

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

-----o0o-----

**TÌM HIỂU MỘT SỐ PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG
VÀ ỨNG DỤNG CHO TRA CỨU ẢNH THEO NỘI DUNG**

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành: Công nghệ thông tin

Hải Phòng - 2011

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

-----o0o-----

**TÌM HIỂU MỘT SỐ PHƯƠNG PHÁP TRÍCH CHỌN ĐẶC TRƯNG
VÀ ỨNG DỤNG CHO TRA CỨU ẢNH THEO NỘI DUNG**

**ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY
NGÀNH CÔNG NGHỆ THÔNG TIN**

Sinh viên : Bùi Thị Thúy Nga
Giáo viên hướng dẫn : PGS. TS Ngô Quốc Tạo
Mã sinh viên : 111217

Hải Phòng, 7/2011

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

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh Phúc
-----oOo-----

NHIỆM VỤ THIẾT KẾ TỐT NGHIỆP

Sinh viên: Bùi Thị Thúy Nga

Mã số: 111217

Lớp: CT1101

Ngành : Công nghệ thông tin

Tên đề tài:

Tìm hiểu một số phương pháp trích chọn đặc trưng và ứng dụng cho tra cứu ảnh theo nội dung.

NHIỆM VỤ CỦA ĐỀ TÀI

1. Nội dung và các yêu cầu cần giải quyết trong nhiệm vụ của đề tài tốt nghiệp

a. Nội dung

- Nghiên cứu và đánh giá các kỹ thuật tra cứu ảnh theo nội dung.
- Đề xuất sử dụng một số kỹ thuật khác và đưa ra sự so sánh với các hệ thống ảnh trước đó.
- Từ đó đưa ra các phương pháp cải tiến và khắc phục những hạn chế hiện tại.
- Thử nghiệm chương trình tra cứu ảnh theo nội dung.

b. Các yêu cầu cần giải quyết

- Tìm hiểu kiến thức về xử lý ảnh.
- Tham khảo các mã nguồn mở về xử lý ảnh.
- Nghiên cứu về tra cứu ảnh theo nội dung và các phương pháp trích chọn đặc trưng dựa trên các tài liệu nghiên cứu trước đó. Từ đó triển khai cài đặt thuật toán tra cứu trên ngôn ngữ lập trình C#.
- Thực nghiệm dựa trên xây dựng và đánh giá các kết quả nghiên cứu.

2. Các số liệu cần để thiết kế, tính toán

Các phương pháp trích chọn đặc trưng sau đó áp dụng vào bài toán tra cứu ảnh theo nội dung.

3. Địa điểm thực tập

LỜI CẢM ƠN

Em xin được bày tỏ lòng biết ơn sâu sắc tới thầy giáo, PGS.TS Ngô Quốc Tạo, Viện Công Nghệ Thông tin thuộc Viện Khoa học và công nghệ Việt Nam là người trực tiếp hướng dẫn, tận tình chỉ bảo em trong suốt quá trình làm đồ án.

Em xin chân thành cảm ơn các thầy cô giáo trong khoa Công nghệ thông tin - Trường Đại học Dân Lập Hải Phòng ,những người đã nhiệt tình giảng dạy và truyền đạt những kiến thức cần thiết trong suốt thời gian em học tập tại trường, để em hoàn thành tốt quá trình tốt nghiệp.

Cuối cùng em xin gửi lời cảm ơn đến gia đình, bạn bè, người thân đã ủng hộ động viên tinh thần để đồ án này được hoàn thành.

Em xin chân thành cảm ơn!

Hải Phòng, ngày 2 tháng 7 năm 2011

Sinh viên

Bùi Thị Thúy Nga

MỤC LỤC

MỤC LỤC	5
DANH MỤC CÁC HÌNH	8
DANH MỤC CÁC CHỮ VIẾT TẮT	9
LỜI NÓI ĐẦU	10
CHƯƠNG 1: TỔNG QUAN	12
1.1. Tổng quan về tra cứu ảnh theo nội dung	12
1.2. Các chức năng của một hệ thống tra cứu ảnh dựa vào nội dung tiêu biểu	12
1.2.1. Truy vấn người sử dụng	14
1.2.1.1. Truy vấn bởi ảnh mẫu (QBE – Query By Example).....	14
1.2.1.2. Truy vấn bởi đặc trưng (<i>QBF – Query By Feature</i>)	14
1.2.1.3. Các truy vấn dựa vào thuộc tính (<i>Attribute – Based queries</i>)	14
1.2.2. Đánh chỉ số nhiều chiều	15
1.3. Đánh giá hiệu năng tra cứu	15
1.4. Những hệ thống tra cứu ảnh theo nội dung	16
1.4.1. Hệ thống QBIC (Query By Image Content) của IBM	17
1.4.2. Hệ thống Virage của công ty Virage	17
1.4.3. Hệ thống RetrievalWare của tập đoàn công nghệ Excalibur	17
1.4.4. Hệ thống VisualSeek và WebSeek của đại học Columbia.....	18
1.4.5. Hệ thống Photobook của phòng thí nghiệm truyền thông MIT	18
1.4.6. Hệ thống Netra của Đại học California, Thư viện ảnh Alexandria	18
1.5. Kết luận	18
CHƯƠNG 2: CÁC KHÁI NIỆM CƠ BẢN	20
2.1. Màu sắc	20
2.1.1. Các không gian màu	20
2.1.1.1. Không gian màu RGB	21
2.1.1.2. Không gian màu CMY	21
2.1.1.3. Không gian màu L^*a^*b	22

2.1.1.4. Không màu HSV	22
2.1.2. Các đặc trưng về màu sắc	23
2.1.2.1. Lược đồ màu (Histogram)	23
2.1.2.2. Vector liên kết màu (Color Coherence Vector)	24
2.1.2.3. Tương quan màu.....	24
2.1.2.4. Các màu trội	25
2.1.2.5. Các mômen màu	25
2.2. Kết cấu	26
2.2.1. Ma trận đồng hiện (Co-occurrence Matrix)	27
2.2.2. Các đặc trưng Tamura	27
2.2.2.1. Thô (Coarseness)	27
2.2.2.2. Độ tương phản.....	28
2.2.2.3. Hướng.....	28
2.2.3. Các đặc trưng Wold.....	29
2.2.4. Mô hình tự hồi qui đồng thời SAR.....	30
2.2.5. Các đặc trưng lọc Gabor.....	31
2.2.6. Các đặc trưng biến đổi sóng	32
2.3. Hình dạng	33
2.3.1. Các bất biến mômen	33
2.3.2. Các góc uốn	34
2.3.3. Các ký hiệu mô tả Fourier	35
2.3.4. Hình tròn, độ lệch tâm, và hướng trục chính.....	36
2.4. Thông tin không gian	36
2.5. Phân đoạn.....	37
2.6. Độ đo	37
2.6.1. Khái niệm	38
2.6.2. Một số độ đo thông dụng.....	38
2.6.2.1. Khoảng cách Minkowsky:.....	38
2.6.2.2. Khoảng cách toàn phương.....	38

2.6.2.3. Khoảng cách Euclid:	38
2.6.2.4. Độ đo khoảng cách min-max	39
CHƯƠNG 3: KỸ THUẬT TRA CỨU ẢNH DỰA THEO NỘI DUNG	40
3.1. Màu sắc.....	40
3.1.1. Lược đồ màu.....	40
3.1.2. Vector liên kết màu (Color Coherence Vector).....	40
3.1.3. Đặc trưng tự tương quan màu (AutoCorrelogram)	41
3.2. Độ đo khoảng cách giữa các lược đồ màu	41
3.2.1. Khoảng cách dạng Minkowsky	41
3.2.2. Khoảng cách toàn phương	42
3.2.3. Độ đo khoảng cách min-max.....	43
3.3. Kỹ thuật dựa vào đặc trưng màu	43
3.3.1. Lược đồ màu toàn cục	44
3.3.2. Lược đồ màu cục bộ	45
CHƯƠNG 4: TRIỂN KHAI & THỰC NGHIỆM.....	46
4.1. Quy trình xây dựng phần mềm tra cứu ảnh tổng hợp.....	46
4.2. Lựa chọn tập mẫu.....	47
4.3. Lựa chọn phương pháp truy vấn ảnh.....	47
4.4. Xây dựng ứng dụng	48
4.5. Kết quả	49
KẾT LUẬN	52
TÀI LIỆU THAM KHẢO	53

DANH MỤC CÁC HÌNH

- Hình 1.1 Mô hình hệ thống tra cứu ảnh theo nội dung
- Hình 2.1 Không gian màu RGB
- Hình 2.2 Không gian màu CMY
- Hình 2.3 Không gian màu HSV
- Hình 3.1 Khoảng cách dạng Minkowsky
- Hình 3.2 Khoảng cách dạng toàn phương
- Hình 3.3 Ba ảnh và biểu đồ màu tương ứng
- Hình 4.1: Quy trình huấn luyện tập dữ liệu ảnh ban đầu
- Hình 4.2 Mô hình của ứng dụng
- Hình 4.3: Tập ảnh cơ sở dữ liệu
- Hình 4.4: Kiến trúc của ứng dụng
- Hình 4.5: Giao diện ứng dụng
- Hình 4.6: Tiến trình tra cứu
- Hình 4.7: Kết quả tra cứu
- Hình 4.8: So sánh giữa hai kỹ thuật sử dụng

DANH MỤC CÁC CHỮ VIẾT TẮT

KÝ HIỆU

DIỄN GIẢI

VIR	Visual Information Retrieval (Tra cứu thông tin)
CBIR	Content Based Image Retrieval (Tra cứu ảnh dựa theo nội dung)
RGB	Red Green Blue (Đỏ, Xanh lục, Xanh lơ)
HSV	Hue Saturation Value (Màu, sắc nét, cường độ)
CMY	Cyan Magenta Yellow (màu lục lam, màu đỏ tươi, màu vàng)
QBE	Query By Example (Truy vấn bởi ảnh mẫu)
QBF	Query By Feature (Truy vấn bởi đặc trưng)
CCV	Color Coherence Vectors (Véc tơ gắn kết màu)
SDF	Spectral Distribution Functions (Hàm phân bố phổ)
MLE	Maximum Likelihood Estimation (Ước lượng khả năng nhất)
MRSAR	multi-resolution simultaneous auto-regressive model (Mô hình tự hồi quy đồng thời đa phân giải)
SAR	Simultaneous Auto-Regressive (Tự hồi quy đồng thời)
MRF	Markov random field (Trường ngẫu nhiên Markov)
PWT	Pyramid-Structured Wavelet Transform (Biến đổi song cấu trúc hình chóp)
TWT	Tree-Structured Wavelet Transform (Biến đổi song cấu trúc cây)
RISAR	Rotation-Invariant SAR model (Mô hình SAR bất biến quay)

LỜI NÓI ĐẦU

Cùng với sự mở rộng của multimedia, với khối lượng hình ảnh và phim lưu trữ ngày càng lớn, cùng với sự phát triển của các thiết bị Điện tử, Tin học và Viễn thông đã thu hút ngày càng nhiều chuyên gia đi sâu vào nghiên cứu những công cụ cung cấp cho việc lấy thông tin dữ liệu ảnh từ nội dung của chúng. Trong thực tế, bài toán tra cứu ảnh có rất nhiều ứng dụng quan trọng. Các ứng dụng trong phức tạp hơn như so sánh mẫu vân tay, tìm kiếm ảnh tội phạm... được áp dụng trong ngành khoa học hình sự, cơ sở dữ liệu ảnh về địa lý, y học... làm cho lĩnh vực nghiên cứu này phát triển nhanh trong công nghệ thông tin.

Tuy nhiên khi số lượng ảnh được lưu trữ trở nên rất lớn thì vấn đề là phải có những phương pháp tổ chức cơ sở dữ liệu ảnh tốt cùng với những kỹ thuật tìm kiếm, tra cứu ảnh hiệu quả, có độ chính xác cao và có hiệu năng tốt. Việc tìm kiếm được một bức ảnh mong muốn trong hàng triệu bức ảnh thuộc đủ loại chủ đề khác nhau là rất khó khăn.

Khi số lượng ảnh trong một bộ sưu tập còn ít, việc nhận diện một bức ảnh hay việc so sánh sự giống và khác nhau giữa nhiều bức ảnh có thể thực hiện được bằng mắt thường, tuy nhiên khi có số lượng rất lớn ảnh thì việc so sánh bằng mắt thường là rất khó khăn, đòi hỏi phải có những phương pháp hiệu quả và chính xác hơn.

Tìm kiếm theo cách thông thường dựa trên văn bản giờ đây được bổ sung bởi truy vấn dựa theo nội dung, tìm vào khía cạnh nhân thức thông tin ảnh. Thực hiện truy vấn ở mức nhận thức đòi hỏi những phương thức mới, cho phép chỉ định đến những thuộc tính liên quan đến thị giác cần tìm. Giao diện đồ họa phải hỗ trợ đặc tả những ví dụ đó như là những mẫu có sẵn. Khi đó người dùng trong một vòng lặp, mô hình giao diện sao cho người dùng có thể truy cập vào sự giống nhau giữa những đối tượng.

Trước năm 1990, người ta thường sử dụng phương pháp tra cứu ảnh theo văn bản (Text Based Image Retrieval). Theo cách này người ta sẽ gán cho mỗi bức ảnh một lời chú thích phù hợp với nội dung hoặc một đặc điểm nào đó của ảnh, sau đó việc tra cứu ảnh được thực hiện dựa trên những lời chú thích này. Phương pháp này khá đơn giản, tuy nhiên lại không thể áp dụng để tra cứu các cơ sở dữ liệu ảnh có số lượng ảnh lớn và kết quả tra cứu thì mang tính chủ quan và cảm ngữ cảnh [8].

Một trong những phương pháp được nhiều người quan tâm nghiên cứu hiện nay là phương pháp “Tra cứu ảnh dựa theo nội dung” (Content-Based Image Retrieval). Ý

tương phương pháp này là trích chọn các đặc điểm dựa vào nội dung trực quan của ảnh như màu sắc, kết cấu, hình dạng và bố cục không gian của ảnh để làm cơ sở cho việc tra cứu, sắp xếp, tổ chức cơ sở dữ liệu ảnh [8].

Nội dung của đề tài này giới thiệu cơ sở lý thuyết và các ứng dụng chính của một số phương pháp tra cứu ảnh, trong đó đi sâu vào giới thiệu phương pháp “tra cứu ảnh theo nội dung”. Trên những cơ sở đó tiến hành thử nghiệm một phương pháp cụ thể để xây dựng một chương trình phần mềm tra cứu ảnh cho phép đọc vào một ảnh mẫu và tìm kiếm những ảnh tương tự với ảnh mẫu trong một tập hợp các ảnh cho trước theo đặc trưng màu sắc.

Nội dung của đề tài gồm các phần chính sau:

Chương 1: Tổng quan về tra cứu ảnh theo nội dung.

