví dụ với ba ảnh L, N, M, muốn sử dụng phương pháp Harbin để tính toán khoảng cách giữa hai ảnh L với M và hai ảnh L với N thì ta phải xây dựng hai đồ thị vô hướng và tìm giá trị đối sánh cực tiểu trong đó. Hình 2.13 chỉ rõ những cạnh được tính trong giá trị đối sánh cực tiểu.

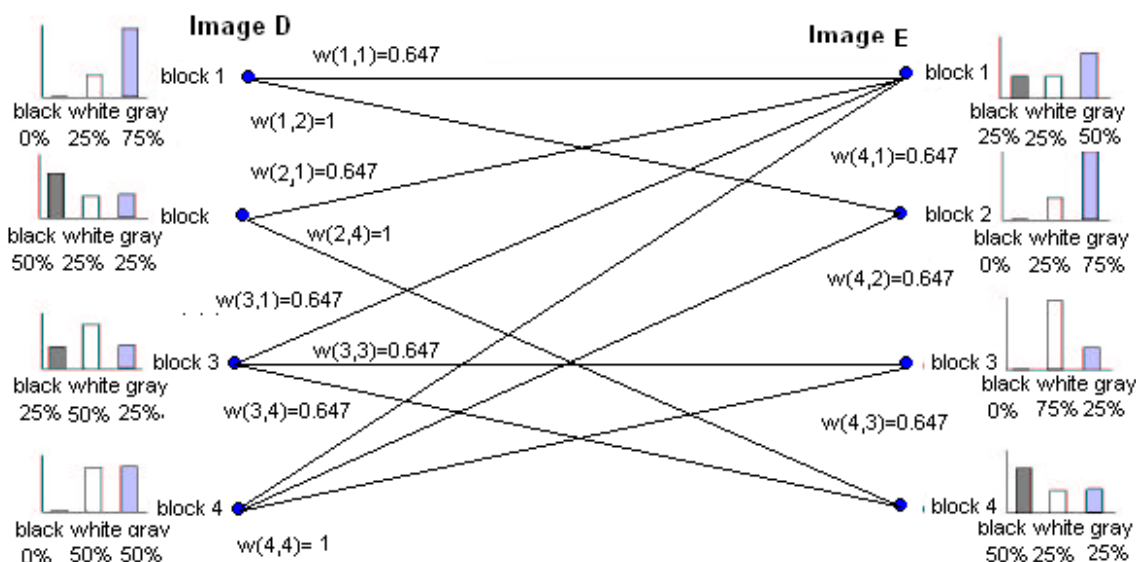
Từ hình vẽ 2.13 chúng ta thấy rằng khoảng cách giữa hai ảnh L&M và L&N là giống nhau và đều bằng 2, điều này có vẻ là không hợp lý (chú ý rằng một nửa hình ảnh L và một nửa hình ảnh N có thể chính xác giống nhau nhưng tất cả bốn cặp hình khối giữa ảnh L và N là hơi khác nhau). Lý do là những cạnh có giá trị bằng 0,5 đã làm tăng nhiều cho kết quả cuối cùng. Để giảm bớt nhiều khi xây dựng đồ thị vô hướng, các tác giả đã sử dụng theo kinh nghiệm để làm thước đo xác định trọng số của các cạnh và họ gọi đó là Δ . Nếu khoảng cách hai khối lớn hơn Δ thì trọng số của cạnh đó sẽ có giá trị cực đại (ở đây giá trị cực đại bằng 1). Nếu sử dụng $\Delta=0.5$ thì trong đồ thị vô hướng sẽ không có cạnh nào có trọng số nằm trong khoảng giữa 0.5 và 1.