Chương 2: Các khái niệm cơ bản.

Chương 3: Kỹ thuật tra cứu ảnh theo nội dung.

Chương 4: Triển khai và thực nghiệm.

Chương 1: TỔNG QUAN

1.1. Tổng quan về tra cứu ảnh theo nội dung

Tra cứu ảnh dựa theo nội dung (CBIR) có nguồn gốc từ năm 1992. Tra cứu ảnh theo nội dung là tra cứu thông tin trực quan (VIR - Visual Information Retrieval). Tra cứu thông tin trực quan là chủ đề nghiên cứu mới trong lĩnh vực công nghệ thông tin. Tương tác với nội dung trực quan là cách thiết yếu nhất để tra cứu thông tin trực quan. Các yếu tố trực quan như màu sắc, kết cấu, hình dạng và bố cục không gian trực tiếp liên quan đến khía cạnh của cảm nhận nội dung ảnh, cùng với các khái niệm ở mức cao như ý nghĩa đối tượng, khung cảnh trong ảnh, được dùng như là manh mối cho tìm kiếm hình ảnh với nội dung tương tự từ cơ sở dữ liệu.

Tra cứu thông tin, mô hình hóa và thể hiện dữ liệu trực quan, phân tích và xử lý phim ảnh, nhận dạng và thị giác máy tính (computer vision), tổ chức cơ sở dữ liệu đa phương tiện (multimedia), lập chỉ mục đa chiều, mô hình hóa tâm lý hành vi người dùng, hệ thống tương tác người - máy và trực quan hóa dữ liệu, là các lĩnh vực nghiên cứu quan trọng nhất đóng góp cho tra cứu thông tin trực quan.

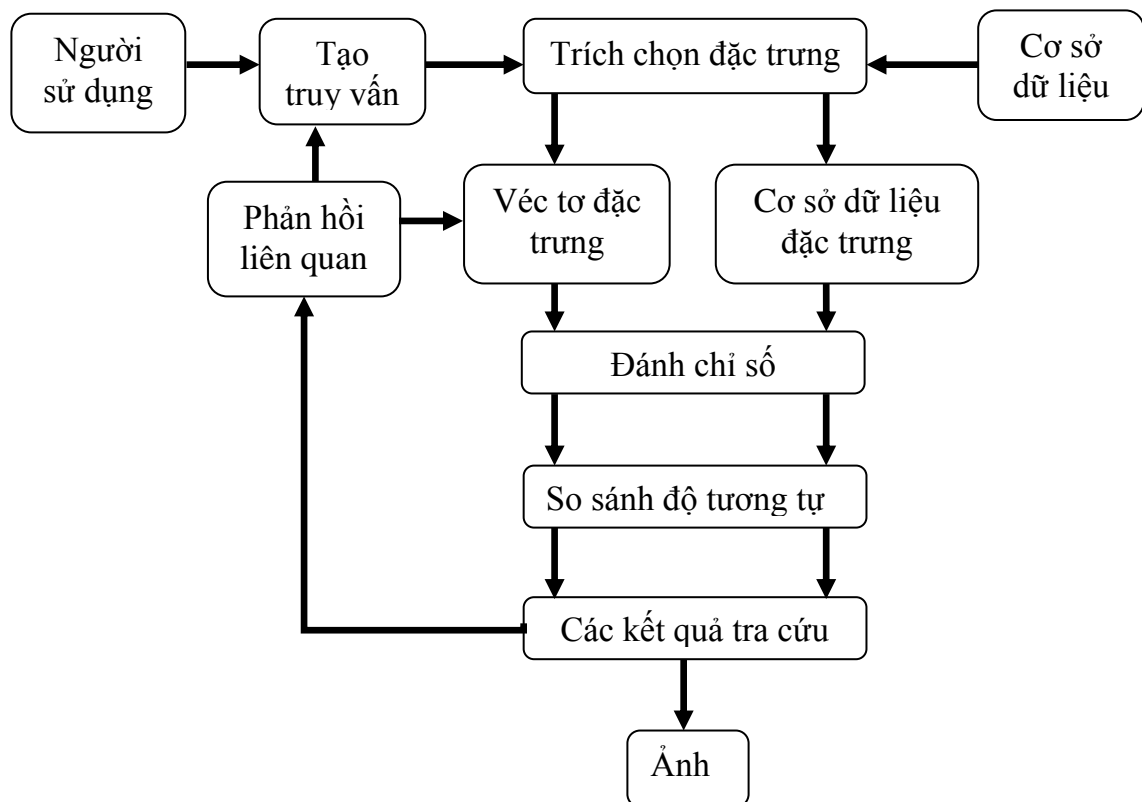
Các yếu tố mô tả nội dung có liên quan đến các đặc trưng cảm nhận như màu sắc, kết cấu, hình dạng, mối liên hệ về không gian chuyển động. Các yếu tố cơ sở nguồn gốc của màu sắc như đối tượng, vai trò, sự kiên, hay các thông tin có liên hệ với màu sắc như cảm giác, cảm tưởng hay ý nghĩa của hình ảnh. Do vậy, phân tích ảnh, nhận dạng và thị giác máy tính đóng vai trò cơ bản trong các hệ thống tra cứu thông tin trực quan. Nó cho phép tự động trích chọn hầu hết các thông tin trực quan, thông qua việc phân tích phân bố điểm ảnh và rút ra các độ đo nội dung trực quan.

Trong đề án này, em chỉ tập trung vào đặc trưng màu cho ứng dụng tra cứu ảnh dựa theo nội dung.

1.2. Các chức năng của một hệ thống tra cứu ảnh dựa vào nội dung tiêu biểu

Một hệ thống tra cứu ảnh dựa theo nội dung (Content-Based Image Retrieval) tiêu biểu không chỉ xử lý các nguồn thông tin khác nhau mà ở các khuôn dạng khác nhau (như văn bản, hình ảnh và video) mà còn giải quyết được các nhu cầu của người sử dụng. Về cơ bản hệ thống phân tích cả các nội dung của nguồn thông tin cũng như các truy vấn sử dụng, và đem so sánh các nội dung này để tra cứu các mục tin liên quan. Các chức năng chính của một hệ thống tra cứu ảnh bao gồm các nội dung sau:

- Phân tích các nội dung của các nguồn thông tin, và biểu diễn các nội dung của các nguồn được phân tích theo cách thích hợp cho so sánh các truy vấn sử dụng. Bước này thông thường cần rất nhiều thời gian do nó phải xử lý lần lượt các thông tin nguồn (các ảnh) trong cơ sở dữ liệu. Tuy nhiên, bước này được thực hiện chỉ một lần.
- Phân tích các truy vấn người sử dụng và biểu diễn chúng ở dạng thích hợp cho so sánh với cơ sở dữ liệu nguồn. Bước này là tương tự bước trước, nhưng chỉ được áp dụng với ảnh truy vấn.
- Định nghĩa một chiến lược để so sánh các truy vấn tìm kiếm với thông tin trong cơ sở dữ liệu được lưu trữ. Tra cứu thông tin liên quan theo một cách hiệu quả. Bước này được thực hiện trực tuyến và được yêu cầu rất nhanh. Các kỹ thuật đánh chỉ số hiện đại có thể được sử dụng để tổ chức lại không gian đặc trưng để tăng tốc quá trình đối sánh.
- Thực hiện các điều chỉnh cần thiết trong hệ thống dựa trên phản hồi từ người sử dụng hoặc các ảnh được tra cứu.



Hình 1.1: Mô hình hệ thống tra cứu ảnh dựa theo nội dung

1.2.1. Truy vấn người sử dụng

Có nhiều cách gửi một truy vấn trực quan. Một phương pháp truy vấn tốt là một phương pháp tự nhiên đối với người sử dụng cũng như thu được đủ thông tin từ người sử dụng để trích rút các kết quả có ý nghĩa. Các phương pháp truy vấn dưới đây sử dụng phổ biến trong nghiên cứu tra cứu ảnh dựa theo nội dung:

1.2.1.1. Truy vấn bởi ảnh mẫu (QBE – Query By Example)

Trong kiểu truy vấn này, người sử dụng hệ thống chỉ rõ một ảnh truy vấn đích dựa vào cơ sở dữ liệu sẽ tìm kiếm được các ảnh tương tự nhất. Ảnh truy vấn có thể là ảnh thường, một ảnh quét có độ phân giải thấp, hoặc một phác thảo sử dụng công cụ mô tả giao diện đồ họa. Ưu điểm của hệ thống này đó là nó là một cách tự nhiên cho những người sử dụng chung và để tra cứu ảnh trong cơ sở dữ liệu.

1.2.1.2. Truy vấn bởi đặc trưng (QBF – Query By Feature)

Trong hệ thống QBF tiêu biểu, những người sử dụng chỉ rõ truy vấn bằng việc chỉ rõ các đặc trưng được quan tâm trong tìm kiếm. Như người dùng có thể truy vấn cơ sở dữ liệu ảnh bằng cách đưa ra một lệnh để tra cứu tất cả các ảnh có góc phần tư trái chứa 25% pixel màu vàng. Truy vấn này được chỉ rõ bằng việc sử dụng các công cụ giao diện đồ họa chuyên dụng. Những người sử dụng hệ thống tra cứu ảnh chuyên nghiệp thì có thể thấy loại truy vấn này là bình thường, nhưng người sử dụng chung thì không thể. QBIC [3] là một ví dụ về hệ thống tra cứu ảnh dựa theo nội dung sử dụng loại phương pháp truy vấn này.

1.2.1.3. Các truy vấn dựa vào thuộc tính (Attribute – Based queries)

Các truy vấn dựa vào thuộc tính sử dụng các chú thích văn bản, trích rút bởi nỗ lực của con người, như một khoá tra cứu chính. Loại biểu diễn này đòi hỏi một độ trừu tượng cao khó để thu được bằng các phương pháp tự động hoàn toàn do một ảnh chứa một số lượng lớn các thông tin mà khó tóm tắt khi sử dụng một số ít các từ khoá. Phương pháp này nhìn chung là nhanh hơn và dễ hơn để thực hiện, nhưng có một độ chủ quan và nhập nhằng cao xuất hiện như chúng ta đã đề cập. Hầu hết nỗ lực nghiên cứu và thương mại tập trung vào xây dựng các hệ thống thực hiện tốt với các phương pháp QBE.

1.2.2. Đánh chỉ số nhiều chiều

Để thực hiện tra cứu ảnh dựa vào nội dung đối với các cơ sở dữ liệu ảnh lớn, các kỹ thuật đánh chỉ số nhiều chiều cần được sử dụng. Có ba cộng đồng nghiên cứu chính đóng góp vào lĩnh vực này: hình học tính toán, quản trị cơ sở dữ liệu, và nhận dạng mẫu. Các kỹ thuật đánh chỉ số nhiều chiều phổ biến đã có gồm thuật toán bucketing, cây k-d, cây k-d ưu tiên, cây tứ phân, cây K-D-B, cây hB, cây R-tree và các biến thể của nó cây R^+ và R^* .

Lịch sử của các kỹ thuật đánh chỉ số nhiều chiều có thể quay lại giữa những năm 1970, khi các phương pháp khối, cây tứ phân, và cây $k + d$ được giới thiệu lần đầu tiên. Tuy nhiên, hiệu năng của nó thì còn xa so với mong muốn. Bị thúc đẩy bởi nhu cầu cấp bách của đánh chỉ số không gian từ các hệ thống GIS và CAD, Guttman đã đề xuất cấu trúc đánh chỉ số R-tree [2]. Dựa trên nghiên cứu của ông, nhiều biến thể khác của R-tree đã được phát triển [1]. Năm 1990, Beckmann và Kriegel đã đề xuất biến thể R-tree động, R^* tree trong [5]. Tuy nhiên, R^* tree không thực hiện tốt với số chiều lớn hơn 20.

1.3. Đánh giá hiệu năng tra cứu

Để đánh giá hiệu năng của hệ thống tra cứu, người ta đưa ra hai số đo đó là độ thu hồi (recall) và độ chính xác (precision). Các số đo này được lấy từ hệ thống tra cứu thông tin truyền thống.

Đối với một truy vấn q , tập hợp các ảnh trong cơ sở dữ liệu thích hợp với truy vấn q được ký hiệu là $R(q)$ còn kết quả tra cứu của truy vấn q được ký hiệu là $Q(q)$.

Độ chính xác của việc tra cứu được định nghĩa là tỉ lệ những kết quả thu được thích hợp với truy vấn.

$$precision = \frac{|Q(q) \cap R(q)|}{|Q(q)|} \quad (1)$$

Độ thu hồi là tỉ lệ những kết quả thích hợp do truy vấn trả lại:

$$recall = \frac{|Q(q) \cap R(q)|}{|R(q)|} \quad (2)$$

Việc lựa chọn $R(q)$ lại rất không ổn định do sự đa dạng của cách hiểu về một bức ảnh. Hơn nữa, khi số lượng ảnh thích hợp lại lớn hơn số lượng ảnh hệ thống tìm

được thì lúc đó khái niệm độ thu hồi trở thành vô nghĩa. Do đó, độ chính xác và độ thu hồi chỉ là các mô tả ở dạng thô về hiệu năng của một hệ thống tra cứu. Gần đây *MPEG7* có đề nghị một cách đánh giá mới về hiệu năng của các hệ thống tra cứu gọi là *ANMRR* (average normalized modified retrieval rank). Theo cách này độ chính xác và độ thu hồi được kết hợp thành một số đo duy nhất. Ký hiệu số lượng ảnh hoàn toàn đúng với truy vấn q là $N(q)$ và số lượng lớn nhất của các ảnh hoàn toàn đúng với tất cả Q truy vấn tức là $\max\{N(q_1), N(q_2), \dots, N(q_0)\}$ là M . Sau đó đối với mỗi truy vấn q thì mỗi bức ảnh hoàn toàn đúng k được gán một giá trị xếp hạng $rank(k)$ giá trị này là thứ hạng của nó trong số những ảnh hoàn toàn đúng nếu ảnh đó nằm trong k kết quả truy vấn đầu tiên ($k = \min\{N(q), M\}$) hoặc có giá trị $k+1$ nếu ảnh đó không nằm trong k kết quả truy vấn đầu tiên.

Thứ hạng trung bình $AVR(q)$ đối với truy vấn q được tính:

$$AVR(q) = \sum_{k=1}^{N(q)} \frac{rank(k)}{N(q)} \quad (3)$$

Thứ hạng tra cứu sửa đổi $MRR(q)$ được tính là:

$$MRR(q) = AVG(q) - 0.5 - 0.5 * N(q) \quad (4)$$

$MRR(q)$ nhận giá trị 0 khi tất cả các ảnh hoàn toàn đúng đều nằm trong k kết quả truy vấn đầu tiên.

Thứ hạng tra cứu sửa đổi và chuẩn hoá $NMRR(q)$ nhận giá trị từ 0 đến 1 được tính:

$$NMRR(q) = \frac{MRR(q)}{k + 0.5 - 0.5 * N(q)} \quad (5)$$

Thứ hạng tra cứu sửa đổi và chuẩn hoá trung bình $ANMRR$ xét trên tất cả các truy vấn Q là:

$$ANMRR = \frac{1}{Q} \sum_{q=1}^Q NMRR(q) \quad (6)$$

1.4. Những hệ thống tra cứu ảnh theo nội dung

Những năm gần đây, có nhiều hệ thống tra cứu ảnh, các hệ thống nghiên cứu và hệ thống thương mại đã được xây dựng. Dưới đây, là một số hệ thống của CBIR đã được phát triển.

1.4.1. Hệ thống QBIC (Query By Image Content) của IBM

QBIC chuẩn cho truy vấn bởi nội dung ảnh, là hệ thống tra cứu ảnh dựa vào nội dung thương mại đầu tiên. QBIC hỗ trợ chính các truy vấn dựa vào các ảnh mẫu, các phác thảo và các bản vẽ được người sử dụng xây dựng, và các mẫu kết cấu và màu được lựa chọn. QBIC là kỹ thuật flood-fill tăng cường. Một công cụ phác thảo để trợ giúp những người sử dụng theo dõi các điểm biên đối tượng dựa trên khái niệm “snakes” được phát triển trong nghiên cứu thị giác máy tính. Công cụ này nhận một đường cong phác thảo của người sử dụng và căn lề tự động nó với các điểm biên ảnh gần cạnh. Nó tìm đường cong cực đại hoá độ lớn gradient ảnh dọc đường cong. Sau khi nhận biết đối tượng, QBIC sẽ tính toán các đặc trưng của mỗi đối tượng và ảnh. Các đặc trưng như sau: Màu, Kết cấu, Hình, Phác thảo.

Ngay khi các đặc trưng được mô tả, các độ đo tương tự được sử dụng để nhận các ảnh tương tự. Trong bước tìm kiếm, QBIC phân biệt giữa “các ảnh” và “các đối tượng”. Một ảnh là một ảnh màu đầy đủ hoặc frame đơn của video và một đối tượng là một phần của ảnh. QBIC tính toán các đặc trưng sau: Các đối tượng, các ảnh. QBIC là một trong số ít các hệ thống nhận vào bản miêu tả đánh chỉ số đặc trưng nhiều chiều.

1.4.2. Hệ thống Virage của công ty Virage

Virage là một máy tìm kiếm ảnh dựa vào nội dung được phát triển tại liên hợp Virage. Tương tự với QBIC, Virage hỗ trợ các truy vấn trực quan dựa trên màu, thành phần cấu tạo (bố cục màu), kết cấu và cấu trúc (thông tin đường bao đối tượng). Nhưng Virage tiến một bước xa hơn QBIC. Nó cũng hỗ trợ các kết hợp tùy ý của các truy vấn trên. Người sử dụng có thể điều chỉnh các trọng số được kết hợp với các đặc trưng theo sự nhấn mạnh riêng của họ. Jeffrey và cộng sự đã đề xuất tiếp một khuôn khổ mở cho quản lý ảnh. Họ đã phân loại các đặc trưng trực quan thành tổng quát (như màu, hình, hoặc kết cấu) và lĩnh vực cụ thể (nhận dạng mặt người, phát hiện khối u, v.v.).

1.4.3. Hệ thống RetrievalWare của tập đoàn công nghệ Excalibur

RetrievalWare là một máy tra cứu ảnh dựa vào nội dung được phát triển bởi tập đoàn công nghệ Excalibur. Máy tìm kiếm gần đây của họ sử dụng màu, hình, kết cấu, độ sáng, bố cục màu, và hướng tỷ lệ của ảnh, như các đặc trưng truy vấn. Người sử

dụng có thể điều chỉnh các trọng số kết hợp với mỗi đặc trưng này trong quá trình tìm kiếm.

1.4.4. Hệ thống VisualSeek và WebSeek của đại học Columbia

VisualSEEK là một máy tìm kiếm đặc trưng trực quan và WebSEEK là một máy tìm kiếm văn bản/ảnh trên web, cả hai sản phẩm đã được phát triển tại đại học Columbia. VisualSEEK hỗ trợ các truy vấn dựa vào cả các đặc trưng trực quan và các quan hệ không gian của chúng. WebSEEK là một máy tìm kiếm hướng Web. Nó gồm ba module chính đó là module tập hợp ảnh/video, module phân loại chủ đề và đánh chỉ số, và module tìm kiếm, duyệt, và tra cứu và hỗ trợ tra cứu dựa trên các từ khóa và nội dung trực quan.

1.4.5. Hệ thống Photobook của phòng thí nghiệm truyền thông MIT

Photobook là một tập các công cụ tương tác để duyệt và tìm kiếm các ảnh được phát triển tại phòng thí nghiệm MIT Media. Hệ thống này cho phép người sử dụng tra cứu ảnh dựa trên màu sắc, hình dạng và kết cấu. Hệ thống như 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. Cho phép người sử dụng trực tiếp đưa những yêu cầu truy vấn của họ với những lĩnh vực khác nhau, mỗi lĩnh vực thì họ có thể thu được những mẫu truy vấn tối ưu. Trong phiên bản Photobook gần đây nhất của nó, Picard và cộng sự đã đề xuất gồm những người sử dụng trong lập và chú thích tra cứu ảnh.

1.4.6. Hệ thống Netra của Đại học California, Thư viện ảnh Alexandria

Netra là một hệ thống tra cứu ảnh nguyên mẫu được phát triển trong dự án ADL. Netra sử dụng thông tin màu, kết cấu, hình, và vị trí không gian trong các vùng ảnh được phân đoạn để tìm kiếm và tra cứu các vùng tương tự từ cơ sở dữ liệu. Các đặc trưng nghiên cứu chính của hệ thống Netra là phân tích kết cấu dựa trên lọc Gabor, xây dựng từ điển ảnh dựa trên mạng neural và phân đoạn vùng dựa vào luồng biên.

1.5. Kết luận

Trong chương này, em đã giới thiệu kỹ thuật cơ bản về tra cứu ảnh dựa vào nội dung, bao gồm: tra cứu thông tin thị giác, các chức năng tiêu biểu của hệ thống tra cứu ảnh dựa vào nội dung, đánh giá hiệu năng tra cứu và trình bày một số hệ thống tra cứu ảnh theo nội dung. Đặc biệt em nhấn mạnh vào các đặc trưng thị giác. Trong đồ án này

em tập trung vào vấn đề nâng cao hiệu năng hệ thống tra cứu ảnh dựa vào nội dung thị giác thông qua sử dụng đặc trưng về màu sắc.

Chương 2: CÁC KHÁI NIỆM CƠ BẢN

2.1. Màu sắc

Màu là đặc trưng trực quan quan trọng đầu tiên và đơn giản nhất cho đánh chỉ số và tra cứu các ảnh. Nó cũng là đặc trưng được sử dụng phổ biến nhất trong lĩnh vực.

Một ảnh màu tiêu biểu được thu từ một camera số, hoặc được tải xuống từ Internet thường có ba kênh màu (các ảnh xám chỉ có một kênh, các ảnh đa phổ có thể có nhiều hơn ba kênh). Tuy nhiên, các giá trị của dữ liệu ba chiều từ ảnh màu không cho chúng ta một mô tả chính xác của màu trong ảnh, nhưng cho vị trí của các pixel này trong không gian màu. Các pixel có các giá trị $(1,1,1)$ sẽ xuất hiện khác nhau về màu trong các không gian màu khác nhau. Vì thế mô tả đầy đủ của một ảnh màu tiêu biểu sẽ bao gồm thông tin không gian hai chiều cho biết nơi của pixel màu là trong miền không gian, chúng ta đang nói đến không gian màu, và dữ liệu màu ba chiều cho biết nơi của pixel màu là trong không gian màu này.

Các tín hiệu màu một hoặc hai chiều cũng được sử dụng rộng rãi trong CBIR đặc biệt trong các ứng dụng mà các điều kiện thu ảnh tương phản mạnh là quan trọng. Thông tin màu (Hue) được sử dụng trong các ứng dụng nơi chỉ sự khác biệt giữa chất liệu của các đối tượng trong cảnh là quan trọng. Nếu coi thông tin màu của một ảnh như một tín hiệu một, hai, hoặc ba chiều, phân tích tín hiệu bằng việc sử dụng ước lượng mật độ xác suất là cách đơn giản nhất để mô tả thông tin màu của ảnh. Lược đồ là công cụ đơn giản nhất. Các cách mô tả thông tin màu khác trong tra cứu ảnh theo nội dung bao gồm sử dụng các không gian màu, véc tơ liên kết màu, tương quan màu, các màu trội, và các mômen màu.

2.1.1. Các không gian màu

Một không gian màu là một mô hình đại diện cho màu về giá trị độ sáng, một không gian màu xác định bao nhiêu thông tin màu được thể hiện. Nó định nghĩa không gian 1, 2, 3 hoặc 4 chiều mà mỗi chiều của nó, còn gọi là thành phần, đại diện cho những giá trị độ sáng. Một thành phần màu còn được gọi là một kênh màu. Mỗi điểm ảnh trong ảnh có thể được đại diện bởi không gian màu 3 chiều. Những không gian màu thường được dùng bao gồm: RGB, CMY, CIE L^*a^*b , HSV,... Cho đến nay vẫn chưa có sự thông nhất không gian nào tốt nhất. Một số không gian màu thường gặp.

2.1.1.1. Không gian màu RGB

Không gian RGB là không gian được sử dụng rộng rãi trong việc hiển thị hình ảnh. Ý tưởng tạo ra không gian màu RGB từ cách mà mắt con người hoạt động. Có những cơ quan cảm nhận để phát hiện ra 3 màu khác nhau: đỏ (red), lục (green), lam (blue). Không gian màu RGB cũng gồm 3 thành phần màu: Red, Green, Blue. Những thành phần này được gọi là màu gốc để cộng vào, vì mỗi màu được tạo nên bằng cách cộng thêm các phần tử Black(0,0,0). Khuôn dạng của không gian RGB là định dạng phổ biến nhất của ảnh số, lý do chính là tính tương thích với màn hình hiển thị chính là màn hình vi tính. Tuy nhiên không gian màu RGB có hạn chế lớn nhất là không phù hợp với cách con người cảm nhận về màu sắc. Do vậy, không phù hợp cho việc tìm kiếm ảnh.



Hình 2.1: Không gian RGB

2.1.1.2. Không gian màu CMY

Không gian màu CMY được dùng trong in ấn. CMY là viết tắt của Cyan-Magenta-Yellow (màu lục lam, màu đỏ tươi, màu vàng), đó là ba màu chính tương ứng với ba màu mực in. Chúng được gọi là những màu gốc để trừ, vì mỗi màu trong không gian CMY được tạo ra thông qua việc hấp thụ độ sáng. Cyan hấp thụ sự chiếu sáng của màu đỏ, Magenta hấp thụ màu xanh lục, Yellow hấp thụ màu xanh dương.

Công thức chuyển đổi từ không gian màu RGB sang không gian màu CMY đó là:

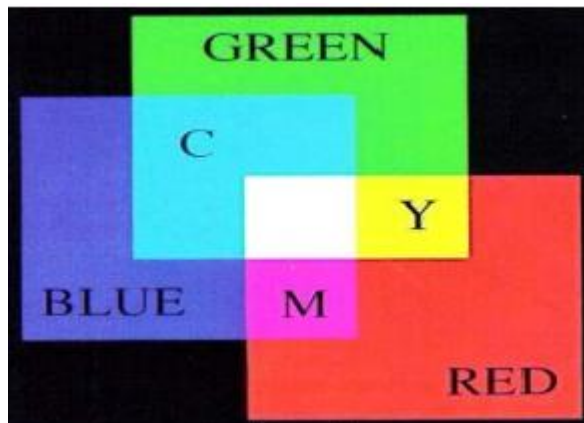
$$C = 1 - R$$

$$M = 1 - G$$

$$Y = 1 - B$$

2.1

Hệ thống màu CMY dường như là một sự đảo ngược của hệ thống màu RGB. Đặc tính của nó là sự đơn giản, ứng dụng nhiều trong thực tế. Tuy nhiên khuyết điểm của nó cũng tương tự như không gian màu RGB, tức là cách mã hóa khác với cách mà con người cảm nhận về màu sắc. Không thích hợp cho bài toán tra cứu ảnh dựa theo nội dung.



Hình 2.2: Không gian màu CMY

2.1.1.3. Không gian màu L*a*b

Mô hình L*a*b được đề cử bởi CIE cho việc lượng hóa sự khác biệt của màu sắc trong vật chiếu sáng của ánh sáng ban ngày. Tuy nhiên nó có một sự chuyển đổi để tính toán cho việc thích nghi với những nguồn sáng. Một màu được xác định bằng hai tọa độ x và y. Độ sáng L* dựa trên độ đo giác quan của thành phần a*, b* là tọa độ màu. Các màu khác nhau theo một hướng duy nhất là xấp xỉ bằng nhau trong không gian màu này. Tuy nhiên, việc chuyển sang không gian màu RGB là không tuyến tính. Đây là hệ thống màu có sự tách riêng ánh sáng và màu sắc. Do đó, cũng có khả năng lớn cho việc tra cứu ảnh dựa vào nội dung.

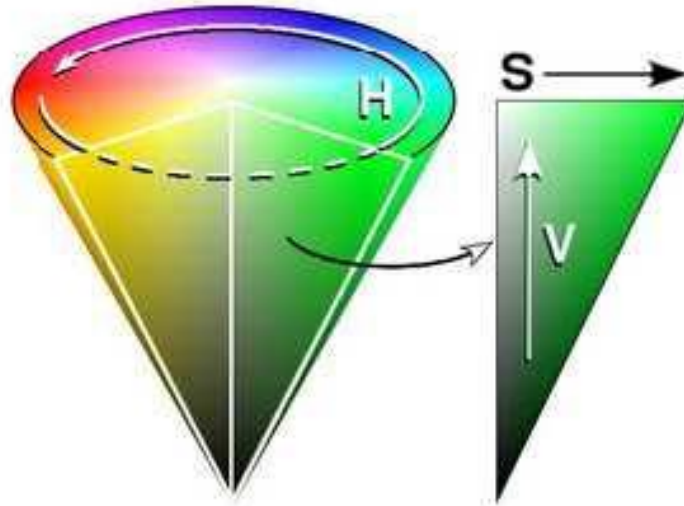
2.1.1.4. Không màu HSV

Mô hình HSV (Hue, Saturation, Value), còn được gọi là HSB (Hue, Saturation, Brightness) định nghĩa một không gian màu gồm 3 thành phần tạo nên:

- Hue: là loại màu (màu đỏ, xanh hay vàng,...)
- Saturation: là độ thuần khiết của màu.

- Value: là độ sáng của màu.