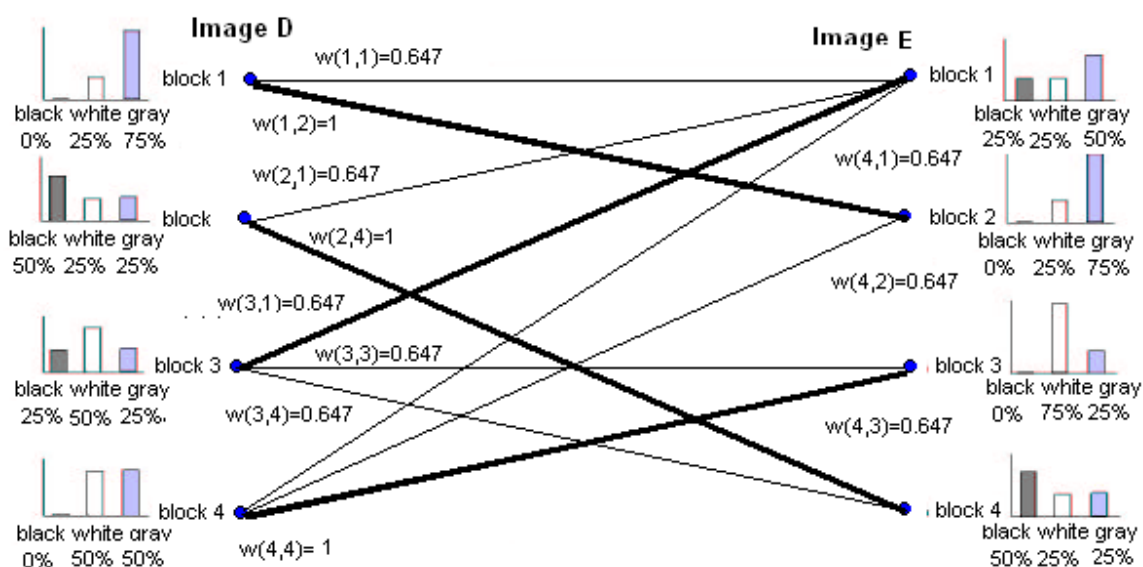
Hình 2.14. Đồ thị vô hướng không trọng số sau khi sử dụng $\Delta=0.5$

Hình vẽ 2.14 mô tả việc sử dụng ngưỡng $\Delta=0.5$ cho đồ thị vô hướng trong hình 2.12. Bằng việc đặt tất cả các khoảng cách có giá trị lớn hơn Δ bằng khoảng cách cực đại các tác giả hy vọng giảm được nhiều và cải thiện được kết quả tra cứu. Mặc dù vậy, nếu sử dụng giá trị Δ quá nhỏ thì sẽ loại bỏ những cách kết nối giữa những khối giống nhau và đây không phải là những gì mà chúng ta mong muốn. Vì vậy, vấn đề là làm thế nào để có được giá trị Δ để có được kết quả tra cứu tốt nhất sẽ phụ thuộc vào kết quả thử nghiệm.

Bài toán đối sánh giá trị cực tiểu có thể biến đổi thành bài toán đối sánh giá trị cực đại bằng cách thay đổi trọng số của các cạnh và phương pháp Harbin cũng có thể được sử dụng để tính độ tương tự giữa các ảnh. Chúng ta vẫn có thể sử dụng ngưỡng Δ để xác định trọng số của các cạnh để giảm nhiều. Nếu khoảng cách giữa hai khối lớn hơn Δ thì đặt độ tương tự giữa hai khối = 0. Nếu cạnh có trọng số = 0 thì ta có thể loại bỏ cạnh này trong đồ thị vô hướng. Với cách này, chúng ta đã làm giảm số cạnh trong đồ thị vô hướng và điều đó có nghĩa là đồ thị vô hướng sẽ không đầy đủ, sau đó chúng ta sẽ tìm đối sánh giá trị cực đại trong những đồ thị vô hướng không đầy đủ và giá trị này được coi là độ tương tự giữa hai ảnh.