Mô hình HSV được tạo ra từ năm 1978 bởi Ray Smith. Nó là một phép biến đổi phi tuyến của không gian màu RGB. Mô hình HSV giúp tách bạch màu (H, S) và độ sáng (V), phù hợp với cảm nhận của con người.



Hình 2.3: Không gian màu HSV

2.1.2. Các đặc trưng về màu sắc

2.1.2.1. Lược đồ màu (Histogram)

Lược đồ màu là đại lượng đặc trưng cho phân bố màu cục bộ của ảnh.

$$H[i] = \frac{n_i}{n} \quad (2.2)$$

Trong đó: i là một bin màu, nếu ảnh xám thì $i \in [0, 255]$, nếu ảnh màu RGB thì $i \in [0, 2^{24}]$

n_i : là số điểm ảnh có giá trị màu là i

n : tổng số các pixel trong ảnh.

Ý nghĩa của lược đồ màu đó là: Đối với mỗi điểm ảnh trong ảnh I , H_c thể hiện xác suất điểm ảnh đó có màu là c_i . Không có mang thông tin về không gian.

Ưu điểm:

- Tính toán lược đồ màu ít tốn chi phí, đơn giản, nhanh chóng.
- Lược đồ màu bất biến với một số phép hình học Affine: tịnh tiến, xoay, co giãn.

Nhược điểm: Lược đồ màu chỉ phân bố toàn cục về màu của ảnh mà không xét đến yếu tố cục bộ về vị trí, làm mất thông tin không gian về quan hệ không gian giữa các màu. Do đó, có thể có nhiều ảnh khác nhau nhưng lại có cùng lược đồ màu.

Ứng dụng: Được dùng nhiều trong việc phân đoạn video và tra cứu thông tin thị giác.

2.1.2.2. Vector liên kết màu (Color Coherence Vector)

Trong [3] đã đề cập đến cách tiếp cận khác liên kết thông tin không gian vào lược đồ màu, các vectơ liên kết màu - CCV (*Color Coherence Vectors*), được đề xuất mỗi bin lược đồ được phân thành hai loại đó là: liên kết nếu nó thuộc về một vùng màu đồng nhất lớn, hoặc không gắn kết nếu nó không thuộc về một vùng màu đồng nhất lớn. Cho α_i biểu thị số các pixel gắn kết trong bin màu thứ i và β_i biểu thị số các pixel không gắn kết trong một ảnh. Thì, Vector liên kết màu của một ảnh được định nghĩa bằng vectơ $\langle (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_N, \beta_N) \rangle$

Trong đó: $\langle \alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_N + \beta_N \rangle$ là lược đồ màu của ảnh.

Do thông tin không gian thêm vào nó, đã được chỉ ra rằng Vector liên kết màu cung cấp các kết quả tra cứu tốt hơn lược đồ màu, đặc biệt cho các ảnh hoặc có phần lớn màu đồng nhất hoặc có phần lớn các vùng kết cấu.

2.1.2.3. Tương quan màu

Tương quan màu không chỉ để mô tả các phân bố màu của các pixel, mà còn tương quan không gian của các cặp màu. Một tương quan màu là một bảng được đánh chỉ số bởi các cặp màu. Với mỗi pixel có màu i trong ảnh, là xác suất tìm thấy một pixel có màu j cách pixel ban đầu một khoảng cách k . Cho I biểu diễn toàn bộ tập các pixel ảnh và $I_{c(i)}$ biểu diễn tập các pixel có màu $c(i)$. Thì tương quan màu được định nghĩa bằng:

$$\gamma_{i,j}^k = \frac{\Pr \{ p_2 \in I_{c(j)} \mid p_1 \in I_{c(i)}, |p_1 - p_2| = k \}}{\Pr \{ p_2 \in I_{c(j)} \}} \quad (3)$$

Trong đó, $i, j \in \{1, 2, \dots, N\}$.

$k \in \{1, 2, \dots, d\}$.

$|p_1 - p_2|$: là khoảng cách giữa các pixel p_1 và p_2 .

Kích thước của Correlogram là $O(n^2 d)$.

Khi chọn d để tính Correlogram, ta cần chú ý vấn đề sau:

- Giá trị d lớn thì cần nhiều chi phí tính toán và không gian lưu trữ.
- Giá trị d nhỏ có thể giảm giá trị lưu trữ của đặc trưng.

So sánh với lược đồ màu và véc tơ liên kết màu, tự tương quan màu cho các kết quả tra cứu tốt hơn, nhưng cũng cho chi phí tính toán cao hơn do nó có chiều cao.

2.1.2.4. Các màu trội

Các lược đồ màu thì rất thưa và thông thường một số nhỏ các màu là đủ để miêu tả thông tin màu trong một ảnh màu, các màu trội được sử dụng để mô tả nội dung màu của một ảnh. Một phân cụm màu được thực hiện để thu các màu trội đại diện của nó và phần trăm tương ứng của nó. Mỗi màu đại diện và phần trăm tương ứng của nó 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 định nghĩa là một tập các cặp thuộc tính:

$$F = \{ \{c_i, p_i\}, i = 1..N \} \quad (4)$$

Trong đó, N là tổng số các cụm màu trong ảnh (N có thể thay đổi phụ thuộc vào ảnh cụ thể).

c_i là một véc tơ màu ba chiều.

p_i là phần trăm của nó.

$$\sum_i p_i = 1.$$

2.1.2.5. Các mômen màu

Các mômen màu là các mômen thống kê của các phân bố xác suất của các màu. Các mômen màu được sử dụng thành công trong nhiều hệ thống tra cứu ảnh như QBIC [2], đặc biệt khi ảnh chứa chính xác đối tượng. Các mômen màu bậc nhất (trung bình), bậc hai (phương sai) và bậc ba (độ lệch), đã được chứng minh là hiệu quả trong biểu diễn các phân bố màu của các ảnh.

Về mặt toán học, ba mômen đầu tiên được xác định bởi công thức:

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

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (2.5)$$

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

Trong đó, f_{ij} là giá trị của thành phần màu thứ i của pixel ảnh j .

N là số các pixel trong ảnh.

Thông thường mômen màu thực hiện tốt hơn nếu nó được xác định bởi không gian màu L*a*b* chứ không chỉ duy nhất bởi không gian màu HSV. Sử dụng thêm mômen bậc ba cải tiến hiệu năng tra cứu so với chỉ sử dụng các mômen bậc nhất và bậc hai. phân loại thành ba lớp đó là dạng Minkowsky và dạng toàn phương. Tuy nhiên, mômen bậc ba này thường làm cho biểu diễn đặc trưng nhạy cảm hơn với sự thay đổi cảnh và do đó có thể giảm hiệu năng.

Do chỉ 9 số (ba mômen cho mỗi một trong ba thành phần màu) được sử dụng để biểu diễn nội dung màu của mỗi ảnh, các mômen màu là một biểu diễn nén so với các đặc trưng màu khác. Do biểu diễn nén này, nó có thể cũng giảm khả năng phân biệt. Thông thường, các mômen màu có thể được sử dụng như sơ duyệt lần đầu để giảm không gian tìm kiếm trước khi các đặc trưng màu phức tạp khác được sử dụng cho tra cứu.

2.2. Kết cấu

Kết cấu là một đặc tính quan trọng khác của ảnh. Các biểu diễn kết cấu đa dạng đã được nghiên cứu trong nhận dạng mẫu và thị giác máy tính. Về cơ bản, các phương pháp biểu diễn kết cấu có thể được phân ra thành hai loại: cấu trúc và thống kê. Các phương pháp cấu trúc bao gồm toán tử hình thái và đồ thị kề, mô tả kết cấu bởi nhận dạng cấu trúc gốc và các luật sắp đặt của chúng. Các phương pháp thống kê bao gồm các kỹ thuật phổ năng lượng Fourier, các ma trận đồng hiện, phân tích thành phần chính bất biến - trượt, đặc trưng Tamura, phân rã Wold, trường ngẫu nhiên Markov, mô hình fractal, và lọc đa phân giải như biến đổi Gabor và sóng, mô tả kết cấu bằng phân bố thống kê của cường độ ảnh. Một số biểu diễn kết cấu, được sử dụng thường xuyên và đã được chứng minh là hiệu quả trong tra cứu ảnh dựa theo nội dung.

2.2.1. Ma trận đồng hiện (Co-occurrence Matrix)

Ma trận đồng hiện 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. Dựa trên độ sáng và khoảng cách giữa các điểm ảnh để tạo ma trận đồng hiện. Định nghĩa phần tử $C_{\Delta x, \Delta y}(i, j)$ trong ma trận đồng hiện: Ảnh có $n * m$ điểm ảnh

$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Trong đó, (p, q) : là tọa độ một điểm ảnh trong không gian hai chiều.

(i, j) : là các mức độ sáng của ảnh.

$\Delta x, \Delta y$: là khoảng cách giữa 2 điểm ảnh lân cận nhau.

$$\begin{aligned} \Delta x &= d \times \cos \theta \\ \Delta y &= d \times \sin \theta \end{aligned} \quad d \text{ đủ nhỏ để } p + \Delta x, q + \Delta y \text{ là điểm lân cận của } (p, q).$$

Các đặc trưng kết cấu suy ra từ ma trận đồng hiện đó là:

$$\text{Độ nhiễu: Entropy} = - \sum_i \sum_j C_{ij} \cdot \log C_{ij} \quad (7)$$

$$\text{Năng lượng Energy} = \sum_i \sum_j C_{ij}^2 \quad (8)$$

$$\text{Độ tương phản Contrast} = \sum_i \sum_j |i - j| C_{ij} \quad (9)$$

$$\text{Độ đồng nhất Homogeneity} = \sum_i \sum_j \frac{C_{ij}}{1 + |i - j|} \quad (10)$$

2.2.2. 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.

2.2.2.1. Thô (Coarseness)

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

$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i, j) / 2^{2k} \quad (2.11)$$

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:

$$\begin{aligned} E_{k,h}(x, y) &= |A_k(x+2^{k-1}, y) - A_k(x-2^{k-1}, y)| \\ E_{k,v}(x, y) &= |A_k(x, y+2^{k-1}) - A_k(x, y-2^{k-1})| \end{aligned} \quad (2.12)$$

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:

$$S_{best}(x, y) = 2^k \quad (2.13)$$

Thô được tính bằng trung bình S_{best} trên toàn bộ ảnh đó là:

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i, j) \quad (2.14)$$

Cải tiến của đặc trưng thô có thể thu được bởi sử dụng một lược đồ để mô tả phân bố của S_{best} . 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.

2.2.2.2. Độ tương phản

Công thức cho tương phản là:

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}} \quad (2.15)$$

Trong đó, $\alpha_4 = \mu_4 / \sigma^4$

μ_4 : là mômen thứ tư về trung bình.

σ^2 : phương sai.

2.2.2.3. Hướng

Độ lớn và góc của véc tơ được định nghĩa như sau:

$$\begin{aligned} |\Delta G| &= \sqrt{|\Delta_H|^2 + |\Delta_V|^2} \\ \theta &= \tan^{-1}(\Delta_V / \Delta_H) + \pi/2 \end{aligned} \quad (2.16)$$

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 $|\Delta G|$ lớn hơn một ngưỡng, một lược đồ của θ , biểu thị bằng H_D , 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:

$$F_{dir} = \sum_p \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi) \quad (2.17)$$

Trong đó, p là tổng các phạm vi này trên n_p đỉnh.

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

ϕ_p là bin màu nhận giá trị đỉnh.

2.2.3. Các đặc trưng Wold

Phân rã Wold [7] cung cấp một cách tiếp cận khác để mô tả các kết cấu về mặt các đặc tính nhận thức. Ba thành phần Wold, điều hoà, tạm thời, và vô định, tương ứng với chu kỳ, hướng, và tính ngẫu nhiên của kết cấu tương ứng. Các kết cấu chu kỳ có một thành phần điều hoà mạnh, các kết cấu hướng cao có một thành phần tạm thời mạnh, và các kết cấu được cấu trúc kém hơn có khuynh hướng có một thành phần không xác định mạnh hơn.

Đối với một trường ngẫu nhiên đều $\{y(m,n), m,n \in Z^2\}$, phân rã Wold cho phép trường được phân rã thành ba thành phần trực giao lẫn nhau: đó là thành phần vô định, tiền định và tạm thời. Được định nghĩa:

$$y(m,n) = u(m,n) + d(m,n) = u(m,n) + h(m,n) + e(m,n) \quad (2.18)$$

Trong đó, $u(m,n)$ là thành phần vô định.

$d(m,n)$ là thành phần tiền định.

$e(m,n)$: là thành phần tạm thời.

$h(m,n)$: là thành phần điều hoà.

Thành phần tạm thời có thể được phân rã tiếp thành $h(m,n)$ và $e(m,n)$.

Trong miền tần số ta có:

$$F_y(\xi, \eta) = F_u(\xi, \eta) + F_d(\xi, \eta) = F_u(\xi, \eta) + F_h(\xi, \eta) + F_e(\xi, \eta) \quad (2.19)$$

Trong đó, $F_y(\xi, \eta), F_u(\xi, \eta), F_d(\xi, \eta), F_h(\xi, \eta), F_e(\xi, \eta)$: là các hàm phân bố phổ (SDF) của $\{y(m, n)\}, \{u(m, n)\}, \{d(m, n)\}, \{h(m, n)\}$ và $\{e(m, n)\}$ tương ứng.

Trong miền không gian, ba thành phần trực giao có thể thu được bởi ước lượng khả năng nhất (MLE), nó gồm sự điều chỉnh một quá trình AR bậc cao, cực tiểu hoá hàm giá, và giải một tập các phương trình tuyến tính. Trong miền tần số, các thành phần Wold có thể thu được bởi ngưỡng toàn cục của các độ lớn phổ Fourier của ảnh. Phương pháp sử dụng trích chọn đỉnh điều hoà và mô hình MRSAR thiếu một phân rã thực sự của ảnh được giới thiệu. Phương pháp này được thiết kế để dung sai sự đa dạng của các sự không đồng nhất trong các mẫu kết cấu tự nhiên.

2.2.4. Mô hình tự hồi qui đồng thời SAR

Mô hình SAR là một thể hiện của các mô hình trường ngẫu nhiên Markov - MRF (Markov random field). Mô hình tự hồi qui đồng thời thành công về mô hình kết cấu trong những thập kỷ qua. So với các mô hình MRF, SAR sử dụng ít các tham số hơn. Trong mô hình SAR, các cường độ pixel nhận được bằng các biến ngẫu nhiên. Cường độ $g(x, y)$ tại pixel (x, y) có thể được ước lượng bằng một kết hợp tuyến tính của các giá trị pixel lân cận $g(x', y')$ và một số hạng nhiễu cộng $\sum(x, y)$ đó là:

$$g(x, y) = \mu + \sum_{(x', y') \in D} \theta(x', y') g(x', y') + \varepsilon(x, y) \quad (2.20)$$

Trong đó, μ là giá trị xiên được xác định bởi trung bình của toàn bộ ảnh.

D là tập lân cận của (x, y) .

$\theta(x', y')$ là tập các trọng số được kết hợp với mỗi pixel lân cận.

$\varepsilon(x, y)$ là một biến ngẫu nhiên độc lập Gaussian với trung bình không.

σ^2 : phương sai.

Kỹ thuật sai số bình phương tối thiểu (LSE) hoặc phương pháp ước lượng khả năng nhất (MLE) thường được sử dụng để đánh giá các tham số của mô hình SAR.

Mô hình SAR là không bất biến quay. Để nhận được một mô hình SAR bất biến quay (RISAR), các pixel nằm trên các đường tròn có bán kính khác nhau có tâm tại mỗi pixel (x, y) đáp ứng bằng tập D lân cận của nó. Như thế cường độ $g(x, y)$ tại pixel (x, y) có thể được ước lượng bằng

$$g(x, y) = \mu + \sum_{i=1}^p \theta_i(x, y) l_i(x, y) + \varepsilon(x, y) \quad (2.21)$$

p : là số lân cận tròn.

Để tạo chi phí tính toán thấp và để thu được bất biến quay tại cùng thời điểm, p không được quá lớn hoặc quá nhỏ. Thông thường $p = 2.l(x, y)$ có thể được tính toán bởi:

$$l_i(x, y) = \frac{1}{8i} \sum_{(x', y') \in N_i} w_i(x', y') g(x', y') \quad (2.22)$$

Trong đó, N_i là lân cận tròn thứ i của (x, y) .

$w_i(x', y')$ là một tập các trọng số được tính trước chỉ ra đóng góp của pixel (x', y') trong vòng tròn thứ i .

Để mô tả các kết cấu có các tính chất hệt khác nhau, mô hình tự hồi qui đồng thời đa phân giải MRSAR (*multi-resolution simultaneous auto-regressive*) được đề xuất để cho phép phân tích kết cấu đa mức. Một ảnh được biểu diễn bởi hình chóp Gaussian đa độ phân giải với lọc thông thấp và lấy mẫu dưới được áp dụng tại một số mức liên tiếp. Sau đó hoặc là mô hình SAR hoặc là mô hình RISAR có thể được áp dụng đối với mỗi mức của hình chóp.

MRSAR đã được chứng minh có hiệu năng trên cơ sở dữ liệu kết cấu Brodatz tốt hơn nhiều đặc trưng kết cấu khác, như phân tích thành phần chính, phân rã Wold, và biến đổi sóng.

2.2.5. 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 [4]. 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:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j Wx\right] \quad (2.23)$$

Trong đó, σ_x : là độ lệch chuẩn của các bao Gaussian dọc theo hướng x .

σ_y : 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 $g(x, y)$:

$$\begin{aligned} g_{mn}(x, y) &= a^{-m} g(x', y') \\ x' &= a^{-m} (-x \cos \theta + y \sin \theta) \\ y' &= a^{-m} (-x \sin \theta + y \cos \theta) \end{aligned} \quad (2.24)$$

Trong đó, $a > 1, \theta = n\pi / K, n = 0, 1, \dots, K-1$, và $m = 0, 1, \dots, S-1$.

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:

$$W_{mn}(x, y) = \int I(x, y) g_{mn}^*(x - x_1, y - y_1) dx_1 dy_1 \quad (2.25)$$

Trong đó, *: chỉ ra số liên hợp phức.

μ_{mn} : là trung bình.

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

$f = [\mu_{00}, \sigma_{00}, \dots, \mu_{mn}, \sigma_{mn}, \Lambda, \mu_{S-1K-1}, \sigma_{S-1K-1}]$ có thể được sử dụng để biểu diễn đặc trưng kết cấu của một vùng kết cấu thuần nhất.

2.2.6. Các đặc trưng biến đổi sóng

Tương tự với lọc Gabor, biến đổi sóng [6] cung cấp một cách tiếp cận đa độ phân giải đối với phân tích kết cấu và phân lớp. Các biến đổi sóng phân rã một tín hiệu với một họ các hàm cơ sở $\psi_{mn}(x)$ thu được thông qua dịch chuyển và sự co giãn của sóng $\psi(x)$:

$$\psi_{mn}(x) = 2^{-\frac{1}{2}} \psi(a^{-m} x - n) \quad (2.26)$$

Trong đó, m và n là các tham số co giãn và dịch chuyển.

Một tín hiệu $f(x)$ có thể được biểu diễn đó là:

$$f(x) = \sum_{mn} c_{mn} \psi_{mn}(x) \quad (2.27)$$

Tính toán các biến đổi sóng của một tín hiệu hai chiều gồm lọc đệ quy và lấy mẫu dưới. Tại mỗi mức, tín hiệu được phân rã thành bốn dải tần số con, LL, LH, HL, và HH, ở đây L biểu thị tần số thấp và H biểu thị tần số cao. Hai loại biến đổi sóng chính được sử dụng cho phân tích kết cấu là biến đổi sóng cấu trúc hình chóp PWT (*pyramid-structured wavelet transform*) và biến đổi sóng cấu trúc hình cây TWT (*tree-structured wavelet transform*). PWT phân rã dải LL một cách đệ quy. Tuy nhiên, với một số kết cấu thông tin quan trọng nhất thường xuất hiện trong các kênh tần số chung. Để khắc phục hạn chế này, TWT phân rã các dải khác như LH, HL hoặc HH khi cần.

Sau khi phân rã, các véc tơ đặc trưng có thể được xây dựng sử dụng trung bình và độ lệch chuẩn của phân bố năng lượng của mỗi dải con tại mỗi mức. Với phân rã ba mức, PWT đưa ra một véc tơ đặc trưng có $3 \times 4 \times 2$ thành phần. Với TWT, đặc trưng sẽ phụ thuộc vào dải con nào tại mỗi mức được phân rã. Một cây phân rã cố định có thể thu được bởi phân rã liên tiếp các dải LL, LH, và HL, và vì thế cho ra một véc tơ đặc trưng có 52×2 thành phần. Lưu ý trong ví dụ này, đặc trưng thu được bởi PWT có thể được coi như tập con của đặc trưng thu được bởi TWT. Hơn nữa, theo so sánh của các đặc trưng biến đổi sóng khác nhau, chọn riêng lọc sóng không là then chốt cho phân tích kết cấu.

2.3. Hình dạng

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.

2.3.1. Các bất biến mômen

Biểu diễn hình cô điển sử dụng một tập các bất biến mômen. Nếu đối tượng R được biểu diễn như một ảnh nhị phân, thì các mômen trung tâm bậc $p+q$ cho hình của đối tượng R được định nghĩa:

$$\mu_{p,q} = \sum_{(x,y) \in R} (x-x_c)^p (y-y_c)^q \quad (2.28)$$

Trong đó, (x_c, y_c) là tâm của đối tượng.

Mômen trung tâm này có thể được chuẩn hoá để bất biến tỷ lệ:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma}, \quad \gamma = \frac{p+q+2}{2} \quad \text{②.29}$$

Dựa trên các mômen này, một tập các bất biến mômen đối với dịch chuyển, quay và tỷ lệ có thể tìm thấy trong:

$$\begin{aligned} \phi_1 &= \mu_{2,0} + \mu_{0,2} \\ \phi_2 &= (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2 \\ \phi_3 &= (\mu_{3,0} - 3\mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2 \\ \phi_4 &= (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} + \mu_{2,1})^2 \\ \phi_5 &= (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2}) \left[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2 \right] \\ &\quad + (\mu_{0,3} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1}) \left[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2 \right] \\ \phi_6 &= (\mu_{2,0} - \mu_{0,2}) \left[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1})^2 \right] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1}) \\ \phi_7 &= (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2}) \left[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2 \right] \end{aligned} \quad \text{②.30}$$

2.3.2. Các góc uốn

Chu tuyến (contour) của một đối tượng hai chiều có thể được biểu diễn bằng một dãy đóng đường bao các pixel liên tiếp (x_s, y_s) , ở đây $0 \leq s \leq N-1$ là tổng số các pixel trên đường biên đóng. Hàm xoay hoặc góc xoay $\theta(s)$ đo góc tang ngược chiều kim đồng hồ như một hàm độ dài cung s theo một điểm tham chiếu trên đường biên đóng của đối tượng, có thể được định nghĩa như:

$$\begin{aligned} \theta(s) &= \tan^{-1} \left(\frac{y'_s}{x'_s} \right) \\ y'_s &= \frac{dy_s}{ds} \\ x'_s &= \frac{dx_s}{ds} \end{aligned} \quad \text{②.31}$$

Một vấn đề chính đối với biểu diễn này là nó biến đổi đối với quay của đối tượng và chọn điểm tham chiếu. Nếu chúng ta trượt điểm tham chiếu dọc theo đường bao của đối tượng bởi một lượng t , thì hàm xoay mới trở thành $\theta(s+t)$. Nếu chúng ta quay đối tượng một góc ω thì hàm mới trở thành $\theta(s) + \omega$.

Do đó, để so sánh sự tương tự hình giữa các đối tượng A và B với các hàm xoay của nó, khoảng cách tối thiểu cần được tính toán trên tất cả các trượt t và các quay ω có thể là:

$$d_p(A, B) = \left(\min_{\omega \in \mathbb{R}, t \in [0, 1]} \int_0^1 |\theta_A(s+t) - \theta_B(s) + \omega|^p ds \right)^{\frac{1}{p}} \quad (2.32)$$

Giả thiết rằng mỗi đối tượng đã tỷ lệ lại sao cho tổng độ dài chu vi là 1. Độ đo này là bất biến với dịch chuyển, quay, và thay đổi tỷ lệ.

2.3.3. Các ký hiệu mô tả Fourier

Các ký hiệu mô tả Fourier mô tả hình của một đối tượng với biến đổi Fourier của đường bao của nó. Xét đường biên đóng của một đối tượng hai chiều bằng một dãy đóng các pixel bao liên tiếp (x_s, y_s) ,

Trong đó, $0 \leq s \leq N-1$ và N : là tổng số các pixel trên đường bao.

Độ cong tại một điểm s dọc theo đường biên đóng được định nghĩa như tỷ lệ thay đổi theo hướng \tan của đường biên đóng:

$$K(s) = \frac{d}{ds} \theta(s) \quad (2.33)$$

Trong đó, $K(s)$: là độ cong.

$\theta(s)$: là hàm xoay của đường biên đóng.

Khoảng cách trọng tâm được định nghĩa bằng hàm khoảng cách giữa các pixel bao quanh và trọng tâm (x_c, y_c) của đối tượng:

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \quad (2.34)$$

Toạ độ phức hợp thu được bởi biểu diễn đơn giản các toạ độ của các pixel bao như các số phức hợp:

$$Z(s) = (x_s - x_c) + j(y_s - y_c) \quad (2.35)$$

Các biến đổi Fourier của ba loại biểu diễn đường biên đóng này sinh ra ba tập hệ số phức hợp, biểu diễn hình của một đối tượng trong miền tần số. Các hệ số tần số thấp hơn mô tả đặc tính hình chung, trong khi các hệ số tần số cao phản ánh các chi tiết hình. Để thu được bất biến quay (mã đường biên đóng không liên quan đến chọn điểm tham chiếu), chỉ độ lớn của các hệ số phức hợp được sử dụng và các thành phần pha bị loại bỏ. Để thu được bất biến tỷ lệ, độ lớn của các hệ số được chia bằng độ lớn của thành phần DC hoặc hệ số khác không đầu tiên. Bất biến dịch chuyển thu được trực tiếp từ biểu diễn đường biên đóng.

Các ký hiệu mô tả Fourier của đường cong là:

$$f_K = \left[|F_1|, |F_2|, \dots, |F_{M/2}| \right] \quad (36)$$

Ký hiệu mô tả Fourier của khoảng cách trọng tâm là:

$$f_R = \left[\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_{M/2}|}{|F_0|} \right] \quad (37)$$

Trong đó, F_i trong (31) và (32) biểu thị thành phần thứ i của các hệ số biến đổi Fourier.

Các biến đổi Fourier của nó có tính đối xứng hay $|F_{-i}| = |F_i|$.

Ký hiệu mô tả Fourier của tọa độ phức hợp là:

$$f_Z = \left[\frac{|F_{-(M/2-1)}|}{|F_1|}, \dots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \dots, \frac{|F_{M/2}|}{|F_1|} \right] \quad (38)$$

Trong đó, F_1 là thành phần tần số khác không đầu tiên được sử dụng để chuẩn hoá các hệ số biến đổi.

Cả hai thành phần tần số dương và âm được xem xét. Hệ số DC là phụ thuộc vào vị trí của hình và nên bị loại bỏ.

Để đảm bảo các đặc trưng hình kết quả của tất cả các đối tượng trong một cơ sở dữ liệu có cùng độ dài, đường bao quanh $((x_s, y_s), 0 \leq s \leq N-1)$ của mỗi đối tượng được lấy mẫu lại với M mẫu trước khi thực hiện biến đổi Fourier.

2.3.4. Hình tròn, độ lệch tâm, và hướng trục chính

Hình tròn được định nghĩa là:

$$\alpha = \frac{4\pi S}{P^2} \quad (39)$$

Trong đó, S là cỡ.

P là chu vi của một đối tượng.

Hướng trục chính có thể được định nghĩa như hướng của vectơ riêng lớn nhất của ma trận bậc hai của một vùng hoặc một đối tượng. Độ lệch tâm có thể được định nghĩa như tỷ lệ của giá trị riêng nhỏ nhất với giá trị riêng lớn nhất.

2.4. Thông tin không gian

Các vùng hoặc đối tượng với các đặc tính màu và kết cấu tương tự có thể được phân biệt dễ dàng bằng việc tận dụng các ràng buộc không gian. Ví dụ, các vùng bầu trời màu xanh và biển xanh có thể có các lược đồ màu tương tự, nhưng các vị trí không gian của chúng trong các ảnh là khác nhau. Do đó, vị trí không gian của các vùng (hoặc các đối tượng) hoặc quan hệ không gian giữa nhiều vùng (hoặc đối tượng) trong một ảnh thì rất hữu ích cho tìm kiếm các ảnh.

Các thao tác như giao và chồng được sử dụng. Bố cục màu kết hợp thông tin không gian với thông tin màu xuất hiện trong ảnh và tạo ra một đặc trưng rất quan trọng trong quá trình tra cứu, gọi là đặc trưng màu - không gian.

Tuy nhiên, tìm kiếm các ảnh dựa trên các quan hệ không gian của các vùng còn lại một vấn đề nghiên cứu khó trong tra cứu ảnh dựa vào nội dung, do phân đoạn tin cậy của các đối tượng hoặc các vùng thường là không khả thi ngoại trừ các ứng dụng rất giới hạn. Mặc dù một số hệ thống chia các ảnh thành các khối đều, chỉ thu được sự thành công khiêm tốn với các lược đồ chia không gian như thế do hầu hết các ảnh tự nhiên không thể ép thành các khối con đều về không gian. Để giải quyết vấn đề này, một phương pháp dựa vào biến đổi radon, tận dụng phân bố không gian của các đặc trưng trực quan không cần phân đoạn phức tạp được đề xuất trong.