Hình 2.15. Sử dụng $\Delta = 0.5$, trọng số của đồ thị mô tả độ tương tự



Hình 2.16. Giá trị đối sánh cực đại sau khi sử dụng $\Delta = 0.5$

Chúng ta biết rằng độ phức tạp tính toán của các thuật toán đối sánh liên quan tới số lượng các cạnh. Vì vậy, sau khi sử dụng Δ và tìm kiếm đối sánh giá trị cực đại thì phương pháp Harbin trở lên hiệu quả hơn. Ví dụ sau khi sử dụng $\Delta = 0.5$ và thay khoảng cách thành độ tương tự trên đồ thị vô hướng đầy đủ ở hình 2.11 chúng ta được đồ thị như hình 2.15

Hình 2.16 chỉ ra đối sánh giá trị cực đại của đồ thị vô hướng trong hình 2.15 với đường kẻ đậm biểu thị cạnh đối sánh. Độ tương tự giữa ảnh D và ảnh E là 2.294. Trước khi tìm kiếm giá trị đối sánh cực tiểu để cực tiểu hoá

khoảng cách tương tự chúng ta phải tìm giá trị đối sánh cực đại để cực đại hoá độ tương tự.

2.6. Cải tiến hiệu quả tra cứu

Phương pháp Harbin khắc phục được thiếu sót của phương pháp LCH là chỉ so sánh những khối trong những vị trí giống nhau dù rằng độ phức tạp tính toán của nó là cao ($O(mn^2)$). So với phương pháp LCH phương pháp Harbin cho hiệu quả hơn. Trong thực tế đối với những cơ sở dữ liệu ảnh lớn, một cấu trúc đánh chỉ số hiệu quả được sử dụng để tránh tìm kiếm tuyến tính thường quan trọng hơn là hiệu quả của chính phương pháp đó. Câu hỏi được đặt ra là không cần cấu trúc đánh chỉ số hiệu quả liệu chúng ta có thể vẫn sử dụng phương pháp Harbin trong thực hành?

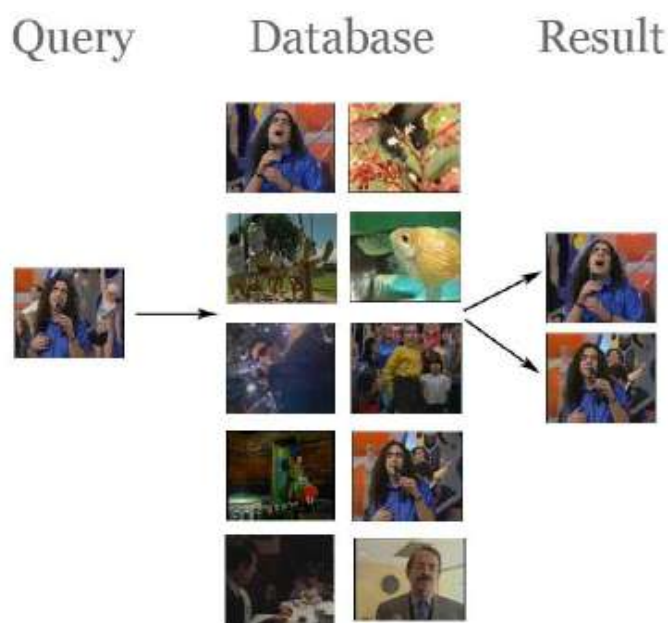
Kỹ thuật GCH không hiệu quả lắm bởi vì nó không có thông tin vùng. Nếu độ tương tự đạt được giữa hai ảnh sử dụng phương pháp Harbin cao thì độ tương tự đạt được khi sử dụng kỹ thuật GCH cũng phải cao. Mặt khác nếu độ tương tự đạt được giữa hai ảnh sử dụng phương pháp GCH là thấp thì độ tương tự đạt được khi sử dụng kỹ thuật Harbin cũng phải thấp. Do đó không cần thiết sử dụng phương pháp Harbin cho những ảnh có độ tương tự thấp khi sử dụng kỹ thuật GCH. Trên cơ sở ý tưởng này các tác giả đã đề xuất sử dụng phương pháp Harbin để cải tiến những kết quả tra cứu đạt được khi sử dụng kỹ thuật GCH. Bước đầu các tác giả sử dụng phương pháp GCH để đạt được những hình ảnh cần thiết (khoảng 100 ảnh), và sau đó sử dụng công nghệ Harbin để sắp xếp lại những ảnh này. Bằng cách này thay vì sử dụng công nghệ Harbin để so sánh những nét giống nhau giữa ảnh truy vấn với tất cả các ảnh trong cơ sở dữ liệu ảnh, các tác giả chỉ sử dụng nó trên những ảnh có nét tương tự cao đạt được khi sử dụng kỹ thuật GCH và vì vậy tránh được việc tìm kiếm tuyến tính. Mặc dù công nghệ GCH không hiệu quả lắm và phương pháp Harbin cũng không hiệu lực lắm nhưng một phương pháp mới kết hợp hiệu quả của