2.5. Phân đoạn

Phân đoạn là quá trình phân ảnh ra thành các vùng mà về lý tưởng nó sẽ tương ứng với các đối tượng xuất hiện trong ảnh. Đây là bước rất quan trọng đối với tra cứu ảnh. Cả đặc trưng hình và đặc trưng bố cục phụ thuộc vào phân đoạn tốt. Trong phần này chúng tôi sẽ mô tả một số kỹ thuật phân đoạn đã có được sử dụng trong cả thị giác máy và tra cứu ảnh.

Một ưu điểm chính của các thuật toán phân đoạn loại này là nó trích chọn các đường bao quanh từ một số lượng lớn các ảnh mà không chiếm thời gian và nỗ lực của con người. Tuy nhiên, trong một lĩnh vực tự nhiên, với các ảnh không có điều kiện tiên quyết, phân đoạn tự động không luôn luôn tin cậy. Một thuật toán có thể phân đoạn trong trường hợp này chỉ là các vùng, mà không là các đối tượng. Để thu được các đối tượng mức cao, nó cần có sự trợ giúp của con người.

Với các đặc trưng hình, phân đoạn chính xác là mong muốn cao trong khi các đặc trưng bố cục, một phân đoạn thô có thể là đủ.

2.6. Độ đo

2.6.1. Khái niệm

Độ đo tương tự là một trong những phương pháp tốt để máy tính phân biệt được các hình ảnh qua nội dung của chúng. Thông thường hệ thống tra cứu ảnh theo nội dung sẽ truy vấn hình ảnh bằng phương pháp đo tương tự dựa trên các chức năng, việc xác định nó có thể dưới nhiều hình thức như phát hiện biên, màu sắc, vị trí điểm ảnh... các phương pháp như histogram, màu sắc và phân tích histogram dòng cột sử dụng biểu đồ để xác định độ tương tự.

Giả sử D : là hàm khoảng cách.

k, l, m : là các đối tượng.

Thì D cần đáp ứng các tiên đề sau:

- $D(k, l) \geq 0 \Leftrightarrow k \equiv l$
- $D(k, l) \geq 0$ khoảng cách là số dương
- $D(k, l) = D(l, k)$ khoảng cách có tính đối xứng
- $D(k, m) \leq D(k, l) + D(l, m)$ bất đẳng thức tam giác

Do đó, độ đo có ý nghĩa quan trọng trong tìm kiếm ảnh dựa theo nội dung. Độ đo mang ý nghĩa quyết định kết quả tìm kiếm sẽ như thế nào, mức độ chính xác ra sao.

2.6.2. Một số độ đo thông dụng

2.6.2.1. Khoảng cách Minkowsky:

$$D(k, l) = \left[\sum_{i=1}^d |x_{ik} - x_{il}|^s \right]^{\frac{1}{s}} \quad (4.40)$$

Trong đó, $D(k, l)$: là khoảng cách từ đối tượng k đến đối tượng l .

d : là số chiều của không gian.

x_{ik} : là tọa độ thứ i của đối tượng k .

x_{il} : là đối tượng i của đối tượng l

2.6.2.2. Khoảng cách toàn phương

$$Intersection(k, l) = \sum_{i=1}^K \sum_{j=1}^K [G_h(a_{ij}) | G_h(i) | G_h(j)] \quad (4.41)$$

2.6.2.3. Khoảng cách Euclid:

Đây là cách tính khoảng cách Euclid thông thường giữa các K bin:

$$Intersection(A, B) = \sum_{j=1}^K \sqrt{(A_j - B_j)^2} \quad (4.42)$$

2.6.2.4. Độ đo khoảng cách min-max

Được thực hiện trên ý tưởng lấy phần giao của hai lược đồ màu cần so sánh, ta sẽ được một lược đồ màu, tính tổng các giá trị có được từ lược đồ này sẽ được độ đo min-max. Khoảng cách min-max thể hiện sự tương tự giữa hai lược đồ màu. Ta có:

$$D(A, B) = \sum_{i=1}^d \min(x_{ik}, x_{il}) \quad (4.43)$$

Chương 3: KỸ THUẬT TRA CỨU ẢNH DỰA THEO NỘI DUNG

3.1. Màu sắc

3.1.1. Lược đồ màu

Lược đồ màu thể hiện màu sắc của ảnh. Màu sắc là thông tin gắn liền với con người nhất. Do đó, việc tìm kiếm theo lược đồ màu sẽ thân thiện với người dùng hơn. Dựa vào lược đồ màu có thể giúp cho người dùng tìm kiếm những bức ảnh có sự giống nhau về màu sắc.

Độ đo tính tương tự giữa màu sắc của lược đồ màu và của ảnh truy vấn $H(Q)$ và lược đồ màu của ảnh trong cơ sở dữ liệu ảnh $H(D)$:

$$D(Q, I_D) = \frac{\sum_{j=1}^M \min(H(Q, j), H(D, j))}{\sum_{j=1}^M H(D, j)} \quad (3.1)$$

Trong đó, M : tổng số bin màu.

3.1.2. Vector liên kết màu (Color Coherence Vector)

Cũng giống như lược đồ màu đặc trưng vector liên kết màu cũng thể hiện màu sắc của ảnh. Nó thể hiện rõ ràng mật độ phân bố màu trong ảnh. Với hai ảnh có thể rất giống nhau về lược đồ màu nhưng do khác nhau về sự phân bố màu sắc, nên trong quá trình tra cứu theo lược đồ màu có thể cho ra nhiều ảnh thừa. Nếu trong quá trình tra cứu mà sử dụng vector liên kết màu thì có thể khắc phục được tình trạng trên.

Với mỗi ô màu, 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:

$$V_c = [(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_n, \beta_n)] \quad (3.2)$$

Trong đó, n là số ô màu.

Độ đo tương tự giữa hai ảnh dựa trên đặc trưng vector liên kết màu:

$$D_c(Q, I_D) = \sum_{j=1}^n (|\alpha_{Q_j} - \alpha_{D_j}| + |\beta_{Q_j} - \beta_{D_j}|) \quad (3.3)$$

3.1.3. Đặc trưng tự tương quan màu (AutoCorrelogram)

Cũng giống như vector liên kết màu, đặc trưng tương quan màu cũng thể hiện sự phân bố màu sắc của ảnh. Nó thể hiện rõ ràng mật độ phân bố màu trong ảnh hơn. Đặc trưng này cũng nhằm giải quyết việc dư thừa trong tra cứu ảnh, giúp tìm những bức ảnh có ngữ nghĩa gần với ảnh truy vấn hơn nhờ các tính chất sau:

- Bao gồm các thông tin về sự tương quan về mặt không gian các màu
- Có thể được dùng để mô tả sự phân bố toàn cục của mối quan hệ không gian cục bộ giữa các màu.
- Kích thước của vector đặc trưng có dung lượng lưu trữ nhỏ

Độ đo tính tương tự về màu sắc giữa đặc trưng tương quan màu của ảnh truy vấn I_Q và ảnh trong cơ sở dữ liệu I_I :

$$D(I_Q, I_I) = \sum_d \sum_{i=1}^N |\alpha_{I_Q}^d - \alpha_{I_I}^d| \quad (3.4)$$

Trong đó, $d = 1, 3, 5, \dots, 2k+1$

N : là tổng bin màu.

3.2. Độ đo khoảng cách giữa các lược đồ màu

3.2.1. Khoảng cách dạng Minkowsky

Độ đo này chỉ so sánh các bin giống nhau giữa các lược đồ màu (xem hình 3.1) và được xác định:

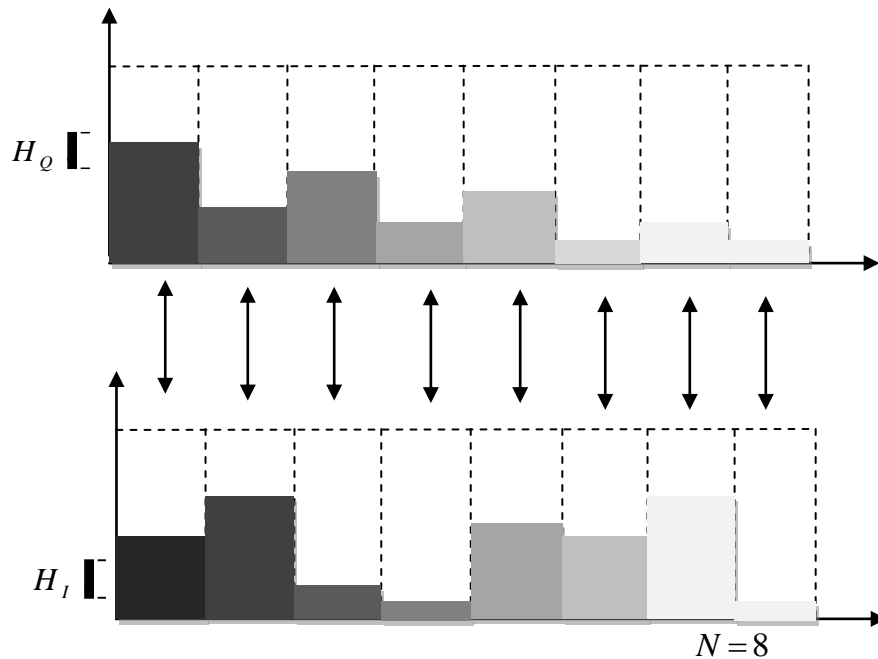
$$d(Q, I) = \sum_{i=1}^N |H_{Q_i} - H_{I_i}| \quad (3.5)$$

Trong đó, Q và I là hai ảnh.

N là số các bin trong lược đồ màu

H_{Q_i} là giá trị bin i trong lược đồ màu H_Q

H_{I_i} là giá trị bin i trong lược đồ màu H_I



Hình 3.1: Khoảng cách dạng Minkowsky.

3.2.2. Khoảng cách toàn phương

Độ đo này không chỉ so sánh các bin giống nhau mà so sánh nhiều bin giữa các lược đồ màu (xem hình 3.2) và được xác định:

$$d(Q, I) = \sum_{i=1}^N |H_Q - H_I| A |H_Q - H_I| \quad (3.6)$$

Trong đó, Q và I là hai ảnh.

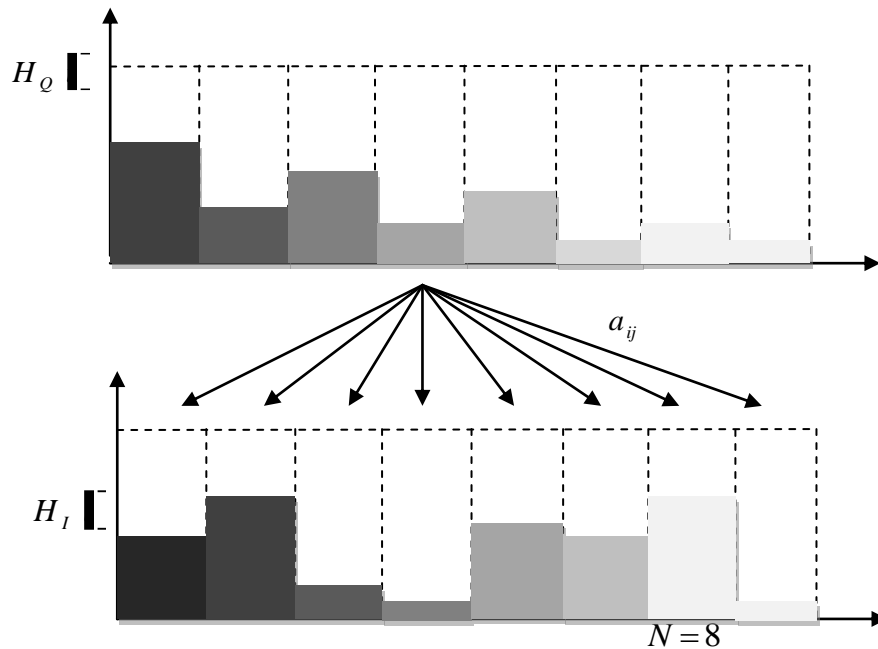
H_Q là lược đồ màu của ảnh Q.

H_I là lược đồ màu của ảnh I.

$A = [a_{i,j}]$ là ma trận $N \times N$.

N là số các bin trong lược đồ màu.

$a_{i,j}$ biểu thị sự tương tự giữa màu i và màu j.



Hình 3.2: Khoảng cách toàn phương.

3.2.3. Độ đo khoảng cách min-max

Được thực hiện lấy phần giao của hai lược đồ cần so sánh, ta sẽ được một lược đồ. Tính tổng các giá trị có được từ lược đồ này cho ta được độ min-max. Khoảng cách min-max thể hiện sự tương tự giữa hai lược đồ. Công thức:

Đối với độ đo min: ta tính dựa vào giá trị min tại mỗi K bin màu

$$Intersection(Q, h) = \sum_{j=1}^K \min(h_j, h_j) \quad (6.7)$$

Đối với độ đo max: ta tính dựa vào giá trị max tại mỗi K bin màu

$$Intersection(Q, h) = \sum_{j=1}^K \max(h_j, h_j) \quad (6.8)$$

$$Matching(Q, h) = \frac{Intersection(Q, h)}{\max\left(\sum_i h_i, \sum_i h_i\right)} \quad (6.9)$$

3.3. Kỹ thuật dựa vào đặc trưng màu

Có hai kỹ thuật thường được sử dụng trong tra cứu ảnh theo nội dung dựa trên đặc trưng màu đó là lược đồ màu toàn cục và lược đồ màu cục bộ.