hai phương pháp này hy vọng sẽ cho hiệu quả và hiệu suất cao hơn trong tra cứu ảnh tương tự.

CHƯƠNG 3: CHƯƠNG TRÌNH THỬ NGHIỆM

3.1. Bài toán

Tra cứu hình ảnh dựa trên nội dung là ứng dụng các kỹ thuật thị giác máy cho truy vấn hình ảnh từ cơ sở dữ liệu lớn các ảnh số. Trong một truy vấn nội dung hình ảnh dựa trên hệ thống truy cập có thể được thực hiện với một ký họa hoặc (thường) bởi một truy vấn hình ảnh (Hình 1.3).



Hình 3.1

3.2. Lựa chọn công cụ

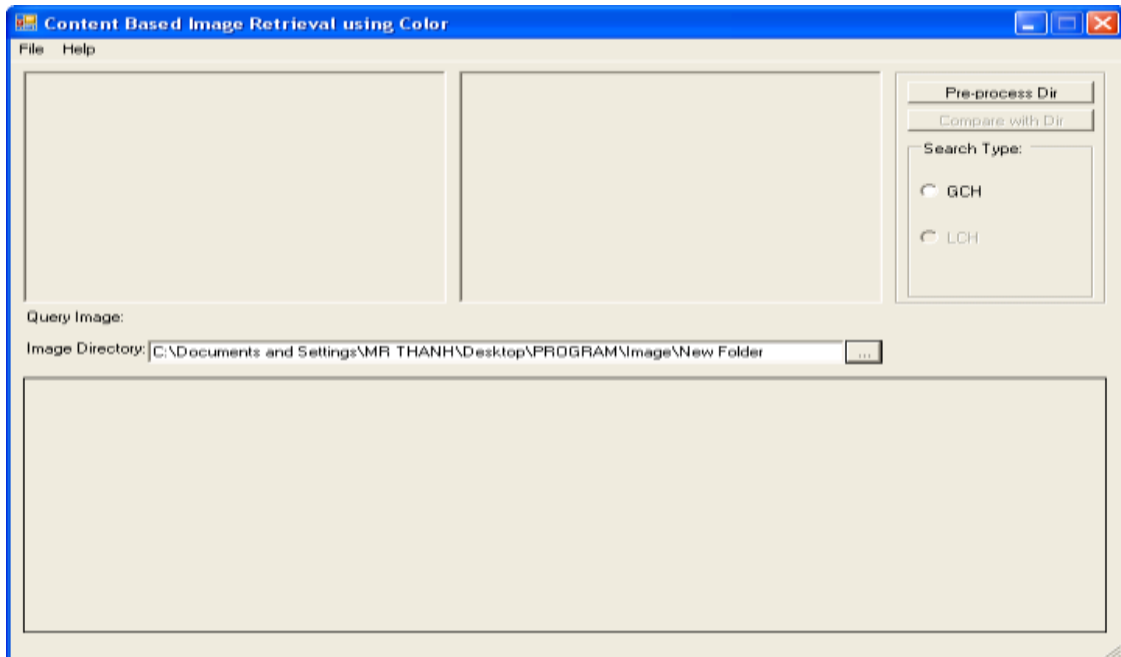
Ngôn ngữ Visual C# là một trong những ngôn ngữ ứng dụng đồ họa mạnh mẽ, rất thích hợp trong các ứng dụng về xử lý ảnh. Chương trình được xây dựng trên ngôn ngữ Visual C#.