3.3.1. Lược đồ màu toàn cục

Lược đồ màu toàn cục mô tả sự phân bố màu sử dụng tập các bin màu. Sử dụng lược đồ màu toàn cục, một ảnh sẽ được mã hóa với lược đồ màu của nó và khoảng cách giữa hai ảnh sẽ được xác định bởi khoảng cách giữa những lược đồ màu của chúng. Với kỹ thuật này ta có thể sử dụng các độ đo khác nhau để tính toán khoảng cách giữa hai lược đồ màu. Ví dụ:

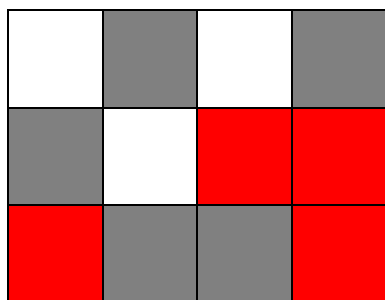


Image A

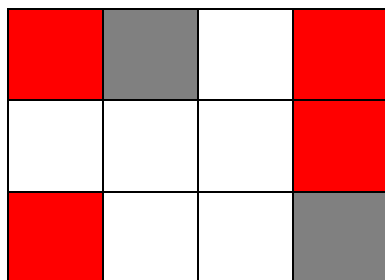
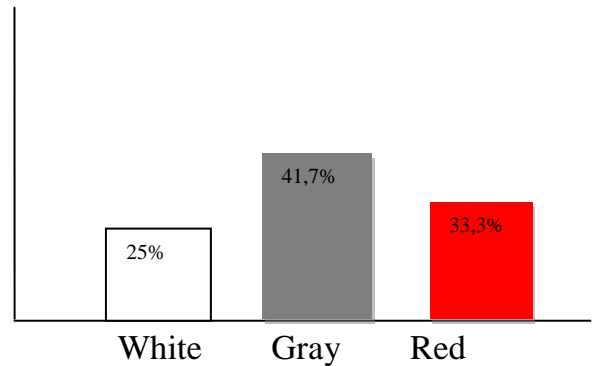


Image B

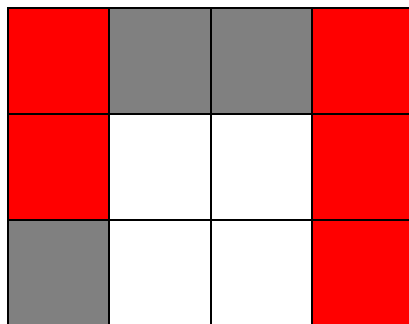
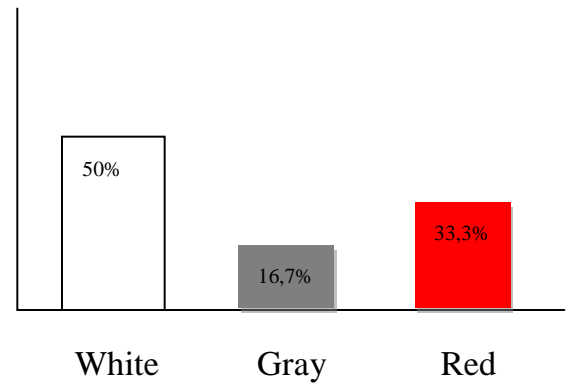
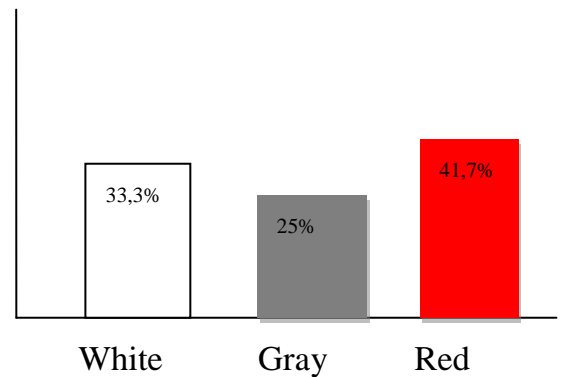


Image C



Hình 3.3 : Ba ảnh và biểu đồ màu tương ứng.

Trong biểu đồ có ba màu: White, Gray and Red. Ta có kí hiệu lược đồ màu như sau:

Image A { 25%, 41.7%, 33.3% }.

Image B { 50%, 16.7%, 33.3% }.

Image C { 33.3%, 25%, 41.7% }

Ta có độ đo khoảng cách giữa hai ảnh A và B (sử dụng độ đo khoảng cách Euclid) cho lược đồ màu toàn cục là:

$$d(A, B) = \sqrt{(0.25 - 0.5)^2 + (0.417 - 0.167)^2 + (0.333 - 0.333)^2} = 0.829$$

$$d(A, C) = 0.203 \quad \text{và} \quad d(B, C) = 0.204$$

Lược đồ màu toàn cục là phương pháp truyền thống cho việc tra cứu ảnh dựa theo màu sắc. Mặt khác, nó không chứa các thông tin liên quan đến sự phân bố màu của các vùng. Do vậy, khoảng cách giữa các ảnh đôi khi chưa thể chỉ ra được sự khác nhau thực sự giữa chúng. Điều này chính là hạn chế của lược đồ màu toàn cục.

3.3.2. Lược đồ màu cục bộ

Phương pháp lược đồ màu cục bộ bao gồm thông tin liên quan đến sự phân bố màu của các vùng. Trước tiên nó phân đoạn ảnh thành nhiều khối và sau đó biểu diễn biểu đồ màu cho mỗi khối, mỗi ảnh sẽ được biểu diễn bởi những biểu đồ này. Khi so sánh hai ảnh, khoảng cách được tính toán bằng cách sử dụng những biểu đồ của chúng giữa một vùng trong 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. Khi sử dụng căn bậc hai của độ đo khoảng cách Euclid để tính toán khoảng cách biểu đồ thì khoảng cách giữa hai ảnh Q và I cho biểu đồ cục bộ là:

$$d(Q, I) = \sum_{k=1}^m \sqrt{\sum_{i=1}^n (H_Q^k - H_I^k)^2} \quad (10)$$

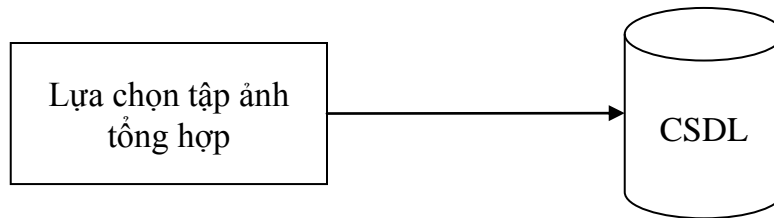
Trong đó, m : là số vùng được phân đoạn trong ảnh.

n : là số mức trong biểu đồ màu.

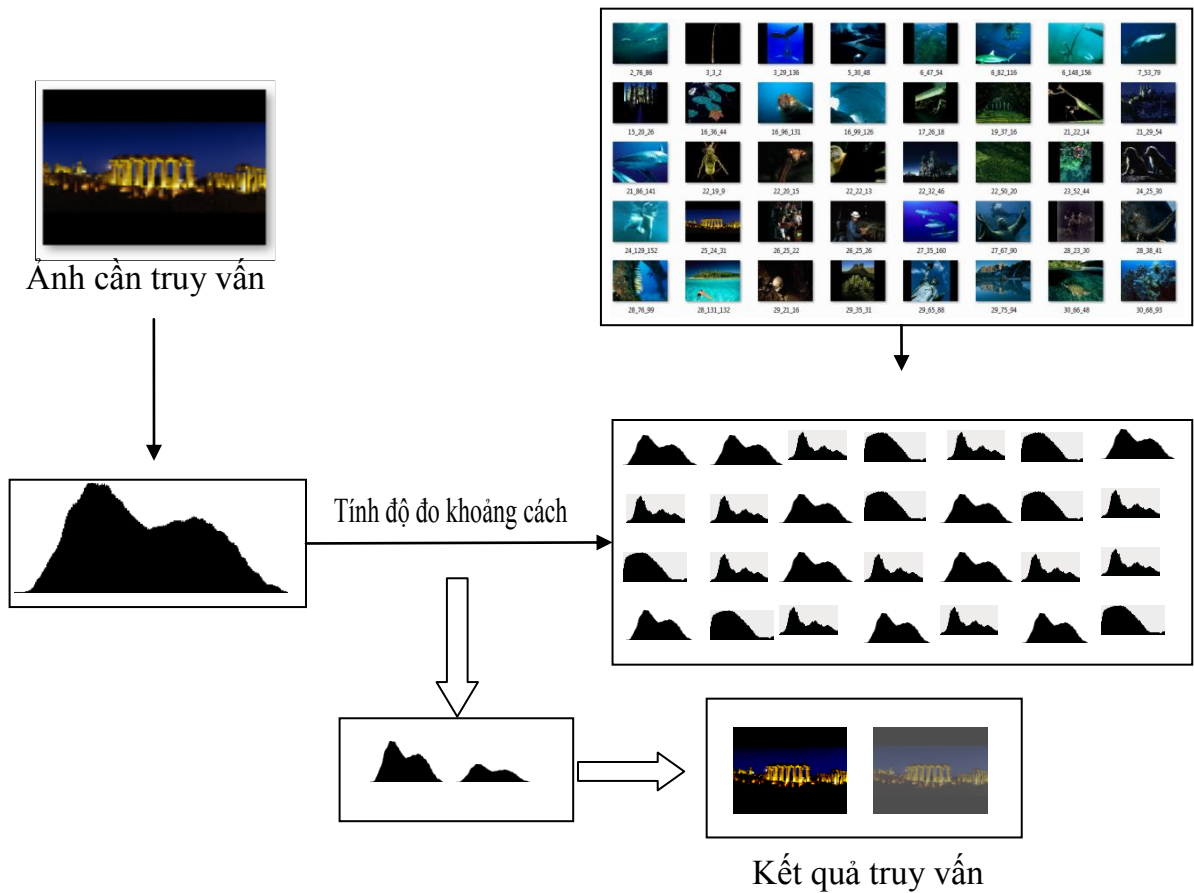
H_i : là giá trị của mức i trong biểu đồ màu đại diện cho vùng k của ảnh.

Chương 4: TRIỂN KHAI & THỰC NGHIỆM

4.1. Quy trình xây dựng phần mềm tra cứu ảnh tổng hợp

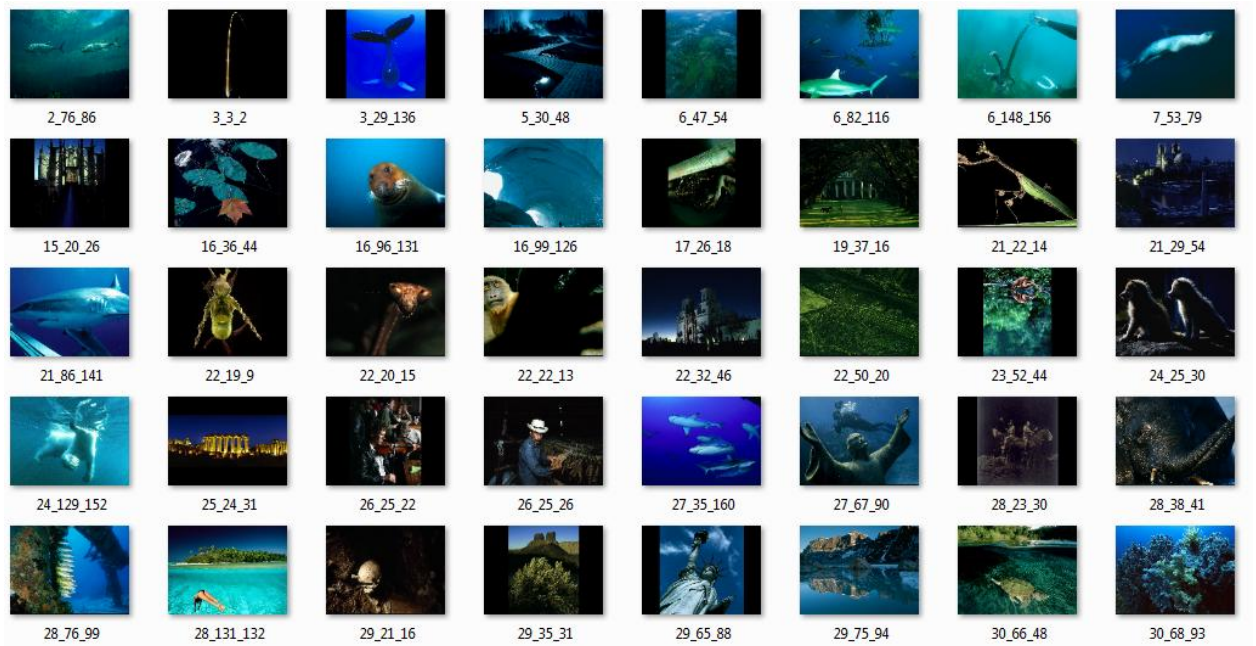


Hình 4.1: Quy trình huấn luyện tập dữ liệu ảnh ban đầu



Hình 4.2 Mô hình của ứng dụng

4.2. Lựa chọn tập mẫu



Hình 4.3: Tập ảnh cơ sở dữ liệu

- Tập ảnh mẫu mà em sử dụng gồm 1.491 ảnh thuộc các thể loại:
 - Các kỳ quan
 - Động vật hoang dã
 - Phong cảnh tự nhiên
 - Dân cư một số vùng ở Châu Phi
 - ...

Được sưu tầm từ trang: <http://www.fuzzywobble.com>

4.3. Lựa chọn phương pháp truy vấn ảnh

Sau khi nghiên cứu các phương pháp truy vấn ảnh theo nội dung (đã trình bày ở chương 3), em nhận thấy phương pháp tra cứu ảnh theo nội dung dựa trên đặc trưng về màu sắc phù hợp với chúng em vì các lý do:

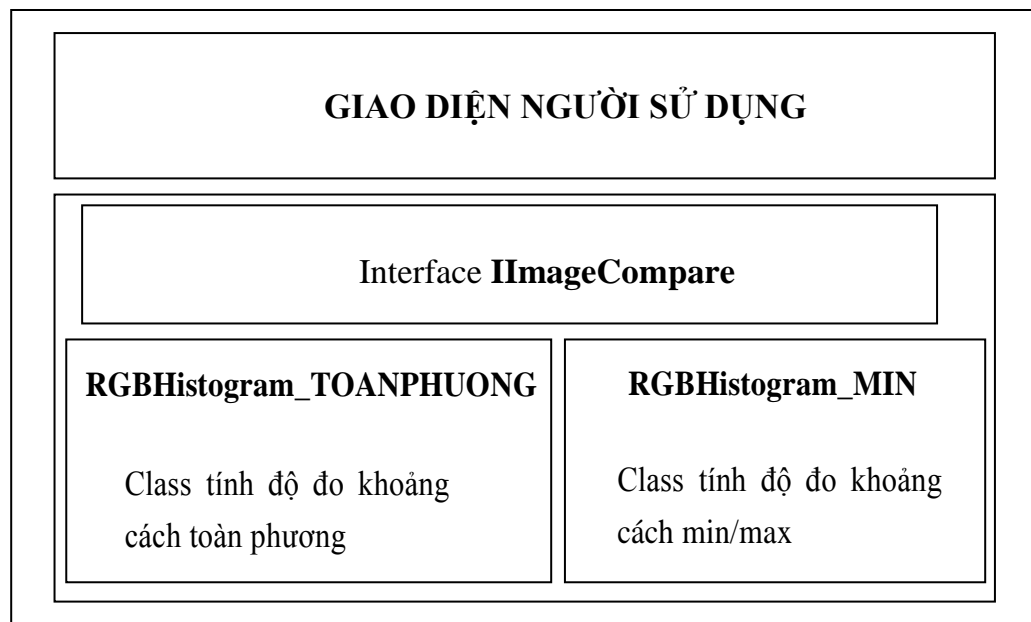
- Dễ tiếp cận
- Dễ cài đặt
- Kết quả tương đối tốt trong một số trường hợp nhất định

Các độ đo lược đồ màu mà em sử dụng là:

- Độ đo khoảng cách toàn phương
- Độ đo khoảng cách min-max

Trong đó, với độ đo khoảng cách toàn phương, em sử dụng thư viện xử lý ảnh OpenCV – với wrapper cho C#.NET là EmguCV để thao tác một số hàm như Tính lược đồ màu, nạp ảnh ... Với độ đo khoảng cách min-max, em không sử dụng thư viện bên ngoài.