Cấu hình máy:

- Bộ vi xử lý Pentium hoặc Pentium Pro trở lên.
- Window 2000 trở lên.
- Bộ nhớ động RAM 256 MB.

3.3. Một số kết quả chương trình

3.3.1 Giao diện chương trình



Hình 3.2

Trong đó:

Pre –process Dir : Tính toán đặc trưng lưu vào Cơ sở dữ liệu

GCH : Chọn chế độ tra cứu ảnh toàn cục

LCH : Chọn chế độ tra cứu ảnh cục bộ

Compare with Dir : Thực hiện chế độ tra cứu ảnh tương tự

3.3.2 Kết quả



Hình 3.3

KẾT LUẬN

Để hoàn thành đề tài đồ án tốt nghiệp “Tra cứu ảnh bằng biểu đồ màu” em đã tìm hiểu về tổng quan về tra cứu ảnh dựa trên nội dung, tra cứu ảnh dựa trên biểu đồ màu, từ đó em đã thu được một số thông tin như sau:

Tổng quan về tra cứu ảnh dựa trên nội dung.

Trích chọn đặc trưng ảnh dựa vào biểu đồ màu ứng dụng cho tra cứu ảnh.

Từ đó em xây dựng chương trình mô phỏng tra cứu ảnh bằng biểu đồ màu bằng ngôn ngữ Visual C#.

Tuy nhiên trong quá trình tìm hiểu bài báo do chưa có nhiều thời gian nên em chưa tìm hiểu hết được các mục tác giả đưa ra trong phần tài liệu tham khảo. Trong thời gian tới đây em sẽ cố gắng đọc các tài liệu đó để hiểu thêm về các thuật toán liên quan về nhận dạng ảnh bằng biểu đồ màu.

TÀI LIỆU THAM KHẢO

- [1]. Nguyễn Thanh Bình, Võ Nguyễn Quốc Bảo. “Xử lý âm thanh, hình ảnh”. *Chương trình đào tạo từ xa PTIT, 2007.*
- [2]. Nguyễn Văn Ngọ. “Giáo trình Xử lý ảnh”. *Giáo trình xử lý ảnh ĐHQGHN, 2001.*
- [3]. Đỗ Năng Toàn, Phạm Việt Bình. “Giáo trình môn học xử lý ảnh”. *ĐH Thái Nguyên, 2007.*
- [4]. Wikipedia article on the RGB colour model, <http://en.wikipedia.org/wiki/RGB>, last visited june 29th2005.
- [5]. R.Russel, PSinhha. *Perceptually based Comparison of Image Similarity Metrics*. MIT AI Memo 2001-014. Massachusetts Institute of Technology, 2001.
- [6]. Wikipedia article on HSV, http://en.wikipedia.org/wiki/HSV_color_space, last visted June 29th2005.
- [7]. GongY, Chuan C.H, Xiaoyi G. *Image indexing and retrival using color histogram*, Multimedia Tools and Applications, vol.2pp.133-156, 1996.
- [8]. Shengjiu Wang, *A Robust CBIR Approach Using Color Histogram*, Technical Report TR 01-03, Department of computing science, University of Alberta, Canada. October 2001.