4.4. Xây dựng ứng dụng



Hình 4.4: Kiến trúc của ứng dụng

- Interface **IImageCompare**: Là giao diện định nghĩa ra các phương thức cần triển khai chung cho các lớp sử dụng để so sánh các lược đồ màu.

Phương thức cần triển khai là:

- o double **GetSimilarity**(Bitmap a, Bitmap b): Phương thức truyền vào 2 lớp Bitmap (là 2 ảnh cần so sánh). Kết quả trả về từ [0-1] là giá trị độ tương tự giữa 2 ảnh.
- Xây dựng một struct RGBHistogram để lưu trữ các thông tin về lược đồ màu.
- Triển khai 2 lớp: **RGBHistogram_ToanPhuong** và **RGBHistogram_MIN** kế thừa từ IImageCompare để thực thi việc so sánh lược đồ màu nhằm thực hiện mục tiêu bài toán.

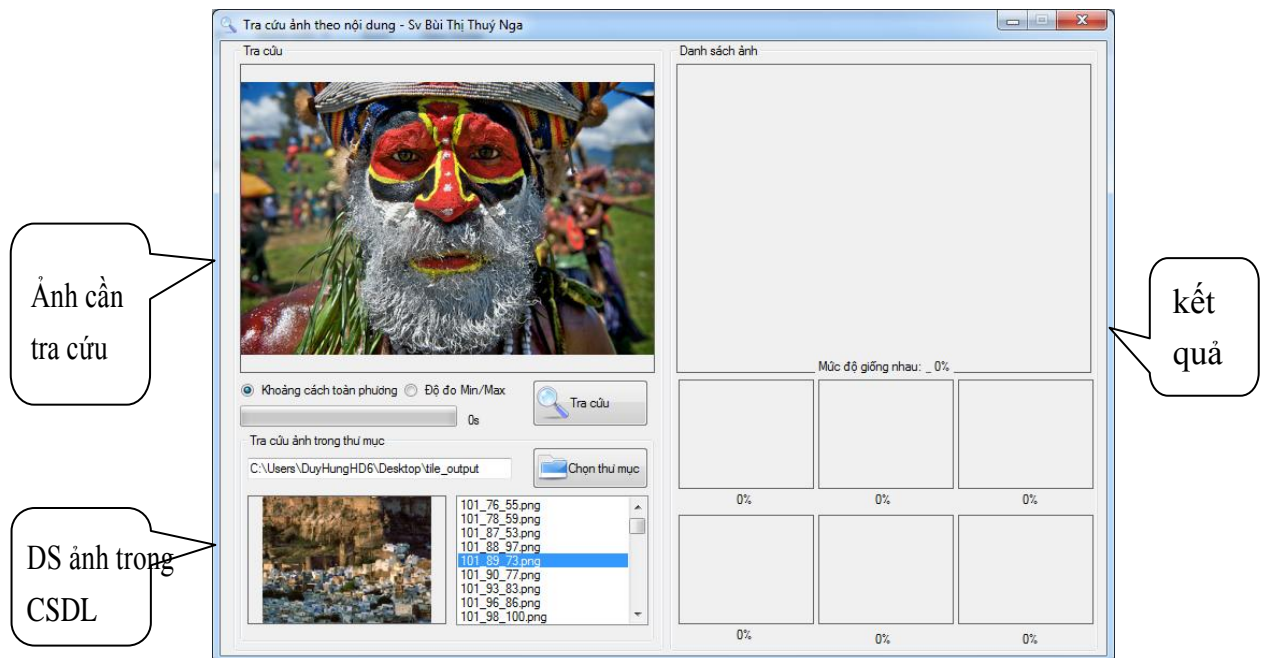
- Xây dựng các phương thức xử lý giao diện, nạp ảnh, hiện thị kết quả ...

Môi trường triển khai:

- Phần cứng: Cấu hình tối thiểu để cài đặt .NET Framework 3.5
- Hệ điều hành: Windows XP, Windows 7
- Môi trường cần: .NET Framework 3.5

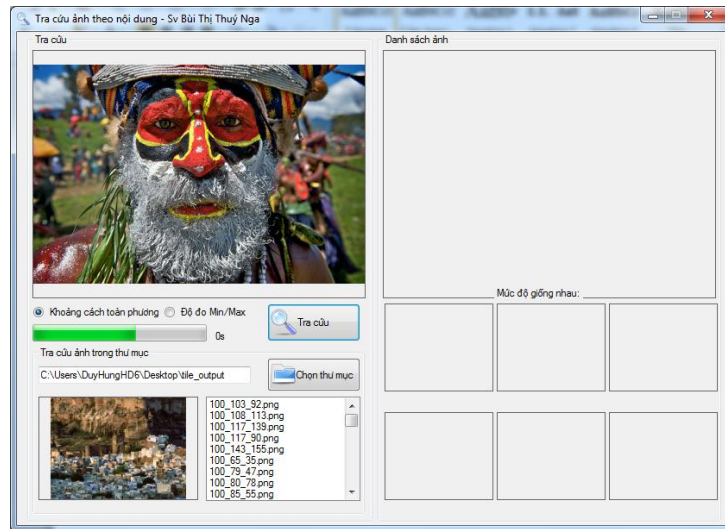
4.5. Kết quả

- Giao diện ứng dụng:

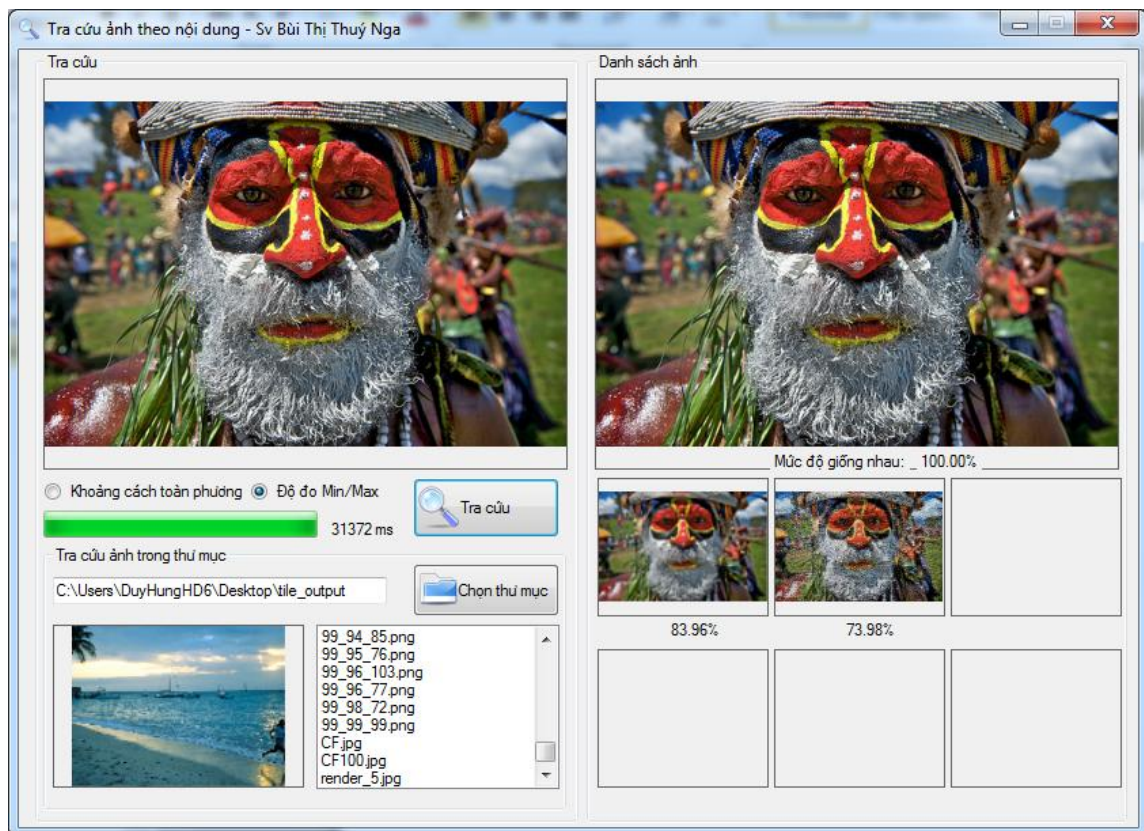


Hình 4.5: Giao diện ứng dụng

- Kết quả thử nghiệm

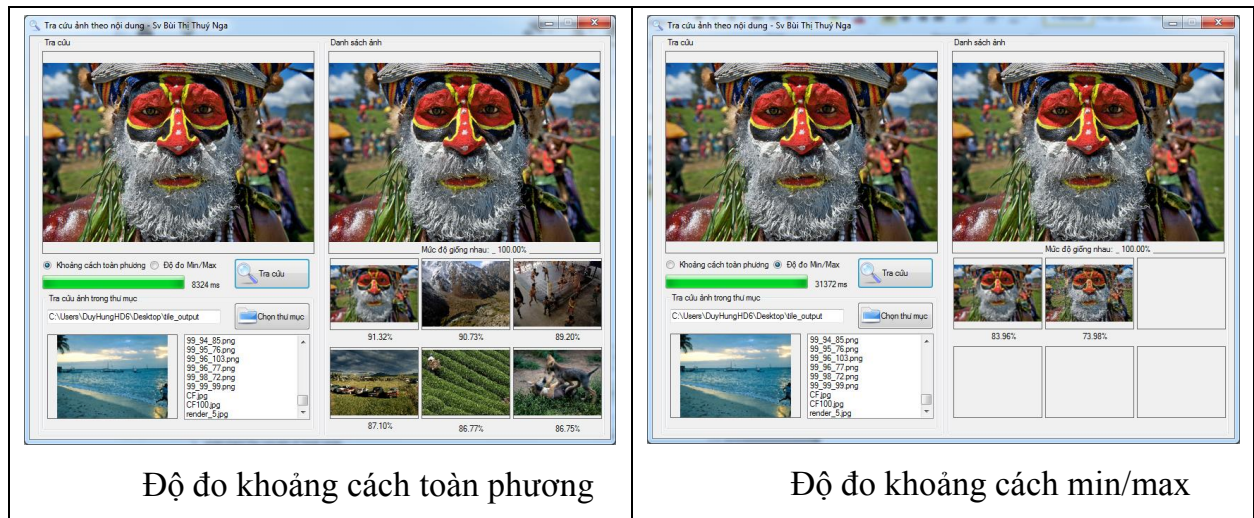


Hình 4.6: Tiến trình tra cứu



Hình 4.7: Kết quả tra cứu

- Thực nghiệm và so sánh giữa kết quả khi sử dụng độ đo khoảng cách min/max và độ đo khoảng cách toàn phương.



Hình 4.8: So sánh giữa hai kỹ thuật sử dụng

Theo kết quả chủ quan, thực nghiệm từ 21 ví dụ. Chúng em nhận thấy, độ đo khoảng cách toàn phương cho phép tìm ảnh tương tự tốt hơn. Nhưng lại tìm ảnh chính xác kém hơn. Nghĩa là nếu có nhiều ảnh giống nhau, chỉ khác nhau về kích thước, độ sáng thì kết quả của kỹ thuật sử dụng độ đo khoảng cách toàn phương cho kết quả kém chính xác hơn.

	ĐỘ ĐO KHOẢNG CÁCH TOÀN PHƯƠNG	ĐỘ ĐO KHOẢNG CÁCH MIN/MAX
Tìm ảnh có đặc trưng màu khá tương tự nhau	Tốt hơn	
Tìm những ảnh chỉ khác nhau về kích thước, độ sáng...		Tốt hơn
Độ phức tạp thuật toán	Cao hơn	
Tốc độ thực nghiệm trong ứng dụng	Nhanh hơn (vì sử dụng thư viện OpenCV với kiểu dữ liệu con trỏ mạnh mẽ)	Chậm hơn vì em chưa tận dụng được các kỹ thuật con trỏ trong C#

KẾT LUẬN

❖ **Đánh giá**

Bài đồ án đã trình bày được một số kỹ thuật nền tảng của hệ thống tra cứu ảnh theo nội dung. Bao gồm, mô tả nội dung trực quan, đánh giá độ tương tự, sơ đồ đánh chỉ số, tương tác với người sử dụng và đánh giá hiệu năng hệ thống. Trong đó, nhấn mạnh vào kỹ thuật vào các kỹ thuật mô tả các đặc trưng trực quan.

❖ **Những mục tiêu đã đạt được**

- Giới thiệu tổng quan về tra cứu ảnh theo nội dung
- Áp dụng phương pháp tra cứu ảnh theo nội dung vào bài toán tra cứu ảnh theo đặc trưng màu.

❖ **Những mục tiêu không khả thi**

- Chưa có chức năng tra cứu ảnh bằng cách kết hợp nhiều đặc điểm.
- Chưa cài đặt được kỹ thuật đánh chỉ số.

❖ **Hướng nghiên cứu trong tương lai**

- Cài đặt kỹ thuật truy vấn dựa trên đặc trưng kết cấu Taumura, lọc Garbor.
- Xây dựng ứng dụng web đáp ứng yêu cầu bài toán.

TÀI LIỆU THAM KHẢO

- [1] Guttman, A. (1984). R-Tree: A dynamic index structure for spatial searching. In Proc. of ACM SIGMOD.
- [2] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by image and video content: The QBIC project. IEEE Computer, 28(9).
- [3] G. Pass, and R. Zabith, "Histogram refinement for content-based image retrieval," IEEE Workshop on Applications of Computer Vision, pp. 96-102, 1996.
- [4] A.K. Jain, and F. Farroknia, "Unsupervised texture segmentation using Gabor filters," Pattern Recognition, Vo.24, No.12, pp. 1167-1186, 1991.
- [5] Beckmann, N., Kriegel, H.-P., Schneider, R., and Seeger, B. (1990). The R*Tree: An efficient and robust access method for points and rectangles. In Proc. of ACM SIGMOD.
- [6] Daubechies, "The wavelet transform, time-frequency localization and signal analysis," IEEE Trans. on Information Theory, Vol. 36, pp. 961-1005, Sept. 1990.
- [7] J. M. Francos. "Orthogonal decompositions of 2D random fields and their applications in 2D spectral estimation," N. K. Bose and C. R. Rao, editors, *Signal Processing and its Application*, pp.20-227. North Holland, 1993.
- [8] Jones, K. S. and Willett, P. (1977). Reading in Information Retrieval. Morgan Kaufmann Pub. Inc.